



# Improving LLM-assisted requirement validation using persona-based prompting

Jati H. Husen<sup>1\*</sup>, Mira Kania Sabariah<sup>2</sup>, Daffa Hilmy Fadhlurohman<sup>3</sup>, Adam Rafif Faqih<sup>4</sup>

<sup>1,2,4</sup>Bachelor of Software Engineering, School of Computing, Telkom University

<sup>4</sup>Tohkimo, Inc., Japan

<sup>1\*</sup>[jatihusen@telkomuniversity.ac.id](mailto:jatihusen@telkomuniversity.ac.id), <sup>2</sup>[mirakania@telkomuniversity.ac.id](mailto:mirakania@telkomuniversity.ac.id),

<sup>3</sup>[daffahilmanafrizal@gmail.com](mailto:daffahilmanafrizal@gmail.com), <sup>4</sup>[adamrafif@student.telkomuniversity.ac.id](mailto:adamrafif@student.telkomuniversity.ac.id)

**Abstract:** Requirements validation is a crucial activity in requirements engineering. Large language models (LLMs) have been proposed to improve the efficiency of requirements validation activity. However, the LLMs validate in a general manner without considering the stakeholders' characteristics of the software development projects that will use the requirements. In this paper, we propose PReVAL, a novel persona-driven LLM-based validation approach that utilizes personas to reflect the characteristics of stakeholders in the prompting of LLMs. A comparative experiment demonstrates the benefits of PReVAL in terms of consistency in validating requirements based on the persona. However, introducing a persona also increases unnecessary aspects of the validation result. These findings contribute to the software engineering body of knowledge by introducing a new approach to subjectively validating software requirements with reduced expert involvement.

**Keywords:** Requirement validation, Large language model, Software requirements

## 1. INTRODUCING

Requirements are important factors in ensuring the success of software engineering activities [1]. It serves as the basis of the correctness of the developed software, both functionally and non-functionally. Failure to provide adequate requirements is considered one of the most critical sources of failures in software development projects [2]. That importance leads to the validation of requirements as one of the key activities in requirements engineering, the phase of software development focused on handling the requirements. Requirement validation ensures that the requirements are not only correct based on the needs of the stakeholders but also the understandability of the requirement specifications for them [3].

One of the approaches to validate requirements is perspective-based reading [4]. This validation technique emphasizes subjective correctness and understanding from each stakeholder, both external and internal. Perspective-based reading allows the specification to be validated based on the specific preferences of the stakeholders rather than generalized metrics, leading to tailored understandability that fits the project [5]. Tailored understandability minimizes the risks of the requirement specifications not being written to the stakeholders' personal preferences despite scoring high on the requirements of quality metrics. However, this technique requires the involvement of the stakeholders in the implementation, making the implementation logistically challenging [5].

Since its rise to maturity, large language models (LLMs) have been utilized to tackle the logistical problem of providing artificial expertise [6]. This situation is also true in requirements engineering, including requirements validation activities. However, the approach for validating requirements using LLM tends to focus on the standardized quality of specification quality [7]. This situation causes the problem of providing requirement specifications that fit the project stakeholders' preferences to persist despite LLM-based support in requirements engineering. An approach to utilize LLMs in a way that allows them to consider different characteristics of stakeholders' preferences is necessary to solve the problem.





This paper proposes PReVaL (Persona-driven Requirement Validation with Large Language Models) as a novel approach in LLM-driven requirement validation to solve that problem by utilizing persona in LLM prompting. A persona-based prompting is an LLM prompt engineering pattern where the LLM agent is assigned a role in completing its task. In this case, the persona assigned is based on the persona of the stakeholders, highlighting their preferences towards the requirement specification. This strategy allows the LLMs to act based on the provided persona, allowing them to mimic a role or a person and reducing the need for experts to implement perspective-based requirements validation. This paper provides the concept of persona commonly used in requirements engineering and user interface and experience design to capture the characteristics of different project stakeholders.

The utilization of LLMs in requirements engineering is a commonly researched area [8]. Several works specifically focused on requirements validation, including Mahbub et al., who explored the capabilities of GPT-4 in detecting ambiguities, inconsistencies, and incompleteness in software requirements [9]. Fantechi et al. conducted a similar preliminary evaluation of inconsistency detection using GPT 3.5 [10]. Arora et al. assessed the role of generative AI, especially LLMs, and stated the possibility of utilizing LLMs to identify potentially overlooked human-centric aspects of the validation [11]. Our approach aims to build on those fundamental works and provide a more technical implementation of LLMs in validating requirements in a human-centric manner.

The usage of persona in LLMs is also a well-discussed field of research. Sun et al. discussed the potential applications of persona in LLMs and its challenges and promises [12]. Hu and Collier investigated the effects of persona in LLM simulations and found significant variance caused by it [13]. Bhandari et al. looked at LLMs' capability to maintain persona throughout the conversation [14]. Our approach utilizes the same concept but, more specifically, requirement validation.

