Leveraging Generative AI for Requirement Gathering in Software Startups

Samuel Oluwagbemiga Ojo, Hussain Hedayati, Tejas Bhimrao Jambharunkar, Ugonna Chimezie

Gisma University of Applied Sciences, Potsdam, Germany
Scientific Work for Computer Science (SS0325)

Abstract

Requirement Engineering (RE) for software startups is usually troubled with insufficient resources, volatile goals, and high market volatility. Traditional RE practices, though formal, are usually bypassed for quick prototyping and user-friendly design. The recent advances in Generative AI (GenAI), and Large Language Models (LLMs) in particular, offer new opportunities to improve and automate RE processes. This study presents a comparative analysis of three GenAI-powered tools, ChatGPT, Elicitron, and ReqBrain, each realizing a distinct style of RE augmentation. ChatGPT provides adaptive prompt-based support for writing requirements; Elicitron leverages agent-based simulations to elicit latent user needs; and ReqBrain employs fine-tuned LLMs to generate ISO 29148-compliant requirements. Drawing on a multi-dimensional assessment framework, we compare each tool's performance across authenticity, adequacy, latent need extraction, and startup fit. Our findings suggest that there is no single, universal tool but that a hybrid solution can effectively supplement the RE lifecycle of startups. This contribution adds to the body of work on AI-aided RE and offers practical guidance for the integration of GenAI into agile, lightweight development environments.

Keywords: Requirement engineering, Software development, Software startups, Generative AI

1. Introduction

Background

Requirements engineering (RE) is a foundational process within software development that involves elicitation, analysis, specification, and validation of software requirements. In software startups, however, RE is largely informal, reactive, and obstructed by limited resources and rapidly changing goals. Studies

have reported that startups tend not to follow traditional RE practices and rather employ direct customer feedback and rapid prototyping, and hence miss requirements, feature creep, and traceability (Paternoster et al., 2014; Klotins et al., 2019).

Although agile and lean methods allow for flexibility, they are not strong enough to formalize tacit knowledge and ensure long-term scalability. This gap has triggered scholars to explore how emerging technologies, primarily Generative AI (GenAI), can support RE in startup settings. Large Language Models (LLMs) such as GPT-4 and Zephyr have been shown to have remarkable capabilities in natural language understanding and generation and are therefore potential candidates for complementing RE.

Rationale

Different approaches to integrating LLMs into RE processes have been explored in more recent research. Marques et al. (2024) deliberated on the use of ChatGPT for RE, its advantages of facilitating brainstorming, documentation, and stakeholder communication but also its limitations in the guise of hallucination and explainability. Ataei et al. (2024) suggested Elicitron, an LLM agent-based framework to simulate different user personas and uncover underlying needs by simulating product experiences using a systematic method. Habib et al. (2025) proposed ReqBrain, a tuned LLM trained on requirements compliant with ISO 29148, which was able to produce standard aligned and original requirements at high precision levels.

Objectives

This study builds on these contributions by performing a comparative assessment of ChatGPT, Elicitron, and ReqBrain. We evaluate the approach, outcome, and usability of each tool in startup environments according to a multi-dimensional framework of authenticity, adequacy, latent need detection, and suitability for startups. By combining these viewpoints, we aspire to provide practical recommendations for integrating GenAI into RE

2. Background

Software startups evolve in environments that are marked by rapid iteration, evolving goals, and scarce resources. Unlike mature organizations, startups tend to lack formalized processes and domain expertise, thus embracing informal and reactive RE practices. Studies such as Paternoster et al. (2014) and Klotins et al. (2019) have shown that startups often omit traditional RE processes in favor of customer feedback, intuition-driven decisions, and rapid prototyping.

Klotins et al. proposed a progression model mapping startup maturity stages (inception, stabilization, growth, maturity) to engineering process domains, indicating that RE in startups is more spontaneous than planned. This results in challenges such as:

- Challenges in capturing tacit and latent requirements.
- Frequent scope changes and feature creep.
- Lack of formal validation and traceability mechanisms.

