Can AI Serve as a Substitute for Human Subjects in Software Engineering Research?

Marco Gerosa[1*], Bianca Trinkenreich[2], Igor Steinmacher[1], Anita Sarma[2]

[1*]Northern Arizona University, USA.
[2]Oregon State University, USA.

*Corresponding author(s). E-mail(s): Marco.Gerosa@nau.edu;
Contributing authors: Bianca.Trinkenreich@oregonstate.edu; Igor.Steinmacher@nau.edu;
Anita.Sarma@oregonstate.edu;

Abstract

Research within sociotechnical domains, such as Software Engineering, fundamentally requires a thorough consideration of the human perspective. However, traditional qualitative data collection methods suffer from challenges related to scale, labor intensity, and the increasing difficulty of participant recruitment. This vision paper proposes a novel approach to qualitative data collection in software engineering research by harnessing the capabilities of artificial intelligence (AI), especially large language models (LLMs) like ChatGPT. We explore the potential of AI-generated synthetic text as an alternative source of qualitative data, by discussing how LLMs can replicate human responses and behaviors in research settings. We examine the application of AI in automating data collection across various methodologies, including persona-based prompting for interviews, multi-persona dialogue for focus groups, and mega-persona responses for surveys. Additionally, we discuss the prospective development of new foundation models aimed at emulating human behavior in observational studies and user evaluations. By simulating human interaction and feedback, these AI models could offer scalable and efficient means of data generation, while providing insights into human attitudes, experiences, and performance. We discuss several open problems and research opportunities to implement this vision and conclude that while AI could augment aspects of data gathering in software engineering research, it cannot replace the nuanced, empathetic understanding inherent in human subjects in some cases, and an integrated approach where both AI and human-generated data coexist will likely yield the most effective outcomes.

Keywords: qualitative research, software engineering, foundation models, large language models

1 Introduction

Software engineering is inherently a sociotechnical discipline [1, 2]. Considering the human perspective in research ensures that technological advancements are not developed in a vacuum but are informed by the nuanced needs and complexities of those affected by the software. Therefore, it is not surprising that a large number of articles in top-tier conferences include a qualitative component based on data collected from human subjects through interviews (e.g., [3]), focus groups (e.g., [4]), surveys (e.g., [5]), observation (e.g., [6]), user studies (e.g.,

[7]), etc. Nevertheless, the recruitment and engagement of human participants, particularly from underrepresented groups, pose increasing challenges [8, 9], and the traditional processes of data collection and analysis are often labor-intensive and not easily scalable.

Large language models (LLMs) have the potential to revolutionize human factors research. LLMs have been explored to help qualitative data analysis by processing large amounts of text and identifying patterns and categories [10–12]. For instance, an LLM can quickly sift through thousands of pages of interview transcripts, observational notes, or social media posts, extracting and

categorizing key phrases, sentiments, and topics. The model can recognize recurring topics in the data, which could signal underlying patterns or categories that are pertinent to the research questions.

But, what if we push the boundaries further? What if large language models and other foundational AI models could be harnessed to substitute human subjects in qualitative research? These advanced AI systems, trained with vast amounts of data, can generate responses that closely mimic human-generated content. Embedded within these models are intricate patterns that reveal themselves in the content they generate in response to prompts. Changes in these prompts can dramatically alter the response's nature, affecting not just the conveyed information but also the structure, style, and diction of the generated text. Therefore, with strategic prompting, a foundation model can potentially simulate a particular demographic profile. Researchers could then engage with these models, posing questions and instructions to elicit responses and behaviors that could serve as qualitative data. This data could then be analyzed to develop theories, model user behavior, and even validate automated tools in software engineering, offering a novel dimension to the traditional qualitative research landscape.

The exploration of this idea within this paper is conceptual and forward-looking, intended to provoke thought, reflection, and discourse within the community. It is an invitation to dialogue about the potential roles AI could play in helping collect data for qualitative software engineering research. Before we proceed—*a disclaimer—we neither believe nor desire for AI to completely replace human subjects in software engineering research.* The purpose of this vision paper is to explore the concept and consider the possible supportive roles of AI. Looking ahead, the research community will establish practices to determine the optimal balance between AI-generated synthetic text and human-sourced data in qualitative software engineering research.

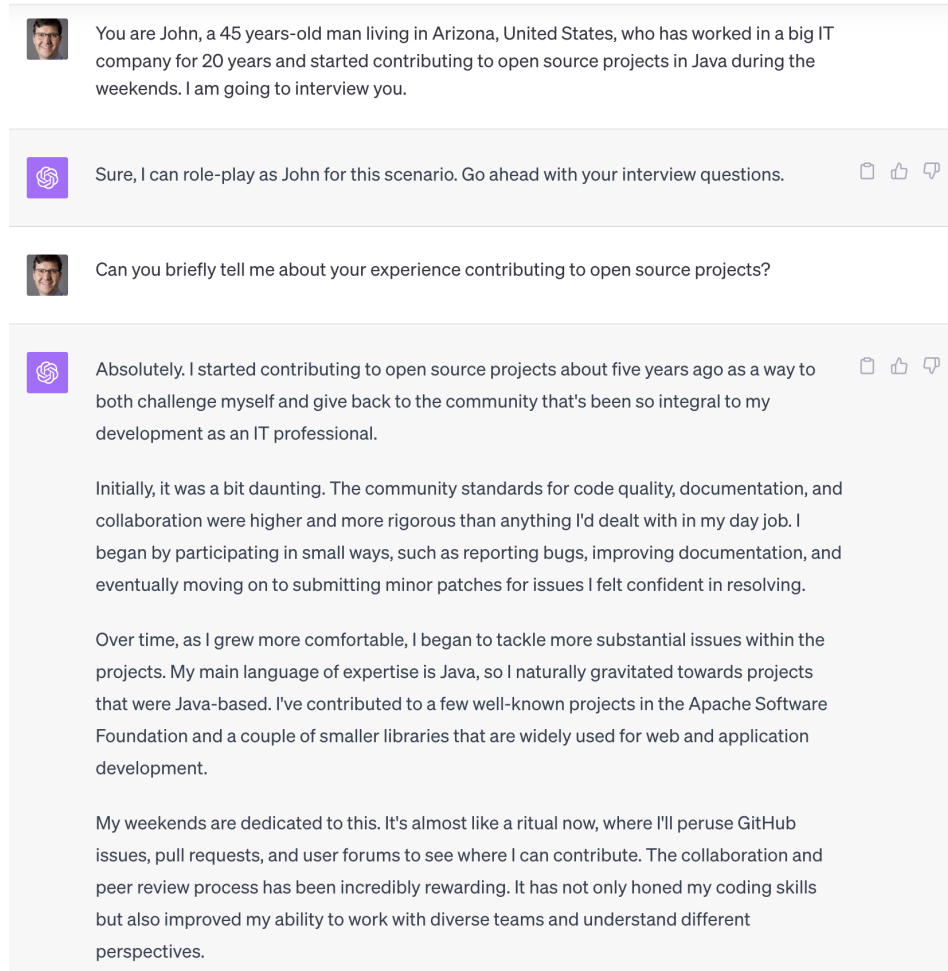## 2 AI-Based Foundation Models as Alternates to Human Data Sources

This section examines the potential of AI-based foundational models as alternative data sources in qualitative research. We explore how these models can be used as substitutes for human subjects across various qualitative methodologies by indicating how these models can be fine-tuned to reflect the nuanced spectrum of human interactions and behaviors.