To guide the discussion of this paper, the following research questions (RQs) are developed:

- RQ1. How effective is PReVaL in mimicking stakeholders' concerns in requirement validation? This research question aims to validate the effect of persona-based prompting compared to one-shot prompting in a requirement validation context.
- RQ2. How useful is the validation result generated through PReVaL for stakeholders? This research question aims to validate the correctness of LLM-based validation by measuring stakeholders' acceptance.

This paper contributes to the software engineering body of knowledge, especially in the requirements engineering area, in the form of an automated perspective-based requirement validation approach.

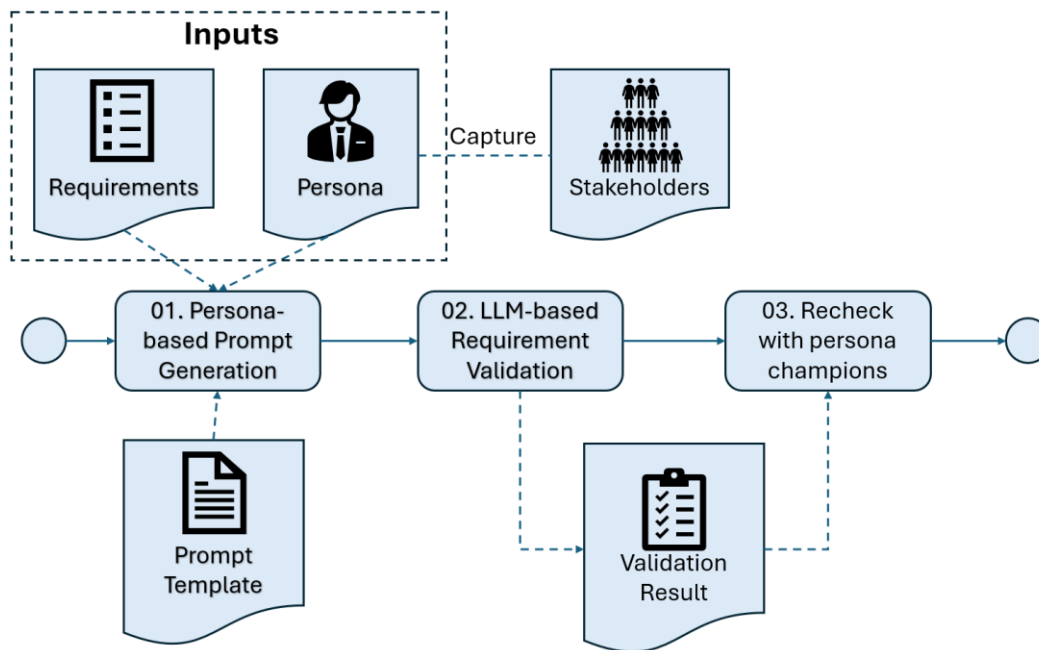
## **2. RESEARCH METHODOLOGY**

This section will discuss the concept of PReVAL and the validation of the approach in two separate sub-sections. The first section will present PReVAL, including its inputs, the design decisions made inside it, outputs, and the need for maintenance to maintain its performance. The second section explains the research methodology designed to evaluate the performance of PReVAL.

### **2.1. Persona-driven Requirement Validation with Large Language Models (PReVAL)**

Figure 1 summarizes the overview of the PReVAL approach, including the expected inputs and the three main processes of the approach. The central concept of the approach is to utilize persona as the representation of one or more individuals from the same group of stakeholders as a guide for the validation inside the LLMs. To achieve that, the persona must first be captured by existing stakeholders. As mentioned in the introduction, the personas in this approach describe the stakeholders as the users of the requirements. This concept means the personas should contain characteristics such as their goals of using the requirements, common challenges in achieving their goals, and stylistic preferences of requirement specifications.





**Figure 1.** Overview of The PReVAL Approach

The requirements to be validated in this approach are specified in a natural language form. LLMs' ability to process free-form texts is important in this part as software requirements tend to be described in semi-structured or completely unstructured. While the current implementation of LLMs as part of more generalized generative artificial intelligence often integrates image recognition technology, it is not part of the current approach. Instead, PReVAL focuses on working with natural language specifications to cover the most used specification approach, regardless of their form.

### 2.1.1. Persona-based Prompt Generation

Prompts are the main denominator of success in utilising LLMs [15]. A systematic approach to designing the prompt is necessary to achieve correct validation using persona-based prompting. Furthermore, a consistent approach to prompting is also necessary. To achieve that, PReVAL includes a template-based approach that facilitates the introduction of persona data into the LLM's knowledge base. The prompt is engineered with zero-shot learning in mind [16]. This design decision is based on the availability of LLM tokens, a common limitation of the current implementation of LLM technology [17].

Fig. 2 presents the prompt utilized in the PReVAL approach, with the parameters of the template indicated using double curly brackets. The parameters are persona data designed based on their importance to software development activities. Those parameters induce the persona data into the LLM's knowledge base. Rather than inducing the persona data in a separate artefact, this approach minimizes the complexity of LLM execution. Ultimately, this approach follows the persona-based prompting approach more faithfully [13].

