



UNIVERSIDADE FEDERAL DO CEARÁ
CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE

ANDREZA HONÓRIO MAGALHÃES

ELICITAÇÃO APRIMORADA DE REQUISITOS DE SOFTWARE POR MEIO
DA INTEGRAÇÃO DO MÉTODO 5W2H COM IA GENERATIVA

RUSSAS

2026

ANDREZA HONÓRIO MAGALHÃES

ELICITAÇÃO APRIMORADA DE REQUISITOS DE SOFTWARE POR MEIO DA
INTEGRAÇÃO DO MÉTODO 5W2H COM IA GENERATIVA

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia de Software
da Universidade Federal do Ceará, como
requisito parcial à obtenção do grau de bacharel
em Engenharia de Software.

Orientadora: Profa. Ms. Valéria Maria
da Silva Pinheiro

RUSSAS

2026

ANDREZA HONÓRIO MAGALHÃES

ELICITAÇÃO APRIMORADA DE REQUISITOS DE SOFTWARE POR MEIO DA
INTEGRAÇÃO DO MÉTODO 5W2H COM IA GENERATIVA

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em Engenharia de Software
da Universidade Federal do Ceará, como
requisito parcial à obtenção do grau de bacharel
em Engenharia de Software.

Aprovada em: 08/07/2026

BANCA EXAMINADORA

Profa. Ms. Valéria Maria da Silva
Pinheiro (Orientadora)
Universidade Federal do Ceará - UFC

Profa. Dra. Anna Beatriz dos Santos Marques
Universidade Federal do Ceará - UFC

Profa. Dra. Patrícia Freitas Campos de Vasconcelos
Universidade Federal do Ceará - UFC

À minha família, por nunca limitar os meus sonhos, por acreditar em mim e por estar ao meu lado em cada etapa da minha caminhada.

AGRADECIMENTOS

Começo agradecendo a Deus, à Santíssima Trindade e à Nossa Senhora, Mãe Santíssima, por me concederem a graça de viver, não apenas, uma vida extraordinária, com inteligência, sabedoria e amor, mas vivências como essa.

Agradeço à minha família por me apoiar em todos os momentos e em cada decisão. Ao meu pai, Jailson Pedro Esteves Magalhães, agradeço por ser minha base, por sempre desejar e proporcionar o melhor para mim, oferecendo-me muito amor, cuidado e proteção. À minha mãe, Marília da Cruz Honório, agradeço por ser, para mim, a personificação do amor, sendo a melhor mãe que eu poderia ter. Ao meu irmão, Paulo Vinícius Honório Magalhães, agradeço por ser meu amigo, por me ouvir em cada dificuldade e em cada conquista e pelos conselhos que somente alguém que cresceu ao meu lado poderia me dar.

Agradeço à minha amiga, confidente e parceira de graduação, Ana Emilly Alves de Oliveira, um dos maiores presentes que a UFC me proporcionou. Ao longo desses anos de estudo, estive ao meu lado em todos os momentos, oferecendo apoio, companheirismo e uma amizade que levarei para a vida.

Agradeço também a todas as minhas amigas, mulheres que admiro e tenho muito amor, que sempre me apoiaram e torceram por mim, apesar de todas as mudanças em nossas vidas. Obrigada, Ananda Vitória, Beatriz, Camilla, Daniele, Gabriele, Iasmin Emillyn, Janara Maria, Lorena, Mariana, Maria Carolina, Nathália Fernandes, Nathália Livya, Nayara Lima, Nayara Soares e Vanessa, por estarem presentes em cada fase da minha vida e por compartilharem comigo tantos momentos ao longo dessa caminhada.

Agradeço a toda a minha rede de apoio, formada por familiares, amigos, colegas, por pessoas que acreditam em mim, confiam no meu potencial e sempre me incentivam a seguir em frente.

