Proceedings of the Design Society, Volume 5: ICED25
https://doi.org/10.1017/pds.2025.10280

Prompt engineering study: comparing pre-service engineers to large language models in requirements generation

Shanae Edwards, Kenny Nonso-Anyakwo, Joshua Summers✉ and Oredola Adebayo

University of Texas at Dallas, USA

✉ joshua.summers@utdallas.edu

ABSTRACT: The objective of this research is to compare the requirements generated by human participants and large language models (LLMs). Requirements are statements that capture the needs and desires from stakeholders and organize them into design parameters. These statements are expressed in natural language which may lead to incompleteness and ambiguity. Due to the recent advancements in the natural language model such as ChatGPT and Gemini as a tool for requirement generation, this study investigates the quantity, variety and completeness of requirements generated by 66 pre-service engineers and 4 LLMs. This is because in some design projects, stakeholder access may be limited. The results show that pre-service engineers outperformed LLMs in variety, quantity and completeness. Future work could involve developing and comparing true human personas to LLMs.

KEYWORDS: artificial intelligence, requirements, computational design methods

1. Motivation: LLMs vs engineers in requirement generation

The role of engineering requirements is fundamental in product development, as they form the basis for meeting the needs, wants, and constraints of stakeholders (Ullman, 2017). These requirements define the problem space and guide the engineering design process [2–4]. Traditionally, requirements are gathered through direct interaction with stakeholders via interviews, surveys, and other elicitation methods (Morkos & Summers, 2009). However, challenges arise when access to these stakeholders is limited, especially in large, complex projects with diverse or geographically dispersed teams, some examples of stakeholders who are difficult to engage may include those from underrepresented or marginalized communities, those from small populations, or those from locations not accessible by the designers (Valerio et al., 2016). In such cases, traditional methods may fall short in capturing the full scope of stakeholder needs

Recent advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP), particularly with Large Language Models (LLMs), have opened new possibilities in automating and enhancing the requirements engineering process (Ha et al., 2024; Hao et al., 2024; Krapp et al., 2024; Liu et al., 2024). LLMs, such as ChatGPT, Gemini, and others, can process vast amounts of data and generate coherent and contextually relevant text (Krapp et al., 2024; Liu et al., 2024). This study explores the potential of using LLMs as surrogate stakeholders in requirements elicitation, comparing the quantity, variety, and completeness of requirements generated by LLMs to those produced by human engineers. By evaluating LLM performance, this research aims to understand whether LLMs can be a viable tool in supporting or supplementing traditional requirements generation methods.

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

2. Background

Requirements engineering ensures products align with stakeholder needs but capturing clear and comprehensive requirements is often challenging. This section discusses key evaluation metrics, the importance of requirements engineering, and the potential of LLMs as surrogate stakeholders.

2.1. Importance of requirements engineering

Requirements engineering is a critical aspect of engineering design that ensures the designed product aligns with stakeholder needs (Darlington & Culley, 2002). Well-defined requirements are key to preventing miscommunication, reducing the risk of failure, and ensuring that the final product meets its intended purpose (Lind et al., 2023). Standards organizations, such as INCOSE (International Council on Systems Engineering), IEEE (Institute of Electrical and Electronics Engineers), and NASA, emphasize the importance of clear, concise, and complete requirements to ensure successful product development (Lind et al., 2023; National Aeronautical and Space Administration, 2016).

The International Council of System Engineers (INCOSE) defines requirements as well-formed textual "shall" statements that communicate in a structured, natural language what an entity must do to realize the intent of the needs from which they were transformed (Institute of Electrical and Electronics Engineers (IEEE), 2018). The Institute of Electrical and Electronics Engineers (IEEE) defines a requirement as a statement that translates or expresses a need, along with its associated constraints and conditions (Institute of Electrical and Electronics Engineers (IEEE), 2018). NASA's Systems Engineering Handbook defines a requirement as the agreed-upon need, desire, want, capability, capacity, or demand for resources or services, expressed as a "shall" statement (National Aeronautical and Space Administration, 2016). These organizations emphasize writing requirements with a single subject and using modal terms such as "shall", "must", or "should" to capture the criticality of a requirement (National Aeronautical and Space Administration, 2016). The requirement should also end with a verb phrase, capturing what the subject must do, along with any necessary information to fulfill the requirement (Spivey, 2019) Based on guidelines from INCOSE, IEEE, and NASA, as well as other research (Joshi & Summers, 2014b, 2014a; Spivey, 2019; Spivey et al., 2021), three metrics to evaluate requirements have been identified. These metrics are quantity, variety and completeness.