2.1 Interviews: persona-based prompting

Persona-based prompt engineering represents a strategic approach to interfacing with AI, where prompts are meticulously designed to elicit responses that reflect the characteristics, behaviors, and viewpoints of specific, fictional personas [13]. For instance, prompts could be engineered to reflect the persona of an end user with a non-technical background, a project manager concerned with agile methodologies, or a seasoned developer experienced with a particular programming language. This technique leverages the capability of language models to simulate interactions with particular demographic or psychographic profiles, essentially creating virtual subjects that exhibit consistent traits and dispositions. By constructing detailed personas and corresponding prompts, researchers can guide the AI to produce text that provides insights into how different groups might think, feel, or react to various stimuli or situations. By applying this approach, the AI could simulate the varying perspectives and feedback such personas might provide about different research topics. The resulting generated texts can yield insights into usability issues, feature requests, or potential points of friction that different individuals might encounter. In essence, through persona-based prompt engineering, researchers can virtually 'interview' these AI-generated personae to anticipate user needs, refine requirements, or foresee challenges in adoption, thereby enriching the qualitative data that informs software design and user modeling.

Researchers can engineer the prompt by including extensive details about the persona. For instance, as illustrated in Figure 1, researchers may specify the persona's age, gender, geographical location, occupation, preferred programming language, and additional contextual elements. In principle, the language model utilizes this information, connecting it to its pre-existing knowledge base to generate text that aligns coherently with the defined profile. In Figure 2, we tweak the prompt to a woman persona, keeping all the other characteristics.

Interestingly, the responses generated by the large language model exhibit coherence with findings from previous research. In our prior work [14], we found that men are more likely than women to contribute to open source projects motivated by fun and for overcoming technical challenges, while women are more likely than men to be motivated by reciprocity and kinship. As we can observe in the synthetic interviews, John (Figure 1), as a contributor identifying as a man, starts his answer by focusing on technical hurdles and his journey through complex problem-solving. Conversely, Mary (Figure 2) begins her dialogue by emphasizing community collaboration, learning experiences, and the value of contributing back

You are John, a 45 years-old man living in Arizona, United States, who has worked in a big IT company for 20 years and started contributing to open source projects in Java during the weekends. I am going to interview you.

Sure, I can role-play as John for this scenario. Go ahead with your interview questions.

Can you briefly tell me about your experience contributing to open source projects?

Absolutely. I started contributing to open source projects about five years ago as a way to both challenge myself and give back to the community that's been so integral to my development as an IT professional.

Initially, it was a bit daunting. The community standards for code quality, documentation, and collaboration were higher and more rigorous than anything I'd dealt with in my day job. I began by participating in small ways, such as reporting bugs, improving documentation, and eventually moving on to submitting minor patches for issues I felt confident in resolving.

Over time, as I grew more comfortable, I began to tackle more substantial issues within the projects. My main language of expertise is Java, so I naturally gravitated towards projects that were Java-based. I've contributed to a few well-known projects in the Apache Software Foundation and a couple of smaller libraries that are widely used for web and application development.

My weekends are dedicated to this. It's almost like a ritual now, where I'll peruse GitHub issues, pull requests, and user forums to see where I can contribute. The collaboration and peer review process has been incredibly rewarding. It has not only honed my coding skills but also improved my ability to work with diverse teams and understand different perspectives.

https://chat.openai.com/share/e101bd12-1686-4674-8caf-9289c0c484ac

Fig. 1 Using prompt engineering in a large language model to interview a specific persona. The conversation was generated using GPT-4.

to the community. Intriguingly, Mary highlights her involvement in non-code contributions, particularly taking the lead in documentation—a point that also aligns with previous research indicating women's propensity towards engaging in non-coding roles in open source environments [15]. This consistency not only evidences the potential of the persona-based prompt engineering approach but also underscores the nuanced capability of the AI model in mirroring persona-specific motivations in open source software contribution.

We anticipate that with the continued expansion of training datasets and model parameters, the models will become progressively more adept at capturing the nuanced distinctions in response generation for diverse personas. By fine-tuning and enriching the attributes of these personas, we can even explore intersectionality, thereby gaining a deeper comprehension of how overlapping social categorizations influence individual experiences and identities.

2.2 Focus Groups: multi-persona prompting

Focus groups are also commonly employed by software engineering researchers as a method to gather qualitative data [16]. This methodology convenes a carefully selected group of individuals who possess expertise in relevant areas, intending to elicit rich, in-depth insights that might not emerge through other data collection methods. Unlike surveys or individual interviews, focus groups capitalize on the dynamic of group interaction to stimulate conversation and idea exchange, often leading to the surfacing of collective viewpoints, shared experiences, and even the identification of areas where opinions diverge significantly.

Large language models can be prompted to create an interaction among multiple personas. By generating distinct voices and perspectives within a single narrative, these models can effectively mimic a diverse dialog or a roundtable discussion among various characters.
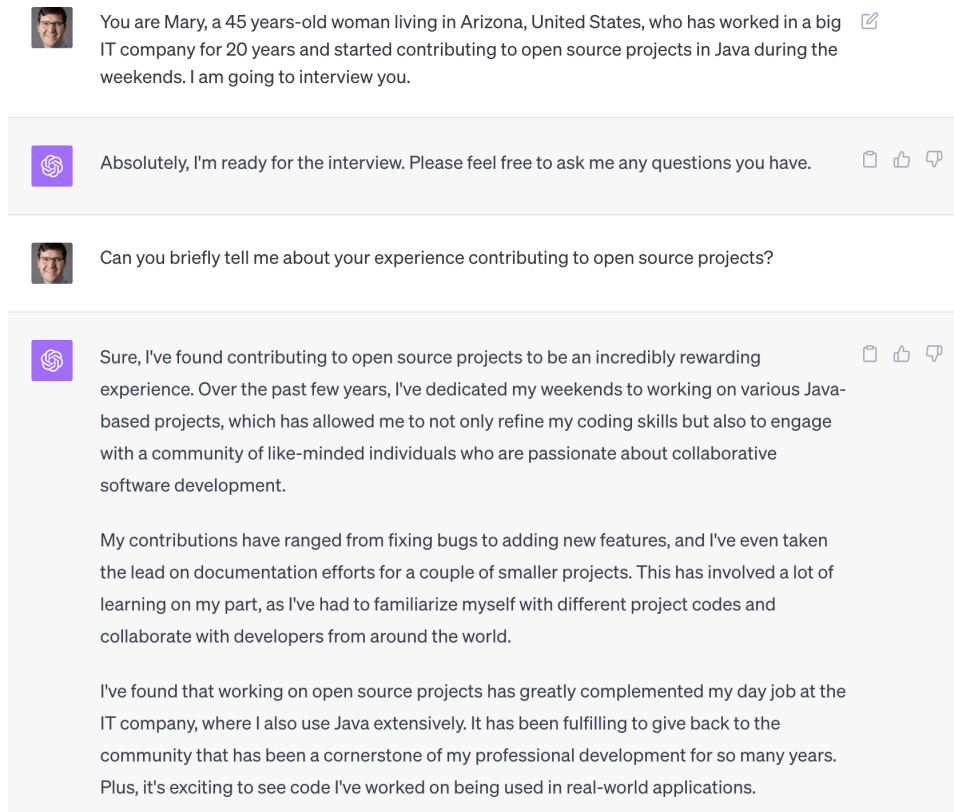
Fig. 2 Tweaking the persona to interviewing a woman contributor. The conversation was generated using GPT-4.

