



UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE COMPUTAÇÃO - ICOMP
PROGRAMA PÓS-GRADUAÇÃO EM INFORMÁTICA - PPGI

Uso de LLMs na Automatização da Geração de Histórias de Usuário de Qualidade

Reine Menezes dos Santos

Manaus - AM

Março de 2026

Reine Menezes dos Santos

Uso de LLMs na Automatização da Geração de Histórias de Usuário de Qualidade

Dissertação submetida ao Programa de Pós-Graduação em Informática do Instituto de Computação da Universidade Federal do Amazonas (PPGI-UFAM) como requisito para a obtenção do título de Mestre em Informática.

Orientadora

Profa. Dra. Ana Carolina Oran

Coorientadora

Profa. Dra. Tayana Uchôa Conte

Universidade Federal do Amazonas - UFAM

Instituto de Computação - IComp

Manaus - AM

Março de 2026

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

S237u Santos, Reine Menezes dos
 Uso de LLMs na Automatização da Geração de Histórias de Usuário de
 Qualidade / Reine Menezes dos Santos. - 2026.
 248 f. : il., color. ; 31 cm.

 Orientador(a): Ana Carolina Oran Rocha.
 Coorientador(a): Tayana Uchôa Conte.
 Dissertação (mestrado) - Universidade Federal do Amazonas, Programa
 de Pós-Graduação em Informática, Manaus, 2026.

 1. Engenharia de Requisitos. 2. Histórias de Usuário. 3. Large Language
 Models - LLMs. 4. Prompt Engineering. 5. Qualidade das Histórias de
 Usuário. I. Rocha, Ana Carolina Oran. II. Conte, Tayana Uchôa. III.
 Universidade Federal do Amazonas. Programa de Pós-Graduação em
 Informática. IV. Título



Ministério da Educação
Universidade Federal do Amazonas
Coordenação do Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

"USO DE LLMS NA AUTOMATIZAÇÃO DA GERAÇÃO DE HISTÓRIAS DE USUÁRIO DE QUALIDADE"

REINE MENEZES DOS SANTOS

Dissertação de Mestrado defendida e aprovada pela banca examinadora constituída pelos Professores:

Prof.^a Dr.^a Ana Carolina Oran Rocha - **Presidente**

Prof. Dr. Márcio de Medeiros Ribeiro - **Membro Externo**

Documento assinado digitalmente
MARCIO DE MEDEIROS RIBEIRO
Data: 24/03/2026 15:03:42-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Awdren de Lima Fontão - **Membro Externo**

Documento assinado digitalmente
AWDREN DE LIMA FONTAO
Data: 25/03/2026 10:15:58-0300
Verifique em <https://validar.iti.gov.br>

Manaus, 20 de Março de 2026.



Documento assinado eletronicamente por **Ana Carolina Oran Rocha, Professor do Magistério Superior**, em 23/03/2026, às 10:38, conforme horário oficial de Manaus, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufam.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3032880** e o código CRC **84727987**.

Avenida General Rodrigo Octávio, 6200 - Bairro Coroado I Campus Universitário Senador Arthur Virgílio Filho, Setor Norte - Telefone: (92) 3305-1181 / Ramal 1193
CEP 69080-900, Manaus/AM, coordenadorppgi@icomp.ufam.edu.br

"Dedico este trabalho aos meus pais, que me ensinaram o valor da honestidade, da persistência e do amor silencioso. Obrigada por serem minha base firme, mesmo quando o caminho parecia incerto. E ao meu namorado, companheiro de todas as horas, que me segurou com paciência quando eu pensei em desistir, me fez rir nos dias mais nublados e acreditou em mim até quando eu não conseguia. Esta conquista também é sua, mô."

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, pela oportunidade que me foi concedida. Todos os dias busco honrar a confiança que depositaste em mim.

Aos meus pais, por todo apoio, pelo exemplo e, principalmente, por me incentivarem a cada dia nessa jornada. Se este sonho hoje, se torna realidade, é graças a vocês. E ao meu namorado, por estar sempre ao meu lado, estudando comigo, me ensinando e me apoiando em todos os momentos. Obrigada por ter vivido comigo cada fase desta etapa da minha vida.

Ao meu querido professor Bruno Gadelha, que me apresentou ao caminho da pesquisa em Engenharia de Software. Sempre foi meu exemplo e guia durante toda a graduação e grande parte do mestrado.

À minha orientadora, profa. Ana Carolina, que foi mais do que uma guia acadêmica, foi inspiração, porto seguro e exemplo de dedicação. Obrigada por acreditar em mim mesmo nos momentos em que duvidei, por sua paciência incansável, pela sabedoria compartilhada e pelo cuidado humano em cada orientação. Sua presença fez diferença não apenas na minha formação profissional, mas também na minha vida pessoal. Levo comigo não só os ensinamentos de pesquisa, mas também o exemplo de generosidade, respeito, amor e paixão pelo que se faz.

À minha querida coorientadora, profa. Tayana Conte, exemplo de pesquisadora, professora e mulher determinada. Obrigada por todo carinho, pela confiança e pela fé que sempre depositou em mim. É uma alegria enorme tê-la em minha vida e um privilégio poder aprender com a senhora.

Aos amigos do grupo USES, que fizeram parte dessa trajetória e conquista,

compartilhando comigo muitos momentos de aprendizado, vitórias e alegrias. Foi uma honra crescer ao lado de vocês. Tenho certeza de que nos encontraremos em parcerias ao longo da nossa vida profissional. E às minhas amigas da secretaria, que sempre me apoiaram e me acompanharam durante essa caminhada. Obrigada pelos cafés, pelas conversas animadas e pelos conselhos tão valiosos.

À Universidade Federal do Amazonas (UFAM) e ao Instituto de Computação (IComp), que foram minha casa desde a graduação. Aqui tive a oportunidade de aprender não apenas o conhecimento técnico que me guia como pesquisadora, mas também valores humanos que levo para a vida. Cada professor deixou em mim um pedaço do que hoje me tornei, seja como profissional, seja como pessoa. Serei eternamente grata por ter crescido nesse espaço e por tudo o que ele representa na minha trajetória.

"Nas grandes dificuldades é que se reconhecem os grandes Espíritos."

Allan Kardec

Uso de LLMs na Automatização da Geração de Histórias de Usuário de Qualidade

Autor: Reine Menezes dos Santos

Orientador: Profa. Dra. Ana Carolina Oran

Coorientadora

Profa. Dra. Tayana Uchôa Conte

Resumo

A fase de especificação de requisitos é o processo de registrar os requisitos do sistema de forma clara, completa e consistente, utilizando linguagem acessível e elementos simples. Porém, pode ser desafiadora na prática, devido a diferentes interpretações e expectativas dos stakeholders. Em metodologias ágeis, esse desafio se amplia, pois exige comunicação contínua, rápida adaptação a mudanças e intensa colaboração entre stakeholders. Nesse cenário, as histórias de usuário consolidaram-se como uma técnica adotada por sua simplicidade, foco no usuário e alinhamento ao desenvolvimento iterativo. No entanto, a escrita manual dessas histórias frequentemente apresenta problemas de qualidade, dificultando a compreensão compartilhada dos requisitos e impactando negativamente o produto final. Com o avanço dos Modelos de Linguagem de Grande Porte (LLMs), como o ChatGPT, surgem novas oportunidades para automatizar a geração de histórias de usuário. Ainda assim, a literatura carece de estudos que investiguem de forma experimental a eficácia dessas abordagens, bem como a qualidade dos textos gerados, a percepção de utilidade pelos profissionais e a aplicação em diferentes contextos. Diante disso, este trabalho tem como objetivo apoiar engenheiros de requisitos, responsáveis pela especificação de requisitos na geração automatizada de histórias de usuário de qualidade. Para isso, foi criado de forma iterativa o US-Prompt, um artefato

baseado na combinação das técnicas de Meta-Prompting e Few-Shot Prompting. Como evolução do artefato, foi também implementada a ferramenta web Story AI Generator, que operacionaliza o US-Prompt em um ambiente interativo para geração e refinamento automatizado de histórias de usuário. Para conduzir esta pesquisa, adotou-se a metodologia Design Science Research (DSR), que foca no desenvolvimento de artefatos para resolver problemas práticos, por meio da condução de estudos experimentais articulados em ciclos de relevância, design e rigor. O artefato desenvolvido foi avaliado com base no framework Quality User Story (QUS), que define critérios de qualidade para histórias de usuário, além da análise da percepção dos usuários quanto à utilidade, facilidade de uso e intenção de uso futuro, fundamentada no modelo TAM3. Os resultados indicam que o uso de LLMs com técnicas de prompting é promissor para apoiar a geração de histórias de usuário mais claras, consistentes e aderentes às boas práticas da engenharia de requisitos. A ferramenta desenvolvida demonstrou potencial de adoção em contextos reais, ampliando a aplicabilidade da abordagem proposta. Conclui-se que a aplicação de LLMs com técnicas como Meta-Prompting e Few-Shot Prompting, aliada à sua operacionalização em ferramentas de apoio, pode contribuir de forma eficaz para a geração automatizada de histórias de usuário de qualidade em contextos ágeis, aumentando a produtividade, a consistência e a qualidade dos requisitos, ao mesmo tempo em que reduz o esforço cognitivo dos profissionais envolvidos.

Palavras-chave: Engenharia de Requisitos, Histórias de Usuário, Modelos de Linguagem de Grande Porte, Engenharia de Prompt, Qualidade das Histórias de Usuário.

Uso de LLMs na Automatização da Geração de Histórias de Usuário de Qualidade

Autor: Reine Menezes dos Santos

Orientador: Profa. Dra. Ana Carolina Oran

Coorientadora

Profa. Dra. Tayana Uchôa Conte

Abstract

The requirements specification phase is the process of recording system requirements in a clear, complete, and consistent manner, using accessible language and simple elements. However, it can be challenging in practice due to differing interpretations and stakeholder expectations. In agile methodologies, this challenge is amplified as it demands continuous communication, rapid adaptation to change, and intense collaboration among stakeholders. In this scenario, user stories have become a widely adopted technique due to their simplicity, user-centric focus, and alignment with iterative development. Nevertheless, the manual writing of these stories often suffers from quality issues, hindering a shared understanding of requirements and negatively impacting the final product. With the advancement of Large Language Models (LLMs), such as ChatGPT, new opportunities arise to automate the generation of user stories. Still, the literature lacks experimental studies investigating the effectiveness of these approaches, the quality of the generated texts, professional perception of utility, and their application in different contexts. Given this, this work aims to support requirements engineers responsible for specification in the automated generation of high-quality user stories. To achieve this, US-Prompt was iteratively created—an artifact based on the combination of Meta-Prompting and Few-Shot Prompting techniques. As an evolution of this artifact,

the web tool Story AI Generator was implemented, which operationalizes US-Prompt in an interactive environment for the automated generation and refinement of user stories. To conduct this research, the Design Science Research (DSR) methodology was adopted, focusing on developing artifacts to solve practical problems through experimental studies articulated in relevance, design, and rigor cycles. The developed artifact was evaluated based on the Quality User Story (QUS) framework, which defines quality criteria for user stories, as well as an analysis of user perception regarding usefulness, ease of use, and intention to use, grounded in the TAM3 model. The results indicate that using LLMs with prompting techniques is promising for supporting the generation of clearer, more consistent user stories that adhere to requirements engineering best practices. The developed tool demonstrated potential for adoption in real-world contexts, expanding the applicability of the proposed approach. It is concluded that the application of LLMs with techniques such as Meta-Prompting and Few-Shot Prompting, combined with their operationalization in support tools, can effectively contribute to the automated generation of quality user stories in agile contexts, increasing productivity, consistency, and requirements quality while reducing the cognitive effort of the professionals involved.

Keywords: Requirements Engineering, User Stories, Large Language Models, Prompt Engineering, Quality of User Stories.

LISTA DE ILUSTRAÇÕES

Figura 1 – Metodologia de desenvolvimento da pesquisa.	28
Figura 2 – Fluxograma do Questionário com Profissionais de Requisitos.	59
Figura 3 – Distribuição dos cargos dos participantes do estudo.	62
Figura 4 – Distribuição do uso de Histórias de Usuários pelos participantes do estudo.	64
Figura 5 – Percentual de participantes que selecionaram cada critério de qualidade para histórias de usuário.	68
Figura 6 – Diagrama das etapas metodológicas da pesquisa.	88
Figura 7 – Comparação entre o número de histórias de usuários geradas por métodos manuais e Free-Form Prompt.	92
Figura 8 – Comparação de taxas de sucesso entre os métodos Manual e Prompt de Forma Livre.	93
Figura 9 – Gráfico apresentando indicadores por meio de representações visuais: gráficos de densidade, boxplots e pontos de dados individuais.	95
Figura 10 – Captura de tela da conversa no ChatGPT, demonstrando o resultado da aplicação do prompt final.	101
Figura 11 – Taxa de sucesso dos métodos Free-Form Prompt e US-Prompt.	106
Figura 12 – Comparação dos métodos em relação às categorias de qualidade.	107
Figura 13 – Histórias de usuário produzidas a partir do prompt padrão, organizadas por participante de cada grupo, com representação em barras coloridas.	121

Figura 14 – Distribuição das taxas de sucesso por participante na geração de histórias de usuários utilizando o US-Prompt Padrão.	122
Figura 15 – Comparação das taxas de sucesso entre o Prompt Padrão e o Prompt Modificado nos seis grupos avaliados.	127
Figura 16 – Estratégias de melhorias adotadas por cada grupo.	128
Figura 17 – Gráfico de boxplot comparando a distribuição dos resultados.	130
Figura 18 – Gráfico apresentando indicadores por meio de representações visuais: gráficos de densidade, boxplots e pontos de dados individuais.	131
Figura 19 – Distribuição das respostas para o construto Facilidade de Uso Percebida.	133
Figura 20 – Distribuição das respostas para o construto Utilidade Percebida.	134
Figura 21 – Distribuição das respostas para o construto Intenção de Uso Futuro.	135
Figura 22 – Guia de uso da ferramenta Story AI Generator.	150
Figura 23 – Formulário de entrada de dados da Story AI Generator.	151
Figura 24 – Interface de saída com histórias geradas.	152
Figura 25 – Registros do processo de condução das entrevistas.	157
Figura 26 – Distribuição do nível de familiaridade com ferramentas de IA generativas.	160
Figura 27 – Distribuição das respostas para o construto Facilidade de Uso Percebida.	163
Figura 28 – Distribuição das respostas para o construto Utilidade Percebida.	164
Figura 29 – Distribuição das respostas para o construto Intenção de Uso Futuro.	164
Figura 30 – Distribuição das respostas para o construto Relevância Percebida.	165
Figura 31 – Distribuição das respostas para o construto Qualidade de Saída Percebida.	166
Figura 32 – Diagrama de relações sobre os benefícios observados ao usar a Story AI Generator.	168
Figura 33 – Diagrama de relações sobre o uso percebido da ferramenta.	169

LISTA DE TABELAS

Tabela 1 – Classificação das Perguntas do Questionário por Seção.	60
Tabela 2 – Descrição dos critérios derivados do framework Quality User Story (QUS).	86
Tabela 3 – Resultados estatísticos da comparação da qualidade das histórias de usuários geradas manualmente e com prompt livre	94
Tabela 4 – Exemplos de histórias de usuários geradas manualmente e com US-Prompt.S	106
Tabela 5 – Associação dos Grupos aos Projetos.	122
Tabela 6 – Resultados estatísticos sobre a qualidade das histórias de usuários com Prompts Padrão e Modificado	129
Tabela 7 – Caracterização dos participantes.	159

LISTA DE ABREVIATURAS E SIGLAS

DSR *Design Science Research* (Pesquisa Baseada em Design)

ER Engenharia de Requisitos

GPT *Generative Pre-trained Transformer*

IA Inteligência Artificial

NLP *Natural Language Processing* (Processamento de Linguagem Natural)

QP Questão de Pesquisa

RaT *Refine and Thought*

RSL Revisão Sistemática da Literatura

UML *Unified Modeling Language* (Linguagem de Modelagem Unificada)

US-Prompt *User Story Prompt*

UST *User Story Tutor*

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Contextualização ou Definição do Problema	21
1.2	Questão de Pesquisa	24
1.3	Objetivos da Pesquisa	25
1.4	Metodologia de Pesquisa	26
1.5	DSR Aplicado nessa Pesquisa	27
1.5.1	Ciclo de Relevância	28
1.5.2	Ciclo de Design	29
1.5.3	Ciclo de Rigor	30
1.5.4	Requisitos de Aceite da Solução	31
1.6	Organização do Trabalho	32
2	CICLO DE RELEVÂNCIA: FUNDAMENTOS	34
2.1	Metodologias Ágeis	34
2.2	Engenharia de Requisitos	36
2.3	User Stories	38
2.4	NLP e LLMs	43
2.5	Tarefas de engenharia de requisitos e LLMs	44
2.5.1	Geração:	45
2.5.2	Avaliação de Qualidade:	45
2.5.3	Análise:	46
2.5.4	Tradução:	46
2.6	Técnicas de Engenharia de Prompts	47
2.7	Trabalhos Relacionados	48

3	CICLO DE RELEVÂNCIA: SURVEY COM PROFISSIONAIS DA INDÚSTRIA	55
3.1	INTRODUÇÃO	55
3.2	METODOLOGIA	56
3.2.1	Desenho da pesquisa	57
3.2.2	Execução	59
3.2.3	Resultados da Pesquisa	61
3.2.4	Discussões dos Resultados do Estudo	73
3.2.4.1	QP1: Como os profissionais da indústria lidam com requisitos em suas atividades cotidianas?	73
3.2.4.2	QP2: De que forma os profissionais utilizam Modelos de Linguagem de Grande Escala (LLMs) no contexto da engenharia de requisitos?	76
3.2.4.3	RQ3: Qual é a percepção dos profissionais sobre a geração automatizada de requisitos, especialmente histórias de usuários, com o uso de LLMs?	77
3.3	Ameaças à Validade	80
3.4	CONCLUSÃO	80
4	1º CICLO DE DESIGN: HISTÓRIAS DE USUÁRIOS: O CHATGPT FAZ MELHOR?	84
4.1	INTRODUÇÃO	84
4.2	MÉTODO	85
4.3	ESTUDO EXPERIMENTAL 1: COMPARAÇÃO ENTRE GERAÇÃO MANUAL E PROMPT LIVRE	89
4.3.1	Desenho do Estudo	89
4.3.2	Execução do Estudo	90
4.3.3	Resultados e Discussão	91
4.4	Abordagem US-Prompt (Versão 1)	96
4.4.1	Design da Abordagem	99
4.4.2	Estrutura de Abordagem	100

4.5	ESTUDO EXPERIMENTAL 2: PROPOSTA E AVALIAÇÃO DA ABORDAGEM US-PROMPT	103
4.5.1	Desenho do Estudo	103
4.5.2	Resultados e Discussão	104
4.6	AMEAÇAS À VALIDADE	109
4.7	CONCLUSÃO	111
5	2º CICLO DE DESIGN: HISTÓRIAS DE USUÁRIOS GERADAS POR IA. ELAS SÃO BOAS O SUFICIENTE?	113
5.1	INTRODUÇÃO	113
5.2	Evolução do Prompt para Geração de Histórias de Usuário	114
5.3	Estudo Experimental	115
5.3.1	Desenho do Estudo	116
5.3.1.1	Hipóteses	116
5.3.1.2	Variáveis	117
5.3.1.3	Participantes	117
5.3.2	Execução do Estudo	117
5.3.3	Coleta de Dados	118
5.3.4	Avaliação das Histórias de Usuários Geradas	119
5.4	Resultados e Discussão	120
5.4.1	Análise Quantitativa	120
5.4.1.1	Análise das Estratégias de Refinamento dos Prompts	122
5.4.1.2	Comparação entre as Taxas de Sucesso dos Prompts utilizados:	126
5.4.1.3	Análise Estatística dos Resultados	128
5.4.2	Avaliação da Aceitação e Percepções dos Participantes	132
5.4.2.1	Percepções Negativas	135
5.4.2.2	Percepções Positivas	138
5.4.3	Recomendações de Melhoria	139
5.5	Limitações do Estudo	140
5.6	Conclusão	141

6	3º CICLO DE DESIGN: FERRAMENTA WEB - STORY AI GENERATOR: CRAFTING USER STORIES	146
6.1	Introdução	146
6.2	Ferramenta Story AI Generator: Crafting User Stories	148
6.2.1	Objetivos da Ferramenta	148
6.2.2	Tecnologias Utilizadas	149
6.2.3	Interface e Fluxo de Uso	149
6.2.4	Geração das Histórias com ChatGPT	152
6.3	Estudo Experimental Exploratório com Profissionais de Engenharia de Software	153
6.3.1	Metodologia do Estudo	153
6.3.1.1	Desenho do Estudo	153
6.3.1.1.1	Participantes:	153
6.3.1.1.2	Instrumentação:	154
6.3.1.2	Execução do Estudo	155
6.3.1.2.1	Execução do Estudo Piloto:	155
6.3.1.2.2	Execução do Estudo Experimental:	156
6.3.2	Resultados do Estudo	157
6.3.2.1	Resultados do Questionário: Avaliação da Aceitação	162
6.3.2.2	Resultados das Entrevistas: Percepções dos Participantes	166
6.4	Limitações do Estudo	176
6.5	Conclusão	177
7	CONSIDERAÇÕES FINAIS E CONTRIBUIÇÕES DA PESQUISA	179
7.1	Considerações Finais	179
7.2	Contribuições da Pesquisa	181
7.3	Produção Científica	182
7.3.1	Artigos Aceitos	182
7.4	Trabalhos Futuros	183
	Referências	185

APÊNDICE A	QUESTIONÁRIO UTILIZADO NO SURVEY COM PROFISSIONAIS DA INDÚSTRIA	193
APÊNDICE B	US-PROMPT PARA GERAÇÃO DE HISTÓRIAS DE USUÁRIO - VERSÃO 1	203
APÊNDICE C	US-PROMPT PARA GERAÇÃO DE HISTÓRIAS DE USUÁRIO - VERSÃO 2	209
APÊNDICE D	CONTEXTOS DOS SISTEMAS DESENVOLVIDOS PELOS PARTICIPANTES	215
APÊNDICE E	EXEMPLOS DE CENÁRIOS DE SISTEMAS UTI- LIZADOS NO ESTUDO DA FERRAMENTA STORY AI GENERATOR	223
APÊNDICE F	QUESTIONÁRIO UTILIZADO NO ESTUDO DA FER- RAMENTA STORY AI GENERATOR	230
APÊNDICE G	CÓDIGOS EMERGIDOS NA ETAPA DE CODIFI- CAÇÃO ABERTA	242

1

INTRODUÇÃO

Este capítulo apresenta a motivação para a realização desta pesquisa, contextualizando o trabalho, além de expor a definição do problema, os objetivos propostos e a metodologia de pesquisa adotada. Por fim, é apresentada a estrutura desta pesquisa.

1.1 Contextualização ou Definição do Problema

No âmbito do desenvolvimento da Engenharia de Software, a etapa de requisitos desempenha um papel crucial, pois define o que precisa ser desenvolvido para atender às necessidades e expectativas dos usuários finais ([SOMMERVILLE, 2011](#)). Entre as atividades que compõem essa etapa, a elicitação de requisitos é frequentemente considerada o processo mais desafiador no desenvolvimento de software, sendo também citada como a principal causa de falha em projetos ([MARTINS et al., 2019](#)). Ela consiste em um processo colaborativo e analítico que busca identificar as necessidades e restrições das diversas partes interessadas de um sistema de software. Essa atividade envolve a realização de tarefas como coletar, descobrir, extrair e definir requisitos. Para isso, é essencial estabelecer uma comunicação direta com as partes interessadas, empregando técnicas variadas para compreender suas necessidades e expectativas. No entanto, esse processo frequentemente enfrenta desafios, como declarações vagas e genéricas, que demandam maior esclarecimento e detalhamento para se transformarem em requisitos específicos e mensuráveis ([TORRES-IGARTUA et al., 2024](#)).

Esses desafios levam à etapa de especificação de requisitos, que é a etapa responsável por registrar os requisitos do usuário e do sistema, garantindo que sejam claros, não ambíguos, compreensíveis, completos e consistentes. Porém, a especificação de requisitos também pode ser desafiadora na prática, devido a diferentes interpretações dos stakeholders e a possíveis conflitos entre os requisitos ([SOMMERVILLE, 2011](#)).

Os requisitos do usuário devem descrever funcionalidades e aspectos não funcionais de forma compreensível por usuários sem conhecimento técnico, para isso, recomenda-se o uso de linguagem natural e elementos simples, como tabelas e diagramas intuitivos ([SOMMERVILLE, 2011](#)). A avaliação da qualidade dos requisitos pode ocorrer em dois níveis: individual, onde analisa-se ambiguidade e conformidade com padrões e documental onde a análise considera o conjunto para detecção de inconsistências e lacunas ([NORHEIM et al., 2024](#)).

Nos últimos anos, o cenário do desenvolvimento de software tem experimentado uma mudança significativa com a consolidação de metodologias ágeis, que enfatizam a flexibilidade, a colaboração e a capacidade de responder rapidamente às mudanças nos requisitos dos projetos. Porém, a natureza dinâmica dos ambientes ágeis, caracterizada por requisitos em constante evolução e pela necessidade de rápida adaptação às necessidades do cliente, torna desafiadora a criação de especificações precisas. Esse desafio é ainda mais intensificado em contextos iterativos, nos quais o desenvolvimento é tanto adaptativo quanto iterativo, exigindo comunicação eficaz. A ênfase em entregas rápidas pode levar a mal-entendidos e dificultar a manutenção de um entendimento compartilhado entre as partes interessadas ([RASHEED, 2021](#)).

Esses fatores contribuem para que equipes ágeis frequentemente enfrentem desafios como negligência em relação a requisitos de qualidade, foco excessivo em documentação mínima e má priorização. Mudanças frequentes nos requisitos causam instabilidade, enquanto especificações e estimativas de esforço imprecisas complicam o planejamento. Segundo [Hoy e Xu \(2023\)](#), problemas de comunicação entre as partes interessadas impedem o progresso dos projetos.

Levando isso em consideração, as histórias de usuário desempenham um papel fundamental, pois elas oferecem uma forma flexível de capturar e expressar requisitos.

Esta prática facilita a comunicação entre desenvolvedores, clientes e outros stakeholders, ajudando a priorizar e entregar recursos em ciclos de desenvolvimento, promovendo colaboração contínua (RAHMAN et al., 2024; RONANKI; CABRERO-DANIEL; BERGER, 2024; OSWAL; KANAKIA; SUKTEL, 2024).

Como as histórias de usuário são essenciais para orientar o desenvolvimento e atingir os resultados desejados, é crucial que elas sejam bem elaboradas para evitar impactos negativos no cliente, projeto e equipe (OSWAL; KANAKIA; SUKTEL, 2024). Entretanto, as histórias de usuário geralmente são mal escritas na prática e exibem defeitos de qualidade inerentes (LUCASSEN et al., 2016a). A qualidade das histórias de usuário é crucial, pois impacta diretamente no design do sistema e no produto final. No entanto, a criação manual dessas histórias, embora permita maior controle e refinamento, é propensa a inconsistências e potencialmente mais demorada (RAHMAN et al., 2024; RONANKI; CABRERO-DANIEL; BERGER, 2024).

Com os avanços em Inteligência Artificial (IA), do inglês *Artificial Intelligence*, particularmente em Modelos de Linguagem de Grande Porte (LLMs), do inglês *Large Language Models*, como por exemplo, o ChatGPT da OpenAI ¹, têm demonstrado um potencial significativo para automatizar diversas tarefas que anteriormente, exigiam uma intervenção humana intensiva (BELZNER; GABOR; WIRSING, 2023). A automação reduz o esforço humano e o tempo gasto em tarefas repetitivas, minimiza erros e aumenta a eficiência, permitindo que os desenvolvedores se concentrem em atividades de maior valor agregado (YARLAGADDA, 2021). Os modelos oferecem novas oportunidades para apoiar processos de engenharia de software, atuando de forma conversacional com os desenvolvedores. Capacitados para interpretar prompts e gerar texto, esses sistemas aprimoram a engenharia de software ao fornecer conhecimento especializado e auxiliar em todo o ciclo de vida do desenvolvimento, desde a engenharia de requisitos até o design do sistema. Dessa forma, tornam-se candidatos ideais para apoiar a captura e o refinamento de requisitos de software (BELZNER; GABOR; WIRSING, 2023).

Como resultado, os LLMs se tornaram ferramentas úteis para capturar e refinar requisitos de software, agilizando e enriquecendo o processo de desenvolvimento (BELZNER; GABOR; WIRSING, 2023). Com foco em abordar as ineficiências ineren-

¹ <https://chatgpt.com/>

tes à criação manual de histórias de usuário, vários estudos estão sendo conduzidos sobre a geração automatizada dessas histórias por meio de LLMs, como uma solução para acelerar o processo e melhorar a qualidade (OSWAL; KANAKIA; SUKTEL, 2024; RAHMAN et al., 2024).

A integração de LLMs na engenharia de requisitos representa uma transformação relevante no desenvolvimento ágil de software, oferecendo a promessa de automatizar atividades e, assim, alcançar maior produtividade e qualidade nos produtos finais. Conforme apontado por Nguyen-Duc et al. (2025), tais tecnologias já vêm sendo exploradas para dar suporte a tarefas técnicas e gerenciais, abrangendo desde a automação parcial de atividades até o apoio à tomada de decisão no ciclo de vida de desenvolvimento de software. No entanto, é essencial investigar não apenas os benefícios, mas também as dificuldades e limitações, para que essas abordagens automatizadas possam ser implementadas de maneira eficaz. A aplicação prática dessas tecnologias levanta questões sobre sua eficácia, adaptabilidade e o equilíbrio entre automação e supervisão humana na gestão ágil de projetos.

1.2 Questão de Pesquisa

O principal desafio abordado neste estudo é o desenvolvimento de uma abordagem automatizada, capaz de gerar histórias de usuários de maior qualidade, alinhadas a critérios de qualidade reconhecidos. Ao enfrentar esse problema, esta pesquisa busca fornecer evidências sobre o potencial da inteligência artificial em complementar, ou até mesmo superar, os esforços humanos na elaboração de histórias de usuários. Assim, motivada pelo contexto anteriormente destacado, esta pesquisa visa responder a seguinte questão de pesquisa:

QP: Como apoiar a geração de histórias de usuários de maior qualidade no contexto do desenvolvimento ágil de software?

Para alcançar o objetivo da pesquisa, inicialmente foram analisados conhecimentos consolidados na literatura e investigado o contexto do problema em cenários

práticos. Em seguida, exploraram-se técnicas que serviram de base para a construção do artefato US-Prompt. Por fim, a abordagem desenvolvida foi refinada e avaliada quanto à sua eficácia e aplicabilidade.

Esta pesquisa está vinculada ao projeto aprovado pelo Comitê de Ética em Pesquisa (CEP) da Universidade Federal do Amazonas, conforme parecer de aprovação (CAAE [829552724.0.0000.5020]). Nas seções a seguir, são apresentados os objetivos de pesquisa e a metodologia adotada para a condução deste estudo.

1.3 Objetivos da Pesquisa

Esta pesquisa tem como principal objetivo: **Apoiar engenheiros de requisitos, product owners ou outros profissionais responsáveis pela especificação de requisitos na geração automatizada de histórias de usuário de maior qualidade.** Para alcançar esse objetivo, estabelecemos os seguintes objetivos específicos:

- Identificar ferramentas e técnicas de automatização aplicadas à especificação de requisitos, em metodologias ágeis;
- Identificar práticas e percepções de profissionais de software no uso de LLMs para a geração automatizada de histórias de usuário, a fim de compreender o estágio atual dessa abordagem.
- Desenvolver e validar iterativamente um artefato especializado baseado em técnicas de engenharia de prompt, visando orientar a geração de histórias de usuários de maior qualidade.
- Identificar os desafios e limitações na implementação de soluções automatizadas para a especificação de requisitos, em times ágeis.

1.4 Metodologia de Pesquisa

A fim de alcançar os objetivos estabelecidos nessa pesquisa, utilizou-se o Design Science Research (DSR) como abordagem metodológica, pois se concentra na criação e avaliação de artefatos, como sistemas, modelos, métodos ou teorias, para resolver problemas complexos no mundo real ([DRESCH et al., 2015](#); [HEVNER et al., 2004a](#)). O DSR é um paradigma de pesquisa que visa ajudar a compreender problemas e melhorar o estado da prática, gerando produtos inovadores ([HEVNER et al., 2004b](#)). Para isso, estudam-se instâncias do problema, envolvendo o contexto do problema, conceitualização, design da solução e validação. [Hevner et al. \(2004b\)](#) propõem os princípios Relevância, Design e Rigor, considerados norteadores para os ciclos que orientam a condução de pesquisas em DSR.

Os três ciclos fundamentais, Relevância, Rigor e Design, asseguram que a investigação aborde problemas pertinentes, ofereça contribuições inovadoras e seja conduzida com rigor metodológico, contribuindo de forma significativa para o campo de estudo. As ações desenvolvidas em cada ciclo podem ser resumidas da seguinte forma:

- **Ciclo de Relevância:** Este ciclo assegura que a pesquisa aborde problemas relevantes e significativos para o contexto organizacional. As ações incluem a identificação e compreensão do problema no ambiente real. Para isso, foram realizados estudos exploratórios com o objetivo de entender o ambiente e o contexto do problema, investigando o design e a avaliação de soluções inovadoras que promovam a automatização na engenharia de requisitos. Nesta etapa, define-se o domínio de aplicação com base nos stakeholders envolvidos, cujas demandas servem de insumo para o design do artefato.
- **Ciclo de Rigor:** O ciclo de rigor estabelece a base de conhecimento que oferece a fundamentação teórica e metodológica necessária para garantir o rigor científico na definição, planejamento, construção e avaliação dos artefatos desenvolvidos durante o ciclo de design ([HEVNER, 2007](#)). Esse ciclo sustenta os demais ciclos na base de conhecimento existente e, por meio das atividades de pesquisa, determina a inclusão de novos conhecimentos a essa base, conforme destacado por ([THUAN;](#)

[DRECHSLER; ANTUNES, 2019](#)).

- **Ciclo de Design:** Este ciclo assegura que a pesquisa ofereça contribuições inovadoras e originais ao campo. Ele consiste em integrar as informações obtidas nos ciclos de Relevância e Rigor para construir a solução proposta. Nessa etapa, diferentes versões dos métodos serão desenvolvidas e refinadas com base nos resultados obtidos nos estudos exploratórios. Além disso, serão conduzidos estudos experimentais para validar a eficiência e a qualidade das técnicas propostas para verificar sua aplicação em contextos reais.

1.5 DSR Aplicado nessa Pesquisa

Inicialmente, foram analisados conhecimentos consolidados na literatura e, em seguida, conduzida uma pesquisa com profissionais da indústria para entender como esses profissionais lidam com requisitos em seu dia a dia e suas percepções sobre o uso de LLMs. Em um segundo momento, exploraram-se técnicas que fundamentaram a construção do artefato US-Prompt. Por fim, a abordagem desenvolvida foi refinada e avaliada quanto à sua eficácia e aplicabilidade prática. Nas seções seguintes, são apresentados os objetivos da pesquisa e a metodologia utilizada para conduzir este estudo.

A definição do domínio de aplicação desta pesquisa ocorreu mediante consultas à literatura, realizadas para delimitar o problema a ser tratado (Seção 1.1), os objetivos a serem alcançados (Seção 1.2) e os critérios de aceite da solução proposta para o problema identificado. A Figura 1, adaptada de [Hevner et al. \(2004b\)](#), ilustra as etapas executadas na condução desta pesquisa.

Para garantir a robustez e a validade do método US-Prompt, a pesquisa foi estruturada em estudos experimentais alinhados aos princípios fundamentais da Design Science Research (DSR), assegurando uma condução metodológica coerente, sistemática e ancorada nos ciclos. A seguir, descreve-se como esses ciclos se conectam aos estudos realizados.

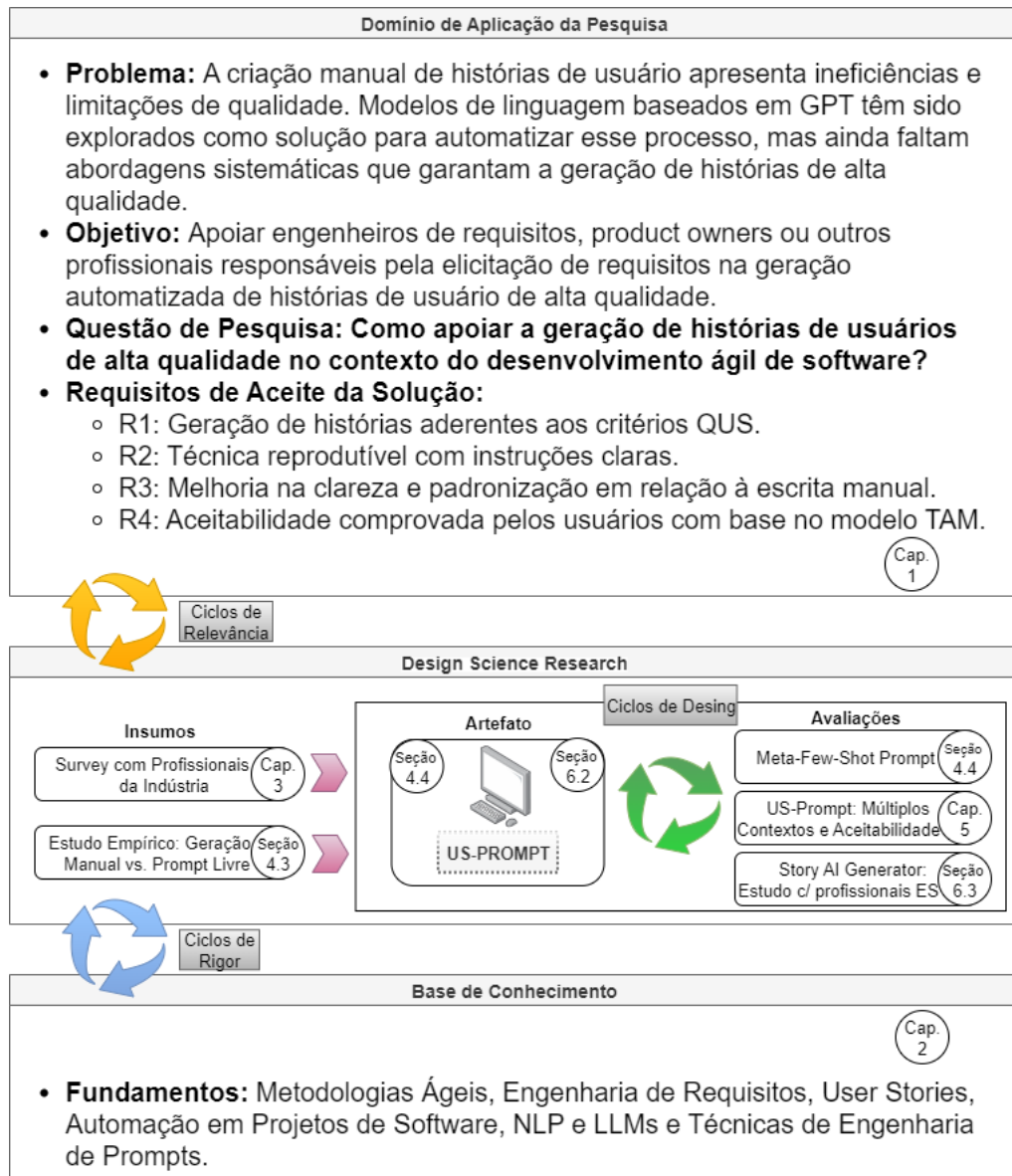


Figura 1 – Metodologia de desenvolvimento da pesquisa.

1.5.1 Ciclo de Relevância

Neste ciclo foram conduzidos dois estudos exploratórios que visaram compreender detalhadamente o contexto e o ambiente em que o problema está inserido, identificando sua relevância e impacto prático. Nesse sentido, a pesquisa buscou investigar fenômenos de interesse relacionados à especificação de requisitos em equipes ágeis, com o objetivo de identificar a viabilidade de um artefato que utiliza LLMs para automatizar atividades de geração nessa área, solucionando um desafio específico e prático da engenharia de software.

O primeiro estudo teve como objetivo compreender como profissionais da indústria de software lidam com atividades de elicitação e especificação de requisitos, além de investigar o uso atual de LLMs no apoio a essas tarefas. Os dados coletados permitiram não apenas confirmar a relevância prática do tema abordado, mas também fornecer subsídios para fundamentar teoricamente os requisitos do artefato, contribuindo para o delineamento do espaço do problema e reforçando sua importância no contexto organizacional. Este estudo está descrito no Capítulo 3 e está diretamente vinculado aos ciclos de Relevância e Rigor.

O segundo estudo foi concebido em uma fase exploratória inicial com o objetivo de investigar se histórias de usuário geradas manualmente por humanos apresentavam maior qualidade do que aquelas produzidas automaticamente via prompting livre com o ChatGPT. A comparação entre essas abordagens evidenciou lacunas importantes em termos de padronização, clareza e eficiência, além de indicar a necessidade de estratégias mais estruturadas de prompting. Os resultados obtidos forneceram embasamento para orientar o delineamento do artefato proposto nos ciclos de design subsequentes, caracterizando-se, assim, como uma fase preparatória essencial para o processo de design. Este estudo está descrito na Seção 4.3.

1.5.2 Ciclo de Design

O ciclo de design compreendeu a criação e avaliação do artefato concebido para atender às necessidades identificadas durante o ciclo de rigor e relevância. Com base nos estudos realizados sobre métodos existentes e no entendimento do ambiente do problema, foi desenvolvido um artefato baseado em técnicas de prompting, com foco em promover a automatização da geração de histórias durante a especificação de requisitos na engenharia de software ágil.

Este ciclo é caracterizado por um processo iterativo, envolvendo refinamento e reavaliação constantes do artefato para garantir sua eficácia e adequação. Além disso, foram definidas estratégias de medição de desempenho para mensurar a eficácia dos métodos propostos, assegurando a validade e aplicabilidade da solução.

Com base nos achados dos Ciclos de Relevância e Rigor, foi realizado o **Ciclo 1 de Design**, responsável pela proposição, construção e avaliação inicial do **US-Prompt**, apresentado na Seção 1.1 e detalhado na Seção 4.4, com o objetivo de apoiar engenheiros de requisitos e product owners na geração automatizada de histórias de usuário com qualidade, alinhadas a critérios sintáticos, semânticos e pragmáticos.

O próximo estudo, descrito no Capítulo 5, teve como base os aprendizados obtidos no Ciclo 1 de Design e consistiu no **Ciclo 2 de Design**. Com foco no refinamento do artefato e na validação externa, essa etapa teve como objetivos avaliar a eficácia do US-Prompt em diferentes contextos de geração de histórias de usuários e mensurar sua aceitabilidade com base no modelo TAM (Technology Acceptance Model). A aplicação de métricas consolidadas, como utilidade percebida, facilidade de uso percebida e intenção de uso futuro, permitiu uma análise abrangente, tanto do desempenho técnico do artefato quanto da percepção por parte dos usuários. Os resultados reforçaram a viabilidade do US-Prompt e indicaram melhorias para automatizar a técnica e torná-la mais prática.

A partir dos resultados do Ciclo 2 de Design, foi conduzido o **Ciclo 3 de Design**, descrito no Capítulo 6, dedicado à concepção e implementação da ferramenta web Story IA Generator, que operacionaliza o US-Prompt em um ambiente interativo para apoio à geração e refinamento automatizado de histórias de usuário. Por fim, a ferramenta também foi avaliada por profissionais de engenharia de software, fornecendo evidências experimentais de sua aplicabilidade prática e reforçando sua relevância para a engenharia de requisitos apoiada por LLMs.

1.5.3 Ciclo de Rigor

O Ciclo de Rigor desta pesquisa tem como objetivo utilizar e gerar conhecimento relevante para a Engenharia de Requisitos e a criação de histórias de usuário. Conforme [Thuan, Drechsler e Antunes \(2019\)](#), esse ciclo fundamenta os demais ciclos por meio da base de conhecimento existente e possibilita a incorporação de novos conhecimentos a partir das atividades de pesquisa.

A base de conhecimento utilizada inclui teorias e práticas de Engenharia de Requisitos, frameworks e métodos de geração e avaliação de histórias de usuário, além de diretrizes de qualidade como o framework *Quality User Story* (QUS).

A principal contribuição desta pesquisa é a abordagem de prompt desenvolvida, o US-Prompt, que aplica técnicas de Meta-Prompting e Few-Shot Prompting para gerar histórias de usuário de maior qualidade. Os estudos experimentais conduzidos como survey com profissionais e estudos comparando geração manual, prompt livre e prompts estruturados, fornecem dados e insights relevantes para futuras pesquisas e práticas na área.

Adicionalmente, a pesquisa contribui com: (i) Conhecimento sobre frameworks e critérios de qualidade de histórias de usuário; (ii) Identificação das necessidades informacionais de diferentes perfis de usuários; (iii) Estratégias de prompting que aumentam a qualidade da geração automatizada; (iv) Aspectos metodológicos para avaliação da qualidade de histórias geradas por LLMs.

A base de conhecimento foi continuamente atualizada ao longo dos ciclos de design e avaliação, garantindo fundamentação teórica robusta para a solução proposta. O Capítulo 2 detalha a base de conhecimento gerada nesta pesquisa.

1.5.4 Requisitos de Aceite da Solução

A fim de validar a efetividade da técnica de prompting proposta (US-Prompt), foram definidos os seguintes requisitos de aceite:

1- A técnica deve permitir a geração de histórias de usuário aderentes aos critérios de qualidade estabelecidos no framework *Quality User Story* (QUS), incluindo aspectos sintáticos, semânticos e pragmáticos.

2- A técnica deve ser formalizada por meio de instruções claras, exemplos e estrutura de prompt reproduzível, possibilitando sua aplicação por diferentes usuários e em diferentes contextos.

3- A técnica deve produzir resultados com maior padronização e clareza do que a geração manual, especialmente por usuários com menor experiência em requisitos.

4- A técnica deve alcançar níveis satisfatórios de aceitabilidade por parte dos usuários, conforme avaliado por meio do modelo TAM3 (*Technology Acceptance Model*), considerando os construtos de utilidade percebida, facilidade de uso e intenção de uso futuro.

1.6 Organização do Trabalho

Este trabalho está organizado em mais cinco capítulos, além deste primeiro capítulo de introdução que apresentou a contextualização, os objetivos da pesquisa e a metodologia. A organização do texto deste trabalho segue a estrutura a seguir:

- **Capítulo 2:** Contém o referencial teórico, com uma visão geral dos conceitos relevantes e dos trabalhos relacionados à área.
- **Capítulo 3:** Apresenta o estudo de pesquisa de opinião com profissionais da indústria, com o objetivo de compreender o ambiente e o contexto do problema.
- **Capítulo 4:** Relata o planejamento, a execução e os resultados de dois estudos exploratórios que comparam a qualidade das histórias geradas manualmente com aquelas produzidas por LLMs. Também propõe a abordagem Meta-Few-Shot Prompt, componente resultante do primeiro ciclo de design da metodologia DSR utilizada.
- **Capítulo 5:** Apresenta o segundo ciclo de design, incluindo o aprimoramento e avaliação do US-Prompt com base no julgamento dos participantes, destacando as melhorias e ajustes realizados ao longo do processo.
- **Capítulo 6:** Apresenta o terceiro ciclo de design, no qual é concebida e implementada a ferramenta web Story IA Generator, baseada na técnica US-Prompt e também descreve o estudo exploratório conduzido com profissionais de engenharia de software para avaliar a ferramenta proposta, investigando sua aceitação e potencial de adoção em contextos reais.

- **Capítulo 7:** Apresenta as considerações finais e as principais contribuições da pesquisa.

2

CICLO DE RELEVÂNCIA: FUNDAMENTOS

Esta seção apresenta a fundamentação teórica utilizada como base para essa proposta. Em especial, com foco na utilização de LLMs para a automação das atividades de especificação de requisitos em times ágeis.

2.1 Metodologias Ágeis

Desde o surgimento do Manifesto Ágil, a metodologia ágil tem sido amplamente estudada e adotada como uma alternativa mais flexível e responsiva ao tradicional modelo em cascata. Embora o desenvolvimento em cascata ainda tenha seu espaço, o uso de métodos ágeis tem mostrado maior efetividade no progresso e no sucesso de projetos de software, principalmente devido à sua capacidade de adaptação, foco em software funcional e colaboração contínua com o cliente ([MISHRA; ALZOUBI, 2023](#); [GHEORGHE; GHEORGHE; IATAN, 2020](#)).

Estudos anteriores demonstram que a adoção de práticas ágeis, como ciclos curtos de entrega, envolvimento frequente dos stakeholders e comunicação informal, contribui significativamente para a melhoria na qualidade dos produtos, redução de custos e prazos, e maior satisfação dos clientes. Diferente do modelo em cascata, que é mais rígido e estruturado, o ágil permite ajustes contínuos ao longo do desenvolvimento, o que é essencial em contextos de constantes mudanças ([MISHRA; ALZOUBI, 2023](#); [COSTA; NUNES, 2023](#)).

No entanto, a abordagem ágil também apresenta desafios. Questões como a

descoberta tardia de falhas críticas, especialmente quando o foco está apenas em novas funcionalidades, e a menor ênfase na documentação podem comprometer a garantia da qualidade se não forem bem gerenciadas ([MISHRA; ALZOUBI, 2023](#)). Dentre os principais frameworks ágeis, destacam-se Scrum, XP e Kanban. O Scrum organiza o trabalho em sprints curtos com entregas regulares e envolvimento ativo do cliente. O XP enfatiza a entrega rápida e práticas como o desenvolvimento orientado a testes. Já o Kanban utiliza quadros visuais para acompanhar o fluxo das tarefas e melhorar a transparência do processo ([TRIVEDI, 2021](#)).

A metodologia ágil surgiu como resposta à necessidade de acelerar o desenvolvimento sem comprometer a qualidade e a adaptabilidade. Equipes ágeis geralmente são compostas por desenvolvedores, testadores, engenheiros de dados, um Scrum Master e um Product Owner, sendo este último, o elo entre a equipe e o cliente, responsável por garantir que os requisitos sejam compreendidos e atendidos. O sucesso de um projeto ágil depende diretamente da comunicação clara, da colaboração contínua e da capacidade de adaptação da equipe ([MISHRA; ALZOUBI, 2023](#); [GHEORGHE; GHEORGHE; IATAN, 2020](#)).

O método ágil remete à habilidade de se adaptar e responder à mudança, e é justamente essa flexibilidade, rapidez e foco na melhoria contínua que fazem da metodologia ágil uma abordagem tão eficaz e valorizada, tornando essas metodologias nas práticas padrão para as indústrias modernas ([GHEORGHE; GHEORGHE; IATAN, 2020](#)). O feedback das iterações contínuas é usado para refinar o produto até sua integração final, promovendo criatividade e liberdade nos processos. Essa abordagem também permite a rápida re-priorização do backlog, e assim agilizando o ciclo de desenvolvimento. Ao dividir o projeto em iterações gerenciáveis, os métodos ágeis melhoram a qualidade geral por meio de desenvolvimento e colaboração aprimorados e elaborados e testes frequentes. A comunicação eficaz entre stakeholders e equipes é essencial, pois o feedback contínuo dos clientes levanda a melhorias constantes e a um produto final mais detalhado ([TRIVEDI, 2021](#)).

As metodologias ágeis apresentam diversos desafios que podem comprometer a qualidade do produto final. Segundo [Martins et al. \(2019\)](#), um estudo anterior iden-

tificou sete principais dificuldades nesse contexto: (1) problemas com estimativas de custo e cronograma; (2) arquitetura inadequada ou mal planejada; (3) negligência de requisitos não funcionais; (4) acesso limitado e baixa participação dos stakeholders; (5) priorização baseada em uma única dimensão, o que dificulta a escalabilidade; (6) verificação insuficiente dos requisitos; e (7) documentação mínima dos artefatos de requisitos. Dentre esses, três desafios estão diretamente relacionados à falta de envolvimento efetivo das partes interessadas na elicitação de requisitos: as dificuldades em estimar esforço e tempo (item 1), a negligência aos requisitos não funcionais (item 3) e o acesso restrito aos stakeholders (item 4).

2.2 Engenharia de Requisitos

A Engenharia de Requisitos (ER) é considerada uma das etapas mais complexas do desenvolvimento de software, pois tem como responsabilidade principal decidir e definir o que precisa ser desenvolvido ([PACHECO; GARCÍA; REYES, 2018](#)). Para isso, o processo de ER busca identificar, modelar, comunicar e documentar os requisitos do sistema, além de estabelecer seu contexto de uso ([ALHAZMI; HUANG, 2020](#)). Esses requisitos representam as funcionalidades, restrições ou qualidades que o sistema deve possuir para atender às necessidades das partes interessadas ([GENTILI; FALESSI, 2025](#)). Assim, a ER surgiu no âmbito da Engenharia de Software com o propósito de enfrentar os desafios envolvidos na definição precisa desses requisitos. De modo geral, esse processo abrange um conjunto de atividades fundamentais, como elicitação, análise, especificação, validação, verificação e gerenciamento ([PACHECO; GARCÍA; REYES, 2018](#)).

A (ER) tem um papel crucial nas metodologias tradicionais, onde os requisitos do sistema são identificados e delineados antes do início do seu desenvolvimento, acarretando em diversos desafios que a abordagem ágil de ER busca superar ([ALHAZMI; HUANG, 2020](#)).

- Engenharia de Requisitos em Métodos Tradicionais: Os métodos tradicionais seguem fases pré-organizadas do ciclo de vida de desenvolvimento de software,

como modelo cascata e modelo V. A ER nesse contexto tem como principal objetivo definir o que construir antes de iniciar o desenvolvimento do sistema, porém existem desafios a serem enfrentados, como lacunas de comunicação e excesso de escopo. O processo de ER é dividido em cinco atividades: elicitacão, análise, documentação, validação e gerenciamento.

- **Engenharia de Requisitos em Métodos Ágeis:** Os métodos ágeis priorizam a interação sobre processos, o software funcional sobre documentação extensa e a colaboração do cliente sobre negociação de contrato. A ER apresenta um aspecto mais informal e interativo, ocorrendo ao longo do desenvolvimento. Os requisitos são definidos de forma incremental, com base em artefatos como épicos e histórias de usuário. Essa abordagem traz flexibilidade e adaptabilidade ao desenvolvimento de software ([ALHAZMI; HUANG, 2020](#)).

Neste último, é utilizado um modelo just-in-time para refinar requisitos de alto nível em tarefas de baixo nível implementáveis pelos desenvolvedores, e algumas técnicas comuns de elicitacão em ambientes ágeis incluem entrevistas, brainstorming, observação e análise de caso ([MARTINS et al., 2019](#)). De modo que, os métodos ágeis se adaptam bem às equipes de desenvolvimento de software nas atividades de elicitacão de requisitos ([MARTINS et al., 2019](#)).

[Kotonya e Sommerville \(1998\)](#) propuseram um modelo conceitual linear para o processo de Engenharia de Requisitos (RE), que reconhece interações iterativas entre as atividades. Eles destacam que, embora o modelo seja apresentado de forma linear, suas atividades frequentemente se sobrepõem e podem ser executadas de forma iterativa ao longo do ciclo de vida do desenvolvimento ([KOTONYA; SOMMERVILLE, 1998](#)). As etapas do processo são:

1. **Estudo de Viabilidade:** Avaliar se o sistema proposto é útil e viável para o negócio.
2. **Levantamento e Análise de Requisitos:** Descobrir e compreender as necessidades e expectativas dos usuários e stakeholders.
3. **Especificação dos Requisitos:** Documentar os requisitos de forma estruturada e padronizada.

4. **Validação dos Requisitos:** Verificar se os requisitos definidos realmente representam o sistema desejado pelo cliente.

Como os stakeholders frequentemente apresentam visões distintas, é essencial adotar uma linguagem comum que possibilite a comunicação e o compartilhamento eficaz de informações entre as partes envolvidas. Por essa razão, a linguagem natural é amplamente utilizada na descrição de requisitos. No entanto, quando utilizada de forma não estruturada, a linguagem natural pode introduzir defeitos recorrentes nas especificações, como ambiguidade, incompletude, inconsistência e incorreção, os quais comprometem o andamento do projeto e podem resultar em retrabalho, atrasos e insatisfação dos usuários (GENTILI; FALESSI, 2025).

