A Prompt Engineering-based Process to Build Proto-personas during Lean Inception

Raul Leão, Fernando Ayach, Vitor Lameirão, Awdren Fontão
Faculty of Computing - Federal University of Mato Grosso do Sul
Campo Grande, MS, Brazil
{raul.leao,fernando.ayach,vitor.lameirao,awdren.fontao}@ufms.br

ABSTRACT

Product discovery approaches such as Lean Inception (LI) typically span five days (40 working hours). During LI, the participants create and refine proto-personas during four working hours to understand user needs. Proto-personas are preliminary, assumption-based representations of ideal users that guide initial design discussions. The accuracy of proto-personas generated in this context has been counterintuitive due to limited time for idea exploration and refinement, for example. There are approaches to building personas (e.g. data-driven, LLMs). However, there is a gap in exploring the use of prompt engineering and proto-persona strategies to support the Product Discovery approaches. Our research investigates the application of a prompt engineering-based approach to building proto-personas during LI. We report an exploratory case study where six participants used our approach to generate proto-personas in a given scenario. The impact of our approach positively influenced the outcome. Most proto-personas developed by our process better represented the target audience than those from LI, despite some inconsistencies. Our process was well accepted by participants and suggestions were made to improve the process. Our approach used an average of 11 minutes of working hours (SD 2.24 minutes), traditionally this time in LI is four hours.

KEYWORDS

prompt engineering, proto-persona, lean persona, product discovery, lean inception

1 INTRODUCTION

Product Discovery approaches stand as a pivotal stage in Requirements Engineering (RE), aiming to decipher user needs and delineate clear initial directions for the project. Lean Inception (LI) stands out among existing approaches, spanning five days (40 working hours). The conduction of product discovery activities helps deliver features that fit customer and business goals [22]. Within LI, participants create and refine proto-personas, allocating a focused four-hour window to grasp user requirements. Proto-personas or Lean Persona can be understood as an initial representation of a target user, based on assumptions and preliminary data, used to kickstart product development before more detailed research [8].

The accuracy of these proto-personas, crafted within the constraints of time and resource limitations, may be counterintuitive [17]. In Product Discovery approaches, tight timelines necessitate rapid creation and iteration of proto-personas. Engaging with end-users and domain experts for persona data can be resource-intensive and challenging due to limited user data in the early stages. Ensuring high accuracy and relevance in proto-personas is crucial for effective product discovery. LLMs (Large Language Models) have become tools that could support software engineering activities, especially in RE [15]. An important process for using LLMs like ChatGPT is formulating appropriate prompts [7]. A prompt can be understood as a set of instructions provided to an LLM, programming it by customizing its capabilities [13]. It influences the output generated by an LLM by providing specific rules and guidelines. Prompt Engineering is how LLMs are systematically programmed via prompts to optimize their results.

In the systematic mapping study on the use of personas in RE, Karolita et al. [10] did not identify any studies involving the use of prompt engineering for the construction of proto-personas. The authors recommend new approaches to creating personas. Marques et al. [15], when exploring the use of ChatGPT in RE, did not identify studies in the context of proto-personas. The authors highlight the need to explore techniques for different stages of software development as a way to support quality improvement in approaches based on prompt engineering.

We perform an exploratory case study involving six participants to advance towards a prompt engineering-based process to build proto-personas in a Product Discovery approach (i.e., LI) based on limited input, facilitating initial ideation and validation. Our unit of analysis is a project for a mobile application to support the dynamics of producers and consumers in cooperatives: "*after 14 months of development, a usability test with 11 users revealed a previously unidentified persona (during the LI phase and Sprint Reviews), necessitating costly redesigns including a UX workshop and interface modifications. The presence of only one UX Researcher led to task overload due to the high demand for revisions caused by the unaccounted-for persona*".

2 RELATED WORK

There have been approaches for persona development, including data-driven personas using media data [9], persona design guidelines tools [12], and personas based on knowledge graphs [25]. However, despite their potential, there remains a dearth of demonstrated value and effectiveness in integrating persona systems into User-Centered Design (UCD). Furthermore, with the rapid advancement of Large Language Models (LLMs) like ChatGPT, leveraging LLMs for persona development has become increasingly significant across various domains such as software engineering [1], UI design [2], and education [16]. And also the existing automated system for generating personas based on LLMs [24]. Nevertheless, there is a lack of exploring the synergy between prompt engineering and proto-persona strategies to support Product Discovery approaches.