À Profa. Dra. Anna Beatriz dos Santos Marques, coordenadora do Labora-

tório de Pesquisa e Desenvolvimento em Usabilidade, Diversidade e Inclusão (LUDI), expresso minha profunda gratidão por confiar no meu trabalho, pelo suporte institucional e por me inspirar a ser uma profissional de excelência.

Agradeço aos meus colegas do LUDI, por tornarem esse final de curso mais agradável, descontraído e tranquilo ao compartilharem risadas e momentos de descontração na nossa sala, além de momentos únicos de confraternização e amizade nas datas especiais.

Agradeço à minha orientadora, Profa. Ms. Valéria Maria da Silva Pinheiro, por me guiar nesta etapa final da graduação, compartilhando seu conhecimento e suas experiências, e por se tornar uma amiga e uma inspiração ao longo desse percurso.

Aos professores da Universidade Federal do Ceará (UFC), Campus de Russas, expresso minha admiração e gratidão por conhecimento compartilhado ao longo dessa jornada. Seus ensinamentos, dedicação e incentivo foram fundamentais para a consolidação da minha formação acadêmica e para o meu desenvolvimento como profissional.

Por fim, mas não menos importante, agradeço à Universidade Federal do Ceará por me proporcionar não apenas uma formação acadêmica de excelência, mas também um crescimento social, político e cultural. Sou grata pelas vivências adquiridas ao longo dessa caminhada, que só foram possíveis graças às oportunidades proporcionadas por essa instituição.

Muito obrigada!

“Se você fizer só o que sabe, nunca será mais do
que já é.”

(Conceição Evaristo)

RESUMO

A Engenharia de Requisitos desempenha um papel fundamental no desenvolvimento de software, pois falhas na elicitação e na especificação de requisitos podem gerar retrabalho, aumento de custos e maior risco de insucesso do projeto. Métodos estruturados, como o 5W2H, auxiliam na compreensão do problema, mas ainda dependem significativamente da experiência humana. Nesse contexto, a IA generativa surge como uma alternativa promissora, contribuindo para a redução de ambiguidades e para a automatização parcial das atividades de elicitação e especificação de requisitos. Este trabalho propõe o desenvolvimento e a avaliação de uma metodologia que integra o 5W2H à IA generativa, com o objetivo de aprimorar a qualidade dos requisitos funcionais e não funcionais, considerando critérios de qualidade estabelecidos na literatura, além de comparar essa abordagem integrada com o processo tradicional de elicitação de requisitos. A pesquisa foi estruturada conforme o processo experimental descrito por Wohlin et al. (2012), contemplando as seguintes fases: (i) definição do escopo, (ii) planejamento, (iii) execução, (iv) análise e interpretação dos resultados e (v) apresentação e empacotamento. Com essa abordagem, foi possível produzir requisitos com maior nível de qualidade, atendendo aos critérios recomendados pela literatura, além de otimizar o tempo de elaboração e oferecer maior suporte cognitivo aos participantes. O estudo também discute as limitações e os riscos associados ao uso da IA generativa, como vieses e dependência excessiva, reforçando a importância da supervisão humana e contribuindo com uma metodologia empiricamente avaliada para a elicitação e especificação de requisitos de software.

Palavras-chave: Engenharia de Requisitos; elicitação de requisitos; especificação de requisitos; método 5W2H.; Inteligência Artificial generativa; qualidade de requisitos de software.