These challenges underline the need for lightweight, adaptive, and scalable RE practices aligned with the rapid pace of startup environments. The emergence of Generative AI (GenAI), particularly Large Language Models (LLMs), offers promising opportunities to bridge these gaps.

2.1 Traditional and Agile RE Techniques: Strengths and Weaknesses

Traditional RE techniques, e.g., stakeholder interviews, use case modeling, and formal specifications, are well tested but too rigid for startup settings. Agile RE techniques such as user stories, backlogs, and feedback loops are more flexible but shallow in requirement analysis and validation.

Although agile methods allow rapid iteration, they are likely to fall short of:

- Uncovering tacit or hidden requirements.
- Formalization or scaling of compliance requirements.
- Traceability between development iterations.

This has led researchers to explore hybrid approaches and AI-based methods that combine agility and analytical accuracy.

2.2. Generative AI in Requirements Engineering

Recent advances in GenAI have created new opportunities for automating and enriching RE activities. Models such as GPT-4, Zephyr, and Mistral have been shown to have natural language understanding and generation abilities, making them suitable for uses like stakeholder simulation, requirement writing, and iteration.

2.2.1 ChatGPT

A comprehensive review of ChatGPT's use in RE was presented by Marques et al. (2024). According to their findings, ChatGPT is effective in:

- Brainstorming and idea exploration.
- Stakeholder communication.

- Rapid documentation and analysis of feedback.

But with constraints:

- Hallucination and factual inaccuracies.
- Transparency and absence of explainability.
- Ethics of bias and over saturation.

Despite these, ChatGPT is readily available and versatile, thus a convenient tool for startups when there is a need for rapid ideation and coarse requirement sketching.

2.2.2 Elicitron

Ataei et al. (2024) introduced Elicitron, an agent-based simulation system using LLMs to generate diverse user personas and simulate product experience. With systematic reasoning (Action, Observation, Challenge), Elicitron uncovers implicit needs that human interviews often miss.

Key findings are:

- Serial generation of agents yields the most diverse user needs.
- Elicitron outperforms human interviews in identifying implicit needs.
- Chain-of-thought reasoning improves classification accuracy (F1 = 0.95).
- Elicitron is particularly well suited for early design and empathic requirement elicitation, offering scalable, low-cost substitutes for traditional user studies.

2.2.3 ReqBrain

Habib et al. (2025) introduced ReqBrain, a task-oriented instruction-tuned LLM trained on ISO 29148-conformant requirements. ReqBrain is optimized to generate authentic and adequate requirements that can be used for formal use.

Highlights are:

- Beat ChatGPT-4o on BERTScore (F1 = 0.89) and FRUGAL (91.2).
- Human evaluators could not distinguish ReqBrain outputs from human-written requirements.
- Auto-generated requirements aligned with, missed from, and supplemented existing specifications.
- ReqBrain is better suited to formal RE tasks, especially for scaling and compliance-driven scenarios.

2.3 Comparative Insights and Research Gaps

While each tool is somewhat strong, there are some gaps:

- Limited empirical validation in actual startup settings.
- No agile toolchain integration (e.g., Jira, Trello).
- Human-AI collaboration frameworks are required.
- Ethical concerns of bias, privacy, and explainability.

These shortcomings support the argument in favor of hybrid workflow that combines the positives of prompt-based (ChatGPT), agent-based (Elicitron), and fine-tuned (ReqBrain) approaches to support the complete RE lifecycle in startups.

3. Research design and methodology

3.1 Research Design

This study uses mixed methods to evaluate the effectiveness of three GenAI-based tools (ChatGPT, Elicitron, ReqBrain) to find out the usefulness of each in supporting requirements engineering (RE) tasks. These are the questions for this research which should be answered by this methodology:

RQ 1: How do ChatGPT, Elicitron, and ReqBrain differ in their approach to RE?

