

# Um Caso de Experiência do Uso do ChatGPT na Especificação de Requisitos de Software

Iuri V. F. Costa<sup>1</sup>, Igor E. F. Costa<sup>1</sup>

<sup>1</sup>iuri.vf.costa@gmail.com; <sup>1</sup>igorcosta@ufpa.br

<sup>1</sup>UFPA - Universidade Federal do Pará

Rua Augusto Corrêa, 01 - Guamá, Belém - PA, 66075-110

**Abstract.** *Requirements specification is a crucial stage in software development. This study reports the experience of using Generative Artificial Intelligence (AI) as a support tool in this process. It explores how such technologies can enhance and accelerate complex tasks, particularly the generation of user stories (US) with acceptance criteria and field dictionaries. The results highlight benefits, challenges, and implications for professional practice, contributing to a deeper understanding of the applicability and effectiveness of generative AIs in software requirements engineering.*

**Keywords—** *Generative AI, Requirements Specification, Software Engineering*

**Resumo.** *A especificação de requisitos é uma etapa crucial no desenvolvimento de software. Este estudo relata a experiência do uso de Inteligência Artificial (IA) generativa como ferramenta de apoio nesse processo. O estudo explora como essas tecnologias podem aprimorar e acelerar tarefas complexas, sendo a geração de histórias de usuário (HU) com critérios de aceitação e dicionário de campos. Os resultados evidenciam benefícios, desafios e implicações à prática profissional havendo contribuições à compreensão da aplicabilidade e eficácia das IAs generativas na engenharia de requisitos de software.*

**Palavras-chave—** *IA Generativa, Especificação de Requisitos, Engenharia de Software*

## 1. Introdução

A aplicação de *Large Language Models* (LLMs) e da inteligência artificial tem se tornado cada vez mais presente no cotidiano, destacando-se pelo potencial de gerar textos coerentes e relevantes a partir de instruções simples (“*prompts*”). Ferramentas como ChatGPT, Gemini e CoPilot têm demonstrado esse potencial, despertando o interesse da sociedade e da indústria de software, que identifica na IA generativa uma aliada no desenvolvimento de soluções e produtos tecnológicos [Khojah and Neto 2024, Feuerriegel and Zschech 2024, Zhang and Liu 2023].

No ciclo de vida de desenvolvimento de software, a **Especificação de Requisitos de Software** (ERS) é uma fase crítica, pois define a base para a entrega de produtos que atendam às necessidades e expectativas das partes interessadas [Marques and Bernadino 2024, Yeow and Majid 2024]. Assim, o objetivo deste estudo é avaliar a produtividade na atividade de especificação de requisitos e seus impactos em tarefas correlatas, comparando o processo tradicional com o apoio da IA generativa.

Este artigo apresenta um **relato de experiência com caráter avaliativo**, conduzido em ambiente real de desenvolvimento de software. Analisa-se o impacto do uso do ChatGPT na produtividade e qualidade da especificação de requisitos, descrevendo a aplicação prática da ferramenta, os resultados obtidos e sua análise sob métricas quantitativas e qualitativas. O estudo contribui para a Engenharia de Requisitos ao demonstrar como o uso de IA generativa pode reduzir o esforço humano, padronizar histórias de usuário e aumentar a completude dos artefatos, alinhando-se aos princípios definidos por [Sommerville 2011]: entendimento claro das necessidades do cliente e documentação precisa dos requisitos.

Dessa forma, a questão de pesquisa que norteia este trabalho é: *Qual o impacto na produtividade da especificação de requisitos com a utilização do ChatGPT e em atividades correlacionadas?* Além desta introdução, a Seção 2 apresenta os trabalhos relacionados; a Seção 3 descreve a metodologia; a Seção 4 discute os resultados; a Seção 5 aborda as limitações do estudo; e a Seção 6 expõe as considerações finais e perspectivas futuras.