ABSTRACT

Requirements Engineering plays a fundamental role in software development, as deficiencies in requirements elicitation and specification can lead to rework, increased costs, and a higher risk of project failure. Structured methods, such as 5W2H, support problem understanding but still rely heavily on human expertise. In this context, Generative Artificial Intelligence (Generative AI) has emerged as a promising alternative, contributing to the reduction of ambiguities and the partial automation of requirements elicitation and specification activities. This study proposes the development and evaluation of a methodology that integrates the 5W2H method with Generative AI to improve the quality of both functional and non-functional requirements, considering quality criteria established in the literature. In addition, it compares the proposed integrated approach with the traditional requirements elicitation process. The research was structured according to the experimental process described by Wohlin et al. (2012), comprising the following phases: (i) scoping, (ii) planning, (iii) operation, (iv) analysis and interpretation of results, and (v) presentation and packaging. The proposed approach enabled the production of higher-quality requirements that satisfy the quality criteria recommended in the literature, while also reducing the time required for requirements development and providing greater cognitive support to participants. Furthermore, the study discusses the limitations and risks associated with the use of Generative AI, such as bias and overreliance, reinforcing the importance of human oversight and contributing an empirically evaluated methodology for software requirements elicitation and specification.

Keywords: Requirements Engineering; requirements elicitation; requirements specification; method 5W2H; generative Artificial Intelligence.; software requirements quality.

LISTA DE FIGURAS

Figura 1 – Ilustração do Método 5W2H	29
Figura 2 – Método de estudo experimental segundo Wohlin et al. (2012)	39
Figura 3 – Página inicial do documento	46
Figura 4 – Página de especificação dos requisitos	47
Figura 5 – Modelos de prompt	48
Figura 6 – Matriz do 5W2H	48
Figura 7 – Etapas da avaliação de cada artefato	56
Figura 8 – Etapas da preparação dos dados qualitativos	79
Figura 9 – Organização dos documentos no Taguette	80
Figura 10 – Resultado da análise qualitativa	82

LISTA DE TABELAS

Tabela 1 – Estatística Descritiva - Perfil do Participante - Parte 1	59
Tabela 2 – Estatística Descritiva - Perfil do Participante - Parte 2	61
Tabela 3 – Teste de normalidade - Perfil dos Participantes	61
Tabela 4 – Teste <i>t</i> de Student para amostras independentes - Perfil do Participante	62
Tabela 5 – Estatística Descritiva - Feedback dos Participantes - Grupos A e B .	63
Tabela 6 – Teste de normalidade - Feedback do Participante - Grupos A e B . .	64
Tabela 7 – Teste <i>t</i> de Student para amostras independentes – Feedback dos parti- cipantes - Grupo A e B	64
Tabela 8 – Estatística Descritiva - Feedback dos Participantes - Grupo B - Parte 1	65
Tabela 9 – Estatística Descritiva - Feedback dos Participantes - Grupo B - Parte 2	67
Tabela 10 – Estatística Descritiva - Feedback dos Participantes - Grupo B - Parte 3	67
Tabela 11 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 1	69
Tabela 12 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 2	70
Tabela 13 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 3	71
Tabela 14 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 4	72
Tabela 15 – Teste de normalidade - Qualidade Documental - Grupos A e B . . .	73
Tabela 16 – Teste U de Mann–Whitney – Qualidade documental - Grupos A e B	74
Tabela 17 – Tamanho de Efeito - Correlação Ranque-Bisserial	75
Tabela 18 – Estatística descritiva – Qualidade documental - Grupo B - Parte 1 . .	76
Tabela 19 – Estatística descritiva – Qualidade documental - Grupo B - Parte 2 . .	77
Tabela 20 – Estatística descritiva – Qualidade documental - Grupo B - Parte 3 . .	78

LISTA DE QUADROS

Quadro 1 – Principais diferenças entre os trabalhos relacionados e a pesquisa proposta	38
Quadro 2 – Definição do objetivo do estudo com base no paradigma GQM. . .	40
Quadro 3 – Hipóteses da pesquisa	41
Quadro 4 – Grupos do estudo	41
Quadro 5 – Variáveis dependentes do experimento	42
Quadro 6 – Critérios de inclusão e exclusão dos participantes	44
Quadro 7 – Instrumentos utilizados no experimento	45
Quadro 8 – Critérios utilizados na avaliação dos documentos de requisitos . . .	53
Quadro 9 – Conjunto de dados após a etapa de limpeza	55
Quadro 10 – Conversão da escala Likert em valores numéricos	57
Quadro 11 – Exemplo de aplicação do <i>reverse scoring</i>	58