RQ 2: What are the comparative strengths and limitations of each tool in generating authentic, adequate, and contextually relevant requirements?

RQ 3: How useful and suitable are these tools for the dynamic and resource limit environment of software startups?

This research is divided into three phases:

1. Literature Synthesis: Reviewing the foundational and recent research in startups and GenAI applications in RE.
2. Comparative Evaluation: studying published studies to analyse each tool, methodology, outputs, and evaluation metrics.
3. Synthesis and Interpretation: Cross case analysis to identify patterns, gaps, and implications for future research.

3.2 Data Sources

The study draws on the following primary sources:
Peer-reviewed and preprint research papers:
Marques et al. (2024) on ChatGPT in RE.
Ataei et al. (2024) on Elicitron.
Habib et al. (2025) on ReqBrain.

Supplementary literature on RE in software startups (e.g., Paternoster et al., Klotins et al.).
ISO/IEC/IEEE 29148:2018 standard for requirements quality.
Evaluation metrics such as BERTScore, FRUGAL, and human expert assessments.

3.3 Evaluation Framework

To ensure a consistent and rigorous comparison, we adopt a multi-dimensional evaluation framework with the following criteria:

Dimension	Description
Authenticity	How human-like and contextually appropriate the generated requirements are.
Adequacy	Alignment with ISO 29148 standards (completeness, clarity, consistency).
Latent Need Discovery	Ability to uncover unspoken or implicit user needs.
Startup Fit	Suitability for fast-paced, resource-constrained environments.
Evaluation Methodology	Rigor and transparency of the tool's validation process.

Each tool is assessed based on how it performs across these dimensions, using both quantitative metrics (e.g., BERTScore, F1) and qualitative insights (e.g., human evaluations, case studies).

3.4 Comparative Analysis Methodology

A comparative analysis follows these steps:
Tool profiling: Summarize the architecture, input/output design, and usage of each tool.
Methodology: Select and compare the test datasets, and evaluation criteria used in each study.
Synthesis of results: Aggregate and analyze the performance outcomes of the tools.
Context mapping: Map the strengths and limitations of each tool to the specific needs of software startups.

4. Results and Discussion

This section recapitulates the comparative findings of ChatGPT, Elicitron, and ReqBrain's analysis against their performance across key RE dimensions:

authenticity, adequacy, latent need discovery, and startup fit. Implications of the findings on RE practice in software startups are also considered.

4.1 Authenticity of Generated Requirements

Authenticity reflects the degree to which the generated requirements are human-like and contextually appropriate. Among the three tools:

ReqBrain was the most authentic. Human evaluators were not able to distinguish ReqBrain-generated from human-authored requirements consistently. It achieved a BERTScore F1 of 0.89 and a FRUGAL score of 91.2, indicating acceptable semantic consistency and fluency.

ChatGPT was moderately authentic. While capable of producing coherent requirements, it was very sensitive to the quality of the prompt and would likely require post-editing.

Elicitron was not so much examined for realism in the traditional sense but for its ability to come up with realistic user personas and experiences. However, its structured outputs (Action, Observation, Challenge) were human-like and contextually rich in narrative structure.

Insight: Fine-tuning (as in ReqBrain) significantly enhances the realism of generated requirements, especially when combined with alignment to formal standards.

4.2 Adequacy and Compliance with Standards

Sufficiency decides to what level the requirements being elicited align with ISO 29148 standards, including clarity, completeness, and consistency.

ReqBrain fared better in this regard. It was specifically trained using ISO-compliant datasets and outperformed both ChatGPT and its own untuned baseline on syntax written and signaling keyword usage. Human evaluations confirmed its outputs to be syntactically correct and standards-compliant.

ChatGPT created sufficient requirements in most situations but with no consistency. Without domain-specific tuning or prompt engineering, it generated too generic or ambiguous statements in most cases.