PReVAL follows LinkedIn's Developer Productivity and Happiness (DPH) Framework to systematically design those persona data parameters, which includes an approach to developing a software developer persona [18]. The parameters include personal background (age, experience, education, responsibilities), working environment (team size, working location, tooling), goals, pain points, values, and further personalization through quotes. While PReVAL does not explicitly define how the data collection and analysis for the developer persona development, a reliable and standardized approach to persona development is recommended to ensure the quality of representation encapsulated through the personas.



You are an expert assistant acting strictly based on the following persona:

Persona Name: **{{Name}}** – **{{Persona Title}}**

Age & Background: **{{Age}}**, with a background in **{{Education/Experience}}**

Current Role: **{{Job Title}}**

Key Responsibilities:

**{{Responsibilities List (3-5 bullets)}}**

Work Environment:

Employment type: **{{e.g., Full-time, WFO/WFH/Hybrid}}**,

Company Size: **{{Company size, e.g., >1,000 employees}}**,

Tools & Platforms: **{{e.g., Jira, GRC, Power BI, Excel, Slack}}**,

Team & Collaboration: **{{Internal/external dynamics}}**,

Goals & Motivations:

**{{Goal 1}}**,

**{{Goal 2}}**,

**{{Goal 3}}**,

Frustrations & Pain Points:

**{{Pain point 1}}**,

**{{Pain point 2}}**,

**{{Pain point 3}}**,

Technical Proficiency & Mindset:

**{{Tool 1}}**, **{{Tool 2}}**, and **{{Tool 3}}** proficiency,

Values: **{{e.g., automation, usability, flexibility}}**,

Representative Quotes:

**"{{Quote 1}}"**

**"{{Quote 2}}"**

---

Your Task:

I will provide you with a list of system or product requirements.

For each requirement, you will:

Validate whether the requirement is clear and relevant for this persona.,

Assess Fit: Based on their goals, frustrations, and tools, how well does this requirement align with their needs?,

Suggest Improvements (if needed): Offer specific rewrites or enhancements to make the requirement more

actionable or aligned with their context.,

Reply with:

[Clear & Relevant],

[Unclear / Needs Improvement] + Suggested rewrite (if applicable),

Optional: Short rationale if helpful,

Wait for the list of requirements before beginning evaluation.

**Figure 2. PReVAL's prompt template. Bold and double curly bracketed indicates parameters to be updated in implementation**



### 2.1.1.2. LLM-based Requirement Validation

The requirements to be validated are prompted into the LLM separately from the initial prompt. This prompt engineering decision is made based on the reliability of the prompting process. By separating the prompts, the correctness of the initial template-based prompting can be validated first prior to the validation of the requirements itself. This strategy minimizes the potential costs when the initial prompting to induce persona to the LLM knowledge base fails, and a reattempt is required.

The output generated from the LLM-based validation comprises two separate parts as prompted by the template in Fig. 2. The first part is whether the requirement is clear and relevant based on the persona. This part is mandatory, and when the requirement is found unclear and relevant, a suggested rewrite based on the persona for the requirement is also generated. The second part is the rationale for validation. Should the decision require further explanation, this part of the output will enlighten the thought process behind the decision.

### 2.1.1.3. Recheck With Persona Champions

Regardless of the quality of PReVAL outputs, all AI-based approaches will carry the risks from data drifts [19]. In this particular approach, the collected personal data may no longer be representative of the actual characteristics of developers in the real world as time goes on. To counter that, the output of PReVAL should be consistently evaluated to monitor the possibility of wrong outputs and the occurrence of performance degradation due to data drifts. Should that happen, an update to existing personas should be conducted to restore the relevancy of the persona data to real-world conditions. The concept of persona champions is borrowed from LinkedIn's DPH Framework, which has the same goal of monitoring changes in the real world [18]. This concept makes monitoring activities more efficient as the champions validate the output rather than the overall population, reducing costs.