2.3 User Stories

Em projetos ágeis de software, os requisitos são frequentemente expressos como histórias de usuários, que são descrições sucintas de funcionalidades do ponto de vista do usuário. Essas breves narrativas capturam os elementos essenciais de um requisito: quem é o usuário, o que se espera do sistema e por que isso é importante (ZHANG et al., 2024). O uso de templates para redigir histórias de usuários é amplamente difundido entre profissionais da área, sendo que no estudo de Lucassen et al. (2016a) apenas uma minoria (15%) declarou não adotar nenhum modelo específico. O formato mais recorrente é o originalmente proposto pela empresa Connextra e amplamente difundido por (COHN, 2004), cuja estrutura segue o padrão (LUCASSEN et al., 2016a):

**“Como um<tipo de usuário>,
eu quero<objetivo>,
para que<algum motivo>.”**

Nesse contexto, o trabalho de Wautelet et al. (2014) buscou unificar e estender os modelos existentes, propondo uma estrutura conceitual consolidada. Com base nesse formato, Lucassen et al. (2016b) define que uma história de usuário deve conter quatro elementos fundamentais:

**"Papel (quem),
meio (o que),
um ou mais fins (por que)."**

Além de um **formato** padronizado. A seguir, cada elemento é detalhado:

- **Papel:** Especifica a parte interessada ou persona que expressa a necessidade, por exemplo, "gerente de contas", "comprador" ou "representante de vendas". Alternativamente, pode-se usar *personas*, personagens fictícios que representam grupos de usuários com características semelhantes.
- **Meios:** Descrevem a ação requerida e, geralmente, seguem três componentes gramaticais:
 - Sujeito + verbo de intenção, como "quero" ou "sou capaz de";
 - Verbo de ação, que indica a operação desejada (por exemplo, "abrir");
 - Objeto direto, sobre o qual a ação se aplica (por exemplo, "o mapa interativo").

Além disso, os meios podem incluir modificadores opcionais, como adjetivos ("visualização maior") ou objetos indiretos ("a partir da página de perfil do usuário").

- **Fins:** Justificam o valor da história, respondendo "por que" aquele meio é necessário. Lucassen2016a identifica três variantes não excludentes de fins bem formulados:
 - Esclarecimento do próprio meio, detalhando melhor a ação;
 - Dependência de outra funcionalidade, ligando a história a um contexto maior;
 - Requisito de qualidade, especificando características não-funcionais exigidas.

A redação eficaz de histórias de usuários é fundamental, pois elas comunicam as necessidades dos usuários e guiam a equipe de desenvolvimento. Cada história de usuário deve incluir os critérios de aceitação que definem as condições que devem ser atendidas para que a história seja considerada aceitável, abrangendo comportamento funcional e aspectos de qualidade (ZHANG et al., 2024).

Assim como diretrizes gerais de qualidade são amplamente utilizadas na engenharia de requisitos, diversas estruturas, como o INVEST (Independent–Negotiable–Valuable–Estimatable–Scalable–Testable)([WAKE; WAKE, 2003](#)), oferecem critérios específicos para avaliar a qualidade das histórias de usuário ([ZHANG et al., 2024](#)). A adoção dessas diretrizes é uma prática comum entre profissionais que lidam com histórias de usuário, sendo o INVEST o conjunto de critérios mais amplamente utilizado, segundo [Lucassen et al. \(2016a\)](#).

O artigo de [Wake e Wake \(2003\)](#) introduz o acrônimo INVEST como uma abordagem para criar histórias eficazes em ambientes ágeis. Ao adotar esses princípios, é possível aprimorar a qualidade das histórias e aumentar a eficiência do desenvolvimento ágil, melhorando a comunicação entre a equipe e as partes interessadas. Ele representa seis características essenciais de boas histórias de usuário, conforme ([WAKE; WAKE, 2003](#)):

- **I – Independente:** Histórias devem ser independentes entre si para facilitar o agendamento e a implementação em qualquer ordem.
- **N – Negociável:** Histórias não devem ser contratos rígidos; os detalhes são definidos em colaboração entre cliente e desenvolvedores ao longo do tempo.
- **V – Valiosa:** Cada história precisa gerar valor real para o cliente, sendo possível fatiar histórias complexas em partes menores que ainda entreguem valor.
- **E – Estimável:** É necessário que a equipe consiga estimar o esforço necessário para implementar a história, o que depende da clareza, tamanho e experiência da equipe.
- **S – Small (Pequena):** Histórias menores são mais fáceis de entender, estimar e implementar, normalmente representando poucos dias ou semanas de trabalho.
- **T – Testável:** A história deve permitir a definição de critérios de aceitação claros, possibilitando a escrita de testes para validar sua implementação.

Outra estrutura é apresentada por [Lucassen et al. \(2016b\)](#) o framework QUS, uma coleção de 13 critérios que focam na qualidade intrínseca do texto da história do usuário

e avaliam a qualidade das histórias de usuários em termos sintáticos, pragmáticos e semânticos:

- **Qualidade sintática:** Refere-se à estrutura textual de uma história de usuário, sem considerar seu significado.
- **Qualidade semântica:** Diz respeito às relações e ao significado de (partes de) texto da história de usuário.
- **Qualidade pragmática:** Considera a interpretação subjetiva do público sobre o texto da história de usuário, além da sintaxe e da semântica.

Os critérios são: Bem-formada, Atômico, Mínima, Conceitualmente Sólida, Orientada a Problemas, Não ambígua, Livre de Conflitos, Frase Completa, Estimável, Única, Uniforme, Independente e Completo ([LUCASSEN et al., 2016b](#)).

- **Bem-formada:** A descrição principal da história de usuário deve incluir um papel (quem realiza a ação) e um meio (a funcionalidade desejada). Exemplo: *“Como membro, quero ver uma mensagem de erro quando não consigo visualizar as recomendações após enviar um artigo.”*
- **Atômica:** A história deve abordar uma única funcionalidade. Combinar múltiplas funcionalidades em uma única história compromete a precisão das estimativas de esforço. Exemplo: *“Como usuário, posso clicar em um local específico do mapa e, assim, realizar uma pesquisa de pontos de referência associados a essa combinação de latitude e longitude.”* Essa história deve ser dividida em duas: (1) *“Como usuário, posso clicar em um local específico no mapa”,* e (2) *“Como usuário, posso ver pontos de referência associados à combinação de latitude e longitude de um local específico.”*
- **Mínima:** A história deve conter apenas o papel, o meio e (opcionalmente) um ou mais fins. Informações adicionais como detalhes de comportamento, comentários ou instruções de teste devem ser colocadas em campos auxiliares. Exemplo: *“Como profissional de saúde, quero ver as horas registradas desta semana.”*
- **Conceitualmente sólida:** O meio e os fins desempenham papéis distintos. O meio descreve uma funcionalidade concreta, enquanto o fim expressa a justificativa

para sua existência. Exemplo: *“Como usuário, desejo abrir o mapa interativo para poder ver a localização dos pontos de referência.”* O fim deve ser específico, e não apenas uma referência genérica a outra funcionalidade.

- **Orientada a problemas:** A história deve descrever o problema ou necessidade do usuário, e não a solução técnica. Caso seja necessário incluir sugestões de implementação, elas devem ser indicadas apenas em comentários ou campos auxiliares.
- **Não ambígua:** A história deve ser clara e precisa, sem margem para múltiplas interpretações — tanto de forma isolada quanto em relação a outras histórias do mesmo sistema.
- **Frase completa:** A história deve ser redigida como uma sentença gramaticalmente correta, sem erros de ortografia ou estrutura textual.
- **Estimável:** A história deve ter um escopo suficientemente claro e restrito para permitir a estimativa de esforço. Exemplo de história difícil de estimar: *“Como profissional de saúde, quero ver minha lista de rotas para os próximos dias, para que eu possa me preparar (por exemplo, ver a que horas devo começar a viajar).”*
- **Única e livre de conflitos:** As histórias devem ser semanticamente únicas, evitando duplicações exatas ou semelhantes. Além disso, não devem conter conflitos entre si, que causem inconsistências nos requisitos. Relações entre histórias, como meios diferentes para um mesmo fim ou mesmos meios para fins distintos, precisam ser analisadas para identificar variações, conflitos ou possibilidades de combinação.
- **Uniformidade:** As histórias devem seguir um formato consistente, alinhado ao padrão adotado pela equipe, garantindo coesão na documentação.
- **Independente:** As histórias devem ser conceitualmente independentes, possibilitando que sejam planejadas e implementadas em qualquer ordem. Embora dependências sejam inevitáveis, elas devem ser explicitamente registradas para evitar confusão.

- **Completeness:** O conjunto de histórias deve cobrir o escopo necessário para que o sistema seja funcional. Histórias que dependem de objetos ou funcionalidades inexistentes indicam lacunas que precisam ser preenchidas para garantir a entrega completa do produto.

Os dados da pesquisa de [Lucassen et al. \(2016a\)](#) mostram que a aplicação de um template e diretrizes de qualidade nas histórias de usuários é capaz de aumentar a produtividade e contribuir para a melhoria da qualidade do trabalho final entregue ([LUCASSEN et al., 2016a](#)).

Essas abordagens asseguram que as histórias de usuários sejam concisas e claras, contribuindo para o sucesso dos projetos de desenvolvimento de software e melhorando a experiência do usuário ([ZHANG et al., 2024](#)).

2.4 NLP e LLMs

O Processamento de Linguagem Natural (NLP) é uma técnica analítica que visa analisar e compreender automaticamente a linguagem humana. Com avanços significativos nos últimos anos, especialmente com o surgimento da Inteligência Artificial (IA), a NLP tem sido aplicada em diversas áreas, como tradução automática, correspondência de padrões, análise de sentimento e reconhecimento de fala. Esses avanços não só impactaram a vida cotidiana das pessoas, mas também revolucionaram as práticas empresariais, permitindo a extração eficiente de insights valiosos de conjuntos de dados textuais ([KANG et al., 2020](#)).

No artigo de [Vaswani et al. \(2017\)](#), é apresentada a arquitetura Transformer, uma nova arquitetura de rede neural baseada inteiramente em mecanismos de atenção. Esta arquitetura revolucionou o campo do Processamento de linguagem natural (NLP) ao utilizar mecanismos de atenção para processar e gerar texto, sem depender de convoluções ou recorrências.

A introdução dos Transformadores possibilitou o treinamento de modelos maiores e mais poderosos, capazes de realizar uma ampla gama de tarefas linguísticas com

alta precisão, servindo de base para o desenvolvimento de Large Language Models (LLMs) ([RADFORD et al., 2018](#); [BROWN et al., 2020](#)).

Esses modelos pertencem à classe de métodos fundamentados em aprendizado de máquina e obtêm desempenho expressivo devido à sua capacidade de aprender representações linguísticas eficazes a partir de grandes volumes de dados textuais. Esse desempenho está diretamente relacionado às tarefas de previsão de palavras subsequentes durante o treinamento, como no caso do BERT (Bidirectional Encoder Representations from Transformers), o que permite aos modelos capturar padrões complexos e nuances da linguagem natural ([NORHEIM et al., 2024](#); [DEVLIN et al., 2018](#); [RADFORD et al., 2018](#)). No entanto, esse processo envolve custos computacionais significativamente elevados. A evolução dos LLMs tem ocorrido de forma acelerada, refletindo avanços sucessivos na arquitetura e escala dos modelos ([NORHEIM et al., 2024](#)).

A OpenAI lançou, em 2018, o modelo Generative Pre-trained Transformer (GPT), um LLM pré-treinado voltado à geração de linguagem natural ([RADFORD et al., 2018](#)). Na sequência, foram disponibilizadas novas versões: o GPT-2 em 2019, o GPT-3 em 2020, o GPT-3.5 em 2022 e o GPT-4 em 2023. Cada nova versão representou um aumento expressivo no número de parâmetros do modelo, resultando em melhorias significativas de precisão, cobertura e capacidade geral ([NORHEIM et al., 2024](#)).

2.5 Tarefas de engenharia de requisitos e LLMs

Diversos estudos têm investigado o uso de Processamento de Linguagem Natural (NLP) na Engenharia de Requisitos (RE), sendo que revisões recentes, como as de [Zhao et al. \(2021\)](#) e [Sonbol, Rebdawi e Ghneim \(2022\)](#), consolidaram esses esforços ao proporem classificações das aplicações de NLP em tarefas de RE. Essas revisões identificaram uma tendência crescente de pesquisas voltadas à aplicação de NLP em RE, organizaram as abordagens conforme as atividades envolvidas e as técnicas utilizadas, além de popularizarem a sigla NLP4RE para designar essa área de estudo.

Com base nessas classificações, [Norheim et al. \(2024\)](#) realizaram uma síntese

e propuseram uma nova organização das tarefas de RE que podem se beneficiar do uso de modelos de linguagem de grande escala (LLMs), agrupando-as em categorias funcionais que destacam como os LLMs podem apoiar diferentes atividades no processo de engenharia de requisitos:

2.5.1 Geração:

A geração de requisitos envolve a criação de requisitos a partir de textos em linguagem natural, incluindo a formulação e elicitação auxiliadas por esses modelos ([HICKEY; DAVIS, 2004](#)). LLMs podem processar informações diversas, como necessidades de stakeholders, avaliações de usuários e especificações prévias, para gerar novos requisitos ou refinar os existentes, além de criar modelos compartilháveis entre produtos ([REINHARTZ-BERGER; KEMELMAN, 2020](#)). Essa abordagem é mais abrangente que simples extração de requisitos e também se relaciona com busca e recuperação de informações. Embora representem cerca de 16% das pesquisas em processamento de linguagem natural aplicado à engenharia de requisitos, as aplicações práticas ainda são limitadas ([SONBOL; REBDAWI; GHNEIM, 2022](#)).

Exemplos incluem o RE-BERT, que classifica avaliações de usuários para identificar funcionalidades existentes e derivar requisitos futuros ([ARAÚJO; MARCACINI, 2021](#)), e assistentes interativos conversacionais ajustados para guiar stakeholders ([RAS-TOGI et al., 2020](#)). Eles podem fortalecer técnicas tradicionais, como histórias de usuários ([LUCASSEN et al., 2016b](#)) e outras abordagens clássicas ([HICKEY; DAVIS, 2004](#)). [Arora, Grundy e Abdelrazek \(2024\)](#) demonstraram esse potencial ao empregar o GPT-4 com estratégias de prompting em cenário zero-shot.

2.5.2 Avaliação de Qualidade:

A avaliação da qualidade dos requisitos pode ocorrer em dois níveis: individual (ambiguidade e conformidade com padrões) e documental (detecção de inconsistências e lacunas). [Sonbol, Rebdawi e Ghneim \(2022\)](#) apontam que cerca de um sexto das pesqui-

sas em NLP4RE foca em problemas de qualidade. Trabalhos recentes, como os [Malik et al. \(2025\)](#), exploram LLMs para identificar similaridade semântica e detectar requisitos redundantes ou conflitantes. [Luitel, Hassani e Sabetzadeh \(2023\)](#) propuseram uma abordagem não supervisionada baseada em modelos de linguagem mascarada para identificar requisitos ausentes, utilizando o conjunto de dados PURE sem necessidade de anotações manuais.

2.5.3 Análise:

A análise é a categoria de tarefas de Engenharia de Requisitos que mais recebeu atenção no uso de LLMs até o momento, especialmente em classificação, rastreabilidade e extração de informação. NoRBERT de [Hey et al. \(2020\)](#) e modelos como o de [Kici et al. \(2021\)](#) demonstraram bons resultados em classificação de requisitos. Para NER, destacam-se SpaceBERT de [Berquand, Darm e Riccardi \(2021\)](#), e os trabalhos de [Ajagbe e Zhao \(2022\)](#) e [Chami, Zoghbi e Bruel \(2019\)](#). A extração de modelos UML a partir de requisitos também foi abordada com GPT-3 por [Bajaj et al. \(2022\)](#). Em rastreabilidade, o RDC-BERT de [Deshpande et al. \(2021\)](#) se destacou na detecção de vínculos entre requisitos, superando métodos tradicionais.

2.5.4 Tradução:

LLMs têm sido aplicados à tradução de requisitos de linguagem natural (NL) para linguagens formais, como LTL, RegEx e Lógica de Primeira Ordem. Embora ainda pouco explorada, essa tarefa tem ganhado atenção. [Hahn et al. \(2022\)](#) utilizaram o T5 para traduzir requisitos em NL para formas formais. [Cosler et al. \(2023\)](#) usaram GPT-3 com few-shot prompting para traduzir 36 especificações geradas por especialistas, alcançando sucesso em 31 casos. Embora tradicionalmente classificadas como modelagem, outras abordagens como as de [Deeptimahanti e Sanyal \(2011\)](#) e [Dawood et al. \(2017\)](#) também se enquadram na ideia de tradução ao transformar linguagem informal em representações formais ou semiformal, estudando a migração de requisitos de software

para diagramas de Linguagem de Modelagem Unificada (UML), nele foi realizado um survey a fim de conhecer o status atual do uso de ferramentas de gerenciamento de requisitos, o conhecimento do uso de UML no desenvolvimento de software, diagramas UML usados com frequência e a metodologia usada para gerar diagramas UML a partir de requisitos.

2.6 Técnicas de Engenharia de Prompts

Com o avanço dos LLMs, como o GPT da OpenAI, emergiram novas oportunidades para sua aplicação em diferentes domínios, incluindo a engenharia de software. No entanto, a obtenção de respostas relevantes e de alta qualidade requer o uso de estratégias específicas para a formulação de entradas adequadas.

Assim, a engenharia de prompts surgiu como uma disciplina fundamental na otimização do desempenho dos LLMs, estruturando entradas para aprimorar a coerência, a precisão e o alinhamento com os objetivos da tarefa. Essa abordagem consiste no desenvolvimento de instruções cuidadosamente elaboradas para orientar o comportamento dos modelos, permitindo que suas respostas atendam com maior precisão às demandas específicas (DEBNATH et al., 2025).

Diversas técnicas têm sido propostas no contexto de engenharia de prompts, como zero-shot, one-shot, few-shot e cadeia de pensamento (Chain-of-Thought), cada uma com aplicações e efetividades distintas conforme a tarefa.

Entre essas abordagens, destaca-se a técnica de Meta-Prompting, que consiste no uso de prompts cuidadosamente estruturados com instruções explícitas, exemplos e critérios que orientam o modelo na execução de uma tarefa. Essa técnica se baseia em um ciclo interno de prompts, no qual o modelo é guiado a seguir um conjunto de instruções organizadas de forma hierárquica, criando um fluxo de raciocínio iterativo. Ao incorporar instruções adicionais diretamente no prompt, o modelo consegue realizar ajustes contínuos em sua resposta, o que resulta em uma saída mais refinada e condizente com os requisitos definidos. Essa abordagem tem se mostrado eficaz, por exemplo, na geração de histórias de usuários mais precisas e alinhadas a frameworks

específicos, como o QUS ([BROWN et al., 2020](#); [ZHANG](#); [YUAN](#); [YAO, 2023](#)).

Outra técnica importante é a Few-Shot Prompting, que explora a capacidade dos LLMs de generalizar a partir de poucos exemplos. Ao fornecer apenas um conjunto reduzido de exemplos relevantes, o modelo é capaz de identificar padrões e produzir respostas adequadas, sem a necessidade de treinamento adicional ou grandes volumes de dados ([BROWN et al., 2020](#); [ZHANG](#); [YUAN](#); [YAO, 2023](#)).

Ambas as técnicas, Meta Prompting e Few-Shot Prompting, maximizam a capacidade de generalização dos modelos, permitindo que atuem com eficácia mesmo em contextos com requisitos variados. Esse potencial é particularmente valioso em ambientes ágeis, nos quais velocidade e precisão são essenciais para a entrega contínua de valor. Assim, essas abordagens emergem como estratégias promissoras para aprimorar a capacidade dos modelos de IA em executar tarefas com qualidade e eficiência ([BROWN et al., 2020](#); [ZHANG](#); [YUAN](#); [YAO, 2023](#)).

Além dessas, outras técnicas também vêm sendo discutidas na literatura. Em um levantamento abrangente, [Debnath et al. \(2025\)](#) apresentam não apenas essas abordagens fundamentais, mas também estratégias avançadas, que ampliam ainda mais o potencial dos LLMs em tarefas complexas e especializadas:

- **Zero-Shot Prompting:** Modelos de linguagem geram respostas sem exemplos. Essa abordagem é utilizada em PLN para realizar tarefas específicas sem que o modelo tenha sido previamente treinado ou exemplificado para aquela tarefa.
- **Few-Shot Prompting:** Técnica sofisticada de PLN que melhora a capacidade do modelo de executar tarefas fornecendo alguns poucos exemplos no próprio prompt.

2.7 Trabalhos Relacionados

Esta seção inicia-se com a revisão conduzida por [Santos, Bouchard e Napoleão \(2024\)](#), que apresenta uma Revisão Sistemática da Literatura (RSL) sobre abordagens de geração automática de histórias de usuário. O objetivo do estudo foi identificar e analisar as

estratégias existentes para automatizar a criação de histórias de usuário, considerando diferentes fontes de entrada, técnicas aplicadas e métodos de avaliação empregados.

Diversos estudos têm explorado a geração automática de histórias de usuário a partir de diferentes fontes, como requisitos textuais, conversas naturais ou modelos de processo ([HENG; TSILIONIS; WAUTELET, 2023](#); [KUMAR; TIWARI; DOBHAL, 2022](#); [RAHARJANA; SIAHAAN; FATICHAH, 2019](#); [SIAHAAN; RAHARJANA; FATICHAH, 2023](#)). No entanto, apenas uma parcela dessas pesquisas realiza avaliações estruturadas da qualidade das histórias geradas. Dentre os estudos revisados, quatro (24%) adotaram abordagens qualitativas para analisar as características das histórias de usuários geradas. [Murtazina e Avdeenko \(2019\)](#) utilizaram regras ontológicas para validar a presença dos principais elementos de uma história (papel, ação e benefício), destacando o valor dessa abordagem para ambientes ágeis. [Resketi et al. \(2020\)](#) integraram a ferramenta AQUASA, baseada no framework Quality User Story (QUS), como uma etapa de pré-processamento para avaliar e aprimorar a qualidade das histórias antes e depois da aplicação de sua técnica. Nesse caso, o conjunto final de histórias foi avaliado estatisticamente e também por profissionais de software por meio de uma pesquisa. [Nistala et al. \(2022\)](#) basearam sua avaliação na experiência de especialistas, que analisaram a saída gerada comparando-a com dados de especificações originais, atribuindo pontuações e comentários. Já [Lam, Hurtado e Portillo \(2022\)](#) conduziram um experimento controlado com uma pesquisa estruturada baseada na norma ISO 9126, avaliando aspectos como completude, correção e verificabilidade das histórias geradas por um chatbot. Os resultados indicaram que as histórias produzidas pela ferramenta superaram aquelas criadas com técnicas tradicionais da Engenharia de Requisitos.

No que diz respeito ao emprego de métodos quantitativos, alguns estudos adotaram métricas para mensurar a qualidade das histórias geradas. [Rodeghero et al. \(2017\)](#) utilizaram métricas clássicas de aprendizado de máquina, como precisão, revocação e taxa de verdadeiros positivos, para avaliar dados extraídos automaticamente em comparação com dados anotados manualmente. [Veitía, Roldán e Vegetti \(2020\)](#) utilizaram o modelo ELMo, treinado com dados extraídos de rastreadores de problemas, com o objetivo de classificar automaticamente se uma issue representa ou não uma

história de usuário, alcançando cerca de 96% de acurácia. Trata-se do único estudo identificado que emprega redes neurais para essa tarefa, o que evidencia o potencial dos modelos de linguagem. [Kumar, Tiwari e Dobhal \(2022\)](#) realizaram uma avaliação estatística utilizando o índice de silhueta para medir a coesão entre grupos de histórias de usuário, propondo uma análise quantitativa da integridade dos conjuntos gerados. [Dwitama e Rusli \(2020\)](#), por sua vez, avaliaram o desempenho de um chatbot por meio de métricas como tempo de resposta e número de interações, além de aplicar uma pesquisa com usuários, embora não tenham avaliado diretamente a qualidade das histórias resultantes. Esses trabalhos demonstram a variedade de estratégias empregadas para validar a geração automática de histórias, variando entre métricas de performance computacional, medidas de agrupamento e usabilidade.

Dentre os trabalhos identificados na revisão sistemática sobre geração automática de histórias de usuário, alguns se destacam por sua maior proximidade com o presente estudo, seja pelo uso de frameworks estruturados de avaliação da qualidade das histórias, seja pela aplicação de experimentos com usuários ou agentes baseados em linguagem natural. O trabalho de [Resketi et al. \(2020\)](#) é o mais alinhado, pois propõe o uso do framework Quality User Story (QUS), que também é adotado nesta pesquisa, para identificar problemas em histórias de usuários e gerar versões aprimoradas com suporte automatizado, além de realizar uma validação empírica com profissionais e aplicar testes estatísticos para analisar os efeitos da intervenção. De forma semelhante, [Lam, Hurtado e Portillo \(2022\)](#) propõem uma abordagem baseada em um agente conversacional treinado com regras heurísticas para gerar histórias de usuário, comparando sua eficácia com técnicas manuais por meio de um experimento controlado com base na norma ISO 9126, que define características de qualidade de software. Ambos os estudos avaliam as histórias geradas com base em critérios claros e replicáveis, característica que também é central no presente trabalho.

Ao final da revisão, os autores concluíram que ainda há amplo espaço para a aplicação de melhores diretrizes qualitativas na geração automática de histórias de usuário. Eles argumentam que a avaliação estatística pode não ser a abordagem mais adequada para mensurar a qualidade dessas histórias, considerando sua relevância no

ciclo de vida do desenvolvimento de software e o impacto potencial de requisitos mal definidos no produto final. De modo geral, a geração automatizada de histórias de usuários mostra-se uma área promissora de pesquisa, no entanto, os estudos identificados nesta RSL indicam que ela ainda está em estágios iniciais de desenvolvimento, o que reforça tanto o caráter exploratório dos estudos atuais quanto a diversidade de técnicas de inteligência artificial que ainda podem ser exploradas nesse contexto.

Diversos outros estudos investigam o uso de LLMs na geração de requisitos de software. [Marques, Silva e Bernardino \(2024\)](#) destaca os benefícios do ChatGPT na engenharia de requisitos, como a automação da documentação, redução de erros e aumento da eficiência das equipes. [White et al. \(2024\)](#) enfatiza a importância de prompts bem estruturados, apresentando técnicas de design de prompts para engenharia de software.

[Arvidsson e Axell \(2023\)](#) exploram como técnicas de engenharia de prompt podem melhorar a qualidade dos outputs dos LLMs, oferecendo diretrizes práticas e um mapeamento para atividades específicas de engenharia de requisitos. [Zhang, Yuan e Yao \(2023\)](#), [Hiraou \(2024\)](#), [Yildirim, Cevik e Başar \(2023\)](#) analisam o impacto do Meta Prompting e do Few-Shot Learning na precisão dos LLMs. [Yildirim, Cevik e Başar \(2023\)](#) sugere uma solução de duas etapas para melhorar a qualidade dos requisitos, aplicando técnicas de few-shot learning e ajustando o modelo com dados de domínio específicos.

[Krishna et al. \(2024b\)](#) compara a geração automática de documentos de Especificação de Requisitos de Software (SRS) usando modelos de linguagem, como GPT-4 e CodeLlama, com o trabalho de engenheiros iniciantes. Os resultados indicam que esses modelos podem produzir rascunhos de SRS comparáveis e identificar problemas nos documentos.

O estudo de [Ronanki, Cabrero-Daniel e Berger \(2024\)](#) avaliou o ChatGPT na replicação de avaliações humanas de histórias de usuários usando um banco de dados de Tõemets com benchmark da ferramenta AQUASA. Foram aplicados sete critérios de qualidade (atomicidade, bem-formado, minimalidade, solidez conceitual, inequívoca, completude e estimabilidade) e o ChatGPT foi testado com prompting one-shot, repetido três vezes para consistência.

[Ronanki, Berger e Horkoff \(2023\)](#) investiga o uso do ChatGPT na elicitación de requisitos, comparando os requisitos gerados pelo modelo com aqueles formulados por especialistas humanos. Os requisitos do ChatGPT mostraram-se altamente abstratos, atômicos, consistentes, corretos e compreensíveis, evidenciando que a IA pode gerar resultados de qualidade comparável ou superior a humanos em certos atributos. No entanto, também apresentaram limitações importantes, como ambiguidade e baixa viabilidade, indicando que a supervisão e revisão humanas são essenciais. Este estudo reforça a utilidade de LLMs para auxiliar na geração de requisitos, ao mesmo tempo que destaca a necessidade de intervenção humana para garantir clareza e aplicabilidade prática.

[Rahman e Zhu \(2024\)](#) introduziram a ferramenta "GeneUS" que utiliza o GPT-4.0 e a técnica "Chain-of-Thought Prompting" (CoT) para automatizar a geração de histórias de usuários. A estratégia "Refine and Thought" (RaT) é empregada para extrair e refinar requisitos, e a qualidade das histórias é avaliada com o questionário RUST.

O estudo de [Ramasamy et al. \(2024\)](#) explora o uso de técnicas da OpenAI para aprimorar a geração de histórias de usuário em ambientes de desenvolvimento ágil de software. Os autores investigam o uso de engenharia de prompt para melhorar a qualidade e a abrangência das histórias geradas, com um foco particular em projetos de capstone acadêmicos. A pesquisa compara equipes de estudantes que geraram histórias de usuário usando métodos tradicionais versus equipes que usaram agentes OpenAI como GPT-4-turbo.

[Sharma e Tripathi \(2025\)](#) investigaram a eficácia dos LLMs na avaliação da qualidade de histórias de usuário, utilizando o framework Quality User Story (QUS). A pesquisa usa três LLMs de última geração (GPT-4o, GPT-4-Turbo e GPT-3.5-Turbo) e dois tipos de estratégias de prompting (contexto mínimo e contexto rico) para avaliar o desempenho em oito critérios de qualidade de histórias de usuário. Os resultados mostram que o GPT-4o e o GPT-4-Turbo se destacam na avaliação das histórias, especialmente nos critérios sintáticos e pragmáticos, com desempenho estável mesmo com aumento de contexto.

O estudo de [Brockenbrough e Salinas \(2024\)](#) investiga o uso de ChatGPT por

estudantes de Ciência da Computação na criação de histórias de usuário a partir de feedback de usuários. Ele compara histórias geradas com e sem assistência de IA, avaliando qualidade e tempo de elaboração. As histórias foram analisadas usando uma versão modificada do framework INVEST, considerando estrutura, independência, valor, testabilidade e gramática. Os resultados indicam que o uso do ChatGPT melhora a qualidade geral das histórias, embora não garanta sempre resultados corretos, já que o modelo pode falhar em identificar dependências ou gerar estruturas incorretas. O estudo ressalta a importância de treinamento e revisão humana para aproveitar efetivamente as histórias assistidas por IA.

O estudo de [Oswal, Kanakia e Suktel \(2024\)](#) propõe uma abordagem baseada em IA utilizando GPT-3.5, capaz de transformar automaticamente textos de requisitos em histórias de usuário padronizadas. A implementação prática inclui um aplicativo que processa os requisitos fornecidos pelo usuário, aplicando técnicas de prompting para gerar histórias de usuário consistentes e estruturadas, demonstrando a viabilidade de automatizar a geração de histórias em contextos ágeis.

[Neo et al. \(2024\)](#) desenvolveram o User Story Tutor (UST), uma aplicação web que verifica a legibilidade de histórias de usuário e recomenda práticas para melhoria. Além disso, estima o esforço necessário para a implementação usando técnicas de aprendizado de máquina. A aceitação do UST foi avaliada por 40 profissionais ágeis utilizando o TAM e o framework AttrakDiff, apresentando bons resultados em variáveis como facilidade de uso e utilidade percebida.

Apesar da similaridade em relação ao objetivo de gerar automaticamente histórias de usuário com boa qualidade, o presente estudo se diferencia em diversos aspectos importantes. Enquanto estudos anteriores como os de [Resketi et al. \(2020\)](#) e [Lam, Hurtado e Portillo \(2022\)](#) focam em abordagens baseadas em regras heurísticas ou integração com frameworks tradicionais de avaliação, esta pesquisa explora diretamente o uso de técnicas avançadas de prompting com LLMs, especialmente Meta Prompting e Few-Shot Prompting, com foco específico na qualidade estrutural e conceitual das histórias. Além disso, a avaliação conduzida neste estudo adota o framework QUS não apenas como ferramenta de análise, como por exemplo no trabalho de [Resketi et al. \(2020\)](#),

mas também como instrumento didático e de refinamento das instruções (prompts), não como validação, como em [Sharma e Tripathi \(2025\)](#) e [Ronanki, Cabrero-Daniel e Berger \(2024\)](#). Diferentemente de [Rahman e Zhu \(2024\)](#), que aplicam Chain-of-Thought Prompting para construir requisitos, aqui o foco está em estratégias de engenharia de prompt para induzir geração de histórias de maior qualidade desde o início, e enquanto eles avaliam a qualidade das histórias com base no questionário RUST, esta pesquisa adota o framework QUS como fundamento conceitual e prático, tanto para orientar a geração quanto para avaliar a qualidade das histórias. Por fim, diferentemente da maioria dos estudos que dependem exclusivamente de avaliação automatizada ou de especialistas, este trabalho também realiza análises quantitativas e qualitativas, buscando entendimento também das percepções dos participantes. Essas contribuições indicam que o presente estudo pode representar um passo promissor no uso sistemático de LLMs para apoiar práticas de engenharia de requisitos, contribuindo para uma geração de histórias de usuário com maior padronização e qualidade.

3

CICLO DE RELEVÂNCIA: SURVEY COM PROFISSIONAIS DA INDÚSTRIA

Este capítulo apresenta os resultados de um survey realizado com profissionais da indústria com experiência em engenharia de requisitos. O objetivo foi investigar como esses profissionais lidam com requisitos em seu dia a dia, especialmente no que se refere à elicitación e à especificação de requisitos, e como percebem o uso de Modelos de Linguagem de Grande Escala (LLMs) em suas atividades, com ênfase no interesse e na viabilidade da geração automatizada de histórias de usuários.

3.1 INTRODUÇÃO

Antes de propor soluções automatizadas, é fundamental compreender como os profissionais da área percebem as tecnologias emergentes e em que medida veem valor em aplicá-las na automação de atividades, como a geração de histórias de usuários. Para isso, buscou-se entender como os requisitos são elicitados na prática, quais ferramentas já são utilizadas e se há interesse ou abertura para o uso de LLMs nesse contexto. Essa motivação sustenta a proposta deste estudo, que busca obter uma visão ampla sobre o cenário atual, de modo que futuras soluções estejam alinhadas às práticas e expectativas do ambiente profissional.

Com base nisso, o objetivo foi investigar como esses profissionais lidam com

requisitos em seu dia a dia, especialmente no que se refere à elicitación e à especificação de requisitos, e como percebem o uso de LLMs em suas atividades, com ênfase no interesse e na viabilidade da geração automatizada de histórias de usuários.

A partir dessa investigação, utilizando as seguintes questões de pesquisa, busca-se compreender as práticas adotadas, os desafios enfrentados e as oportunidades percebidas por profissionais da indústria que atuam diretamente com requisitos em contextos reais de desenvolvimento de software.

QP1: Como os profissionais da indústria lidam com requisitos em suas atividades cotidianas?

QP2: De que forma os profissionais utilizam Modelos de Linguagem de Grande Escala (LLMs) no contexto da engenharia de requisitos?

QP3: Qual é a percepção dos profissionais sobre a geração automatizada de requisitos, especialmente histórias de usuários, com o uso de LLMs?

3.2 METODOLOGIA

Pesquisas de opinião estão entre os métodos de pesquisa mais utilizados, especialmente em estudos que buscam investigar opiniões, atitudes ou padrões de comportamento de um grupo. Mais do que um simples questionário ou lista de verificação, um survey constitui um sistema abrangente de coleta de informações, com o objetivo de descrever, comparar ou explicar conhecimentos, percepções e ações. Esse processo envolve uma série de atividades interdependentes que estruturam a condução de um survey (KITCHENHAM; PFLEEGER, 2008; PFLEEGER; KITCHENHAM, 2001).

Neste estudo, optou-se por aplicar survey de forma online pois oferecem diversas vantagens significativas, conforme apontado por (PUNTER et al., 2003; LINÅKER et al., 2015). Esse formato de survey não só simplifica o acesso dos respondentes ao questionário, mas também reduz o esforço necessário para o preenchimento por parte dos participantes. Além disso, a navegação dinâmica do questionário, que se adapta às respostas anteriores, e a facilidade na análise dos dados registrados eletronicamente são outros pontos fortes que tornam as pesquisas online uma ferramenta eficiente. A seguir,

detalhamos o desenho da pesquisa, a execução e a análise dos resultados.

3.2.1 Desenho da pesquisa

O desenho adotado neste estudo seguiu as diretrizes propostas por (KITCHENHAM; PFLEEGER, 2008), que recomendam, inicialmente, a definição clara dos objetivos da pesquisa, garantindo que sejam específicos e mensuráveis. Em seguida, realiza-se o desenho da pesquisa, no qual são tomadas decisões sobre a abordagem metodológica e a população-alvo. Com isso definido, desenvolve-se o instrumento de pesquisa, geralmente um questionário, que deve ser cuidadosamente elaborado. A etapa seguinte consiste na avaliação desse instrumento, visando assegurar sua clareza, validade e confiabilidade. Após a coleta, é fundamental garantir que os dados obtidos sejam válidos e consistentes. Por fim, procede-se à análise dos dados, a partir da qual os resultados são interpretados e utilizados para responder às três perguntas de pesquisa. Para tanto, as principais questões relacionadas ao desenho da pesquisa são: a definição da população-alvo e a caracterização dos respondentes, bem como o desenho do questionário, que serão explicados a seguir.

O estudo teve como objetivo investigar como esses profissionais lidam com os requisitos em seu trabalho diário, particularmente em relação às atividades de elicitação e especificação, e como eles percebem o uso de requisitos nessas tarefas. A população-alvo é o grupo ou indivíduos aos quais a pesquisa se aplica, selecionados de forma a garantir que os participantes tenham experiência suficiente para responder às questões propostas. Assim, a população foi composta por profissionais da indústria com experiência prévia em atividades de elicitação de requisitos. Para atingir esse perfil desejado, foram recrutados sujeitos alinhados aos objetivos do estudo.

Os participantes foram selecionados com base em critérios de elegibilidade, como atuação profissional na indústria de software e experiência prévia em atividades de elicitação de requisitos. Assim, a amostragem adotada foi não probabilística por conveniência, conforme caracterizado por (KITCHENHAM; PFLEEGER, 2008), o que é comum em estudos exploratórios na área de Engenharia de Software quando há neces-

cidade de acesso a especialistas com perfil específico. Embora esse tipo de amostragem limite a generalização dos resultados, ele é adequado para obter percepções qualificadas de profissionais experientes sobre práticas reais de engenharia de requisitos.

Já com o instrumento de pesquisa para a coleta de dados, optou-se pela utilização de um questionário online, por se tratar de um instrumento eficiente para alcançar participantes geograficamente distribuídos e com maior conveniência de resposta. Segundo (LINÅKER et al., 2015), ressaltam que o uso de questionários online podem preservar a confidencialidade das respostas, evita a influência do pesquisador sobre os participantes e reduz os custos operacionais. Tais características foram determinantes para a escolha deste formato, considerando o público-alvo e os objetivos da pesquisa.

Para coletar de forma eficaz as informações necessárias para responder às três perguntas de pesquisa, foi elaborado um questionário dividido em quatro seções. A Figura 2 apresenta um fluxograma que representa a lógica de navegação do questionário. Nesse fluxograma, os losangos indicam perguntas condicionais que direcionam os participantes a diferentes ramificações, de acordo com suas respostas (por exemplo, "Sim" ou "Não"). Esse esquema facilita a visualização da progressão e dos filtros aplicados ao longo do preenchimento do questionário. A Tabela 1 apresentada a descrição das questões de cada uma das quatro seções que compõem o questionário, totalizando 28 perguntas (Q1 a Q28).

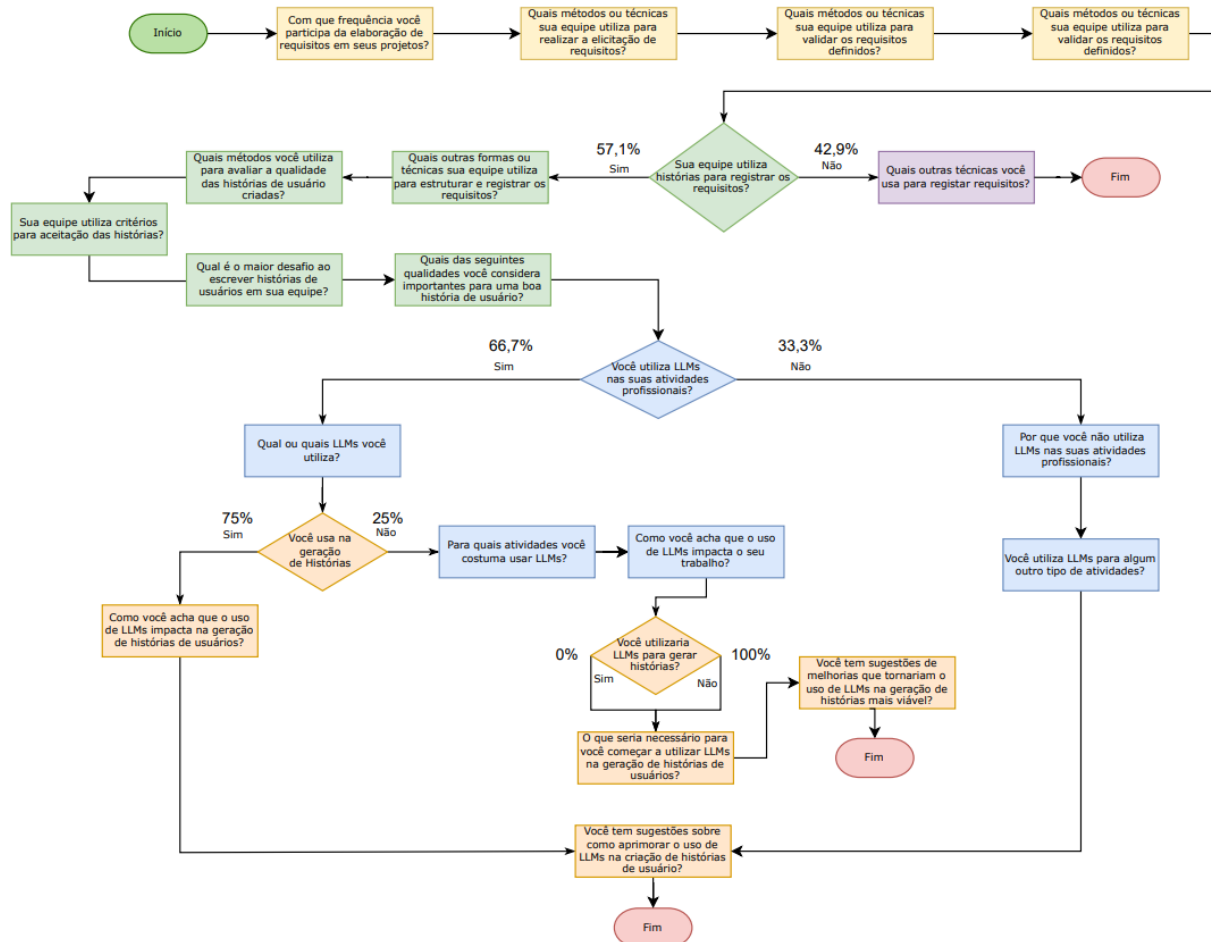


Figura 2 – Fluxograma do Questionário com Profissionais de Requisitos.

3.2.2 Execução

O questionário online, foi criado utilizando a ferramenta Google Forms, uma escolha motivada pela praticidade de uso, ampla acessibilidade e potencial para alcançar uma maior amplitude de participantes. Ao final da elaboração do questionário, foi conduzido um estudo piloto com dois participantes profissionais e pesquisadores da área, ambos com experiência prévia em requisitos e pesquisas relacionadas. O objetivo dessa etapa foi avaliar a clareza, a relevância e a compreensibilidade das perguntas do questionário, bem como identificar possíveis ambiguidades ou interpretações divergentes. Essa verificação preliminar foi fundamental para assegurar que os participantes da etapa principal do estudo compreendessem corretamente os questionamentos, garantindo

Tabela 1 – Classificação das Perguntas do Questionário por Seção.

Seção	ID	Pergunta Completa
Informações Demográficas	Q1	Nome
	Q2	Email
	Q3	Qual é o seu cargo atual?
	Q4	Há quanto tempo você trabalha na área de desenvolvimento de software?
	Q5	Qual é o nome da empresa em que você trabalha?
	Q6	Onde ela está localizada (cidade e estado)?
Práticas de Engenharia de Requisitos	Q7	Com que frequência você participa da elaboração de requisitos em seus projetos?
	Q8	Quais métodos ou técnicas sua equipe utiliza para realizar a elicitação de requisitos?
	Q9	Quais métodos ou técnicas sua equipe utiliza para validar os requisitos definidos?
	Q10	Sua equipe utiliza histórias de usuários para estruturar e registrar os requisitos gerados?
	Q11	Além de histórias de usuários, quais outras formas ou técnicas sua equipe utiliza para estruturar e registrar os requisitos?
	Q12	Quais procedimentos ou métodos você utiliza para avaliar a qualidade das histórias de usuário criadas?
	Q13	Sua equipe utiliza critérios para aceitação das histórias? Se sim, quais são os critérios e quais informações são incluídas para que a história seja considerada pronta?
	Q14	Na sua opinião, qual é o maior desafio ao escrever histórias de usuários em sua equipe?
	Q15	Quais das seguintes qualidades você considera importantes para uma boa história de usuário?
	Q16	Quais formas ou técnicas sua equipe utiliza para estruturar e registrar os requisitos?
Uso de LLMs	Q17	Você utiliza Modelos de Linguagem (LLMs) nas suas atividades profissionais?
	Q18	Qual ou quais LLMs você utiliza?
	Q19	Você utiliza LLMs na geração de histórias de usuários?
	Q20	Para quais atividades você costuma usar LLMs?
	Q21	Como você acha que o uso de LLMs impacta o seu trabalho? Explique os impactos positivos e/ou negativos:
	Q22	Você utilizaria LLMs para gerar histórias de usuários?
	Q23	O que seria necessário para você começar a utilizar LLMs na geração de histórias de usuários?
	Q25	Como você acha que o uso de LLMs impacta na geração de histórias de usuários? Explique os impactos positivos e/ou negativos:
	Q26	Por que você não utiliza LLMs nas suas atividades profissionais?
	Q27	Você utiliza LLMs para algum outro tipo de atividades?
Sugestões e Opiniões Finais	Q24	Você tem sugestões de melhorias que tornariam o uso de LLMs na geração de histórias mais viável?
	Q28	Você tem sugestões, ideias ou percepções sobre como aprimorar o uso de LLMs na criação de histórias de usuário?

que o instrumento de coleta estivesse alinhado aos objetivos da pesquisa.

A versão final do questionário foi disponibilizada por um período de 14 dias. A divulgação ocorreu por meio de publicações abertas na plataforma LinkedIn e por convites diretos via e-mail e aplicativo de mensagens instantâneas, utilizando o critério de conveniência. Em todos os canais, foi solicitado que apenas profissionais com experiência em engenharia de requisitos participassem da pesquisa, de modo a ampliar o alcance da pesquisa e garantir a adequação do perfil dos respondentes aos objetivos do estudo. Ao todo houve a participação de 21 respondentes. O conteúdo completo do questionário, incluindo as perguntas aplicadas e a lógica condicional de navegação, encontra-se disponível no Apêndice [A](#).

3.2.3 Resultados da Pesquisa

A seguir, são apresentados os resultados do questionário, separadas por seção.

Seção 1 – Informações Demográficas

Os resultados obtidos indicam que os participantes ocupam, majoritariamente, os cargos de Product Owner (PO) e Desenvolvedor, ambos representando 28,6% da amostra. Além disso, 19% dos respondentes atuam como analistas de requisitos. Esses dados demonstram uma predominância de profissionais diretamente envolvidos com a definição e implementação de requisitos, o que reforça a adequação da amostra ao objetivo da pesquisa.

Os demais participantes ocupam cargos como gerente de produto (Product Manager), testador, UX designer e especialista técnico, cada um representando 4,8% da amostra. A presença desses profissionais contribui para ampliar a diversidade da amostra, incorporando perspectivas complementares sobre as práticas de engenharia de requisitos (Figura 3).

Quanto à experiência, os dados revelam que 65% dos participantes possuem mais de cinco anos de experiência na área de desenvolvimento de software, sendo que 23,8% atuam há mais de 20 anos, 23,8% há mais de 10 anos e 14,3% há mais de 5 anos. Os demais respondentes possuem entre 1 e 5 anos de experiência, sendo 19% com 1 a 3

Qual é o seu cargo atual?

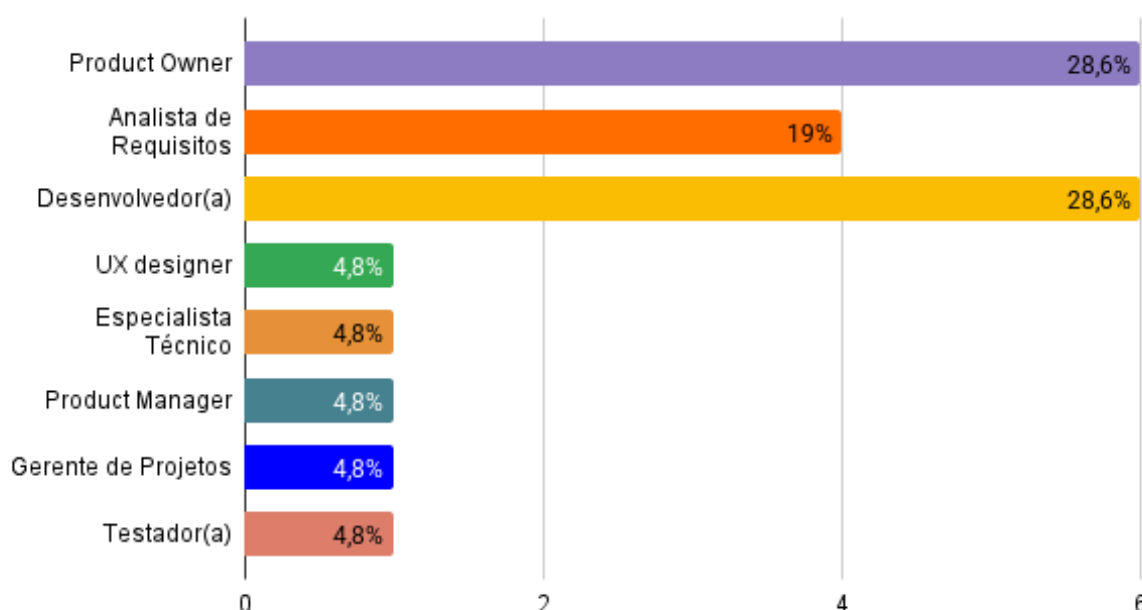


Figura 3 – Distribuição dos cargos dos participantes do estudo.

anos e 14,3% com 3 a 5 anos de atuação. E a grande maioria dos participantes (85,7%) atua em empresas sediadas em Manaus, Amazonas. Os demais trabalham de forma remota para empresas localizadas em Florianópolis (SC) e Belo Horizonte (MG).

Seção 2 – Práticas de Engenharia de Requisitos

Com o foco em compreender a frequência que os participantes atuam na elaboração de requisitos, observou-se que a maioria dos participantes (95,2%) participa frequentemente da elaboração de requisitos em seus projetos. Apenas um participante (4,8%) declarou participar ocasionalmente. Esse resultado reforça que a amostra é composta majoritariamente por profissionais fortemente envolvidos nas atividades de engenharia de requisitos.

Através de perguntas abertas e descritivas, foi possível identificar quais métodos ou técnicas suas equipes utilizam para realizar a elicitação de requisitos e observar padrões recorrentes.

Os dados revelam que entrevistas com stakeholders são a técnica de elicitação mais utilizada técnica de elicitação, com 17 menções diretas, frequentemente descritas

como uma etapa inicial ou contínua ao longo dos ciclos de desenvolvimento, muitas vezes combinadas com outras práticas. A análise de documentos ocupa o segundo lugar, com 12 menções, utilizada para compreender o contexto e complementar os dados coletados. Outras técnicas relevantes incluem brainstorming (7 menções), comum em workshops e sessões de ideação, e Lean Inception (5 menções), destacada como uma abordagem estruturada para alinhamento estratégico. A observação direta (4 menções) também foi relatada, principalmente em cenários que exigem compreensão do processo em campo.

Da mesma forma, as respostas abertas referentes aos métodos e técnicas utilizados para validar os requisitos permitiram identificar que a Prototipagem é a prática mais recorrente e central no processo de validação. O uso de protótipos, mockups ou wireframes, foram amplamente descritos (em pelo menos 12 respostas), com protótipos de alta ou baixa fidelidade, como forma de verificar se os requisitos atendem às necessidades do projeto e dos stakeholders. Em segundo lugar, destacam-se Revisões com Stakeholders também foram mencionadas (em cerca de 10 respostas), frequentemente associadas tanto a reuniões formais, quanto em interações informais ou homologações, nas quais os envolvidos no projeto avaliam e ajustam os requisitos com base em suas expectativas e feedbacks. Outras práticas incluem verificações baseadas em documentos (aproximadamente 4 a 5 menções), como análise de requisitos formalizados, comparação com normas e rastreabilidade. Esses achados indicam uma forte adoção de abordagens iterativas e colaborativas tanto na elicitação quanto na validação de requisitos.

Por fim, a fase de testes, homologação e validação em ambientes controlados também foi mencionada, evidenciando que o feedback obtido em testes de usabilidade e nos processos de homologação com os stakeholders contribui de forma significativa para verificar a qualidade e a adequação dos requisitos antes da implementação final.

Quando questionados sobre o uso de histórias de usuários para estruturar e registrar os requisitos gerados, os dados revelaram um cenário equilibrado, como visto no gráfico da Figura 4, com 57,1% dos participantes afirmando utilizar histórias de usuários, contra 42,9% que disseram não utilizar. Apesar da diferença estreita, o resultado demonstra que as histórias de usuários estão mais presentes do que se

imaginava, consolidando-se como uma prática cada vez mais comum na estruturação e registro de requisitos. O uso dessa abordagem indica uma tendência de amadurecimento e adoção mais ampla entre as equipes de desenvolvimento.

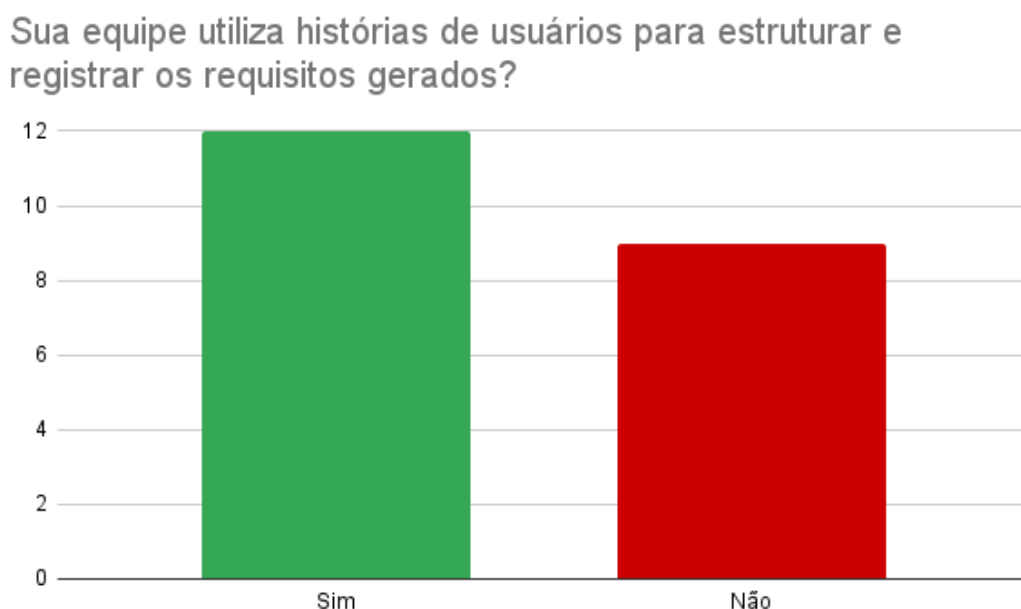


Figura 4 – Distribuição do uso de Histórias de Usuários pelos participantes do estudo.

Entre os 57,1% dos participantes que declararam utilizar histórias de usuário, também foram identificadas outras formas ou técnicas empregadas pelas equipes para estruturar e registrar os requisitos. Verificou-se que os diagramas representam outra técnica mais empregada para estruturar e registrar requisitos, sendo mencionados por 83,3% desse grupo. Esse resultado evidencia a preferência por recursos gráficos para descrever funcionalidades, fluxos e a arquitetura dos sistemas. Casos de uso e mockups/protótipos também se destacam como práticas recorrentes, ambos com 66,7% de adoção, evidenciando uma abordagem mista entre descrição textual estruturada e elementos de interface.

Esses resultados mostram que, mesmo com o crescimento das histórias de usuários, as equipes frequentemente adotam uma estratégia complementar, combinando técnicas tradicionais e recursos visuais para representar os requisitos de forma mais clara e acessível.

Entre os 42,9% dos participantes que declararam não utilizar histórias de usuário,

observou-se maior presença de métodos mais convencionais. Mockups e protótipos foram a técnica mais citada nesse grupo, mencionados por 77,8%, o que aponta uma forte preferência por artefatos de interface como meio de explorar e validar funcionalidades.