Elicitron was not developed for standards compliance but for ideation at an early stage. Its results were rich in context of the user but needed to be converted into formal requirements.

Insight: For startups planning to scale up or be compliance-aware, such tools as ReqBrain are clearly advantageous in producing high-quality, well-defined requirements.

4.3 Latent Need Discovery

Implicit or unconscious user needs are difficult to obtain through normal techniques. **Elicitron** performed most effectively in this sense. Agent-based simulations and empathic role-play were used by it to elicit a broader spectrum of needs than with conventional interviews. Controlled tests identified much more latent needs, especially when using chain-of-thought reasoning and guiding prompts, with it.

ChatGPT could elicit latent needs to the surface when cued with well-crafted prompts but lacked the structured process of Elicitron.

ReqBrain was not developed to specifically support latent need discovery but can facilitate it indirectly by iterative refinement and stakeholder simulation.

Insight: Elicitron is most valuable during the early stages of product development, where insight into user empathy and edge cases is critical.

4.4 Startup Fit and Pragmatism

This dimension assesses the degree to which each tool meets the software startup's limitations and workflows, speed, flexibility, and limited resources.

ChatGPT is highly flexible and usable, ideal for rapid ideation, writing, and low-fuss RE work. Having a low barrier to entry and integration with applications like Slack or Notion, it is suitable for startups.

Elicitron gives deeper insight but requires more configuration, computation, and scenario creation. It is best suited to design-intensive startups or those with UX research capacity.

ReqBrain produces high-quality, standards-compliant outputs but assumes fine-tuned models are available and may require integration with agile toolchains (e.g., Jira, GitHub).

Insight: ChatGPT is utilized by early-stage, rapid-fire startups; user-centered design utilizes Elicitron; and ReqBrain is utilized for scaling and formalization.

4.5 Summary of Comparative Findings

Feature	ChatGPT	Elicitron	ReqBrain
Core Approach	Prompt-based LLM (general-purpose)	Agent-based simulation using LLMs	Fine-tuned LLM trained on ISO 29148-compliant requirements
Primary Use Case	Rapid ideation, drafting, stakeholder communication	Discovery of latent/tacit user needs via simulated personas	Generation of formal, standards-compliant requirements
Interaction Style	Conversational, flexible	Structured simulation (Action, Observation, Challenge) + interviews	Instruction-based generation with task-specific tuning
Strengths	Easy to use Versatile Fast prototyping	High diversity in user needs Effective latent need discovery Empathic simulation	High authenticity (BERT F1 = 0.89) ISO 29148 compliance Human-indistinguishable output
Limitations	Hallucinations No standards compliance Prompt-sensitive	Requires setup and tuning Not standards-focused Less suitable for formal specs	Needs fine-tuning Less flexible for ideation Requires integration for agile workflows
Evaluation Metrics Used	Human judgment, literature review	Convex hull, silhouette score, latent need count, F1	BERTScore, FRUGAL, human evaluation, ISO compliance
Empirical Validation	Literature-based, some expert surveys	3 experiments: diversity, latent needs, classification	4-part study: benchmarking, human evaluation, ISO compliance, adequacy
Startup Fit	High (low barrier, fast iteration)	Medium (best for UX-heavy or early-stage design)	Medium-High (best for scaling, compliance, formalization)
Toolchain Integration	Easy (Slack, Notion, etc.)	Not integrated	Not integrated (but RAG-enabled version in development)
Unique Capabilities	Prompt engineering flexibility Stakeholder simulation	Empathic lead user simulation Chain-of-thought reasoning Diversity clustering	Instruction tuning ISO syntax and keyword compliance Human indistinguishability
Best Stage in RE Lifecycle	Early-stage ideation and drafting	Early-stage user research and latent need discovery	Mid-to-late stage formalization and compliance

Visual comparison of ChatGPT, Elicitron, and ReqBrain across key dimensions.