- **Quantity:** This refers to the total number of requirements generated during the elicitation phase. Having a sufficient number of requirements ensures that all aspects of the project are covered.
- **Variety:** The diversity of requirements is essential to ensure that different categories are included, such as functional, non-functional, performance, safety, and regulatory requirements. A broad variety helps ensure the product is suitable for its intended environment.
- **Completeness:** A good requirement should be fully detailed, including the subject, verb, modal terms (e.g., "shall" or "must"), and target values. Completeness refers to the depth and specificity of each requirement to ensure clarity and traceability.

Despite the importance of these metrics, capturing stakeholder needs effectively remains challenging (Huzzard, 2021). In projects, especially new product development, requirements are often incomplete or inaccurate due to stakeholders' difficulty in articulating their needs (Huzzard, 2021). As products evolve from conception to prototype, latent needs emerge that further complicate requirements elicitation. The process is resource-intensive, particularly when the product is novel, and its design is uncertain.

2.2. Challenges in requirement elicitation

Eliciting accurate and comprehensive requirements is a complex task (Christophe et al., 2011). The challenge is exacerbated by the fact that stakeholders often have limited time and availability, and their needs can evolve over time (Nilsson & Fagerström, 2006). Traditional methods, such as one-on-one interviews, focus groups, and surveys, can be time-consuming, costly, and dependent on the skill of facilitators (Morkos & Summers, 2009). These methods can also fail to capture the full spectrum of requirements, particularly in large-scale systems where stakeholders may have conflicting interests or differing perspectives (Khade et al., 2022; Ulrich et al., 2020).

Furthermore, the richness of the information stakeholders provide can vary significantly based on their role and understanding of the project (Nilsson & Fagerström, 2006). For example, end-users might

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

provide insights into usability, while regulatory bodies might focus on safety and compliance (Khade et al., 2022; Ulrich et al., 2020). This variability in input complicates the task of establishing a comprehensive set of requirements (Morkos & Summers, 2010). Additionally, as the project progresses and new information surfaces, previously gathered requirements may need to be refined, updated, or even discarded [14].
The complexity of modern projects, combined with evolving stakeholder needs, makes it clear that traditional methods have limitations(Yasin et al., 2024). As a result, there is growing interest in exploring whether automated tools like LLMs could offer a more efficient and effective means of gathering and refining requirements (Wei, 2024).

2.3. pPotential of LLMs as surrogate stakeholders

Large language models (LLMs) represent a promising tool for assisting in requirements engineering (Liu et al., 2024). Their advanced capabilities in text generation and context understanding make them potentially valuable for simulating stakeholder input (Liu et al., 2024). By carefully designing prompts and providing context, LLMs can generate requirements that align with typical stakeholder concerns, potentially serving as surrogate stakeholders in the requirements elicitation process (Marvin et al., 2024). Prompt engineering plays a crucial role in optimizing the performance of LLMs (Gao, 2023; Marvin et al., 2024; Wang et al., 2024). This process involves crafting task-specific prompts to guide LLMs in producing relevant, coherent, and high-quality outputs (Marvin et al., 2024). Research has demonstrated that well-designed prompts can significantly enhance the capabilities of LLMs in generating domain-specific outputs, improving their ability to match the needs of the project at hand (Marvin et al., 2024). However, while LLMs have shown promise, their ability to generate complete and varied requirements remains an open question. Initial studies have demonstrated that LLMs can produce contextually relevant text, but whether they can consistently generate requirements that are as comprehensive, diverse, and detailed as those developed through traditional methods is still under investigation. A study examining the use of LLMs for generating safety requirements for autonomous vehicles found that LLMs could generate requirements comparable to human-written ones, but certain areas still required refinement to fully capture the complexity of the system (Liu et al., 2024).
This study aims to explore the comparative effectiveness of LLMs in generating requirements, focusing on key metrics such as quantity, variety, and completeness. By identifying areas where LLMs excel or fall short, the research will contribute to understanding how these tools can best support or supplement human-driven requirements engineering efforts.