In the existing secondary studies on personas in Requirements Engineering [10], personas in Agile methodologies [14], use cases for design personas [21], LLMs for Software Engineering [7] and ChatGPT in Requirements Engineering [15], no studies are addressing prompt engineering for the generation of proto-personas in the context of Product Discovery. Considering this, it is worth investigating due to the impacts on projects generated by failures in building proto-personas.

3 THE APPROACH

Our approach was modeled based on the Prompt Patterns cataloged by White et al. [23] and OpenAI[1]. The approach was refined through pilot studies by the authors using the ChatGPT 3.5. Four versions of the model were created. For each version, we performed a business process modeling (BPMN) and a formal description[2], containing details of each activity: name, responsible, prompt pattern used, goal, prompt to be inserted, and expected ChatGPT response.

The current version, Figure 1, comprises two events, five activities, and three artifacts. The process begins with accessing the LLM (ChatGPT) and ends with building a set of proto-personas. Additionally, based on the conducted tests, activities 2, 3, and 4 do not necessarily need to be performed in that order but were defined as such to establish a unified process. Activities 1 and 5 need to be respectively at the beginning and end, as they represent the purpose of the prompts and the actual generation.

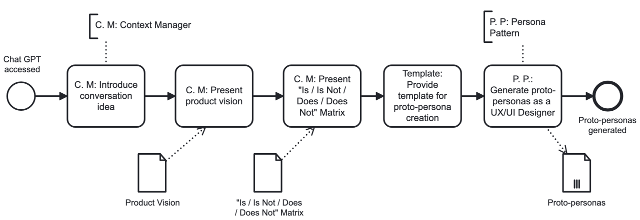

Figure 1: Our proposed approach

Activity 1: *Introduce conversation idea.* The first prompt aims to establish an information context for the LLM and limit its response for proto-persona generation in a Lean Inception-based product discovery. ***The prompt pattern used was the Context Manager, which aims to restrict the AI analysis to a specific context.*** This pattern is crucial in this initial stage because ChatGPT is a generic text generator, and we need to guide the model's line of reasoning to work with information from our context, avoiding hallucinations and off-topic responses [3]. This step uses the product discovery context, which will guide subsequent activities.

Activity 2: *Present product vision.* In this, we input the product vision as a prompt for the LLM, aiming to restrict the LLM's response to the context of the software product that includes the product vision and adds more application context. ***Once again, we utilize the Context Manager as the prompt pattern***, given its similarity to the previous topic. This step uses the product vision input for subsequent activities.

Activity 3: *Present "Is/Is Not/Does/Does Not" Matrix.* This is the final contextual activity of the model and aims to input the "Is/Is Not/Does/Does Not" LI matrix as a prompt for the LLM, restricting the LLM's response to the context of the software product. Similar to the previous activities, ***it also uses the Context Manager*** and is equally necessary for proto-persona generation in the last stage of the process, as notions of the product vision and its functionalities are essential for proto-persona conception. This step uses input on what the product does and does not do for subsequent activities.

Activity 4: *Provide the desired template for proto-persona creation.* After contextualization, we supply the LLM with the template of the proto-persona we want to generate, in this case, that of Lean Inception (profile, needs, and behaviors), to ensure that the generated response aligns with the LI's specification. ***We use the Template Pattern as the prompt pattern*** because it fulfills the function we desire for this activity: formatting the generated output into a specific format. This step prepares the proto-persona template to be used in the last activity of the process.

Activity 5: *Ask the LLM to act as a UI/UX Designer with experience in proto-persona creation and request that it generate the proto-personas for the software.* Finally, here we generate the artifacts, making the LLM act as an experienced professional in building software proto-personas. ***The prompt pattern used is the Persona Pattern***, aiming to incorporate a role into the LLM, causing it to assume (or at least attempt to assume) the line of thinking of the specified role. It can improve the focus of the model's line of reasoning. This last step yields the proto-personas of the specified product, based on the inputs of product vision, "Is/Is Not/Does/Does Not" LI matrix, and proto-persona template.

Reflections on the approach conception: The approach went through four versions, and certain activities were removed during refinements. One of these was the validation activity, which aimed to refine the proto-personas generated via a prompt. We attempted to use an approach without additional context inputs (since the idea is to use minimal information possible, anticipating generation) based on a method called validation document [24]. This method suggests formulating questions to be answered about the generated proto-personas to validate them about the product in question. Considering this context, we attempted to formulate a prompt based on the Reflexion pattern, where the LLM created questions about the proto-personas it generated, answered them, and applied improvement points to an updated set.

The outcome was a reiteration of information from the approach, exhibiting negligible deviations from the initial dataset. We preliminarily concluded that, without supplementary inputs (such as additional context or feedback from experts or end users), conducting a validation step for the dataset is unwarranted.