## 2. Trabalhos Relacionados

No estudo de [Marques and Bernadino 2024], são discutidos os benefícios, desafios e considerações éticas do uso do ChatGPT 3.5 na ERS, destacando como essa ferramenta pode remodelar as práticas de ERS e sugerindo recomendações estratégicas para sua aplicação. No trabalho de [Ronanki and Horkoff 2023] os requisitos gerados pelo ChatGPT são abstratos, atômicos, consistentes, corretos e compreensíveis, embora apresentem problemas de ambiguidade, comparando com a atuação de especialista humano. Para [Zhang and Liu 2023] o desempenho do ChatGPT em tarefas de recuperação de informações de requisitos, revela sua capacidade promissora de recuperar informações relevantes, mas com limitações na precisão para informações mais específicas.

O presente estudo se diferencia ao focar na geração de especificações de requisitos de software, incluindo critérios de aceite e dicionário de campos, utilizando um agente de ChatGPT especialista na geração de requisitos. O objetivo foi avaliar a redução de esforço e a melhoria na qualidade das especificações, analisada pelo time de desenvolvimento, bem como utilizando a Matriz SWOT (*Strengths, Weaknesses, Opportunities e Threats*) para a análise qualitativa.

## 3. Metodologia da Pesquisa

A pesquisa foi conduzida como um estudo de caso experimental, numa empresa de tecnologia com mais de 100 colaboradores, certificada no modelo MPS-BR (nível F para serviços e C para software). Para isso, houveram cinco etapas, conforme ilustrado na Figura 1.

Etapa 1 – As histórias de usuário (HU) foram geradas por analistas de requisitos, sem o uso do ChatGPT, estabelecendo a linha de base de produtividade.

Etapa 2 – Foram entrevistados 5 especialistas (2 seniores e 3 plenos) de um total de 20, pertencentes ao setor de fábrica de software, para compreender as práticas e dificuldades no processo de especificação.

Etapa 3 – Realizou-se a configuração de um agente do ChatGPT personalizado

ao contexto e linguagem técnica da organização, houve o treinamento desse agente personalizado, bem como a supervisão das saídas e, por fim, a geração das HUs. Para este último, as tarefas automatizadas foram: a) geração de histórias de usuário; b) definição de critérios de aceitação; c) criação do dicionário de campos.

Etapa 4 – Mediante a geração das HUs ocorreu a prototipação, pois percebeu-se que tal atividade seria melhor desempenhada ao obter a HU concluída após suas revisões.

Etapa 5 – Foram coletados dados quantitativos, sendo o tempo médio de execução, em horas, para tarefas de especificação e prototipagem nos meses de março (sem ChatGPT) e maio (com ChatGPT). Além disso, houve a avaliação qualitativa sendo obtido a partir de uma análise SWOT com os especialistas.

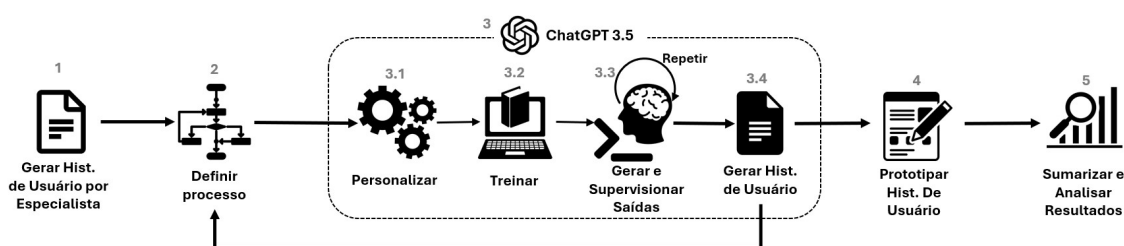


Figura 1. fluxograma do método da pesquisa

A partir do treinamento da ferramenta, desenvolveu-se alguns modelos de *prompts* para obter o melhor aproveitamento da eficiência da GenAI. A seguir apresentamos tais modelos *prompts* utilizados no de forma sequenciais para especificação dos requisitos de uma história de usuário. i) "crie uma história de usuário para [breve descrição da HU. Ex.: 'cadastro de colaborador']"; ii) "altere o perfil de usuário para [nome dos perfis]"; iii) "o produto é um sistema [web e/ou móvel], a ser utilizado pelo [descrição do cliente. Ex.: 'Órgão público de [...]']. Crie (ou ajuste) os critérios de aceitação"; iv) "Crie o dicionário de campos organizado em tabelas de Campos de Saídas, Campos de Entradas, Botões. As tabelas devem conter as colunas: Nome, Descrição, Natureza, Tamanho, Obrigatório, Validações, Regras". Contudo, é indispensável a supervisão do especialista e, se necessário, repetições das interações de *prompt*.