SUMÁRIO

1	INTRODUÇÃO	18
2	OBJETIVO DO TRABALHO	21
2.0.1	<i>Objetivo geral</i>	<i>21</i>
2.0.2	<i>Objetivos específicos</i>	<i>21</i>
3	FUNDAMENTAÇÃO TEÓRICA	22
3.1	Requisitos de Software	22
3.2	Engenharia de Requisitos	24
3.2.1	<i>Processos da Engenharia de Requisitos</i>	<i>24</i>
3.2.1.1	<i>Elicitação de Requisitos</i>	<i>25</i>
3.2.2	<i>Técnicas de Elicitação</i>	<i>26</i>
3.2.3	<i>Especificação de Requisitos</i>	<i>27</i>
3.2.3.1	<i>Formas de especificação de requisitos</i>	<i>27</i>
3.3	O método 5W2H	28
3.3.1	<i>Potencial do 5W2H na especificação de requisitos</i>	<i>31</i>
3.3.2	<i>Limitações do método</i>	<i>31</i>
3.4	Inteligência Artificial Generativa	32
3.4.1	<i>Aplicações da IA generativa na Engenharia de Software</i>	<i>34</i>
4	TRABALHOS RELACIONADOS	35
5	METODOLOGIA	39
5.1	Definição do Escopo (Scoping)	39
5.2	Planejamento (Planning)	40
5.2.1	<i>Hipóteses</i>	<i>40</i>
5.2.2	<i>Tipo do Experimento</i>	<i>41</i>
5.2.3	<i>Variáveis Identificadas</i>	<i>41</i>
5.2.3.1	<i>Variáveis Independentes</i>	<i>42</i>

5.2.3.2	<i>Variáveis Dependentes</i>	42
5.2.3.3	<i>Variáveis de Controle</i>	43
5.2.4	<i>Participantes</i>	43
5.2.5	<i>Contexto Experimental</i>	44
5.2.6	<i>Instrumentação</i>	45
5.2.7	<i>Análise de Dados</i>	49
5.2.8	<i>Ameaças à Validade</i>	49
5.2.8.1	<i>Validade Interna</i>	49
5.2.8.2	<i>Validade Externa</i>	50
5.2.8.3	<i>Validade de Construção</i>	50
5.2.8.4	<i>Validade de Conclusão</i>	50
5.2.9	<i>Resultado Esperado</i>	50
5.3	<i>Operação (Operation)</i>	51
5.3.1	<i>Apresentação & Empacotamento (Presentation & Package)</i>	51
5.3.2	<i>Critérios de avaliação dos requisitos</i>	52
6	ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS	54
6.1	Dados Quantitativos	54
6.1.1	<i>Limpeza dos Dados Quantitativos</i>	54
6.1.1.1	<i>Etapa 1 - Cruzamento dos dados - Formulários de Perfil e Feedback</i> .	54
6.1.1.2	<i>Etapa 2 - Análise dos artefatos - Grupos A e B</i>	55
6.1.1.3	<i>Etapa 3 - Preparação dos dados</i>	56
6.1.2	<i>Resultados das análises quantitativas</i>	58
6.1.2.1	<i>Perfil dos participantes</i>	58
6.1.2.2	<i>Feedback dos Participantes</i>	62
6.1.2.2.1	<i>Feedback dos Participantes - Grupo B</i>	65
6.1.3	<i>Avaliação da Qualidade dos Artefatos</i>	68
6.1.3.1	<i>Avaliação dos artefatos - Grupos A e B</i>	68