Além disso, diagramas e casos de uso foram igualmente mencionados por 55,6% dos respondentes, reforçando o papel dessas ferramentas na comunicação e estruturação dos requisitos, mesmo fora do contexto das histórias de usuários.

Os profissionais que utilizam histórias de usuário relataram a adoção de diferentes técnicas para garantir que essas histórias sejam claras, viáveis e alinhadas às necessidades dos usuários e do negócio. Dentre as técnicas mais citadas, destacam-se as revisões em equipe (peer review), mencionadas por praticamente todos os respondentes. Essa prática colaborativa permite identificar ambiguidades, inconsistências e oportunidades de melhoria antes da implementação. Outra técnica recorrente é a validação com stakeholders, considerada um elemento central nos processos adotados. Essa validação ocorre por meio de reuniões, apresentações de protótipos e interações diretas com os usuários finais, assegurando que as histórias de usuário reflitam com precisão os requisitos reais do projeto.

Em relação aos critérios priorizados na avaliação das histórias, os participantes destacaram aspectos como clareza, alinhamento com os requisitos do usuário e viabilidade técnica. Como exemplificado pelo participante P10: *"Durante a avaliação, priorizamos clareza, alinhamento com os requisitos do usuário e viabilidade de implementação dentro dos ciclos de desenvolvimento ágil."* Já o participante P18 complementa: *Revisões em equipe, validação com stakeholders, critérios de aceitação, priorização e refinamento contínuo."*

Além dessas práticas, observou-se a adoção de atividades complementares ao longo do processo de desenvolvimento, como reuniões utilizadas para o detalhamento técnico e definição de critérios de aceitação. Como citado por P16 que relata: *"...No nosso time fazemos reuniões de Grooming para entendimento da funcionalidade a ser desenvolvida. Após essa reunião o time de upstream continua com a definição das regras e prototipação..."*.

Algumas equipes também integram o refinamento e discussão das histórias em dailys e retrospectivas, favorecendo o alinhamento com o planejamento das sprints e com o feedback contínuo dos stakeholders. Evidenciando um esforço consistente das

equipes em adotar boas práticas de avaliação de requisitos, combinando abordagens colaborativas, validações com usuários e ciclos iterativos de melhoria contínua.

Também foram identificadas práticas adotadas pelas equipes em relação à aceitação das histórias de usuário. Os profissionais relataram utilizar uma variedade de critérios, em diversas equipes, isso ocorreu por meio de checklists ou Definition of Done (DoD).

O participante P1 afirma que considera como critérios de aceitação todos os elementos que garantam que a história esteja completa: *“... tudo que eu considero que a US siga para ser considerada completa. Regras de negócio, comportamentos, requisitos de qualidade (escalabilidade, etc.).”*

Para o P7, os critérios devem ser definidos conforme o projeto, e devem incluir o máximo de informação relevante: *“... os critérios dependem muito do projeto, mas incluímos o máximo de informações técnicas, de negócio e de UX de forma mensurável.”* Já o P10 detalha dois critérios fundamentais para validar histórias: *“Funcionalidade Atendida – A história deve implementar corretamente a funcionalidade descrita, atendendo ao objetivo do usuário. Testes de Aceitação – Devem estar definidos e ser executáveis, garantindo que a funcionalidade possa ser validada por meio de testes automatizados ou manuais.”*

O P13 enfatiza que os critérios são definidos de acordo com a especificidade de cada história: *“Critérios são definidos diante da especificidade de cada história.”* A participação do PO e do time também é destacada, como explica o P15: *“Os critérios são definidos pelo PO. A equipe é consultada e cada membro pontua sua história.”* Complementando, o P16 relata: *“...Criamos uma seção de DoD (Definition of Done) nos cards com os critérios de aceite.”*

Entre os itens recorrentes nos critérios, o P17 lista que *“A história de usuário deve ser descrita de forma clara e compreensível para todos os envolvidos. Regras de negócios e critérios de aceitação definidos com as condições necessárias para que a história seja considerada concluída. Qualquer dependência técnica ou funcional deve ser documentada, evitando surpresas durante a implementação. A história também deve ser pequena o suficiente para ser concluída em uma única iteração (sprint), e entregar valor claro para o usuário ou para o negócio.”*

O planejamento e a verificação dos critérios também fazem parte do processo. O P19 explica que *“o critério é feito durante a Planning de cada sprint, em que o time sinaliza se*

há informações mínimas suficientes para início da task.” O P20 reforça a ideia de verificação com base em checklist: *“... uma história é aceita se atender a um checklist, onde, por exemplo, são descritos os critérios de aceite (da atividade quando concluída).”* Por fim, o P21 destaca que *“os critérios de aceitação são as aplicações dos cenários de teste verificando se o resultado é conforme o esperado.”*

Os resultados indicam que os critérios mais frequentemente citados para considerar uma história pronta incluem a clareza na descrição, a definição das regras de negócio e a testes de aceitação bem definidos. Esses elementos garantem que a história seja compreendida por todos e validada corretamente. Também foram mencionados critérios como adequação ao tamanho da sprint, requisitos de qualidade (como escalabilidade e usabilidade) e boas práticas de engenharia, reforçando o compromisso das equipes com a entrega de histórias bem definidas e alinhadas às expectativas técnicas e de negócio.

Em uma questão de múltiplas respostas, onde os participantes puderam selecionar mais de uma alternativa, questionou-se sobre os desafios encontrados ao escrever histórias de usuários nas suas equipes, e os principais desafios identificados foram: falta de informações claras sobre as necessidades dos usuários (75%), problemas na comunicação entre as equipes envolvidas (58,3%) e falta de tempo para elaborar histórias detalhadas (50%).

Esses resultados indicam que os principais obstáculos enfrentados pelas equipes estão associados tanto à compreensão insuficiente das necessidades dos usuários quanto a limitações organizacionais e de processo. Isso evidencia a importância de tornar o processo de escrita das histórias mais claro, preciso e eficiente, de modo a facilitar o alinhamento entre os envolvidos e a entrega de valor de forma mais ágil.

Na sequência, pediu-se que os participantes selecionassem todos os critérios que consideravam relevantes para uma boa história de usuário (questão de múltipla escolha). Para esta análise, foram consideradas as opções que atingiram ao menos 50% de seleção. A Figura 5 apresenta os critérios mais mencionados:

Seção 3 – Uso de LLMs

Em relação ao uso de LLMs, a pesquisa revelou que 66,7% dos participantes

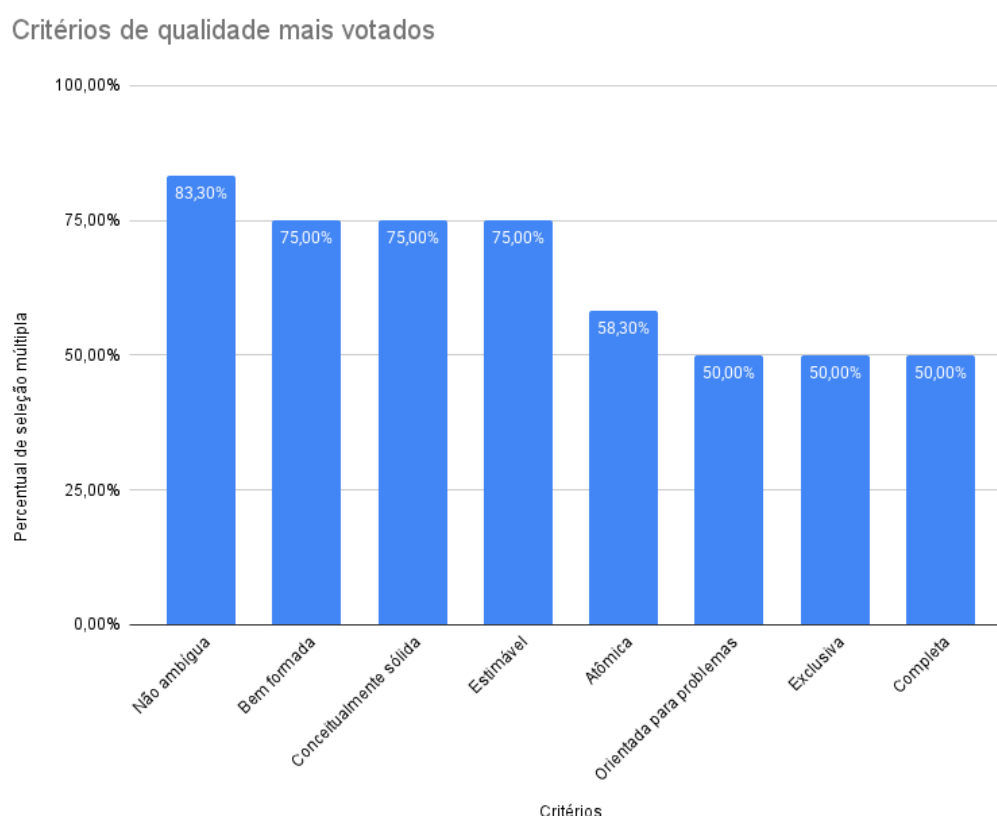


Figura 5 – Percentual de participantes que selecionaram cada critério de qualidade para histórias de usuário.

utilizam Modelos de Linguagem (LLMs), como ChatGPT, Gemini, LLama e Copilot, em suas atividades profissionais, enquanto 33,3% não fazem uso dessas ferramentas. O ChatGPT foi o modelo mais citado, mencionado em todas as 8 respostas, ou seja, 100% de frequência, seguido pelo Copilot, com 3 menções, ou 37,5% das respostas. Os outros dois modelos mencionados foram DeepSeek e Gamma.

Dos quatro participantes que não utilizam LLMs em suas atividades profissionais, três justificaram a decisão com base em preocupações relacionadas à segurança e privacidade, especialmente quanto ao risco de violação da LGPD (Lei Geral de Proteção de Dados). O participante P21 argumentou: “Não sei se consigo usar essas ferramentas sem violar a LGPD”. O P15 apontou a existência de restrições organizacionais e a necessidade de proteger informações sensíveis: “Existem regras na empresa que limitam uso de tais ferramentas.” Já o P13 foi direto ao mencionar: “Evitar compartilhamento de informações confidenciais.” De modo geral, a segurança e o controle sobre dados confidenciais se

destacam como os principais fatores para a não adoção dos LLMs.

Mas quando questionados sobre o uso de LLMs em outras atividades que não profissionais, os mesmos três participantes apresentaram respostas distintas. O participante P15 afirmou que não utiliza essas ferramentas para nenhuma outra finalidade. Já o P21 relatou um uso limitado: *“Só para coisas pessoais.”*. Por fim, o P13 mencionou um uso pontual para apoio linguístico: *“Às vezes para revisão textual em outras necessidades.”*

Dentre os participantes que utilizam LLMs em suas atividades profissionais, 75% afirmaram empregá-las na geração de histórias de usuários. Os relatos indicam uma percepção positiva sobre os impactos dessas ferramentas nesse processo.

Os participantes destacaram diversos impactos positivos do uso de LLMs na geração de histórias de usuário. Entre eles, a agilidade e o ganho de tempo aparecem como um dos principais benefícios do uso de LLMs. O participante P1 afirmou: *“Reduz o trabalho, faço mais rápido e com mais detalhes...”*. Já o P10 ressaltou: *“Agilidade na Escrita – LLMs podem gerar rapidamente histórias de usuário bem estruturadas, economizando tempo da equipe.”* Na mesma linha, o P18 mencionou: *“Aceleração do Processo...”*, enquanto o P19 apontou: *“O ganho de tempo é um impacto muito positivo.”* Por fim, o P20 reforçou essa percepção: *“Agiliza e facilita o trabalho, atualmente o uso de LLMs é de grande ajuda em face dos prazos cada vez mais reduzidos.”*

Além da agilidade, os LLMs também foram valorizados como apoio à criação e ao brainstorming. O P18 relatou: *“Uso bastante como uma ferramenta de apoio! Em especial na geração de histórias de usuários que acelera muito o processo, além de ajudar a inspirar novas ideias.”* O P1 complementou: *“Inclusive com ideias da própria IA para completar, se eu der o devido contexto.”* Ainda segundo o P18: *“Inspiração, brainstorming...”*.

A padronização também foi mencionada como um impacto relevante. O P10 declarou: *“Consistência – Garante um padrão uniforme, ajudando a manter clareza e coesão entre as histórias.”* O P18 reforçou: *“Padronização”*, e o P16 afirmou: *“Já monta um modelo padrão para as histórias.”* e *“Ajuda na estruturação.”* O suporte à qualidade das histórias foi outro aspecto ressaltado. O P10 mencionou: *“Validação de Critérios – Podem verificar automaticamente se as histórias seguem boas práticas.”*

Entre os impactos negativos, a limitação mais recorrente apontada foi a dificul-

dade das LLMs em lidar com o contexto específico dos projetos. O P10 observou: *“Falta de Contexto Específico – LLMs podem não compreender completamente as nuances do projeto, gerando histórias genéricas ou desalinhadas.”* e *“Possíveis Inconsistências – Se os prompts não forem bem elaborados, as histórias geradas podem não estar alinhadas com os requisitos reais do produto.”* Já o P18 reforçou: *“Mas ainda é necessário refinar para garantir que estejam alinhadas com o contexto do projeto e as necessidades do negócio.”*, e também acrescentou: *“Superficialidade, falta de Alinhamento com o Negócio”*.

Alguns participantes apontaram riscos associados à dependência excessiva. O P16 alertou: *“Pode deixar POs e Analista de Requisitos preguiçosos, ou dependentes destas ferramentas, pois com a comodidade pode ser que os mesmos não dediquem tanto tempo para análise de requisitos.”* O P19 também ressaltou: *“Alguns profissionais apenas copiam sugestões do ChatGPT sem levar em consideração o seu contexto”*. Por fim, o P19 levantou uma preocupação relacionada a questões organizacionais, observando que: *“Em outros contextos possa ser a licença para utilização do ChatGPT”*.

De forma geral, os participantes reconhecem o potencial das LLMs para acelerar e apoiar o processo de escrita, especialmente na agilidade, estruturação das histórias e na geração de ideias. No entanto, os relatos também alertam para a importância de revisão crítica e alinhamento com o contexto do projeto, a fim de evitar dependência excessiva ou produções desajustadas e genéricas.

Embora dois participantes (P5 e P7) tenham afirmado que não utilizam LLMs na geração de histórias de usuários, eles relataram utilizá-las em outras atividades. Entre essas, a mais comum foi a geração de ideias iniciais para brainstorming, mencionada por dois respondentes. Outras atividades também foram citadas, como geração de código, escrita de documentação técnica e análise ou refinamento de requisitos, cada uma mencionada por um participante. Isso sugere que os LLMs estão sendo empregados de forma diversificada ao longo do ciclo de desenvolvimento.

Quanto aos impactos positivos, estes estão principalmente relacionados à agilidade e otimização do tempo. O participante P5 destacou que *“...o desenvolvimento se torna mais rápido, otimizando o tempo de entrega”*, indicando ganhos de produtividade com o uso das LLMs. Outro ponto positivo mencionado por P7 foi a percepção de que

os resultados produzidos com auxílio da IA tendem a ser mais completos.

No entanto, também foram apontados efeitos negativos. P5 alertou para o risco de “...deixar o desenvolvedor mal-acostumado, apenas colando o código gerado sem realmente entender a solução proposta.”. Outro participante mencionou que, embora os resultados possam ser mais completos, o processo pode acabar levando mais tempo, o que contraria a ideia de eficiência em certos contextos. Essas observações reforçam a importância de um uso crítico e equilibrado das LLMs no apoio ao desenvolvimento de software.

Apesar de ambos terem afirmado que não utilizam LLMs para gerar histórias de usuários atualmente, suas respostas indicam condições que poderiam motivar essa adoção, o que revela barreiras relacionadas à falta de tempo ou de estrutura adequada para o uso, e não uma rejeição à tecnologia em si.

O participante P7 destacou que seria necessário “ter mais tempo para dedicar a isso e o time participar mais na escrita da história”, sugerindo que, com uma técnica já pronta e validada, essa barreira poderia ser superada. Já o participante P5 apontou como principal necessidade “criar o prompt necessário para extração de uma boa história de usuário”, o que demonstra um interesse prático no uso da ferramenta, condicionado à disponibilidade de um modelo de prompt eficaz. Esses relatos indicam que a adoção de LLMs pode ser facilitada por metodologias prontas e validadas, que reduzam o esforço inicial e promovam confiança no processo automatizado.

Seção 4 – Sugestões e Opiniões Finais

Segundo os participantes, uma das formas mais mencionadas de aprimorar o uso de LLMs na criação de histórias de usuário está relacionada à melhoria da qualidade das histórias geradas. O participante P1 acredita que os LLMs podem ser úteis para revisar e refinar histórias, sugerindo que é importante “utilizar LLMs para revisar histórias e sugerir melhorias com base em critérios de aceitação”. O participante P1 também destaca o potencial de automatizar validações, recomendando “implementar verificações automáticas para garantir que as histórias sejam claras e testáveis”.

O participante P19 vê os LLMs como aliados na otimização do tempo durante esse processo: “usar LLMs para ser um aliado na criação das histórias de usuário, para que seja possível ganhar tempo”.

Uma das preocupações evidenciadas pelos participantes está relacionada ao fornecimento de contexto adequado para que os LLMs possam gerar histórias de usuário mais alinhadas às necessidades reais do projeto. O participante P1 destacou que *“ela [a ferramenta] precisa ter contexto e possivelmente estar linkada com a ferramenta de gestão das histórias”*, o que indica a percepção de que a integração com sistemas como Jira ou Trello pode facilitar o fornecimento automático de informações relevantes para a geração das histórias.

O participante P20 propõe uma abordagem concreta nesse sentido: *“fornecer, para treinamento da LLMs, todos os artefatos possíveis para contexto”*. Segundo o participante, essa prática tem sido adotada ativamente: *“tenho utilizado essa estratégia, não só para criar as estórias, mas também outros artefatos que compõem a documentação do projeto”*. P20 descreve ainda uma aplicação mais avançada, voltada ao suporte contínuo da equipe: *“a criação de um prompt, treinado com todos os artefatos do projeto, para que os desenvolvedores tirem dúvidas sobre requisitos”*. Essa estratégia sugere o uso de LLMs como um ponto de consulta contextualizado, contribuindo não apenas para a escrita das histórias, mas também para a compreensão contínua dos requisitos ao longo do desenvolvimento.

Quanto ao participante P21 reforça a importância do contexto ao sugerir a necessidade de *“treinar o modelo desde o início do projeto”*, evidenciando que o fornecimento contínuo de informações desde as etapas iniciais é essencial para que os LLMs apoiem adequadamente a execução das tarefas.

No entanto, esse cenário ideal encontra limitações importantes. Como aponta o participante P18, *“a IA recebe menos contexto”* devido às restrições impostas pela LGPD ao compartilhamento de dados sensíveis e estratégicos. Essa limitação *“impacta diretamente a qualidade e o alinhamento das histórias de usuário geradas”* e, segundo o participante, *“pode levar a resultados genéricos ou desalinhados com as necessidades reais do negócio”*. Além disso, exige um esforço adicional para *“anonimizar e mascarar informações sem comprometer o contexto”* e revela o quão crítico é esse equilíbrio, pois *“é preciso equilibrar a proteção dos dados com a clareza necessária para que a IA contribua de forma eficaz na aceleração do processo.”*

O participante P13 também aborda essa preocupação, observando que *“o uso*

de LLM pode ser bom, mas ao mesmo tempo pode expor o avanço tecnológico ou informações restritas para a ferramenta.” Ele reforça ainda a importância de “entender como proteger o tipo de informação para trazer mais segurança ao usuário de uma ferramenta com este objetivo, o que ajuda à adesão.”

Dessa forma, embora o fornecimento de contexto seja amplamente reconhecido como essencial para a eficácia dos LLMs, as exigências de proteção de dados representam um desafio significativo que precisa ser enfrentado para garantir a adoção segura e responsável dessas ferramentas em ambientes corporativos.

3.2.4 Discussões dos Resultados do Estudo

Esta seção apresenta a análise e interpretação dos dados coletados com relação as questões de pesquisa formuladas para este estudo. O objetivo é compreender como os profissionais da indústria lidam com atividades relacionadas à engenharia de requisitos e como percebem e utilizam LLMs em seus contextos profissionais. A discussão está organizada com base nas três perguntas norteadoras do estudo, explorando as práticas atuais, o uso de tecnologias emergentes e os impactos percebidos da automação no processo de especificação de requisitos.

3.2.4.1 **QP1: Como os profissionais da indústria lidam com requisitos em suas atividades cotidianas?**

Os dados coletados revelam como profissionais da indústria lidam com requisitos em suas rotinas, destacando práticas, decisões técnicas e adaptações conforme o contexto organizacional e o perfil dos envolvidos.

Os resultados indicam que a maioria dos participantes possui ampla vivência na área, o que reforça a confiabilidade das respostas, quanto às práticas e percepções relacionadas à engenharia de requisitos.

Os dados mostram que, ao todo, 65% dos profissionais têm mais de cinco anos de experiência na área de desenvolvimento de software. Essa significativa experiência

contribui para a robustez das informações coletadas sobre as práticas cotidianas e desafios enfrentados na engenharia de requisitos.

Além disso, com base nessa amostra, ao analisar o perfil das organizações dos participantes que se dispuseram a responder essa pesquisa, verificou-se que 85,7% dos respondentes atuam em empresas privadas. Essa distribuição indica que os resultados refletem majoritariamente práticas e percepções oriundas do setor privado, onde acredita-se que há maior incentivo para adoção de abordagens ágeis e uso de métodos modernos para gestão de requisitos.

Os dados revelam que a participação dos profissionais na elaboração de requisitos é significativa, 95,2% dos respondentes indicaram que participam frequentemente da elicitação de requisitos em seus projetos. Apenas um participante (4,8%) relatou envolvimento ocasional. Esse resultado reforça que a amostra é composta majoritariamente por profissionais fortemente engajados em atividades de engenharia de requisitos.

As principais técnicas utilizadas para elicitação de requisitos incluem entrevistas, reuniões com clientes e brainstorming, práticas que reforçam a importância da comunicação direta entre as partes interessadas. Mostrando que, embora as entrevistas se destaquem como a prática central, há uma diversidade de técnicas sendo adotadas, adaptadas ao contexto de cada projeto e equipe.

Em relação à validação dos requisitos, a maioria citou o uso de protótipos, e a revisões com stakeholders, demonstrando uma abordagem colaborativa e iterativa. Dessa maneira, evidencia-se que as equipes combinam a construção de protótipos com ciclos contínuos de revisões e testes práticos, promovendo uma validação iterativa e colaborativa, capaz de garantir o alinhamento dos requisitos com os objetivos do projeto.

A estruturação dos requisitos é realizada majoritariamente por meio de histórias de usuários (em 57,1% dos casos), mas também há uso complementar de casos de uso, documentos de especificação e protótipos.

Os dados apontam que profissionais com mais de 5 anos de experiência tendem a usar histórias de usuário com maior frequência, alcançando 100% na faixa “Mais de 5 anos”. Já as faixas mais experientes, “Mais de 10 anos” e “Mais de 20 anos”, apresentam

uma taxa semelhante de uso, em torno de 66%. Por outro lado, os menos experientes (1–5 anos) apresentam menor uso, entre 33% e 40%.

Esses dados reforçam que, embora as histórias de usuários estejam ganhando espaço, com 57,1% das equipes fazendo uso, há ainda uma parcela significativa que se apoia em abordagens mais tradicionais ou visuais, como casos de uso e diagramas, para documentar e compartilhar os requisitos ao longo do desenvolvimento.

Os profissionais também relataram desafios comuns na especificação dessas histórias, com destaque para a falta de tempo para elaborar histórias. Também foi possível identificar uma falta de padronização na definição de critérios de aceitação, a subjetividade na avaliação da qualidade. Tais desafios reforçam a complexidade inerente à prática da engenharia de requisitos, mesmo em ambientes que adotam abordagens ágeis.

Dentre os participantes cujas equipes utilizam histórias de usuário, observa-se uma valorização de critérios que refletem tanto a preocupação com a boa estruturação da história, como "Bem Formada" e "Atômica", quanto com sua aplicação prática no contexto do desenvolvimento ágil, como "Estimável" e "Conceitualmente Sólida". Esses critérios indicam que os participantes que usam histórias têm foco em escrever histórias claras, bem definidas, estruturadas, adequadas ao planejamento ágil e viáveis de serem planejadas e executadas pelas equipes.

Já entre os participantes que não utilizam histórias de usuário, os critérios mais citados, como "Não Ambíguo", "Livre de Conflitos", "Exclusiva" e "Independente", revelam uma preocupação voltada à consistência interna, à ausência de ambiguidade e à gestão de requisitos conflitantes ou redundantes. Isso sugere uma abordagem mais teórica ou preventiva, centrada na organização e controle dos requisitos, em vez do uso prático e contínuo de histórias no processo de desenvolvimento.

O critério "Não Ambíguo" foi valorizado tanto por quem usa quanto por quem não usa histórias de usuário, indicando que a clareza é uma preocupação comum. Por outro lado, os critérios "Exclusiva" e "Independente", típicos do método INVEST, são mais citados por quem não utiliza histórias, sugerindo que esses profissionais conhecem ou valorizam a teoria para evitar sobreposições e dependências.

Em síntese, os profissionais da indústria lidam com requisitos combinando técnicas clássicas de elicitação e validação com métodos ágeis de documentação e gestão. Suas práticas são moldadas por fatores como experiência, setor organizacional e acesso a tecnologias, resultando em uma dinâmica, onde abordagens estruturadas convivem com experimentações voltadas à agilidade e inovação.

3.2.4.2 **QP2: De que forma os profissionais utilizam Modelos de Linguagem de Grande Escala (LLMs) no contexto da engenharia de requisitos?**

A análise dos dados revela que aproximadamente 66,7% dos participantes já utilizam LLMs em suas rotinas profissionais. As ferramentas mais mencionadas incluem o ChatGPT e o GitHub Copilot, com referências pontuais a outros modelos, como DeepSeek e Gamma. Ainda assim, um terço dos respondentes relatou não utilizar essas ferramentas, sobretudo devido a preocupações com privacidade, segurança da informação e conformidade com a LGPD, fatores esses, especialmente relevantes em ambientes organizacionais com políticas institucionais mais restritivas.

Ao cruzar esses dados com a experiência profissional, observa-se que o uso de LLMs não se restringe a profissionais iniciantes. Pelo contrário, aqueles com mais de 10 anos de experiência compõem uma parcela significativa entre os usuários dessas ferramentas, o que evidencia uma abertura à inovação mesmo entre profissionais com trajetórias consolidadas na indústria. Por outro lado, profissionais com menos de 5 anos de atuação também figuram entre os usuários, sugerindo que o contato com LLMs pode ocorrer desde os estágios iniciais da carreira, especialmente em contextos mais inovadores.

A forma como os profissionais lidam com requisitos também varia conforme o setor de atuação. Quando questionados sobre o uso de LLMs em suas atividades profissionais, 87,5% pertencem a empresas privadas e 12,5% a empresas públicas. Essa proporção provavelmente reflete que ambientes privados tendem a apresentar maior abertura a abordagens ágeis e ao uso de tecnologias emergentes, como LLMs. Apenas

um participante vinculado a uma instituição pública (12,5%) descreveu o uso combinado de histórias de usuário e LLMs para a geração automatizada de requisitos, o que pode indicar uma adoção gradual de soluções inteligentes mesmo em contextos tradicionalmente mais conservadores.

Entre os profissionais que utilizam LLMs, os relatos indicam uma variedade de aplicações, como apoio à escrita técnica, brainstorming de ideias, geração de código e assistência em tarefas gerais de desenvolvimento. Embora apenas um participante tenha mencionado explicitamente o uso de LLMs na documentação de requisitos, esse tipo de aplicação ainda aparece de forma pontual e não sistematizada, o que sugere que a utilização direta de LLMs no domínio da engenharia de requisitos permanece em estágio inicial. O uso mais frequente citado foi o brainstorming, reforçando a percepção de que os profissionais têm explorado os LLMs principalmente como ferramentas de apoio criativo e suporte técnico, e não necessariamente como mecanismos voltados à formalização de requisitos.

3.2.4.3 RQ3: Qual é a percepção dos profissionais sobre a geração automatizada de requisitos, especialmente histórias de usuários, com o uso de LLMs?

Os resultados indicam uma percepção positiva sobre o uso de LLMs na geração de histórias de usuários. Observa-se que 75% dos participantes que utilizam LLMs em suas rotinas profissionais aplicam essas ferramentas para esse fim, o que pode evidenciar uma adoção significativa dessa tecnologia no contexto da engenharia de requisitos.

Os profissionais destacam benefícios claros na utilização das LLMs para essa finalidade, com ênfase na agilidade e no ganho de tempo durante a especificação das histórias, como apontados por P1 e P10.

Além disso, as LLMs são valorizadas como ferramentas de apoio à criação e ao brainstorming, contribuindo para a inspiração e para a estruturação das histórias, conforme relatado pelos participantes P18 e P10.

Outro impacto positivo relevante é a padronização proporcionada pelas LLMs,

que ajudam a garantir consistência e clareza nas histórias geradas, como destacado pelos participantes P10, P16 e P18. O suporte à qualidade das histórias, por meio da validação de critérios, também é mencionado como um benefício adicional.

Por outro lado, os profissionais reconhecem limitações importantes no uso dessas tecnologias, especialmente a dificuldade das LLMs em capturar o contexto específico dos projetos, o que pode levar à geração de histórias desalinhadas. Essa preocupação é evidenciada nos relatos dos participantes P10 e P18, que ressaltam a necessidade de refinar e ajustar as histórias para garantir seu alinhamento com o negócio. Além disso, riscos relacionados à dependência excessiva das ferramentas foram apontados, incluindo a possibilidade de redução do esforço analítico por parte dos profissionais, conforme alertas dos participantes P16 e P19.

Em síntese, a percepção geral é de que as LLMs oferecem um importante potencial para acelerar e apoiar a geração automatizada de histórias de usuários, mas seu uso deve ser acompanhado de práticas críticas de revisão e contextualização para assegurar a qualidade e a adequação dos requisitos gerados.

A percepção dos participantes que ainda não utiliza LLMs para esse fim, não expressa rejeição à tecnologia, mas revela entraves práticos, como falta de tempo, ausência de participação da equipe na escrita e dificuldade em elaborar prompts eficazes. Esses relatos sugerem que a adoção pode ser facilitada por técnicas prontas e validadas, que reduzam o esforço inicial e aumentem a confiança dos profissionais no processo automatizado.

De modo geral, a geração automatizada de histórias de usuários com LLMs é percebida como uma possibilidade, especialmente quando utilizada como ferramenta de apoio, e não como substituto integral da análise humana. O uso crítico e contextualizado é apontado como fator essencial para garantir que os benefícios superem os riscos.

Além disso, ao analisarmos as sugestões e opiniões dos respondentes para a última pergunta, aberta, observamos uma ênfase recorrente na importância do fornecimento de contexto como fator essencial para o bom desempenho dos LLMs na geração de histórias de usuário. Esse fornecimento pode ser automatizado, abrangente, incluindo todos os artefatos do projeto, e iniciado desde o começo do desenvolvimento.

Tal prática é vista como crucial para garantir que os LLMs produzam histórias mais relevantes, claras e alinhadas às necessidades reais do projeto.

No entanto, embora o fornecimento de contexto seja amplamente reconhecido como essencial para a eficácia dessas ferramentas, as exigências de proteção de dados, especialmente em ambientes corporativos privados, representam um desafio significativo. As restrições legais, como as impostas pela LGPD, limitam o compartilhamento de informações sensíveis, exigindo esforços adicionais para anonimizar ou mascarar dados sem comprometer o contexto necessário.

Esses relatos sugerem que, para viabilizar a adoção segura e eficaz de LLMs no contexto corporativo, é necessário ir além da capacidade técnica das ferramentas, estabelecendo estratégias claras para garantir a privacidade e a segurança da informação. A preocupação com a exposição de dados sensíveis atravessa diferentes níveis de experiência e reforça a importância do desenvolvimento de soluções que simplifiquem esse processo, permitindo um equilíbrio entre proteção e desempenho.

Essa percepção pode revelar uma barreira à adoção: a desconfiança em relação à segurança dos dados ainda se mostra um obstáculo, mesmo entre aqueles que reconhecem o potencial da tecnologia, especialmente em contextos corporativos que lidam com informações sigilosas ou reguladas por normas como a LGPD.

Para mitigar esses riscos, uma solução viável é o uso de prompt engineering com dados mínimos, priorizando a identificação do contexto essencial necessário para a tarefa. Isso inclui apenas informações como o tipo de usuário, os objetivos e o cenário geral do sistema, evitando nomes, identificadores, dados financeiros ou pessoais.

Outra abordagem complementar é o uso de LLMs locais ou privados, que, ao serem executados em ambientes controlados (on-premises), asseguram que os dados não sejam transmitidos a serviços externos, ampliando o controle e a segurança sobre a informação utilizada nos processos automatizados de geração de requisitos.

3.3 Ameaças à Validade

Este estudo enfrentou ameaças à validade comuns em pesquisas survey ([RALPH et al., 2020](#)).

Validade de constructo: Termos como “qualidade da história de usuário” podem ter sido interpretados de maneira diferente pelos participantes. Para reduzir ambiguidades, um piloto foi conduzido, permitindo ajustes antes da aplicação principal do instrumento.

Validade interna: Por se tratar de um estudo descritivo, não foram exploradas relações causais. Exemplos podem ter influenciado as respostas, e as respostas abertas estão sujeitas a viés de interpretação dos participantes.

Validade externa: A amostra (21 participantes) foi recrutada por conveniência, o que limita a generalização dos resultados. Apenas profissionais com experiência autodeclarada em engenharia de requisitos foram incluídos.

Confiabilidade: Nenhum teste estatístico de confiabilidade foi realizado, mas as perguntas foram padronizadas. Os dados ausentes foram mínimos e tratados individualmente por pergunta.

Objetividade: Os procedimentos de coleta e análise de dados foram documentados e aplicados de forma consistente para reduzir o viés do pesquisador.

3.4 CONCLUSÃO

Este capítulo apresentou um survey realizado com profissionais da indústria com experiência em engenharia de requisitos. O objetivo foi investigar como esses profissionais lidam com requisitos no dia a dia, especialmente em atividades de elicitação e especificação, e como percebem o uso de LLMs em suas práticas, com ênfase no interesse e na viabilidade da geração automatizada de histórias de usuários.

O questionário foi estruturado em quatro seções interdependentes e abordou aspectos contextuais, práticos e perceptivos. Foi aplicado de forma online para ampliar o alcance e garantir a diversidade dos participantes. Essa abordagem possibilitou a coleta de dados relevantes para responder às questões de pesquisa propostas e fornecer

uma visão representativa das práticas atuais na indústria.

Participaram 21 profissionais com experiência significativa, majoritariamente de Manaus, que relataram combinar diversas técnicas de elicitação, como entrevistas, análise documental, brainstorming e Lean Inception. Na validação dos requisitos, destacam-se práticas como prototipagem, revisões com stakeholders e testes em ambientes controlados. No que diz respeito à estruturação dos requisitos, o uso de histórias de usuários é relativamente comum, com 57,1% dos respondentes adotando essa abordagem, o que indica uma tendência de consolidação no mercado, embora haja espaço para maior disseminação.

Os profissionais que utilizam histórias de usuário valorizam critérios que aliam boa estrutura textual e aplicabilidade prática, como “Bem Formada”, “Atômica”, “Estimável” e “Conceitualmente Sólida”. Esses critérios indicam que os participantes que usam histórias têm foco em escrever histórias claras, bem definidas, estruturadas, adequadas ao planejamento ágil e viáveis para serem planejadas e executadas pelas equipes. Por outro lado, aqueles que não utilizam essa técnica priorizam critérios como “Não Ambíguo”, “Exclusiva” e “Independente”, sugerindo uma abordagem mais voltada à consistência formal e à prevenção de conflitos e sobreposições entre requisitos.

Em relação ao uso de Modelos de Linguagem de Grande Escala (LLMs), observou-se que aproximadamente dois terços (66,7%) dos participantes já integram essas ferramentas em suas rotinas, sendo o ChatGPT o modelo mais citado. O uso não está restrito a profissionais iniciantes, evidenciando uma aceitação transversal entre diferentes níveis de experiência. Apesar da adoção, o uso direto das LLMs em atividades profissionais, ainda é incipiente, com poucos relatos de aplicação sistemática na documentação, sendo mais comuns as utilizações para apoio criativo, brainstorming e auxílio em tarefas técnicas. Os que ainda não adotam LLMs justificam principalmente por preocupações com segurança e privacidade, em especial relacionadas à LGPD e às políticas internas das organizações. Além do uso profissional, o emprego de LLMs em outras atividades é pouco frequente, dentre os participantes.

Entre os participantes que utilizam de LLMs para geração de histórias de usuários (75%), destacam-se benefícios como agilidade na escrita, inspiração, padronização e

suporte na revisão da qualidade das histórias. O participante P10 exemplificou: “Agilidade na escrita, Consistência e Validação de Critérios.” Contudo, limitações apontadas incluem falta de contexto, que pode gerar histórias genéricas, necessidade de refinamento, riscos de dependência excessiva e questões de licenciamento. A revisão crítica e o alinhamento com o contexto do projeto são essenciais para mitigar esses desafios. Os profissionais que não utilizam LLMs citam a falta de tempo e a ausência de técnicas prontas como barreiras, mas demonstram interesse em adotá-los mediante o desenvolvimento de modelos de prompt eficazes.

As sugestões dos participantes enfatizam a importância do fornecimento de contexto adequado às LLMs, além do treinamento da IA com todos os artefatos do projeto para melhorar a precisão e utilidade das histórias geradas. A proteção de dados, principalmente pela LGPD, limita o compartilhamento de informações sensíveis, impactando a qualidade das histórias. Assim, é fundamental equilibrar segurança e clareza, visando a viabilidade do uso.

Essas questões reforçam a necessidade de desenvolver diretrizes e estratégias para o uso das LLMs em ambientes corporativos, configurando uma importante oportunidade para estudos futuros. Além disso, a pesquisa contribui com uma visão local sobre as práticas de engenharia de requisitos e o uso emergente de LLMs, abrindo espaço para investigações comparativas e aprofundadas sobre métodos que viabilizam a geração automatizada de histórias de usuários, também em diferentes contextos.

No contexto do Ciclo de Relevância, a análise dos dados coletados neste survey revelou práticas consolidadas, desafios recorrentes e percepções diversificadas acerca da engenharia de requisitos na indústria, especialmente no que tange ao uso de histórias de usuários. Esse cenário evidencia tanto lacunas quanto oportunidades para explorar soluções baseadas em LLMs, como o ChatGPT, no apoio à geração automatizada dessas histórias.

Considerando benefícios como agilidade, padronização e ênfase na qualidade, o próximo capítulo apresenta, inicialmente, um estudo exploratório com prompting livre, voltado à compreensão aprofundada do problema e do contexto. Em seguida, com base nas limitações e aprendizados obtidos, tem início a primeira iteração do segundo ciclo

(Design), no qual se propõe uma abordagem estruturada, fundamentada em técnicas de prompting descritas na literatura, capaz de orientar o ChatGPT na criação de histórias pautadas em critérios de qualidade reconhecidos. Essa investigação busca compreender o impacto da estratégia sobre a qualidade das histórias geradas, comparando-as com aquelas produzidas manualmente.

4

1º CICLO DE DESIGN: HISTÓRIAS DE USUÁRIOS: O CHATGPT FAZ MELHOR?

Este capítulo apresenta dois estudos experimentais realizados para investigar se as histórias de usuários geradas pelo ChatGPT, utilizando a abordagem US-Prompt, exibem qualidade superior em comparação às histórias escritas manualmente ou produzidas por prompts de formato livre. O capítulo explora e avalia as diferenças em termos de qualidade, contribuindo para o avanço das técnicas de geração automatizada no contexto da Engenharia de Requisitos. Os estudos foram publicados no artigo internacional “**User Stories: Does ChatGPT Do It Better?**”, aceito na **Proceedings of the 27th International Conference on Enterprise Information Systems - Volume 2: ICEIS (2025)**.

4.1 INTRODUÇÃO

Considerando o que foi apresentado no capítulo anterior, o problema central dessa pesquisa foi explorar se as histórias de usuários geradas pelo ChatGPT, utilizando técnicas de prompting, alcançam uma qualidade superior em comparação às criadas por humanos. Essa comparação foi realizada com base no framework Quality User Story (QUS) ([LUCASSEN et al., 2016b](#)), que avalia critérios-chave como atomicidade, minimalidade e solidez. Ao abordar esse problema, a pesquisa buscou fornecer insights sobre o potencial da IA para complementar ou, até mesmo, superar os esforços humanos

na geração de histórias de usuários.

A pergunta de pesquisa é: **De que forma a estratégia de prompting influencia a qualidade das histórias de usuários geradas por IA em comparação às escritas manualmente?**

Esta pesquisa foi dividida em dois estudos: o primeiro comparou histórias de usuários criadas manualmente com aquelas geradas pelo ChatGPT, utilizando prompt de formato livre. O segundo estudo comparou o prompt de formato livre com o Meta-Few-Shot Prompt, um prompt estruturado composto por técnicas de prompting projetadas para automatizar a geração de histórias de usuários de maior qualidade. A partir deste momento, o prompt para a geração automatizada utilizado nesta pesquisa, será chamado de User Story Prompt (US-Prompt).

4.2 MÉTODO

Esta pesquisa analisou se histórias de usuários geradas utilizando técnicas de prompting, com o ChatGPT, atingem uma qualidade superior em comparação com aquelas criadas manualmente, por humanos. Para isso, o estudo avaliou a capacidade de cada abordagem em produzir histórias de usuários que melhor aderem aos critérios de qualidade QUS, ou seja, identifica qual técnica demonstra maior eficácia na geração de histórias de qualidade, sejam elas criadas manualmente ou assistidas pelo ChatGPT.

Para avaliar a eficácia das técnicas, foram seguidos os mesmos critérios de qualidade utilizados no estudo de [Ronanki, Cabrero-Daniel e Berger \(2024\)](#), baseado no framework QUS, proposto por [Lucassen et al. \(2016a\)](#), que oferecem uma base sólida para a comparação dos métodos, enfatizando a qualidade e a conformidade das histórias de usuários com requisitos bem definidos.

Como os critérios aplicáveis a histórias individuais permitem identificar problemas específicos em cada história, como falta de clareza ou ambiguidades, e facilitam a obtenção de métricas detalhadas, como a porcentagem de histórias que atendem a cada critério, é provável que estes tenham sido os critérios utilizados no estudo conduzido por [Ronanki, Cabrero-Daniel e Berger \(2024\)](#). Dado que seu objetivo era determinar

a eficácia do ChatGPT na avaliação de histórias de usuários, a utilização de critérios individuais possibilitou uma análise mais precisa e granular, avaliando cada história de forma isolada antes de considerar o conjunto completo da especificação. Por esse motivo, optou-se por manter os critérios, garantindo a mesma consistência na avaliação de cada história de usuário.

A Tabela 2 apresenta os critérios, com descrições detalhadas para garantir maior clareza e precisão durante a avaliação.

Tabela 2 – Descrição dos critérios derivados do framework Quality User Story (QUS).

Qualidade	Critério	Descrição
Sintática	Bem-formada	Se a história inclui pelo menos um papel (persona), que define quem está solicitando a funcionalidade e um meio (ação), que é a atividade ou funcionalidade específica que o sistema deve permitir.
	Atomicidade	Se a história aborda uma única funcionalidade específica.
	Minimalidade	Se a história contém apenas um papel, um meio (ex: “quero abrir o mapa”), e um ou mais fins, que são os objetivos que o usuário pretende alcançar usando a funcionalidade (ex: “para localizar facilmente destinos turísticos”), sem informações desnecessárias.
Semântica	Solidez Conceitual	Se o meio expressa uma funcionalidade, e os fins explicam a razão dessa funcionalidade.
	Não Ambiguidade	Se ela evita termos ou abstrações que possam gerar múltiplas interpretações.
Pragmática	Frase Completa	Se a história é completa, bem formada e fornece contexto suficiente para ser claramente compreendida.
	Estimabilidade	Se a história não denota um requisito de granulação grossa que seja difícil de planejar e priorizar.

Cada história de usuário foi avaliada pelos pesquisadores com base nos critérios do QUS, utilizando um sistema de pontuação binária, onde uma pontuação de 1 indicava satisfação e 0 indicava não conformidade, determinando se uma história de usuário atendia ou não ao critério de qualidade avaliado, em alinhamento com o objetivo do estudo de medir a aderência estrita aos padrões do QUS (LUCASSEN et al., 2016b). A pontuação total de efetividade para um conjunto de histórias de usuários foi calculada como uma taxa de sucesso por meio de três etapas principais. Primeiro, somou-se o número total de critérios atendidos em todas as histórias avaliadas. Em seguida, determinou-se o total de sucessos possíveis, multiplicando o número de histórias

avaliadas (N) pelo número de critérios de qualidade (C , que é 7). Por fim, o total de critérios atendidos foi dividido pelo total de sucessos possíveis e multiplicado por 100 para obter a taxa de sucesso. Essa taxa representou a aderência geral das histórias de usuários aos critérios de qualidade do QUS, refletindo a efetividade da técnica.

$$\text{Taxa de Sucesso (\%)} = \frac{\sum_{i=1}^N \text{Critérios Atendidos pela História}_i}{N \times C} \times 100$$

Descrição das Variáveis:

- N : Número total de histórias de usuários avaliadas.
- C : Número total de critérios de qualidade avaliados.
- Critérios Atendidos pela História _{i} : Número de critérios atendidos pela i -ésima história de usuário.

Por exemplo, o grupo A gerou manualmente 11 histórias, resultando em 77 possíveis sucessos (11 histórias $\times$ 7 critérios). Dessas, 52 critérios foram atendidos, resultando em uma taxa de sucesso de 67,53%.

Da mesma forma, as taxas de sucesso foram calculadas para as histórias geradas utilizando o prompt livre e o US-Prompt, facilitando uma comparação direta entre as técnicas. Essa abordagem permitiu a representação dos dados em forma de porcentagens de sucesso, fornecendo uma base quantitativa e objetiva para comparar a qualidade das histórias de usuários, produzidas pelas diferentes técnicas.

Nesta pesquisa, o foco foi avaliar a qualidade das histórias geradas com base nos critérios do QUS, sem a formalização adicional de critérios de aceitação ou a definição de "pronto", o que permitiu concentrar-se na análise das histórias de usuários em termos de suas características estruturais e qualitativas.

As três etapas metodológicas estão descritas abaixo, consistindo em dois estudos comparativos e no desenvolvimento de uma nova abordagem, conforme mostrado na Figura 6. A pesquisa seguiu uma abordagem sistemática, organizada em três fases principais:

1. **Estudo Experimental 1 (Seção 4.3):** Este estudo foca na comparação da qualidade das histórias de usuários geradas manualmente versus aquelas geradas de forma

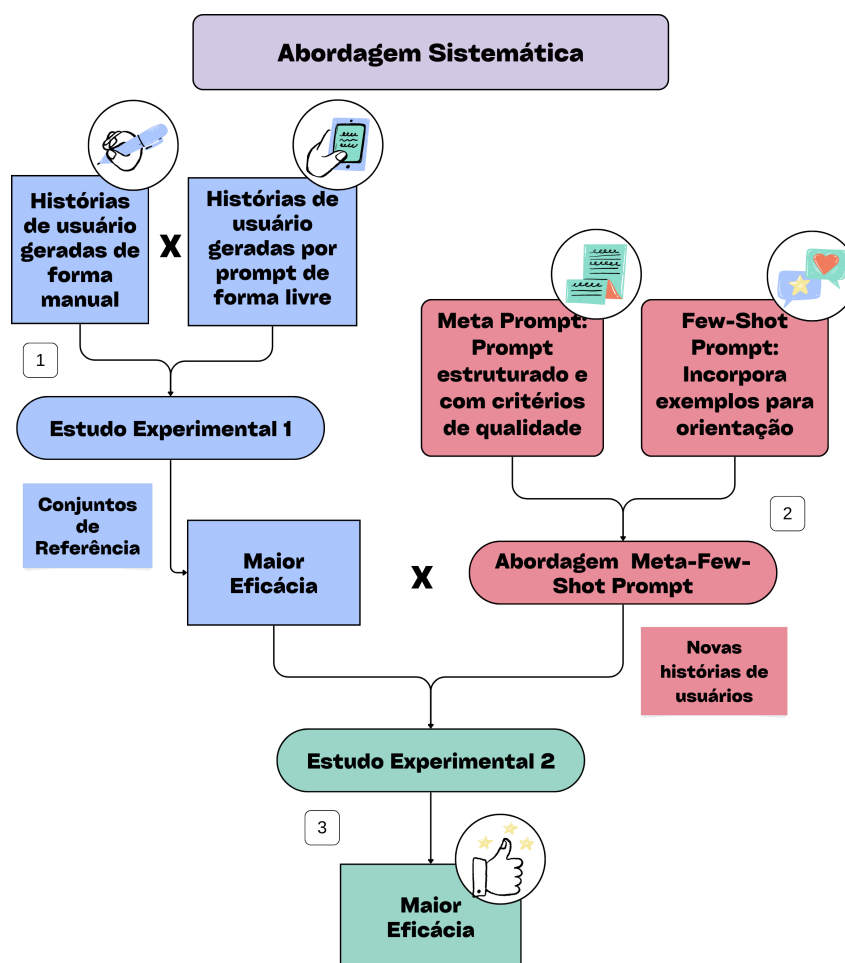


Figura 6 – Diagrama das etapas metodológicas da pesquisa.

automatizada utilizando o ChatGPT, para determinar qual abordagem foi mais eficaz. Dois conjuntos de histórias de usuários foram criados e analisados, sendo utilizados como referência para comparações posteriores.

2. **Design do US-Prompt (Seção 4.4):** Esta fase envolveu a análise de padrões e deficiências nas histórias de usuários geradas durante o Estudo Experimental 1, o que levou ao desenvolvimento de um prompt estruturado para guiar o ChatGPT a melhorar a precisão e a qualidade das histórias de usuários. O método de US-Prompt foi aplicado ao ChatGPT para gerar um novo conjunto de histórias de usuários. Este método combina:

- Meta Prompt: Utilização de prompts estruturados com critérios específicos de qualidade (BROWN et al., 2020; ZHANG; YUAN; YAO, 2023).

- **Few-Shot Prompt:** Fornecimento de exemplos para guiar as respostas do modelo (BROWN et al., 2020; ZHANG; YUAN; YAO, 2023).
3. **Estudo Experimental 2 (Section 4.5):** Neste estudo, um novo conjunto de histórias de usuários foi gerado pelo ChatGPT, utilizando o método recém-desenvolvido de US-Prompt. Este novo conjunto foi avaliado e comparado aos conjuntos de referência do Estudo Experimental 1. Essa comparação permitiu uma análise da qualidade e da eficácia do método US-Prompt.

4.3 ESTUDO EXPERIMENTAL 1: COMPARAÇÃO ENTRE GERAÇÃO MANUAL E PROMPT LIVRE

Este estudo foi projetado para simular cenários do mundo real enfrentados por engenheiros de software iniciantes, garantindo condições consistentes de coleta de dados entre todos os grupos de participantes.

4.3.1 Desenho do Estudo

O objetivo deste estudo foi comparar a qualidade das histórias de usuários geradas manualmente com aquelas produzidas pelo ChatGPT, com ênfase no framework QUS. Especificamente, o estudo avaliou a eficácia de cada abordagem, medida pela taxa de sucesso na geração de histórias de usuários que atendam aos padrões de qualidade definidos pelo QUS.

A pesquisa buscou determinar se o uso de LLMs pode aprimorar a geração de histórias de usuários, no contexto da engenharia de requisitos. Para isso, o estudo explorou a seguinte questão de pesquisa (QP): **Como a qualidade das histórias geradas pelo ChatGPT com prompt de formato livre se compara às escritas manualmente?**

Hipóteses: O estudo foi desenhado com as seguintes hipóteses:

- **Hipótese Nula (H01):** Não há diferença significativa na qualidade das histórias de usuários geradas manualmente em comparação com aquelas produzidas utili-

zando prompts livres no ChatGPT.

- **Hipótese Alternativa (HA1):** Há uma diferença significativa na qualidade das histórias de usuários geradas manualmente em comparação com aquelas produzidas utilizando prompts livres no ChatGPT.

Variáveis: A variável independente neste estudo foi o método de geração das histórias de usuários, que inclui dois tratamentos: geração manual (tratamento 1) e geração automatizada com o ChatGPT (tratamento 2). A variável dependente foi a qualidade das histórias de usuários, avaliada com base na aderência aos critérios do framework QUS.

Seleção dos participantes:

Os participantes foram recrutados por amostragem por conveniência a partir da disciplina de Engenharia de Requisitos e Análise de Sistemas (ERAS) ofertada no 5º período do curso de Engenharia de Software da Universidade Federal do Amazonas (UFAM). Por se tratar de uma disciplina intermediária, todos os alunos já possuíam familiaridade com conceitos introdutórios de engenharia de software, entretanto, a engenharia de requisitos ainda era uma área totalmente nova para a maioria. Mas, durante as aulas da disciplina, os participantes receberam treinamento sobre as diretrizes do framework QUS para a criação de boas histórias de usuários.

Desenho Experimental: Todos os participantes assinaram um termo de consentimento livre e esclarecido, que garantiu o direito de se retirar do estudo a qualquer momento sem penalidades. Em seguida, os participantes foram instruídos a formar grupos e especificar histórias de usuários para um Sistema de Agência de Viagens utilizando duas abordagens distintas: geração manual (sem assistência tecnológica) e geração automatizada com prompts livres no ChatGPT. Os dados coletados foram anonimizados para garantir sua confidencialidade.

4.3.2 Execução do Estudo

A execução do estudo foi dividida em duas etapas:

- **Fase 1 - Geração Manual:** Os grupos geraram histórias de usuários manualmente, no período de 16 a 21 de maio, aplicando os critérios aprendidos durante as aulas. Cada grupo submeteu suas histórias por meio de uma plataforma digital fornecida;
- **Fase 2 - Geração Automatizada com Prompt Livre:** Os grupos foram orientados a utilizar o ChatGPT (versão 3.5 ou superior) para gerar novas histórias de usuários, no período de 27 a 29 de maio. Os participantes empregaram diferentes estratégias de prompting: alguns forneceram apenas o enunciado básico, enquanto outros realizaram consultas adicionais ou forneceram mais contexto antes de solicitar a geração das histórias, mas sem utilizar exemplos ou diretrizes específicas de uma estrutura definida e detalhada, como por exemplo, apenas: “Gere uma história de usuário.” A intenção era simular a liberdade de um desenvolvedor para criar histórias de usuários de forma autônoma. As histórias geradas e os prompts usados foram, então, submetidos digitalmente para análise;

Inicialmente, doze grupos de participantes foram formados, mas apenas sete, compostos por 30 participantes, completaram ambas as fases do estudo.

Os dados coletados incluíram 14 conjuntos de histórias de usuários dos sete grupos. Cada grupo forneceu um conjunto de histórias geradas manualmente e outro gerado com prompt livre, totalizando 126 histórias (62 manuais e 64 automatizadas). Apesar das variações no comprimento das histórias, todas mantiveram níveis semelhantes de complexidade, uma vez que todos os grupos basearam seu trabalho no mesmo enunciado fornecido durante o estudo.

4.3.3 Resultados e Discussão

Os dados foram compilados a partir das avaliações das histórias geradas manualmente e automaticamente pelos grupos de participantes. A análise incluiu o cálculo da taxa de sucesso para cada critério de qualidade atendido. As histórias de cada grupo, geradas manualmente e criadas utilizando prompt livre no ChatGPT, foram avaliadas com base nos critérios de qualidade.

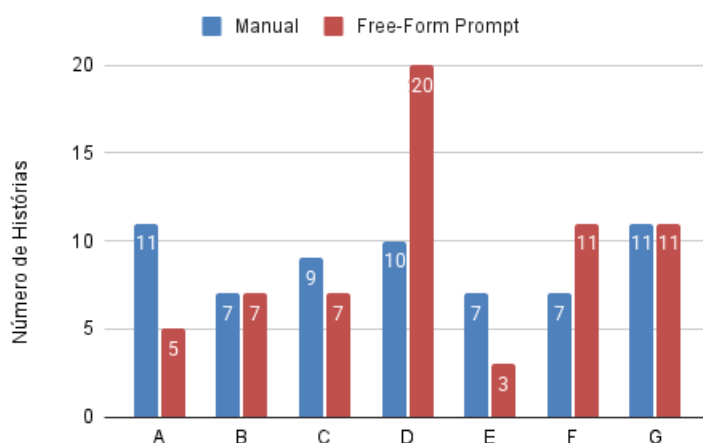


Figura 7 – Comparação entre o número de histórias de usuários geradas por métodos manuais e Free-Form Prompt.