3. Experimental design

This study investigates the comparison between engineering requirements generated by preservice engineers (undergraduate engineering students) and those produced by large language models (LLMs). The objective is to evaluate the effectiveness of LLMs in generating engineering requirements and assess how their output compares to human-generated requirements in terms of **quantity, variety,** and **completeness.** This section details the experimental setup, including participant selection, procedures, and evaluation criteria, to systematically analyse the generated requirements. By employing a controlled experimental framework, this study aims to provide insights into the potential of artificial intelligence to complement traditional methods of requirements generation in engineering design.
The study focuses on two primary independent variables: preservice engineers and LLMs. These variables are analyzed to determine their respective strengths in generating engineering requirements, with the comparison grounded on three metrics: quantity, variety, and completeness. A standardized rubric based on requirement guidelines outlined in (Spivey et al., 2021) serves as the evaluation tool for both groups.

3.1. Design prompts

Two distinct design prompts were created to elicit requirements from both participant groups. **Design Prompt 1** was administered to both pre-service engineers (undergraduate engineering students) and LLMs. **Design Prompt 2 with Persona** was exclusively administered to the LLMs. This variant incorporated a persona intended to mimic the context and behavior of preservice engineers. These design prompts are detailed in Table 1.

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

Table 1. Design prompts

Design Prompt 1 (LLMs and Pre-Service Engineers)
In order to help people in wheelchairs to grab books from the highest level of the bookshelf (at 6 ft or above), a mechanism needs to be developed. The device must be safe to use, convenient, and operate smoothly without damaging the books. The assembly should be relatively simple so that it can be installed on most existing bookshelves. Generate requirements that solves the problem above.
Design Prompt with Persona (LLMs only)
Imagine you are a first-year mechanical engineering freshman at a public university who is also a part of the FSAE team. You are employed by an engineering design firm and have been tasked with generating requirements for a design that solves the following problem. *In order to help people in wheelchairs to grab books from the highest level of the bookshelf (at 6 ft or above), a mechanism needs to be developed. The device must be safe to use, convenient, and operate smoothly without damaging the books. The assembly should be relatively simple so that it can be installed on most existing bookshelves*

3.2. Human participants

The study was conducted under an approved experimental protocol through the local Institutional Review Board, ensuring no risks to the participants. The participants included **116 preservice engineers** enrolled in four sections of an "Introduction to Engineering and Computer Science" course at a large public research university in the United States. While the study initially involved 116 participants, 50 of them were given a different design prompt, resulting in **66 participants** included in the analysis. The course, a required first-semester class for all engineering and computer science students, comprised approximately half computer science majors, with the remaining students enrolled in bioengineering, electrical engineering, or mechanical engineering programs. About four-fifths of the students identified as male, and only 5% were international students, with an age range typically between 17 and 20 years. Specific demographic data were not collected, as it was deemed outside the scope of the experiment.

While this study had implications for both educational and industrial settings, pre-service engineering students from the School of Engineering and Computer Science were recruited to ensure a consistent baseline of engineering knowledge. Prior research demonstrated that pre-service engineers performed similarly to practitioners with at least three years of experience when generating requirements (Elena, 2019). Additionally, another study found no significant difference in solution quality or technical feasibility between freshman and senior engineering students when generating innovative solutions (Genco et al., 2012).

The four sections involved in the study had enrollments ranging from 170 to 300 students and met in a traditional lecture hall on Tuesdays and Thursdays. Participants were asked to refrain from discussing the in-class activity with others. The experiment was conducted during regular class time, following a lecture on requirements as part of the coursework. While each section had a different instructor, the content was standardized with a common guest lecture delivered by the same researcher. The lecture covered topics such as the role of requirements in engineering design, methods for measuring requirements, common issues in requirement formulation, and the involvement of stakeholders.

After the lecture, participants were given detailed instructions for the requirements generation task, specifying the required format, structure, and key considerations. They were presented with a problem statement and given 20 minutes to generate requirements. Once the time limit elapsed, the required documents were collected. Independent observers, unaffiliated with the participants, supervised the activity through real-time observation and periodic check-ins to ensure adherence to experimental protocols. This monitoring process was designed to be non-intrusive.

The participant packets contained two sections: the first requested the creation of a unique identifier, while the second included instructions for the requirements generation task and the design problem prompt. Participants were asked to imagine working for an engineering design firm tasked with generating requirements to address the given problem.