6.1.3.2	<i>Avaliação dos artefatos - Grupo B</i>	75
6.2	Dados Qualitativos	79
6.2.1	<i>Preparação dos dados</i>	79
6.2.2	<i>Resultados da análise qualitativa</i>	81
6.3	Discussão dos resultados	85
6.3.1	<i>Ameaças à validade</i>	86
7	CONCLUSÕES E TRABALHOS FUTUROS	88
	REFERÊNCIAS	90
	APÊNDICES	95
	APÊNDICE A–MODELO GQM DO ESTUDO	95
	APÊNDICE B–FORMULÁRIO - PERFIL DO PARTICIPANTE .	103
	APÊNDICE C–FORMULÁRIO - FEEDBACK DO PARTICIPANTE	111
	APÊNDICE D–CENÁRIO DO EXPERIMENTO	132
	APÊNDICE E–FORMULÁRIO - AVALIAÇÃO DA QUALIDADE DOS DOCUMENTOS	137
	APÊNDICE F–PROMPT PARA GERAÇÃO DO CODEBOOK	161
	APÊNDICE G–LIVRO DE CÓDIGOS - ANÁLISE QUALITATIVA	162

1 INTRODUÇÃO

O desenvolvimento de software envolve um conjunto de etapas que variam conforme o modelo de processo adotado, como os modelos prescritivos, iterativos e incrementais, bem como as abordagens ágeis. Independentemente dessas variações, a especificação e a elicitação de requisitos constituem uma etapa comum e essencial a todos os processos de desenvolvimento, sendo determinantes para o sucesso do produto final (SOMMERVILLE, 2016; POHL, 2016). A elicitação de requisitos tem como objetivo compreender as necessidades dos stakeholders e traduzi-las em requisitos claros, completos e verificáveis, servindo de base para as demais atividades do ciclo de vida do software.

Diversos estudos apontam que falhas nessa etapa, bem como requisitos ambíguos, incompletos ou mal documentados, estão entre as principais causas de atrasos, retrabalho, aumento de custos e baixa qualidade em projetos de software (WIEGERS; BEATTY, 2013; KOTONYA; SOMMERVILLE, 1998). Nesse contexto, foi identificado que métodos estruturados podem servir de apoio à elicitação, como o 5W2H, que apresenta perguntas que podem auxiliar na compreensão do problema, do contexto e das necessidades do sistema de forma organizada (SULTAN; MIRANSKYY, 2015; ALFLEN; PRADO, 2021) e assim, acaba por contribuir na construção de um documento mais completo.

Paralelamente, o avanço recente das inteligências artificiais generativas, especialmente dos modelos de linguagem de grande porte (*Large Language Models* – LLMs), abriu novas possibilidades para apoiar atividades tradicionalmente conduzidas de forma manual na Engenharia de Requisitos. Esses modelos demonstram capacidade de interpretar descrições em linguagem natural, sugerir alternativas, reduzir ambiguidades e apoiar a estruturação de documentos de requisitos (ZHAO *et al.*, 2021; OZKAYA, 2023).

Apesar da existência de métodos e técnicas consolidados para a elicitação de

requisitos, a literatura evidencia limitações significativas quanto à garantia sistemática da qualidade dos requisitos conforme descrito no IEEE Std 830-1998, especialmente no que se refere a atributos como clareza, completude, correção, verificabilidade e rastreabilidade, sendo essas limitações ainda mais evidentes no tratamento de requisitos não funcionais (GLINZ, 2007; CHUNG *et al.*, 2012). Além disso, muitos métodos tradicionais dependem fortemente da experiência do analista e estão sujeitos a ambiguidades linguísticas inerentes à comunicação em linguagem natural (BERRY *et al.*, 2006).

Pesquisas recentes, como as de Zhao *et al.* (2021), Ozkaya (2023), têm explorado o uso de técnicas de processamento de linguagem natural e de modelos de linguagem em atividades da Engenharia de Software, incluindo a Engenharia de Requisitos. No entanto, esses estudos concentram-se majoritariamente em análises exploratórias e conceituais, carecendo de avaliações empíricas sistemáticas que investiguem o impacto dessas abordagens na qualidade da elicitação e documentação de requisitos. Observa-se, portanto, uma lacuna na literatura quanto à investigação do impacto da integração estruturada entre métodos clássicos de elicitação e IA generativa na qualidade da documentação de requisitos.