## 2.2. Validation Method

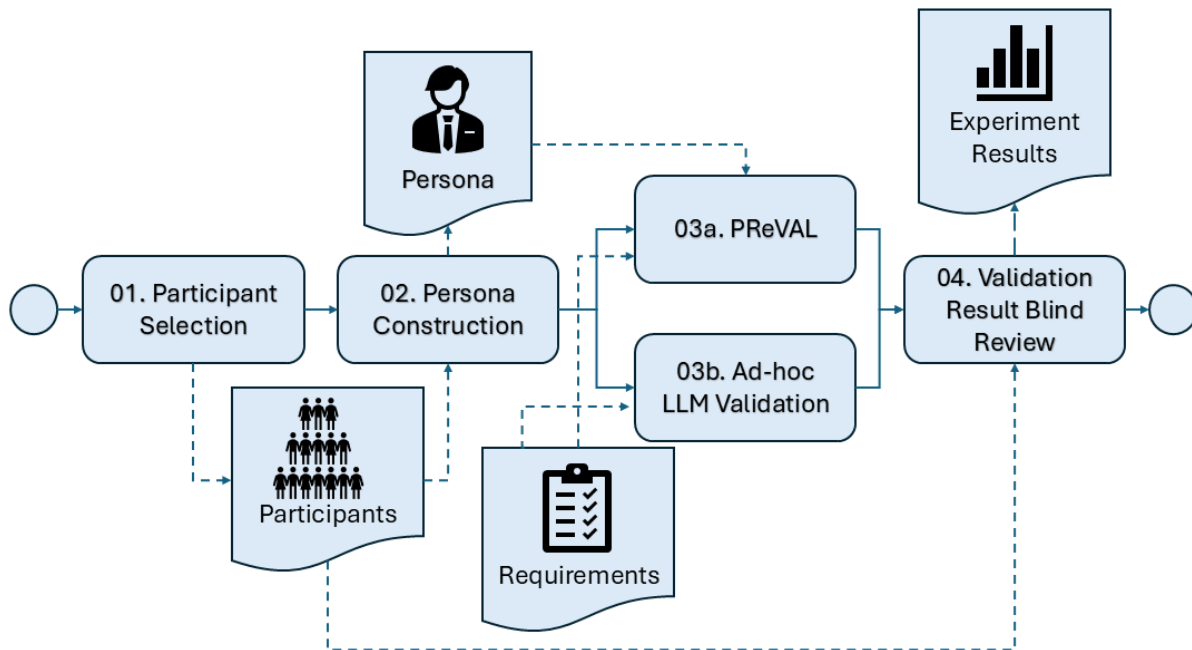
Regardless of the quality of PReVAL outputs, all AI-based approaches will carry the risks from data drifts [19]. In this particular approach, the collected personal data may no longer be representative of the actual characteristics of developers in the real world as time goes on. To counter that, the output of PReVAL should be consistent. A comparative experiment is conducted to evaluate the effectiveness of the PReVAL approach with ad-hoc LLM-based validation directed by the RQs defined in Section I. The experiment is designed to be a single-blind comparative study, where the participants do not know which validation comes from the PReVAL or ad-hoc approach [20]. The experiment is also designed to be an experimental simulation [21]. The experimental simulation brings the advantage of achieving more precision in measuring the participants' behaviour in exchange for less generalizability over the participants' characteristics. The implication of the research design on the result will be discussed further in the latter section.

Figure 3 summarizes the overview of the experiment flow. The experiment starts with the construction of a persona for each participant. A set of requirements is then randomly split in half before PReVAL validates half of them, while the rest are validated by ad-hoc LLM-based validation. On the PReVAL utilization, validation is done for all personas from the participants. Due to its popularity, OpenAI's ChatGPT 4o generates requirement validation. Each participant was then tasked to evaluate the validation result from both PReVAL, with their persona utilized as a part of it, and the ad-hoc approach in a blind fashion.

The evaluation is conducted using a single-item Likert scale for each requirement. The question is, "How much do you agree with the validation of requirement n?" with n notating the requirement number. Using a single-item Likert scale reduces the chance of participant dropouts due to task overload caused by the number of validation results to be evaluated. However, this also introduces another limitation to the result of this experiment that will be discussed further in the latter section. The distribution of Likert-scale responses on each validation result will be used to answer the RQs. The distribution between requirements validation results from PReVAL and ad-hoc approaches will



answer RQ2. On top of that, a top-two box analysis will be conducted to further analyze the answer to RQ1.



**Figure 3.** Overview of the comparative experiment

### 2.2.1. PARTICIPANTS

Eight participants are invited to the comparative experiment. The selection of the participants is conducted randomly except for their work experience. The random selection keeps the probability of selection bias to be low. Table I summarizes the participants' experience and background. Detailed information of the participants is obscured to protect their personal information.

**Table 1.** List of Participants of The Experiment

No.	Job background	Years of Experience
1	QA Engineer	0.5 Year
2	QA Engineer	1 Year
3	Technical QA Engineer	6 Years
4	ODP IT	0.5 Year
5	IT Control & Compliance Analyst	2 Years
6	Application Developer	1 Year
7	IT Solution Architect	3 Years
8	Project Manager	1 Year

### 2.2.2. REQUIREMENTS

A set of requirements from a reliable case is necessary to mimic real-life situations as much as possible. To achieve that, ten requirements from PURE dataset are randomly selected as the requirements to be utilized in the comparative experiment [22]. The random selection is done to reduce the chance of selection bias. The decision to include ten requirements is made with the balance of case representativeness and participants' workload in the experiment in mind. Table II presents the requirements selected to be used in the experiment. Requirements identified with P



indicate requirements to be validated through PReVAL, while N indicates validation through an ad-hoc approach.

**Table 2.** List of Requirements Used in The Experiment

ID	Method Utilized	Requirements
P1	PReVAL	The system shall provide the ability to individualize alerts to address a patient's specific clinical situation.
P2	PReVAL	The system shall associate (store and link) key identifier information (e.g. system ID, medical record number) with each patient record.
P3	PReVAL	The system shall provide the ability to filter, search or order notes by the provider who finalized the note.
P4	PReVAL	The system shall support protection of confidentiality of all Protected Health Information (PHI) delivered over the Internet or other known open networks via encryption using triple-DES (3DES) or the Advanced Encryption Standard (AES) and an open protocol such as TLS, SSL, IPsec, XML encryptions, or S/MIME or their successors.
P5	PReVAL	The system shall provide the ability to check for potential interactions between medications to be prescribed or ordered and current medications and alert the user at the time of medication prescribing or ordering if potential interactions exist.
N1	Ad-hoc	When access to a chart is restricted, the system shall provide a means for appropriately authorized users to "break the glass" for emergency situations.
N2	Ad-hoc	The system shall provide the ability to associate orders, medications, and notes with one or more problems or diagnoses.
N3	Ad-hoc	The system shall provide the ability to record the prescribing of medications including the identity of the prescriber.
N4	Ad-hoc	The system shall provide the ability to record the user ID and date of all updates to the problem or diagnosis list.
N5	Ad-hoc	The system shall provide authorized administrators with the capability to read all audit information from the audit records in one of the following two ways: 1) The system shall provide the audit records in a manner suitable for the user to interpret the information.