#### 4. Resultados e Discussões

Na Figura 2, observa-se que, em março, o tempo médio de execução das atividades foi de 3h01min, sendo 1h18min na especificação de requisitos (43,02%) e 1h43min na prototipagem (56,98%). Em maio, com a utilização do ChatGPT 3.5, o tempo médio reduziu para 1h50min, distribuído em 1h01min para especificação (56,08%) e 48min para prototipagem (43,92%), evidenciando ganhos de produtividade em ambas as atividades.

Foram adotadas as seguintes métricas de produtividade: (i) **Tempo médio de execução (TME)**, medido em horas e minutos; (ii) **Varição percentual de produtividade ( $\Delta\%$ )**, obtida pela redução do TME entre março e maio; (iii) **Complexidade média (SP)**, mensurada em *story points* atribuídos pelo time de desenvolvimento; e (iv) **Qualidade percebida**, avaliada quanto à completude, clareza e ausência de ambiguidade. Os resultados indicaram aumento de produtividade de 21,11% na especificação de requisitos e 53,35% na prototipagem. A complexidade média das HU geradas com o ChatGPT

(2,00 SP) foi superior àquelas criadas sem o uso da ferramenta (1,83 SP), refletindo maior qualidade percebida pelo time de desenvolvimento.

Na análise SWOT, os especialistas destacaram como **forças** a agilidade na execução das tarefas, a maior completude das HU e a redução de retrabalho; como **fraquezas**, a necessidade de treinamento frequente e atualização da base de conhecimento da ferramenta; como **ameaças**, a dependência de supervisão humana para garantir coerência e completude; e como **oportunidades**, o uso ampliado do ChatGPT em outras tarefas da engenharia de software, como prototipagem, geração e validação de cenários de teste, desenvolvimento de código e apoio às etapas iniciais de levantamento de requisitos.

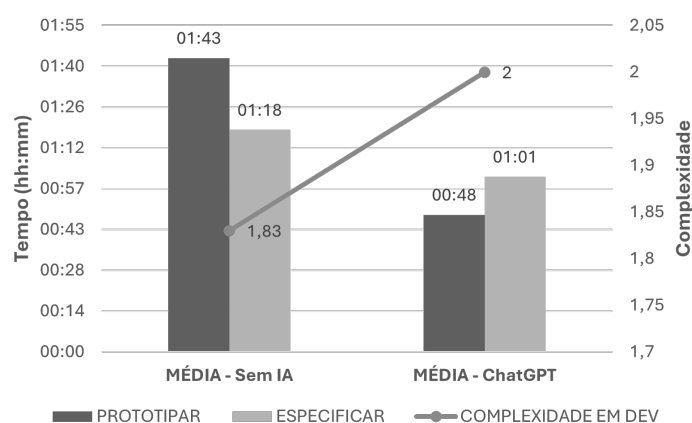


Figura 2. Média de tempo e complexidade nos meses de março e maio.

A **qualidade dos artefatos** produzidos foi analisada em três dimensões: (i) *Completeness*, observada em 92% das HU, que apresentaram todos os elementos estruturais esperados (título, descrição, critérios e campos); (ii) *Consistência*, com 8% apresentando ambiguidades terminológicas corrigidas durante revisão técnica; e (iii) *Aderência às práticas de ERS*, evidenciada pelo aumento médio de 9% nos *story points*, indicando maior riqueza informacional.