A Figura 7 ilustra o número de histórias de usuários geradas pelos sete grupos participantes do estudo, comparando o desempenho do método manual com o uso do ChatGPT utilizando prompt livre. Entre os grupos analisados, quatro (B, D, F e G) geraram um número de histórias maior ou igual ao das histórias produzidas manualmente ao empregar a ferramenta automatizada. Esses resultados sugerem que o ChatGPT ajudou a aumentar o número de histórias de usuários geradas. Embora alguns grupos tenham enfrentado desafios específicos, a maioria obteve um volume de produção maior com a automação.

Em sequência, foi calculado o percentual de sucesso para cada grupo. A Figura 8 apresenta as diferenças nos percentuais de respostas corretas para cada critério de qualidade das histórias geradas, comparando os métodos manual e automatizado, com ChatGPT, utilizando prompt livre. As barras representam o percentual de respostas corretas para cada grupo e método, com variações de desempenho, destacadas em verde.

De modo geral, observou-se uma melhoria no desempenho em três dos sete grupos ao utilizar a geração automatizada, com os grupos E e F alcançando melhorias superiores a 5%, com taxas de 6,12% e 8,91%, respectivamente. Valores positivos indicam que o ChatGPT superou o método manual na geração de histórias de usuários, sugerindo maior eficácia na aplicação de critérios de qualidade como consistência e ato-

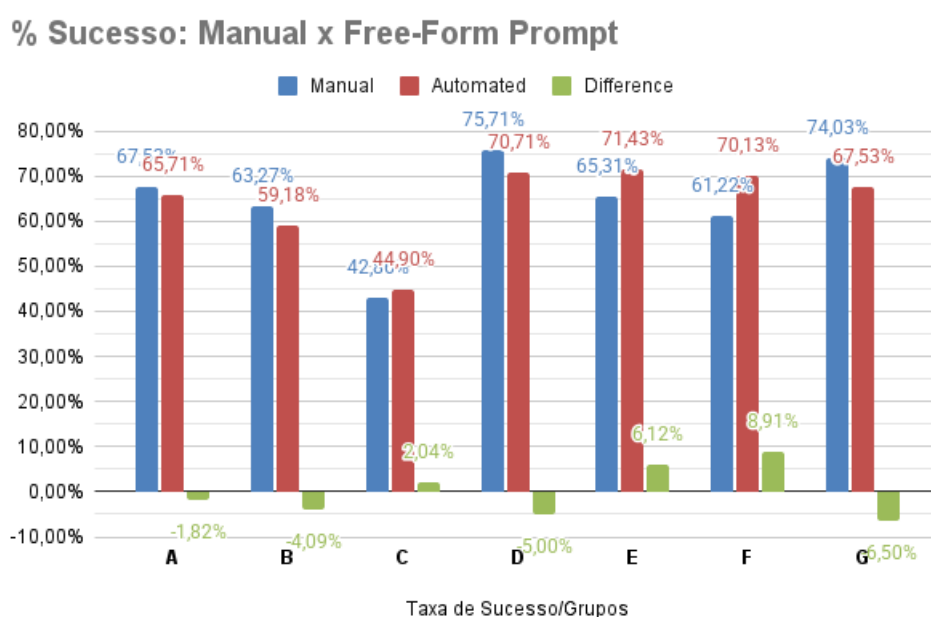


Figura 8 – Comparação de taxas de sucesso entre os métodos Manual e Prompt de Forma Livre.

micidade. Por outro lado, valores negativos, como os observados no grupo A, indicam que o método manual foi mais eficaz em alguns casos, o que pode estar relacionado à experiência dos participantes ou à qualidade dos prompts fornecidos.

Variações negativas em grupos como D e G, com valores de 5,00% e 6,50%, respectivamente, podem sugerir que, embora o ChatGPT tenha aumentado o número de histórias de usuários geradas para alguns grupos, a geração automatizada não atendeu às expectativas de qualidade. Essa variação pode ser atribuída a fatores específicos que influenciaram seu desempenho, como a complexidade das histórias geradas ou diferenças na forma como os grupos interagiram com a ferramenta, afetando, por fim, os resultados finais. Em particular, a ausência de instruções claras e específicas nos prompts pode ter comprometido a qualidade das respostas. Alguns prompts forneceram apenas a declaração sem diretrizes adicionais, o que pode ter levado a uma geração menos eficaz de histórias de usuários.

Os dados quantitativos foram analisados para avaliar a eficácia dos métodos de geração de histórias de usuários, utilizando análise estatística realizada com um software gratuito e de código aberto JASP (Just Another Statistical Program), versão

0.19.1¹. A eficácia média das histórias de usuários geradas manualmente foi calculada em 64,28% (arredondado) e comparada à eficácia das histórias geradas automaticamente com prompt livre, que teve uma média de 64,23% (arredondado).

Tabela 3 – Resultados estatísticos da comparação da qualidade das histórias de usuários geradas manualmente e com prompt livre

	Manual	Free Form
Mediana	65.310	67.530
Média	64.276	64.227
Desvio Padrão	10.852	9.494

Embora a média das histórias geradas manualmente tenha sido ligeiramente superior, o teste estatístico não encontrou evidências suficientes para afirmar que a qualidade das histórias geradas pelos métodos manuais foi significativamente diferente daquelas geradas automaticamente. Assim, pode-se concluir que ambos os métodos produzem resultados comparáveis em termos de qualidade das histórias de usuário.

O teste de Shapiro-Wilk foi utilizado para avaliar a normalidade dos dados, um pré-requisito para a aplicação de testes paramétricos, como o teste T, a fim de garantir resultados estatisticamente válidos.

Neste estudo, o p-value obtido para o teste Shapiro foi de ($p = 0.463$). Como o valor de (p) foi maior do que o nível de significância estabelecido de ($\alpha = 0.05$), pode-se concluir que os dados analisados apresentaram uma distribuição normal. Isso justificou o uso do teste T de Student.

O teste T de Student obteve um valor de ($p = 0.983$) sugerindo que não houve diferença estatisticamente significativa entre os dois métodos. Assim, a hipótese nula (H_0) não foi rejeitada, indicando que ambos os métodos geraram histórias de usuário de qualidade semelhante.

Além disso, os desvios-padrão foram de 10.9% para as histórias manuais e 9.5% para as histórias automatizadas. O menor desvio-padrão no método automatizado sugere maior homogeneidade nas respostas geradas com o prompt livre. Essa homogeneidade significa que, ao utilizar técnicas automatizadas, as histórias tendem a ser mais

¹ Disponível em: <<https://jasp-stats.org/>>

consistentes em termos de qualidade e formato, o que pode ser vantajoso em contextos onde a uniformidade é desejada.

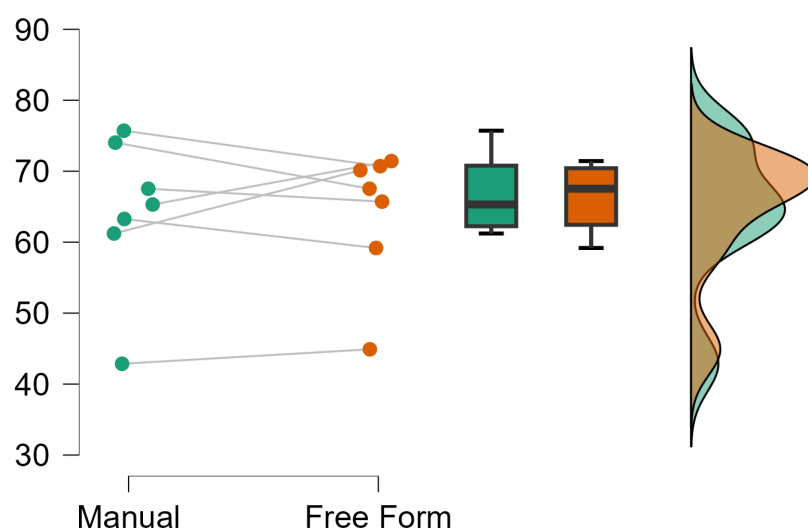


Figura 9 – Gráfico apresentando indicadores por meio de representações visuais: gráficos de densidade, boxplots e pontos de dados individuais.

O gráfico na Figura 9 ajuda a identificar potenciais diferenças entre as médias e a variabilidade dos grupos. As curvas de densidade ilustram a distribuição dos dados, destacando a concentração de valores próximos a 70%, com distribuições sobrepostas, sugerindo certa semelhança entre os métodos. No entanto, as diferenças nos picos de densidade indicam variações nas distribuições das pontuações, que, juntamente com os resultados do teste de normalidade de Shapiro, sugerem que a distribuição é, de fato, aproximadamente normal. A distribuição gerada pelo método de prompt livre exibiu uma forma mais simétrica e bem definida, com um pico claro, o que indica uma maior consistência na qualidade das histórias geradas. Em contraste, a distribuição do método manual tem uma forma mais achatada, sem um pico pronunciado, o que sugere uma maior dispersão e maior variabilidade na qualidade das histórias geradas.

Os boxplots resumem a distribuição central, enfatizando a mediana e os quartis. Embora as distribuições dos dados para ambos os métodos tenham sido em grande parte semelhantes, existiram pequenas diferenças nas pontuações mais altas e mais baixas (pontuações extremas) e na maneira como os dados se espalharam (variação

geral). De acordo com o gráfico, o prompting livre teve uma mediana de 67,53%, em comparação com 65,31% para a geração manual, o que pode indicar uma vantagem de eficácia para o método automatizado com prompting livre.

Os pontos de dados individuais, ou "jitter", mostram cada par de dados de forma distinta, proporcionando uma visão clara da dispersão geral e facilitando a identificação de padrões descendentes nos dados.

Em resumo, embora a eficácia média geral não tenha demonstrado uma melhoria substancial em relação ao método manual, a maior homogeneidade das histórias geradas automaticamente sugere que as técnicas automatizadas podem ajudar a aumentar o número de histórias de usuário geradas e contribuir para uma geração mais uniforme e padronizada de histórias de usuário. Este resultado foi consistente com as descobertas de [Krishna et al. \(2024a\)](#), que também identificaram uma diferença marginal na qualidade entre documentos gerados automaticamente e manualmente. Em seu estudo, Krishna explorou a capacidade dos LLMs de produzir rascunhos coerentes e estruturados de documentos de Especificação de Requisitos de Software (SRS), seus resultados sugeriram que os LLMs podem igualar a qualidade de saída de um engenheiro de software iniciante para gerar um SRS, entregando rascunhos completos e consistentes.

Concluindo, os resultados deste estudo se alinharam com as conclusões de [Krishna et al. \(2024a\)](#), demonstrando que as técnicas automatizadas podem atingir uma qualidade semelhante aos métodos manuais, ao mesmo tempo em que garantem maior uniformidade e padronização na geração das histórias de usuário.

4.4 Abordagem US-Prompt (Versão 1)

Os resultados do estudo empírico na Seção 4.3 mostraram que não houve diferença estatisticamente significativa entre as histórias de usuários criadas pelos participantes, tanto manualmente, quanto geradas automaticamente com prompts de formato livre. Isso destacou a necessidade de melhorar a abordagem de design de prompts para produzir histórias de usuários de maior qualidade, levando ao desenvolvimento da abordagem mais estruturada: o US-Prompt.

Essa abordagem combinou as técnicas Meta-Prompting ([BROWN et al., 2020](#); [ZHANG; YUAN; YAO, 2023](#)) e Few-Shot Prompting ([BROWN et al., 2020](#); [ZHANG; YUAN; YAO, 2023](#)) para criar uma técnica especializada de engenharia de prompts voltada para aprimorar a precisão e a qualidade das histórias de usuários geradas por LLMs. Aproveitando a orientação estruturada do Meta-Prompt junto às capacidades adaptativas do Few-Shot Prompt, o US-Prompt permite que o modelo siga critérios específicos por meio de prompts que incorporam tanto instruções explícitas quanto exemplos ilustrativos.

Com o uso do Meta-Prompting, o modelo recebe prompts contendo critérios de qualidade do framework QUS de [Lucassen et al. \(2016a\)](#), com instruções projetadas para atender a padrões como não-ambiguidade, atonicidade e estrutura bem-formada. Simultaneamente, o Few-Shot Prompt fornece um conjunto de exemplos bem definidos de histórias de usuários, tanto corretos quanto incorretos, o que permite ao modelo identificar e replicar os padrões desejados com maior consistência.

Ao integrar o US-Prompt, o modelo deve ser capaz de efetivamente generalizar a partir de um conjunto limitado de exemplos, adaptando sua saída a diversos contextos enquanto produz histórias de usuários que atendem de forma consistente aos padrões de qualidade estabelecidos.

Para embasar o desenvolvimento do modelo da abordagem US-Prompt, utilizou-se os artigos de [Mao, He e Chen \(2025\)](#) e de [Leão et al. \(2024\)](#). O artigo de [Mao, He e Chen \(2025\)](#) apresenta uma análise abrangente da estrutura e composição de templates de prompts em aplicações reais de LLMs. O artigo construiu um conjunto de dados a partir de repositórios open-source do GitHub, incluindo contribuições de empresas como Uber e Microsoft, categorizando componentes, placeholders e padrões de composição dos templates.

Os principais componentes no template identificado, incluíram:

- **Profile/Role** – o papel ou função que o modelo deve assumir;
- **Directive** – o objetivo principal do prompt, geralmente na forma de instrução ou pergunta;

- **Workflow** – os passos e processos que o modelo deve seguir para completar a tarefa;
- **Context** – informações de fundo e contexto que o modelo deve considerar;
- **Examples** – exemplos de como a resposta esperada deve se parecer;
- **Output Format/Style** – tipo, formato ou estilo da saída;
- **Constraints** – restrições que o modelo deve respeitar ao gerar a resposta.

Mao, He e Chen (2025) também ressaltou que a organização adequada dos templates influencia significativamente a aderência aos objetivos das tarefas e a consistência das respostas. Além disso, templates bem estruturados e compostos segundo padrões específicos melhoraram a capacidade dos LLMs de seguir instruções e gerar respostas precisas. Esses resultados forneceram diretrizes para o design de templates robustos que elevam a qualidade das saídas em diversas aplicações de LLMs.

O segundo artigo utilizado como base, foi o de Leão et al. (2024), cujo método utilizado foi o padrão de prompt Gerenciador de Contexto (Context Manager), que visa restringir a análise de IA a um contexto específico. O método foi desenvolvido com base na modelagem do processo de negócio (BPMN) e em uma descrição formal detalhando: nome da atividade, responsável, padrão de prompt utilizado, objetivo, prompt a ser inserido e resposta esperada do LLM.

Com isso, a estrutura foi composta de cinco atividades, iniciando com o acesso ao LLM e finalizando com a construção de um conjunto de proto-personas.

A sequência de atividades utilizada por Leão et al. (2024) foi:

1. **Introdução do contexto:** estabelece a ideia da conversa e delimita o escopo informativo para a geração de proto-personas. O padrão de prompt aplicado foi o Context Manager, cujo objetivo é restringir a análise do LLM a um contexto específico, evitando respostas fora do tema.
2. **Visão do produto:** insere a visão do produto como prompt para o LLM, ampliando o contexto aplicativo, utilizando novamente o Context Manager.

3. **Matriz Is/Is Not/Does/Does Not:** fornece informações sobre o que o produto é e não é, complementando o contexto necessário para a geração de proto-personas, mantendo o uso do Context Manager.
4. **Fornecimento do template de proto-persona:** apresenta o formato de saída desejado (por exemplo, Lean Inception: perfil, necessidades e comportamentos), utilizando o Template Pattern para formatar a saída de forma específica.
5. **Geração final:** solicita que o LLM atue como um UI/UX Designer experiente e produza os proto-personas, aplicando o Persona Pattern para induzir o modelo a assumir o papel especificado.

Além disso, no estudo de [Leão et al. \(2024\)](#), foram realizados testes aplicando a técnica Chain of Thought – Zero Shot, para auxiliar na geração sequencial de proto-personas.

Com base nessas duas diretrizes de design de templates e focando em um contexto específico para elevar a qualidade das saídas em aplicações de LLMs, foi possível fundamentar o desenvolvimento e propor o a abordagem US-Prompt.

4.4.1 Design da Abordagem

Com base nos resultados obtidos no estudo empírico na Seção 4.3, as seguintes etapas foram adotadas para projetar a abordagem US-Prompt para a geração de histórias de usuários. Essas etapas tiveram como objetivo desenvolver o prompt para melhorar a qualidade das histórias geradas.

O desenvolvimento do US-Prompt seguiu as seguintes etapas:

- **Análise das Histórias Anteriores:** As histórias geradas manualmente e automaticamente foram avaliadas de acordo com os critérios do framework QUS. A análise revelou lacunas, como a falta de atomicidade e minimalidade, indicando a necessidade de ajustes no processo de geração automatizada.
- **Criação de um Prompt Refinado:** Com base nas lacunas identificadas, foi desenvolvido um prompt detalhado, baseado no template apresentado, que fornecia ao

ChatGPT templates para as histórias e orientações específicas sobre os critérios de qualidade. A abordagem Few-Shot incluiu exemplos explícitos de histórias que atendiam aos critérios de qualidade, enquanto o Meta Prompting ofereceu um contexto mais amplo sobre o que era solicitado.

- **Testes e Refinamentos:** Após a criação do prompt inicial, ele foi testado de forma iterativa para gerar novas histórias de usuários. A cada iteração, o prompt sofreu ajustes e melhorias contínuas ao longo do tempo, com base nos resultados dos testes anteriores. Ou seja, após cada teste, o prompt foi modificado ou refinado para melhorar os resultados, repetindo este ciclo até atingir o melhor desempenho possível, buscando melhorar a clareza das instruções e aprimorar os resultados de maneira progressiva. Além disso, tanto os critérios de avaliação quanto as orientações fornecidas no prompt foram refinados para garantir clareza e precisão, orientando a geração e revisão das histórias, enquanto asseguravam que elas atendiam aos critérios de qualidade estabelecidos. Exemplos de histórias bem formadas foram incluídos para facilitar a compreensão do que se era esperado.
- **Aplicação da Abordagem US-Prompt:** Após as iterações e refinamentos, o prompt final foi executado no ChatGPT, resultando na geração de 15 novas histórias de usuários, que foram então avaliadas usando os mesmos critérios das etapas anteriores, garantindo consistência na avaliação da qualidade.

A Figura 10 apresenta a captura de tela da conversa gerada durante a aplicação direta do US-Prompt no ChatGPT, registrando a solicitação final e as respostas obtidas. A conversa completa está disponível em: <<https://chatgpt.com/share/dd6b591b-df07-4c52-b228-9126247c4349>>.

4.4.2 Estrutura de Abordagem

O prompt foi implementado seguindo a abordagem apresentada por [Leão et al. \(2024\)](#), de forma estruturada para aprimorar a compreensão e a criação de histórias de usuários. A estrutura, organização e atividades do template original foram mantidas de forma similar, garantindo que a base metodológica fosse consistente e eficaz, e assim facilitando

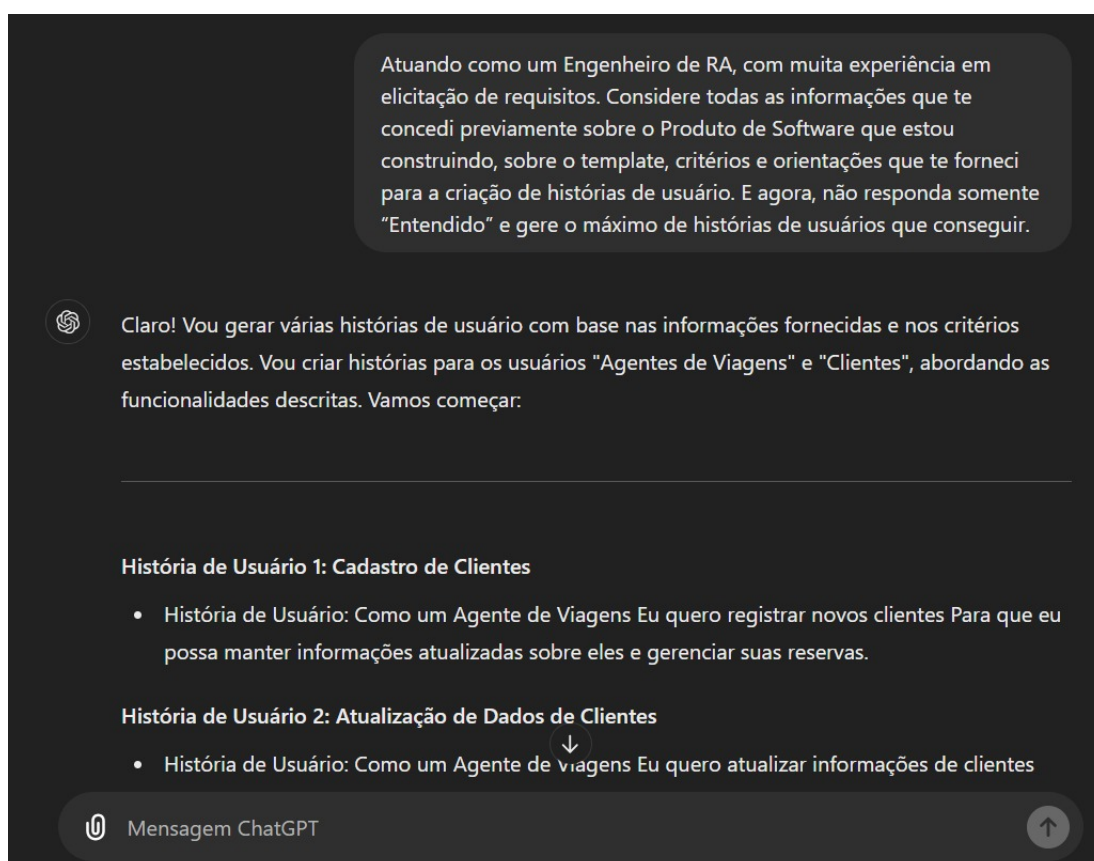


Figura 10 – Captura de tela da conversa no ChatGPT, demonstrando o resultado da aplicação do prompt final.

a geração de histórias mais alinhadas aos requisitos do produto. Abaixo está um resumo da abordagem inserida no **US-Prompt**:

1. **Orientações para Execução:** As orientações iniciais fornecem um guia sobre como o leitor deve proceder durante o processo de geração das histórias de usuários. Primeiro, o documento pede que o leitor leia atentamente todas as orientações antes de iniciar as atividades. Em seguida, são apresentadas instruções para garantir que o leitor compreenda os passos seguintes, que incluem a personalização do prompt com dados específicos do produto de software em questão e interação com o ChatGPT, seguindo cada etapa do processo de geração das histórias.
2. **Atividades:** O documento divide o processo em atividades específicas:
 - **Introdução ao Processo:** Esta atividade consiste em apresentar a ideia da conversa ao ChatGPT, fornecendo o contexto da dinâmica a seguir. O ChatGPT

deve ser instruído a estudar e reter as informações fornecidas, respondendo “Entendido” após cada solicitação.

- **Apresentação do Produto:** Esta atividade solicita uma visão geral do produto para um sistema de software, por exemplo, um sistema de gestão para uma agência de turismo, o qual o ChatGPT deve reconhecer.
- **Apresentação dos Requisitos:** Esta atividade solicita os requisitos do sistema, incluindo diferentes papéis de usuários (como Agentes de Viagem e Clientes) e principais ações desejadas, como o cadastro de clientes, gestão de reservas e comunicação entre cliente e agente.
- **Template de História de Usuário:** Esta atividade informa a estrutura desejada para a criação das histórias de usuário. Cada história deve incluir uma persona, uma ação e um objetivo.
- **Critérios de Qualidade e Orientações Adicionais:** Esta atividade define os critérios específicos para as histórias de usuários, garantindo que elas atendam a estes padrões exigidos. Cada história deve seguir sete critérios: bem formada, atômica, mínima, conceitualmente sólida, não ambígua, completa e estimável. Além dos critérios, também são fornecidas orientações adicionais, que visam esclarecer e direcionar o entendimento do ChatGPT sobre como aplicar corretamente os critérios em cada história gerada.
- **Exemplos de História de Usuário:** Esta atividade apresenta exemplos de histórias bem formadas, que servem como referência para auxiliar na criação de novas histórias de usuário, garantindo que elas atendam aos critérios estabelecidos.
- **Geração de História de Usuário:** Esta atividade instrui o ChatGPT a atuar como um Engenheiro de Requisitos, utilizando as informações fornecidas para gerar histórias de usuário para o software com base no template e critérios, em vez de responder com “Entendido”.

Em resumo, essa abordagem fornece um guia procedural para a geração sistemática de histórias de usuário precisas e estruturadas, aderindo a critérios predefinidos, no

contexto da descoberta de produtos de software. A descrição completa da abordagem do **US-Prompt (Versão 1)** pode ser consultada no Apêndice [G](#).

4.5 ESTUDO EXPERIMENTAL 2: PROPOSTA E AVALIAÇÃO DA ABORDAGEM US-PROMPT

Com base nas descobertas do Estudo Empírico na Seção [4.3](#), que revelou lacunas na qualidade das histórias de usuário geradas por métodos automatizados, este estudo apresenta a abordagem US-Prompt. Ao combinar as vantagens do Meta Prompting e do Few-Shot Prompt, conforme explorado no estudo de [Hiraou \(2024\)](#), essa técnica orienta o ChatGPT a produzir histórias de usuário mais precisas e de maior qualidade.

4.5.1 Desenho do Estudo

Este estudo teve como objetivo avaliar a eficácia do método US-Prompt na geração de histórias de usuário, com foco em melhorias em aspectos-chave como não ambiguidade, atomicidade e conformidade com padrões de qualidade predefinidos. Diferentemente do estudo anterior, que envolveu grupos de participantes, este estudo foi conduzido exclusivamente pelo pesquisador, que aplicou o método US-Prompt para gerar novas histórias de usuário e avaliar os resultados. Para isso, o estudo explorou a seguinte questão de pesquisa (QP): **Como a qualidade das histórias geradas com o US-Prompt se compara às geradas com prompt de formato livre?**

Hipóteses: Para responder a essa questão, o estudo foi estruturado com as seguintes hipóteses:

- **Hipótese Nula (H02):** Não há diferença significativa na qualidade das histórias de usuário geradas por meio de prompting livre em comparação com aquelas produzidas utilizando US-Prompt.
- **Hipótese Alternativa (HA2):** Há uma diferença significativa na qualidade das histórias de usuário geradas por meio de prompting livre em comparação com

aquelas produzidas utilizando US-Prompt.

Variáveis: A variável independente neste estudo é o método de geração das histórias de usuário, que inclui dois tratamentos: prompting livre (tratamento 1) e US-Prompt (tratamento 2). A variável dependente é a qualidade das histórias de usuário, avaliada com base na aderência ao framework QUS.

Desenho Experimental: Inicialmente, foram estabelecidos dois conjuntos de referência de histórias de usuário, compostos por histórias geradas manualmente e por meio de prompting livre (usados no Estudo Empírico 1). Esses conjuntos serviram como pontos de comparação para avaliar a eficácia do método US-Prompt. Em seguida, o pesquisador aplicou o método US-Prompt ao ChatGPT para gerar um novo conjunto de histórias de usuário. As novas histórias geradas foram então analisadas para avaliar sua qualidade em comparação com os conjuntos de referência. A qualidade das histórias geradas foi avaliada com base nos mesmos critérios estabelecidos no primeiro estudo, garantindo uma avaliação consistente para fins de comparação.

Para avaliar a eficácia do método US-Prompt, foi realizada uma análise comparativa entre as histórias geradas por esse método e os dois conjuntos de referência, geradas manualmente e por prompting livre. Além disso, foi conduzida uma análise qualitativa para examinar melhorias na qualidade das histórias com base nos critérios predefinidos. Por fim, foi realizada a análise estatística para determinar se os diferentes métodos produziram diferenças significativas na qualidade das histórias.

4.5.2 Resultados e Discussão

O prompt final, o US-Prompt, foi aplicado para guiar o modelo na geração de histórias de usuário. Essa abordagem combina a estruturação do Meta-Prompt com a flexibilidade do Few-Shot Prompt, permitindo que o modelo receba diretrizes específicas e exemplos de histórias anteriores. O resultado foi a geração de 15 histórias, elaboradas para atender aos critérios de qualidade estabelecidos.

A aplicação do US-Prompt proporcionou um contexto rico e detalhado, auxiliando o modelo na captura de nuances importantes dos requisitos do usuário. Cada

história gerada foi projetada para refletir a funcionalidade desejada de maneira simples, coesa e mínima. Essa técnica teve como objetivo aprimorar a relevância e a qualidade das histórias, tornando o processo de produção mais eficiente e eficaz, pois não apenas garante a criação de histórias relevantes e de alta qualidade, mas também possibilita a geração dessas histórias de forma ágil, utilizando apenas os poucos recursos disponíveis, contribuindo para um processo mais otimizado e simplificado no ambiente de desenvolvimento ágil.

As histórias resultantes foram então avaliadas quanto à conformidade com os mesmos critérios de qualidade. Como forma de exemplificação, a Tabela 4 apresenta exemplos produzidos por dois métodos de geração manual e a geração automatizada por meio da abordagem US-Prompt. A comparação evidencia as diferenças entre as histórias produzidas.

Observa-se, por exemplo, que na História 1, a versão manual compromete os critérios de atomicidade, minimalidade e estimabilidade, ao agregar múltiplas funcionalidades e informações em uma única história. Em contraste, a versão gerada com o US-Prompt não infringe esses critérios, principalmente por se concentrar em uma única funcionalidade, apresentando uma estrutura mais concisa.

De forma semelhante, na História 2, a versão manual repete os mesmos problemas de qualidade identificados anteriormente, enquanto a versão gerada pelo US-Prompt apresenta melhor delimitação, embora ainda infrinja o critério de minimalidade, ao incluir elementos adicionais que não são estritamente necessários à compreensão da história.

O desempenho das histórias geradas, em comparação com aquelas geradas por prompting livre, é apresentado na Figura 11, destacando a eficácia do US-Prompt na geração de histórias de maior qualidade.

Os métodos automatizados utilizados na geração de histórias de usuários apresentaram eficiências variando de 46,94% a 75,32%. Isso indica que, embora esses métodos tenham produzido resultados variados, eles não alcançaram uma eficiência consistentemente alta.

Em contraste, o método US-Prompt se destacou ao atingir uma eficiência de

Tabela 4 – Exemplos de histórias de usuários geradas manualmente e com US-Prompt.S

Geração Manual	US-Prompt
História 1 Como um agente de viagens, eu quero pesquisar e sugerir opções de viagens baseadas no orçamento e nas preferências dos meus clientes, para oferecer opções diversificadas, viáveis financeiramente e adequadas aos diferentes meios de locomoção.	História 1 Como um agente de viagens, eu quero pesquisar opções de viagem, para que eu possa sugerir as melhores alternativas com base nas preferências e no orçamento dos clientes.
História 2 Como cliente, eu quero solicitar cotações e realizar reservas, para planejar minha viagem e garantir os serviços desejados.	História 2 Como cliente, eu quero fazer reservas diretamente pelo sistema, para garantir os serviços de viagem desejados de forma rápida e eficiente.

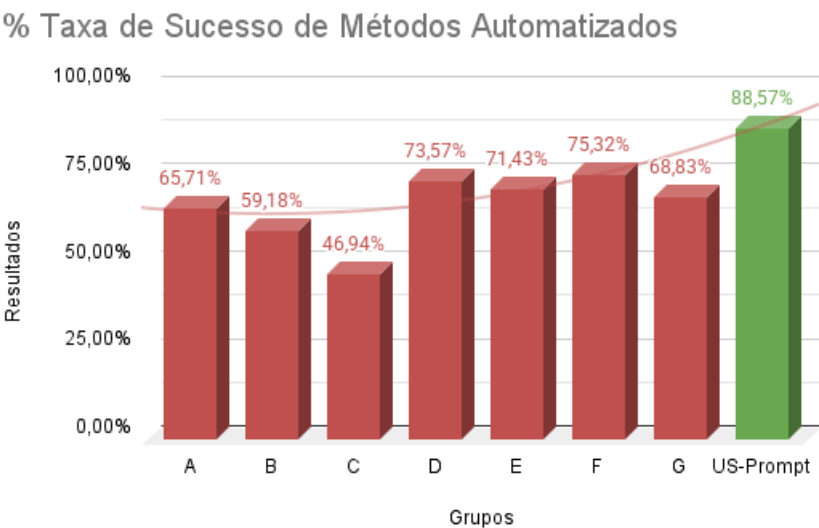


Figura 11 – Taxa de sucesso dos métodos Free-Form Prompt e US-Prompt.

88,57%, calculada com base no número de respostas corretas em relação ao total de critérios avaliados. Essa métrica fornece uma visão clara da eficácia do método na geração de histórias de usuários de alta qualidade. A diferença significativa na eficiência sugere que o US-Prompt é uma abordagem mais eficaz para a geração de histórias em comparação com os métodos automatizados simplificados.

Ao comparar a eficácia média dos métodos automatizados (65,85%) com o US-Prompt, observa-se que este último foi 22,72% mais eficaz.

Também foi realizada uma análise estatística utilizando o teste t para uma amostra. Observou-se que, ao comparar a geração manual com o US-Prompt, o valor de p foi 0,001. Além disso, ao comparar a geração automatizada utilizando o método Free-Form Prompt com o US-Prompt, o valor de p foi $< 0,001$. Esses resultados fornecem evidências estatísticas suficientes para rejeitar a hipótese nula (H_0) e concluir que a taxa de sucesso alcançada pelo US-Prompt é significativamente diferente tanto da taxa de sucesso da geração manual quanto da geração automatizada utilizando Free-Form Prompt. Portanto, a hipótese alternativa (H_A) é válida.

Essa diferença substancial destaca o potencial do US-Prompt como uma técnica superior, capaz de gerar histórias de usuários com maior qualidade e eficácia.

Esses resultados indicam que, ao aplicar a combinação de Meta Prompting e Few-Shot Prompt, o modelo não apenas melhora a qualidade das histórias geradas, mas também demonstra um aumento significativo na eficiência. Esses achados corroboram os resultados do estudo de [Hiraou \(2024\)](#), que enfatiza como a aplicação de técnicas avançadas, como o US-Prompt, pode levar a melhorias significativas na qualidade e eficácia das respostas geradas por modelos de linguagem, destacando seu potencial em contextos práticos, como engenharia de requisitos e ambientes ágeis, onde velocidade e qualidade são essenciais.

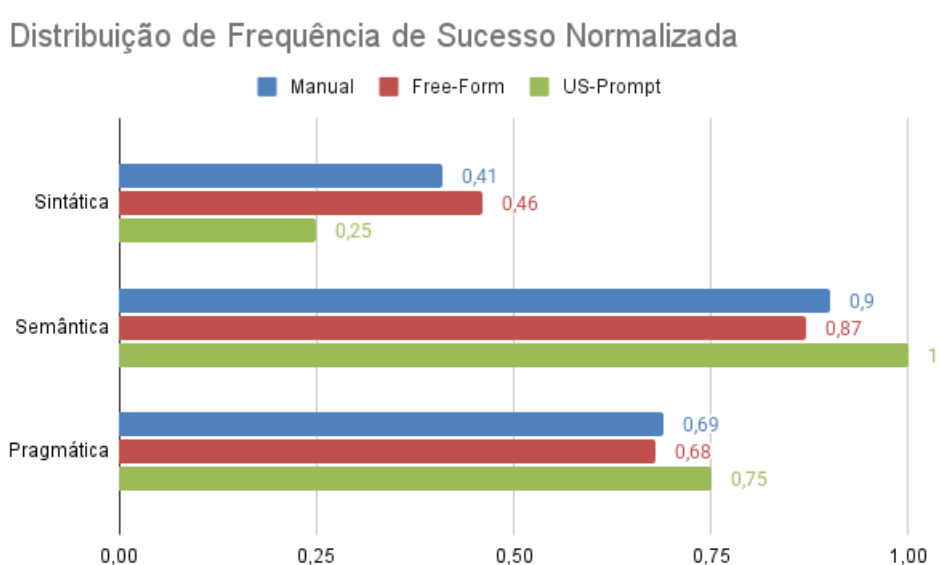


Figura 12 – Comparação dos métodos em relação às categorias de qualidade.

Na Figura 12, a frequência normalizada de respostas corretas revela a proporção de histórias que atenderam aos critérios de qualidade em três categorias do framework QUS, sintática, semântica e pragmática, para os métodos manual, automatizado com Free-Form Prompt e com US-Prompt.

As histórias geradas automaticamente apresentaram uma leve melhora na correção sintática, 0,46, em comparação com as histórias manuais, 0,41. No entanto, foram observadas pequenas reduções nas categorias semântica (0,87 versus 0,90) e pragmática (0,68 versus 0,69), o que sugere que erros semânticos e pragmáticos continuam sendo desafios relevantes, apesar da melhora na sintaxe.

Histórias manuais e aquelas geradas com Free-Form Prompt apresentaram taxas de sucesso de 0,41 e 0,46 na categoria sintática, respectivamente. Embora o método US-Prompt tenha demonstrado um desempenho inferior nessa categoria (0,25), ele compensou essa limitação com melhorias notáveis nas categorias semântica e pragmática.

Na categoria semântica, houve uma melhoria significativa, com o US-Prompt alcançando uma taxa de acerto de 1, em contraste com 0,90 e 0,87 dos métodos manual e Free-Form, respectivamente. Na categoria pragmática, o US-Prompt também apresentou aumento, atingindo uma taxa de sucesso de 0,75, em comparação com 0,69 e 0,68 dos métodos manual e Free-Form Prompt.

Assim, o método US-Prompt apresentou uma melhora substancial na taxa de sucesso em comparação aos métodos manual e Free-Form Prompt. Especificamente, o US-Prompt obteve uma taxa de acerto de 1 na categoria semântica, superando os métodos manual (0,90) e Free-Form (0,87). Na categoria pragmática, o US-Prompt também apresentou aumento, com taxa de sucesso de 0,75, em comparação com 0,69 (manual) e 0,68 (Free-Form).

Sessenta por cento das histórias geradas com o US-Prompt atenderam completamente aos critérios de qualidade estabelecidos, demonstrando avanços nos aspectos semântico e pragmático. No entanto, o erro sintático mais comum foi a dificuldade em garantir a atomicidade, essencial para a formulação clara e simples das histórias. Histórias não atômicas comprometeram a minimalidade, tornando-as mais complexas e difíceis de estimar e priorizar. Esse problema resultou em uma taxa de erro de 40% nas

histórias geradas. Metade desses erros, especificamente em três histórias, ocorreram devido a dificuldades em manter a atomicidade, impactando diretamente tanto a clareza sintática quanto a viabilidade de implementação e priorização dessas histórias dentro do planejamento ágil.

Os resultados deste estudo indicam que as histórias de usuário geradas pelo ChatGPT, especialmente utilizando a abordagem US-Prompt, são de qualidade superior às escritas manualmente ou geradas por meio de free-form prompt. Isso está alinhado com os achados de [Brockenbrough e Salinas \(2024\)](#), que observaram que histórias de usuário criadas com o auxílio de IA superaram aquelas desenvolvidas sem o uso dessa ferramenta, ao mesmo tempo em que ampliaram seu escopo ao incorporar prompts estruturados com exemplos claros.

O US-Prompt aborda de forma eficaz aspectos como minimalidade e clareza semântica, superando o free-form prompt tanto em qualidade semântica quanto pragmática. Isso reforça sua capacidade de superar as limitações dos métodos anteriores e melhorar a qualidade geral das histórias de usuário geradas.

4.6 AMEAÇAS À VALIDADE

Todos os estudos enfrentam ameaças que podem comprometer a validade dos resultados ([WOHLIN et al., 2012](#)), as quais foram categorizadas em ameaças à validade interna, externa, de conclusão e de constructo, com as respectivas estratégias de mitigação descritas a seguir:

Validade Interna: A variabilidade nas habilidades dos participantes para gerar histórias de usuário manual e automaticamente representa uma ameaça. Embora todos os participantes tenham recebido treinamento prévio focado na criação de histórias de usuário, a ausência de uma estrutura rígida de prompt para alguns pode ter levado a uma maior variabilidade nos resultados, especialmente para aqueles menos familiarizados com a geração automatizada de histórias. Um dos principais benefícios do uso do US-Prompt no estudo é sua capacidade de reduzir essa dependência da experiência, ao fornecer um prompt estruturado que qualquer participante pode seguir.

Validade Externa: A amostra da pesquisa foi limitada a estudantes do curso de Engenharia de Software, o que pode restringir a generalização das competências para papéis fora desse contexto acadêmico. No entanto, estudos como os de [Höst, Regnell e Wohlin \(2000\)](#) e [Salman, Misirli e Juristo \(2015\)](#) demonstraram que estudantes podem representar adequadamente profissionais da indústria, tornando essa amostra adequada para a avaliação do framework.

Validade de Constructo: As métricas utilizadas para avaliar a qualidade das histórias de usuário, com base no framework QUS, oferecem um conjunto robusto de critérios para avaliar diferentes dimensões da qualidade. Embora essas métricas sejam confiáveis e assegurem a qualidade geral das histórias geradas, elas podem não capturar todas as nuances em contextos mais complexos. No entanto, o uso do QUS minimiza significativamente esse risco.

Validade de Conclusão: Inclui o tamanho reduzido da amostra, que limita a generalização, a variabilidade nas habilidades dos participantes, o risco de erros estatísticos devido ao número limitado de observações e o baixo poder estatístico, como evidenciado pela semelhança nos resultados. Além disso, o viés na análise, decorrente da avaliação subjetiva do pesquisador, pode comprometer a objetividade das conclusões. Para minimizar a influência do pesquisador, o autor foi responsável por identificar e analisar todos os resultados do estudo. Posteriormente, todos os dados e achados foram revisados e validados por dois outros pesquisadores: o orientador, especialista em técnicas de prompting, e o coorientador, especialista em Engenharia de Requisitos, assegurando a validade e confiabilidade das conclusões.

Ameaças de Alucinação: Uma preocupação ao utilizar modelos de linguagem é a alucinação, que se refere à geração de informações imprecisas ou fabricadas que parecem plausíveis, mas carecem de fundamento. Isso ocorre porque modelos como o ChatGPT são treinados com grandes volumes de dados online, permitindo-lhes extrapolar informações, interpretar prompts ambíguos ou modificar dados sem uma base sólida ([HUANG et al., 2024](#)). Como resultado, as histórias de usuário geradas podem conter detalhes errôneos que não refletem a realidade.

4.7 CONCLUSÃO

Este estudo investigou a aplicação do método US-Prompt para a geração automatizada de histórias de usuário, comparando seu desempenho com métodos manuais e free-form prompting. A abordagem foi avaliada com base em três categorias de qualidade: sintática, semântica e pragmática. O objetivo foi determinar se a combinação de Meta Prompting e Few-Shot Learning poderia superar as limitações dos métodos tradicionais, aumentando tanto a eficácia quanto a qualidade das histórias geradas.

Os resultados demonstraram que o US-Prompt trouxe melhorias significativas em relação aos métodos anteriores. Ele se destacou particularmente na categoria semântica, onde eliminou completamente os erros, evidenciando a capacidade do modelo de interpretar e traduzir com precisão os requisitos do usuário em histórias coerentes. Além disso, 60% das histórias geradas atenderam integralmente aos critérios de qualidade estabelecidos, superando as médias dos métodos manuais e de free-form prompting, que apresentaram taxas de sucesso inferiores em diversas categorias.

No entanto, o método enfrentou desafios na categoria sintática devido a dificuldades em garantir a atomicidade das histórias de usuário. Histórias que não atendem a esse critério tendem a ser mais complexas e difíceis de estimar e priorizar, impactando negativamente tanto a qualidade sintática quanto a pragmática. Metade dos erros observados nas histórias geradas por US-Prompt foi atribuída à dificuldade em manter a atomicidade, o que sugere que ainda há espaço para melhorias na estruturação de histórias concisas e claras.

Esses resultados indicam que, embora o US-Prompt ofereça uma solução promissora para a geração automatizada de histórias de usuário, especialmente em termos de qualidade semântica e pragmática, há oportunidades para otimizar ainda mais o processo. A técnica se mostrou significativamente mais eficaz do que os métodos manuais e automatizados, mas existem possibilidades de aperfeiçoamentos que podem levar a uma geração automatizada mais eficiente e precisa.

Em conclusão, os achados deste estudo estão alinhados com os de [Brockenbrough e Salinas \(2024\)](#), que indicam que histórias de usuário geradas com o auxílio do ChatGPT apresentam qualidade superior em comparação com aquelas criadas sem o suporte da

IA. Embora ambos os estudos demonstrem que a eficácia do ChatGPT é ampliada por meio de técnicas de prompting estruturadas, os resultados mostram inequivocamente que, quando utilizado corretamente, o ChatGPT melhora significativamente a qualidade das histórias de usuário.

Este capítulo marca o início da primeira iteração do Ciclo de Design, com a construção do artefato US-Prompt, sucedendo o Ciclo de Relevância, que envolveu o survey, o estudo exploratório com prompting livre e a primeira experimentação estruturada. Esta etapa teve como objetivo projetar, aplicar e avaliar o artefato, de forma que ele respondesse aos desafios identificados na geração de histórias de usuários com apoio de LLMs. Em seguida, apresenta-se a segunda iteração do Ciclo de Design, que consolida o processo de refinamento e reavaliação do artefato proposto.

5

2º CICLO DE DESIGN: HISTÓRIAS DE USUÁRIOS GERADAS POR IA. ELAS SÃO BOAS O SUFICIENTE?

Este capítulo apresenta um estudo experimental que investigou a eficácia da técnica US-Prompt na geração automatizada de histórias de usuários em diferentes contextos. O capítulo foca na comparação entre métodos que buscam melhorias na qualidade das histórias geradas, bem como na avaliação da aceitação da abordagem pelos usuários. Os resultados contribuem para o avanço das técnicas de geração automatizada de histórias de usuários no contexto da Engenharia de Requisitos, destacando como a aplicação do US-Prompt pode otimizar e ampliar o processo de especificação de requisitos. Este estudo foi publicado no artigo nacional **“AI-Generated User Stories: Are They Good Enough?”**, aceito na trilha de **Ideias Inovadoras e Resultados Emergentes do XXXIX Simpósio Brasileiro de Engenharia de Software (SBES 2025)**.

5.1 INTRODUÇÃO

Como visto no capítulo anterior, o US-Prompt surge como uma técnica especializada que permite gerar histórias de usuários de forma automatizada, mantendo um maior

padrão de qualidade baseado no framework Quality User Story (QUS) de [Lucassen et al. \(2016b\)](#). Entretanto, além da qualidade técnica, é crucial entender como essa abordagem é percebida por profissionais da área, o que pode influenciar diretamente sua adoção em contextos industriais ([RAHMAN; ZHU, 2024](#); [OSWAL; KANAKIA; SUKTEL, 2024](#)).

Este estudo visa preencher essa lacuna, ao realizar um estudo experimental com participantes da disciplina de Engenharia de Requisitos e Análise de Sistemas (ERAS) do curso de Engenharia de Software da Universidade Federal do Amazonas (UFAM). Durante o experimento, os participantes utilizaram o US-Prompt (V2) para criar histórias de usuários, e os resultados foram avaliados tanto pela qualidade das histórias quanto pela aceitação da técnica, utilizando o modelo TAM3 (Technology Acceptance Model) ([VENKATESH; BALA, 2008](#)).

O TAM3 é um modelo teórico amplamente utilizado para medir a aceitação de novas tecnologias, considerando fatores como Utilidade Percebida (Perceived Usefulness), Facilidade de Uso Percebida (Perceived Ease of Use) e Intenção de Uso Futuro (Behavioral Intention) ([VENKATESH; BALA, 2008](#)).

Os principais objetivos deste estudo incluem avaliar a viabilidade da técnica de US-Prompt na geração automatizada de histórias de usuários, identificar oportunidades de melhoria e fornecer insights práticos sobre como essa abordagem pode ser aplicada na prática da engenharia de requisitos.

5.2 Evolução do Prompt para Geração de Histórias de Usuário

A Versão 1 do US-Prompt era composta por 5 atividades, em sua estrutura inicial, disponível no Apêndice [G](#). Nessa versão, os critérios de qualidade, o template e os exemplos estavam agrupados em uma única atividade, e o prompt aplicava um exemplo fixo de Agência de Turismo, usado na fase de elaboração e nos testes iniciais. A instrução final para o ChatGPT era gerar o máximo de histórias possíveis com base nas informações fornecidas.

Após os testes iniciais, o foco passou a ser viabilizar o prompt para uso por parti-

participantes reais. Para isso, foram necessárias algumas mudanças, como a generalização do conteúdo: exemplos prontos foram retirados e placeholders foram criados para indicar onde os participantes deveriam inserir suas informações.

Em seguida, foi realizado um estudo piloto com um participante, posteriormente descartado da amostra. O objetivo do piloto foi validar se as instruções eram suficientemente claras para que o participante pudesse entender o procedimento, seguir o passo a passo e executar sozinho a criação das histórias.

Com base no feedback do estudo piloto, percebeu-se a necessidade de melhorar tanto as instruções quanto a estrutura do prompt. Observou-se que havia dados concentrados em poucas atividades, então optou-se por separar o conteúdo em atividades menores. Além disso, a descrição de cada atividade foi refinada, incluindo instruções diretas para o leitor, prompt e o uso de placeholders no lugar de exemplos fixos (produto, usuários, ações).

A Versão 2 passou a ter 7 atividades em sua estrutura, cada uma com: descrição, instruções, prompt e resposta esperada. A última atividade foi atualizada para solicitar, além das histórias padrão, a criação de histórias adicionais que melhorem a experiência do usuário e a eficiência do sistema. A Versão 2 está disponível no Apêndice C.

5.3 Estudo Experimental

O estudo busca responder à seguinte questão de pesquisa: **“A adoção do US-Prompt é capaz de gerar histórias de usuários de maior qualidade em diferentes projetos?”** Para isso, foi explorada a aplicação da técnica em um cenário experimental, fornecendo insights sobre sua eficácia na geração de histórias de qualidade e aceitação. A metodologia segue as diretrizes de [Wohlin et al. \(2012\)](#) para estudos experimentais, permitindo uma análise fundamentada dos resultados.

O estudo foi realizado com 24 participantes, organizados em seis grupos, no contexto de uma disciplina de Engenharia de Software. O estudo investigou a eficácia e a aceitação do US-Prompt (V2) para a geração automatizada de histórias de usuários, com o uso do ChatGPT, através da técnica de meta-few-shot prompt, aplicada a diferentes

contextos.

Além disso, a aceitação foi avaliada utilizando o modelo TAM3 (VENKATESH; BALA, 2008). Ao todo, foram geradas 457 histórias de usuários, das quais 87,5% apresentaram taxas de sucesso superiores a 75% em aderência aos critérios de qualidade do QUS. Os resultados indicam que, mesmo em sua versão padrão, o US-Prompt se mostra eficaz na geração de histórias consistentes e de maior qualidade.

5.3.1 Desenho do Estudo

O estudo foi projetado para simular cenários do mundo real enfrentados por engenheiros de software iniciantes, garantindo condições consistentes de coleta de dados entre os grupos. Para isso, cada grupo desenvolveu um sistema completo, planejado e implementado seguindo práticas de desenvolvimento ágil, incluindo definição de problemas realistas, elicitação de requisitos e implementação iterativa, como fariam em um projeto profissional real.

O objetivo principal foi avaliar a eficácia do US-Prompt na geração de histórias de usuários de maior qualidade, bem como identificar fatores que influenciam sua aceitação entre futuros profissionais da área.

Para isso, o estudo foi estruturado em três fases: (1) geração e avaliação das histórias utilizando o US-Prompt padrão pré-fornecido, (2) geração e avaliação das histórias utilizando com uma versão modificada pelos próprios participantes do US-Prompt e (3) avaliação da aceitação do US-Prompt utilizando o modelo TAM3. A eficácia foi medida pela taxa de sucesso, que é a proporção de histórias que atendem aos critérios QUS e a aceitação foi analisada pela utilidade percebida, facilidade de uso e intenção de uso.

5.3.1.1 Hipóteses

O estudo foi desenhado com as seguintes hipóteses:

- **Hipótese Nula (H01):** Não há diferença significativa na qualidade das histórias de

usuários geradas com o US-Prompt utilizando o prompt padrão em comparação ao prompt modificado pelos participantes.

- **Hipótese Alternativa (HA1):** Existe uma diferença significativa na qualidade das histórias de usuários geradas com o US-Prompt utilizando o prompt padrão em comparação ao prompt modificado pelos participantes.

5.3.1.2 Variáveis

A variável independente do estudo é o método de geração automatizada das histórias de usuários com US-Prompt, o qual é dividido em dois tratamentos: (1) a utilização do US-Prompt padrão e (2) a utilização do US-Prompt modificado pelos participantes. Já a variável dependente é a qualidade das histórias de usuários, avaliada pelo framework QUS.

5.3.1.3 Participantes

Os participantes foram recrutados na disciplina de Tópicos Avançados em Engenharia de Software (TAES), ofertada em conjunto para estudantes de graduação e pós-graduação do curso de Engenharia de Software da UFAM. Ao todo, 29 participantes aceitaram participar da pesquisa e assinaram o termo de consentimento livre e esclarecido, com a opção de se retirar do estudo a qualquer momento, garantindo a ética e o voluntariado no processo.

5.3.2 Execução do Estudo

Inicialmente, os participantes se organizaram em sete equipes de forma autônoma, com base em afinidades mútuas e disponibilidade. No entanto, apenas seis equipes completaram todas as fases propostas do estudo, totalizando 24 participantes. Destes, 6 são estudantes que também atuam como profissionais de software.

Os grupos receberam a orientação para definir o contexto do sistema que dese-

javam desenvolver. As descrições dos contextos dos sistemas definidos pelas equipes encontram-se no Apêndice E. E na execução do estudo, foi fornecido um treinamento inicial sobre o uso de prompts especializados. Os grupos foram orientados a utilizar o GPT-4o (Mini) com configurações padrão (temperature 0.7, top-p 1.0, penalidades de frequência e presença 0) para gerar as histórias de usuários. O estudo foi conduzido em três fases:

- **Fase 1 - Geração com Prompt Padrão:** Os participantes utilizaram US-Prompt com o prompt padrão (tratamento 1), e as histórias geradas foram avaliadas pelos pesquisadores, quanto à qualidade com o framework QUS.
- **Fase 2 - Geração com Prompt Modificado:** Os participantes puderam realizar modificações no prompt e geraram novas histórias de usuários utilizando o prompt modificado (tratamento 2). Novamente, as histórias foram avaliadas com o framework QUS.
- **Fase 3 - Avaliação de Aceitação (TAM):** Após a solicitação para geração das histórias, foi aplicado o modelo TAM3 para medir a aceitação da técnica, considerando a utilidade percebida, a facilidade de uso e a intenção de uso futuro.

5.3.3 Coleta de Dados

Os dados foram coletados por meio da avaliação das histórias de usuários geradas, utilizando o framework QUS, e de questionários de aceitação baseados no modelo TAM3. A análise incluiu métricas quantitativas, referentes à eficácia da geração das histórias, e qualitativas, considerando o feedback dos participantes sobre o uso da técnica e sugestões de melhorias.

- **Fase 1 - Geração com Prompt Padrão:** Na primeira fase, cada participante gerou individualmente um conjunto de histórias de usuários utilizando o prompt padrão fornecido pelo estudo. As histórias de usuários foram criadas com base no contexto do sistema previamente definido por cada grupo. Como cada um dos 24

participantes contribuiu com um conjunto de histórias, esta fase resultou em 24 conjuntos de histórias de usuários, totalizando 372 histórias, distribuídas entre as equipes da seguinte forma: Equipe A (59 histórias), Equipe B (42 histórias), Equipe C (69 histórias), Equipe D (75 histórias), Equipe E (51 histórias) e Equipe F (76 histórias).

- **Fase 2 - Geração com Prompt Modificado:** Na segunda fase, os participantes trabalharam de forma colaborativa com seus grupos para revisar e aprimorar colaborativamente o US-Prompt padrão. Após modificar o prompt, cada equipe gerou um novo conjunto de histórias de usuários usando sua versão revisada. Cinco das seis equipes produziram 5 novos conjuntos de histórias, resultando em 85 novas histórias. Uma equipe optou por não fazer nenhuma modificação no prompt e manteve as histórias geradas na fase anterior. Todas as equipes utilizaram a mesma descrição do sistema em ambas as fases.

5.3.4 Avaliação das Histórias de Usuários Geradas

As histórias de usuários geradas foram avaliadas por meio de métricas quantitativas, com foco na eficácia do processo de geração.

A eficácia do método é medida pela capacidade de gerar histórias de usuários que atendam aos critérios de qualidade definidos (bem formada, atômica, mínima, conceitualmente sólida, não ambígua, frase completa e estimável).

Os pesquisadores avaliaram cada história de usuário de acordo com os critérios do QUS utilizando o mesmo sistema de pontuação binário, sendo que, uma pontuação de "1" indicava conformidade com o critério, enquanto "0" indicava não conformidade. Isso determinou se cada história atendia ao critério de qualidade específico. Por exemplo, a história do Grupo C, “Como cinéfilo, quero filtrar o catálogo de filmes por gênero e diretor”, deveria ser dividida em duas histórias atômicas: “Como cinéfilo, quero filtrar o catálogo de filmes por gênero” e “Como cinéfilo, quero filtrar o catálogo de filmes por diretor.”