It is important to highlight that the final result, based on testing, is sensitive to how prompts are formulated. An example is the addition of the prompt technique Chain Of Thought - Zero shot, which suggests using the phrase *"Let's think step by step"* to improve the LLM's response in step-by-step algorithms, such as in the generation of proto-personas in our process. However, when implementing this, instead of possibly improving the development of artifacts, the ChatGPT returns only one proto-persona, and we constantly need to ask for more to generate all of them.

[1]https://platform.openai.com/docs/guides/prompt-engineering.
[2]https://tinyurl.com/bdfc25dk.

4 CASE STUDY DESIGN

We used the guidelines of Runeson et al. [19] to conduct and report an exploratory case study, which is applied in the search for new insights, generating ideas and hypotheses for research when there is little information about the phenomenon being studied.

4.1 Case and Units of analysis

The case analyzed in this study refers to the use of a prompt engineering-based process to build proto-personas within a mobile application project during the initial phases of a product discovery approach. In this project, Lean Inception is used as the discovery approach to generate proto-personas (four working hours to diverge and converge during the proto-personas building and refinement). ***It is characterized as a single case*** because we are exploring a phenomenon in a specific context. ***As the unit of analysis***, we used a mobile application project that supports the Solidarity Economy development. The mobile application was developed by a team of 12 software engineers and three specialists in the fields of administration and economics. As part of the product discovery phase, the Lean Inception [5] was used, generating proto-personas in this process. After 14 months of development, a usability test was conducted with 11 users, revealing a previously uncataloged persona during the Lean Inception phase and subsequent project progression and persona refinement. This discovery necessitated a comprehensive redesign of the application, including a UX workshop and modifications to workflows and interfaces. Such changes are exceedingly costly for any project and could have been mitigated with a precise initial definition of the system's personas.

Furthermore, it is important to note the presence of a single UX researcher in the unit of analysis, leading to an overload of tasks for this professional due to the high demand for revisions prompted by the lack of persona mapping identified in the usability test. This situation underscores a broader issue of insufficient human resources for tasks such as persona formulation, which is likely a challenge not only for this analysis unit but also for other projects.

4.2 Rationale

Given the case described above, the use of external tools such as LLMs (e.g. ChatGPT) can be useful to at least minimize such problems [1]. However, the casual use of tools like ChatGPT is not sufficient for the development of more complex tasks, such as persona generation [15]. Therefore, prompt engineering serves as a means to maximize the effectiveness of the approach's response [13, 23], and our process focuses precisely on this in this case study.

4.3 Research Question and Study's Goal

Our goal in this case study is to investigate the use of "our process" to enhance the quality of the proto-personas designed during the proto-persona generation activity in a Lean Inception. To support it we developed the following research questions and metrics:

RQ1: How effective is our proto-persona generation process? **Metric 1:** Similarity (calculated using cosine similarity technique [11] and qualitative manual verification) between the proto-persona set generated by "our process" and the conventional Lean Inception activity. **Metric 2:** Similarity between the proto-persona set generated by "our process" and the validated project personas.

RQ2: What is the acceptance of "our process" of proto-persona generation? **Metric 1:** Perceived Usefulness (PU). **Metric 2:** Perceived Ease of Use (PEoU). **Metric 3:** Attitude Toward Use (ATU). **Metric 4:** Behavioral Intention to Use (BItU).

4.4 DATA COLLECTION & EXECUTION

At the end of each study session, we collected data regarding the set of generated proto-personas and the evaluation of the process. This evaluation was carried out by the participant during and after the study. During the study session, the researchers made notes related to the application of the think-aloud method.

4.4.1 Artifacts. **As a preliminary phase to the study execution**, we collected and refined the artifacts produced during Lean Inception, including the "Product Vision" and the "Is/Is Not/Does/Does Not" LI matrix of the unit of analysis. They are essential inputs for the process execution. Natalia Arsand developed the persona description template used [4], which serves as a reference during the "Describe the persona" stage of Lean Inception. We chose to have the artifacts ready for use by the participants, as our evaluation did not aim to verify the collection of inputs before the process execution. All the data should be available during the evaluation of LLMs [20]. Then along the paper, we share the artifacts link. **The process evaluation questions** were based on the Technology Acceptance Model (TAM). For this study, we chose the most relevant variables for evaluating our process using a five-point Likert Scale: Perceived Usefulness (PU), Perceived Ease of Use (PEoU), Attitude Toward Use (ATU), and Behavioral Intention to Use (BItU).