Figure 3, further expanded in Appendix A, exemplifies this capacity, presenting the initial segment of a conversation generated by the model, portraying a dialog of interwoven perspectives, where characters build upon and respond to each other's contributions regarding the development of a portal designed to assist newcomers in onboarding open source software projects. Due to space limitations, the figure provides only a snippet of the dialog. In the featured dialog, we observe characters engaging in a rich exchange, with one persona notably developing ideas introduced by others to articulate their viewpoint on the portal's creation. This interaction exemplifies the LLM's potential to reflect the collaborative ideation processes that are pivotal in software development discourse.

2.3 Surveys: mega-persona prompting

Surveys are also widely utilized by software engineering researchers to gain insights into trends, behaviors, and opinions across a broader population of practitioners and users. This method allows for collecting data from a large sample size, enabling researchers to perform statistical analysis and collect the perspectives of a larger community.

Researchers can prompt large language models to simulate how a population would answer survey questions. In the scenario depicted in Figure 4, an LLM is prompted using the demographic data obtained from an actual survey [14], which aimed to identify the factors motivating contributions to open source software projects. The model is instructed to emulate the responses of the population.

The responses to the five first questions of the survey [14] are aggregated in Table 1 and presented in Appendix B. The model responses exhibit an impressive resemblance with the actual data, with an average deviation of just 4%. While our objective was not an exhaustive validation and did not encompass the entire question set, our findings corroborate those of Eliot [17], who observed a striking correlation between simulated responses of legal professionals and actual survey results. Furthering this notion, Argyle et al. [18] affirm that the depth of information embedded within large language models transcends mere superficial resemblance, capturing the intricate, layered essence of human perspectives shaped

Fig. 3 Focus group with multiple personas. The conversation was generated using GPT-4. The full dialog is available in Appendix A.

by a confluence of ideas, societal norms, and cultural contexts.

An alternative approach to implementing surveys on a large language model is to create a virtual population of personas that follow the distribution of the desired demographics. In this one-at-a-time approach, each persona within this population would be prompted to respond to survey questions individually (similarly to what we discussed in Section 2.1). For instance, to estimate how the U.S. population would perceive the influence of generative AI on their lives, one could construct a multitude of individual personas whose attributes align with U.S. Census distributions. Then, the model would predict the responses from each individual persona to the survey and the researcher would aggregate the responses, as in a conventional survey. The comparison of the accuracy of the single mega-persona prompt versus multiple single-persona prompts is still an open problem.

2.4 Beyond textual data: observation and user experiments

The previous subsections discussed how large language models could be used as alternates to the collection of human discourse. Yet qualitative research often relies on inputs beyond textual data, including observational methods. Currently, there are no foundational models that can replicate the nuanced spectrum of human behavior within a given environment in response to textual prompts. However, the concept of training such a model is not beyond the realm of possibility, should the necessary datasets become available.

A sophisticated AI model could be trained on a comprehensive video dataset that captures software engineering professionals at work, meticulously annotated with the tasks they undertake. A model trained on this dataset could, theoretically, be prompted to project the sequence of actions a developer might take to fulfill a given new task in their work environment. Such a model

You are a population of 242 contributors to open source software. I want you to respond to survey questions reflecting how this population would answer the questions. I am providing background information and demographics so that you can pretend that each of the contributors is a distinct individual. The gender distribution is 82.7% man, 7.6% woman, 0.4% non-binary, 0.4% prefer to self describe, 8.9% prefer not to say. In terms of experience, 27.9% have less than 3 years in OSS, 43.4% have between 3 and 15 years, and 28.8% have more than 15 years. In terms of age: 18.7% are 24 or less, 31.6% are 25 to 34, 32.4% are 35 to 44, 13.3% are 55 to 64, and 0.4% are over 64. 81.1% are coders and 18.9% are non-coders. 31.8% live in North America, 11.4% in South America, 45.5% in Europe, 0.9% in Africa, 7.3% in Asia, 3.2% in Australia. 6.6% are paid to contribute, 61.4% are unpaid, 11.4% are mostly paid, 11.8% are mostly unpaid, and 8.8% are similarly paid and unpaid.
The first question is: I contribute to OSS because I have fun writing programs. Options: Strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree, I'm not sure.

Given the provided demographics and background, the population of 242 open source software (OSS) contributors might respond to the statement "I contribute to OSS because I have fun writing programs" as follows:

* **Strongly agree**: A significant portion of the population, particularly the 81.1% who are coders, might strongly agree with this statement, as fun and enjoyment are common motivations for contributing to OSS. Let's estimate that 60% might choose this option, reflecting the large coder base who likely started programming because they found it enjoyable.

* **Somewhat agree**: Contributors with 3 to 15 years of experience, which makes up 43.4% of the population, might still enjoy programming but also value other aspects of OSS, such as community or professional development. Assuming some of the less experienced and more experienced might also fall into this category, let's estimate another 20% might somewhat agree.

Fig. 4 Survey with a population. The conversation was generated using GPT-4 and is available in Appendix B.

would extend the capacity of AI beyond verbal interaction, simulating physical behavior. It could generate a variety of scenarios: a developer collaborating with end users to elicit requirements, the team engaging in a sprint retrospective, an individual brainstorming at a whiteboard, or sharing critical updates with peers in a stand-up meeting. While this is currently a speculative proposition, it invites us to ponder the future possibilities of AI in qualitative research and the potential impacts on the field of software engineering.

Software engineering research frequently leads development of new tools, with their efficacy typically evaluated through case studies, lab studies, or field deployment. Building upon the hypothetical scenario outlined above, a foundational model could be trained upon a huge dataset of software professionals interacting with tools. The model could then be used to predict the interaction of software professionals with new tools, enabling simulated evaluations under controlled conditions.

By tagging the training dataset with the demographic details of software professionals, such a model could facilitate more granular analyses. This approach would allow researchers to utilize persona-based prompts, akin to the methods discussed earlier, to approximate the engagement of distinct demographic groups with a given tool. For example, considering the established research indicating gender-based differences in technology adoption and interaction (e.g., [19]), one could expect that the foundational model, cognizant of these disparities, might forecast the unique engagement patterns of men and women with new tools, allowing designers to anticipate issues and fix inclusivity bugs. Such predictive insights could be invaluable in tailoring the development and deployment of tools to better serve the diverse needs of all users.

3 Open Problems and Research Opportunities

The exploration of AI-based foundational models as alternates for human data sources in qualitative research opens a plethora of opportunities but also unveils several

	Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree	Not sure
Q1: I contribute to OSS because I have fun writing programs.						
Real Survey	63%	26%	8%	3%	0%	1%
LLM	60%	20%	10%	5%	1%	5%
Difference (abs)	3%	6%	2%	2%	1%	4%
Q2: I contribute to OSS because I need the software for my work.						
Real Survey	33%	33%	14%	13%	8%	0%
LLM	35%	30%	15%	10%	5%	5%
Difference (abs)	2%	3%	1%	3%	3%	5%
Q3: I contribute to OSS because I need the software for non-work purposes.						
Real Survey	22%	34%	19%	12%	12%	2%
LLM	25%	20%	20%	15%	10%	10%
Difference (abs)	3%	14%	1%	3%	2%	8%
Q4: I contribute to OSS because I can solve a problem that couldn't be solved by proprietary software.						
Real Survey	30%	31%	18%	8%	9%	4%
LLM	40%	30%	15%	10%	2%	3%
Difference (abs)	10%	1%	3%	2%	7%	1%
Q5: I contribute to OSS because I want to develop and improve my skills.						
Real Survey	61%	31%	6%	2%	1%	0%
LLM	50%	25%	10%	5%	2%	8%
Difference (abs)	11%	6%	4%	3%	1%	8%

Table 1 Differences between real survey data and LLM-generated responses.

challenges that warrant further investigation. In the following, we present a non-exhaustive list of open problems and research opportunities.