3.3. AI participants

Four prominent LLMs were evaluated in this study: CoPilot by Microsoft, Gemini by Google, ChatGPT by Open AI, and Claude by Anthropic. These LLMs were selected based on their demonstrated capabilities in supporting engineering design activities, specifically in requirements generation. CoPilot has been effectively used for assisting with requirement formulation tasks, such as writing requirements

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

for a "tennis ball launcher" (Valentine, n.d.). Gemini has been evaluated to have advanced capabilities in understanding complex instructions (Rane et al., 2024). ChatGPT has been explored with respect to improving software requirements engineering (Marques et al., 2024). Claude has been recognized for its ability to process multimodal prompts, including text, code, and structured data files (Caruccio et al., 2024). These LLMs were chosen for their demonstrated utility in engineering design and their distinct approaches to requirement generation. The inclusion of multiple LLMs allows for comparative analysis and highlights variations in their outputs.

4. Results and discussions

This section presents the responses of the human and AI participants. The human participants responded to the baseline prompt, while large language models (LLMs) were given both the baseline prompt (B) and a persona-enhanced prompt (P). Responses were evaluated based on quantity, variety, and completeness of requirements. Differences and similarities between human and LLM responses are also discussed.

4.1. Quantity coding

Quantity refers to the number of requirements generated by the participant, reflecting their ability to propose design specifications. To standardize analysis, requirements were split into atomic statements, each expressing a single thought (Spivey et al., 2021). For example, consider the requirement:

- *"While in operation, the system must cool the motor and remain under 60 decibels."*

The improved requirement would be written as two requirement statements:

- *"While in operation, the system must cool the motor",*
- *"While in operation, the system must remain under 60 decibels."*

This splitting process adhered to established criteria, including the presence of multiple verbs, adjectives, or conditional expressions. Exceptions to splitting were made when clauses explained a requirement's purpose or when conditions applied collectively rather than independently. A comprehensive description of the splitting criteria can be found in (Spivey et al., 2021).

4.2. Quantity analysis

The initial count of requirements generated totalled 842, distributed across four course sections with varying participant numbers: 12, 9, 25, and 21. Statistical analysis using ANOVA confirmed no significant differences in the number of requirements generated between sections ($\alpha = 0.05$). Following the initial count, requirements were split into atomic units to facilitate detailed analysis.

Table 2 presents the initial and final counts of requirements, along with the average number of requirements per participant and the percentage of splits. While human participants generated more preliminary requirements than the LLMs, splitting revealed that LLMs using the persona-enhanced prompt produced more refined and focused outputs. The results indicate that the persona prompt improved the LLMs' ability to generate a higher quantity of meaningful requirements.

Table 2. Number of split requirements for human participants

	Human	LLM (Baseline)	LLM (Persona)
Number of Participants	66	4	4
Preliminary Count	842	69	71
Final Count	1023	83	82
Average Requirements per Person (Final Count)	15.5	17.25	17.75
Number of Times Split (% of preliminary requirements)	145 (17.22%)	14 (20%)	11(15%)

These results demonstrate that while human participants generated a larger volume of requirements, LLMs using the persona prompt achieved better efficiency, with fewer splits needed to clarify requirements.

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

Table 3. Examples of requirement categorizations

Requirement	Categorization	Source
Device must have a minimum reach of 6 feet	Geometry	Human
Device shall convert electrical to mechanical energy	Operation, Energy	Human
Device must have a maximum manufacturing cost of $20	Cost	Human
The mechanism must be stable prevent tipping	Kinematics	LLM
Device must be safe to use, with no sharp edges or pinch points	Safety, Geometry	LLM
The materials used should be able to withstand regular use	Material	LLM

4.3. Variety coding

Variety evaluates the range of categories covered by participants' requirements. Categories, adapted from (Pahl et al., 2007), include geometry, kinematics, forces, energy, material, signal, safety, ergonomics, production, quality control, assembly, transport, operation, maintenance, recycling, cost, and schedules. If a requirement does not fall within these categories, it is coded as "NA". Examples of requirement categorizations can be found in Table 4.

4.4. Variety analysis

Variety was chosen as a key metric in this study to evaluate the breadth of requirements generated. Research indicates that greater variety suggests a more comprehensive understanding of the problem domain (Viswanathan & Linsey, 2012). Variety was assessed by categorizing requirements based on the split requirements approach detailed in Section 4.1. Each requirement was assigned to a maximum of three categories to minimize overgeneralization and ensure relevance.