4.4.2 Participants selection. Participant selection was based on the following categories: *(a) Academic professional with experience in User Research* (*P3* - PhD. 16 years of industry experience. 11 years as a researcher, *P6* - PhD in progress. 24 years of industry experience); *(b) Industry professional with experience in User Research* (*P1* - Specialist. 6 years of industry experience); *(c) Newly integrated member of the unit of analysis* (*P4* - Undergraduate in progress. 4 months of experience); *(d) Experienced member of the unit of analysis* (*P2* - Undergraduate in progress. 1 and a half years of experience); and, *(e) Domain expert of the unit of analysis* (*P5* - PhD. 7 years of experience). The categories were chosen based on the experience variable, both about the project and knowledge of User Research. We opted for this type of sample stratification to reflect the usage scenarios of the process by different project members (following Rainer and Wohlin recommendations [18]): those with UX/UI expertise, with domain experts, or with members without experience in the previous two criteria. Additionally, we chose an industry and an academic User Research Specialist to gain insights into the process.

We conducted the study with six participants, each representing one of the categories mentioned earlier. The study took place both online and in person. Before conducting the study, we provided a form to the participant[3], which contained the consent form, a field to submit the proto-personas generated in the process, and the evaluation questions.

4.4.3 Study execution. We provided the participant with an execution script[4], a document that contains the prerequisites to start the

[3]https://tinyurl.com/mtnkb3ys.
[4]https://tinyurl.com/yessmtpz.

process and the activities to be performed. Similarly, we prepared a script to support the researchers[5]. We conducted the execution of the process with one participant at a time and with two responsible researchers. One researcher guided the process, explaining the study proposal, while the other was designated for observation and note-taking. To enrich the observations, we used the think-aloud method, whereby we instructed the participant to narrate their actions, thoughts, and emotions during the process. We recorded audio, screen, and the total execution time.

4.4.4 Data Analysis. **To address RQ1**, we grouped the sets of generated proto-personas, project proto-personas/personas. Based on the identified similarities, we conducted two comparisons: (1) a comparison between the generated proto-personas and the project proto-personas, and; (2) a comparison between the generated proto-personas and the project personas.

Calculating the similarity between the set of proto-personas generated by "our process" and the conventional Lean Inception activity: Firstly, we quantitatively assessed how much the generated proto-persona text resembled that of the project proto-persona, i.e., human-generated content, (if it was covered by the generation). To do this, we chose one of the most common: cosine similarity [11], an overall efficient method in text mining tasks which provides a value from 0 to 1 relative to the equivalence between two texts. While this mechanism already gives us an idea of similarity, we decided to perform a qualitative manual comparison between the artifacts because relying solely on pure text similarity can lead to false negatives, as a proto-persona may represent the same context using different words. In this way, we chose cosine similarity in the first place as a complementary automatic measure to supplement RQ1 similarity metric. The manual verification aimed to observe discrepancies between the two artifacts, analyzing similarities between the information contained in the proto-persona profile (age range, profession, and education), needs (pain points and means to alleviate them), and behaviors (daily activities, hobbies). If some generated proto-persona was not a LI-covered one, we analyzed if it made sense in the project's domain.

Calculating the similarity between the set of proto-personas generated by "our process" and the validated personas of the project: The procedure conducted here was the same as the comparison with the project proto-personas. However, this comparison is more critical than the previous one, as we are verifying whether what our process generated aligns with a validated final artifact of the project, an analysis that more clearly demonstrates the effectiveness of the generation.

To answer RQ2, after assuring the similarity between the LLM generated proto-persona and the project one during RQ1, we collected data from the TAM questionnaire and analyzed each statement. It aims to understand if there were trends toward agreement, disagreement, or neutrality. In the case of agreement, we verified if there was partiality or full agreement. In the case of disagreement, we discussed hypotheses that could generate the disagreement, based on qualitative data (think-aloud observations and participant feedback on the process) collected during the studies, extracting refinement ideas for the process from this. For cases of neutrality, we analyzed the reasons behind the participant's non-positioning, also based on the other qualitative data mentioned. We also calculated the average execution time of the proto-persona generation process and compared it with the time for the same activity in the conventional LI activity. We also analyzed feedback from the process, the TAM questionnaire, and the proto-persona generation. We extracted transcription units, preliminary codes, and axial codes [6] to understand the relationship between participant feedback and process metrics (e.g., proto-persona quality, ease of use). Interviews were conducted and coded by two or more authors over several iterative cycles.