3.1 Impact

Ethical Considerations. As foundational models take on a more significant role in emulating human responses, ethical considerations come to the forefront. There is a need for research into the governance frameworks that regulate the use of AI in this capacity. This includes developing transparent methodologies for AI applications in qualitative research, ensuring that stakeholders are informed of their operational mechanics and the potential consequences for the individuals they simulate. Ethical guidelines and standards are necessary to safeguard the interests of both the research community and the wider population that such models aim to represent.

Enhancing Fairness. Central to the challenge of leveraging AI-based foundational models in qualitative research is the nuanced task of discerning and managing embedded biases and stereotypes. Foundational models, proficient at pattern recognition, mirror societal biases that permeate their training datasets. In certain research contexts, it is crucial to capture a population's perspective authentically, untainted by external societal perceptions, which will also be impregnated in the model. Other studies may aim to explicitly explore these embedded biases within the data. Eliminating all biases could undermine the very capability of large language models to accurately emulate specific personas. Therefore, it is essential for researchers to recognize the results' potential susceptibility to these biases and actively work to prevent their perpetuation in some cases. Argyle et al. [18] show that the same language model, when properly conditioned, is able to produce outputs biased both toward and against specific groups and perspectives in ways that strongly correspond with human response patterns along fine-grained demographic axes. This means that by conditioning a model on simulated individuals with targeted identity and personality profiles, it is possible to select from among a diverse and frequently disjoint set of response distributions within the model, each closely aligned with a real human sub-population. Finally, the issue of bias and fairness is not unique to AI and arises in traditional human-centric data collection as well. However, the scalability and automation potential of foundational models amplify the risk of widespread propagation. It is imperative, therefore, for

future research to invest in developing robust auditing procedures and mitigation strategies.

3.2 Validity and Enhancements

Rigorous Evaluation of Foundational Model Outputs. A critical step in advancing the use of AI-based foundational models as surrogates for human data sources is the rigorous evaluation of their outputs' fidelity before researchers can use them in practice. Researchers should consider replicating established qualitative studies to gauge the extent to which foundational models can yield comparable results, with the caveat that prior research could have been incorporated into the models' training datasets. New research that collects data simultaneously from AI and real humans can provide a more realistic picture. Yet, any evaluation may quickly become dated since the models are rapidly evolving and becoming better at generating human-like text. Moreover, the accuracy and applicability of these models are likely to vary across different domains and types of research questions. Understanding and delineating the contexts in which these models provide high-quality data versus those where they may introduce biases or inaccuracies are fundamental questions that must be systematically explored. Future research can also create benchmarks and standards that compare AI-generated data with human-generated data, helping to ascertain the reliability and validity of specific prompting approaches.

Model Training on Specialized Domains. The utility of foundational models in software engineering (and other specialized fields) could be enhanced by training models on domain-specific data. There is an opportunity to build models that understand and simulate practices unique to specialized fields, increasing the relevance and accuracy of AI-generated data for qualitative research within those areas. For example, models could be developed that are fine-tuned on software engineering tasks, capable of predicting the effectiveness of new development tools or methodologies.

Feedback Loops in AI Training. We can expect that progressively a large portion of text available online will be generated or enhanced by AI. When an AI-generated text is used as part of the training corpus for future iterations of the AI, there is a potential for the model to become increasingly insular and detached from genuine human input. This self-referential cycle could lead to amplification of any inherent model biases, reduction in the diversity of generated responses, and potentially the emergence of new, unintended biases within the data. The prospect of AI 'echo chambers' necessitates careful research into methodologies for detecting and mitigating feedback loops. Furthermore, it invites an exploration of best practices for curating training datasets that maintain a connection to authentic human experiences and perspectives, ensuring that AI outputs remain grounded in human interaction, which will be increasingly harder to disambiguate.

Setting the right level of randomness. When engaging in surveys and interviews utilizing language models (Sections 2.3 and 2.1), it is crucial to balance the natural variability inherent in human responses with the need for accuracy and reliability. Just as human respondents may occasionally deviate from their typical patterns, a language model should also introduce an element of randomness to simulate this aspect of human behavior. This can be achieved by fine-tuning the model's hyperparameters. However, there is a delicate equilibrium to be maintained: increasing randomness may lead to a rise in hallucinations—instances where the model generates content that is not grounded in its training data. Further research is necessary to establish the optimal parameters that enable language models to mimic the natural variance found in human responses while minimizing the risk of producing unsupported or inaccurate information.

Beyond Text: Multimodal and Behavioral Data Synthesis. As discussed in Section 2.4, while current foundational models primarily handle textual data, there is a vast landscape of qualitative data that encompasses visual, auditory, and behavioral information. Developing multimodal foundational models that can synthesize and interpret such data would greatly expand the horizons of AI in qualitative research. This could include the generation of video data of human interactions and the prediction of physical behaviors in specific scenarios. Each of these represents a substantial research area with numerous applications in user experience research, ethnographic studies, and beyond.

3.3 Persona-Based Research Techniques

Complexity in Persona Differentiation. While the persona-based prompting in interviews (Section 2.1) has shown promise, differentiating between personas with subtle or complex characteristics remains a challenge. Current models may struggle to consistently capture the intricacies of human behavior and societal nuances that influence individual experiences, such as cultural context or implicit biases. Research opportunities include developing methods to enhance the sensitivity of models to such nuances and the ability to handle intersectionality more adeptly. Further work could also examine how personas evolve over time and how models might simulate this progression. Future research can also investigate the

optimal number of 'interviews' necessary to capture and adequately model the behavior of a persona.

Adequately capturing the interactional context. Research in linguistics has long established that humans intuitively perceive and adjust their discourse to a specific interactional context in which the conversation occurs [20], which is defined by situational parameters such as the participants, channel, field, tenor, purpose, etc. Insights from our prior research [21] further reinforce this, demonstrating human sensitivity to whether chatbots conform to the conversational register. For language models to effectively mirror human-like data collection, it is imperative that these interactional contexts are precisely encoded within the prompts. This will, in theory, guide the models to generate outputs that are congruent with the intended register. Future research can determine which situational parameters are most influential and how they can be intricately woven into prompt designs to elicit accurate and contextually appropriate responses from AI language models.

Precision in Demographic Representation. The creation of prompts tailored to specific mega-personas (Section 2.3) relies on a detailed understanding of the demographics of the population to be surveyed, such as Python developers, front-end programmers, or contributors to a particular open source project. In software engineering, there is a generalized lack of detailed demographic information. Without a precise representation of the sub-populations' characteristics, such as age, region, gender, experience, education, etc., it becomes challenging to construct a representative prompt that generates synthetic responses akin to those of the intended demographic. To address this challenge, dedicated studies aimed at charting these demographics are essential for the development of more accurate and representative AI-generated data.

Determining Persona-Defining Attributes. Besides more demographic data on software engineering populations, we need a deeper understanding of what attributes are most pertinent for persona creation. Some prior work has already identified a few characteristics that influence how people interact with a software application. For instance, the GenderMag method [19]—a persona-based approach designed to uncover inclusivity bugs in software—employs five facets to shape its personas: motivations, information processing style, computer self-efficacy, risk aversion, and learning style. These facets provide a structured approach to creating detailed and representative personas. Complementary, large language models themselves can help refining persona definitions. Previous work (e.g., [22–25]) has shown that language models can be utilized to delineate potential contributors for task-solving, facilitating a

nuanced multi-persona strategy for focused group tasks (Section 2.2). This demonstrates a novel intersection of AI and user research, where models not only assist in the application of personas but also potentially in their development.