A pontuação geral de eficácia para um conjunto de histórias de usuários foi

calculada como a taxa de sucesso, seguindo o mesmo procedimento do capítulo anterior, na Seção 4.4, representando a porcentagem de critérios satisfeitos em relação ao total avaliado.

Primeiro, somou-se o número total de critérios satisfeitos em todas as histórias de usuários avaliadas. Em seguida, determinou-se o número máximo de sucessos possíveis, multiplicando o número de histórias avaliadas pelo número de critérios de qualidade, que se manteve o mesmo (7). Por fim, o número de critérios satisfeitos foi dividido pelo total de sucessos possíveis e multiplicado por 100 para obter a taxa de sucesso. Essa taxa reflete a aderência geral das histórias de usuários aos critérios de qualidade do QUS e indica a eficácia da técnica.

5.4 Resultados e Discussão

O estudo analisou dados quantitativos e qualitativos. Os resultados quantitativos foram baseados nas taxas de sucesso da geração de histórias de usuários. Para os dados qualitativos, as respostas abertas do questionário baseado no TAM3 foram lidas e analisadas manualmente pelo primeiro autor. As respostas foram agrupadas em categorias temáticas emergentes, como clareza das instruções, utilidade percebida e sugestões de melhoria. Embora nenhum software de análise qualitativa tenha sido utilizado, a codificação foi realizada manualmente para identificar padrões recorrentes nas percepções dos participantes.

5.4.1 Análise Quantitativa

Os dados foram compilados a partir das avaliações das histórias geradas com o prompt padrão pré-fornecido e com o prompt modificado, após as mudanças feitas pelos participantes. A análise incluiu o cálculo da taxa de sucesso de cada critério de qualidade atendido. As histórias de cada grupo, independentemente do método de geração, foram avaliadas com base nos critérios de qualidade.

A Figura 13 ilustra o número de histórias de usuários geradas pelos seis grupos

participantes do estudo na fase 1, utilizando o prompt padrão. A quantidade média de histórias geradas foi de aproximadamente 14,88, sendo que a menor quantidade foi gerada pelo participante P6, com 9 histórias. Isso mostra uma variação na produção das histórias, mas a maioria dos participantes gerou 14 ou mais histórias. Nos grupos A, D e E, todos os participantes geraram 14 ou mais histórias, sugerindo que o prompt padrão teve um efeito consistente na geração de histórias, com algumas variações pontuais, como o caso de P6. Esses resultados apoiam a capacidade da versão do prompt de manter um volume estável de geração de histórias de usuários.

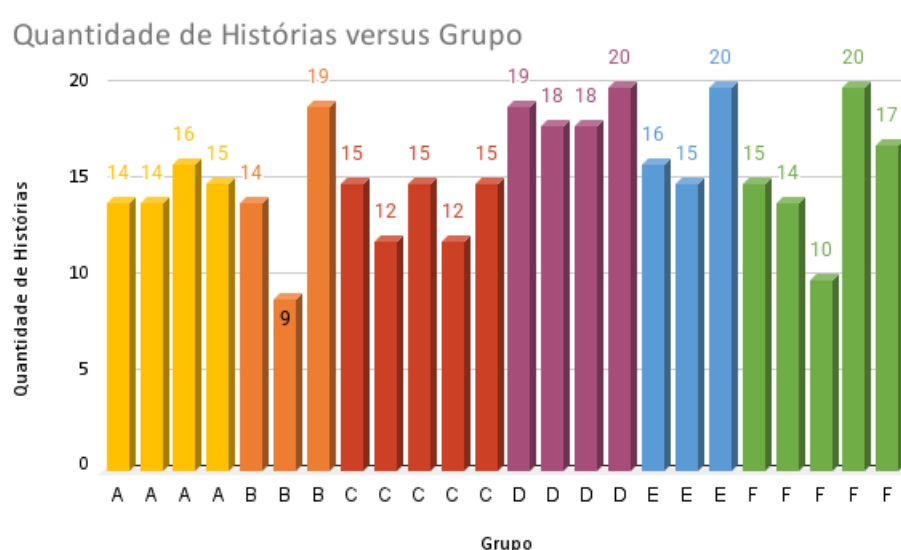


Figura 13 – Histórias de usuário produzidas a partir do prompt padrão, organizadas por participante de cada grupo, com representação em barras coloridas.

O US-Prompt na versão Padrão demonstrou uma alta eficácia na geração de histórias de usuários. Segundo o gráfico da Figura 14, dos 24 conjuntos de histórias avaliados, apenas 3 apresentaram uma taxa de sucesso abaixo de 75%, conforme mostrado no gráfico. Esse percentual reflete apenas o desempenho observado na amostra e não foi definido como limiar ou threshold previamente estabelecido. Isso significa que 87,5% dos conjuntos de histórias atingiram bons resultados, reforçam a consistência e a eficácia da abordagem adotada. Essa elevada taxa de acerto evidencia o potencial do prompt padrão em gerar histórias de usuários aderentes aos critérios de qualidade estabelecidos, mesmo antes de quaisquer melhorias ou personalizações.

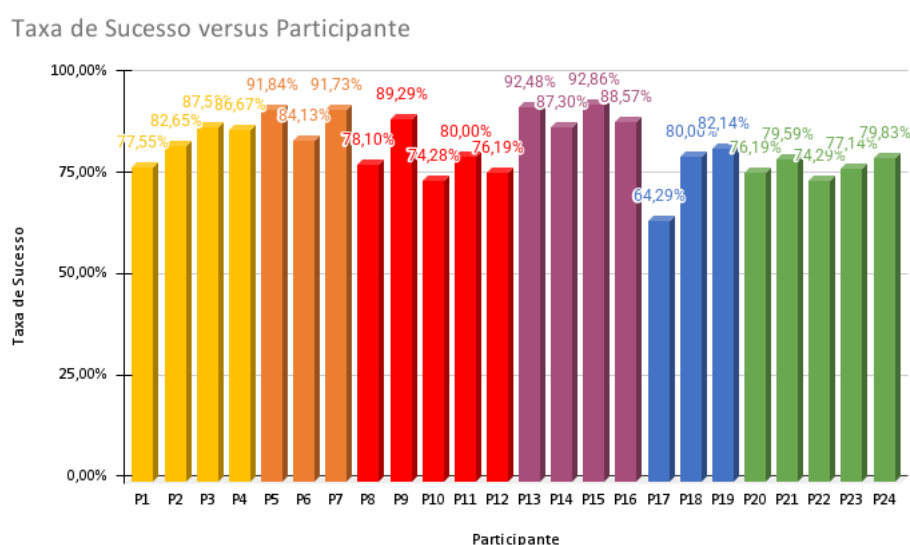


Figura 14 – Distribuição das taxas de sucesso por participante na geração de histórias de usuários utilizando o US-Prompt Padrão.

5.4.1.1 Análise das Estratégias de Refinamento dos Prompts

A Tabela 5 mostra a associação de cada grupo ao tema do projeto desenvolvido, oferecendo uma visão geral dos sistemas estudados.

Tabela 5 – Associação dos Grupos aos Projetos.

Grupos	Temas dos Projetos
Grupo A	Sistema de Restaurante com Controle de pedidos, Preparação e Entrega
Grupo B	Sistema de Horas Complementares
Grupo C	Aplicação de Sugestão de Filmes
Grupo D	Sistema de Gestão de Finanças Pessoais para Idosos
Grupo E	TeseLab - Uma ferramenta Web para Realização de Debate de Teses
Grupo F	Comparador de Aplicativos de Serviço de Transporte

Todos os grupos iniciaram a atividade utilizando o mesmo prompt-base padrão, elaborado com técnicas de Meta-Prompting e Few-Shot Prompting, o US-Prompt, composto por 7 atividades e que está disponível no Apêndice C. Após essa etapa inicial, foi concedida autonomia para que os grupos pudessem modificar o prompt conforme julgassem necessário, com o objetivo de explorar formas mais eficazes de orientar a geração das histórias de usuários. Abaixo, apresenta-se uma análise comparativa das principais alterações realizadas:

Grupo A: optou por um tom mais informal e conversacional, incorporando emojis ("😊", "😄") como forma de verificação de entendimento. Além disso, o grupo inseriu uma orientação direta de execução ao final do prompt, com o comando "CRIE". Essa abordagem sugere uma tentativa de reforçar a clareza da ação esperada pelo modelo.

Grupo B: manteve o prompt inalterado. A ausência de modificações pode indicar que o grupo considerou o prompt inicial suficientemente claro ou não identificou oportunidades de melhoria, o que também serve como referência de controle para análise comparativa com os demais grupos.

Grupo C: realizou modificações com foco no enriquecimento das funcionalidades descritas no prompt. Foram acrescentados detalhes como "formulário de registro" e "upload de foto", o que contribui para uma melhor compreensão das funcionalidades esperadas. Essa modificação sugere um passo adicional na construção do prompt, atuando como uma forma de alinhamento entre a intenção do usuário e a interpretação do modelo, tornando explícito aquilo que poderia ser inferido de maneira ambígua ou incompleta.

Grupo D: incorporou Proto-Personas (João e Maria) ao prompt, oferecendo informações sobre comportamento e necessidades dos usuários finais. Essa adição promove um direcionamento mais empático e centrado no usuário, o que poderia auxiliar na criação de histórias mais aderentes ao contexto real de uso do sistema.

Grupo E: modificou o prompt original de forma significativa, fornecendo uma visão de produto mais detalhada, incluindo funcionalidades como criação de salas, notificações e distribuição automática de papéis.

Além disso, houve um refinamento na definição dos papéis envolvidos, que passaram a ser descritos com mais clareza e granularidade: *Usuário geral*, *Moderador* (responsável por criar e gerenciar salas), *Participante* (que ingressa nas salas) e *Revisor*, com responsabilidades específicas dentro do sistema. O grupo também introduziu elementos novos ao processo de elicitação, como:

- Um prompt dedicado às regras de negócio, abordando aspectos como notificações automáticas e verificação de prazos.

- Modificação do template, incluindo campos adicionais como “Critérios de aceitação” e “Regras de negócio”, promovendo uma abordagem mais completa, como mostrado no quadro abaixo.
- Adaptação dos exemplos ao contexto da aplicação, como “Realizar Cadastro”, facilitando a compreensão e a geração de histórias mais precisas.

08

Eu vou lhe fornecer alguns exemplos de histórias de usuários bem formadas para lhe auxiliar na criação das histórias. Utilize-os para estruturar sua resposta, sempre que for solicitado a gerar o máximo de histórias de usuário possíveis.

Exemplo 1:**História de Usuário 3: Realizar Cadastro**

Descrição: Como usuário, eu quero ser capaz de realizar o cadastro na plataforma para poder utilizar todas as funcionalidades.

Critérios de aceitação:

- O sistema só deve permitir que o usuário siga com o cadastro caso todos os campos estejam preenchidos.
- O sistema deve verificar as informações do usuário para garantir sua validade.
- O sistema não deve permitir que o usuário siga com o cadastro se o e-mail ou cpf já estiver sendo utilizado em outra conta.

Regras de negócio:

- Ao fazer o cadastro, é obrigatório informar Nome de usuário, Nome completo, Email, Senha (no mínimo 8 caracteres alfanuméricos), Data de Nascimento e CPF válidos.

Outro aspecto interessante foi a duplicação dos prompts de contextualização: um voltado ao processo de debate de teses e outro à visão do produto em si (*TeseLab*), enriquecendo a construção do modelo antes da geração das histórias.

Em contrapartida, o grupo optou por remover o prompt com o framework QUS e as orientações adicionais fornecidas inicialmente. Essa decisão sugere que o grupo priorizou a clareza e a especificidade funcional, em detrimento da especificidade conceitual. No entanto, é importante considerar o aprendizado prévio do modelo durante a execução do prompt padrão, o que pode ter contribuído para o bom desempenho do modelo mesmo sem os critérios explícitos no prompt.

Grupo F: este grupo adotou uma abordagem mais refinada e iterativa, introduzindo modificações que demonstram um entendimento mais amplo dos desafios relacionados à atomicidade e à organização das histórias de usuários. A equipe promoveu uma clara estruturação do processo de geração, por meio da reformulação da Atividade 7 (última atividade do prompt padrão) e da criação de uma atividade nova, a Atividade 8.

Na Atividade 7 (modificada), o grupo reestruturou o prompt original para separar as histórias de usuários em duas categorias distintas:

- **Versão base:** histórias que refletem diretamente as funcionalidades que o usuário deseja para seu sistema;
- **Versão adicional:** histórias geradas com base em aspectos complementares que podem melhorar a experiência do usuário e a eficiência do sistema.

Essa separação estratégica facilita a triagem das histórias geradas, especialmente em cenários onde o escopo do sistema já está bem definido e o usuário deseja evitar um excesso de sugestões não planejadas. Ao fazer isso, o grupo proporcionou maior controle sobre os resultados da geração automatizada, aumentando sua utilidade prática.

Na Atividade 8 (nova), o grupo criou um novo prompt com o objetivo de refinar ainda mais a atomicidade das histórias de usuários. A motivação decorreu da observação de que o ChatGPT frequentemente gera histórias que incluem múltiplas funcionalidades em uma única estrutura. Para resolver isso, o grupo instruiu o modelo a revisar as histórias da versão base e gerar novas histórias mais granulares, com foco na separação de funcionalidades e na melhoria do nível de atomicidade.

Essa abordagem traz um ganho importante de qualidade ao promover a reavaliação das histórias geradas, favorecendo a modularidade dos requisitos produzidos. O processo proposto também reflete práticas reais de refinamento de backlog em metodologias ágeis, o que torna a estratégia particularmente relevante para projetos em estágios de especificação detalhada.

De forma geral, as modificações aplicadas pelos grupos evidenciam diferentes estratégias de refinamento do prompt, como aumento da empatia e detalhamento funcional. Esses ajustes revelam o potencial do US-Prompt como ponto de partida personalizável, embora também ressaltem os riscos de perda de qualidade quando alterações são feitas sem considerar os critérios de avaliação.

5.4.1.2 Comparação entre as Taxas de Sucesso dos Prompts utilizados:

Após analisar as taxas de sucesso dos Prompts Padrão e Modificado nos diferentes grupos, foi possível observar padrões interessantes que oferecem insights valiosos para o aprimoramento do US-Prompt.

De modo geral, ao combinar todos os resultados no gráfico da Figura 15, pode-se concluir que, para a maioria dos grupos, o Prompt Padrão obteve desempenho melhor ou igual ao Prompt Modificado, em termos de taxa de sucesso. Apenas em dois casos o US-Prompt Modificado obteve uma taxa de sucesso superior, nos Grupos E (P19) e F (P21). Em contrapartida, em três casos o Prompt Padrão superou o Prompt Modificado de forma significativa, nos Grupos A (P3), C (P9) e D (P15). Em um caso, os resultados foram idênticos (Grupo B, P5). A diferença mais pronunciada foi observada no Grupo C (P9), onde o Prompt Padrão obteve 89,29%, enquanto o Prompt modificado caiu drasticamente para 65,71%.

Essa análise revela a importância das modificações feitas pelos participantes, mas também destaca que elas precisam ser mais refinadas. Embora as melhorias propostas tenham gerado aumentos no desempenho em alguns casos, em outros, elas resultaram em uma diminuição da taxa de sucesso. Isso sugere que as alterações não foram uniformemente benéficas e precisam ser compreendidas e ajustadas conforme o contexto e os tipos de histórias geradas.

Ao testar o US-Prompt com um número maior de histórias e diferentes contextos, foi possível identificar mais erros e áreas que precisam ser melhoradas, especialmente no critério Bem-Formada. As diferenças nas taxas de sucesso indicam que as melhorias feitas pelos participantes oferecem oportunidades para aprimorar o US-Prompt, visando um desempenho mais consistente e de qualidade.

A Figura 15 mostra a maior pontuação em QUS de cada grupo na Fase 1, com o prompt padrão, representando o melhor desempenho possível (“limite superior”), ao lado da pontuação correspondente na Fase 2, com prompt modificado. Essa visão integrada evidencia, ao mesmo tempo, o potencial máximo de cada abordagem de prompting e as diferenças de desempenho observadas na prática, revelando tanto sua eficácia quanto sua variabilidade.

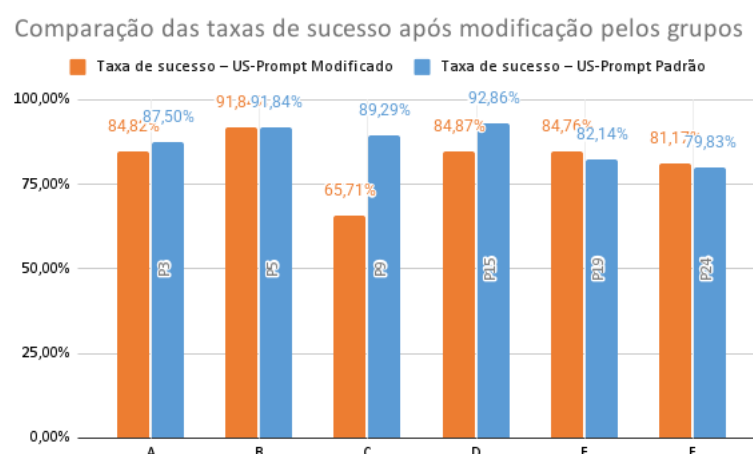


Figura 15 – Comparação das taxas de sucesso entre o Prompt Padrão e o Prompt Modificado nos seis grupos avaliados.

As modificações realizadas revelam diferentes estratégias para o refinamento da performance da geração automática. Enquanto alguns grupos apostaram na inclusão de contexto e detalhamento funcional (Grupos C, D, E), outros buscaram aumentar a clareza da tarefa (Grupos A, F) ou reforçar a qualidade técnica (Grupo F com foco na atomicidade). Essas estratégias refletem o nível de compreensão dos grupos sobre como o design do prompt influencia diretamente na saída do modelo.

Além disso, as análises comparativas entre as versões modificadas e os resultados das histórias geradas permitem inferir que intervenções mais estruturadas, como as realizadas pelos Grupos D, E e F, tendem a produzir histórias mais alinhadas com os critérios de qualidade definidos.

Sobre os grupos que apresentaram uma taxa de sucesso superior à do prompt padrão, no Grupo E, a melhoria está relacionada à resposta que introduziu uma visão de produto mais completa, definição clara de papéis (usuário, moderador, participante) e a inclusão de campos como critérios de aceitação e regras de negócio, além de dois prompts separados para contextualizar o produto. Já no Grupo F, os avanços se deram por refinamentos orientados à qualidade estrutural das histórias de usuários e vieram de duas respostas distintas: uma que separou histórias com base em funcionalidades fornecidas e outras geradas automaticamente (versão base e versão adicional), e outra que propôs um novo prompt para refinar a atomicidade das histórias, levando à criação

de uma versão definitiva mais coesa. O gráfico da Figura 16 mostra um resumo das estratégias adotadas por cada grupo.

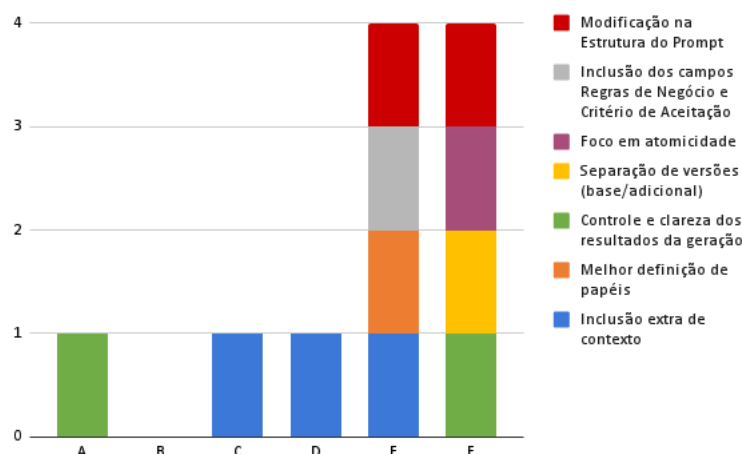


Figura 16 – Estratégias de melhorias adotadas por cada grupo.

Esses achados são particularmente relevantes para a viabilização do emprego LLMs na engenharia de requisitos, pois mostram que nem toda personalização leva automaticamente a resultados melhores. Um refinamento eficaz do prompt requer compreensão dos critérios de qualidade e validação contínua dos resultados.

Abordagens que combinam clareza, contexto e reestruturação orientada à qualidade, mesmo quando aplicadas de forma isolada entre os grupos, demonstraram maior potencial para guiar o modelo na geração de histórias mais úteis e aderentes às boas práticas da engenharia de requisitos.

Esse processo de refinamento contínuo é crucial para garantir que as alterações realmente agreguem valor ao desempenho do modelo, e essas observações servirão como base para o próximo ciclo de aprimoramento do US-Prompt.

5.4.1.3 Análise Estatística dos Resultados

Para avaliar a eficácia na geração das histórias de usuários utilizando os Prompts Padrão e Modificado, foi realizada uma análise estatística com o software JASP, versão 0.19.3¹. A eficácia média das histórias de usuários geradas com o Prompt Padrão foi de 87,24%,

¹ Disponível em: <<https://jasp-stats.org/>>

enquanto a eficácia média com o Prompt Modificado foi de 82,20% conforme mostrado na Tabela 6.

Tabela 6 – Resultados estatísticos sobre a qualidade das histórias de usuários com Prompts Padrão e Modificado

	Prompt Padrão	Prompt Modificado
Mediana	88.395	84.790
Média	87.243	82.195
Desvio Padrão	5.252	8.792
P-value de Shapiro-Wilk	0.498	0.096

Ao comparar as médias da qualidade das histórias de usuários geradas com prompts padrão e modificado, observou-se que o Prompt Padrão apresentou uma média de 87,24%, enquanto o Prompt Modificado alcançou uma média de 82,20%.

Essa diferença nas médias sugere que o Prompt Padrão foi mais eficaz na geração de histórias de usuários, indicando uma certa vantagem na utilização prompt padrão no contexto do estudo, mas para determinar se essa diferença é significativa, foi realizada uma análise estatística. Os resultados obtidos permitirão verificar se a diferença observada nas médias reflete uma variação significativa ou se pode ser atribuída ao acaso, reforçando ou refutando a hipótese inicial sobre a qualidade das histórias dos prompts analisados.

O teste de Shapiro-Wilk foi aplicado para verificar a normalidade dos dados, um pré-requisito para a aplicação de testes paramétricos, como o Teste T Pareado, garantindo a validade estatística dos resultados.

Os resultados do teste de normalidade indicaram que os dados apresentam uma distribuição normal, com os valores de p sendo 0,498 para o Prompt Padrão e 0,096 para o Prompt Modificado. Como ambos os valores de p são superiores ao nível de significância estabelecido ($\alpha = 0,05$), foi possível empregar o Teste T Pareado (Paired T-test) para comparar as taxas de sucesso entre os dois tipos de prompts, considerando cada participante individualmente.

O teste T obteve um valor de ($p = 0.264$) indicando que a diferença nas médias não é estatisticamente significativa, considerando o nível de significância estabelecido de ($\alpha = 0,05$). Assim, a hipótese nula (H_0) não foi rejeitada, indicando que ambos

os prompts geram histórias de usuário de qualidade semelhante. E, embora o Prompt Padrão tenha apresentado uma eficácia média ligeiramente superior ao Prompt Modificado, essa diferença não reflete em uma vantagem estatisticamente robusta entre os dois métodos.

Além disso, os desvios padrão foram de 5.252% para o Prompt Padrão e 8.792% para o Prompt Modificado. O Prompt Modificado apresenta um desvio padrão maior, o que indica uma maior variabilidade nos resultados em torno da média, quando comparado ao Prompt Padrão, com menor desvio. Isso sugere maior consistência nos resultados do Prompt Padrão, gerando respostas mais estáveis.

Boxplots

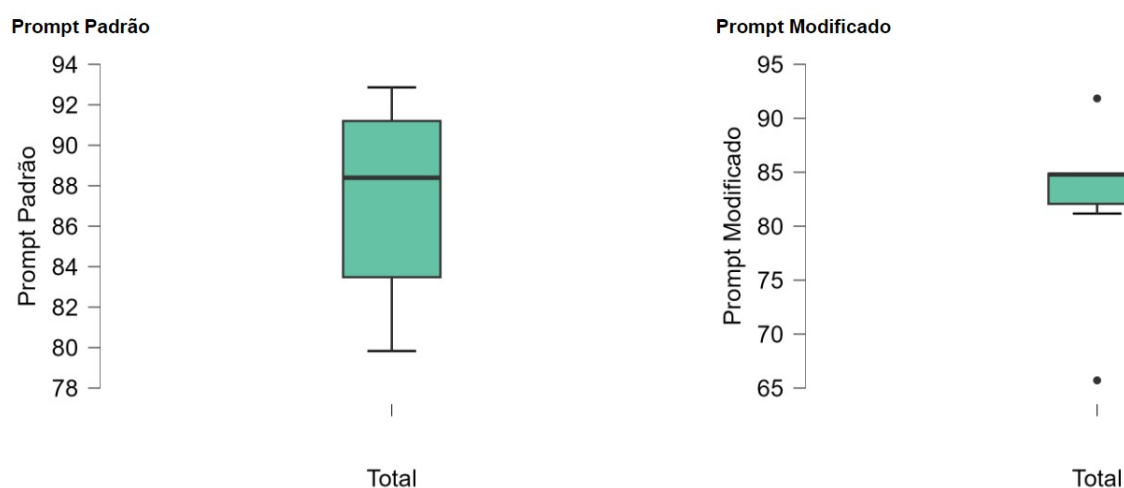


Figura 17 – Gráfico de boxplot comparando a distribuição dos resultados.

O gráfico da Figura 17, mostra os boxplots, destacando a mediana, que é representada pela linha central dentro da caixa de cada grupo. Para o Prompt Padrão, a mediana é de 88.395, enquanto para o Prompt Modificado, a mediana é de 84.790.

A seguir, é apresentado gráfico raincloud (Figura 18) que agrupa gráficos de densidade, boxplots e pontos de dados individuais. A distribuição das taxas de sucesso no gráfico de densidade mostra que o "Prompt Padrão" apresenta uma amplitude maior, com maior concentração de dados em torno de valores próximos a 90%. Em contraste, o "Prompt Modificado" exibe uma distribuição mais estreita, indicando menor variabilidade e uma média mais baixa. O pico mais acentuado da densidade do

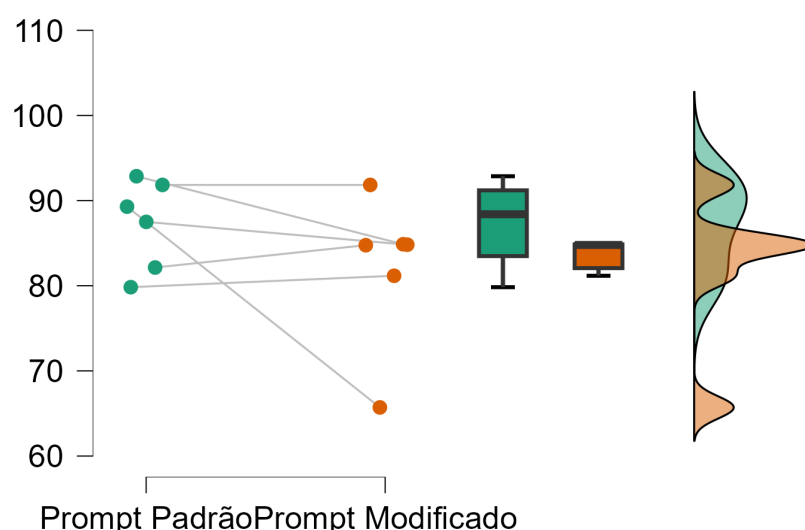


Figura 18 – Gráfico apresentando indicadores por meio de representações visuais: gráficos de densidade, boxplots e pontos de dados individuais.

"Prompt Modificado" sugere maior consistência nos resultados, pois os valores estão mais agrupados. No entanto, essa maior consistência não implica necessariamente um desempenho melhor, já que os dados estão concentrados em uma faixa inferior. A forma mais "achatada" e simétrica da distribuição do "Prompt Padrão" se aproxima de um padrão normal, mas reflete uma menor variação.

O boxplot verde, representando o "Prompt Padrão", revela uma mediana mais alta e um intervalo interquartil mais amplo, indicando que a maior parte dos dados está em uma faixa superior de sucesso. Por outro lado, o boxplot laranja, referente ao "Prompt Modificado", apresenta uma mediana mais baixa e uma dispersão menor. Embora o "Prompt Modificado" mostre uma variação reduzida, a menor dispersão observada, com um desvio padrão de 8,792, reforça a consistência dos resultados, mas em um nível de desempenho inferior.

As linhas que conectam os pontos no gráfico indicam a trajetória de desempenho dos mesmos participantes entre as duas versões de prompts. Na maioria dos casos, observou-se uma queda no desempenho ao migrar para o "Prompt Modificado", sugerindo que essa mudança não resultou em maior eficácia na geração das histórias de usuários.

O "Prompt Padrão" demonstrou um desempenho consistentemente superior, tanto na média quanto na dispersão dos dados, com menos quedas bruscas de desempenho individual. Já o "Prompt Modificado", apesar de apresentar maior consistência nos resultados, não só possui uma mediana mais baixa, como também sua distribuição está mais concentrada em valores mais baixos. Com base nas métricas observadas, o "Prompt Padrão" se mostrou mais eficaz na geração de histórias de usuários.

5.4.2 Avaliação da Aceitação e Percepções dos Participantes

O questionário de acompanhamento foi formulado para coletar dois principais tipos de informações sobre facilidade de uso da técnica proposta e identificar oportunidades de melhoria. É importante destacar que não foram realizadas entrevistas, uma vez que elas tendem a levar os participantes a restringirem informações, especialmente sobre as desvantagens percebidas. Assim, ao utilizar questionários, espera-se permitir que os participantes expressem seus sentimentos com maior liberdade. Além disso, considerando o número de participantes, o uso de questionários mostrou-se uma abordagem mais viável para a obtenção de dados qualitativos.

Quanto ao primeiro tipo de dado, utilizamos as respostas da primeira parte do questionário, construído com base no Modelo TAM3 ([VENKATESH; BALA, 2008](#)). Foram utilizados três construtos principais: Utilidade Percebida, Facilidade de Uso Percebida e Intenção de Uso Futuro. As afirmações foram elaboradas para captar o grau de concordância dos participantes em relação a esses três aspectos.

Para essa parte do questionário, utilizou-se a escala Likert, proposta por [Likert \(1932\)](#), amplamente adotada em pesquisas sociais e comportamentais por sua capacidade de mensurar atitudes e percepções de maneira estruturada e interpretável. As respostas foram fornecidas em uma escala de cinco níveis: Concordo fortemente, Concordo, Não concordo nem discordo, Discordo e Discordo fortemente. A inclusão de um nível intermediário teve como objetivo permitir que os participantes expressassem atitudes genuinamente neutras. Essa decisão é respaldada por [Kankaraš e Capecchi \(2025\)](#), que demonstraram que a presença de um ponto neutro pode melhorar as propriedades

psicométricas das escalas. Segundo os autores, tanto a confiabilidade quanto a validade tendem a aumentar quando mais opções de resposta são disponibilizadas, sugerindo que o ponto médio pode ser útil para captar nuances nas percepções dos respondentes.

Para compreender as respostas às questões baseadas no modelo TAM3, também incluímos uma pergunta aberta no questionário de acompanhamento. Essa pergunta foi apresentada após as escalas Likert, com o objetivo de captar a opinião dos participantes sobre o prompt utilizado na geração das histórias de usuários e como ele poderia ser modificado.

A seguir, são apresentadas as distribuições das respostas, por meio de gráficos de barras empilhadas horizontais, às afirmações baseadas nos três construtos do modelo TAM. Cada gráfico representa a frequência de respostas em uma escala de Likert de 5 pontos, variando de Discordo Fortemente a Concordo Fortemente.

As respostas apresentadas no gráfico da Figura 19 indicam uma tendência positiva quanto à facilidade de uso do prompt. A maioria dos participantes concordou fortemente ou concordou com as afirmações relacionadas à clareza da interação (F1), ao baixo esforço cognitivo necessário (F2), à facilidade geral de uso (F3) e à simplicidade para utilizar o prompt na geração de histórias (F4). Ainda que em menor número, há respostas indicando neutralidade e discordância, especialmente em (F2) e (F4), sugerindo que alguns usuários enfrentaram barreiras pontuais ao interagir com o prompt.

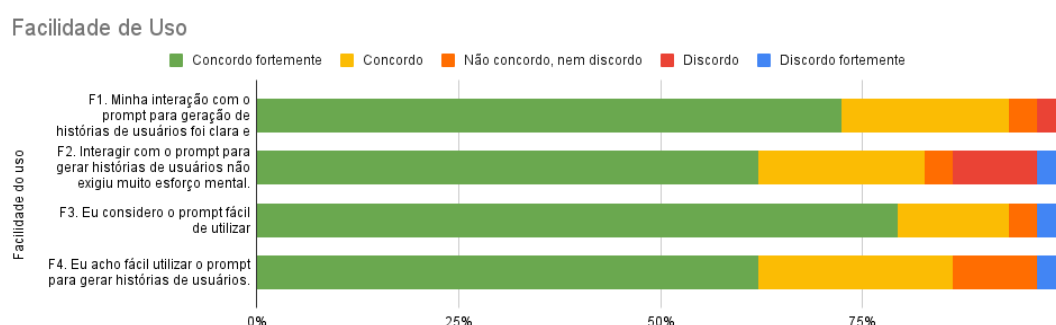


Figura 19 – Distribuição das respostas para o construto Facilidade de Uso Percebida.

As respostas no gráfico da Figura 20 também revelam uma percepção positiva quanto à utilidade do prompt. Grande parte dos participantes concordou fortemente que o uso do prompt melhora seu desempenho (U1), produtividade (U2) e eficácia na

geração de histórias de usuários de qualidade (U3). Entretanto, comparado ao construto de facilidade de uso, nota-se um aumento na proporção de respostas neutras em todas as afirmações de utilidade, especialmente em U3, indicando que embora o prompt seja percebido como fácil de usar, sua aplicabilidade não foi evidente para todos os participantes.

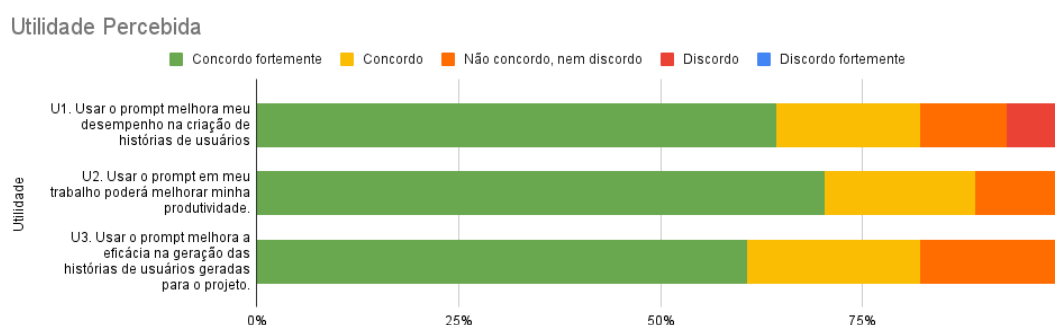


Figura 20 – Distribuição das respostas para o construto Utilidade Percebida.

No gráfico referente ao construto Intenção de Uso Futuro (Figura 21), observou-se que a resposta neutra manteve-se elevada e as discordâncias aumentam em I3 (“pretendo usar nos próximos meses”), o que pode indicar uma menor disposição em utilizar a técnica em contextos reais. Uma possível explicação para essa resposta é que, embora a ferramenta tenha sido considerada útil e fácil de usar, há incerteza prática sobre o uso recorrente no curto prazo, possivelmente porque muitos participantes não têm oportunidades de gerar histórias de usuários com frequência, em seu dia a dia. Assim, mesmo que a experiência tenha sido positiva, a intenção de uso futuro pode ter sido impactada pela falta de aplicabilidade prática imediata da técnica no contexto profissional dos respondentes, reforçando a ideia de que a intenção é elevada em cenários hipotéticos, mas diminui quando se considera a aplicação imediata.

Em suma, apesar de o prompt ser considerado claro, útil e fácil de usar, sua aplicabilidade concreta ainda não está totalmente consolidada para todos os usuários, especialmente em termos de uso prático no dia a dia profissional.

Ao final do questionário, incluiu-se a seguinte pergunta aberta: “Você tem comentários ou sugestões de como melhorar o Prompt para Histórias de Usuários?”, com o objetivo de captar percepções qualitativas sobre possíveis melhorias. As respostas

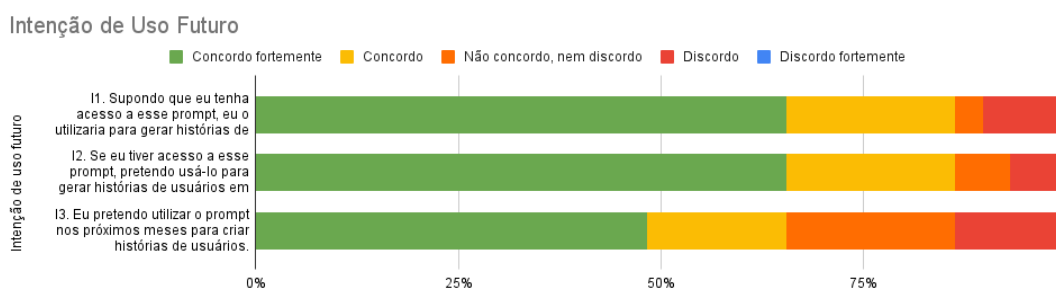


Figura 21 – Distribuição das respostas para o construto Intenção de Uso Futuro.

à pergunta aberta foram fundamentais para contextualizar e interpretar os resultados quantitativos obtidos pelas escalas do modelo TAM3. Enquanto os gráficos e as médias das escalas (Utilidade Percebida, Facilidade de Uso e Intenção de Uso) forneceram uma visão geral sobre a aceitação do método proposto, a pergunta aberta permitiu identificar nuances e justificativas subjetivas para as avaliações atribuídas pelos participantes.

Os comentários em formato livre revelaram não apenas críticas diretas ao funcionamento do prompt, mas também expectativas não atendidas, frustrações com resultados, e sugestões práticas de melhoria.

Dessa forma, a análise da pergunta aberta revelou-se essencial para compreender aspectos da experiência dos usuários que não seriam captados exclusivamente por indicadores quantitativos. A seguir, detalhamos as principais percepções negativas e positivas relatadas pelos participantes.

5.4.2.1 Percepções Negativas

Em relação à **utilidade percebida** do prompt, a maioria dos participantes concordou que ele é útil. No entanto, alguns relatos indicam experiências divergentes.

Os participantes P22 e P21 discordaram que a interação com o prompt exigiu pouco esforço mental.

“... Eu senti que eu mesmo já estava escrevendo as user stories...” — P22

O participante P21 também discordou quanto à **facilidade de uso** do prompt e à simplicidade de gerar histórias de usuários com ele, indicando uma experiência mais

negativa de forma geral.

Além disso, os participantes P19, P21 e P17 discordaram que utilizariam o prompt para gerar histórias de usuários, mesmo que tivessem acesso a ele. P22 e P17 declararam não pretender usar o prompt em projetos futuros, enquanto P22, P19 e P17 afirmaram não ter intenção de utilizá-lo nos próximos meses para criação de histórias de usuários. Esses dados apontam para uma **intenção de uso futuro** mais baixa entre esses participantes, apesar da experiência prévia.

“... Informei dois atores (professor, aluno) e nas histórias de usuário geradas o chat adicionou um ator extra (administrador)...” — P17

O participante P17 destacou como negativo o fato de o ChatGPT-4o ter extrapolado os elementos informados, ao incluir um ator não solicitado. Essa percepção pode ter contribuído para a sua avaliação negativa tanto em relação à utilidade quanto à intenção de uso da ferramenta, especialmente se a aderência às instruções fornecidas for um fator crítico no contexto de uso do participante.

O P2, apesar de ter atribuído avaliações positivas nos construtos de Facilidade de Uso e **Utilidade**, relatou uma experiência negativa relacionada à forma como o prompt interpretou os dados fornecidos:

“...A gente colocou ‘pessoas que fazem pedidos’ e a IA colocou ‘Como um Pessoas que fazem o pedido, Eu quero ...’” — P2

Mesmo mantendo uma opinião neutra quanto à intenção de uso nos próximos meses, essa experiência pode indicar uma frustração com a literalidade da resposta gerada, sugerindo que pequenos desvios na formulação pelo modelo afetam a confiança do usuário na ferramenta.

Já o P19, destacou um ponto importante sobre a inconsistência entre versões do modelo:

“...Existem muitas diferenças dependendo da versão do ChatGPT e até entre as mesmas versões...” — P19

Essa percepção pode ter influenciado sua avaliação negativa tanto em relação à utilidade percebida quanto à intenção de uso, pois a falta de previsibilidade nos resultados compromete a confiança e dificulta a adoção em cenários reais.

No entanto, P15 complementa que algumas histórias poderiam ser quebradas em mais histórias, apontando uma limitação na atomicidade e na granularidade dos resultados. Isso sugere que, embora útil como etapa inicial, o prompt ainda demanda retrabalho para alcançar o nível de refinamento necessário em contextos mais exigentes, embora isso não tenha comprometido sua avaliação geral.

“...Acredito que algumas poderiam ser quebradas em mais histórias do usuário, mas servem como uma bom início.” — P15

O participante P22, ao relatar o sentimento de estar fazendo o trabalho sozinho, expressa a percepção de que o prompt não estava agregando valor, gerando resultados muito parecidos com os que ele próprio faria. Essa percepção de baixa utilidade, somada à frustração por não perceber ganho de eficiência, contribui para uma tendência de não adoção futura. Já o participante P17 destaca a adição de conteúdo não solicitado, onde o modelo extrapolou o escopo da entrada fornecida. Embora isso possa parecer uma contribuição útil, essa extrapolação pode ser percebida como perda de controle sobre o resultado. Em contextos que exigem precisão e aderência estrita às instruções, tal comportamento pode ser interpretado como um erro.

P15 também observa que o prompt não gerou histórias suficientemente atômicas, o que compromete a qualidade dos resultados pela baixa aderência aos critérios de boas histórias de usuário. O participante P19 concentrou sua crítica na inconsistência entre versões do modelo, apontando que essa variabilidade prejudica a reprodutibilidade dos resultados. Por fim, o P2 relata um erro de interpretação literal, que resultou em uma estrutura textual incoerente ao não adaptar corretamente os papéis informados.

Em suma, as experiências relatadas convergem em alguns pontos centrais: falta de percepção de valor agregado e ausência de ganho de eficiência; baixa previsibilidade e controle sobre o comportamento do prompt; erro na aderência aos critérios de qualidade; e falhas na geração que comprometem a confiança no resultado. Essas experiências indicam que a percepção de facilidade de uso, por si só, não garante a adoção

da ferramenta. Quando o usuário não percebe melhora concreta em sua produtividade ou qualidade do trabalho, a utilização tende a ser desconsiderada. Para que o uso seja recorrente e significativo, é necessário que o prompt gere valor claro e percebido, seja confiável e previsível, e respeite os limites e o controle definidos pelo usuário.

5.4.2.2 Percepções Positivas

Apesar das críticas, vários participantes relataram experiências positivas com o uso do prompt, destacando ganhos de produtividade, simplicidade no uso e a utilidade como ponto de partida para a criação de histórias de usuário.

O participante P4, por exemplo, avaliou de forma direta e positiva:

“... ele foi eficiente e intuitivo” — P4

A percepção de que o prompt contribuiu para acelerar o processo também foi destacada por P9:

“Acredito que otimizou o tempo e as respostas do chat foram satisfatórias para o nosso sistema...” — P9

P5 reforça a utilidade da ferramenta, destacando tanto sua facilidade de uso quanto a qualidade das histórias geradas:

“Achei fácil e produtivo de utilizar ... O Chat gerou boas histórias de usuário com ele” — P5

O apoio inicial fornecido pelo prompt também é reconhecido por P15, que destacou sua capacidade de gerar uma base inicial significativa de histórias:

“Utilizar prompts facilita a criação de histórias do usuário, servindo como um pontapé inicial...” — P15

Ainda que P15 tenha sugerido que algumas histórias poderiam ser melhor fragmentadas, o uso do prompt foi considerado produtivo e eficaz na etapa inicial de geração.

P8 também reconheceu um ganho de esforço cognitivo reduzido, sendo indicador positivo de usabilidade:

“... no caso desse eu só copieei e coleii, então não me esforcei mentalmente como normalmente” — P8

5.4.3 Recomendações de Melhoria

Vários participantes sugeriram melhorias que poderiam tornar o uso do prompt ainda mais eficaz. A distribuição das categorias de melhorias sugeridas pelos participantes revela três áreas principais de aprimoramento:

- Exemplificação mais clara de papéis e contextos (P2, P17);
- Ajuste no formato de entrada (P22);
- Granularidade insuficiente em algumas histórias (P15).

A seguir, são apresentadas as recomendações feitas por cada participante, organizadas conforme as categorias identificadas na lista acima:

O P2 recomenda a inclusão de exemplos mais claros de como informar corretamente os papéis dos usuários:

“... poderia deixar um exemplo de como colocar os usuários... aí podemos nos atentar melhor como devemos colocar” — P2

P17 reforça essa recomendação ao sugerir o fornecimento prévio de um modelo mais detalhado de como descrever o contexto do sistema:

“Poderia ser fornecido previamente um modelo de como o contexto do sistema deve ser escrito para inserir no prompt” — P17

Essas sugestões indicam que, para muitos participantes, a efetividade do método está diretamente relacionada à qualidade das instruções fornecidas antes da geração.

Elas reforçam a importância de oferecer exemplos mais claros, contextualizados e completos, a fim de reduzir ambiguidades e aumentar a aderência do modelo às intenções do usuário.

P22 sugere mudanças no formato da entrada para o prompt, com o objetivo de evitar a sensação de estar fazendo o trabalho sozinho:

“Seria melhor se a atividade 3 fosse preenchida com um texto corrido e não uma lista de ações...” — P22

Essa sugestão indica que a estrutura do input influencia diretamente na percepção de valor do resultado gerado. Ao receber ações fragmentadas, o modelo pode parecer apenas estar organizando ou repetindo algo que o próprio usuário já estruturou.

Já P15 propõe uma melhoria na granularidade dos resultados, observando que algumas histórias geradas poderiam ser mais atômicas:

“...Acredito que algumas poderiam ser quebradas em mais histórias do usuário...” — P15

A facilidade de uso por si só não garante a adoção do prompt. A falta de benefícios percebidos, problemas de qualidade e imprevisibilidade reduzem a confiança. Para uso repetido, o prompt deve oferecer valor, confiabilidade e atender às expectativas do usuário. Essas considerações indicam melhorias desejáveis tanto na entrada quanto na saída do prompt, visando um maior alinhamento com as expectativas dos usuários

5.5 Limitações do Estudo

Toda pesquisa tem limitações de seus resultados (WOHLIN et al., 2012). Esta seção discute as principais limitações consideradas neste estudo:

Validade de Constructo: Embora a avaliação tenha sido baseada no framework Quality User Story (QUS), que fornece critérios robustos e bem estabelecidos, ela pode não captar nuances em situações mais complexas ou específicas de determinado contexto. Para mitigar essa limitação, dois pesquisadores independentes revisaram as histórias e as avaliações.

Validade Interna: LLMs, como o ChatGPT, são conhecidos por gerar conteúdo incorreto ou fabricado (alucinações) segundo [Huang et al. \(2024\)](#). Para minimizar essa limitação, um especialista em requisitos realizou uma revisão final de todas as histórias geradas de forma automatizada antes da avaliação. Além disso, o estudo não controlou completamente fatores externos, como a experiência dos participantes, familiaridade com as ferramentas ou o ambiente específico em que as tarefas foram realizadas, o que pode ter influenciado os resultados.

Validade Externa: A amostra foi composta majoritariamente por estudantes de Engenharia de Software. Embora isso possa limitar a generalização dos resultados, pesquisas anteriores sugerem que estudantes podem representar adequadamente profissionais da indústria em estudos empíricos em engenharia de software ([HÖST; REGNELL; WOHLIN, 2000](#); [SALMAN; MISIRLI; JURISTO, 2015](#)).

Validade de Conclusão: O tamanho relativamente pequeno da amostra limita o poder estatístico da análise e aumenta a possibilidade de viés. Isso foi parcialmente mitigado por meio da avaliação com revisores independentes.

5.6 Conclusão

Este estudo experimental investigou a qualidade das histórias de usuários geradas e a aceitação da técnica de US-Prompt na geração automatizada de histórias de usuários, aplicada a diferentes contextos, com o uso de modelos de linguagem. A partir da condução estruturada do experimento em três fases, envolvendo 24 participantes organizados em grupos, foi possível observar um volume de 457 histórias geradas, distribuídas entre diferentes versões de prompts.

Os resultados das análises demonstraram que o US-Prompt, mesmo em sua forma padrão, é altamente eficaz para a geração de histórias de usuários aderentes aos critérios de qualidade definidos. A análise quantitativa revelou que 87,5% dos conjuntos de histórias gerados com o Prompt Padrão atingiram taxas de sucesso superiores a 75%, evidenciando sua consistência mesmo antes de qualquer refinamento. Dessa forma, as histórias de usuários geradas por IA podem ser consideradas boas o suficiente, aten-

dendo de maneira satisfatória aos critérios de qualidade estabelecidos pelo framework QUS.

Entretanto, ao permitir que os grupos participantes otimizassem o prompt, observou-se que as melhorias propostas nem sempre resultaram em ganhos de qualidade. Em apenas dois dos casos analisados, o Prompt modificado superou o desempenho do Prompt Padrão. Por outro lado, em três casos, a performance caiu de forma significativa após as modificações. Esses dados indicam que, embora haja potencial para personalizações que aumentem a aderência às necessidades contextuais de cada grupo, ajustes mal planejados podem comprometer a estrutura e os critérios de qualidade das histórias geradas.

Além disso, a variação nos resultados sugere que o sucesso das modificações pode se beneficiar em algum grau da clareza, especificidade e aderência aos princípios de atomicidade, empatia com o usuário e alinhamento com práticas ágeis. Estratégias como a inclusão de Proto-Personas, visão de produto detalhada e separação entre funcionalidades base e adicionais demonstraram maior potencial de melhoria, quando comparadas a alterações superficiais.

Portanto, a comparação entre os dois métodos revela que o US-Prompt é eficaz em sua forma original, servindo como um ponto de partida eficiente. As tentativas de otimização podem aumentar ainda mais a qualidade das histórias, desde que sejam conduzidas com atenção às boas práticas de engenharia de requisitos e aos critérios de avaliação definidos. Tais achados reforçam a necessidade de diretrizes mais claras para o refinamento de prompts em contextos de elicitação de requisitos, sobretudo em ambientes mediados por LLMs.

A análise conjunta dos dados qualitativos revela uma compreensão abrangente sobre a aceitação do método proposto para a geração automatizada de histórias de usuários. Os resultados obtidos por meio da escala TAM indicam níveis satisfatórios de aceitação em termos de facilidade de uso percebida e utilidade percebida. A maioria dos participantes avaliou positivamente a experiência de interação com o prompt, destacando que o sistema é fácil de entender, requer pouco esforço mental e permite obter resultados de forma rápida e prática.

Essa percepção é reforçada pelas respostas, nas quais os termos “simples”, “fácil” e “intuitivo” foram frequentemente utilizados. A clareza do prompt e a organização das instruções foram apontadas como elementos facilitadores, especialmente por parte dos participantes com menor familiaridade. Tais evidências sugerem que o sistema atende bem ao princípio de facilidade de uso, um dos principais determinantes da aceitação segundo o modelo TAM.

Contudo, apesar da maioria considerar que o sistema pode melhorar a produtividade e apoiar a elaboração de histórias de usuários, alguns participantes relataram problemas de qualidade nas respostas geradas pelo modelo, como falta de aderência ao formato esperado, generalizações excessivas, interpretações imprecisas do contexto e inconsistências na saída. Esses fatores reduziram a percepção de confiabilidade, especialmente em cenários que demandam alto grau de precisão ou domínio técnico.

Esses achados indicam que, embora o sistema seja promissor como suporte inicial à especificação de histórias, ainda não foi plenamente validado como substituto ao processo manual, sobretudo para tarefas críticas.

Por fim, a intenção de uso futuro mostrou-se moderada, com parte dos participantes demonstrando interesse em continuar utilizando a abordagem como ferramenta complementar, enquanto outros expressaram cautela. Essa hesitação está associada à necessidade de um controle mais rigoroso sobre a qualidade das respostas e à adaptação do sistema às necessidades específicas do domínio.

Os dados sugerem que há espaço para evolução, especialmente no refinamento do comportamento do modelo frente ao prompt e na adaptação do método às necessidades específicas dos usuários. Com base nas observações feitas durante o experimento e nas contribuições dos participantes, identificamos um conjunto de recomendações para aprimorar o US-Prompt:

- **Reduzir a complexidade textual sem comprometer a clareza:** A leitura do prompt pode ser cansativa, especialmente para participantes com menos familiaridade com ferramentas de IA. Recomenda-se a simplificação da execução da tarefa.
- **Melhorar os exemplos fornecidos:** A presença de exemplos bem formados se mostrou essencial, mas há a possibilidade de que os participantes se restrinjam ao

campo de aplicação específico dos exemplos fornecidos, o que pode introduzir viés nos resultados, ao limitar a diversidade dos contextos explorados e comprometer a generalização das conclusões. Recomenda-se a diversificação dos exemplos, abrangendo diferentes contextos, para estimular a generalização.

- **Oferecer uma estrutura modular para o prompt:** Em grupos com domínios mais complexos, foi comum a necessidade de adaptar o prompt. Assim, propõe-se uma versão modular, com seções opcionais que podem ser ativadas conforme a complexidade do produto, promovendo flexibilidade sem sobrecarregar os usuários menos experientes.
- **Introduzir a reescrita iterativa das histórias:** Alguns participantes beneficiaram-se da revisão das histórias geradas, refinando a atomicidade. Recomenda-se a inclusão de instruções ou etapas que incentivem essa reescrita, por exemplo, orientando o modelo a dividir histórias muito amplas.
- **Integrar os critérios QUS de forma leve e funcional:** Em vez de apresentar os critérios completos aos usuários, o sistema deve avaliá-los automaticamente por meio de perguntas-guia para cada história gerada (ex.: “Está claro o papel e a ação?”, “A história é específica ou aborda múltiplas funcionalidades?”). Essa abordagem permite fornecer feedback imediato sobre a qualidade das histórias com base no framework, sem exigir esforço adicional dos usuários.

Os resultados obtidos indicam que a adoção do US-Prompt, especialmente quando associada a melhorias colaborativas no prompt, apresenta potencial para gerar histórias de usuários de maior qualidade, conforme avaliado pelos critérios do framework QUS. Além disso, a aplicação do modelo TAM revelou percepções positivas quanto à utilidade, facilidade de uso e intenção de uso da técnica, sugerindo a viabilidade de sua adoção em contextos reais.

Estudos anteriores sobre histórias de usuários geradas por LLMs indicam boa qualidade geral, mas destacam desafios relacionados à especificabilidade e aos detalhes técnicos (RAHMAN et al., 2024). Brockenbrough e Salinas (2024) reportam uma melhora de 88% nos critérios de testes de aceitação e no valor das histórias. Já Yamani, Baslyman

e Ahmed (2025) observaram que histórias elaboradas por humanos ainda superam as geradas por LLMs em alguns aspectos. Em nosso estudo, a taxa de sucesso no QUS atingiu 87,24% nos grupos de melhor desempenho, o que evidencia qualidade consistente e equilibrada.

Dessa forma, este estudo demonstra o potencial do US-Prompt como ponto de partida eficaz para a geração automatizada de histórias de usuários, integrando avaliação objetiva de qualidade e percepção dos usuários, e reforçando sua viabilidade como ferramenta robusta e adaptável.

Por fim, neste capítulo, apresentou-se a segunda iteração do Ciclo de Design, aprofundando a conexão com o Ciclo de Rigor da pesquisa. Para tanto, foram empregadas metodologias e fundamentações teóricas robustas no desenvolvimento, aplicação e avaliação dos prompts. As métricas de desempenho utilizadas foram essenciais para garantir a validade, a confiabilidade e a utilidade do artefato desenvolvido.

Em síntese, este capítulo detalhou o estudo experimental realizado com o artefato US-Prompt, focando na validação externa e na capacidade de generalização da abordagem em diferentes contextos de uso. Além disso, foram analisados aspectos relacionados à aceitação da técnica pelos usuários participantes. Com base nos resultados obtidos e nos problemas identificados, o capítulo seguinte apresenta uma proposta de ferramenta para o US-Prompt, visando aperfeiçoar seu desempenho e atenuar as limitações reveladas neste estudo.

6

3º CICLO DE DESIGN: FERRAMENTA WEB - STORY AI GENERATOR: CRAFTING USER STORIES

Este capítulo apresenta a ferramenta Story AI Generator, desenvolvida para operacionalizar a abordagem US-Prompt e apoiar, de forma guiada e simplificada, a geração automatizada de histórias de usuário sem exigir conhecimento avançado em engenharia de prompts. Além disso, descreve o estudo exploratório conduzido com profissionais de Engenharia de Software, com o objetivo de avaliar a ferramenta quanto à aceitação e percepção de uso. Os resultados obtidos fornecem evidências sobre a aplicabilidade prática da solução e seu potencial para apoiar o processo de especificação de requisitos na Engenharia de Requisitos.