### 3. RESULT AND DISCUSSIONS

This section is comprised of four sub-sections. The first sub-section presents the developed persona and results of requirements validation. The second sub-section presents the result of validating the proposed approach, while the third sub-section discusses the result as answers to the research questions. Finally, the fourth sub-section explains the limitations of the experimentation in the form of validity threats.

#### 3.1. Developed Persona and Requirements Validation Result

Figure 4 shows an example of persona-induced prompts constructed during the experiment based on the persona of participant no. 1. Based on the template shown in Fig.2, the prompt is constructed through filling the parameters with persona data. Figure 5 shows the validation of requirement P1 through PReVAL based on the persona of participant no. 1. It can be seen that the validation explicitly addresses the business goal as the persona stated that it wants strong emphasis on business goals rather than simple requirements.

#### 3.2. Review Results

Figure 6 summarizes the distribution of participants' responses for each validation result in a diverging bar chart [23]. Positive responses are located to the right while negatives are to the left,





which means validation results with more positive results skewed more to the right. Each validation result is identified with the same ID as shown in Table II, where P indicates review results from PReVAL. In contrast, N indicates review results from the ad-hoc approach. The review results are sorted in descending order based on the positiveness of the participants' responses.

You are an expert assistant acting strictly based on the following persona:

Persona 1: B – The Reflective Problem-Solver

Age & Demographics:

24 years old, background in Software Engineering, early-career professional seeking clarity on future goals.

Job Title & Responsibilities: UI/UX Designer

Responsible not only for designing interfaces but also understanding user psychology, conducting research, analyzing problems, and supporting documentation for business and technical teams.

Work Environment (tools, team structure, workflow):

- Full-time, hybrid setup (home + office)
- Part of a large corporation (>300 employees, 20+ divisions)
- Works in the Business Solution Team, within a hierarchical corporate structure
- Frequently collaborates with external partners, business managers, and developers
- Tools used: Figma, ChatGPT, Microsoft Word, Slack, Rive

Goals & Motivations:

- Deliver user-centered design with strong business value
- Maintain work-life balance and personal exploration through hybrid flexibility
- Advocate for design maturity and better cross-functional understanding

Frustrations & Pain Points:

- Misconceptions about UI/UX being limited to visuals
- Poor task delegation and inefficient collaboration
- CRM system is unintuitive and leads to data input errors
- Design inconsistency due to teammates ignoring system guidelines
- Inadequate documentation practices and poorly structured requirements

Technical Proficiency / Tools Used:

High proficiency: Figma, Rive, documentation tools

Self-directed learning style: reads documentation, explores tools independently, asks peers if needed

Quotes or Representative Statements:

“Many people think that UI/UX is only responsible for designing a feature, but the process should include research, problem analysis, and testing. The process should include research, problem analysis, and testing.”

**Figure 4.** Example of prompt generated during the experiment with induced persona



It can be seen from Fig. 6 that while the responses are pretty similar between the PReVAL and ad-hoc, ad-hoc validation edges the PReVAL by a small margin through responses to N4. The only occurrence of a "Strongly disagree" response happens in P5, while there is no occurrence of a "Strongly agree" response. PReVAL and ad-hoc approaches have one validation result that ranks lowest in P4 for PReVAL and N1 for ad-hoc. Most importantly, a significant number of neutral responses, such as "Neither agree or disagree," need to be addressed for better analysis.

#### Clarity & Completeness Check

- What does "individualize" mean? Does it imply user-customizable alerts, or automatic triggering based on patient data?
- What is meant by "specific clinical situation"? Is it referring to conditions, medications, or something else?
- No clear business goal: Lacks an explanation of how it impacts the clinical workflow or patient safety.

#### Consideration

- The requirement needs clarification on the business objective, especially whether it's aimed at improving diagnosis accuracy, regulatory compliance, or patient safety.
- The technical details need to align with the end-user's needs and the clinical process.