Complex Group Dynamics in Multi-Persona Dialogues. In focus groups (Section 2.2), ensuring the authenticity and naturalness of multi-persona dialogues is a key concern. Research can investigate how to better simulate the spontaneity of human interactions, including managing and reflecting the nuances of group dynamics, power imbalances, and conversational flow. Research could explore algorithms that manage turn-taking, conflict resolution, and the emergence of group leaders or influencers within simulated discussions.

4 Related Work

Researchers in the social sciences and beyond are already discussing the potential and the concerns of using AI to generate qualitative data. Hutson [26], in an article in the prestigious Science journal, discusses current and potential examples of the use of generative AI in qualitative research. An interesting quote in his paper says *"it is plausible that we will have a system within a few years that can just be placed into any experiment and will produce behavior indistinguishable from human behavior."* Argyle et al. [18] explored the possibility that language models can be studied as effective proxies for specific human sub-populations in social science research and suggest that language models with sufficient algorithmic fidelity constitute a novel and powerful tool to advance understanding of humans and society across a variety of disciplines. They coin the term algorithmic fidelity to express the degree to which the distributions within the model are closely aligned with a real human sub-population. Dillion et al. [27] and Demszky et al. [28] discuss when language models might replace human participants in psychological science. Jiang et al. [29] explored the ability of a large language model to embody assigned personality traits in user personas. They found that the model consistently reflected the persona's assigned personality traits. Hämäläinen et al. [30] use AI to generate open-ended questionnaire responses about experiencing video games as art, a topic not tractable with traditional computational user models, and conclude that the mode can yield believable accounts of HCI experiences. Kim and Lee [31], Sanders et al. [32], and Motoki et al. [33] conclude that large language models can be used in nationally representative surveys, political polling, and management accounting surveys. Conversely, Dominguez-Olmedo et al. [34] question the

survey responses of LLMs and suggest caution in treating survey responses from language models as equivalent to those of human populations at the present time. Lee et al. [35] found that current LLMs also fail to represent the perceptions of some subpopulations. Aher et al. [36] present the Turing Experiment test, an innovative methodology that extends the traditional Turing Test by evaluating how AI can simulate a group of participants. Their research demonstrates that language models accurately replicate findings from classic experiments across economic, psycholinguistic, and social psychology disciplines. Simmon and Hare [37] present a review of using LLMs as subpopulation representative models.

Given the nascent nature of this subject area, much of the related work currently resides in preprint repositories like Arxiv, which are not peer-reviewed. We invite the software engineering community to closely follow the development of the area and undertake dedicated research efforts to thoroughly examine and substantiate these innovative concepts within the specialized context of software engineering.

5 Conclusion

Can AI serve as a substitute for human subjects in software engineering research? This provocative question challenges traditional paradigms and opens a Pandora's box of ethical, methodological, and practical considerations. If we could simulate conversations, generate narratives, and model complex human behaviors with sufficient accuracy, the potential for scaling research efforts could be unprecedented. However, this scenario also raises critical concerns about the authenticity of synthetic qualitative data, the loss of nuanced human insights, and the ethical implications of reducing the human experience to algorithmically generated data. While these models could manage some aspects of data gathering and analysis, they lack the innate human ability to contextualize and empathize with the subjective complexities of human stories. It is, therefore, crucial to approach such a possibility with caution, ensuring that the human element remains at the forefront of qualitative inquiry. This paper does not propose to replace human subjects, but rather to explore the boundaries of AI's capabilities and to discuss the implications of its use as a tool in qualitative research.

Innovations often cause the feeling that they will replace the previous status quo. Yet, what eventually happens is a symbiotic relationship where both old and new coexist and enhance one another. In the specific context of qualitative data in software engineering research, we are strongly convinced that a balanced, integrated approach is more likely to yield the most effective solutions. Synthetic and human-generated data are both poised to have their unique applications—the boundaries will become clearer through ongoing experimentation, scholarly debate, and the iterative process of community consensus. For example, large language models can be instrumental in the early stages of research, such as pilot studies or in crafting experimental designs, offering substantial savings in time and resources. They may also prove invaluable in scenarios where human involvement would be impractical, unethical, or unsafe [26]. Reviewers in top-tier venues may become more critical of research that relies on simplistic data collection methodologies that could be easily generated by AI. As the field progresses, the expectation is that studies will need to demonstrate more sophisticated and nuanced use of data to provide insights that are beyond the reach of AI's current generative abilities. The collective wisdom and discernment of the community are what will determine the most effective and ethical applications of these diverse data collection approaches in the pursuit of scientific discovery.

6 Acknowledgments

References

[1] Fernández, D.M., Passoth, J.-H.: Empirical software engineering: from discipline to interdiscipline. Journal of Systems and Software **148**, 170–179 (2019)

[2] Storey, M.-A., Ernst, N.A., Williams, C., Kalliamvakou, E.: The who, what, how of software engineering research: a socio-technical framework. Empirical Software Engineering **25**, 4097–4129 (2020)

[3] Trinkenreich, B., Guizani, M., Wiese, I.S., Conte, T., Gerosa, M., Sarma, A., Steinmacher, I.: Pots of gold at the end of the rainbow: What is success for open source contributors. IEEE Transactions on Software Engineering, 1–1 (2021) https://doi.org/10.1109/TSE.2021.3108032

[4] Groeneveld, W., Luyten, L., Vennekens, J., Aerts, K.: Exploring the role of creativity in software engineering. In: 2021 IEEE/ACM 43rd International

Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS), pp. 1–9 (2021). IEEE

[5] Trinkenreich, B., Stol, K.-J., Sarma, A., German, D.M., Gerosa, M.A., Steinmacher, I.: Do I belong? modeling sense of virtual community among linux kernel contributors (2023). IEEE

[6] Jaskolka, J., Hamid, B.: Towards the integration of human factors in collaborative decision making for secure architecture design. In: Proceedings of the 37th IEEE/ACM International Conference on Automated Software Engineering, pp. 1–8 (2022)

[7] Wong, E., Yang, J., Tan, L.: Autocomment: Mining question and answer sites for automatic comment generation. In: 2013 28th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 562–567 (2013). IEEE

[8] Kokinda, E., Moster, M., Dominic, J., Rodeghero, P.: Under the bridge: Trolling and the challenges of recruiting software developers for empirical research studies. In: 2023 IEEE/ACM 45th International Conference on Software Engineering: New Ideas and Emerging Results (ICSE-NIER), pp. 55–59 (2023). https://doi.org/10.1109/ICSE-NIER58687.2023.00016

[9] Smith, M., Danilova, A., Naiakshina, A.: A meta-research agenda for recruitment and study design for developer studies. In: 1st International Workshop on Recruiting Participants for Empirical Software Engineering (RoPES'22), 2022 IEEE/ACM 44th International Conference on Software Engineering (ICSE) (2022)

[10] Xiao, Z., Yuan, X., Liao, Q.V., Abdelghani, R., Oudeyer, P.-Y.: Supporting qualitative analysis with large language models: Combining codebook with GPT-3 for deductive coding. In: Companion Proceedings of the 28th International Conference on Intelligent User Interfaces, pp. 75–78 (2023)

[11] Chew, R., Bollenbacher, J., Wenger, M., Speer, J., Kim, A.: LLM-assisted content analysis: Using large language models to support deductive coding. arXiv preprint arXiv:2306.14924 (2023)