6.1 Introdução

A ferramenta *Story AI Generator: Crafting User Stories with LLMs* foi desenvolvida a partir dos resultados obtidos nos capítulos anteriores, que investigaram formas de aumentar a eficácia e aprimorar a aplicação de prompts otimizados para a geração de histórias de usuários. Com o objetivo de apoiar engenheiros de requisitos, product owners e demais profissionais envolvidos na elicitação de requisitos, oferecendo geração automatizada de histórias de usuários de maior qualidade.

Integrando o US-Prompt, em sua segunda versão, apresentado na Seção 5.2 e disponível no Apêndice C, a ferramenta proporciona uma experiência prática e facilitada, acelerando a fase de especificação de requisitos. Isso permite que as equipes produzam histórias de forma mais consistente, estruturada e em menor tempo, beneficiando diferentes perfis do time de desenvolvimento.

Para avaliar sua aplicabilidade prática, torna-se igualmente importante coletar e analisar o feedback de profissionais com experiência no uso de histórias de usuários, cujo conhecimento é fundamental para uma avaliação mais aprofundada da qualidade das histórias geradas.

Nesse sentido, foi conduzido um estudo exploratório com especialistas, composto por profissionais de Engenharia de Software experientes no uso de user stories, com o propósito de coletar evidências sobre a aceitação da ferramenta e seu potencial de apoio às atividades de Engenharia de Requisitos. Durante o experimento, os participantes utilizaram a ferramenta Story AI Generator para criar histórias de usuários, e os resultados foram avaliados com base na experiência de uso e na aceitação da ferramenta, utilizando o modelo TAM3 (Technology Acceptance Model) (VENKATESH; BALA, 2008), considerando os construtos Utilidade Percebida (Perceived Usefulness), Facilidade de Uso Percebida (Perceived Ease of Use), Intenção de Uso Futuro (Behavioral Intention), Perceived Relevance (PR) e Perceived Output Quality (POQ).

A análise das percepções dos participantes foi conduzida a partir das respostas abertas do questionário e das entrevistas, empregando procedimentos da *Grounded Theory* (GT), um método qualitativo que permite a identificação de categorias emergentes diretamente fundamentadas nas falas dos participantes (GLASER; STRAUSS; STRUTZEL, 1968). Essa abordagem possibilitou compreender, de forma indutiva, os aspectos positivos, limitações percebidas e oportunidades de melhoria da ferramenta, sob a perspectiva dos especialistas. Os objetivos deste estudo são avaliar a aceitação da ferramenta por especialistas, identificar oportunidades de melhoria e fornecer insights práticos sobre sua aplicação no contexto da Engenharia de Requisitos.

Este capítulo está organizado em duas partes principais: na primeira, é apresentada a ferramenta *Story AI Generator*, incluindo suas funcionalidades, interface, fluxos de

interação e componentes técnicos; na segunda, é descrito o estudo exploratório conduzido com profissionais de Engenharia de Software, incluindo o desenho do experimento, os instrumentos de coleta de dados e a análise dos resultados obtidos.

6.2 Ferramenta Story AI Generator: Crafting User Stories

A ferramenta Story AI Generator, atualizada e iterativa, inclui instruções mais claras, exemplos adicionais, uma etapa de validação e mecanismos para reduzir alucinações. Tais mudanças visam simplificar a geração de user stories e alinhá-la com os padrões de qualidade. Nesse contexto, a ferramenta foi construída com os seguintes objetivos principais:

6.2.1 Objetivos da Ferramenta

A ferramenta Story AI Generator foi construída com os seguintes objetivos principais:

- Apoiar a geração automatizada de histórias de usuários com base em critérios de qualidade;
- Facilitar o uso do US-Prompt em contextos práticos, reduzindo a complexidade e promovendo redução de esforço;
- Organizar e orientar a entrada dos dados, guiando o usuário na inserção de informações relevantes;
- Permitir a customização da entrada, incluindo usuários, funcionalidades, contexto e nível de liberdade na geração (mais flexível ou mais restrita);
- Padronizar o formato das histórias segundo critérios QUS, incluindo uma etapa de validação integrada;
- Incorporar diretrizes de meta-prompting e few-shot prompting na estrutura do prompt utilizado;

6.2.2 Tecnologias Utilizadas

A Story AI Generator foi desenvolvida integralmente em Python, utilizando o framework Django para o backend e frontend.

O backend é responsável por processar requisições, gerenciar dados e integrar-se a APIs externas, como a da OpenAI, por meio de chamadas HTTP seguras. O frontend foi construído utilizando o sistema de templates do Django, fornecendo um formulário simples e responsivo para entrada de dados. Esse frontend é complementado por HTML5, CSS3 e JavaScript para melhor usabilidade.

Em relação ao serviço de hospedagem utilizado, a aplicação foi implantada utilizando o serviço da Amazon Web Services (AWS), permitindo com que a aplicação seja acessada via web de qualquer lugar e em qualquer dispositivo por meio do endereço eletrônico www.storyaigenerator.com.br. Foi também utilizado para controle de versão o Git e GitHub, assegurando rastreabilidade das alterações e colaboração eficiente.

6.2.3 Interface e Fluxo de Uso

A interface da Story AI Generator foi projetada para ser simples, intuitiva e de fácil navegação, permitindo que usuários de diferentes perfis consigam utilizar a ferramenta sem necessidade de treinamento prévio.

A seguir, são apresentadas imagens das principais telas do frontend, essas imagens visam demonstrar a organização visual, a disposição dos elementos e a clareza dos campos de entrada, contribuindo para uma experiência fluida e objetiva.

A Figura 22 apresenta um guia de uso da ferramenta, cujo objetivo é guiar o usuário no processo de geração de histórias de usuário. São apresentadas as principais etapas de inclusão de informações no formulário e também são apresentados exemplos simples de descrições de sistemas.

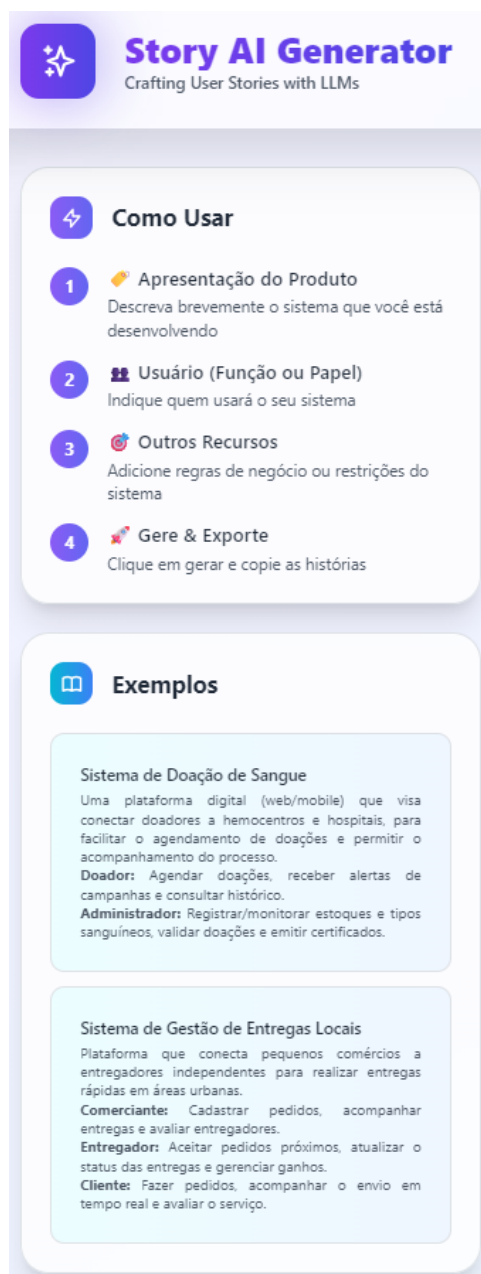


Figura 22 – Guia de uso da ferramenta Story AI Generator.

Logo abaixo na Figura 23 é exibido o principal formulário web da ferramenta, dividido nas seguintes seções:

- **Apresentação do Produto:** breve descrição do produto em questão.
- **Informações do Usuário:**
 - **Usuário (função ou papel):** função ou papel do usuário (ex.: *Administrador*, *Usuário comum*).

- **Descrição do Usuário:** breve descrição do perfil do usuário.
- **Ações do Usuário:** principais ações que o usuário deseja realizar.
- **Informações Adicionais:** ações adicionais que são necessárias para o funcionamento do sistema.

The image shows a web form titled "Apresentação do Produto" with a "Copiar Dados" button. It contains three main sections: "Informações do Usuário" with fields for user role, description, and actions; "Informações Adicionais" for additional criteria; and a large blue button at the bottom labeled "Gerar Histórias com IA".

Apresentação do Produto Copiar Dados

Qual é o propósito geral do sistema que você pretende desenvolver e qual problema ele busca resolver? (Visão do Produto)
Ex: Um sistema para facilitar a gestão de informações dentro de uma organização...

Informações do Usuário

Usuário (Função ou Papel)
Quem vai utilizar o sistema? Ex: Colaborador, Gestor...

Descrição do Usuário
Descreva quem é este usuário.

Ações do Usuário
O que esse usuário precisa fazer dentro do sistema? (Principais ações esperadas)
Ex: Cadastrar informações, consultar registros, atualizar dados, gerar relatórios...

+ Adicionar usuário

Informações Adicionais
Adicione critérios de aceitação, regras de negócio ou restrições do sistema...

Gerar Histórias com IA

Figura 23 – Formulário de entrada de dados da Story AI Generator.

Após o preenchimento, o usuário clica no botão "Gerar Histórias com IA", acionando o backend para processar os dados e enviar um prompt estruturado à API da OpenAI.

6.2.4 Geração das Histórias com ChatGPT

O backend, ao receber os dados do formulário, cria automaticamente o prompt, disponível no Apêndice C, e o envia à API do ChatGPT, a integração com a API da OpenAI é realizada via requisições HTTP autenticadas. O modelo retorna o máximo de histórias de usuário seguindo os critérios esperados.

Esse retorno é então renderizado diretamente na interface web, possibilitando que o usuário revise ou exporte as histórias geradas, como mostrado na Figura 24.

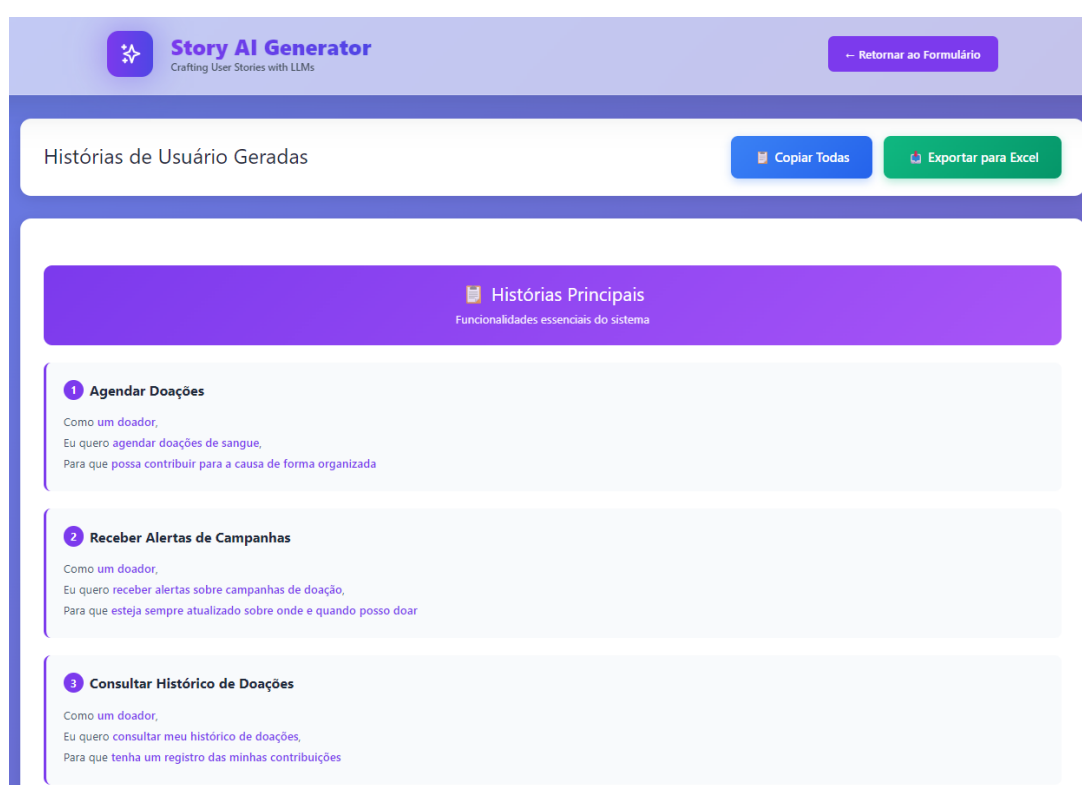


Figura 24 – Interface de saída com histórias geradas.

Dessa forma, a ferramenta Story AI Generator faz a operacionalização das técnicas desenvolvidas nesta pesquisa. Ao encapsular as melhores práticas de engenharia de prompts, aliadas ao poder dos LLMs, ela visa não apenas facilitar a adoção dessas abordagens por profissionais da área, mas também promover a geração de requisitos de maior qualidade, de forma mais rápida e com menos esforço cognitivo.

6.3 Estudo Experimental Exploratório com Profissionais de Engenharia de Software

6.3.1 Metodologia do Estudo

O estudo buscou responder à seguinte questão de pesquisa **QP: “Quais percepções, limitações e oportunidades de melhoria são identificadas pelos profissionais de Engenharia de Software a partir do uso da ferramenta Story AI Generator?”**. Para isso, foi realizado um estudo exploratório, seguindo as diretrizes metodológicas propostas por [Wohlin et al. \(2012\)](#), com profissionais de Engenharia de Software, no qual os participantes utilizaram a ferramenta e forneceram feedback sobre sua experiência de uso.

6.3.1.1 Desenho do Estudo

O estudo foi conduzido seguindo um protocolo estruturado, conforme as diretrizes de experimentação em Engenharia de Software propostas por [Wohlin et al. \(2012\)](#), de modo a garantir a sistematização das atividades e a confiabilidade dos dados coletados, considerando os seguintes aspectos: definição clara do objetivo do estudo, caracterização dos participantes, descrição dos instrumentos de coleta de dados, procedimentos de execução, bem como a estratégia de análise dos dados. Essas diretrizes foram adotadas com o intuito de garantir a replicabilidade do estudo e a validade dos resultados obtidos. O protocolo foi organizado nas etapas descritas a seguir.

6.3.1.1.1 Participantes: Os participantes foram recrutados presencialmente durante o Simpósio Brasileiro de Qualidade de Software (SBQS), o principal fórum nacional dedicado à Qualidade de Software, amplamente reconhecido por sua relevância para a Engenharia de Software no Brasil. O evento reúne profissionais de excelência da academia e da indústria, promovendo a troca qualificada de conhecimentos e experiências que impulsionam o desenvolvimento de métodos, práticas e soluções aplicadas, contribuindo para a inovação e para a melhoria contínua da qualidade de software no país. Esses aspectos, aliados à diversidade do público participante, caracterizam o SBQS

como um ambiente oportuno e apropriado para a condução de estudos exploratórios e avaliativos, favorecendo a coleta de percepções qualificadas sobre ferramentas e técnicas emergentes.

Os participantes foram selecionados com base em um critério de inclusão, segundo o qual deveriam possuir experiência prévia na criação ou utilização de histórias de usuários, seja em atividades acadêmicas, projetos de pesquisa ou no desenvolvimento de software em contextos profissionais. Ao todo, 11 participantes aceitaram participar da pesquisa e assinaram o termo de consentimento livre e esclarecido, com a opção de se retirar do estudo a qualquer momento, garantindo a ética e o voluntariado no processo.

6.3.1.1.2 Instrumentação: A instrumentação do estudo foi composta por quatro elementos principais:

- (i) O Termo de Consentimento Livre e Esclarecido (TCLE).
- (ii) A ferramenta Story AI Generator, utilizada para a geração automatizada de histórias de usuários¹;
- (iii) Um documento contendo cenários de exemplo, disponibilizado aos participantes durante a interação com a ferramenta;
- (iv) Um questionário de coleta de dados baseado no modelo TAM3, adaptado ao contexto do estudo;

Antes do início do experimento, todos os participantes receberam e assinaram o TCLE, garantindo o caráter voluntário da participação, o anonimato das respostas e o uso dos dados exclusivamente para fins acadêmicos.

O questionário de acompanhamento avalia a aceitação da técnica e identifica melhorias, utilizando como base teórica o modelo TAM3 (Technology Acceptance Model 3) (VENKATESH; BALA, 2008). Foram considerados cinco construtos principais: Utilidade Percebida, Facilidade de Uso Percebida, Intenção de Uso Futuro, Relevância Percebida e Qualidade de Saída Percebida. As afirmações associadas a cada construto

¹ Disponível em: <<https://www.storyaigenerator.com.br/>>. Story AI Generator.

foram formuladas de modo a captar o grau de concordância dos participantes em relação a esses aspectos.

Para a coleta das respostas quantitativas, utilizou-se a escala Likert ([LIKERT, 1932](#)) com respostas em uma escala de cinco pontos: Concordo fortemente, Concordo, Não concordo nem discordo, Discordo e Discordo fortemente.

Com o intuito de aprofundar a compreensão das respostas relacionadas aos construtos do TAM3, foi incluída, para cada construto, uma pergunta aberta de caráter qualitativo. Essas questões foram apresentadas após as afirmações em escala Likert, visando captar as percepções, justificativas e impressões dos participantes em relação às respostas fornecidas. Na sequência, são apresentados os gráficos que ilustram a distribuição das respostas às afirmações associadas aos cinco construtos do modelo TAM. O questionário está disponível no Apêndice [F](#).

Além do questionário, foi utilizado um roteiro de entrevista semiestruturada, elaborado com o objetivo de aprofundar a compreensão das percepções gerais dos participantes, bem como identificar limitações e oportunidades de melhoria relacionadas ao uso da ferramenta que não seriam plenamente capturadas por meio das questões baseadas nos construtos do TAM3.

6.3.1.2 Execução do Estudo

O estudo foi conduzido de forma presencial e seguiu a mesma sequência padronizada de execução para todos os participantes, com o objetivo de garantir a uniformidade do procedimento e reduzir possíveis vieses durante a coleta de dados.

6.3.1.2.1 Execução do Estudo Piloto: Antes da execução do estudo principal, foi conduzido um estudo piloto com dois participantes: um pesquisador doutor na área de Informática, com atuação em Engenharia de Software e Inteligência Artificial, e um profissional da indústria de software vinculado a um instituto de pesquisa e desenvolvimento tecnológico reconhecido. Ambos possuem experiência prévia no uso de histórias de usuários.

O objetivo dessa etapa foi avaliar a clareza e a compreensibilidade dos instru-

mentos utilizados, bem como do uso da ferramenta Story AI Generator. Adicionalmente, buscou-se identificar possíveis ambiguidades, interpretações divergentes e problemas relacionados à geração das histórias de usuários.

A realização do estudo piloto permitiu verificar se os participantes compreendiam corretamente as funcionalidades da ferramenta e se os instrumentos de coleta de dados estavam claros. Como resultado dessa etapa, foram realizadas melhorias na ferramenta, incluindo ajustes no formato de apresentação das histórias geradas, a inclusão de uma funcionalidade para salvar o contexto inserido pelos usuários a fim de evitar perdas e aprimoramentos relacionados à conectividade com a internet. Não foram apontadas mudanças nos instrumentos de coleta de dados.

6.3.1.2.2 Execução do Estudo Experimental: Inicialmente, todos os participantes receberam uma explicação sobre os objetivos do estudo e sobre as atividades que seriam realizadas. Em seguida, foi apresentado e coletado o Termo de Consentimento Livre e Esclarecido (TCLE), garantindo a participação voluntária e o anonimato das informações fornecidas. Após o consentimento, os participantes receberam uma breve explicação sobre a ferramenta Story AI Generator, incluindo uma breve apresentação das informações disponibilizadas em sua interface. Em seguida, tiveram acesso ao documento contendo os cenários de exemplo.

Cada participante utilizou a ferramenta de forma individual para gerar histórias de usuários, podendo optar por utilizar os cenários previamente disponibilizados ou elaborar histórias a partir de cenários definidos por eles próprios. Ao final da interação, os participantes responderam ao questionário de caracterização e ao instrumento de avaliação baseado no modelo TAM3, composto por questões fechadas e abertas, com foco na experiência de uso, na aceitação da ferramenta e nas percepções gerais. Por fim, foi realizada uma entrevista semiestruturada, conduzida pelo pesquisador, com o objetivo de aprofundar as respostas fornecidas no questionário, esclarecer pontos levantados pelos participantes e identificar limitações e oportunidades de melhoria da ferramenta. As entrevistas foram registradas para posterior análise qualitativa. A Figura 25 apresenta alguns registros desse processo. Todo o procedimento foi conduzido em um único momento, o tempo médio de execução foi de 28 minutos e 49 segundos.



(a) Registro da entrevista com participante P7.



(b) Registro da entrevista com participante P5.

Figura 25 – Registros do processo de condução das entrevistas.

6.3.2 Resultados do Estudo

A análise dos dados coletados neste estudo foi realizada por meio de uma abordagem mista, combinando técnicas de análise quantitativa e qualitativa, de forma a obter uma compreensão mais abrangente das percepções dos participantes sobre o uso da ferramenta Story AI Generator.

A análise quantitativa foi aplicada aos dados provenientes das questões fechadas do questionário baseado no modelo TAM3. Os dados foram analisados por meio de estatísticas descritivas com o objetivo de avaliar os construtos relacionados à aceitação da tecnologia, tais como utilidade percebida, facilidade de uso percebida e intenção de uso.

A análise qualitativa foi conduzida a partir das respostas às questões abertas do questionário e dos dados obtidos nas entrevistas semiestruturadas. Esses dados foram analisados por meio de análise temática, buscando identificar padrões recorrentes, percepções, limitações e oportunidades de melhoria relacionadas ao uso da ferramenta. O processo de análise envolveu a leitura sistemática dos dados, a codificação inicial e o

agrupamento dos códigos em temas representativos.

A análise qualitativa foi conduzida pela própria autora, com revisão de duas pesquisadoras doutoras, a partir da categorização temática das respostas abertas do questionário e das entrevistas semiestruturadas, com apoio da ferramenta ATLAS.ti².

O conjunto de dados que fundamentam a análise deste estudo, estão em um material suplementar contendo as histórias de usuários geradas na íntegra pela ferramenta Story AI Generator, as transcrições das respostas abertas do modelo TAM e o registro textual detalhado das entrevistas semiestruturadas realizadas³.

Seção 1 - Perfil demográfico dos participantes

Embora o ambiente de coleta de dados do estudo tenha ocorrido em São José dos Campos (SP), durante uma conferência de alcance nacional, observou-se uma distribuição geográfica diversa entre os participantes, como visto na Tabela 7. O estado do Amazonas apresentou a maior frequência, com quatro menções à cidade de Manaus. Em seguida, destacam-se os estados do Ceará e do Rio Grande do Sul, enquanto São Paulo, Paraná e Distrito Federal foram mencionados uma vez cada.

Essa diversidade geográfica sugere que as percepções coletadas não se restringem a um contexto regional específico, o que contribui para ampliar o escopo de análise e reduzir vieses associados a práticas locais de Engenharia de Requisitos.

A maioria dos participantes (81,8%) tenha indicado utilizar histórias de usuários tanto em ambientes acadêmicos quanto industriais.

Seção 2 - Experiência dos participantes com histórias de usuários

Considerando que o critério de inclusão do estudo foi a experiência prévia com histórias de usuários, observou-se que 81,8% dos participantes afirmaram utilizá-las tanto em contextos acadêmicos quanto industriais. Esse aspecto é particularmente relevante, uma vez que avaliações realizadas por usuários experientes tendem a apresentar uma visão mais ampla e maior rigor crítico em relação à qualidade das histórias geradas e à adequação das ferramentas ao contexto real de uso.

A proximidade dos participantes com a prática de especificação de histórias também se evidencia pela frequência de uso: 72,8% dos participantes relataram especifi-

² ATLAS.ti Scientific Software Development GmbH. ATLAS.ti. Disponível em: <<https://atlasti.com/>>.

³ [Material Suplementar](#)

Tabela 7 – Caracterização dos participantes.

Participante	Localização	Contexto de Uso	Frequência	Experiência
P1	São Borja, RS	Ambos	Mensalmente	Mais de 10 anos
P2	Porto Alegre, RS	Ambos	Semanalmente	Mais de 5 anos
P3	Campo Mourão, PR	Ambos	Esporadicamente	Menos de 5 anos
P4	Rio de Janeiro, RJ	Ambos	Diariamente	Menos de 5 anos
P5	Brasília, DF	Ambos	Semanalmente	Mais de 10 anos
P6	Morada Nova, CE	Academia	Esporadicamente	Menos de 5 anos
P7	Fortaleza, CE	Ambos	Diariamente	Mais de 5 anos
P8	Manaus, AM	Indústria	Esporadicamente	Mais de 20 anos
P9	Manaus, AM	Ambos	Semanalmente	Menos de 5 anos
P10	Manaus, AM	Ambos	Mensalmente	Mais de 5 anos
P11	Manaus, AM	Ambos	Mensalmente	Menos de 5 anos

car histórias de usuários de forma recorrente, seja diariamente ou mensalmente. Esse dado reforça que a maioria da amostra está inserida em contextos nos quais a escrita de histórias de usuários faz parte do fluxo regular de trabalho ou de suas atividades acadêmicas.

No que se refere ao tempo de experiência, 54,6% dos participantes declararam possuir mais de cinco anos de atuação na especificação de requisitos por meio de histórias de usuários e 45,5% possuem menos de 5 anos de experiência. Ainda assim, uma parcela expressiva apresenta menor tempo de experiência, o que pode ser interpretado à luz da adoção relativamente recente dessa prática em larga escala, impulsionada principalmente pela disseminação de metodologias ágeis. Essa heterogeneidade do perfil contribui para a coleta de percepções tanto de profissionais mais experientes quanto daqueles menos experientes.

Seção 3 - Familiaridade com IAs generativas

Outro aspecto relevante refere-se à familiaridade dos participantes com tecnologias de IA generativa, apresentada na Figura 26. Foi utilizada uma classificação simplificada buscando abranger os diferentes níveis de familiaridade com modelos LLMs, considerando desde o uso básico em interfaces prontas até o desenvolvimento e avaliação técnica dessas tecnologias.

- **Nenhuma:** Não utiliza LLMs.

- **Baixa:** Utiliza LLMs para tarefas simples, por meio de interfaces prontas, como o ChatGPT.
- **Média:** Utiliza prompts estruturados e aplica técnicas como few-shot learning para melhorar a qualidade das respostas.
- **Alta:** Ajusta estratégias de prompting, integra interfaces APIs e aplica LLMs em processos e fluxos de trabalho.
- **Muito Alta:** Avalia, treina ou realiza ajustes em LLMs e desenvolve soluções, como frameworks, baseadas em inteligência artificial generativa.

Pressupõe-se que usuários mais familiarizados com essas tecnologias tendem a apresentar expectativas mais elevadas quanto à qualidade e utilidade das saídas geradas. Os resultados indicam um alto nível de familiaridade: 45,5% dos participantes declararam possuir familiaridade muito alta com IAs generativas. Esse dado sugere que as avaliações realizadas não partem de usuários iniciantes, mas de indivíduos capazes de analisar criticamente as soluções geradas por IAs.

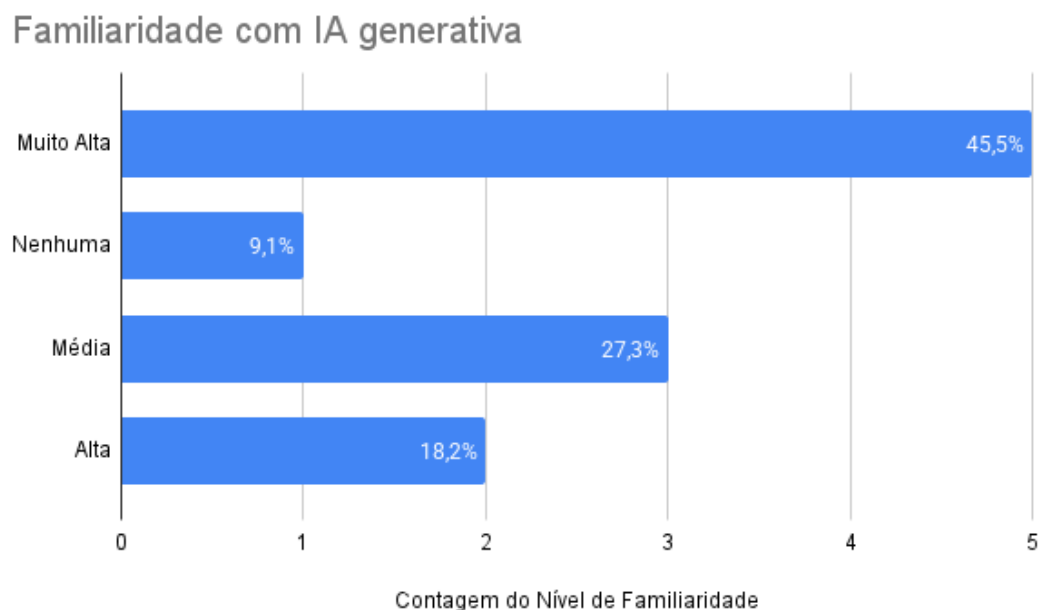


Figura 26 – Distribuição do nível de familiaridade com ferramentas de IA generativas.

Aproveitando-se a alta familiaridade, questionou-se também, se os participantes fazem uso de alguma ferramenta automatizada ou de IA para apoiar a escrita de histórias, especificamente. Obteve-se que 63,6% já fizeram este uso e 36,4% não utilizaram. Dentre os participantes que possuem muito alta ou alta familiaridade com IAs, apenas um deles não fez esse uso para gerar histórias, o que pode estar associado ao fato de esse participante especificar histórias de maneira esporádica, ou mesmo por seu campo de atuação ser o ambiente industrial, o que por vezes, pode levar a limitações nesses usos.

Além disso, a análise conjunta da frequência de especificação de histórias e do nível de familiaridade com IAs generativas sugere que o uso dessas tecnologias está mais fortemente associado ao grau de familiaridade do usuário com IA do que à frequência com que realiza a atividade de especificação. Esse achado indica que barreiras cognitivas e culturais relacionadas ao uso de IA podem ser mais determinantes do que fatores puramente operacionais.

Entre os participantes que especificam histórias de usuários com maior frequência, 62,5% relataram já ter utilizado IAs para apoiar esse processo. Em contrapartida, entre os participantes com maior familiaridade com IAs generativas, esse percentual aumenta para 85,7%, reforçando a influência da familiaridade com IA sobre a adoção dessas ferramentas. Ainda assim, destaca-se que os dois participantes que especificam histórias diariamente (P4 e P7) afirmaram já ter utilizado IAs para apoiar a escrita de histórias, mesmo classificando sua familiaridade como alta e média, respectivamente.

No que se refere às ferramentas utilizadas, o ChatGPT foi mencionado por seis participantes, seguido por Gemini, Copilot e DeepSeek. Esses resultados evidenciam o uso predominante de modelos de propósito geral para apoiar a escrita de histórias de usuários. Nesse contexto, a menção à ferramenta Collab2US ([DORNELLES; PARIZI, 2024](#)), desenvolvida pelo participante P1, merece destaque, por se tratar de uma solução especializada voltada especificamente à geração colaborativa de histórias de usuários.

Esse achado reforça a motivação para o desenvolvimento e a avaliação da Story AI Generator, ao indicar uma demanda crescente por ferramentas que aliem geração automatizada às boas práticas nas atividades de especificação de requisitos.

6.3.2.1 Resultados do Questionário: Avaliação da Aceitação

O questionário de avaliação capturou dados relacionados à caracterização dos participantes, bem como respostas a questões fechadas estruturadas com base nos construtos do modelo TAM3 e questões abertas voltadas à coleta de percepções qualitativas sobre a experiência de uso e a aceitação da ferramenta.

Facilidade de Uso: As respostas apresentadas no gráfico sobre Facilidade de Uso Percebida (Figura 27) indicam uma avaliação amplamente positiva quanto à facilidade de uso da ferramenta. Em todas as afirmações do construto, ao menos 81,8% dos participantes concordaram fortemente em relação à clareza da interação (F1. Minha interação com a ferramenta para geração de histórias de usuários foi clara e compreensível), ao baixo esforço cognitivo requerido (F2. Interagir com a ferramenta para gerar histórias de usuários não exigiu muito esforço mental), à facilidade geral de uso (F3. Eu considero a ferramenta fácil de utilizar) e à simplicidade para gerar histórias de usuários (F4. Eu acho fácil fazer com que a ferramenta gere histórias de usuários).

Esses resultados sugerem que, sob a percepção da maioria dos participantes, a ferramenta é fácil de utilizar e apresenta uma interação intuitiva. Essa interpretação é corroborada pelas respostas abertas, como a de P1, que destacou que a ferramenta *“...gerou uma quantidade interessante de histórias de usuários de maneira rápida e sem esforço”*, e a de P3, que apontou que *“A interface é bem organizada e clean, facilitando o seu entendimento e a interação”*.

Apesar da predominância de avaliações positivas, foi observada uma resposta neutra à afirmação F4, o participante P7 declarou: *“... justifico minha posição pois são vários passos para chegar ao resultado final, o que em uma ferramenta de LLM padrão, um único prompt seria o suficiente”*. Esse resultado indica que, embora a ferramenta seja percebida como fácil de usar, o fluxo estruturado pode ser interpretado como mais trabalhoso por alguns usuários.

Utilidade: Em relação à Utilidade Percebida (Figura 28), os resultados indicam que a maioria dos participantes concordou ou concordou fortemente que a ferramenta contribui positivamente para seu desempenho (U1. Usar a ferramenta melhora meu desempenho na criação de histórias de usuários), produtividade (U2. Usar a ferramenta

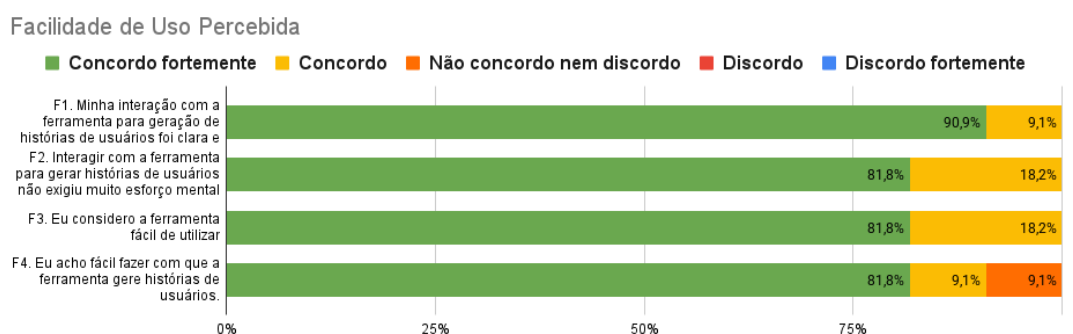


Figura 27 – Distribuição das respostas para o construto Facilidade de Uso Percebida.

em meu trabalho poderá melhorar minha produtividade), eficácia na criação de histórias de usuários (U3. Usar a ferramenta melhora a eficácia na geração das histórias de usuários geradas para o meu projeto) e utilidade no seu trabalho (U4. Considero a ferramenta útil no meu trabalho). As respostas abertas reforçam essa percepção, como exemplificado por P3, que destacou: *"A ferramenta é um ótimo suporte para agilizar a atividade e apoia no processo criativo"* e por P10, que ressaltou que: *"A ferramenta é útil porque ajuda a detalhar as histórias de usuário do sistema. Esse tipo de suporte melhora minha performance para a escrita destas"*.

Ainda que a avaliação geral tenha sido positiva, alguns participantes apontaram limitações relacionadas a funcionalidades não contempladas pela ferramenta, indicando que a ausência desses recursos pode impactar a percepção de utilidade, como dito por P11: *"A ferramenta é útil para um estágio inicial de definição das USs. A medida em que as USs exigem mais detalhes, como o critério de aceite (DOD) e as regras de negócio associadas a ferramenta não consegue detalhar..."*.

Intenção de Uso Futuro: Quanto à Intenção de Uso Futuro, no gráfico da Figura 29 observou-se que 63,6% dos participantes concordaram fortemente com as afirmações relacionadas ao uso da ferramenta em cenários futuros, em todas as afirmações (I1. Supondo que eu tenha acesso a essa ferramenta, I2. Se eu tiver acesso a essa ferramenta, pretendo usá-lo para gerar minhas histórias de usuários em futuros projetos e I3. Eu pretendo utilizar a ferramenta nos próximos meses para criar histórias de usuários eu a utilizaria para gerar minhas histórias de usuários). As respostas evidenciam tanto o interesse em contextos educacionais, como destacado por P2: *"Adoraria usar com meus*

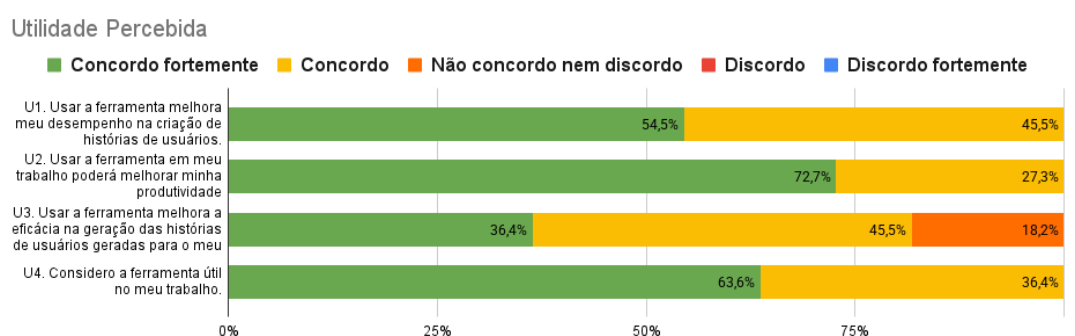


Figura 28 – Distribuição das respostas para o construto Utilidade Percebida.

alunos na disciplina de Requisitos e até para eu mesma testar meu próprio conhecimento”, quanto em ambientes de pesquisa e desenvolvimento, conforme P5: “...usar a ferramenta irá me apoiar na escrita das Histórias para que possa apresentá-la ao stakeholder... ”.

Entretanto, as respostas neutras indicam que a intenção de uso futuro é condicionada a fatores contextuais, destacando-se restrições organizacionais do ambiente industrial, como pontuado por P7: *“Dentro da visão da indústria, o uso deve ser feito com cautela e validado pelo empregador...”* e ausência de elementos considerados essenciais, como relatou P9: *“Eu poderia usar no futuro se a ferramenta mostrasse o critério de aceitação para cada história gerada”*. Esses aspectos influenciam diretamente a disposição dos participantes em adotar a ferramenta, evidenciando que a intenção de uso depende tanto das condições organizacionais quanto do atendimento às expectativas práticas dos usuários.

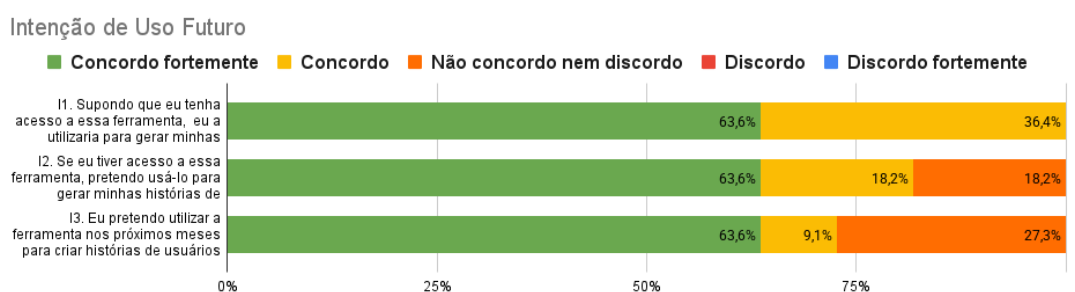


Figura 29 – Distribuição das respostas para o construto Intenção de Uso Futuro.

Relevância: Em relação a Relevância Percebida (Figura 30), os resultados indicam uma avaliação positiva, porém acompanhada de ressalvas. Observa-se uma maior dis-

persão das respostas, com aumento na proporção de avaliações neutras, especialmente na afirmação relacionada à importância (R1: No meu trabalho, o uso desta ferramenta é importante). Na afirmação R1, apenas 36,4% dos participantes concordaram fortemente que o uso da ferramenta é importante em seu trabalho. O que sugere que, embora a ferramenta seja percebida como útil e fácil de usar, nem todos os participantes a consideraram essencial ou indispensável em suas atividades atuais, como P9 comenta: *"...não posso afirmar que o momento atual o uso desta ferramenta é importante no meu projeto"*.

Por outro lado, as afirmações relacionadas à relevância (R2: No meu trabalho, o uso desta ferramenta é relevante) e à pertinência (R3: O uso desta ferramenta é pertinente às minhas diversas tarefas relacionadas à especificação de histórias de usuários), apresentaram avaliações mais positivas. Esses resultados indicam que a ferramenta foi percebida como relevante, ao agregar valor às atividades de especificação de requisitos, e pertinente, por se adequar às tarefas desempenhadas pelos participantes, como apontado por P5: *"Poderei me dedicar a outras atividades, já que a ferramenta gerou as históricas de usuários a partir do cenário fornecido, de maneira clara e rápida"*. Assim, os resultados indicam que a Story AI Generator foi percebida como um recurso de apoio às atividades de especificação de requisitos.

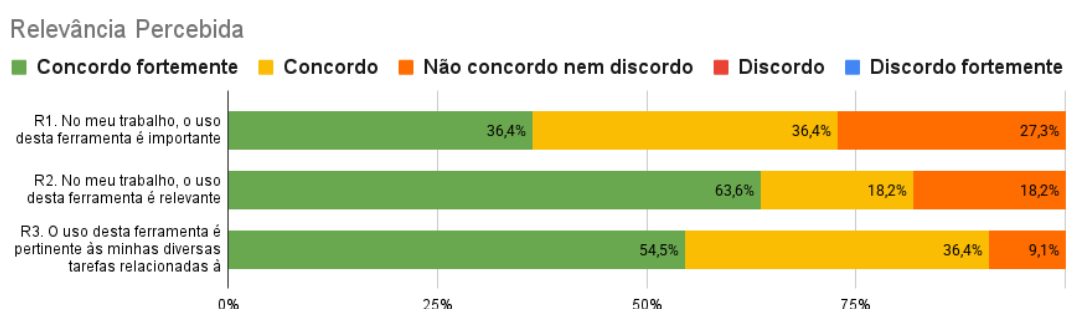


Figura 30 – Distribuição das respostas para o construto Relevância Percebida.

Qualidade de Saída Percebida: Os resultados do construto Qualidade de Saída Percebida (Figura 31) indicam uma avaliação positiva das histórias de usuários geradas pela ferramenta. Nas afirmações O1 (A qualidade da saída que obtenho com a ferramenta é alta) e O2 (Não tenho problemas com a qualidade da saída da ferramenta), a maioria dos participantes concordou e considerou que as histórias atenderam ao con-

texto proposto, ainda que tenham indicado limitações que exigem refinamento, tal como P1: *"As histórias geradas suprimiram o cenário ao qual o contexto fora fornecido"* e P2: *"...Apenas observar os requisitos não funcionais e pensar como a ferramenta poderia reconhecê-los"*.

As únicas respostas de discordância foram em O2 e O3 (Considero a qualidade das saídas da ferramenta excelente), registradas pelo participante P6, que, embora reconheça o alinhamento das histórias ao contexto informado, destacou: *"As histórias de usuário geradas estão bem alinhadas com o contexto que passei para ela. Entretanto, o refinamento das histórias precisa ser feito, considerando histórias semelhantes e épicas (e suas histórias de usuário)"*. As demais respostas situam-se na neutralidade, refletindo uma avaliação satisfatória, porém com necessidade de aprimoramentos.

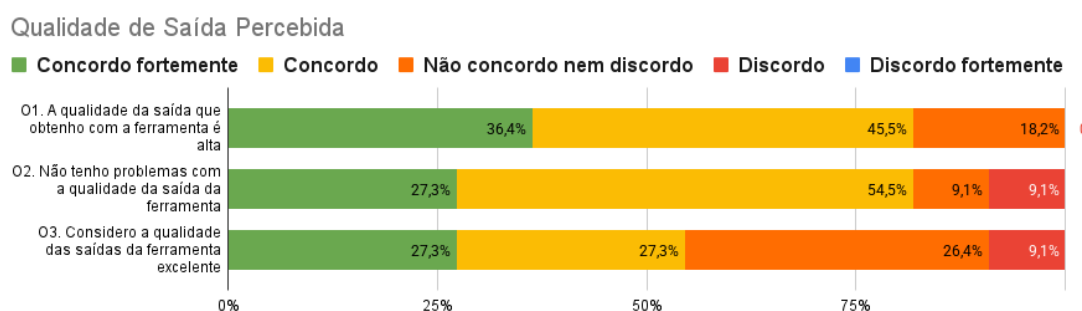


Figura 31 – Distribuição das respostas para o construto Qualidade de Saída Percebida.

6.3.2.2 Resultados das Entrevistas: Percepções dos Participantes

A análise dessas percepções foi conduzida a partir das respostas abertas do questionário e das entrevistas semiestruturadas realizadas ao final do experimento. A análise dos dados foi conduzida com base em procedimentos da Grounded Theory (GT), método qualitativo que emprega procedimentos sistemáticos de coleta e análise de dados para a construção e validação de teorias substantivas relacionadas a fenômenos sociais (GLASER; STRAUSS; STRUTZEL, 1968). Conforme proposto por Glaser, Strauss e Strutzel (1968), as teorias substantivas são específicas a determinados contextos e não têm como objetivo a generalização ampla. Nesse sentido, Corbin e Strauss (2014) definem teoria como um conjunto integrado de proposições que explicam a variação na ocorrência de

um fenômeno social subjacente ao comportamento de um grupo ou à interação entre grupos. Embora a GT tenha como finalidade a construção de teorias, seu uso pode ser parcial, de acordo com os objetivos da pesquisa (CORBIN; STRAUSS, 2014).

O processo de análise fundamentou-se na codificação dos dados, realizada em duas etapas: codificação aberta e codificação axial. A codificação aberta envolveu a fragmentação, análise, comparação, conceitualização e categorização dos dados, permitindo a identificação inicial de códigos e categorias. Em seguida, a codificação axial foi empregada para examinar as relações entre as categorias, estabelecendo vínculos de causa e efeito entre elas, conforme recomendado na literatura de pesquisa qualitativa baseada em GT (CORBIN; STRAUSS, 2014).

Como forma de garantir a confiabilidade da análise, os códigos e categorias identificados pela autora, foram posteriormente revisados pelas duas pesquisadoras orientadoras, com o objetivo de auditar o processo de codificação. Os principais achados sobre as percepções com a ferramenta e suas respectivas implicações estão descritos a seguir.

Benefícios Observados ao Usar a Ferramenta - As respostas dos participantes evidenciam uma percepção amplamente favorável dos participantes em relação à facilidade de uso da ferramenta, à usabilidade, à estrutura da interface e à clareza da saída geradas pela Story AI Generator, como é possível observar no diagrama de relações de causa e efeito, construído com base nessa categoria, apresentado na Figura 32.

Os participantes destacaram que a ferramenta exige baixo esforço cognitivo, apresenta interface intuitiva e permite a geração rápida de histórias de usuários. Essa percepção é ilustrada por P1, que afirmou que a ferramenta *“gerou uma quantidade interessante de histórias de usuários de maneira rápida e sem esforço”*. Além disso, os participantes indicaram que a ferramenta contribui para a melhoria da performance ao fomentar a identificação de novas funcionalidades e ao agilizar atividades manuais e repetitivas.

Além da usabilidade, os participantes perceberam a ferramenta como um apoio eficaz à geração de histórias de usuários, segundo P7: *“...fomentando novas funcionalidades e agilizando trabalhos manuais e repetitivos”*. As histórias geradas foram consideradas alinhadas ao contexto informado, conforme o relato de P3, que indicou que: *“Sem ela, a*

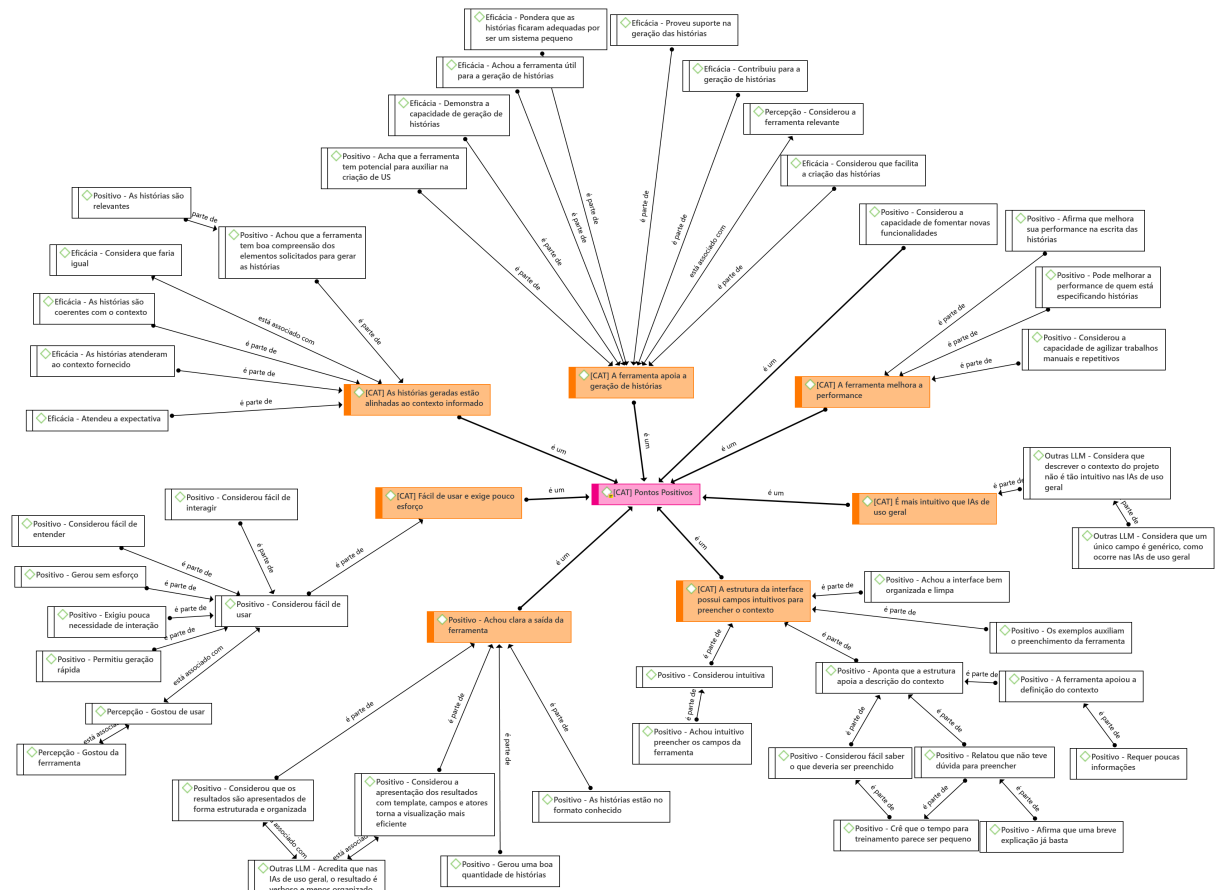


Figura 32 – Diagrama de relações sobre os benefícios observados ao usar a Story AI Generator.

ferramenta, eu faria igual minhas tarefas, mas com ela, pela amostragem, ficou excelente conferir os resultados que eu mesma traria...”

Outro aspecto recorrente refere-se à estrutura da interface, composta por campos orientados à descrição do contexto. Segundo P3: “...A estruturação da ferramenta em campos (descrição, atores e informações adicionais) facilita no momento de descrever o contexto do projeto...”. Essa abordagem foi percebida como mais intuitiva quando comparada a ferramentas de IA de uso geral, normalmente baseadas em um único campo genérico. Os participantes indicaram que essa estrutura, aliada aos exemplos fornecidos, apoia o preenchimento do contexto e reduz o tempo de treinamento prévio.

Quanto à clareza da saída gerada, os participantes ressaltaram que as histórias são apresentadas de forma estruturada e organizada, seguindo o template de histórias de usuários, o que facilita a visualização e compreensão. Em comparação com ferramentas de IA de propósito geral, cuja saída tende a ser mais verbosa e menos organizada, P1

comenta que: *"Achei as histórias efetivas, elas são bem formadas no formato que a gente conhece e todas elas atendiam ao contexto..."*.



Figura 33 – Diagrama de relações sobre o uso percebido da ferramenta.

Uso Percebido da Ferramenta como Apoio - A análise evidenciou que os participantes perceberam a ferramenta principalmente como um ponto de partida para a atividade de especificação de requisitos, e não como um artefato final, como mostra o diagrama de relações de causa e efeito, construído com base nessa categoria, apresentado na Figura 33. Nesse sentido, a ferramenta foi associada ao apoio inicial tanto para analistas de requisitos quanto para product owners, auxiliando na compreensão do problema e na estruturação inicial das histórias de usuários. P6 destacou: *"...a ferramenta é excelente como ponto de partida para a geração das histórias de usuário, logo, tanto na academia quanto na indústria, esse pontapé inicial é valioso"*.

O uso no contexto acadêmico também foi recorrente, sendo associado a atividades de teste e treinamento. Os participantes indicaram que a ferramenta pode ser utilizada por professores em exercícios práticos e em disciplinas relacionadas à Engenharia de Requisitos. P2, por exemplo, afirmou que: *"Adoraria usar com meus alunos na disciplina de Requisitos e até para eu mesma testar meu próprio conhecimento"*.

No contexto industrial, eles ressaltaram o uso da ferramenta como apoio à validação com stakeholders. P5 indicou que: *"...usar a ferramenta irá me apoiar na escrita das Histórias para que possa apresentá-la ao stakeholder logo e ter sua validação para ir para*

a especificação dos requisitos com as regras mais claras”, contribuindo para a definição de regras mais bem compreendidas antes da especificação formal dos requisitos.

Embora os participantes tenham demonstrado intenção de uso da ferramenta, foi recorrente a percepção de que as histórias geradas deveriam passar por uma revisão posterior, reforçando seu papel como apoio ao processo de especificação, e não como substituta da atuação do analista.

Recomendações Adicionais Feitas pela IA - A ocorrência de recomendações adicionais geradas pela IA durante a execução da ferramenta, caracteriza um fenômeno esperado no contexto do estudo, cujo objetivo era tornar as histórias de usuários mais completas e explorar possibilidades além do escopo inicialmente informado.

A maioria dos participantes percebeu esse comportamento de forma positiva ou neutra. Os relatos indicam que as recomendações foram consideradas coerentes com o contexto do problema e, em alguns casos, vistas como um diferencial da ferramenta. P8 reflete que a IA “... abrangeu até mais do que eu estava esperando então, eu achei bem legal, parabéns!”, enquanto P5 aponta: “...Então ela já serviu também no processo criativo”, contribuindo com ideias que não haviam sido inicialmente consideradas.

Em contraste, alguns participantes manifestaram ressalvas, como P7 contrapõe: “...porém a depender do perfil do Analista de requisito ou PO que utilizar, pode ver como um ponto negativo”. De forma semelhante, P10 discordou de algumas histórias sugeridas pela IA, considerando-as inadequadas, segundo ele: “...Embora ela tenha gerado no contexto do sistema mas, por exemplo, eu não quero avaliar problema reportado, seu só quero saber a quantidade, então ela poderia ser substituída por outra...”.

Esses resultados evidenciam um equilíbrio delicado entre o potencial criativo das recomendações adicionais e a necessidade de controle e aderência ao escopo original, sugerindo que esse comportamento da IA pode ser percebido tanto como um facilitador da ideação quanto como um risco à precisão da especificação, a depender do perfil e das expectativas do usuário. Como resposta a essa divergência, emerge a sugestão de permitir que o usuário indique seu interesse, a partir da reflexão de P2 sobre receber tais recomendações: “...pra que eu possa deliberar e até eu possa saber e voltar pro usuário e negociar e ver se o usuário teria interesse naquilo...”, tal sugestão, foca em equilibrar o

potencial criativo da ferramenta com a necessidade de controle e fidelidade ao escopo original, deixando a critério do usuário esta decisão.

Informações Adicionais, Critérios de Aceitação e RNFs - Este fenômeno evidencia limitações percebidas pelos participantes quanto à capacidade da ferramenta de incorporar plenamente o contexto fornecido, especialmente no que se refere à geração de critérios de aceitação e ao tratamento de requisitos não funcionais (RNFs).