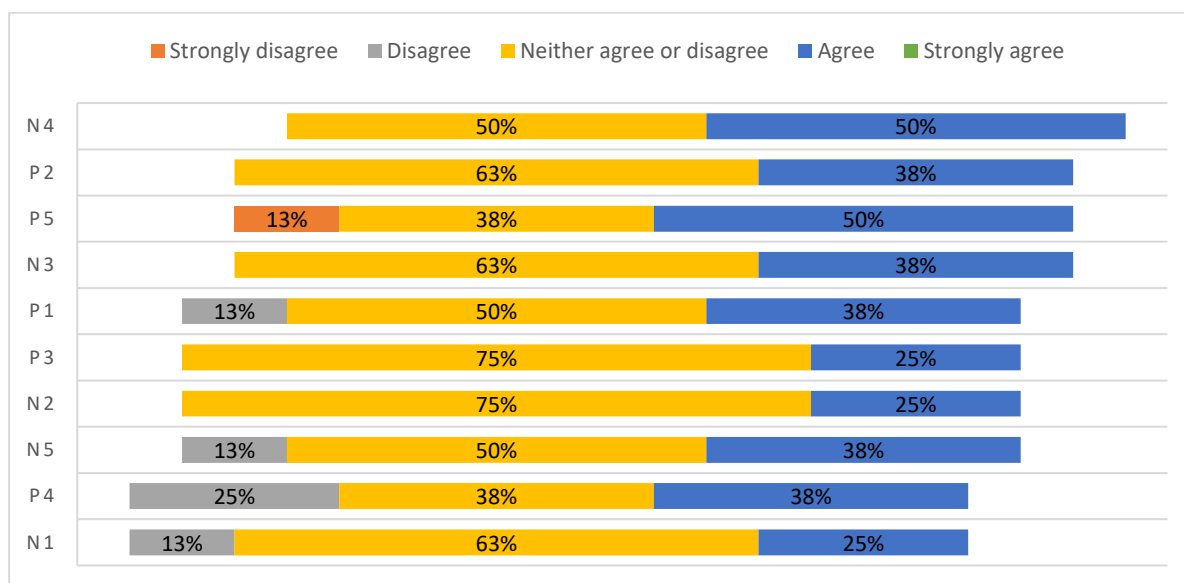
#### Implication

- Ambiguity could lead to misalignment between stakeholders and development teams, especially regarding the scope and purpose of the alerts.
- Without a clear business context, there's a risk of developing a feature that does not adequately meet the user's needs or improve clinical outcomes.

#### Suggested Rewrite

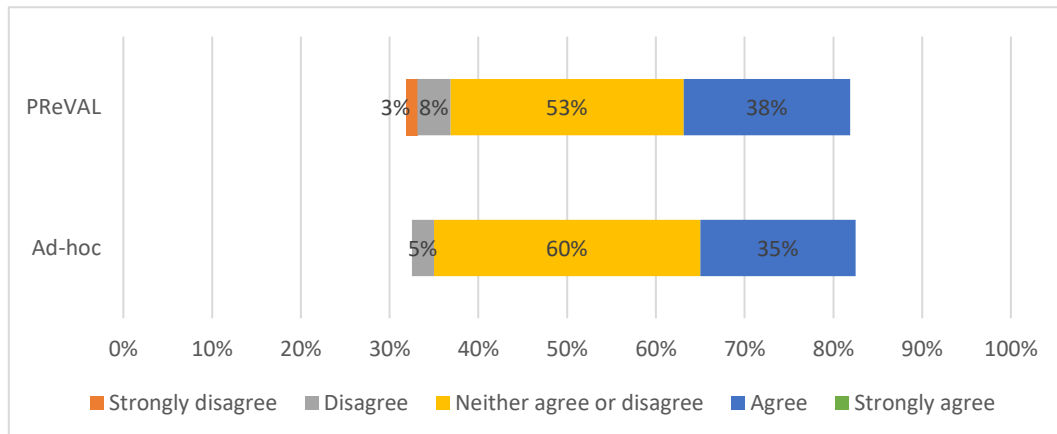
As a healthcare provider, I want the system to automatically check patient data (e.g., medications, allergies, clinical history) for potential issues during prescribing or ordering, So that I can receive personalized alerts about possible clinical situations (e.g., drug interactions, allergies) relevant to the patient and take action accordingly.

**Figure 5.** Example of PReVAL results on experiment



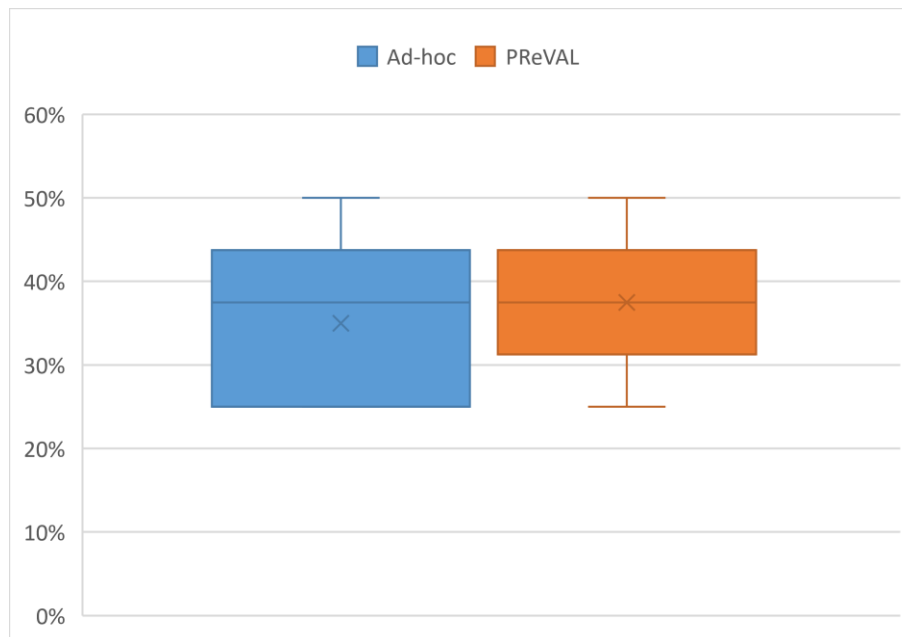
**Figure 6.** Participants' responses to each requirement validation. P indicates validation result from PReVAL while N indicates validation result from ad-hoc LLM-based approach.