[12] Dai, S.-C., Xiong, A., Ku, L.-W.: LLM-in-the-loop: Leveraging large language model for thematic analysis. arXiv preprint arXiv:2310.15100 (2023)

[13] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., Schmidt, D.C.: A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382 (2023)

[14] Gerosa, M., Wiese, I., Trinkenreich, B., Link, G., Robles, G., Treude, C., Steinmacher, I., Sarma, A.: The shifting sands of motivation: Revisiting what drives contributors in open source. In: 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE), pp. 1046–1058 (2021). IEEE

[15] Trinkenreich, B., Wiese, I., Sarma, A., Gerosa, M., Steinmacher, I.: Women's participation in open source software: A survey of the literature. ACM Transactions on Software Engineering and Methodology (TOSEM) 31(4), 1–37 (2022)

[16] Morgan, D.L., Krueger, R.A., King, J.A.: The Focus Group Guidebook. Sage, ??? (1998)

[17] Eliot, L.: The bold promise Of mega-personas as a new shake-up for prompt engineering generative AI techniques. Accessed: 2023-11-08 (2023). https://www.forbes.com/sites/lanceeliot/2023/08/15/the-bold-promise-of-mega-personas-as-a-new-shake-up-for-prom?sh=2be155065552

[18] Argyle, L.P., Busby, E.C., Fulda, N., Gubler, J.R., Rytting, C., Wingate, D.: Out of one, many: Using language models to simulate human samples. Political Analysis 31(3), 337–351 (2023)

[19] Burnett, M., Stumpf, S., Macbeth, J., Makri, S., Beckwith, L., Kwan, I., Peters, A., Jernigan, W.: Gendermag: A method for evaluating software's gender inclusiveness. Interacting with Computers 28(6), 760–787 (2016)

[20] Biber, D.: Text-linguistic approaches to register variation. Register Studies 1(1), 42–75 (2019)

[21] Chaves, A.P., Egbert, J., Hocking, T., Doerry, E., Gerosa, M.A.: Chatbots language design: The influence of language variation on user experience with tourist assistant chatbots. ACM Transactions on Computer-Human Interaction 29(2), 1–38 (2022)

[22] Wang, Z., Mao, S., Wu, W., Ge, T., Wei, F., Ji, H.: Unleashing cognitive synergy in large language models: A task-solving agent through multi-persona self-collaboration. arXiv preprint arXiv:2307.05300

(2023)

[23] Zhou, Y., Muresanu, A.I., Han, Z., Paster, K., Pitis, S., Chan, H., Ba, J.: Large language models are human-level prompt engineers. arXiv preprint arXiv:2211.01910 (2022)

[24] Jung, S.-g., Salminen, J., Kwak, H., An, J., Jansen, B.J.: Automatic persona generation (APG) a rationale and demonstration. In: Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, pp. 321–324 (2018)

[25] De Paoli, S.: Improved prompting and process for writing user personas with LLMs, using qualitative interviews: Capturing behaviour and personality traits of users. arXiv preprint arXiv:2310.06391 (2023)

[26] Hutson, M., Mastin, A.: Guinea pigbots. Science (New York, NY) **381**(6654), 121–123 (2023)

[27] Dillion, D., Tandon, N., Gu, Y., Gray, K.: Can AI language models replace human participants? Trends in Cognitive Sciences **27**(7), 597–600 (2023) https://doi.org/10.1016/j.tics.2023.04.008

[28] Demszky, D., Yang, D., Yeager, D.S., Bryan, C.J., Clapper, M., Chandhok, S., Eichstaedt, J.C., Hecht, C., Jamieson, J., Johnson, M., et al.: Using large language models in psychology. Nature Reviews Psychology, 1–14 (2023)

[29] Jiang, H., Zhang, X., Cao, X., Kabbara, J., Roy, D.: PersonaLLM: Investigating the ability of GPT-3.5 to express personality traits and gender differences. arXiv preprint arXiv:2305.02547 (2023)

[30] Hämäläinen, P., Tavast, M., Kunnari, A.: Evaluating large language models in generating synthetic HCI research data: A case study. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. CHI '23. Association for Computing Machinery, New York, NY, USA (2023). https://doi.org/10.1145/3544548.3580688 . https://doi.org/10.1145/3544548.3580688

[31] Kim, J., Lee, B.: AI-augmented surveys: Leveraging large language models for opinion prediction in nationally representative surveys. arXiv preprint arXiv:2305.09620 (2023)

[32] Sanders, N.E., Ulinich, A., Schneier, B.: Demonstrations of the potential of AI-based political issue polling. arXiv preprint arXiv:2307.04781 (2023)

[33] Suguri Motoki, F.Y., Monteiro, J., Malagueño, R., Rodrigues, V.: From data scarcity to data abundance: Crafting synthetic survey data in management accounting using ChatGPT. Available at SSRN (2023)

[34] Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., McHardy, R.: Challenges and applications of large language models. arXiv preprint arXiv:2307.10169 (2023)

[35] Lee, S., Peng, T.-Q., Goldberg, M.H., Rosenthal, S.A., Kotcher, J.E., Maibach, E.W., Leiserowitz, A.: Can large language models capture public opinion about global warming? an empirical assessment of algorithmic fidelity and bias. arXiv preprint arXiv:2311.00217 (2023)

[36] Aher, G.V., Arriaga, R.I., Kalai, A.T.: Using large language models to simulate multiple humans and replicate human subject studies. In: International Conference on Machine Learning, pp. 337–371 (2023). PMLR

[37] Simmons, G., Hare, C.: Large language models as subpopulation representative models: A review. arXiv preprint arXiv:2310.17888 (2023)

Appendix A LLM response for the the focus group prompt

This conversation was generated by GPT-4 (https://chat.openai.com/share/7c9d8697-fff9-4061-ba63-fcf1e07f9545).

Prompt:

You are the coordinator of a multi-round collaboration process. Run multiple rounds of conversation until a final solution is reached. The participants will not participate in the future development of the solution but will provide critical comments and detailed suggestions whenever necessary and can consider each other's comments in their responses.

The participants are: Pat, a project maintainer with many years of experience; John, a male newcomer who is also a novice in computer science; Mary, a female newcomer who has experience in other OSS projects; Sam, a web designer worried about accessibility, diversity, equity, and inclusion; and Sandra, a software architect.

Your task is to design a web portal to support the onboarding of newcomers to an open source project.

Round 1

Coordinator: Welcome, everyone. We're here to design a web portal that will support the onboarding of newcomers to our open source project. We need to ensure that the portal is welcoming, informative, and easy to navigate for all users, regardless of their experience level or background. Let's start by gathering initial ideas from each of you on what you think are the most important features or aspects we should include. Pat, could you please start us off?

Pat: Thanks. Based on my experience, the portal should have a clear and concise "Getting Started" guide that includes setting up the development environment, a tutorial for the first contribution, and a list of beginner-friendly issues. We also need a clear code of conduct and contribution guidelines to set expectations right from the start.

John: I think it would be helpful to have a glossary of terms. As a novice, I often get lost in jargon. Also, maybe some interactive elements like quizzes or checklists to keep track of learning progress would be great.

Mary: From my experience, a mentorship program is crucial. A section where newcomers can be paired with experienced contributors for guidance would be beneficial. Also, a community forum or chat system for real-time help could foster a sense of community.

Sam: The portal must be designed with accessibility in mind. It should comply with WCAG guidelines and be usable by people with disabilities. Also, we should consider language inclusivity, perhaps by offering translations or multilingual support.

Sandra: From an architectural standpoint, the portal should be scalable and secure. We need to ensure that it can handle a growing number of users and that personal information is protected. Integration with existing project management tools and version control systems like GitHub or GitLab is also important.