Conclui-se que o ChatGPT gera artefatos com qualidade equivalente ou superior às práticas humanas, desde que supervisionado por especialistas. Ressalta-se que o experimento foi conduzido entre março e maio de 2024, período em que o modelo GPT-3.5 era o mais estável e amplamente disponível para uso organizacional, garantindo reprodutibilidade e consistência dos resultados.

## 5. Limitações do Estudo

As principais limitações identificadas são: i) **Tamanho da Amostra e Escopo**: O estudo foi conduzido com um número limitado de participantes, o que pode restringir a generalização dos resultados para contextos mais amplos; ii) **Duração Curta**: A análise comparativa foi realizada em um período de apenas dois meses, o que pode não ser suficiente para capturar tendências de longo prazo ou variações sazonais no processo de especificação de requisitos; iii) **Ausência de Revisão por Pares**: A falta de uma revisão por pares mais abrangente limita a validação das percepções de produtividade e qualidade observadas no estudo; iv) **Potencial Viés**: Não foram implementadas medidas específicas para mitigar o efeito *Hawthorne*, onde os participantes podem melhorar seu desempenho

por estarem sendo observados, o que pode ter influenciado positivamente os resultados e; v) **Contexto Específico:** O estudo foi limitado a projetos de software para o setor público, o que pode não refletir comportamentos ou resultados em outros setores ou tipos de projetos de software.

## 6. Considerações Finais e Trabalhos Futuros

Este estudo apresentou a utilização do ChatGPT, em projetos de software de órgãos públicos, desenvolvidos por uma empresa de tecnologia utilizando métodos ágeis. Observou-se um aumento na produtividade da ERS, com impactos positivos em tarefas relacionadas, como a prototipagem. Em síntese, o estudo contribui para as áreas acadêmica e profissional, principalmente em inovação e otimização na ERS, ao fornecer um entendimento detalhado sobre a aplicabilidade da IA na engenharia de requisitos e ao explorar suas implicações práticas e teóricas, encorajando novas pesquisas e práticas mais eficientes no desenvolvimento de software.

Para trabalhos futuros planeja-se avaliar a eficácia do ChatGPT na ERS num período de mínimo de 6 meses, inclusive com versões mais recentes do ChatGPT, a fim de identificar tendências para uma análise mais abrangente. Busca-se também ampliar a amostra, incluindo uma diversidade maior de projetos e profissionais de diferentes setores e realizar um estudo comparativo entre o ChatGPT e outras ferramentas de IA. Além disso, serão desenvolvidas métricas específicas para avaliar a qualidade dos requisitos gerados, com revisões por pares e feedback dos stakeholders.

## Referências

- Feuerriegel, S.; Hartmann, J. J. C. and Zschech, P. (2024). Generative ai. In *Bus Inf Syst Eng* 66, 111–126. <https://doi.org/10.1007/s12599-023-00834-7>.
- Khojah, R.; Mohamad, M. L. P. and Neto, F. (2024). Beyond code generation: An observational study of chatgpt usage in software engineering practice. In *Proc. ACM Softw. Eng.* 1, FSE, Article 81. <https://doi.org/10.1145/3660788>.
- Marques, N.; Silva, R. and Bernadino, J. (2024). Using chatgpt in software requirements engineering: A comprehensive review. In *Future Internet Journal*. <https://doi.org/10.3390/fi16060180>.
- Ronanki, K.; Berger, C. and Horkoff, J. (2023). Investigating chatgpt's potential to assist in requirements elicitation processes. In *2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. <https://10.1109/SEAA60479.2023.00061>.
- Sommerville, I. (2011). Software engineering. In *9th Edition*. Addison-Wesley, Boston.
- Yeow, J.; Muhammad, R. and Majid, N. (2024). An automated model of software requirement engineering using gpt-3.5. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*. <https://10.1109/ICETISIS61505.2024.10459458s>.
- Zhang, J.; Chen, Y. N. N. W. Y. and Liu, C. (2023). Empirical evaluation of chatgpt on requirements information retrieval under zero-shot setting. In *Proceedings of the 2023 International conference on intelligent computing and next generation networks (ICNGN)*. <https://10.1109/ICNGN59831.2023.10396810>.