De forma recorrente, os participantes relataram a ausência de critérios de aceitação nas histórias de usuários geradas. Essa percepção foi apontada por P6, que destaca: *"...um requisito não funcional não foi considerado, nem os critérios de aceitação"*, levando-o a associar essa ausência à seguinte conclusão: *"Notei que a ferramenta não considerou todo o contexto que descrevi..."*, bem como P7 ao comenta sobre a qualidade das histórias: *"...poderiam ter especificado um pouco mais, coloquei algumas informações que as user stories acabaram não compreendendo"*, que baseou essa conclusão no fato de a ferramenta ter abstraído informações não suportadas.

Em relação aos requisitos não funcionais, os participantes observaram que a ferramenta não os considerou da forma esperada. P2 identificou que: *"Duas user stories geradas representam requisitos não funcionais, então aqui que é preciso pensar como tratar quando for requisito desta categoria"*. Isso foi interpretado pelo participante como uma inadequação, pois ele não esperava que tais requisitos assumissem esse formato.

No que se refere às informações adicionais fornecidas, P3 relatou que determinadas especificações não foram refletidas diretamente nas histórias geradas, exemplificando a sua expectativa sobre uma história relacionada a um dashboard para visualização das notas dos alunos. Segundo o participante, esse termo foi abstraído para uma funcionalidade mais genérica, como a visualização de informações ou relatórios. Posteriormente, P3 ponderou: *"...Mas é uma questão tecnológica assim, então não impactaria tanto na hora da gente colocar essa história do usuário lá pra pra gerar as tarefas em cima, porque é uma questão tecnológica então, poderia se mudar dentro do projeto pra uma outra forma de apresentar esses dados então não afetaria tanto..."*, o que poderia justificar a abstração realizada pela ferramenta. Ainda assim, o episódio evidencia que as especificações detalhadas, inseridas nas informações adicionais, nem sempre são traduzidas de forma

explícita no resultado final, embora as histórias geradas sejam funcionalmente coerentes, a possibilidade de ofertar uma experiência mais completas, pode permitir que a ferramenta receba menos avaliações negativas.

Essas limitações levaram alguns participantes a expressarem dúvidas durante o uso da ferramenta, especialmente quanto aos critérios adotados pela IA para priorização dos elementos, ao conjunto de informações efetivamente considerado e ao nível de granularidade necessário para a geração das histórias. De modo geral, embora a ferramenta seja percebida como útil para a geração inicial de histórias de usuários, há limitações relacionadas à incorporação de informações adicionais, à representação de critérios de aceitação e ao tratamento de requisitos não funcionais, apontando oportunidades de evolução da ferramenta.

Desafios na Geração de Histórias de Usuários - Esta análise evidencia que os desafios percebidos pelos participantes não se restringem a um único aspecto da ferramenta, mas emergem de um conjunto de limitações ao longo do processo de geração das histórias de usuários.

De forma central, os relatos indicam dificuldades na incorporação adequada do contexto também estão associadas à qualidade das histórias geradas. Os participantes relataram histórias genéricas, pouco detalhadas, como relatado por P4: *"...e aí algumas eu senti que foram mais genéricas, por exemplo, na descrição eu definia os alertas eh necessários pro sistema, por exemplo, que era chamada, era alarme, campainha e alerta de incêndio. E aí nas USs... a descrição estava assim: faz eh... eh emite notificações importantes ou então avisa perigo, só que essas notificações elas elas já tinham sido definidas na descrição..."*. Além disso, também apresentam inconsistências, duplicidades e nível de granularidade inadequado, como alertado por P6: *"...o refinamento das histórias precisa ser feito, considerando histórias semelhantes e épicas (e suas histórias de usuário)"*.

De forma recorrente, observou-se que os campos de informações adicionais não foram refletidos nas histórias finais, resultando na ausência de critérios de aceitação e no tratamento inadequado de requisitos não funcionais. Essa limitação contribuiu para a percepção de perda da importância da ferramenta.

Essas lacunas levaram os participantes a dúvidas durante o uso, novamente

P3 comenta: *"...Minha dúvida fica no nível de granularidade que preciso fornecer para a ferramenta para que ela gere histórias adequadas..."*, e *"...Porém, fica a dúvida sobre os elementos que são considerados ou não no momento de gerar as US"*. Tais dúvidas apontam para a possibilidade de aumentar a transparência no funcionamento da ferramenta.

Além disso, foram apontadas dificuldades relacionadas à estrutura de entrada, considerada como excessivamente segmentada, segundo P7: *"...Eu acho que se a gente comparar entre uma ferramenta como essa e ferramentas de LLM padrão, né? Já como Gemini, com o prompt só, talvez ela saia perdendo dada impaciência do usuário, né?...."*. Ele também destacou preocupações quanto à adoção em contextos industriais: *"No geral, ferramentas que utilizam dados devem ser usadas com validação das empresas envolvidas"*, restrições organizacionais: *"...o uso da indústria deve ser tratado com cautela e validado anteriormente"* e privacidade dos dados: *"...porém na indústria, a gente tem que ter, eh. . . , tem que pensar um pouco na privacidade dos dados que a gente informa, né?..."*

O participante P2 também mencionou sua preocupação: *"E lá na indústria eu acho ariscado ser usada com o risco das pessoas ficarem preguiçosas e não quererem pensar"*, reduzindo a qualidade na definição de requisitos.

Expectativas de Evolução da Ferramenta - Os participantes não apenas identificaram limitações na apresentação atual da ferramenta, mas também refletiram sobre caminhos concretos para sua evolução. As sugestões concentram-se em aprimoramentos na interface, no processo de geração baseado em prompt e na ampliação dos elementos gerados, indicando uma expectativa de maior controle, transparência e completude das histórias de usuários.

Os participantes sugeriram melhorias na interface com o objetivo de facilitar a interação e a validação dos resultados, visualizando o texto de entrada, como sugerido por P10: *"...Uma sugestão é que a ferramenta disponibilize para os usuários a descrição fornecida na etapa anterior para que esse possa avaliar as histórias geradas em comparação com o texto fornecido..."*. Também a possibilidade de editar e refinar histórias, como pensado por P2: *"Mas se eu for descrevendo o cenário que eu sei, se ela daí vai conversando Comigo e me gerando mais, né? Porque veio tudo assim. . . não sei se conceitualmente tem diferença, tá? Mas talvez me daria a sensação de que eu estou conversando com alguém e tendo ideias e debatendo, né? Não*

sei se é necessário”.

E selecionar histórias para exportação, além de indicar a obrigatoriedade dos campos. No nível do processo de geração, foram apontadas sugestões para tornar a interação com a IA mais iterativa, permitir maior direcionamento por parte do usuário e explicitar as recomendações geradas automaticamente, conforme P2 sugere: *“...eu acho que dá pra distinguir entre aquilo que é solicitado e o que ela, pela inteligência que ela trabalha, ela recomenda...”*.

Também emergiu a expectativa de expansão funcional da ferramenta, com a geração de novos elementos relevantes à Engenharia de Requisitos. P9 sugeriu incluir critérios de aceitação: *“Uma sugestão seria colocar os critérios de aceitação para cada história gerada”*, como também dito por P6: *“Estou ansioso para ver a versão da ferramenta considerando os critérios de aceitação e também os requisitos não funcionais”*, assim como P2: *“Apenas observar os requisitos não funcionais e pensar como a ferramenta poderia reconhecê-los”*. Também foi sugerida a geração de documentação por P5, que expressa: *“...com a evolução da ferramenta possamos adicionar mais instruções sobre os requisitos e a documentação possa ser gerada automaticamente”*. P3 relata que: *“..opcionalmente, potenciais entregáveis associados às histórias agregaria valor...”*. Paralelamente, também foi destacada a necessidade de melhorias relacionadas à proteção de dados e privacidade como condição essencial para o uso da ferramenta em contextos industriais, P7 sugere: *“...talvez utilizar uma nuvem privativa...para que os dados não sigam para outra instância e se mantenham ali dentro da indústria...”*.

Os participantes vislumbraram diferentes formas de utilização da ferramenta, incluindo validação de histórias externas, P2 sugeriu: *“...eu já tenho escrito as minhas próprias us, e ela me ajudar a refinar...”* e geração de cenários a partir de histórias existentes, além de sugerirem a realização de novos estudos com sistemas mais complexos. Esses achados sugerem que os participantes não apenas buscaram melhorias pontuais, mas também articularam uma visão ampliada de como a ferramenta poderia evoluir para apoiar diferentes etapas do processo de Engenharia de Requisitos. Essa visão de futuro indica um engajamento positivo com a proposta da ferramenta, bem como a capacidade dos participantes de projetar usos mais abrangentes a partir de sua experiência inicial.

A análise qualitativa evidencia que a Story AI Generator foi amplamente aceita como apoio à geração inicial de histórias de usuários, destacando-se pela facilidade de uso, pela estrutura da interface e pela clareza da saída. Contudo, emergiram limitações na incorporação do contexto, especialmente quanto à ausência de critérios de aceitação, ao tratamento de requisitos não funcionais e à abstração de informações adicionais, reforçando a necessidade de maior transparência no processo de geração. As percepções indicam que a ferramenta é compreendida como um instrumento de suporte à Engenharia de Requisitos, útil para ideação, ensino e validação inicial, mas ainda dependente do julgamento do analista para assegurar qualidade, completude e aderência ao contexto. As expectativas de evolução reforçam esse entendimento, ao apontar seu potencial de expansão para apoiar etapas mais amplas do processo.

Estudos recentes têm investigado o uso de LLMs no apoio à Engenharia de Requisitos, embora com focos distintos. Trabalhos como o de [Jin et al. \(2024\)](#) e ReqInOne de [Zhu, Cordeiro e Sun \(2025\)](#) concentram-se principalmente na avaliação técnica de artefatos gerados, realizada por especialistas externos com base em critérios como completude e corretude, sem explorar a experiência de uso ou a percepção dos profissionais. Em outro contexto, [Sampaio et al. \(2024\)](#) analisam qualitativamente o uso do ChatGPT no ensino de Engenharia de Requisitos, destacando que a rapidez e o potencial criativo da ferramenta não eliminam a necessidade de refinamento humano, ainda que em ambiente educacional. [Rani, Svensson e Feldt \(2025\)](#) mostram que, embora a maioria dos profissionais apresente percepção positiva sobre o uso de IA na Engenharia de Requisitos, há preferência por abordagens colaborativas humano-IA, devido à percepção de que a IA tende a produzir soluções genéricas e enfrenta limitações na incorporação de contexto de domínio. De forma alinhada a essas evidências, os participantes deste estudo relataram dificuldades da ferramenta em compreender plenamente o contexto fornecido, associando esse fator à geração de histórias mais genéricas e à ausência de certos elementos da especificação, conforme suas expectativas e necessidades no momento do uso.

6.4 Limitações do Estudo

Toda pesquisa tem limitações de seus resultados (WOHLIN et al., 2012). Esta seção discute as principais limitações consideradas neste estudo:

Natureza exploratória e qualitativa da amostra: O estudo foi concebido com caráter exploratório e avaliativo, com foco na compreensão das percepções de especialistas a partir do uso da ferramenta. Nesse contexto, o número de participantes é compatível com estudos em Engenharia de Software que empregam estatística descritiva e análise qualitativa para investigação inicial de fenômenos emergentes. Ainda assim, os resultados não têm como objetivo a generalização estatística para a população de profissionais de Engenharia de Software, devendo ser interpretados como evidências exploratórias.

Amostragem por conveniência e contexto de coleta: Os participantes foram recrutados durante o SBQS 2025, o que caracteriza uma amostragem por conveniência e pode introduzir viés de auto-seleção. Embora tenha sido observada diversidade geográfica, o perfil dos participantes reflete o contexto de uma conferência científica, podendo não representar plenamente profissionais inseridos em diferentes realidades organizacionais ou setores industriais.

Dependência de percepções subjetivas: A avaliação da ferramenta baseou-se predominantemente em percepções dos participantes, coletadas por meio de questionários baseados no modelo TAM3 e de entrevistas semiestruturadas. Não foram utilizadas métricas objetivas de desempenho, como tempo de execução, número de revisões necessárias ou avaliação da qualidade das histórias por avaliadores independentes, o que limita inferências mais precisas. Estudos futuros podem incorporar medidas quantitativas, como tempo de execução, retrabalho e avaliação independente da qualidade das histórias.

Escopo da avaliação da qualidade das histórias: A avaliação concentrou-se na qualidade geral das histórias de usuários geradas, não incluindo a análises de critérios de aceitação nem do tratamento explícito de requisitos não funcionais. Esses aspectos são considerados como oportunidades para trabalhos futuros.

6.5 Conclusão

Este capítulo apresentou um estudo exploratório com profissionais de Engenharia de Software, com o objetivo de avaliar a aceitação da ferramenta Story AI Generator e compreender as percepções dos usuários quanto ao seu potencial de apoio às atividades de Engenharia de Requisitos. A análise permitiu investigar aspectos de aceitação, experiências de uso e limitações percebidas.

Em relação ao objetivo, os resultados quantitativos e qualitativos indicam uma aceitação predominantemente positiva da ferramenta. Os participantes destacaram a facilidade de uso, a clareza da interface e a organização das histórias geradas, fatores que contribuíram para percepções favoráveis de utilidade e intenção de uso. Esses achados sugerem que a ferramenta atende aos requisitos básicos de usabilidade e apoio inicial à geração de histórias de usuários.

Quanto à questão de pesquisa “Quais percepções, limitações e oportunidades de melhoria são identificadas pelos profissionais de Engenharia de Software a partir do uso da ferramenta Story AI Generator?”, a ferramenta foi compreendida como um apoio eficaz nas fases iniciais de elicitação e estruturação das histórias, sendo vista como um ponto de partida em contextos acadêmicos e industriais. Por outro lado, foram identificadas limitações relevantes, especialmente relacionadas à incorporação de informações adicionais, à ausência de critérios de aceitação e ao tratamento de requisitos não funcionais, considerados essenciais.

As oportunidades de melhoria apontadas reforçam e expandem achados de estudos anteriores. Parte das recomendações previamente propostas, como a redução da complexidade textual, simplificação da interação e o uso de exemplos para apoiar o entendimento do contexto, foi efetivamente endereçada. Entretanto, outras necessidades reapareceram, incluindo a modularização do processo de geração, o suporte à reescrita iterativa das histórias e a integração mais sistemática de critérios de qualidade, agora validadas pela experiência prática dos usuários.

De forma integrada, os achados evidenciam que a evolução da ferramenta ocorre de maneira incremental, guiada por evidências experimentais e feedback dos usuários. Observou-se também um trade-off entre o potencial criativo da IA e a necessidade

de maior controle e aderência ao escopo, indicando que diferentes perfis de usuários podem perceber esse comportamento de forma distinta. Ainda assim, os participantes demonstraram engajamento com a proposta e projetaram usos futuros, como apoio à ideação, validação com stakeholders e aplicação em contextos acadêmicos.

Embora a ferramenta apresente potencial para apoiar a geração de histórias de usuários de forma útil e eficiente, é necessária sua evolução e ampliação de seu escopo, seu uso efetivo depende de maior transparência, controle do usuário e integração de elementos essenciais da especificação. Assim, os resultados reforçam a importância de abordagens híbridas, nas quais ferramentas baseadas em LLMs atuam como suporte ao analista, sem substituir seu papel crítico no processo de Engenharia de Requisitos.

Por fim, este capítulo apresentou a terceira e última iteração do Ciclo de Design, em conformidade com o framework de Design Science Research de Hevner, descrevendo a avaliação do artefato *Story AI Generator* por profissionais de Engenharia de Software quanto à sua aceitação e percepções. Os resultados obtidos permitiram avaliar o artefato seguindo os ciclos de Rigor e Relevância, evidenciando suas contribuições e limitações, e encerram o ciclo de design desta pesquisa.

7

CONSIDERAÇÕES FINAIS E CONTRIBUIÇÕES DA PESQUISA

Concluindo esta Qualificação de Mestrado, este capítulo apresenta um resumo do progresso alcançado até o momento (Seção 7.1) e destacando as contribuições desta dissertação (Seção 7.2).

7.1 Considerações Finais

Esta pesquisa teve como objetivo apoiar engenheiros de requisitos e product owners na geração automatizada de histórias de usuário com qualidade, investigando como o uso de LLMs pode influenciar a qualidade desses artefatos e contribuir para o avanço das práticas em Engenharia de Requisitos. O estudo foi conduzido com base nos princípios do Design Science Research (DSR), estruturado em quatro ciclos iterativos de projeto e avaliação.

No ciclo de relevância, foi realizada uma pesquisa com profissionais da indústria de software para compreender práticas atuais de elicitação e especificação de requisitos e o uso de LLMs nesse contexto. Os resultados indicaram que LLMs já são utilizadas como apoio inicial na geração de histórias de usuário, oferecendo ganhos de agilidade e padronização, embora ainda com limitações relacionadas à ausência de contexto, privacidade e necessidade de supervisão humana. Esses achados evidenciaram a necessidade de técnicas de prompting mais estruturadas e alinhadas a critérios de qualidade.

Com base nessas evidências, foi conduzido um estudo exploratório comparando a geração manual de histórias com a geração automatizada por meio de prompting livre. Os resultados não indicaram diferenças estatisticamente significativas, mas demonstraram maior padronização e produtividade com o uso de LLMs, reforçando a necessidade de abordagens de prompting mais orientadas.

No primeiro ciclo de design, foi proposto o US-Prompt, uma técnica que combina meta-prompting e few-shot prompting para orientar a geração de histórias de usuário conforme os critérios do framework Quality User Story (QUS). A avaliação experimental demonstrou que o US-Prompt superou as abordagens manual e de prompting livre em qualidade semântica e pragmática, com maior padronização e clareza, embora ainda apresente desafios na dimensão sintática, especialmente quanto à atomicidade.

Na segunda iteração do ciclo de design, o US-Prompt foi avaliado em diferentes contextos de uso e quanto à sua aceitação com base no modelo TAM. Os resultados indicaram boa usabilidade, baixo esforço cognitivo e alta intenção de uso, embora tenham sido identificadas limitações relacionadas à necessidade de revisão humana e à sensibilidade a variações de prompt.

Como evolução dos ciclos anteriores, foi conduzido um terceiro ciclo de design, no qual o US-Prompt foi operacionalizado na ferramenta web Story AI Generator, permitindo a geração e o refinamento interativo de histórias de usuário. A ferramenta foi avaliada com profissionais de Engenharia de Software por meio de abordagem mista, com resultados que indicam elevada aceitação, utilidade percebida e potencial de adoção em contextos reais, apesar da necessidade de ajustes e revisões manuais.

Ao longo de todos os ciclos, a pesquisa assegurou rigor metodológico por meio de experimentos controlados, métricas reconhecidas e avaliação de aceitação, garantindo validade e confiabilidade aos resultados obtidos. Com base nas evidências, conclui-se que é possível desenvolver técnicas de prompting capazes de orientar LLMs na geração de histórias de usuário de alta qualidade, alinhadas a critérios sintáticos, semânticos e pragmáticos.

Como contribuição, esta pesquisa entrega não apenas evidências empíricas sobre o uso de LLMs em Engenharia de Requisitos, mas também um artefato validado, o

US-Prompt e sua implementação na ferramenta *Story AI Generator*, capaz de apoiar profissionais e estudantes na elaboração de histórias de usuário com maior qualidade, padronização e produtividade. De forma geral, os resultados indicam que o uso de LLMs, quando orientado por estratégias estruturadas de prompting e materializado em ferramentas de apoio, constitui uma abordagem promissora para a Engenharia de Requisitos.

7.2 Contribuições da Pesquisa

Esta pesquisa contribui para a área de Engenharia de Requisitos, abrangendo aspectos teóricos, práticos e de produção científica. Até o momento, foram submetidos três artigos e dois deles já foram aceitos. A seguir, as principais contribuições categorizadas:

Contribuições Teóricas:

Esta pesquisa apresentou diversas contribuições teóricas relevantes. Primeiramente, forneceu evidências experimentais sobre o uso de LLMs na especificação de histórias de usuários. Em seguida, realizou uma análise experimental comparativa entre a geração manual e a geração automatizada de histórias de usuários com LLMs.

Outro destaque foi a proposição e validação da técnica US-Prompt, fundamentada em princípios de qualidade de histórias de usuários. Por fim, contribuiu conceitualmente para os estudos sobre a aceitação de LLMs no contexto da Engenharia de Requisitos, especialmente no que se refere à automatização de tarefas de especificação de requisitos.

Contribuições Práticas:

Entre as contribuições práticas desta pesquisa, destaca-se o desenvolvimento de um prompt especializado e reutilizável (US-Prompt), adaptável a diferentes domínios e contextos de projeto. A proposta também oferece subsídios para sua aplicação pedagógica em disciplinas de Engenharia de Requisitos, permitindo que estudantes pratiquem a elaboração de histórias de usuário com feedback estruturado, baseado nos critérios de qualidade do framework QUS.

Também foi desenvolvida a ferramenta web *Story AI Generator*, que operaciona-

liza o US-Prompt em um ambiente interativo, possibilitando a geração, visualização e o refinamento automatizado de histórias de usuário. A ferramenta foi avaliada com profissionais de Engenharia de Software, apresentando evidências de utilidade percebida e potencial de adoção em contextos reais

A abordagem proposta também apresenta potencial de integração com ferramentas de apoio à especificação e à validação de requisitos, especialmente em ambientes de desenvolvimento ágil. Como apoio à pesquisa e ao ensino, foi criado um repositório de histórias de usuário geradas por LLMs, que pode ser utilizado para análise, treinamento e experimentação.

Além disso, a pesquisa contribuiu com a definição de um conjunto de critérios de qualidade para histórias de usuário, priorizados a partir da percepção e da experiência de profissionais da indústria, servindo como referência para orientar a avaliação e a geração automatizada desses artefatos.

Por fim, foi realizado o mapeamento das práticas e técnicas utilizadas na indústria para elicitación, estruturação e validação de requisitos, evidenciando como os LLMs vêm sendo incorporados a esses processos em cenários reais de desenvolvimento ágil.

7.3 Produção Científica

7.3.1 Artigos Aceitos

- **Artigo Internacional:**

- Santos, R., Freitas, G., Steinmacher, I., Conte, T., Oran, A. e Gadelha, B. (2025). *User Stories: Does ChatGPT Do It Better?*. In: *Proceedings of the 27th International Conference on Enterprise Information Systems - Volume 2: ICEIS*. SciTePress, pp. 47–58. ISBN: 978-989-758-749-8. ISSN: 2184-4992. DOI: <https://doi.org/10.5220/0013365500003929>.

Artigo aceito na Trilha de Pesquisa da [27th International Conference on Enterprise Information Systems \(ICEIS 2025\)](#).

- **Artigo Nacional:**

- *AI-Generated User Stories: Are They Good Enough?*, aceito na trilha de Ideias Inovadoras e Resultados Emergentes do [XXXIX Simpósio Brasileiro de Engenharia de Software \(SBES 2025\)](#). Autores: Santos, R., Steinmacher, I., Conte, T., Oran, A., Gadelha, B.
- Banner *Menos Esforço, Mais Qualidade: Automatizando a Geração de Histórias de Usuário com LLMs*, que apresenta um resumo da pesquisa sobre geração automatizada de histórias de usuário, aceito na 5ª *Escola Latino-Americana de Software - Trilha Principal*, evento integrante do [Congresso Brasileiro de Software: Teoria e Prática \(CBSOFT\)](#).
- *Adoption of LLMs in Requirements Engineering: What Practitioners Are Worried About?*, aceito antecipadamente na trilha de Trabalhos Técnicos do [XXIV Simpósio Brasileiro de Qualidade de Software \(SBQS 2025\)](#). Autores: Santos, R., Steinmacher, I., Conte, T., Oran, A., Gadelha, B.

7.4 Trabalhos Futuros

Com base nos resultados obtidos ao longo dos ciclos de design e nas evidências coletadas no estudo com profissionais de Engenharia de Software, são propostas as seguintes direções para trabalhos futuros, visando evoluir o artefato desenvolvido e ampliar sua aplicabilidade prática:

1. **Evolução da ferramenta Story AI Generator:** Aprimorar a ferramenta interativa baseada no US-Prompt, incorporando melhorias sugeridas pelos participantes do estudo, tais como: exemplificação mais clara de papéis e contextos, ajustes no formato de entrada e mecanismos para refinamento da granularidade em histórias insuficientemente detalhadas.
2. **Estruturação modular do US-Prompt:** Desenvolver uma arquitetura modular para o prompt, permitindo sua adaptação a diferentes tipos e modalidades de projetos de software, garantindo flexibilidade tanto para cenários simples quanto para contextos mais complexos de engenharia de requisitos.

3. **Ciclo iterativo de geração e avaliação de histórias:** Implementar na ferramenta um fluxo iterativo de geração, revisão e melhoria de histórias de usuário, permitindo que usuários criem novas histórias ou importem histórias externas e as submetam a avaliações automáticas baseadas em LLMs e nos critérios de qualidade definidos.
4. **Ampliação da entrada de dados contextuais:** Evoluir os mecanismos de entrada da ferramenta para suportar informações mais ricas de contexto e multi-modais, como regras de negócio, personas, visão do produto e requisitos de alto nível, possibilitando uma geração mais precisa e alinhada ao domínio do problema.
5. **Integração explícita dos critérios QUS no processo de validação:** Incorporar os critérios do framework *Quality User Story* (QUS) como perguntas-guia e *checklists* funcionais dentro da ferramenta, auxiliando os usuários a verificar se cada história gerada atende aos padrões sintáticos, semânticos e pragmáticos esperados.
6. **Avaliações em cenários reais de uso:** Conduzir estudos adicionais em ambientes industriais e em projetos reais, avaliando o impacto da ferramenta na produtividade, qualidade dos requisitos e aceitação ao longo do tempo.

REFERÊNCIAS

- AJAGBE, M.; ZHAO, L. Retraining a bert model for transfer learning in requirements engineering: A preliminary study. In: IEEE. *2022 IEEE 30th International Requirements Engineering Conference (RE)*. [S.l.], 2022. p. 309–315. [46](#)
- ALHAZMI, A.; HUANG, S. Survey on differences of requirements engineering for traditional and agile development processes. In: *2020 SoutheastCon*. [S.l.]: IEEE, 2020. p. 1–9. [36](#), [37](#)
- ARAÚJO, A. F. de; MARCACINI, R. M. Re-bert: automatic extraction of software requirements from app reviews using bert language model. In: *Proceedings of the 36th annual ACM symposium on applied computing*. [S.l.: s.n.], 2021. p. 1321–1327. [45](#)
- ARORA, C.; GRUNDY, J.; ABDELRAZEK, M. Advancing requirements engineering through generative ai: Assessing the role of llms. In: *Generative AI for Effective Software Development*. [S.l.]: Springer, 2024. p. 129–148. [45](#)
- ARVIDSSON, S.; AXELL, J. *Prompt engineering guidelines for LLMs in Requirements Engineering*. 2023. Retrieved from <https://hdl.handle.net/2077/77967>. [51](#)
- BAJAJ, D. et al. Muce: a multilingual use case model extractor using gpt-3. *International Journal of Information Technology*, Springer, v. 14, n. 3, p. 1543–1554, 2022. [46](#)
- BELZNER, L.; GABOR, T.; WIRSING, M. Large language model assisted software engineering: prospects, challenges, and a case study. In: *International Conference on Bridging the Gap between AI and Reality*. Cham: Springer Nature Switzerland, 2023. p. 355–374. [23](#)
- BERQUAND, A.; DARM, P.; RICCARDI, A. Spacetransformers: Language modeling for space systems. *IEEE access*, IEEE, v. 9, p. 133111–133122, 2021. [46](#)
- BROCKENBROUGH, A.; SALINAS, D. Using generative ai to create user stories in the software engineering classroom. In: *2024 36th International Conference on Software Engineering Education and Training (CSEET)*. [S.l.: s.n.], 2024. p. 1–5. [52](#), [109](#), [111](#), [144](#)
- BROWN, T. B. et al. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020. Disponível em: <https://arxiv.org/pdf/2005.14165>>. [44](#), [48](#), [88](#), [89](#), [97](#)
- CHAMI, M.; ZOGHBI, C.; BRUEL, J.-M. A first step towards ai for mbse: Generating a part of sysml models from text using ai. *A First Step towards AI*, 2019. [46](#)

- COHN, M. *User stories applied: For agile software development*. [S.l.]: Addison-Wesley Professional, 2004. 38
- CORBIN, J.; STRAUSS, A. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. [S.l.]: Sage publications, 2014. 166, 167
- COSLER, M. et al. nl2spec: Interactively translating unstructured natural language to temporal logics with large language models. In: SPRINGER. *International Conference on Computer Aided Verification*. [S.l.], 2023. p. 383–396. 46
- COSTA, R. A.; NUNES, T. S. O impacto da transformação ágil na cultura organizacional: das práticas e valores organizacionais a gestão da mudança. *Perspectivas em Ciência da Informação*, SciELO Brasil, v. 28, p. e29487, 2023. 34
- DAWOOD, O. S. et al. From requirements engineering to uml using natural language processing–survey study. *European Journal of Engineering and Technology Research*, v. 2, n. 1, p. 44–50, 2017. 46
- DEBNATH, T. et al. A comprehensive survey of prompt engineering techniques in large language models. *TechRxiv*, 2025. 47, 48
- DEEPTIMAHANTI, D. K.; SANYAL, R. Semi-automatic generation of uml models from natural language requirements. In: *Proceedings of the 4th India Software Engineering Conference*. [S.l.: s.n.], 2011. p. 165–174. 46
- DESHPANDE, G. et al. Is bert the new silver bullet?-an empirical investigation of requirements dependency classification. In: IEEE. *2021 IEEE 29th international requirements engineering conference workshops (REW)*. [S.l.], 2021. p. 136–145. 46
- DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. Disponível em: <<https://doi.org/10.48550/arXiv.1810.04805>>. 44
- DORNELLES, P. L.; PARIZI, R. Collab2us: a tool for collecting user stories from collaborative boards. In: WER. [S.l.: s.n.], 2024. 161
- DRESCH, A. et al. *Design science research*. [S.l.]: Springer, 2015. 26
- DWITAMA, F.; RUSLI, A. User stories collection via interactive chatbot to support requirements gathering. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, v. 18, n. 2, p. 890–898, 2020. 50
- GENTILI, E.; FALESSI, D. Practitioners’ perceptions on requirements smells. *Information and Software Technology*, v. 187, p. 107823, 2025. ISSN 0950-5849. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0950584925001624>>. 36, 38
- GHEORGHE, A.-M.; GHEORGHE, I. D.; IATAN, I. L. Agile software development. *Informatica Economica*, v. 24, n. 2, 2020. 34, 35
- GLASER, B. G.; STRAUSS, A. L.; STRUTZEL, E. The discovery of grounded theory; strategies for qualitative research. *Nursing research*, LWW, v. 17, n. 4, p. 364, 1968. 147, 166

HAHN, C. et al. Formal specifications from natural language. *arXiv preprint arXiv:2206.01962*, 2022. 46

HENG, S.; TSILIONIS, K.; WAUTELET, Y. Building user stories and behavior driven development scenarios with a strict set of concepts: Ontology, benefits and primary validation. In: *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*. [S.l.: s.n.], 2023. p. 1422–1429. 49

HEVNER, A. R. A three cycle view of design science research. *Scandinavian journal of information systems*, v. 19, n. 2, p. 4, 2007. 26

HEVNER, A. R. et al. *Design Research in Information Systems: Theory and Practice*. New York: Springer, 2004. ISBN 9780387245520. Disponível em: <https://books.google.com/books?id=89w-scN7_8MC>. 26

HEVNER, A. R. et al. Design science in information systems research. *MIS Quarterly*, Management Information Systems Research Center, University of Minnesota, v. 28, n. 1, p. 75–105, 2004. ISSN 02767783. Disponível em: <https://wise.vub.ac.be/sites/default/files/thesis_info/design_science.pdf>. 26, 27

HEY, T. et al. Norbert: Transfer learning for requirements classification. In: IEEE. 2020 *IEEE 28th international requirements engineering conference (RE)*. [S.l.], 2020. p. 169–179. 46

HICKEY, A. M.; DAVIS, A. M. A unified model of requirements elicitation. *Journal of management information systems*, Taylor & Francis, v. 20, n. 4, p. 65–84, 2004. 45

HIRAOU, S. R. *Optimising hard prompts with few-shot meta-prompting*. 2024. arXiv:2407.18920. Retrieved from <https://arxiv.org/abs/2407.18920>. 51, 103, 107

HÖST, M.; REGNELL, B.; WOHLIN, C. Using students as subjects—a comparative study of students and professionals in lead-time impact assessment. *Empirical Software Engineering*, Springer, v. 5, p. 201–214, 2000. 110, 141

HOY, Z.; XU, M. Agile software requirements engineering challenges-solutions—a conceptual framework from systematic literature review. *Information*, v. 14, n. 6, p. 322, 2023. Disponível em: <<https://doi.org/10.3390/info14060322>>. 22

HUANG, L. et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Trans. Inf. Syst.*, Association for Computing Machinery, New York, NY, USA, nov. 2024. ISSN 1046-8188. Just Accepted. Disponível em: <<https://doi.org/10.1145/3703155>>. 110, 141

JIN, D. et al. Mare: Multi-agents collaboration framework for requirements engineering. *arXiv preprint arXiv:2405.03256*, 2024. 175

KANG, Y. et al. Natural language processing (nlp) in management research: A literature review. *Journal of Management Analytics*, v. 7, n. 2, p. 139–172, 2020. Disponível em: <<https://doi.org/10.1080/23270012.2020.1756939>>. 43

KANKARAŠ, M.; CAPECCHI, S. Neither agree nor disagree: use and misuse of the neutral response category in likert-type scales. *METRON*, Springer, v. 83, n. 1, p. 111–140, 2025. ISSN 2281-695X. Disponível em: <<https://doi.org/10.1007/s40300-024-00276-5>>. 132

- KICI, D. et al. A bert-based transfer learning approach to text classification on software requirements specifications. In: *Canadian AI*. [S.l.: s.n.], 2021. 46
- KITCHENHAM, B. A.; PFLEEGER, S. L. Pesquisas de opinião pessoal. In: SHULL, F.; SINGER, J.; SJØBERG, D. I. K. (Ed.). *Guide to Advanced Empirical Software Engineering*. London: Springer, 2008. 56, 57
- KOTONYA, G.; SOMMERVILLE, I. *Requirements Engineering: Processes and Techniques*. Wiley, 1998. (Worldwide Series in Computer Science). ISBN 9780471972082. Disponível em: <<https://books.google.com.br/books?id=Up1QAAAAMAAJ>>. 37
- KRISHNA, M. et al. Using LLMs in Software Requirements Specifications: An Empirical Evaluation . In: *2024 IEEE 32nd International Requirements Engineering Conference (RE)*. Los Alamitos, CA, USA: IEEE Computer Society, 2024. p. 475–483. Disponível em: <<https://doi.ieeecomputersociety.org/10.1109/RE59067.2024.00056>>. 96
- KRISHNA, M. et al. *Using LLMs in Software Requirements Specifications: An Empirical Evaluation*. 2024. arXiv:2404.17842. Retrieved from <https://arxiv.org/abs/2404.17842>. 51
- KUMAR, B.; TIWARI, U.; DOBHAR, D. C. User story splitting in agile software development using machine learning approach. In: IEEE. *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*. [S.l.], 2022. p. 167–171. 49, 50
- LAM, L. K.; HURTADO, C. A. L.; PORTILLO, L. W. Framework for automating requirement elicitation using a chatbot. In: IEEE. *2022 IEEE Engineering International Research Conference (EIRCON)*. [S.l.], 2022. p. 1–4. 49, 50, 53
- LEÃO, R. et al. A prompt engineering-based process to build proto-personas during lean inception. In: *Proceedings of the 38th Brazilian Symposium on Software Engineering, SBES 2024, Curitiba, Brazil, September 30 - October 4, 2024*. [s.n.], 2024. p. 588–594. Disponível em: <<https://doi.org/10.5753/sbes.2024.3562>>. 97, 98, 99, 100
- LIKERT, R. A technique for the measurement of attitudes. *Archives of Psychology*, v. 22, n. 140, p. 1–55, 1932. 132, 155
- LINÅKER, J. et al. Guidelines for conducting surveys in software engineering. Department of Computer Science, Lund University, 2015. 56, 58
- LUCASSEN, G. et al. Improving agile requirements: The quality user story framework and tool. *Requirements Engineering*, Springer, v. 21, n. 4, p. 383–403, 2016. 23, 38, 40, 43, 85, 97
- LUCASSEN, G. et al. The use and effectiveness of user stories in practice. In: DANEVA, M.; PASTOR, O. (Ed.). *Requirements Engineering: Foundation for Software Quality*. [S.l.]: Springer, Cham, 2016, (Lecture Notes in Computer Science, v. 9619). p. 187–202. 38, 40, 41, 45, 84, 86, 114
- LUITEL, D.; HASSANI, S.; SABETZADEH, M. Using language models for enhancing the completeness of natural-language requirements. In: SPRINGER. *International working conference on requirements engineering: foundation for software quality*. [S.l.], 2023. p. 87–104. 46

- MALIK, G. et al. *Supervised Semantic Similarity-based Conflict Detection Algorithm: S3CDA*. 2025. Disponível em: <<https://arxiv.org/abs/2206.13690>>. 46
- MAO, Y.; HE, J.; CHEN, C. From prompts to templates: A systematic prompt template analysis for real-world llmapps. In: *Proceedings of the 33rd ACM International Conference on the Foundations of Software Engineering*. New York, NY, USA: Association for Computing Machinery, 2025. (FSE Companion '25), p. 75–86. ISBN 9798400712760. Disponível em: <<https://doi.org/10.1145/3696630.3728533>>. 97, 98
- MARQUES, N.; SILVA, R. R.; BERNARDINO, J. Using chatgpt in software requirements engineering: A comprehensive review. *Future Internet*, v. 16, n. 6, p. 180, 2024. Disponível em: <<https://doi.org/10.3390/fi16060180>>. 51
- MARTINS, H. F. et al. Design thinking: Challenges for software requirements elicitation. *Information*, v. 10, n. 12, 2019. ISSN 2078-2489. Disponível em: <<https://www.mdpi.com/2078-2489/10/12/371>>. 21, 35, 37
- MISHRA, A.; ALZOUBI, Y. I. Structured software development versus agile software development: a comparative analysis. *International Journal of System Assurance Engineering and Management*, v. 14, n. 4, p. 1504–1522, 2023. ISSN 0976-4348. Disponível em: <<https://doi.org/10.1007/s13198-023-01958-5>>. 34, 35
- MURTAZINA, M.; AVDEENKO, T. An ontology-based approach to the agile requirements engineering. In: SPRINGER. *International Andrei Ershov Memorial Conference on Perspectives of System Informatics*. [S.l.], 2019. p. 205–213. 49
- NEO, G. et al. User story tutor (ust) to support agile software developers. In: *Proceedings of the 16th International Conference on Computer Supported Education*. SCITEPRESS - Science and Technology Publications, 2024. p. 51–62. Disponível em: <<http://dx.doi.org/10.5220/0012619200003693>>. 53
- NGUYEN-DUC, A. et al. Generative artificial intelligence for software engineering—a research agenda. *Software: Practice and Experience*, Wiley Online Library, v. 55, n. 11, p. 1806–1843, 2025. 24
- NISTALA, P. V. et al. Towards digitalization of requirements: generating context-sensitive user stories from diverse specifications. *Automated Software Engineering*, Springer, v. 29, n. 1, p. 26, 2022. 49
- NORHEIM, J. J. et al. Challenges in applying large language models to requirements engineering tasks. *Design Science*, v. 10, p. e16, 2024. 22, 44
- OSWAL, J. U.; KANAKIA, H. T.; SUKTEL, D. Transforming software requirements into user stories with gpt-3.5-: An ai-powered approach. In: IEEE. *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. [S.l.], 2024. p. 913–920. 23, 24, 53, 114
- PACHECO, C.; GARCÍA, I.; REYES, M. Requirements elicitation techniques: a systematic literature review based on the maturity of the techniques. *IET Software*, v. 12, n. 4, p. 365–378, 2018. Disponível em: <<https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-sen.2017.0144>>. 36

- PFLIEGER, S. L.; KITCHENHAM, B. A. Principles of survey research: Part 1: Turning lemons into lemonade. *ACM SIGSOFT Software Engineering Notes*, ACM, v. 26, n. 6, p. 16–18, 2001. 56
- PUNTER, T. et al. Conducting on-line surveys in software engineering. In: IEEE. 2003 *International Symposium on Empirical Software Engineering*, 2003. ISESE 2003. *Proceedings*. [S.l.], 2003. p. 80–88. 56
- RADFORD, A. et al. Improving language understanding by generative pre-training. San Francisco, CA, USA, 2018. 44
- RAHARJANA, I. K.; SIAHAAN, D.; FATICHAH, C. User story extraction from online news for software requirements elicitation: A conceptual model. In: IEEE. 2019 *16th international joint conference on computer science and software engineering (JCSSE)*. [S.l.], 2019. p. 342–347. 49
- RAHMAN, T.; ZHU, Y. *Automated user story generation with test case specification using large language model*. 2024. arXiv. Retrieved from <https://arxiv.org/abs/2404.01558>. 52, 54, 114
- RAHMAN, T. et al. Take loads off your developers: Automated user story generation using large language model. In: 2024 *IEEE International Conference on Software Maintenance and Evolution (ICSME)*. [S.l.: s.n.], 2024. p. 791–801. 23, 24, 144
- RALPH, P. et al. *ACM SIGSOFT Empirical Standards*. 2020. 80
- RAMASAMY, V. et al. Enhancing user story generation in agile software development through open ai and prompt engineering. In: 2024 *IEEE Frontiers in Education Conference (FIE)*. [S.l.: s.n.], 2024. p. 1–8. 52
- RANI, L. M.; SVENSSON, R. B.; FELDT, R. Ai for requirements engineering: Industry adoption and practitioner perspectives. In: IEEE. 2025 *40th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*. [S.l.], 2025. p. 244–251. 175
- RASHEED, I. Requirement engineering challenges in agile software development. *Mathematical Problems in Engineering*, Hindawi, v. 2021, p. 1–18, 2021. Disponível em: <<https://doi.org/10.1155/2021/6696695>>. 22
- RASTOGI, A. et al. Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. In: *Proceedings of the AAAI conference on artificial intelligence*. [S.l.: s.n.], 2020. v. 34, n. 05, p. 8689–8696. 45
- REINHARTZ-BERGER, I.; KEMELMAN, M. Extracting core requirements for software product lines. *Requirements Engineering*, Springer, v. 25, n. 1, p. 47–65, 2020. 45
- RESKETI, M. R. et al. Automatic summarising of user stories in order to be reused in future similar projects. *IET software*, Wiley Online Library, v. 14, n. 6, p. 711–723, 2020. 49, 50, 53
- RODEGHERO, P. et al. Detecting user story information in developer-client conversations to generate extractive summaries. In: IEEE. 2017 *IEEE/ACM 39th International Conference on Software Engineering (ICSE)*. [S.l.], 2017. p. 49–59. 49

- RONANKI, K.; BERGER, C.; HORKOFF, J. Investigating chatgpt's potential to assist in requirements elicitation processes. In: IEEE. 2023 49th Euromicro conference on software engineering and advanced applications (SEEA). [S.l.], 2023. p. 354–361. [52](#)
- RONANKI, K.; CABRERO-DANIEL, B.; BERGER, C. Chatgpt as a tool for user story quality evaluation: Trustworthy out of the box? In: KRUCHTEN, P.; GREGORY, P. (Ed.). *Agile Processes in Software Engineering and Extreme Programming – Workshops*. Cham: Springer, 2024. (Lecture Notes in Business Information Processing, v. 489). Disponível em: https://doi.org/10.1007/978-3-031-48550-3_17. [23](#), [51](#), [54](#), [85](#)
- SALMAN, I.; MISIRLI, A. T.; JURISTO, N. Are students representatives of professionals in software engineering experiments? In: IEEE. 2015 IEEE/ACM 37th IEEE international conference on software engineering. [S.l.], 2015. v. 1, p. 666–676. [110](#), [141](#)
- SAMPAIO, S. S. et al. Exploring the use of large language models in requirements engineering education: An experience report with chatgpt 3.5. In: *Proceedings of the XXIII Brazilian Symposium on Software Quality*. [S.l.: s.n.], 2024. p. 624–634. [175](#)
- SANTOS, C. A. dos; BOUCHARD, K.; NAPOLEÃO, B. M. Automatic user story generation: a comprehensive systematic literature review. *International Journal of Data Science and Analytics*, Springer, p. 1–24, 2024. [48](#)
- SHARMA, A.; TRIPATHI, A. K. Evaluating user story quality with llms: a comparative study. *Journal of Intelligent Information Systems*, Apr 2025. ISSN 1573-7675. Disponível em: <https://doi.org/10.1007/s10844-025-00939-3>. [52](#), [54](#)
- SIAHAAN, D.; RAHARJANA, I. K.; FATICHAH, C. User story extraction from natural language for requirements elicitation: Identify software-related information from online news. *Information and software technology*, Elsevier, v. 158, p. 107195, 2023. [49](#)
- SOMMERVILLE, I. *Software Engineering*. 9th. ed. Boston: Pearson, 2011. ISBN 978-0-13-703515-1. Disponível em: https://engineering.futureuniversity.com/BOOKS%20FOR%20IT/Software-Engineering-9th-Edition-by-Ian-Sommerville.pdf?utm_source=chatgpt.com. [21](#), [22](#)
- SONBOL, R.; REBDAWI, G.; GHNEIM, N. The use of nlp-based text representation techniques to support requirement engineering tasks: A systematic mapping review. *Ieee Access*, IEEE, v. 10, p. 62811–62830, 2022. [44](#), [45](#)
- THUAN, N. H.; DRECHSLER, A.; ANTUNES, P. Construction of design science research questions. *Communications of the Association for Information Systems*, v. 44, n. 1, p. 20, 2019. [27](#), [30](#)
- TORRES-IGARTUA, M. J. et al. Analysis of natural language processing techniques and tools for requirements elicitation: A systematic literature review. In: 2024 12th International Conference in Software Engineering Research and Innovation (CONISOFT). [S.l.: s.n.], 2024. p. 48–57. [21](#)
- TRIVEDI, D. Agile methodologies. *International Journal of Computer Science Communication*, v. 12, n. 2, p. 91–100, 2021. [35](#)
- VASWANI, A. et al. Attention is all you need. In: *Advances in Neural Information Processing Systems*. [s.n.], 2017. v. 30. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. [43](#)

- VEITÍA, F. J. P.; ROLDÁN, L.; VEGETTI, M. User stories identification in software's issues records using natural language processing. In: IEEE. 2020 IEEE Congreso Biental de Argentina (ARGENCON). [S.l.], 2020. p. 1–7. 49
- VENKATESH, V.; BALA, H. Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, Wiley, v. 39, n. 2, p. 273–315, 2008. 114, 116, 132, 147, 154
- WAKE, B.; WAKE, B. *Invest in good stories, and Smart Tasks*. 2003. Accessed: 28 April 2023. Disponível em: <<https://xp123.com/articles/invest-in-good-stories-and-smart-tasks/>>. 40
- WAUTELET, Y. et al. Unifying and extending user story models. In: JARKE, M. et al. (Ed.). *Advanced Information Systems Engineering*. Cham: Springer International Publishing, 2014. p. 211–225. ISBN 978-3-319-07881-6. 38
- WHITE, J. et al. Chatgpt prompt patterns for improving code quality, refactoring, requirements elicitation, and software design. In: _____. *Generative AI for effective software development*. Springer, 2024. p. 71–108. Disponível em: <https://doi.org/10.1007/978-3-031-55642-5_4>. 51
- WOHLIN, C. et al. *Experimentation in software engineering*. [S.l.]: Springer, 2012. v. 236. 109, 115, 140, 153, 176
- YAMANI, A.; BASLYMAN, M.; AHMED, M. Leveraging llms for user stories in ai systems: Ustai dataset. In: *Proceedings of the 21st International Conference on Predictive Models and Data Analytics in Software Engineering*. New York, NY, USA: Association for Computing Machinery, 2025. (PROMISE '25), p. 21–30. ISBN 9798400715945. Disponível em: <<https://doi.org/10.1145/3727582.3728689>>. 145
- YARLAGADDA, R. T. Software engineering automation in it. *International Journal of Innovations in Engineering Research and Technology*, 2021. Retrieved from <https://ssrn.com/abstract=3797346>. 23
- YILDIRIM, S.; CEVIK, M.; BAŞAR, A. Few-shot learning approaches to software requirement quality prediction. In: *Proceedings of the 33rd Annual International Conference on Computer Science and Software Engineering*. IBM Corp., 2023. p. 155–160. Disponível em: <<https://dl.acm.org/doi/abs/10.5555/3615924.3615941>>. 51
- ZHANG, Y.; YUAN, Y.; YAO, A. C.-C. Meta prompting for ai systems. *arXiv preprint arXiv:2311.11482*, 2023. 48, 51, 88, 89, 97
- ZHANG, Z. et al. Llm-based agents for automating the enhancement of user story quality: An early report. In: *International Conference on Agile Software Development*. [S.l.]: Springer Nature Switzerland, 2024. p. 117–126. 38, 39, 40, 43
- ZHAO, L. et al. Natural language processing for requirements engineering: A systematic mapping study. *ACM Computing Surveys (CSUR)*, ACM New York, NY, USA, v. 54, n. 3, p. 1–41, 2021. 44
- ZHU, T.; CORDEIRO, L. C.; SUN, Y. Requinone: A large language model-based agent for software requirements specification generation. In: IEEE. 2025 IEEE 33rd International Requirements Engineering Conference (RE). [S.l.], 2025. p. 449–457. 175

A

QUESTIONÁRIO UTILIZADO NO SURVEY COM PROFISSIONAIS DA INDÚSTRIA

Pesquisa de Opinião: Profissionais de Requisitos

Esta pesquisa tem como objetivo compreender como os profissionais da indústria lidam com requisitos e utilizam Modelos de Linguagem de Grande Escala (LLMs) em suas atividades, além de identificar o interesse na geração automatizada de histórias de usuários.

* Indica uma pergunta obrigatória

1. Nome *

2. E-mail *

3. Qual é o seu cargo atual? *

Marcar apenas uma oval.

☐ Analista de Requisitos

☐ Desenvolvedor(a)

☐ Gerente de Projetos

☐ Product Owner

☐ Scrum Master

☐ Testador(a)

☐ Outro:

4. Há quanto tempo você trabalha na área de desenvolvimento de software? *

Marcar apenas uma oval.

- ☐ Menos de 1 ano
- ☐ 1-3 anos
- ☐ 3-5 anos
- ☐ Mais de 5 anos
- ☐ Mais de 10 anos
- ☐ Mais de 20 anos

5. Qual é o nome da empresa em que você trabalha? *

6. Onde ela está localizada (cidade e estado)? *

Práticas de Engenharia de Requisitos

Queremos entender melhor como você e sua equipe realizam o processo de Engenharia de Requisitos. Suas respostas nos ajudarão a compreender as abordagens e métodos que utilizam no desenvolvimento de requisitos de software.

7. Com que frequência você participa da elaboração de requisitos em seus projetos? *

Marcar apenas uma oval.

- ☐ Frequentemente
- ☐ Ocasionalmente
- ☐ Raramente
- ☐ Nunca

8. Quais métodos ou técnicas sua equipe utiliza para realizar a elicitação de requisitos (ex: entrevistas com stakeholders, brainstorming, análise de documentos, observação)? Por favor, descreva brevemente o processo: *

9. Quais métodos ou técnicas sua equipe utiliza para validar os requisitos definidos (ex: revisões com stakeholders, prototipagem)? Por favor, descreva brevemente o processo: *

10. Sua equipe utiliza histórias de usuários para estruturar e registrar os requisitos gerados?

Marcar apenas uma oval.

☐ Sim *Pular para a pergunta 11*

☐ Não *Pular para a pergunta 16*

Histórias de Usuários

Queremos entender melhor como você e sua equipe criam e utilizam histórias de usuário no processo de Engenharia de Requisitos. Suas respostas nos ajudarão a identificar as abordagens, desafios e práticas adotadas para gerar histórias de usuário eficazes no desenvolvimento de software.

11. Além de histórias de usuários, quais outras formas ou técnicas sua equipe utiliza para estruturar e registrar os requisitos? (Marque todas que se aplicam) *

Marque todas que se aplicam.

- ☐ Casos de uso
- ☐ Diagramas
- ☐ Mockups e Protótipos
- ☐ Outro: _____

12. Quais procedimentos ou métodos você utiliza para avaliar a qualidade das histórias de usuário criadas? (ex: uso de frameworks, revisões em equipe, validação com stakeholders). Por favor, descreva como essa avaliação é conduzida e quais aspectos são priorizados: *

13. Sua equipe utiliza critérios para aceitação das histórias? Se sim, quais são os critérios e quais informações são incluídas para que a história seja considerada pronta? *

14. Na sua opinião, qual é o maior desafio ao escrever histórias de usuários em sua equipe? (Marque todas que se aplicam) *

Marque todas que se aplicam.

- ☐ Falta de informações claras sobre as necessidades dos usuários
- ☐ Dificuldade em transformar requisitos técnicos em histórias compreensíveis
- ☐ Falta de tempo para elaborar histórias detalhadas
- ☐ Problemas na comunicação entre as equipes envolvidas
- ☐ Falta de prioridade na escrita das histórias de usuário por parte da equipe
- ☐ Outro: _____

15. Quais das seguintes qualidades você considera importantes para uma boa história de usuário? (Marque todas que se aplicam) *

Marque todas que se aplicam.

- ☐ Bem Formada (Inclui papel (persona) e ação específica)
- ☐ Atômica (Aborda uma única funcionalidade específica)
- ☐ Mínima (Contém apenas informações essenciais)
- ☐ Conceitualmente sólida (A funcionalidade tem um propósito claro e bem explicado)
- ☐ Orientada para Problemas (Especifica apenas o problema, não a solução)
- ☐ Não ambígua (Evita termos ou abstrações que possam gerar múltiplas interpretações)
- ☐ Livre de Conflitos (Não deve ser inconsistente com nenhuma outra história de usuário)
- ☐ Frase completa (É uma frase completa e bem formada)
- ☐ Estimável (Não possui um requisito de granulação grossa e pode ser planejada e priorizada de forma viável)
- ☐ Exclusiva (Cada história de usuário é única sem duplicatas)
- ☐ Uniforme (Todas as histórias empregam o mesmo modelo)
- ☐ Independente (Não possui dependências inerentes de outras histórias)
- ☐ Completa (A implementação de um conjunto de histórias de usuário cria um aplicativo completo, sem etapas faltando)
- ☐ Outro: _____

Pular para a pergunta 17

Outras Técnicas de Registro dos Requisitos

16. Quais formas ou técnicas sua equipe utiliza para estruturar e registrar os requisitos? (Marque todas que se aplicam)

Marque todas que se aplicam.

- ☐ Casos de uso
☐ Diagramas
☐ Mockups e Protótipos
☐ Outro: _____

Uso de Modelos de Linguagem de Grande Escala (LLMs)

17. Você utiliza Modelos de Linguagem (LLMs), ex: ChatGPT, Gemini, LLama e Copilot, nas suas atividades profissionais? *

Marcar apenas uma oval.

- ☐ Sim
☐ Não *Pular para a pergunta 26*

Uso de LLMs (Modelos de Linguagem de Grande Escala)

Queremos explorar como os LLMs estão impactando o processo de Engenharia de Requisitos, especialmente na geração de histórias de usuário.

18. Qual ou quais LLMs, ex: ChatGPT, Gemini, LLama e Copilot, você utiliza? *

19. Você utiliza LLMs na geração de histórias de usuários?

Marcar apenas uma oval.

- ☐ Sim *Pular para a pergunta 25*
☐ Não *Pular para a pergunta 20*

Não Uso de LLMs na Geração de Histórias de Usuários

20. Para quais atividades você costuma usar LLMs? (Marque todas que se aplicam) *

Marque todas que se aplicam.

- ☐ Geração de código
- ☐ Escrita de documentação técnica
- ☐ Análise ou refinamento de requisitos
- ☐ Ideias iniciais para brainstorming
- ☐ Outro: _____

21. Como você acha que o uso de LLMs impacta o seu trabalho? Explique os impactos positivos e/ou negativos: *

22. Você utilizaria LLMs para gerar histórias de usuários?

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não

23. O que seria necessário para você começar a utilizar LLMs na geração de histórias de usuários?

24. Você tem sugestões de melhorias que tornariam o uso de LLMs na geração de histórias mais viável?

Uso de LLMs na Geração de Histórias

25. Como você acha que o uso de LLMs impacta na geração de histórias de usuários? Explique os impactos positivos e/ou negativos:

Pular para a pergunta 28

Não Uso de LLMs (Modelos de Linguagem de Grande Escala)

Justifique sua resposta

26. Por que você não utiliza LLMs nas suas atividades profissionais? *

27. Você utiliza LLMs para algum outro tipo de atividades?

Pular para a pergunta 28

Sugestões e Opiniões Finais

28. Você tem sugestões, ideias ou percepções sobre como aprimorar o uso de LLMs na criação de histórias de usuário? *

B

US-PROMPT PARA GERAÇÃO DE HISTÓRIAS DE USUÁRIO - VERSÃO 1

US-Prompt (Versão 1)

Orientações para a Execução:

1. Leitura Inicial: Leia atentamente todas as definições fornecidas antes de iniciar o processo.
2. Execução do Processo: Após compreender as definições, prossiga para executar todas as atividades apresentadas.
3. Atividades: O processo é composto de várias atividades. Para cada atividade, você encontrará: Descrição da Atividade, Prompt a ser inserido e Resposta esperada do ChatGPT.
4. Personalização: O texto em **vermelho** representa dados que variam de projeto para projeto, enquanto que o texto em preto são instruções padrão do processo.