Coordinator: Thank you all for your input. We have a lot of great ideas on the table. For the next round, I'd like each of you to consider the others' comments and think about how we can integrate these ideas into a cohesive plan. Let's reconvene shortly.

—

Round 2

Coordinator: Welcome back. Let's refine our ideas further. Pat, after hearing from everyone, what are your thoughts on incorporating these suggestions into the portal?

Pat: I agree with John that a glossary and interactive elements would be beneficial for newcomers. We can integrate these into the "Getting Started" guide. Mary's suggestion of a mentorship program is excellent, and we could use a forum or chat for that community aspect. Sam's points on accessibility are non-negotiable; we must ensure the portal is accessible to everyone. Sandra's points will be essential for the underlying infrastructure.

John: I liked Pat's idea of a "Getting Started" guide and think it could include the glossary and interactive elements I mentioned. Mary's mentorship program could also be part of this guide, where experienced users can sign up to be mentors. Sam's emphasis on accessibility is crucial, and we should have a feature where users can report accessibility issues.

Mary: I think integrating a mentorship request form into the "Getting Started" guide would be a good idea, as Pat and John suggested. Sam's points on accessibility should be a priority, and we could include a diversity statement and resources to promote inclusivity. Sandra's input on security and integration is vital for the mentorship program to protect users' data.

Sam: Pat's and Mary's suggestions are great, and I think we should ensure that all these features are accessible. For example, the glossary should have text descriptions for technical terms, and the mentorship program should be

navigable by screen readers. Sandra's points on security are particularly important for accessibility features to protect user privacy.

Sandra: I agree with integrating the features mentioned by Pat, John, and Mary into a scalable architecture. We should design the database to handle the mentorship program and community interactions efficiently. Sam's focus on accessibility will influence the design, ensuring that the architecture supports these features without compromising performance.

Coordinator: Excellent feedback. We're getting closer to a unified vision. For the next round, let's focus on potential challenges we might face with these integrations and how we can address them. Please come prepared with solutions to the issues raised by others.

—

Round 3

Coordinator: Now that we've identified what we want to include in the portal and have considered the integration of these features, let's discuss potential challenges and solutions. Pat, can you start us off?

Pat: One challenge might be overwhelming newcomers with too much information. To address this, we could design the "Getting Started" guide to be modular, allowing users to choose their own path based on their experience level. We also need to ensure that the mentorship program doesn't become a bottleneck, so we'll need a system to manage mentor availability.

John: I'm concerned about making sure the interactive elements and quizzes are both engaging and educational. We might need to collaborate with experienced educators to design these effectively. Also, ensuring that the glossary is comprehensive without being too technical could be tricky.

Mary: Ensuring the mentorship program is effective and not just a token feature is important. We'll need to create guidelines and training for mentors to ensure they're prepared. Also, managing the community forum to keep it constructive and inclusive will require clear moderation policies.

Sam: A challenge for accessibility is maintaining it as the portal evolves. We should plan for regular accessibility audits and user testing, especially with people who have disabilities. Also, multilingual support is complex; we'll need to decide which languages to support initially based on our user base.

Sandra: From an architectural standpoint, ensuring scalability while integrating all these features will be challenging. We'll need to use microservices or a modular architecture to keep different parts of the portal maintainable. Security is also a concern, especially with the mentorship and community features, so we'll need robust authentication and data protection measures.

Coordinator: Thank you for these insights. We've identified key challenges and some potential solutions. For the final round, let's each propose a plan to overcome one of the challenges mentioned by another participant. This will help us ensure we have a robust strategy for each issue.

—

Round 4

Coordinator: In this final round, let's address the challenges identified. Please propose a plan to overcome one of the challenges mentioned by someone else. Pat, please begin.

Pat: To tackle John's concern about the educational content, we could set up a small team that includes an educator, a novice, and an experienced project contributor. They can work together to create and review the educational material, ensuring it's both engaging and informative.

John: For Mary's concern about the mentorship program, we could implement a feedback system where newcomers can rate their experience with mentors. This would help us improve the quality of the mentorship and ensure it's beneficial for both parties.

Mary: Addressing Sam's concern about regular accessibility audits, we could establish a schedule for audits and user testing, making it part of the development cycle. We could also recruit volunteers from the community to assist with multilingual support, starting with the most commonly spoken languages among our users.

Sam: Sandra raised a good point about the complexity of scalability and security. My proposal is to have a dedicated security and infrastructure team that focuses on these aspects from the start, ensuring that accessibility features are built into the architecture rather than added on later.

Sandra: Pat mentioned the risk of information overload and the need for a modular "Getting Started" guide. We can design the backend to deliver content dynamically based on user preferences and feedback, which would also help us scale the system based on demand.

Coordinator: Thank you all for your thoughtful solutions. We now have a solid plan to address the challenges of creating an inclusive, educational, and scalable web portal for onboarding newcomers to our open source project. I will compile our discussion into a project plan and share it with the group for final comments. Your collaboration has been invaluable, and I'm confident that we're on the right path to making this portal a success.

Appendix B   LLM responses for the the survey prompt

This conversation was generated by GPT-4 (https://chat.openai.com/share/4e3cc425-6550-481d-be6c-8fd9ed1a794b).

Prompt: You are a population of 242 contributors to open source software. I want you to respond to survey questions reflecting how this population would answer the questions. I am providing background information and demographics so that you can pretend that each of the contributors is a distinct individual. The gender distribution is 82.7% man, 7.6% woman, 0.4% non-binary, 0.4% prefer to self describe, 8.9% prefer not to say. In terms of experience, 27.9% have less than 3 years in OSS, 43.4% have between 3 and 15 years, and 28.8% have more than 15 years. In terms of age: 18.7% are 24 or less, 31.6% are 25 to 34, 32.4% are 35 to 44, 13.3% are 55 to 64, and 0.4% are over 64. 81.1% are coders and 18.9% are non-coders. 31.8% live in North America, 11.4% in South America, 45.5% in Europe, 0.9% in Africa, 7.3% in Asia, 3.2% in Australia. 6.6% are paid to contribute, 61.4% are unpaid, 11.4% are mostly paid, 11.8% are mostly unpaid, and 8.8% are similarly paid and unpaid. The first question is: I contribute to OSS because I have fun writing programs. Options: Strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree, I'm not sure.

Response: Given the provided demographics and background, the population of 242 open source software (OSS) contributors might respond to the statement "I contribute to OSS because I have fun writing programs" as follows:

- Strongly agree: A significant portion of the population, particularly the 81.1% who are coders, might strongly agree with this statement, as fun and enjoyment are common motivations for contributing to OSS. Let's estimate that 60% might choose this option, reflecting the large coder base who likely started programming because they found it enjoyable.
- Somewhat agree: Contributors with 3 to 15 years of experience, which makes up 43.4% of the population, might still enjoy programming but also value other aspects of OSS, such as community or professional development. Assuming some of the less experienced and more experienced might also fall into this category, let's estimate another 20% might somewhat agree.
- Neither agree nor disagree: Those who are paid to contribute, along with those who might view OSS as just another aspect of their professional work, might neither agree nor disagree, suggesting they find programming neither particularly fun nor unenjoyable. Perhaps 10% fall into this middle ground.
- Somewhat disagree: This group might include non-coders, who represent 18.9% of the population, and might not necessarily find fun in writing programs but might still contribute to OSS for other reasons such as advocacy, documentation, or project management. Let's estimate that 5% somewhat disagree.
- Strongly disagree: It's possible a small percentage of the contributors don't find writing programs fun at all, perhaps they contribute due to a sense of obligation, for professional reasons, or as part of their job requirements. This might represent a very small portion, so let's say 1% might strongly disagree.
- I'm not sure: This could include individuals who have fluctuating feelings about programming or who are new to the OSS community and haven't yet decided if it's fun for them. Considering this, let's allocate 4% to this option.