4.5 Results' Analysis and Discussion

We extracted preliminary and axial codes[6] [6] from think-aloud, TAM questions and proto-personas feedback from the collected qualitative data (Table 1). The analysis reinforced Likert scale results. We found that there was some little confusion about the process (as Table 1 exemplifies), but nothing that prevented the participants from executing with the facility. The axial codes from the proto-persona feedback revealed acceptance from the participants, but some inconsistencies were discovered (RQ1).

Transcription Unit	"Asking LLM to act as... what is LLM?"
Preliminary Code	Doubt about the definition of LLM.
Axial Code	Questions about definitions in the script

Table 1: Code extraction example - Participant 2

4.5.1 (RQ1) How effective is "our proto-persona generation process? **Similarity between the set of proto-personas generated by "our process" and the conventional Lean Inception activity:** The unit of analysis involved the development of two proto-personas in LI: an institution coordinator (demander) and a small/medium-sized production farmer (offeror). The first proto-persona was covered in all 6 executions (100%). The cosine similarity ranged between 24.87% and 46.6%, with an average of 35.57% similarity. The second proto-persona was also covered in all 6 executions (100%). The cosine similarity ranged between 29.83% and 50.66%, with an average of 43.26% similarity. In total, 20 proto-personas were generated, and we identified 5 main categories: "Offeror" (6), "Demander" (9), "Supporter of the Solidarity Economy" (1), "Community Leader" (1), and "Conscious Consumer" (3). Out of the 5 categories, 2 proto-personas ("Demander" and "Offeror") were covered by the LI. Of the remaining 3, "Supporter of the Solidarity Economy" and "Community Leader" are within the unit of analysis domain. The "Conscious Consumer" proto-persona needs further research to validate its relevance in the domain.

Some inconsistencies were found in the generated proto-personas, particularly regarding the acceptance and habitual use of technology. In the unit analysis domain, the farmer persona typically is not accustomed to technology, but ChatGPT often assumes the opposite, e.g., "Active on social networks, searching for technological

[5]https://tinyurl.com/jh6fpa78.

[6]https://tinyurl.com/5n8hyhju.
https://tinyurl.com/3m8y8jw3.
https://tinyurl.com/yc62ps7t.

solutions that facilitate your commercial activity.". However, this inconsistency was not uniform. ChatGPT occasionally addressed the difficulty with technology: "Has a smartphone with internet access, but has limited familiarity with technology." Despite these inconsistencies, the generated proto-personas were generally adequate for the project domain.

Similarity between the set of proto-personas generated by "our process" and the project's validated personas: At all 6 instances of the case study, the generation achieved to cover the unique persona of the unit of analysis: the farmer (offeror). The cosine similarity varied between 42.02% and 54.76%, with an average of 49.66% similarity, which represents a moderate coverage of the validated persona. In the manual verification, we found out that the proto-personas are within the unit of analysis context. Even though the generation was not the same as the validated persona, it showed relevant topics in the domain. However, the inconsistency in proto-persona behavior regarding technology use, as previously noted in comparison with LI results, remains and affects the accuracy of the real persona description. Despite not being a validated persona for the project, the generation process highlighted the role of "Demander", which we believe could be a viable persona for the application. We found that the generated proto-personas were closer to validated personas than those developed by LI.

4.5.2 ***(RQ2) What is the acceptance of "our process" of generating proto-personas?*** In this section, for each TAM topic, we present the results of the Likert scale and perform an analysis to understand the level of agreement.

Perceived Usefulness (PU): Using our process makes it easier for you to build new proto-personas in your projects: Five participants (83%) agreed and one participant (17%) disagreed. **Analysis:** By analyzing the feedback provided by P1 who expressed disagreement regarding the process and outcome of the generation, additional efforts are required to review our process results. Our process may not consistently serve as a significant facilitator. P1 made an analogy of LLM with an intern, whose inputs need to be validated: "*So, at first, I have to see him as my intern who is creating for your persona, and then I will Check it out. So, at first, it helps me with the process, but it will also give me some work later.*". P3 expressed a similar opinion regarding the response verification stage, although he/she agreed that it depends on a case-by-case basis: "*So we have to very critically analyze the responses received. So sometimes it can even make it difficult, a little bit. But depending on the task, if it's simpler, it makes it easier.*".

We understand that our process, especially in more complex projects, which require a more detailed specification of the proto-personas, may require work after generation since ChatGPT responses are not always 100% coherent with the given context. This review stage was already a hypothesis we had regarding our process. In other versions of the process model, we tried to insert this validation via prompt, but we did not obtain any improvements in the results. Therefore, this activity after the generation of proto-personas is still an open topic to be explored in other works.