Diante dessas lacunas, integrar um método estruturado e amplamente difundido, como o 5W2H, com o uso de IA generativa representa uma oportunidade relevante para aprimorar a elicitação e a documentação de requisitos de software. Essa integração combina a orientação humana e sistemática do método com a capacidade da IA de processar grandes volumes de informação, identificar inconsistências, sugerir refinamentos e apoiar a padronização da documentação.

Este trabalho está organizado da seguinte forma: o Capítulo 2 apresenta o objetivo do trabalho e elenca os objetivos específicos atrelados a ele. O Capítulo 3 apresenta a fundamentação teórica, abordando conceitos de Engenharia de Requisitos, elicitação, qualidade de requisitos, o método 5W2H e o uso de IA generativa nesse contexto. No Capítulo 4 é apresentado os trabalhos relacionados a este estudo. O Capítulo 5 descreve

a metodologia adotada, incluindo o desenho experimental, os instrumentos de coleta e os critérios de avaliação. O Capítulo 6 apresenta e discute os resultados obtidos a partir do experimento conduzido. Por fim, o Capítulo 7 apresenta as conclusões do estudo, suas contribuições, limitações e sugestões para trabalhos futuros.

2 OBJETIVO DO TRABALHO

2.0.1 *Objetivo geral*

O objetivo geral deste trabalho é avaliar um método de elicitação e documentação de requisitos de software que integre o método 5W2H com IA generativa, visando melhorar a qualidade dos requisitos funcionais e não funcionais em termos de clareza, completude, corretude, verificabilidade, testabilidade e rastreabilidade.

2.0.2 *Objetivos específicos*

Como objetivos específicos, este trabalho propõe-se a:

- Propor um modelo de integração entre o método 5W2H e prompts de IA generativa para apoiar a elicitação de requisitos;
- Avaliar a qualidade dos requisitos gerados pela metodologia proposta, utilizando métricas de avaliação de requisitos e comparando-os com a elicitação tradicional;
- Analisar as vantagens, limitações e riscos da abordagem, considerando aspectos como viés da IA e a necessidade de supervisão humana;
- Elaborar um artefato de apoio que viabilize o uso prático da metodologia proposta em projetos de software.

3 FUNDAMENTAÇÃO TEÓRICA

A busca bibliográfica teve como objetivo identificar estudos relevantes relacionados à Engenharia de Requisitos integrada à estruturação do método 5W2H com o uso de IA generativa, a fim de subsidiar a fundamentação teórica e a análise dos resultados. Para isso, foi adotada uma estratégia sistemática de busca em bases de dados científicas reconhecidas na área de Engenharia de Software.

As buscas foram realizadas nas bases Scopus, Web of Science, IEEE Xplore, ACM Digital Library e Google Scholar, no Portal de Periódicos da CAPES e nos periódicos da Sociedade Brasileira de Computação (SBC), considerando publicações no período de 2015 a 2025, nos idiomas português e inglês.

Os termos de busca foram definidos a partir da combinação de palavras-chave derivadas do problema de pesquisa e dos objetivos do estudo, utilizando operadores booleanos (AND e OR) e variações terminológicas em português e inglês, com o intuito de ampliar a abrangência e a sensibilidade da recuperação dos resultados.