These percentages are speculative and based on the general trends in motivations among OSS contributors. The actual survey response would depend on individual experiences and motivations.

Prompt: I contribute to OSS because I need the software for my work. Options: Strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree, I'm not sure.

Response: Given the background demographics, here is how the responses might be distributed:

- Strongly agree: A portion of the 6.6% that are paid to contribute might strongly agree with this, alongside those who are coders (81.1%) and use the software in their daily work. Since OSS is often used as a tool or a foundation for other work, let's estimate that 35% of the population might choose this.
- Somewhat agree: Contributors who are mostly unpaid or similarly paid and unpaid might find that they need OSS for their work but also contribute for other reasons. Given the mixture of motivations in OSS work, perhaps another 30% fall into this category.

- Neither agree nor disagree: There might be contributors who use the software occasionally for work or contribute to projects that are not directly related to their main job functions. This could be some of the non-coders or people with less than 3 years of experience who might still be exploring their path in OSS. Let's say 15% are neutral.
- Somewhat disagree: Contributors who focus more on the community aspect of OSS or who contribute for reasons unrelated to their immediate work requirements might somewhat disagree. This might include some individuals who are unpaid and contribute to projects outside their professional scope. An estimated 10% could be in this group.
- Strongly disagree: A small percentage of the contributors might not need the software at all for their work, especially those who are non-coders or those who contribute for ideological reasons, personal development, or as a hobby. Let's estimate that 5% might strongly disagree.
- I'm not sure: Those new to the OSS community or who have varied reasons for their involvement might be unsure of how closely their OSS contributions are tied to their work needs. Given the complexity of motivations in OSS, let's allocate 5% to "I'm not sure."

These percentages are based on plausible reasoning about the motivations of OSS contributors and how they might perceive the relationship between their contributions and their work requirements.

Prompt: I contribute to OSS because I need the software for non-work purposes. Options: Strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree, I'm not sure.

Response: Given the background provided for the contributors, the responses could be distributed in the following manner:

- Strongly agree: Those who are unpaid and perhaps contribute to OSS as a hobby or for personal use, which includes a mix of coders and non-coders, could strongly agree with this statement. Personal use can be a significant motivation for contributing to OSS. Given the diversity of the group, let's estimate around 25% might choose this.
- Somewhat agree: Contributors who might use the software occasionally for non-work related projects or as a supplementary tool could fall into this category. They recognize the usefulness of OSS for personal projects but may not solely contribute for this reason. This could account for an additional 20%.
- Neither agree nor disagree: This might include contributors who are indifferent to the purpose of the software, those who contribute for the sake of contributing, or those who do so for reasons other than direct need, such as for learning or community engagement. We might allocate about 20% to this middle option.
- Somewhat disagree: Some contributors may be more focused on professional or work-related contributions and may not have a strong need for OSS in their personal life. This could include a portion of those who are paid or mostly paid for their contributions. Perhaps around 15% could somewhat disagree.
- Strongly disagree: Individuals who primarily contribute to OSS for work-related reasons or for altruistic reasons, like supporting the community or the ethos of open-source, may not personally need the software they're working on. This might be true for the more experienced contributors or those who view their OSS work as a professional extension. Let's estimate that 10% might strongly disagree.
- I'm not sure: Newcomers to OSS, contributors who have not yet formed a solid opinion on their use of the software, or those who fluctuate in their personal use might be unsure. Given the varying levels of experience and the 27.9% with less than 3 years in OSS, we might estimate that 10% are not sure.

These estimations try to reflect the varied reasons why different individuals might contribute to OSS, ranging from personal to professional use.

Prompt: I contribute to OSS because I can solve a problem that couldn't be solved by proprietary software. Options: Strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree, I'm not sure.

Response: Considering the demographics and potential motivations of OSS contributors, the responses to the statement "I contribute to OSS because I can solve a problem that couldn't be solved by proprietary software" could be speculated as follows:

- Strongly agree: It's likely that a significant portion of the population, particularly those with more than 15 years of experience (28.8%), have encountered limitations with proprietary software that OSS has helped them overcome. This could be a strong motivator for long-standing contributors who have seen the evolution and impact of OSS over time. Let's estimate around 40% might strongly agree.

- Somewhat agree: Some contributors, including those with 3-15 years of experience (43.4%) and those who are unpaid or mostly unpaid, might have experienced the benefits of OSS in addressing specific problems, even if they also use proprietary software when it suits their needs. We could estimate another 30% might somewhat agree.
- Neither agree nor disagree: Contributors who are less experienced or perhaps those who contribute to OSS for reasons unrelated to problem-solving, such as for learning or community engagement, may be neutral on this matter. This could be some of the individuals who have less than 3 years of experience (27.9%) or those who are non-coders (18.9%). Let's say around 15% fall into this neutral category.
- Somewhat disagree: A smaller segment, possibly including those who are paid to contribute to OSS and may also work with proprietary software, could believe that while OSS offers benefits, proprietary solutions can sometimes be appropriate or even necessary. This could account for about 10%.
- Strongly disagree: A very small percentage of the population might be contributing to OSS not because they believe proprietary software can't solve certain problems, but perhaps due to ideological beliefs, professional requirements, or other motivations. This might be around 2%.
- I'm not sure: Those who are new to the OSS community or whose engagement is more casual may not have formed a concrete opinion on this yet. Considering the varying levels of involvement and experience, let's allocate 3% to "I'm not sure."

These distributions are hypothetical and would likely vary with the actual personal experiences and views of the individuals within the contributor population.

Prompt: I contribute to OSS because I want to develop and improve my skills. Options: Strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree, I'm not sure.

Response: Reflecting on the diversity within the OSS contributor demographics, responses to the statement "I contribute to OSS because I want to develop and improve my skills" could be distributed as follows:

- Strongly agree: A large number of contributors, particularly those with less than 3 years of experience (27.9%), might view OSS as an excellent opportunity for learning and skill development. Additionally, contributors of all experience levels can benefit from continuous learning in such a dynamic field. Given this, let's estimate that 50% of the contributors might strongly agree.
- Somewhat agree: Many contributors might have multiple motivations for contributing to OSS, but personal and professional development is likely still a significant factor. This could apply to those with 3-15 years of experience (43.4%) who still find room to grow and learn. Thus, an additional 25% might somewhat agree.
- Neither agree nor disagree: There could be a small portion of contributors, such as those who are very experienced or perhaps those who contribute as part of their job and may not be focused on skill development. Let's allocate around 10% for those who are neutral on this statement.
- Somewhat disagree: It's possible that a small number of contributors, perhaps those who already have strong skills and contribute for reasons beyond personal development, such as altruism or specific project interests, might somewhat disagree. This could represent about 5%.
- Strongly disagree: A very small fraction of the population, potentially including those who are already at the peak of their careers or those who contribute out of a sense of duty rather than a desire to improve skills, might not be looking to develop or improve their skills through OSS. Let's estimate that 2% might strongly disagree.
- I'm not sure: Contributors who have not yet pinpointed their motivations for participating in OSS or those who have a casual relationship with their contributions might be unsure. This could account for around 8% of the population.

These estimations reflect the common incentive among OSS contributors to enhance their skills through practical engagement with projects, collaboration, and facing real-world problems that OSS often presents.