4.6. Implications for Startups

The findings suggest there isn't a one-tool-fits-all approach. Instead, startups can use a hybrid approach:

- Use Elicitron for initial ideation and empathic exploration.
- Use ChatGPT for rapid drafting and stakeholder collaboration.
- Use ReqBrain for formalization and calibration of requirements as the product matures.

4.7 Responsible and Ethical Use of GenAI in Requirements Engineering

Generative AI application in Requirements Engineering is also accompanied by threats such as hallucinated output, bias, inexplicability, overreliance, and data privacy are of significant concern since RE deals with sensitive stakeholders and system data. To avoid such threats, organizations can opt to employ a human-in-the-loop (HITL) paradigm, apply bias-reducing practices, enhance explainability with chain-of-thought models, and emphasize on-premise or open-source tools. Ethics guidelines such as IEEE's Ethically Aligned Design [IEEE, 2022], EU AI Act [EU, 2024], and ISO/IEC 42001 [ISO, 2023] establish formal

guidance in the areas of transparency, accountability, and risk management. Compliance with these guidelines renders the RE practice universally implementable and ensures trust in AI-driven development.

4.8 Limitations and Validity

While this study offers a structured comparative analysis of ChatGPT, Elicitron, and ReqBrain, several limitations must be acknowledged. First, the evaluation is based on secondary literature rather than direct experimentation, which may introduce inconsistencies due to differing methodologies and metrics across studies. Second, the absence of a unified benchmarking framework limits the precision of performance comparisons. Third, the tools were not tested in live startup environments, so contextual generalizability remains theoretical. Lastly, the rapid evolution of GenAI tools may affect the long-term relevance of the findings.

To ensure validity, a multi-dimensional evaluation framework was employed, grounded in RE literature and ISO 29148 standards. Internal validity was supported through systematic synthesis and tool profiling, while external validity is cautiously extended to agile and resource-constrained environments. All sources were peer-reviewed or preprint publications, and the methodology is replicable, supporting future empirical validation.

5. Conclusion

RE in software startups is a challenging task. Startups exist in rapidly evolving, resource-constrained environments with applications constantly changing. Traditional requirements engineering approaches are incapable of keeping pace. They cannot encapsulate the tacit knowledge of the team members, cannot cope with constant change, and are unable to track the evolution of requirements.

In this study, we compared three GenAI tools, ChatGPT, Elicitron, and ReqBrain, each following a unique approach towards facilitating RE:

ChatGPT: A versatile and multi-faceted LLM suited for quickly coming up with ideas and informally scribbling down requirements.

Elicitron: Uses agent-based simulation and formal reasoning to uncover underlying user needs and aid user-centered design.

ReqBrain: It is aligned to ISO 29148-conformant data and works best in producing high-quality, formalized requirements written using an expert voice.

5.1 Results and a layered approach:
Our results show that there is no one tool that can satisfy all needs. Rather, applying them together at different stages of the process yields the best results for new businesses. Elicitron works best in early stages in collecting user requirements. ChatGPT does best in iterative writing and stakeholder communication. ReqBrain does best in late stages in formalizing requirements and adherence to standards.

This multi-stage process mimics the operation of real startups: fast, iterative, and responsive. It enables teams to move fast without sacrificing quality or customer focus. But there always needs to be human intervention. Humans have to review the output created by AI to ensure that it is coherent, contextually appropriate, and conforms to ethical and practical needs.

In the future, greater time needs to be devoted to analyzing these tools longitudinally in real startup environments. Integration with popular development tools like Jira and GitHub will make them more usable. Discovering ways for humans and AI to communicate more effectively with each other during RE is also vital. And issues related to bias, transparency, and privacy should continue to be addressed.

In short, GenAI does have real potential to transform RE in startups. Applied wisely and in a brilliant mix, these technologies can complete the missing links between ad-hoc startup practices and proper engineering processes. And, naturally, help teams deliver better products faster, with fewer errors, and with improved market fit.