3.1 Requisitos de Software

É necessário compreender completamente o que são Requisitos de Software (RS) e sua importância no processo de desenvolvimento de uma aplicação. Segundo o "Glossário padrão de terminologia de engenharia de software do Institute of Electrical and Electronic Engineers"(IEEE. . . , 1990), os requisitos de software são definidos da seguinte forma:

"(i) Uma condição ou capacidade necessária a um usuário para resolver um problema ou atingir um objetivo. (ii) Uma condição ou capacidade que deve ser atendida ou possuída por um sistema ou componente de sistema para satisfazer um contrato, norma, especificação ou outros do-

cumentos formalmente impostos. (iii) Uma representação documentada de uma condição ou capacidade, conforme (i) ou (ii)."(IEEE. . . , 1990, p. 62).

Atrelada a essa definição, Sommerville (2019) estabelece que requisitos são uma descrição do que o sistema deve fazer e das restrições às quais sua operação estará submetida. A literatura classifica os requisitos em diferentes categorias, com o objetivo de facilitar sua organização, compreensão e validação. Entre as classificações mais aceitas estão os requisitos funcionais, não funcionais e de domínio (SOMMERVILLE, 2019; PRESSMAN; MAXIM, 2021).

Os Requisitos Funcionais (RF) descrevem os serviços, comportamentos ou funcionalidades que o sistema deve oferecer aos usuários. Eles especificam ações, entradas, saídas e interações esperadas, estando diretamente relacionados às funcionalidades percebidas pelo usuário final (SOMMERVILLE, 2019). Trabalhos acadêmicos brasileiros destacam que requisitos funcionais mal definidos são uma das principais causas de falhas em projetos de software (BARROS, 2023).

Os Requisitos Não Funcionais (RNF) referem-se às restrições e atributos de qualidade do sistema, como desempenho, segurança, usabilidade, confiabilidade e manutenibilidade. Embora não descrevam funcionalidades diretamente, esses requisitos exercem forte influência sobre a arquitetura e as decisões de projeto (PRESSMAN; MAXIM, 2021). Pesquisas publicadas em anais da SBC indicam que a negligência na especificação de requisitos não funcionais compromete significativamente a qualidade do produto final (FONSECA *et al.*, 2024).

Por fim, os Requisitos de Domínio (RD), também conhecidos como Regras de Negócio (RN), derivam das características específicas do contexto no qual o sistema será utilizado, como normas legais, regras organizacionais ou padrões técnicos do setor. Esses requisitos são frequentemente implícitos e difíceis de identificar, exigindo atenção especial durante a especificação para evitar inconsistências e omissões (SOMMERVILLE,

2019).

3.2 Engenharia de Requisitos

A Engenharia de Requisitos é compreendida como o conjunto de atividades, processos e artefatos que buscam identificar, documentar, validar e gerenciar as necessidades dos stakeholders em relação a um sistema de software (ISO/IEC/IEEE... , 2018; SOMMERVILLE, 2015; LAMSWEERDE, 2009; WIEGERS; BEATTY, 2013; NUSEIBEH; EASTERBROOK, 2000). Pesquisas recentes indicam que requisitos mal elicitados ou especificados ainda figuram entre as principais causas de falhas em projetos, resultando em retrabalho, custos elevados e baixa qualidade dos sistemas desenvolvidos (FONSECA *et al.*, 2024). Nesse sentido, a comunidade científica tem explorado abordagens que vão desde a elicitação de requisitos para sistemas complexos até a análise de requisitos em ambientes ágeis (CARVALHO *et al.*, 2021; HOY; XU, 2023).

A Engenharia de Requisitos (ER) continua sendo uma disciplina essencial e em constante evolução, enfrentando desafios na elicitação, especificação e validação de requisitos. Pesquisas recentes no Brasil destacam avanços no ensino e na prática, como propostas pedagógicas inovadoras (SANTANA *et al.*, 2023), análises sobre lacunas no ensino de requisitos não funcionais e de experiência do usuário, além de experimentos que investigam métodos de elicitação voltados à qualidade de software (FONSECA *et al.*, 2024).