Passo a passo para a Execução:

1. Atividade - Apresentação da Dinâmica

- Descrição da atividade: Apresentar a ideia da conversa.
- Prompt a inserir:

“Olá, Chat! Quando eu te solicitar uma ação considere que estou conduzindo uma dinâmica de descoberta de um produto de software. Para toda e qualquer informação que eu lhe fornecer, estude, guarde e me responda única e exclusivamente: "Entendido" quando terminar.”
- Resposta esperada do ChatGPT: Entendido.

2. Atividade - Apresentação do Produto

- Descrição da atividade: Apresentar visão de produto.
- Prompt a inserir:

“Considere também que o software que estou construindo possui a seguinte visão de produto: **"A agência de turismo "Roteiros & Aventuras" é uma empresa estabelecida, localizada em uma cidade turística movimentada. A agência é conhecida por sua equipe experiente e seu compromisso em proporcionar experiências de viagem inesquecíveis para seus clientes. Com uma ampla gama de serviços, incluindo venda de passagens aéreas, terrestres e fluviais, reservas em hotéis e organização de passeios nas cidades, a agência está buscando implementar um sistema de gestão para otimizar suas operações e melhorar o atendimento ao cliente."**
- Resposta esperada do ChatGPT: Entendido.

3. Atividade - Apresentação dos Requisitos

- Descrição da atividade: Apresentação do documento de requisitos para o software.
- Prompt a inserir:

“A partir de agora vou lhe fornecer um documento com requisitos ainda não validados, gerado com base em entrevistas ou conversas informais com o cliente, contendo os usuários, ações e os principais funcionalidades desejadas para este sistema:

Usuários:

- Agentes de Viagens: Responsáveis por atender os clientes, fazer reservas, emitir passagens e vouchers de hotel, além de organizar os passeios.
- Clientes: Indivíduos que buscam serviços de viagem, incluindo passagens, hospedagem e passeios turísticos.

Ações:

- Agentes de Viagens:
 - Registrar novos clientes e manter informações atualizadas.
 - Pesquisar e sugerir opções de viagem com base nas preferências e orçamento do cliente.
 - Emitir passagens aéreas, terrestres e fluviais.
 - Fazer reservas em hotéis, incluindo seleção de quartos e datas.
 - Organizar pacotes de passeios em cidades turísticas, incluindo transporte e guias turísticos, se necessário.
 - Gerenciar pagamentos e documentação relacionada à viagem.
- Clientes:
 - Pesquisar opções de viagem, incluindo passagens, hotéis e passeios turísticos.
 - Solicitar cotações e fazer reservas.
 - Pagar pelos serviços contratados.
 - Visualizar itinerários e vouchers de viagem.
 - Comunicar-se com os agentes de viagens para tirar dúvidas ou solicitar alterações.

Requisitos Funcionais:

- Cadastro de Clientes: O sistema deve permitir que os agentes de viagens registrem novos clientes e mantenham suas informações atualizadas, incluindo dados pessoais, preferências de viagem e histórico de reservas.
- Pesquisa e Sugestão de Viagens: Os agentes de viagens devem ter a capacidade de pesquisar e sugerir opções de viagem com base nas preferências e orçamento do cliente, utilizando critérios como destinos, datas de viagem e tipo de transporte.
- Reserva de Serviços: Os agentes de viagens devem ser capazes de emitir passagens aéreas, terrestres e fluviais, fazer reservas em hotéis (incluindo seleção de quartos e datas) e organizar pacotes de passeios em cidades turísticas, conforme solicitado pelos clientes.

- Gerenciamento de Pagamentos e Documentação: O sistema deve permitir que os agentes de viagens gerenciem pagamentos relacionados aos serviços contratados pelos clientes e mantenham a documentação necessária para cada reserva, como recibos, vouchers e itinerários.
- Pesquisa e Reservas por Parte dos Clientes: Os clientes devem ter a capacidade de pesquisar opções de viagem, solicitar cotações e fazer reservas diretamente através do sistema, utilizando uma interface amigável e intuitiva.
- Visualização de Itinerários e Vouchers: Os clientes devem poder visualizar seus itinerários de viagem, incluindo detalhes sobre passagens, hospedagem e passeios, bem como acessar e imprimir vouchers de viagem para cada serviço reservado.
- Comunicação Cliente-Agentes: O sistema deve facilitar a comunicação entre clientes e agentes de viagens, permitindo que os clientes tirem dúvidas, solicitem alterações em suas reservas e recebam assistência durante todo o processo de planejamento e realização da viagem.”

- Resposta esperada do ChatGPT: Entendido.

4. Atividade - Template dos User Stories

- Descrição da atividade: Informar o template e os critérios desejados para criação das histórias de usuário.

- Prompt a inserir:

“Eu vou lhe fornecer um template para a criação de histórias de usuário. Utilize-o como base sempre que for solicitado a gerar o máximo de histórias de usuário possíveis. Observe que as histórias devem seguir rigorosamente os sete critérios descritos abaixo. Antes de finalizar, revise cada história para garantir que ela esteja claramente definida e atenda a todos os critérios.

- **Critérios:**

1- **Bem Formada:** Se a história inclui pelo menos um papel (persona), que define quem está solicitando a funcionalidade e um meio (ação), que é a atividade ou funcionalidade específica que o sistema deve permitir.

2- **Atômica:** Se cada história aborda uma única funcionalidade específica.

3- **Mínima:** Se a história contém apenas um papel, um meio (ex: “quero abrir o mapa interativo”), e um ou mais fins, que são os objetivos que o usuário pretende alcançar usando a funcionalidade (ex: “para localizar facilmente destinos turísticos”), sem informações desnecessárias.

4- **Conceitualmente Sólida:** Se o meio expressa uma funcionalidade, e os fins explicam a razão dessa funcionalidade.

5- **Não Ambígua:** Se ela evita termos ou abstrações que possam gerar múltiplas interpretações.

6- **Frase Completa:** Se a história é completa, bem formada e fornece contexto suficiente para ser claramente compreendida.

7- **Estimável:** Se a história não denota um requisito de granulação grossa que seja difícil de planejar e priorizar.

- **Orientações:**

- Identificação de Funcionalidades: Identifique claramente qual é a tarefa ou recurso que está sendo solicitado. Certifique-se de que cada funcionalidade seja isolada e clara.
 - Criação de Histórias Atômicas: Cada história deve tratar de uma única ação específica. Isso facilita o planejamento, a implementação e a estimativa.
 - Ajuste da Estrutura: Se uma história inclui duas ou mais funcionalidades distintas, haverá dois ou mais verbos de ação ligados por um “e”, se houver, você pode separá-la em duas ou mais histórias separadas, por exemplo, a história composta “Como um cliente, eu quero consultar meu saldo e sacar dinheiro, para que eu possa saber quanto tenho disponível e retirar o valor necessário”. Deverá ser divididas em:
 - História Atômica 1: Como um cliente, eu quero consultar meu saldo, para que eu possa saber quanto tenho disponível.
 - História Atômica 2: Como um cliente, eu quero sacar dinheiro, para que eu possa retirar o valor necessário.
 - Porém outros verbos de ação podem ser considerados como fins, quando não descrevem a ação principal da história. Eles geralmente vêm após um “para” e descrevem o benefício ou propósito da ação. Por exemplo, no exemplo abaixo, “visualizar” é o verbo principal e “verificar” e “repetir” são os resultados esperados.
 - História Atômica 3: Como um cliente, eu quero visualizar meu histórico de pedidos para que eu possa verificar compras passadas e repetir pedidos facilmente.
 - Validação dos Critérios: Verifique se cada história atende a todos os critérios: bem-formado, atomicidade, minimalidade, solidez conceitual, não ambiguidade, frase completa e estimabilidade. Se não atender, ajuste a história até que todos os critérios sejam satisfeitos.
 - Revisão e Refinamento: Após criar as histórias, revise cada uma para garantir que atenda a todos os critérios, especialmente a atomicidade. Se uma história ainda aborda múltiplas funcionalidades, divida-a em partes menores.
- **Exemplos de Histórias Bem Definidas (Few-Shot Prompting):**
 - História de Usuário Atômica 4: Criar Tarefas
Como um gerente
Eu quero criar tarefas
Para que eu possa adicionar novas atividades e distribuí-las para a equipe.
 - História de Usuário Atômica 5: Atualizar Tarefas
Como um gerente
Eu quero atualizar tarefas

Para que eu possa modificar detalhes das atividades existentes conforme necessário.

Tudo que está em CAPS é um placeholder. Toda vez que você gerar um texto tente encaixar em um dos placeholders que eu listei. Por favor, preserve o formato exato do template geral e ao final responda com “Entendido”.

História de Usuário X : NOME DA HISTÓRIA DE USUÁRIO

- História de Usuário: Como um [TIPO DE USUÁRIO]

Eu quero [REALIZAR UMA AÇÃO]

Para que [RESULTADO ESPERADO] ”

- Resposta esperada do Chat GPT: Entendido.

5. Atividade - Geração dos User Stories

- Descrição da atividade: Pedir para a LLM gerar as histórias de usuário para o software.

- Prompt a inserir:

“Atuando como um Engenheiro de RA, com muita experiência em elicitação de requisitos. Considere todas as informações que te concedi previamente sobre o Produto de Software que estou construindo, sobre o template, critérios e orientações que te forneci para a criação de histórias de usuário. E agora, não responda somente “Entendido” e gere o máximo de histórias de usuários que conseguir.”

- Resposta esperada do Chat GPT: Conjunto de histórias de usuário de acordo com template e com os requisitos.

Execução do prompt no dia 16 de agosto de 2024.

Link: <https://chatgpt.com/share/dd6b591b-df07-4c52-b228-9126247c4349>

C

US-PROMPT PARA GERAÇÃO DE HISTÓRIAS DE USUÁRIO - VERSÃO 2

US-Prompt (Versão 2)

Link para o ChatGPT: <https://openai.com/chatgpt/>

Orientações para a Execução:

- 1. Leitura Inicial:** Leia atentamente todas as orientações fornecidas antes de iniciar o processo.
- 2. Execução do Processo:** Após compreender as orientações, prossiga para executar todas as atividades apresentadas.
- 3. Atividades:** O processo é composto de várias atividades. Para cada atividade, você encontrará: Descrição da atividade, Instruções sobre o que você deve fazer em cada atividade, o Prompt a ser inserido e Resposta esperada do ChatGPT.
- 4. Personalização:** O texto em **vermelho** representa dados referentes ao seu produto de software, e devem ser alterados no prompt, antes de ser inserido no ChatGPT.
- 5. Interação com o ChatGPT:** O texto em cor preta são instruções padrão do processo, para elas você deve apenas ler o prompt da atividade atual, copiar e colar o prompt no Chat GPT e enviar.

Passo a passo para a Execução

1. Atividade 1 (Apresentação da Dinâmica)

- **Descrição da atividade:** Apresenta a ideia da conversa para o ChatGPT.
- **Instruções da atividade:** Insira o prompt no ChatGPT.

- **Prompt a inserir:**

[Olá, Chat! Quando eu te solicitar uma ação considere que estou conduzindo uma dinâmica de descoberta de um produto de software. Para toda e qualquer informação que eu lhe fornecer, estude, guarde e me responda única e exclusivamente: "Entendido" quando terminar.]

- **Resposta esperada do chat GPT: Entendido.**

2. Atividade 2 (Apresentação do Produto)

- **Descrição da atividade:** Apresenta a visão de produto para o ChatGPT.
- **Instruções da atividade:** Altere suas informações e insira o prompt no ChatGPT.

- **Prompt a inserir:**

[A partir de agora vou lhe fornecer dados que foram coletados com base em entrevistas ou conversas informais com o cliente. Considere que o software que estou construindo possui a seguinte visão de produto: "Insira aqui a ideia do produto de software que você deseja desenvolver".]

- **Resposta esperada do ChatGPT: Entendido.**

3. Atividade 3 (Apresentação do Requisitos do Produto)

- **Descrição da atividade:** Apresenta os requisitos do software.
- **Instruções da atividade:** Insira o prompt no ChatGPT.

- **Prompt a inserir:**

[Considere que coletei as seguintes informações sobre os principais usuários e as principais ações desejadas para este sistema:

Usuários: Liste quem são os principais usuários (Quem vai usar o software?)

- Usuário 1
- Usuário 2

Ações: Liste as principais ações de cada usuário (O que cada um pode fazer?)

Usuário 1:

- Liste aqui as ações deste tipo de usuário

Usuário 2:

- Liste aqui as ações deste tipo de usuário

]

- **Resposta esperada do ChatGPT: Entendido.**

4. Atividade 4 (Template dos User Stories)

- **Descrição da atividade:** Informa o template desejado para criação das histórias de usuário.
- **Instruções da atividade:** Insira o prompt no ChatGPT.
- **Prompt a inserir:**

[Eu vou lhe fornecer um template para a criação de histórias de usuário. Utilize-o para estruturar sua resposta, sempre que for solicitado a gerar o máximo de histórias de usuário possíveis.

Tudo que está em CAPS é um placeholder. Toda vez que você gerar um texto tente encaixar em um dos placeholders que eu listei. Por favor, preserve o formato exato do template geral e ao final responda com “Entendido”.

História de Usuário X : NOME DA HISTÓRIA DE USUÁRIO

**- História de Usuário: Como um {TIPO DE USUÁRIO},
Eu quero {REALIZAR UMA AÇÃO},
Para que {RESULTADO ESPERADO}.]**

- **Resposta esperada do Chat GPT: Entendido.**

5. Atividade 5 (Critérios de Qualidade dos User Stories)

- **Descrição da atividade:** Informa os critérios desejados para criação das histórias de usuário.
- **Instruções da atividade:** Insira o prompt no ChatGPT.

- **Prompt a inserir:**

[Eu vou lhe fornecer sete critérios de qualidade e suas definições, baseadas no framework Quality User Story (QUS) de Lucassen et al. (2016). O objetivo é orientar a criação de histórias de usuários bem definidas e aderentes a esses critérios. Observe que as histórias devem seguir rigorosamente cada critério e é importante revisar cada uma delas para garantir que estejam claramente definidas e em conformidade com todas as definições antes de considerá-las concluídas. Antes de finalizar, faça uma nova revisão de cada história para assegurar a qualidade e aderência aos critérios das histórias de usuários geradas.

- **Critérios de Qualidade:**

- 1- **Bem Formada:** Se a história inclui pelo menos um papel (persona), que define quem está solicitando a funcionalidade e um meio (ação), que é a atividade ou funcionalidade específica que o sistema deve permitir.
- 2- **Atômica:** Se cada história aborda uma única funcionalidade específica.
- 3- **Mínima:** Se a história contém apenas um papel, um meio (ex: “quero abrir o mapa interativo”), e um ou mais fins, que são os objetivos que o usuário pretende alcançar usando a funcionalidade (ex: “para localizar facilmente destinos turísticos”), sem informações desnecessárias.
- 4- **Conceitualmente Sólida:** Se o meio expressa uma funcionalidade, e os fins explicam a razão dessa funcionalidade.
- 5- **Não Ambígua:** Se ela evita termos ou abstrações que possam gerar múltiplas interpretações.

6- **Frase Completa:** Se a história é completa, bem formada e fornece contexto suficiente para ser claramente compreendida.

7- **Estimável:** Se a história não denota um requisito de granulação grossa que seja difícil de planejar e priorizar.

- **Orientações Adicionais:**

- Identificação de Funcionalidades: Identifique claramente qual é a tarefa ou recurso que está sendo solicitado. Certifique-se de que cada funcionalidade seja isolada e clara.
- Criação de Histórias Atômicas: Cada história deve tratar de uma única ação específica. Isso facilita o planejamento, a implementação e a estimativa.
- Ajuste da Estrutura: Se uma história inclui duas ou mais funcionalidades distintas, haverá dois ou mais verbos de ação ligados por um “e”, se houver, você pode separá-la em duas ou mais histórias separadas. Por exemplo, a história composta “Como um cliente, eu quero consultar meu saldo e sacar dinheiro, para que eu possa saber quanto tenho disponível e retirar o valor necessário”. Deverá ser dividida em:
 - História Atômica 1: Como um cliente, eu quero consultar meu saldo, para que eu possa saber quanto tenho disponível.
 - História Atômica 2: Como um cliente, eu quero sacar dinheiro, para que eu possa retirar o valor necessário.
 - Porém outros verbos de ação podem ser considerados como fins, quando não descrevem a ação principal da história. Eles geralmente vêm após um “para” e descrevem o benefício ou propósito da ação. Por exemplo, no exemplo abaixo, “visualizar” é o verbo principal e “verificar” e “repetir” são os resultados esperados.
 - História Atômica 3: Como um cliente, eu quero visualizar meu histórico de pedidos para que eu possa verificar compras passadas e repetir pedidos facilmente.
- Validação dos Critérios: Verifique se cada história atende a todos os critérios: bem-formado, atomicidade, minimalidade, solidez conceitual, não ambiguidade, frase completa e estimabilidade. Se não atender, ajuste a história até que todos os critérios sejam satisfeitos.
- Revisão e Refinamento: Após criar as histórias, revise cada uma para garantir que atenda a todos os critérios, especialmente a atomicidade. Se uma história ainda aborda múltiplas funcionalidades, divida-a em partes menores.]

- **Resposta esperada do Chat GPT: Entendido.**

6. Atividade 6 (Few-Shot Prompt)

- **Descrição da atividade:** Apresenta alguns exemplos de histórias bem formadas para auxiliar na criação das histórias de usuário.
- **Instruções da atividade:** Insira o prompt no ChatGPT.

- **Prompt a inserir:**

[Eu vou lhe fornecer alguns exemplos de histórias de usuários bem formadas para lhe auxiliar na criação das histórias. Utilize-os para estruturar sua resposta, sempre que for solicitado a gerar o máximo de histórias de usuário possíveis.

- **Exemplos de Histórias Bem Definidas:**

- História de Usuário Atômica 4: Criar Tarefas

- Como um gerente,
Eu quero criar tarefas,
Para que eu possa adicionar atividades e distribuí-las para a equipe.

- História de Usuário Atômica 5: Atualizar Tarefas

- Como um gerente,
Eu quero atualizar tarefas,
Para que eu possa modificar detalhes das atividades existentes conforme necessário.]

- **Resposta esperada do Chat GPT: Entendido.**

7. Atividade 7 (Geração dos User Stories)

- **Descrição da atividade:** Solicita a geração das histórias de usuário.

- **Instruções da atividade:** Insira o prompt no ChatGPT.

- **Prompt a inserir:**

[Atuando como um Engenheiro de RA, com muita experiência em elicitação de requisitos. Considere todas as informações que te concedi previamente sobre o produto de software que estou construindo, sobre o template, critérios, orientações e exemplos que te forneci para a criação de histórias de usuário, e agora, não responda somente "Entendido" e gere o máximo de histórias de usuários que conseguir, por favor gere também mais histórias que considerem os aspectos adicionais que possam melhorar a experiência do usuário e a eficiência do sistema, sinalize que essas são histórias adicionais.]

- **Resposta esperada do Chat GPT:** Conjunto de histórias de usuário de acordo com template e com os requisitos.

D

CONTEXTOS DOS SISTEMAS DESENVOLVIDOS PELOS PARTICIPANTES

Contextos do Sistemas Desenvolvidos

Link para a planilha:

https://docs.google.com/spreadsheets/d/1B6P5XTe_lqbW2iComtYACWg4vzmFFSKL/edit?usp=sharing&ouid=116129903176169444833&rtpof=true&sd=true

Projetos:

1. Grupo A - Sistema de Restaurante com controle de pedidos, preparação e entrega

• Visão de Produto:

Primeiramente, ao abrir o sistema, aparecerá uma tela para efetuar o login. Para o usuário acessar, é necessário inserir seu número de telefone e uma senha previamente estabelecida durante o registro, e depois clicar no botão “entrar”. Caso ainda não tenha um registro, abaixo do botão estará escrito “se você não possui um registro, clique aqui para criar”, permitindo que você clique nessa afirmação, que o levará à tela de registro.

A tela de registro conterá campos para que o usuário preencha os seguintes dados: Nome completo, Número de Telefone, senha e confirmação de senha, seguida de um botão escrito “terminar o registro”. Assim que o login for concluído, abrir-se-á uma tela que exibirá a logo do Ineide no canto superior esquerdo, uma barra de pesquisa na parte superior central, e no canto superior direito haverá uma imagem de uma cesta de compras, com a palavra “carrinho” (que será um ícone). Ao clicar nessa imagem, abrirá uma tela dividida em uma tabela que conterá: quantidade (com ícones de menos e mais ao lado esquerdo e direito da quantidade de produtos), item (nome do produto) e valor (preço do produto). No final, haverá uma linha com a palavra “total” seguida do valor total dos produtos selecionados e uma pequena caixinha escrito “OBS sobre o pedido” para escrever observações sobre o pedido, e um botão ao final escrito “pagar!”, que levará para a tela de pagamento por Pix ou cartão.

Saindo do carrinho e voltando para a tela principal, teremos um “cardápio”. O cardápio será organizado em diversas categorias (doces, sanduíches, bebidas, etc.) e em cada uma delas haverá vários quadrados com a foto do produto, o preço, o nome e uma descrição. Nos cantos superiores direito e esquerdo das fotos, respectivamente, estarão pequenas cestinhas com símbolos de positivo e negativo (pequenos ícones para adicionar ou remover o produto do carrinho).

Algumas informações não mencionadas anteriormente estão relacionadas ao cliente e serão destacadas agora. Na tela de login, haverá um pequeno ícone que se assemelha a uma lojinha no canto inferior direito. Ao clicar nesse ícone, abrirá uma tela com a mensagem “essa parte é para o dono e funcionários do estabelecimento”, onde será necessário inserir uma senha muito complexa, que apenas o proprietário do sistema saberá (NEIDINHA).

Após clicar em “entrar”, serão apresentadas duas opções na parte superior para selecionar: “pedidos” e “cardápio”. No entanto, a tela já abrirá na aba “pedidos”, onde veremos vários quadrados dispostos em fila, representando os pedidos feitos pelos usuários no sistema. Esses quadrados serão clicáveis, permitindo uma verificação mais detalhada do que foi solicitado e das informações do cliente e do produto (nome, telefone e OBS sobre o pedido). Também haverá um botão escrito “Pedido Feito”, que, ao ser clicado, mostrará na tela de pedidos um ícone de verificado no quadrado, e a opção “Entregue” removerá esse pedido da fila.

Ao clicar em “cardápio”, será exibida a mesma tela que o usuário vê, mas com pequenas diferenças. As fotos dos produtos terão um “X” no lado direito, que servirá para remover aquele produto da venda, e um lápis para representar a edição das informações do produto do lado esquerdo (neste caso, ao clicar, o sistema redirecionará para a mesma tela de cadastro de um novo produto, que será detalhada posteriormente, mas já exibindo as informações previamente inseridas). No final dos produtos de cada categoria, haverá um “+” que, ao ser clicado, levará à tela de cadastro de um novo produto. Nela, haverá espaço para inserir uma imagem, um nome e uma descrição, e em seguida você poderá clicar em “Pronto” para adicionar o novo produto (ou editar o já existente).

Usuários:

1. Dono do estabelecimento/ funcionários
2. Pessoas que farão pedidos

Ações:

1. Dono do estabelecimento/ funcionários Gerenciar
 - Gerenciar
 - Pedidos Marcar Pedido Como Feito
 - estabelecer Pedido como entregue
 - Editar Produto do Cardápio
 - Adicionar Novo Produto ao Cardápio
 - Remover Produto do Cardápio
2. Pessoas que farão pedidos
 - Login no Sistema
 - Registro de Novo Usuário
 - Visualizar Carrinho de Compras
 - Adicionar Produto ao Carrinho
 - Remover Produto do Carrinho
 - Pagar Pedido
 - Acessar o Cardápio
 - Inserir Observação no Pedido

2. Grupo B - Sistema de Horas Complementares

• Visão de Produto:

Gerenciar as horas complementares de alunos da graduação.

Usuários:

1. Alunos
2. Professores
3. Administradores

Ações:

1. Alunos

- Submeter atividades complementares: Enviar documentos e informações sobre cursos, palestras, projetos, etc.
- Consultar horas acumuladas: Visualizar o total de horas já aprovadas e as pendentes.
- Acompanhar status de submissões: Verificar o andamento da análise das atividades submetidas (aprovado, pendente, rejeitado).
- Editar ou cancelar submissões: Modificar ou retirar submissões enviadas caso haja algum erro.
- Consultar regras: Ver as regras e orientações para submissão de atividades complementares (como o máximo de horas por atividade).

2. Professores

- Avaliar submissões: Validar ou rejeitar as atividades submetidas pelos alunos.
- Verificar documentação: Consultar e revisar os documentos anexados pelos alunos.
- Enviar feedback: Dar justificativa para aprovação ou rejeição de atividades.
- Consultar histórico de avaliações: Acessar o histórico de horas aprovadas e rejeitadas de cada aluno.

3. Administradores

- Gerenciar usuários: Criar, editar ou remover perfis de alunos e professores.
- Configurar regras de submissão: Ajustar as regras e limites de horas por tipo de atividade.
- Visualizar relatórios gerais: Acessar relatórios globais de todas as horas complementares submetidas e aprovadas no sistema.
- Gerenciar permissões: Definir o que cada usuário (aluno, professor, ou outro) pode ou não fazer no sistema.

3. Grupo C - Aplicação de Sugestão de Filmes

• **Visão de Produto:**

O sistema a ser desenvolvido é uma aplicação de sugestão de filmes que utilizará algoritmos de recomendação para indicar filmes relevantes a cada usuário. A aplicação permitirá o cadastro de usuários, personalização de perfis e as indicações.

Usuários:

1. Usuário Comum (Cliente)

Ações:

1. Usuário Comum (Cliente)

Ações

- O usuário comum é aquele que utiliza o sistema para descobrir novos filmes e avaliar os filmes que já assistiu.
- Ele pode navegar pelas sugestões fornecidas, fazer buscas, visualizar informações detalhadas dos filmes e registrar feedback sobre eles (como notas ou comentários).

Ações no Sistema

- o Cadastro/Login
- o Preencher perfil de preferências (gêneros, filmes favoritos, diretores, etc.)
- o Visualizar sugestões de filmes
- o Avaliar e comentar sobre os filmes assistidos
- o Salvar filmes como "para assistir" ou marcar como "assistido"
- o Receber sugestões com base no histórico de visualizações e preferências

4. Grupo D - Sistema de Gestão de Finanças Pessoais para Idoso

• **Visão de Produto:**

Sistema de Gestão de Finanças Pessoais para Idoso.

Usuários:

1. Idosos

Ações:

1. Idosos

- Registro de novos usuários.
- Perfis personalizados para cada usuário.
- Login.
- Logout.
- Adição, edição e exclusão de receitas e despesas.

- Categorização de despesas (ex: alimentação, transporte, lazer).
- Definição de metas de economia.
- Alerta de pagamentos.
- Geração de relatórios mensais e anuais.
- Gráficos de receitas, despesas e saldo FAQs.

5. Grupo E - TeseLab - Uma ferramenta web para realização de Debate de Teses

• Visão de Produto:

Para professores e estudantes universitários que participam de debates acadêmicos em sala de aula, o TeseLab é uma ferramenta de gestão e automação de debates de teses que organiza, distribui e monitora o processo de discussão e revisão de teses de forma eficiente. Diferentemente do método manual baseado em planilhas e e-mails, o nosso produto automatiza a distribuição de teses e revisores, gerencia os prazos, redistribui papéis em caso de atrasos e garante o anonimato dos alunos, proporcionando uma experiência de debate estruturada e imparcial.

Usuários:

1. Professor
2. Aluno

Ações:

1. Professor

- O sistema deve permitir que professores e alunos realizem cadastro com e-mail e senha.
- O sistema deve categorizar usuários como "professores" ou "alunos", com permissões distintas.
- O sistema deve permitir que o professor crie novas "salas" de debate, especificando o título, descrição, número de teses e data limite.
- O sistema deve permitir ao professor adicionar ou remover alunos de uma sala a qualquer momento.
- O sistema deve permitir ao professor ajustar os prazos das atividades a qualquer momento.
- O sistema deve permitir ao professor visualizar o progresso de cada participante e revisor.
- O sistema deve fornecer ao professor a capacidade de adicionar ou remover alunos durante o debate.
- O sistema deve gerar relatórios com o conteúdo das argumentações e revisões realizadas.
- O sistema deve permitir que o professor acompanhe as interações, comentários e argumentações feitas pelos alunos.
- O sistema deve fornecer uma visão geral do progresso de cada participante em relação aos prazos.

- O sistema deve garantir que apenas o professor tenha acesso a todas as identidades e argumentações.
- O sistema deve permitir que o professor acesse e exporte o conteúdo dos debates finalizados (em PDF ou outro formato) para arquivamento ou revisão posterior.

2. Aluno

- O sistema deve permitir que professores e alunos realizem cadastro com e-mail e senha.
- O sistema deve categorizar usuários como "professores" ou "alunos", com permissões distintas.
- O sistema deve distribuir automaticamente as teses entre os alunos, designando cada um como autor de uma tese.
- O sistema deve designar dois revisores (A e B) aleatórios para cada tese, garantindo que um participante não revise sua própria tese ou seja revisor em duas teses como A ou B.
- O sistema deve redistribuir os revisores automaticamente caso haja atrasos ou abandono de alunos.
- O sistema deve enviar notificações aos alunos informando sua tese, os papéis de revisor, e as datas limite para cada fase do debate.
- O sistema deve notificar os alunos quando o prazo para submissão da argumentação inicial ou revisão estiver se aproximando.
- O sistema deve permitir que os alunos enviem sua argumentação inicial na aba correspondente à sua tese.
- O sistema deve permitir que os alunos A e B enviem suas revisões nas teses atribuídas após o envio da argumentação inicial pelo autor.
- O sistema deve garantir que todas as interações entre os alunos sejam anônimas.
- O sistema deve notificar os alunos para enviar uma argumentação final, considerando os feedbacks dos revisores.
- O sistema deve permitir que os alunos editem e finalizem suas argumentações finais após a revisão.
- O sistema deve verificar, na data limite, quais alunos não realizaram suas tarefas (argumentação ou revisão).
- O sistema deve redistribuir automaticamente os papéis de revisor entre os alunos ativos caso haja desistências ou atrasos.
- O sistema deve garantir que os alunos não possam identificar quem são seus revisores ou quem revisou suas tese.
- O sistema deve armazenar um histórico dos debates e permitir que os alunos acessem suas argumentações após o encerramento do debate.

6. Grupo F - Comparador de Aplicativos de Serviço de Transporte

• Visão de Produto:

Um app que compara as informações de viagens entre diferentes serviços de transporte (Uber, InDrive e 99) para ajudar o usuário escolher a melhor viagem.

Usuários:

1. Pessoa que quer um serviço de transporte barato

Ações:

1. Pessoa que quer um serviço de transporte barato
 - Inserir endereço de origem e destino da viagem
 - Seleciona os serviços de transporte em que quer buscar a viagem
 - Escolher o tipo de transporte (moto ou carro)
 - Escolher o tipo de viagem (passageiro ou entrega)
 - Obter a lista de viagens de cada serviço de transporte
 - Escolher a viagem que mais me agrada e ser redirecionado ao app do serviço de transporte escolhido
 - Salvar minhas prioridades na busca de viagens (preço baixo, tempo para chegar até mim curto, avaliação do motorista) para ordenar os resultados da busca
 - Cadastrar os destinos favoritos, como casa, trabalho, escola etc.
 - Ver o histórico de viagens (incluindo detalhes como valor, trajeto, avaliação de motorista) selecionadas
 - Ver qual serviço de transporte eu costumo selecionar mais.

E

**EXEMPLOS DE CENÁRIOS DE SISTEMAS
UTILIZADOS NO ESTUDO DA
FERRAMENTA STORY AI GENERATOR**

Exemplos de Cenários de Sistemas

(Copie e cole nos campos apropriados)



1. Sistema de Gestão de Entregas Locais

Visão do Produto:

Um aplicativo que conecta pequenos comércios a entregadores independentes para realizar entregas rápidas em regiões urbanas.

Principais Usuários:

- Comerciante – Proprietário de um pequeno negócio local
- Entregador – Profissional autônomo
- Cliente - Comprador do produto

Principais Ações:

Comerciante:

- Registrar pedidos e endereços de entrega.
- Acompanhar o status das entregas.
- Avaliar entregadores.

Entregador:

- Receber e aceitar pedidos disponíveis próximos.
- Atualizar status da entrega.
- Visualizar histórico de entregas e ganhos.

Cliente:

- Acompanhar o status da entrega em tempo real.
 - Avaliar a entrega e o serviço.
-



2. Plataforma de Acompanhamento Escolar

Visão do Produto:

Um sistema para facilitar a comunicação entre professores, alunos e responsáveis, permitindo acompanhar notas, atividades e desempenho escolar.

Principais Usuários:

- Professor – Docente
- Aluno – Criança matriculada
- Responsável – Pai, mãe ou guardião legal
- Coordenador Pedagógico – Membro da gestão escolar

Principais Ações:**Professor:**

- Lançar notas e presenças.
- Enviar atividades e materiais.
- Enviar comunicados à turma e aos responsáveis.

Aluno:

- Visualizar notas, tarefas e materiais.
- Enviar respostas e trabalhos.
- Receber feedback dos professores.

Responsável:

- Acompanhar o desempenho do aluno.
- Receber notificações de faltas e responder à comunicados.

Coordenador:

- Gerar relatórios de turmas e desempenho geral.
 - Cadastrar professores e disciplinas.
-

**3. Sistema de Coleta de Resíduos Recicláveis****Visão do Produto:**

Um aplicativo que conecta moradores a cooperativas de reciclagem, permitindo solicitar coletas e acompanhar o impacto ambiental gerado.

Principais Usuários:

- Morador – Morador de algum bairro
- Coletor da Cooperativa – Funcionário da cooperativa
- Gestor Ambiental – Responsável pelo direcionamento das estratégias adotadas.

Principais Ações:**Morador:**

- Cadastrar endereço e tipo de resíduo.
- Solicitar coleta.
- Acompanhar histórico de coletas.

Coletor da Cooperativa:

- Visualizar solicitações na região.
- Aceitar coletas e registrar conclusão.
- Gerar relatórios de materiais recolhidos.

Gestor Ambiental:

- Acompanhar métricas de reciclagem.
 - Gerar relatórios por bairro ou tipo de resíduo.
-



4. Sistema de Solicitação de Serviços Públicos

Visão do Produto:

Uma plataforma que permite aos cidadãos registrar solicitações de serviços públicos (como iluminação, limpeza ou buracos nas vias), acompanhando o andamento da resolução.

Principais Usuários:

- Cidadão - Morador de algum bairro
- Servidor da Prefeitura - Servidor do Departamento de Obras/Serviços Públicos
- Gestor Municipal - Coordenador de Departamento de Obras/Serviços Públicos

Principais Ações:

Cidadão:

- Registrar ocorrências com foto e localização.
- Acompanhar o status do pedido.
- Avaliar a resolução do problema.

Funcionário da Prefeitura:

- Visualizar e priorizar solicitações.
- Atualizar status da solicitação.
- Registrar ações realizadas.

Gestor Municipal:

- Gerar relatórios de demandas por bairro.
 - Analisar tempo médio de atendimento.
 - Redirecionar recursos e equipes.
-



5. Sistema de Adoção de Animais

Visão do Produto:

Uma aplicação que conecta abrigos e protetores independentes a pessoas interessadas em adotar animais de estimação, promovendo a adoção responsável e reduzindo o número de animais em situação de abandono.

Principais Usuários:

- Adotante – Pessoa interessada em adotar um animal de estimação.

- Protetor do Abrigo – Responsável pelo cuidado e disponibilização dos animais para adoção.
- Administrador – Responsável pela gestão geral da plataforma.

Principais Ações:

Adotante:

- Buscar animais disponíveis por espécie, porte e localização.
- Preencher um perfil de adoção com informações sobre o ambiente e rotina.
- Agendar visitas ou entrevistas com protetores e abrigos.

Protetor do Abrigo:

- Cadastrar novos animais com fotos, histórico e informações de saúde.
- Atualizar o status da situação dos animais.
- Avaliar e validar candidatos com base nos perfis enviados.

Administrador:

- Gerenciar cadastros de usuários e abrigos.
- Supervisionar campanhas de adoção e eventos.
- Monitorar estatísticas de adoções e acessos.

Regras de Negócio:

- Cada adoção deve ser registrada com o consentimento do protetor responsável.
- Animais só podem ser marcados como “adotados” após aprovação formal.



6. Sistema de Gestão de Moradias Estudantis

Visão do Produto:

Um sistema para universidades administrarem residências estudantis, com controle de vagas, solicitações, manutenções e acompanhamento de ocupações, garantindo eficiência e transparência na gestão.

Principais Usuários:

- Estudante – Morador ou candidato à moradia estudantil.
- Administrador da Moradia – Funcionário responsável pela gestão das residências.
- Profissional da Manutenção – Profissionais responsáveis pelos reparos nas residências.

Principais Ações:

Estudante:

- Solicitar vaga na residência informando seus dados e justificativas.

- Registrar problemas de infraestrutura em seu quarto ou bloco.
- Acompanhar o status da solicitação ou do reparo solicitado.

Administrador da Moradia:

- Aprovar ou recusar pedidos de vaga com base nos critérios institucionais.
- Gerenciar quartos, blocos e registros de ocupação.
- Emitir relatórios sobre taxa de ocupação e demandas de manutenção.

Profissional da Manutenção:

- Receber ordens de serviço geradas pelo sistema.
- Atualizar o status dos reparos realizados (pendente, em andamento, concluído).
- Notificar o administrador e o estudante após a conclusão do serviço.

Regras de Negócio:

- Cada estudante pode ter apenas uma solicitação ativa por período.
 - Ordens de manutenção devem ser registradas e rastreadas até sua conclusão.
-

7. Sistema de Monitoramento Ambiental

Visão do Produto:

Um sistema voltado para o monitoramento da qualidade ambiental em áreas urbanas e rurais, integrando dados sobre poluição, desmatamento e condições climáticas para apoiar políticas públicas e ações sustentáveis.

Principais Usuários:

- Técnico Ambiental – Profissional responsável pela coleta e análise de dados em campo.
- Gestor Ambiental – Responsável pelo direcionamento das estratégias adotadas.
- Cidadão – Morador preocupado com os dados ambientais locais.

Principais Ações:**Técnico Ambiental:**

- Registrar medições de qualidade do ar, água e solo.
- Inserir relatórios e fotografias georreferenciadas.
- Atualizar status de monitoramento das áreas acompanhadas.

Gestor Ambiental:

- Gerar relatórios e gráficos de tendência ambiental por região.
- Identificar áreas críticas e propor planos de ação.

- Integrar informações com órgãos públicos e universidades.

Cidadão:

- Consultar indicadores ambientais em tempo real.
- Receber alertas sobre níveis de poluição ou riscos ambientais.
- Acompanhar projetos de recuperação ambiental na sua região.

Regras de Negócio:

- As medições devem ser validadas por técnicos credenciados antes da publicação.
 - O sistema deve manter histórico de dados para comparações anuais.
-

F

**QUESTIONÁRIO UTILIZADO NO ESTUDO
DA FERRAMENTA STORY AI
GENERATOR**

Ferramenta Story AI Generator - TAM

* Indica uma pergunta obrigatória

1. E-mail *

Caracterização do Participante

2. Onde você utiliza ou já utilizou Histórias de Usuários? *

Marcar apenas uma oval.

☐ Na academia

☐ Na indústria

☐ Ambos

3. Com que frequência você especifica Histórias de Usuários? *

Marcar apenas uma oval.

☐ Diariamente

☐ Semanalmente

☐ Mensalmente

☐ Esporadicamente

4. Quantos anos de experiência você possui em especificação de requisitos com Histórias de Usuários? *

Marcar apenas uma oval.

- ☐ Menos de 1 ano
- ☐ Menos de 5 anos
- ☐ Mais de 5 anos
- ☐ Mais de 10 anos
- ☐ Mais de 20 anos

5. Em qual instituição você trabalha atualmente? *

6. Em qual cidade/estado/país está localizada sua instituição? *

7. Qual seu grau de familiaridade com ferramentas baseadas em IA generativa (por exemplo, ChatGPT, Copilot): *

Marcar apenas uma oval.

- ☐ Nenhuma
- ☐ Baixa
- ☐ Média
- ☐ Alta
- ☐ Muito Alta

8. Já utilizou alguma ferramenta automatizada ou IA para apoiar a escrita de Histórias de Usuários?

*

Marcar apenas uma oval.

☐ Sim

☐ Não

9. Se sim, quais ferramentas já utilizou?

Modelo de Aceitação de Tecnologia (TAM)

As escalas são organizadas em categorias que avaliam a percepção do usuário sobre a utilidade e a facilidade de uso de uma tecnologia. Por favor, responda com base em suas percepções.

10. Informe a descrição de qual sistema você utilizou na geração das histórias de usuário na ferramenta *StoryAI Generator*:

11. Sobre a utilidade da ferramenta para gerar Histórias de Usuários *

Marcar apenas uma oval por linha.

	Concordo fortemente	Concordo	Não concordo, nem discordo	Discordo	Discordo fortemente
U1. Usar a ferramenta melhora meu desempenho na criação de histórias de usuários.	☐	☐	☐	☐	☐
U2. Usar a ferramenta em meu trabalho poderá melhorar minha produtividade.	☐	☐	☐	☐	☐
U3. Usar a ferramenta melhora a eficácia na geração das histórias de usuários geradas para o meu projeto.	☐	☐	☐	☐	☐
U4. Considero a ferramenta útil no meu trabalho.	☐	☐	☐	☐	☐

12. Se possível, explique o motivo da sua resposta anterior:

13. Sobre a facilidade do uso da ferramenta para gerar Histórias de Usuários *

Marcar apenas uma oval por linha.

	Concordo fortemente	Concordo	Não concordo, nem discordo	Discordo	Discordo fortemente
P1. Minha interação com a ferramenta para geração de histórias de usuários foi clara e compreensível.	☐	☐	☐	☐	☐
P2. Interagir com a ferramenta para gerar histórias de usuários não exigiu muito esforço mental.	☐	☐	☐	☐	☐
P3. Eu considero a ferramenta fácil de utilizar.	☐	☐	☐	☐	☐
P4. Eu acho fácil fazer com que a ferramenta gere histórias de usuários.	☐	☐	☐	☐	☐

14. Se possível, explique o motivo da sua resposta anterior:

15. **Sobre intenção de uso futuro da ferramenta para Histórias de Usuários ***

Marcar apenas uma oval por linha.

	Concordo fortemente	Concordo	Não concordo, nem discordo	Discordo	Discordo fortemente
I1. Supondo que eu tenha acesso a essa ferramenta, eu a utilizaria para gerar minhas histórias de usuários.	☐	☐	☐	☐	☐
I2. Se eu tiver acesso a essa ferramenta, pretendo usá-lo para gerar minhas histórias de usuários em futuros projetos.	☐	☐	☐	☐	☐
I3. Eu pretendo utilizar a ferramenta nos próximos meses para criar histórias de usuários.	☐	☐	☐	☐	☐

16. Se possível, explique o motivo da sua resposta anterior:

17. **Sobre a relevância da ferramenta para o trabalho com Histórias de Usuários**

*

Marcar apenas uma oval por linha.

	Concordo fortemente	Concordo	Não concordo, nem discordo	Discordo	Discordo fortemente
R1: No meu trabalho, o uso desta ferramenta é importante.	☐	☐	☐	☐	☐
R2: No meu trabalho, o uso desta ferramenta é relevante.	☐	☐	☐	☐	☐
R3: O uso desta ferramenta é pertinente às minhas diversas tarefas relacionadas à especificação de histórias de usuários.	☐	☐	☐	☐	☐

18. Se possível, explique o motivo da sua resposta anterior:

19. **Sobre a qualidade da saída da ferramenta para Histórias de Usuários ***

Marcar apenas uma oval por linha.

	Concordo fortemente	Concordo	Não concordo, nem discordo	Discordo	Discordo fortemente
O1: A qualidade da saída que obtenho com a ferramenta é alta.	☐	☐	☐	☐	☐
O2: Não tenho problemas com a qualidade da saída da ferramenta.	☐	☐	☐	☐	☐
O3: Considero a qualidade das saídas da ferramenta excelente.	☐	☐	☐	☐	☐

20. Com base na sua resposta anterior, responda se você ficou satisfeito com as histórias de usuário geradas pela ferramenta? Elas representam bem os requisitos esperados, na sua opinião? *

21. Você tem comentários ou sugestões de como melhorar a ferramenta *StoryAI Generator*?

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

G

CÓDIGOS EMERGIDOS NA ETAPA DE CODIFICAÇÃO ABERTA

Codebook
[CAT] A estrutura da interface possui campos intuitivos para preencher o contexto
[CAT] A ferramenta apóia a geração de histórias
[CAT] A ferramenta enfrenta dificuldades para uso na indústria
[CAT] A ferramenta melhora a performance
[CAT] Abstração das Informações Adicionais passadas
[CAT] As histórias geradas estão alinhadas ao contexto informado
[CAT] As recomendações foram negativas
[CAT] As recomendações foram positivas/neutras
[CAT] Diferentes formas de usar a ferramenta
[CAT] É mais intuitivo que IAs de uso geral
[CAT] Estrutura de entrada da ferramenta não é positiva
[CAT] Fácil de usar e exige pouco esforço
[CAT] Falta de qualidade nas histórias geradas
[CAT] Falta de RNF
[CAT] Falta dos Critérios de Aceitação
[CAT] Fenômeno Informações Adicionais
[CAT] Fenômeno Melhorias Futuras
[CAT] Fenômeno Recomendações adicionais feitas pela IA
[CAT] Fenômeno Uso da Ferramenta
[CAT] Geração de novos elementos
[CAT] Gerou dúvidas durante o uso
[CAT] Melhoria na proteção dos dados para uso na indústria
[CAT] Melhorias na interface da ferramenta
[CAT] Melhorias no prompt/processo de geração atual
[CAT] Pontos Positivos
[CAT] Realizar novos estudos
[CAT] Uso com revisão posterior
[CAT] Uso como ponto de partida
[CAT] Uso em teste e treinamento, na academia
[CAT] Uso para gerar histórias e melhorar a performance
[CAT] Uso para validação com stakeholder
Abstração - Apontou que o campo de informações adicionais não foi refletido nas histórias geradas
Abstração - Considerou a história gerada com base nas informações adicionais genérica
Abstração - Destacou que uma questão tecnológica, não afetaria tanto nas histórias
[CAT] Eficácia da ferramenta
[CAT] Fenômeno Pontos Negativos
Critérios - Acredita que a inclusão de critérios agregaria valor
Critérios - Apontou que a ferramenta menciona, mas não gera os critérios de aceitação
Critérios - Crê que a falta de critérios de aceitação reduz a importância da ferramenta
Critérios - Sentiu falta dos critérios de aceitação das histórias
Dúvida - Teve dúvida sobre a priorização dos elementos feita pela IA
Dúvida - Teve dúvida sobre o nível de granularidade que deve ser fornecido

Dúvida - Teve dúvida sobre os elementos que são considerados para gerar as histórias
Eficácia - Achou a ferramenta útil para a geração de histórias
Eficácia - As histórias atenderam ao contexto fornecido
Eficácia - As histórias são coerentes com o contexto
Eficácia - Atendeu a expectativa
Eficácia - Considera que faria igual
Eficácia - Considerou que facilita a criação das histórias
Eficácia - Contribuiu para a geração de histórias
Eficácia - Demonstra a capacidade de geração de histórias
Eficácia - Pondera que as histórias ficaram adequadas por ser um sistema pequeno
Eficácia - Proveu suporte na geração das histórias
Eficácia - Expõe que a ferramenta trouxe as dez histórias exatamente como ela havia feito
Eficácia - A ferramenta apresentou todos os possíveis sub-cenários, derivados do cenário principal
Eficácia - A ferramenta conseguiu receber os dados e gerar as histórias
Eficácia - A ferramenta gerou as histórias alinhadas, de maneira clara e rápida
Eficácia - As histórias geradas foram exatamente as que ela gerou no seu projeto
Eficácia - As histórias são coerentes com o cenário especificado
Eficácia - As histórias trouxeram possíveis funcionalidades
Eficácia - Conhecia a saída, e foi a mesma gerada por ela
Eficácia - Crê que usar a ferramenta irá lhe apoiar na escrita das histórias
Eficácia - Destaca que conhece bem o cenário
Eficácia - Pondera que faltou incluir nas regras de negócio para não gerar algumas histórias
Método - Achou parecido com a criação de sistemas no Lean Inception
Negativo - A ferramenta gerou histórias parecidas
Negativo - Acredita a segmentação da entrada não é confortável para todos
Negativo - Considera o risco de profissionais ficarem preguiçosos
Negativo - Não gostou de uma história gerada
Negativo - Percebeu algumas inconsistências nas histórias
Negativo - Percebeu que a ferramenta perdeu informações adicionais
Negativo - QL A ferramenta gerou épicos
Negativo - QL A ferramenta gerou histórias parecidas
Negativo - Se manteve neutro, pois são vários passos para chegar ao resultado final
Negativo - Sentiu falta de maior detalhamento nas histórias
Negativo - Sentiu que algumas histórias foram genéricas
Negativo - Usaria se mostrasse os critérios de aceitação
Outras LLM - Acredita que nas IAs de uso geral, o resultado é verboso e menos organizado
Outras LLM - Apontou a impaciência do usuário, comparada com outras LLMs
Outras LLM - Considera que descrever o contexto do projeto não é tão intuitivo nas IAs de uso geral
Outras LLM - Considera que um único campo é genérico, como ocorre nas IAs de uso geral
Outras LLM - Contrapôs que em uma LLM, só precisaria de um prompt
Percepção - Considerou a ferramenta relevante

Percepção - Gostou da ferramenta
Percepção - Gostou de usar
Percepção - Gostou do layout da ferramenta
Positivo - A ferramenta apoiou a definição do contexto
Positivo - A ferramenta apoiou o processo criativo
Positivo - A ferramenta é simples de entender
Positivo - A ferramenta já é bastante útil só gerando histórias
Positivo - A qualidade depende dos dados inseridos
Positivo - Acha que a ferramenta ajudará a poupar tempo
Positivo - Acha que a ferramenta ajudará a reduzir o esforço cognitivo
Positivo - Acha que a ferramenta tem potencial para auxiliar na criação de US
Positivo - Achou a ferramenta interessante por automatizar a tarefa
Positivo - Achou a interface bem organizada e limpa
Positivo - Achou clara a saída da ferramenta
Positivo - Achou intuitivo preencher os campos da ferramenta
Positivo - Achou que a ferramenta tem boa compreensão dos elementos solicitados para gerar as histórias
Positivo - Afirma que as saídas foram boas
Positivo - Afirma que facilitará o trabalho do analista de requisitos
Positivo - Afirma que melhora sua performance na escrita das histórias
Positivo - Afirma que uma breve explicação já basta
Positivo - Afirma que, com a ferramenta, não precisará mais escrever as histórias
Positivo - Aponta que a estrutura apoia a descrição do contexto
Positivo - Aponta que a estrutura dividida em campos facilita o uso
Positivo - Aponta que a estruturação das respostas veio na sequência lógica esperada
Positivo - Aponta que apenas lendo as instruções nos campos foi o suficiente para lhe direcionar
Positivo - Aponta que as orientações na interface facilitam o uso da ferramenta
Positivo - As histórias estão no formato conhecido
Positivo - As histórias são relevantes
Positivo - Comenta a possibilidade de copiar os dados
Positivo - Confessa que colocou pouca informação para testar a qualidade e mesmo assim o resultado foi bom
Positivo - Considerou a apresentação dos resultados com template, campos e atores torna a visualização mais eficiente
Positivo - Considerou a capacidade de agilizar trabalhos manuais e repetitivos
Positivo - Considerou a capacidade de fomentar novas funcionalidades
Positivo - Considerou boa a interação com a ferramenta
Positivo - Considerou fácil de entender
Positivo - Considerou fácil de interagir
Positivo - Considerou fácil de usar
Positivo - Considerou fácil saber o que deveria ser preenchido
Positivo - Considerou intuitiva
Positivo - Considerou que o resultado foi bem abrangente
Positivo - Considerou que o resultado gerado foi bom

Positivo - Considerou que os resultados são apresentados de forma estruturada e organizada
Positivo - Crê que o tempo para treinamento parece ser pequeno
Positivo - Exigiu pouca necessidade de interação
Positivo - Funcionou como suporte para agilizar a atividade de geração
Positivo - Gerou sem esforço
Positivo - Gerou uma boa quantidade de histórias
Positivo - Os exemplos auxiliam o preenchimento da ferramenta
Positivo - Permitiu geração rápida
Positivo - Pode melhorar a performance de quem está especificando histórias
Positivo - Possibilitará se dedicar a outras atividades
Positivo - Quase todas poderiam ser usadas no projeto
Positivo - Relatou que não teve dúvida para preencher
Positivo - Requer poucas informações
Positivo - Se disse satisfeita com a saída da ferramenta
Positivo - Se esforçou apenas em pensar no cenário do projeto e nas regras de negócio
Recomendações - A ferramenta deu sugestões que não haviam sido pensadas
Recomendações - As recomendações feitas pela IA estão dentro do contexto do problema
Recomendações - Considerou incorreta uma história sugerida pela IA
Recomendações - Considerou que as recomendações feitas pela IA apoiaram no processo criativo
Recomendações - Contrapôs que analistas e POs podem achar negativa a iniciativa da IA
Recomendações - Gostou que a ferramenta abrangeu mais do que o esperado
Recomendações - Sugestões além do escopo original
Recomendações - Viu como negativa as sugestões feitas pela IA
Recomendações - Viu como positiva as sugestões feitas pela IA
Restrições de Uso - Ressalta que ferramentas que usam dados exigem validação das empresas
Restrições de Uso - Ressalta que o uso da ferramenta na indústria precisa ser aprovado
Restrições de Uso - Ressalta que para o uso na indústria exige-se pensar na privacidade dos dados
RNF - Considerou incorreto o formato dos não funcionais
RNF - Necessidade de tratar requisitos não funcionais
RNF - QL A ferramenta não considerou requisito não funcional
RNF - QL A ferramenta não considerou todo o contexto descrito
Sugestão - Editar/refinar as histórias na ferramenta
Sugestão - Rotular o que a IA gerou
Sugestão - Acha que as informações adicionais poderiam ser armazenadas e mostradas na planilha entregue
Sugestão - Acredita que a inclusão opcional de entregáveis associados às histórias agregaria valor
Sugestão - Considera que tratando dados sensíveis, a ferramenta terá maior potencial de uso na indústria
Sugestão - Dar a opção de retornar ao formulário para restringir mais
Sugestão - Disponibilizar a descrição fornecida para que possa avaliar as histórias geradas

em comparação com o texto fornecido
Sugestão - Fazer o contrário, a ferramenta poderia gerar cenários com base em histórias
Sugestão - Incluir a geração da documentação
Sugestão - Incluir a geração da especificação dos requisitos
Sugestão - Indicar a obrigatoriedade dos campos
Sugestão - Possibilitar adicionar mais instruções sobre os requisitos
Sugestão - QL Melhoria futura - Gerar critérios de aceitação
Sugestão - Selecionar as histórias para exportar
Sugestão - Testar com sistemas mais complexos e mais longos
Sugestão - Um processo iterativo de geração/refinamento
Sugestão - Usar servidor para proteger dados sensíveis
Sugestão - Validar histórias externas, prontas
Sugestão - Verificar se o usuário tem interesse nas recomendações feitas pela IA
Sugestões - Tratar informações sensíveis
Uso - Considera o uso valioso tanto na academia quanto na indústria
Uso - Considera que poderia usar a ferramenta como teste, na academia
Uso - Considera revisar posteriormente as histórias
Uso - Considera útil para entender sistemas reais
Uso - Considerou a ferramenta como um ponto de partida
Uso - Considerou como um ponto de entrada para analista de requisitos e product owner, na academia
Uso - Considerou que a ferramenta tem bom uso na academia
Uso - Disse que professores poderiam usar em exercícios
Uso - Disse que usaria para apoiar sua performance
Uso - Teve interesse em usar com alunos e treino
Uso - Usaria no futuro
Uso - Usaria no seu trabalho
Uso - Usaria novamente para gerar histórias
Uso - Usaria para apresentar e validar com o stakeholder rapidamente
Uso - Usaria para definir melhor as regras antes da especificação dos requisitos