Using our process speeded up the process of building proto-personas: four participants (66%) totally agreed and two (34%) partially agreed. **Analysis:** The average execution time of our process is 11 minutes (SD 2.24). Compared to the 4 hours of the LI, it represents a considerable time saving: 95.42% faster. P5, who agreed, expressed: "*... I can't say it made it faster in terms of finalizing the process, but in terms of obtaining proto-personas it made it easier, yes.*". P1 and P3 expressed similar opinions.

Even if the results are insufficient for some projects, the time savings justify using generated proto-personas. Although generation may be inconsistent, it provides a solid starting point for development. It allows professionals in Lean Inception (LI) to focus on refining content rather than creating proto-personas from scratch. Thus, project members can quickly obtain initial proto-personas and use the remaining time to refine them, accelerating persona development. Alternatively, the generated proto-personas could be deemed sufficient, enabling the team to proceed to the next LI activities, such as developing user journeys.

Perceived Ease of Use (PEoU): Understanding the instructions of our process was easy for me: five participants (83%) totally agreed and one (17%) partially agreed. **Analysis:** Our goal with the process was to make it as simple and clear as possible. Observing the qualitative data, the agreement given by the scale was reinforced. P3 highlights: "*Understanding this was very easy.*". P5 reinforces: "*Learning the instructions of our process was easy for me, totally. The instructions themselves were easy.*". P1, the only one who partially agreed, defends: "*Again, there were the points that I mentioned, which seemed like a lot. I agree that I had the help, but totally not because it was quite extensive, you know...*". There were some doubts about understanding the process pointed out by P6, as described in "*Now the last activity, asking him to generate the proto-personas, again, right, here he is, right now, it's because it hasn't been requested yet, now he's going to ask.*", but as the transcription shows, the participant understood the instruction by himself/herself later.

Attitude Toward Use (ATU): **I believe that using our process is a great idea for generating proto-personas:** two participants (34%) agreed and four participants (66%) partially agreed. **Analysis:** Although some participants reported inconsistencies and indicated refinements in the process, everyone agreed that its use is a great idea for a proto-persona generation. P5 said: "*... highlighting that the process has to be refined to have some points, or at least add some related information, such as these warnings, etc.*" Another factor that led to a partial agreement was distrust in the quality of results in more complex contexts, as P4 said: "*So if it were a more complex project, maybe. That's why I don't know, I don't totally agree...*". P3 reinforces the uncertainty in the process: "*... It's a belief, it's not a certainty.*". The participants liked the idea and result of the generation, as reinforced by P6: "*... I think he did it right, I thought it looked cool.*", but many tests remained, and a single execution of the process was insufficient to ensure certainty.

I believe it is much better to use our process, rather than a classic proto-persona generation process: We observed a tendency towards neutrality: four participants (66%) were neutral, one participant (17%) disagreed and another (17%) totally agreed. **Analysis:** It was the question with the most variations in responses from participants, most likely due to the weight of the statement. Our hypothesis about the four participants (66%) who presented neutrality is that, despite having liked the results of the process, the participants did not have experience using our process in other projects, thus, they were not sure whether the quality of the results would hold in different cases. P5 expresses this idea: "*I think that depending on the way the process classic is conducted, we manage to generate better proto-personas.*". P3 was the one who disagreed with this statement, with the argument that our process does not replace a classic process: "*... if I have something more complex, it can generate very simple results. So, sometimes, you can use it as an auxiliary and complementary platform in the creation process, and not just exchange one for the other.*". Thus, from the previous report, it is understood that our process would be more useful as a support tool, an opinion shared by P1 and P5. P2 believes it is better than a classic activity: "*For other people, it's a bit boring to do a classic process.*". We can infer that our process can include those people who do not like classic activity in LI, which contributes to generating proto-personas that they would not be able to carry out the LI activity, probably in a scenario with inexperienced members.

I like to use AI tools and processes to help with my UX activities: We observed a tendency towards partial agreement: one participant (17%) totally agreed and five (83%) partially agreed. **Analysis:** The partial agreement observed is due to the participants' view of AI tools as more of a support and with use with certain reservations. All six participants believe these tools and processes are welcome, but there should not be too strong an attachment to them, as tools like ChatGPT do not guarantee that the answers are 100% true. P4 expresses this frustration in: "*I just won't totally agree, because I'm still a little afraid of the tools.*". P1 has a similar opinion: "*I'm not going to give a complete one because, man, these tools aren't yet... The ones we have on the market today, especially the design part, aren't that good yet, but the business has been evolving quite.*". The perception of AI as something more auxiliary (P3, P5).