The analysis began by examining the variety of requirements generated by human participants and large language models (LLMs). For each participant group, the number of categories addressed was counted. To account for variations in the total number of requirements generated, percentages were calculated by dividing the frequency of each category by the total number of categories across all requirements. This approach normalized the results, enabling comparisons between participants who generated different quantities of requirements. To facilitate a more nuanced comparison, categories were grouped into five percentile ranges. The highest frequency categories were assigned a score of 5, while the lowest received a score of 1. These percentiles provided a clear basis for comparing the diversity of responses between human participants and LLMs.

Table 4 provides a summary of the variety in responses between the human participants and the LLMs. It includes the percentage distribution of requirements across the categories and the assigned percentiles. The categories, adapted from (Pahl et al., 2007) are listed on the x-axis, and the variety percentages for humans, baseline LLM responses, and persona-driven LLM responses are shown.

Table 4. Summary of the variety in responses between the human participants and the LLMs

	Geometry	Kinematics	Forces	Energy	Material	Signals	Safety	Ergonomics	Production	Quality Control	Assembly	Transport	Operation	Maintenance	Recycling	Cost	Schedules	N/A	Total
Human	9.39%	8.58%	4.67%	2.19%	4.55%	4.90%	10.94%	11.23%	0.81%	5.93%	6.22%	0.29%	23.44%	2.30%	0.92%	2.76%	0.46%	0.40%	100.00%
Sum	163	149	81	38	79	85	190	195	14	103	108	5	407	40	16	48	8	7	1736
Percentile	2	2	1	1	1	1	3	3	1	2	2	1	5	1	1	1	1	1	
LLM - B	2.90%	7.25%	0.00%	2.90%	4.35%	0.00%	5.80%	14.49%	11.59%	4.35%	11.59%	1.45%	23.19%	4.35%	0.00%	1.45%	4.35%	0.00%	100.00%
Sum	2	5	0	2	3	0	4	10	8	3	8	1	16	3	0	1	3	0	69
Percentile	1	1	1	1	1	1	4	3	1	1	4	1	5	2	1	1	1	0	
LLM - P	2.70%	8.11%	0.00%	5.41%	8.11%	1.35%	8.11%	13.51%	12.16%	1.35%	8.11%	1.35%	21.62%	5.41%	1.35%	1.35%	0.00%	0.00%	100.00%
Sum	2	6	0	4	6	1	6	10	9	1	6	1	16	4	1	1	0	0	74
Percentile	1	1	1	1	2	1	5	3	1	1	4	1	5	1	1	1	1	0	

Figure 1 visually represents the data from Table 4 illustrating the percentage of categories occupied by the responses from human participants and LLMs. The blue bars represent the categories covered by the human participants, the orange bars represent the baseline LLM responses, and the green bars represent the persona-driven LLM responses.