Figure 7 summarizes the cumulative distribution of responses between PReVAL and ad-hoc validation in a diverging bar chart. The figure shows that PReVAL has more positive responses than ad-hoc validation but has significantly more negative responses. On the other hand, it can be seen that ad-hoc validation received significantly more neutral responses. This situation shows that PReVAL is likely to produce validations with more contrasting responses compared to ad-hoc LLM-based validation. This finding will be discussed further to answer RQ1.



**Figure 7.** Comparison of cumulative response between PReVAL and Ad-hoc.

Finally, Fig.8 shows the distribution of the top two box values of each question from PReVAL and ad-hoc validation in a box and whisker plot. It can be seen that ad-hoc validation and PReVAL have the same minimum, maximum, and median. However, PReVAL showed a more stable performance in this respect with a lower interquartile range compared to ad-hoc's. Additionally, both approaches show a similar mean that are close to their respective median.



**Figure 8.** Box and whisker plot of top two box value of PReVAL and ad-hoc approaches.



### 3.3. Discussion

This sub-section will answer the previously defined research questions based on findings presented in the previous sub-section.

#### **3.3.1. RQ1. How effective is PReVAL in mimicing stakeholders' concerns in requirement validation?**

The distribution of two box values shown in Fig. 8 indicates that while it has shown similar effectiveness to ad-hoc LLM-based validation, it also shows more consistency. The consistency shown puts PReVAL's efficacy in a better position for utilization due to the controllability it showed. However, further exploration is required as more experimentation of the quality of personas utilized in PReVAL is needed to understand its influence on overall effectiveness and consistency.

#### **3.3.2. RQ2. How useful is the validation result generated through PReVAL for the stakeholders?**

The distribution of responses shown in Fig. 6 and Fig. 7 demonstrated the extremities of PReVAL's results from stakeholders' perspectives. On the other hand, ad-hoc LLM-based validation shows more uniform responses, albeit mostly a neutral answer. The neutrality of ad-hoc LLM-based validation is in line with the recent findings of common commercial LLMs' tendency to answer in general manner and avoid more polarizing generation. The utilization of persona-induced prompt in PReVAL managed to push the LLMs into generating more tailored results in validation but introducing less desirable validation as well.

### 3.4. Threat to Validity

There are two main threats to the validity of the findings. The first one is selection bias, both in the selection of participants and the selection of requirements. As previously explained, the single-blind approach was selected to minimize the bias the participants might have towards specific approaches. Random selection in participants, requirements, and validation approaches in the experiment also specifically chosen to limit the influence of selection bias in the result.

The second threat is the small number of participants and the requirements to be validated. This limitation can limit the generality of the results. The explicit requirement of having industrial experience was designed to increase the quality of the findings despite the small number of participants. The limited number of requirements was also a deliberate decision, considering minimizing the chance of participants dropping out of the experiment due to exhaustion.

## 4. CONCLUSION

This paper proposed PReVAL (Persona-driven Requirement Validation with Large Language Models) and evaluated its potency through a comparative experiment. The experiment showed that PReVAL managed to push the limitations of ad-hoc LLM-based requirement validation by introducing a persona-induced prompt. By doing so, PReVAL managed to mimic experts in validating requirements consistently. However, it also shows tendencies to show more mistakes compared to ad-hoc approaches.

In the future, duplication and replication of the experiment are necessary to properly understand the benefits and limitations of PReVAL in a more general manner. The ability of PReVAL to handle variations of specification format needs to be evaluated to understand the ideal condition for utilizing PReVAL. On top of that, further improvements in persona construction methods are necessary to ensure that the aim of PReVAL can be fully achieved. Different styles of personas may have different effects on the validation results.