Behavorial Intention to Use (BItU): I intend to use our process whenever possible: five participants (83%) agreed, of which 3 (50%) totally agreed and 2 (33.33%) partially agreed. Furthermore, one participant (17%) was neutral. **Analysis**: even those who had their reservations regarding generation, for example, concerns in more complex cases, agreed with the statement, given the openness that the issue brings. P5 reinforces this idea: "*I agree. Given the way it is, I still have maybe a few saved, as I said, but it's easy and if I'm going to put it in whenever possible, I think so, it's actually quite simple to use.* ". However, 1 participant (17%) was neutral. The participant's qualitative data indicate that he/she does not frequently perform the persona generation activity. It combined with P1's knowledge in generating proto-personas, suggests the hypothesis he/she is unsure whether he/she needs the process. Another statement from P1 highlights this uncertainty: "*We are clear that there was no need to carry out a new study. So, it depends, okay? So I neither agree nor disagree*.". P1 considers using the process in an academic scenario.

I would adopt new tools similar to our process in the future: Five participants (83%) totally agreed and one (17%) partially agreed. **Analysis:** We observed that the participants liked our process, which makes it consistent that the vast majority (83%) totally agreed with the statement. P3 highlights: "*I totally agree. If there are tools that are reliable that use AI, obviously I will want to use them*.". Some participants also highlighted that they would like to adapt our process for their projects and were excited by the ease and speed it brings. P3 himself comments: "*I intend to use it and I even intend to adapt it.*". Even those who were more critical, like P1, expressed: "*I personally if I had the opportunity, I would even recommend this system of yours.*". P5, who commented on process refinements, said: "*Yes, good tools are always welcome. I would adopt new tools similar to our process in the future, yes, for sure.*". Therefore, we analyzed that the proposal of our process has the potential to be a good alternative for generating proto-personas in a Lean Inception. Just some little additions have to be made, as including pictures to the generated proto-personas, as suggested by P6: "*I know personas, they usually have some other characteristics, they have an age, I don't know, other additional information, it usually generates a photo, you could work on that from there, but I think it turned out pretty well.*".

5 LIMITATIONS

Given the specific nature of the unit of analysis and the participants involved, some findings are likely specific to this context. The application and its user interactions present unique characteristics that may not generalize to other types of applications or user groups. Additionally, the proto-persona generation process using Lean Inception and prompt engineering may yield different results in other settings or with different facilitators. The reliance on self-reported data from participants also introduces the potential for bias.

6 CONCLUSION AND FUTURE WORK

Most proto-personas developed by our process better represented the target audience than those from LI, despite some inconsistencies in technological behavior. The generated proto-personas were quantitatively closer to the validated persona, with a moderate cosine similarity ranging between 42.02% and 54.76,% and an average of 49.66%. ChatGPT's responses varied and sometimes diverged from the domain, but the process sped up the initial creation, allowing more time for team discussion and refinement, which enhanced the time spent on creative activities. Our process was well accepted by participants, who generally agreed with most TAM questions. Due to limited testing, there was neutrality on this point. There are limitations such as handling critical inconsistencies or complex projects with very brief proto-personas.

Future Work: It is essential to continue reflecting about the usage of LLMs in RE and Product Discovery. We highlight the research involving the analysis of conformity of the results with the requirements and business rules of the projects. Exploring other strategies to evaluate similarity. Adapting the process to receive different types of inputs besides product vision and the "Is/Is not/ Does/ Does not" matrix. Adapting the process to other LI activities, such as user journeys. Applying tests and refinements to the process, as well as using other AI models, such as ChatGPT 4.0. Try to specify more clearly the persona template, including age and picture. Although our process has been refined over several cycles, we emphasize the importance of comparing different types of prompts to evaluate and improve their effectiveness. Additionally, it is necessary to adopt more rigorous validation mechanisms to enhance the approach. Finally, it is vital to recognize the importance of exploring our approach in cases and projects that are more complex than the one presented.

REFERENCES

[1] Chetan Arora, John Grundy, and Mohamed Abdelrazek. 2023. Advancing requirements engineering through generative ai: Assessing the role of llms. *arXiv preprint arXiv:2310.13976* (2023).

[2] Stephen Atlas. 2023. ChatGPT for higher education and professional development: A guide to conversational AI. (2023).

[3] Lenz Belzner, Thomas Gabor, and Martin Wirsing. 2023. Large language model assisted software engineering: prospects, challenges, and a case study. In *International Conference on Bridging the Gap between AI and Reality*. Springer, 355–374.