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

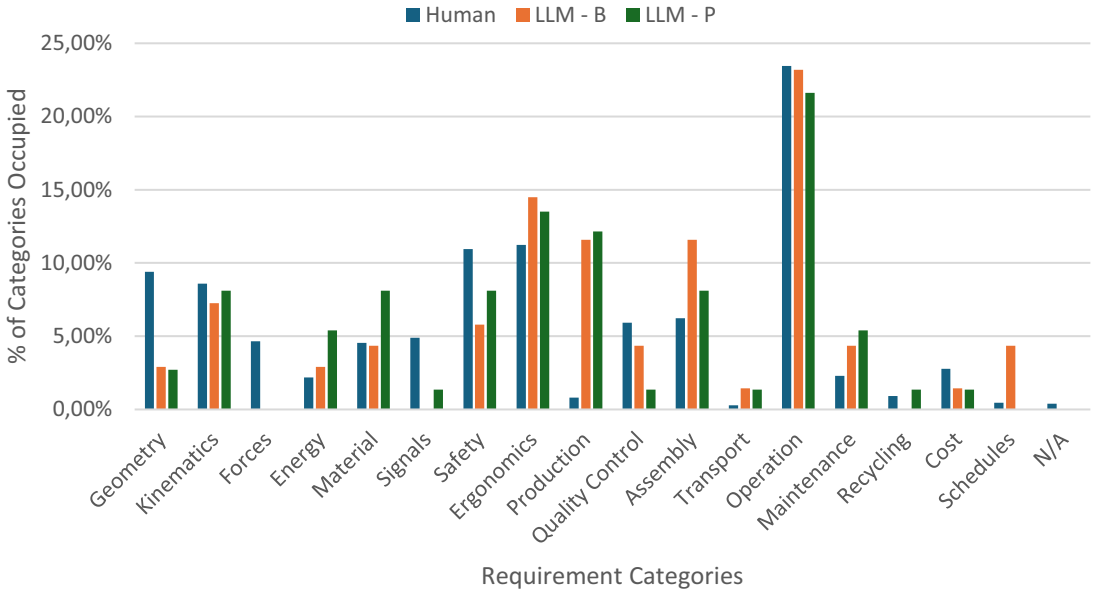


Figure 1. Number of categories occupied

The results in Figure 1 demonstrate that human participants generated requirements with higher variety compared to the LLMs. Categories such as geometry, kinematics, forces, safety, operation, and cost were more extensively addressed by human participants, likely reflecting their familiarity with these areas. Conversely, the LLMs, particularly when prompted with a persona, generated more requirements in energy, material, production, and maintenance categories. Notably, human participants produced minimal requirements related to production and transport, which may stem from limited manufacturing experience.

LLM responses to the baseline prompt failed to include force or recycling requirements, whereas persona-augmented prompts resulted in more assembly-related requirements but excluded schedule requirements entirely. Similarly, the signals category was not covered in baseline responses but was partially addressed by persona-augmented prompts. These findings suggest that while LLMs can expand their coverage with appropriate prompting, they still lag behind human participants in variety across critical categories.

4.5. Completeness coding

Completeness, another critical criterion, measures how well requirements encapsulate all necessary functionalities, constraints, and specifications. A complete requirement should include a subject, modality, and verb phrase to avoid ambiguity and ensure clarity [54][55]. Additional components, such as modifiers, objects, and target values, further enhance the precision and applicability of the requirements. Complete requirements are instrumental in facilitating effective communication and project planning [33].

4.6. Completeness analysis

Completeness was evaluated by analysing the structural components of the requirements generated by both human participants and LLMs. The presence of a subject, modality, and verb phrase served as the baseline for completeness, while the inclusion of modifiers, objects, and target values was also assessed. Percentages were calculated for each component relative to the total number of requirements generated. Table 5 presents the results, highlighting the disparities between the human-generated and LLM generated requirements. Human participants generated 842 requirements, compared to 159 of the LLMs. Across all components, the human-generated requirements exhibited higher percentages. For instance, 75.27% of human-generated requirements contained a subject, compared to only 25% and 30% for the baseline and persona-augmented LLM responses, respectively. Similarly, 98.53% of human-generated requirements included a verb, far exceeding the 38% and 33% observed for the LLMs.

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

Modifiers, objects, and target values were significantly underrepresented in the LLM-generated requirements. Only 1% of baseline LLM responses included modifiers, and none were observed in persona-augmented responses. In contrast, human participants included modifiers in 56.01% of their requirements. Similarly, objects appeared in 44.09% of human-generated requirements, compared to 38% and 32% for the LLMs. Target values, a critical indicator of specificity, were present in 22.09% of human-generated requirements but only 3% of LLM responses.

Table 5. Percentages of requirement components

Contained Subject			**Contained Verb**			**Contained Modal**		
Human	*LLM - B*	*LLM - P*	*Human*	*LLM - B*	*LLM - P*	*Human*	*LLM - B*	*LLM - P*
75.27%	25%	30%	98.53%	38%	33%	81.23%	28%	30%
Contained Modifier			**Contained Object**			**Contained Target Value**		
Human	*LLM - B*	*LLM - P*	*Human*	*LLM - B*	*LLM - P*	*Human*	*LLM - B*	*LLM - P*
56.01%	1%	0%	44.09%	38%	32%	22.09%	3%	3%

The disparity in completeness can be attributed to the training provided to human participants, who received a lecture on requirement writing before the task. This preparation likely equipped them to generate more comprehensive and structured requirements. In contrast, the LLMs, while capable of producing coherent responses, struggled to incorporate critical components consistently, highlighting a gap in their ability to replicate human-level understanding and detail.

5. Conclusions

This study explored the potential of Large Language Models (LLMs) as surrogate stakeholders in requirements engineering, comparing their performance to human engineers. The design team may not be able to engage with all stakeholders, therefore surrogates may be necessary. While LLMs demonstrated the ability to generate a substantial quantity of requirements and explore a variety of categories, their coverage and depth remain limited compared to human participants. Human engineers consistently outperformed LLMs in producing well-structured, detailed, and comprehensive requirements.