This study was able to answer its potential three most significant research questions by way of a systematic comparative examination of ChatGPT, Eliciton, and ReqBrain. This study began with explicit contrasts between each tool's (**RQ1**) reaction towards requirements engineering in the form of ChatGPT's prompt-based flexibility, Eliciton's agent-based unobtrusive need elicitation, and ReqBrain's calibrated compliance-minded generation. Secondly, the study determined the strengths and weaknesses of each instrument relative to criteria such as authenticity, startup fit, and adequacy (**RQ2**) by utilizing empirical and literature-based measures such as BERTScore and FRUGAL. Third, the suitability of these tools to startup environments (**RQ3**) was examined, and it was determined that the most effective and scalable approach is a hybrid approach with layers where each tool is applied in separate phases of the RE lifecycle. These results offer

practical guidance for the proper and effective use of GenAI in agile, resource-scarce development environments.

References

1. Arora, C., Grundy, J., & Abdelrazek, M. (2023). Advancing Requirements Engineering through Generative AI
2. Habib, M. K., Graziotin, D., & Wagner, S. (2025). ReqBrain: Task-Specific Instruction Tuning of LLMs for Requirements Generation
3. Marques, N., Silva, R. R., & Bernardino, J. (2024). Using ChatGPT in Software Requirements Engineering: A Comprehensive Review
4. Ronanki, K., Berger, C., & Horkoff, J. (2023). Investigating ChatGPT's Potential to Assist in Requirements Elicitation Processes
5. Ataei, M., Cheong, H., Grandi, D., Wang, Y., Morris, N., & Tessier, A. (2024). Elicitron: LLM Agent-Based Simulation for Design Requirements Elicitation
6. IEEE. (2022). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (1st ed.). Institute of Electrical and Electronics Engineers.
7. European Commission. (2024). Regulation (EU) 2024 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
8. International Organization for Standardization. (2023). ISO/IEC 42001:2023 Artificial intelligence — Management system — Requirements. ISO.
9. Paternoster, N., Giardino, C., Unterkalmsteiner, M., Gorschek, T., & Abrahamsson, P. (2014). Software development in startup companies: A systematic mapping study. Information and Software Technology, 56(10), 1200–1218.
10. Klotins, E., Gorschek, T., & Fagerholm, F. (2019). Requirements engineering in software startups: Practices and challenges. Empirical Software Engineering, 24(5), 3089–3123.

Appendix:

Summary of Key Literature and Theoretical Frameworks This appendix summarizes the key literature reviewed and theoretical frameworks utilized in the study.

1. Marques et al. (2024) conducted a comprehensive review of ChatGPT in software requirements engineering.
2. Ataei et al. (2024) introduced Elicitron, a multi-agent LLM framework for simulating diverse user personas

3. Habib et al. (2025) proposed ReqBrain, a fine-tuned LLM trained on ISO 29148-compliant requirements.
4. Arora et al. (2023) provided a SWOT analysis of LLMs in requirements engineering, identifying strengths
5. Ronanki et al. (2023) compared ChatGPT-generated requirements with those from human experts, finding Theoretical Frameworks:
   - ISO/IEC/IEEE 29148:2018 standard for quality requirements.
   - Chain-of-thought prompting for reasoning in LLMs.
   - Human-in-the-loop validation for AI-assisted RE.
   - Instruction tuning and low-rank adaptation (LoRA) for fine-tuning LLMs.

These studies collectively inform the design of GenAI-augmented workflows for software startups

AI Contribution Form

Acknowledgement

We would like to acknowledge the use of ChatGPT as an important source of information when conducting the study for this assignment. Although the data produced by the application wasn't directly used in the assignment, it was nevertheless a useful resource that helped students gain a deeper grasp of the topic. The tool was used for a variety of things, for instance,

1. The tool provides feedback on the validity and coherence of our research design and findings.
2. It provides clarity on how ReqBrain and Elicitron work.
3. The tool also provides suggestions for merging or renaming sections and improving clarity.