## 5. REFERENCES

- [1] Wiegers K. and Beatty J., 2013, *Software Requirements*, Ed.3, Microsoft Press, Washington
- [2] The Standish Group (2020). *CHAOS Report 2020*. The Standish Group
- [3] Maalem, Sourour & Zarour, Nacer eddine. (2016). Challenge of validation in requirements engineering. *Journal of Innovation in Digital Ecosystems*. 3. 10.1016/j.jides.2016.05.001.
- [4] Maldonado, J.C., Carver, J., Shull, F. *et al.* Perspective-Based Reading: A Replicated Experiment Focused on Individual Reviewer Effectiveness. *Empir Software Eng* 11, 119–142 (2006). <https://doi.org/10.1007/s10664-006-5967-6>
- [5] F. Shull, I. Rus and V. Basili, "How perspective-based reading can improve requirements inspections," in *Computer*, vol. 33, no. 7, pp. 73-79, July 2000, doi: 10.1109/2.869376
- [6] Vogelsang, A., Fischbach, J. (2025). Using Large Language Models for Natural Language Processing Tasks in Requirements Engineering: A Systematic Guideline. In: Ferrari, A., Ginde, G. (eds) *Handbook on Natural Language Processing for Requirements Engineering*. Springer, Cham. [https://doi.org/10.1007/978-3-031-73143-3\\_16](https://doi.org/10.1007/978-3-031-73143-3_16)
- [7] S. Lubos et al., "Leveraging LLMs for the Quality Assurance of Software Requirements," in *2024 IEEE 32nd International Requirements Engineering Conference (RE)*, Reykjavik, Iceland, 2024, pp. 389-397, doi: 10.1109/RE59067.2024.00046.
- [8] Cheng, Haowei & Husen, Jati & Peralta, Sien & Jiang, Bowen & Yoshioka, Nobukazu & Ubayashi, Naoyasu & Washizaki, Hironori. (2024). *Generative AI for Requirements Engineering: A Systematic Literature Review*. 10.48550/arXiv.2409.06741.
- [9] T. Mahbub, D. Dghaym, A. Shankarnarayanan, T. Syed, S. Shapsough and I. Zualkernan, "Can GPT-4 Aid in Detecting Ambiguities, Inconsistencies, and Incompleteness in Requirements Analysis? A Comprehensive Case Study," in *IEEE Access*, vol. 12, pp. 171972-171992, 2024, doi: 10.1109/ACCESS.2024.3464242.
- [10] A. Fantechi, S. Gnesi, L. Passaro and L. Semini, "Inconsistency Detection in Natural Language Requirements using ChatGPT: a Preliminary Evaluation," *2023 IEEE 31st International Requirements Engineering Conference (RE)*, Hannover, Germany, 2023, pp. 335-340, doi: 10.1109/RE57278.2023.00045.
- [11] Arora, C., Grundy, J., Abdelrazek, M. (2024). *Advancing Requirements Engineering Through Generative AI: Assessing the Role of LLMs*. In: Nguyen-Duc, A., Abrahamsson, P., Khomh, F. (eds) *Generative AI for Effective Software Development*. Springer, Cham. [https://doi.org/10.1007/978-3-031-55642-5\\_6](https://doi.org/10.1007/978-3-031-55642-5_6)
- [12] Sun, G., Zhan, X., & Such, J. (2024). *Building Better AI Agents: A Provocation on the Utilisation of Persona in LLM-based Conversational Agents*. *Proceedings of the 6th ACM Conference on Conversational User Interfaces*.
- [13] Hu, T., & Collier, N. (2024). *Quantifying the Persona Effect in LLM Simulations*. *ArXiv, abs/2402.10811*.
- [14] Bhandari, P., Fay, N., Wise, M., Datta, A., Meek, S., Naseem, U., & Nasim, M. (2025). *Can LLM Agents Maintain a Persona in Discourse?*. *arXiv preprint arXiv:2502.11843*.
- [15] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D.C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. *ArXiv, abs/2302.11382*.
- [16] Siddiq, M., Samee, A., Azgor, S.R., Haider, M.A., Sawraz, S.I., & Santos, J.C. (2023). *Zero-shot Prompting for Code Complexity Prediction Using GitHub Copilot*. *2023 IEEE/ACM 2nd International Workshop on Natural Language-Based Software Engineering (NLBSE)*, 56-59.
- [17] Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). *Attention is All you Need*. *Neural Information Processing Systems*.
- [18] LinkedIn, "LinkedIn Developer Productivity and Happiness Framework," [linkedin.com. https://github.com/linkedin/dph-framework/](https://github.com/linkedin/dph-framework/) (accessed Apr. 20, 2025).
- [19] Kore, A., Abbasi Babil, E., Subasri, V., Abdalla, M., Fine, B., Dolatabadi, E., & Abdalla, M. (2024). *Empirical data drift detection experiments on real-world medical imaging data*. *Nature Communications*, 15.
- [20] Day, S. J., & Altman, D. G. (2000). *Statistics notes: blinding in clinical trials and other studies*. *BMJ (Clinical research ed.)*, 321(7259), 504. <https://doi.org/10.1136/bmj.321.7259.504>
- [21] Stol, K., & Fitzgerald, B. (2018). *The ABC of Software Engineering Research*. *ACM Transactions on Software Engineering and Methodology (TOSEM)*, 27, 1 - 51.
- [22] Ferrari A., Spagnolo O. G., and Gnesi S., 2017, *PURE: A Dataset of Public Requirements Documents*, *Proceeding of 2017 IEEE 25th International Requirement Engineering Conference*, Lisbon, September 2017
- [23] Heiberger, R., & Robbins, N. (2014). *Design of Diverging Stacked Bar Charts for Likert Scales and Other Applications*. *Journal of Statistical Software*, 57(5), 1–32. <https://doi.org/10.18637/jss.v057.i05>