A key limitation of this study is the LLMs were not given any lecture on requirements like the human participants, which may not fully capture the diversity of their capabilities. Future research should explore a wider range of LLMs and prompt engineering techniques to further assess their potential.

Future research could investigate the effectiveness of combining human and LLM efforts in a collaborative approach. By leveraging the strengths of both, it may be possible to achieve a more efficient and effective requirements engineering process. Additionally, exploring the impact of different personas and prompt engineering techniques on LLM performance could provide valuable insights for optimizing their use in practical applications.

References

Caruccio, L., Cirillo, S., Polese, G., Solimando, G., Sundaramurthy, S., & Tortora, G. (2024). Claude 2.0 large language model: Tackling a real-world classification problem with a new iterative prompt engineering approach. *Intelligent Systems with Applications*, 21. https://doi.org/10.1016/j.iswa.2024.200336

Christophe, F., Wang, M., Coatanéa, E., Zeng, Y., & Bernard, A. (2011). Grammatical and Semantic Disambiguation of Requirements at Elicitation and Representation Stages. *Volume 9: 23rd International Conference on Design Theory and Methodology; 16th Design for Manufacturing and the Life Cycle Conference*, 17–31. https://doi.org/10.1115/DETC2011-48139

Darlington, M. J., & Culley, S.J. (2002). Current research in the engineering design requirement. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 216(3), 375–388. https://doi.org/10.1243/0954405021520049

Elena, M. V. (2019). *Understanding Requirement Generation: Studies on Interventions and Comparison between Novices and Practitioners* [Clemson University]. https://open.clemson.edu/all_theses/3076/

Gao, A. (2023). Prompt Engineering for Large Language Models. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4504303

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

Genco, N., Hölttä-Otto, K., & Seepersad, C.C. (2012). An Experimental Investigation of the Innovation Capabilities of Undergraduate Engineering Students. *Journal of Engineering Education*, 101(1), 60–81. https://doi.org/10.1002/j.2168-9830.2012.tb00041.x

Ha, J., Jeon, H., Han, D., Seo, J., & Oh, C. (2024). CloChat: Understanding How People Customize, Interact, and Experience Personas in Large Language Models. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–24. https://doi.org/10.1145/3613904.3642472

Hao, H., Zhang, X., & Zhou, A. (2024). *Large Language Models as Surrogate Models in Evolutionary Algorithms: A Preliminary Study*.

Huzzard, T. (2021). Achieving impact: exploring the challenge of stakeholder engagement. *European Journal of Work and Organizational Psychology*, 30(3), 379–389. https://doi.org/10.1080/1359432X.2020.1761875

Institute of Electrical and Electronics Engineers (IEEE). (2018). *Systems and software engineering-Life cycle processes-Requirements engineering*. https://doi.org/10.1109/IEEESTD.2018.8559686

Joshi, S., & Summers, J. D. (2014a, August 17). Impact of Requirements Elicitation Activity on Idea Generation: A Designer Study. *Volume 7: 2nd Biennial International Conference on Dynamics for Design; 26th International Conference on Design Theory and Methodology*. https://doi.org/10.1115/DETC2014-35022

Joshi, S., & Summers, J. D. (2014b, August 17). Tracking Project Health Using Completeness and Specificity of Requirements: A Case Study. *Volume 3: 16th International Conference on Advanced Vehicle Technologies; 11th International Conference on Design Education; 7th Frontiers in Biomedical Devices*. https://doi.org/10.1115/DETC2014-35020

Khade, V., Masoudi, N., Acena, D., Freeman, G., Rai, R., Gorsich, D., Rizzo, D., & Castanier, M. (2022, October 30). Requirements Elicitation: Impacts of Gamification on Variety, Novelty, and Completeness. *Volume 4: Biomedical and Biotechnology; Design, Systems, and Complexity*. https://doi.org/10.1115/IMECE2022-96016

Krapp, E., Neuhaus, R., Hassenzahl, M., & Laschke, M. (2024, October 13). In a Quasi-Social Relationship With ChatGPT. An Autoethnography on Engaging With Prompt-Engineered LLM Personas. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/3679318.3685501

Lind, E., Gonzalez-Huerta, J., & Alégroth, E. (2023). *Requirements Quality vs. Process and Stakeholders' Well-Being: A Case of a Nordic Bank* (pp. 17–37). https://doi.org/10.1007/978-3-031-31488-9_2

Liu, Y., Sharma, P., Oswal, M.J., Xia, H., & Huang, Y. (2024). *PersonaFlow: Boosting Research Ideation with LLM-Simulated Expert Personas*.