[4] Paulo Caroli. 2015. *Direto ao ponto: criando produtos de forma enxuta*. Editora Casa do Código.

[5] Paulo Caroli. 2017. Lean inception. *São Paulo, BR: Caroli. org* (2017).

[6] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. sage.

[7] Angela Fan, Beliz Gokkaya, Mark Harman, Mitya Lyubarskiy, Shubho Sengupta, Shin Yoo, and Jie M Zhang. 2023. Large language models for software engineering: Survey and open problems. *arXiv preprint arXiv:2310.03533* (2023).

[8] Jeff Gothelf. 2013. *Lean UX: Applying lean principles to improve user experience.* " O'Reilly Media, Inc.".

[9] Soon-Gyo Jung, Jisun An, Haewoon Kwak, Moeed Ahmad, Lene Nielsen, and Bernard J Jansen. 2017. Persona generation from aggregated social media data. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 1748–1755.

[10] Devi Karolita, Jennifer McIntosh, Tanjila Kanij, John Grundy, and Humphrey O Obie. 2023. Use of personas in Requirements Engineering: A systematic mapping study. *Information and Software Technology* (2023), 107264.

[11] Baoli Li and Liping Han. 2013. Distance weighted cosine similarity measure for text classification. In *Intelligent Data Engineering and Automated Learning–IDEAL 2013: 14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013. Proceedings 14*. Springer, 611–618.

[12] Jianing Liu, Jia Shi, Jun Xie, Xinyun Zhang, Zichuan Zhang, John Grundy, and Tanjila Kanij. 2022. A curated personas and design guidelines tool for better supporting diverse end-users. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 1606–1613.

[13] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.

[14] Patricia Losana, John W Castro, Xavier Ferre, Elena Villalba-Mora, and Silvia T Acuña. 2021. A systematic mapping study on integration proposals of the personas technique in agile methodologies. *Sensors* 21, 18 (2021), 6298.

[15] Nuno Marques, Rodrigo Rocha Silva, and Jorge Bernardino. 2024. Using ChatGPT in Software Requirements Engineering: A Comprehensive Review. *Future Internet* 16, 6 (2024), 180.

[16] Anh Nguyen-Duc, Beatriz Cabrero-Daniel, Adam Przybylek, Chetan Arora, Dron Khanna, Tomas Herda, Usman Rafiq, Jorge Melegati, Eduardo Guerra, Kai-Kristian Kemell, et al. 2023. Generative Artificial Intelligence for Software Engineering–A Research Agenda. *arXiv preprint arXiv:2310.18648* (2023).

[17] Eduardo Gouveia Pinheiro, Larissa Albano Lopes, Tayana Uchôa Conte, and Luciana Aparecida Martinez Zaina. 2018. The contribution of non-technical stakeholders on the specification of UX requirements: an experimental study using the proto-persona technique. In *Proceedings of the XXXII Brazilian Symposium on Software Engineering*. 92–101.

[18] Austen Rainer and Claes Wohlin. 2022. Recruiting credible participants for field studies in software engineering research. *Information and Software Technology* 151 (2022), 107002.

[19] Per Runeson and Martin Höst. 2009. Guidelines for conducting and reporting case study research in software engineering. *Empirical software engineering* 14 (2009), 131–164.

[20] June Sallou, Thomas Durieux, and Annibale Panichella. 2024. Breaking the silence: the threats of using llms in software engineering. In *Proceedings of the 2024 ACM/IEEE 44th International Conference on Software Engineering: New Ideas and Emerging Results*. 102–106.

[21] Joni Salminen, Kathleen Wenyun Guan, Soon-Gyo Jung, and Bernard Jansen. 2022. Use cases for design personas: A systematic review and new frontiers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–21.

[22] Stefan Trieflinger, Dominic Lang, Selina Spies, and Jürgen Münch. 2023. The discovery effort worthiness index: How much product discovery should you do and how can this be integrated into delivery? *Information and software technology* 157 (2023), 107167.

[23] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).

[24] Xishuo Zhang, Lin Liu, Yi Wang, Xiao Liu, Hailong Wang, Chetan Arora, Haichao Liu, Weijia Wang, and Thuong Hoang. 2024. Auto-Generated Personas: Enhancing User-centered Design Practices among University Students. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[25] Xishuo Zhang, Lin Liu, Yi Wang, Xiao Liu, Hailong Wang, Anqi Ren, and Chetan Arora. 2023. PersonaGen: A Tool for Generating Personas from User Feedback. In *2023 IEEE 31st International Requirements Engineering Conference (RE)*. IEEE, 353–354.