Marques, N., Silva, R.R., & Bernardino, J. (2024). Using ChatGPT in Software Requirements Engineering: A Comprehensive Review. In *Future Internet* (Vol. 16, Issue 6). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/fi16060180

Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2024). *Prompt Engineering in Large Language Models* (pp. 387–402). https://doi.org/10.1007/978-981-99-7962-2_30

Morkos, B., & Summers, J. D. (2009). Elicitation and Development of Requirements Through Integrated Methods. *Volume 2: 29th Computers and Information in Engineering Conference, Parts A and B*, 1007–1015. https://doi.org/10.1115/DETC2009-87720

Morkos, B., & Summers, J. D. (2010). Requirement Change Propagation Prediction Approach: Results From an Industry Case Study. *Volume 1: 36th Design Automation Conference, Parts A and B*, 111–121. https://doi.org/10.1115/DETC2010-28562

National Aeronautical and Space Administration. (2016). *NASA Systems Engineering Handbook*. https://www.nasa.gov/reference/systems-engineering-handbook/

Nilsson, P., & Fagerström, B. (2006). Managing stakeholder requirements in a product modelling system. *Computers in Industry*, 57(2), 167–177. https://doi.org/10.1016/j.compind.2005.06.003

Pahl, G., Beitz, W., Feldhusen, J., & Grote, K.-H. (2007). *Engineering Design*. Springer London. https://doi.org/10.1007/978-1-84628-319-2

Rane, N., Choudhary, S., & Rane, J. (2024). Gemini versus ChatGPT: applications, performance, architecture, capabilities, and implementation. *Journal of Applied Artificial Intelligence*, 5(1), 69–93. https://doi.org/10.48185/jaai.v5i1.1052

Spivey, N. (2019). *An Exploratory Study of the Influence of Design Process Ordering on the Requirement Generation of Novice Designers*. Clemson University.

Spivey, N., Ortiz, J., Patel, A., Davenport, B., & Summers, J. D. (2021). Analysis of the Impact of Requirement-Sketch Sequencing on Requirement Generation in Conceptual Design. *Journal of Mechanical Design*, 143(12). https://doi.org/10.1115/1.4051079

Ullman, D. G. (2017). *The Mechanical Design Process* (6th ed.). David Ullman LLC.

Ulrich, K., Eppinger, S., & Yang, M.C. (2020). *Product design and development* (7th Edition). McGraw Hill.

Valentine, R. (n.d.). *26 TH INTERNATIONAL CONFERENCE ON ENGINEERING AND PRODUCT DESIGN EDUCATION PROMPT ENGINEERING ON THE ENGINEERING DESIGN PROCESS*.

Valerio, M. A., Rodriguez, N., Winkler, P., Lopez, J., Dennison, M., Liang, Y., & Turner, B.J. (2016). Comparing two sampling methods to engage hard-to-reach communities in research priority setting. *BMC Medical Research Methodology*, 16(1), 1–11. https://doi.org/10.1186/s12874-016-0242-z

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press

Viswanathan, V. K., & Linsey, J.S. (2012). Physical models and design thinking: A study of functionality, novelty and variety of ideas. *Journal of Mechanical Design*, 134(9). https://doi.org/10.1115/1.4007148

Wang, L., Chen, X., Deng, X., Wen, H., You, M., Liu, W., Li, Q., & Li, J. (2024). Prompt engineering in consistency and reliability with the evidence-based guideline for LLMs. *Npj Digital Medicine*, 7(1), 41. https://doi.org/10.1038/s41746-024-01029-4

Wei, B. (2024). *Requirements are All You Need: From Requirements to Code with LLMs*.

Yasin, A., Fatima, R., JiangBin, Z., Ali Khan, J., & Ali Khan, A. (2024). Gamifying requirements: An empirical analysis of game-based technique for novices. *Journal of Software: Evolution and Process*, 36(6). https://doi.org/10.1002/smr.2617

https://doi.org/10.1017/pds.2025.10280 Published online by Cambridge University Press