3.2.1 Processos da Engenharia de Requisitos

Segundo Sommerville (2019), o processo de engenharia de requisitos segue um caminho espiral, tendo como pontos focais a elicitação de requisitos, a especificação, a validação e a documentação de requisitos, sendo esta a fase final do processo. Dentro de cada ponto focal, há um conjunto estruturado de atividades que buscam transformar as

necessidades dos stakeholders em requisitos de sistema claros, consistentes e rastreáveis ao longo do desenvolvimento de software. Essas atividades são iterativas e interdependentes, estando no cerne da qualidade dos produtos de software entregues (NECULA *et al.*, 2024). Como o foco deste estudo está na elicitação e especificação de requisitos, estas serão destacadas.

3.2.1.1 Elicitação de Requisitos

De acordo com Sommerville (2019, p. 96):

Os objetivos do processo de elicitação de requisitos são compreender o trabalho que os *stakeholders* realizam e entender como usariam um novo sistema para apoiar o trabalho deles.

Nesse contexto, o autor explica que a elicitação de requisitos é um processo composto por quatro atividades (SOMMERVILLE, 2019, p. 97):

Descoberta e compreensão dos requisitos: momento em que ocorre a interação com os *stakeholders*, com o fito de descobrir seus requisitos. Nessa atividade, podem ser aplicadas as técnicas anteriormente citadas;

Classificação e organização dos requisitos: essa atividade consiste em "pegar os dados brutos" descobertos anteriormente, classificá-los e agrupá-los em conjuntos de requisitos semelhantes.

Priorização e negociação dos requisitos: comumente há vários *stakeholders* envolvidos na descoberta dos requisitos, e chega um momento em que é necessário reuni-los novamente para negociar os requisitos que devem ser priorizados naquele momento. Tudo isso ocorre nesta atividade.

Documentação dos requisitos: os requisitos descobertos nessa etapa do processo de ER são documentados para servirem de base para as etapas seguintes, atuando como um rascunho inicial.

Essas atividades contribuem para a compreensão do domínio do problema e dos objetivos que o sistema deve alcançar, por meio de diversas técnicas, como entrevistas, workshops, prototipagem e observações (NECULA *et al.*, 2024). Estudos recentes ampliam as estratégias tradicionais ao integrar tecnologias inteligentes, como sistemas de recomendação para apoiar a identificação de requisitos e reduzir vieses humanos (AKRAM *et al.*, 2024). Em projetos reais, a elicitação continua sendo apontada como uma das etapas com maior potencial para falhas ou incompletudes quando negligenciada (AKHTAR; AKHTAR, 2025).

3.2.2 *Técnicas de Elicitação*

Em seu livro *Engenharia de Software* (10ª edição), Sommerville (2019) descreve diferentes técnicas de elicitação de requisitos, cada uma adequada a determinados contextos e objetivos. A escolha da técnica depende do contexto do projeto, do domínio da aplicação e das características dos participantes envolvidos. Em muitos projetos, recomenda-se a combinação de diferentes técnicas, de forma a obter uma visão mais completa do domínio do problema. A seguir, são apresentadas as técnicas citadas pelo autor:

Entrevistas: consistem na realização de reuniões entre o engenheiro de requisitos e os *stakeholders*, com o objetivo de compreender necessidades, expectativas, problemas e restrições do sistema. Podem ser estruturadas ou semiestruturadas, conforme o nível de planejamento das perguntas.

Etnografia: baseia-se na observação do ambiente de trabalho e das atividades desempenhadas pelos usuários, permitindo identificar requisitos implícitos e compreender processos que muitas vezes não são relatados durante entrevistas.

Histórias e cenários: utilizam descrições narrativas de situações reais ou hipotéticas de utilização do sistema para representar como os usuários executam suas tarefas. Essa técnica facilita a comunicação entre stakeholders e desenvolvedores, além de

contribuir para a identificação e validação dos requisitos.

3.2.3 *Especificação de Requisitos*

A especificação de requisitos pode ser definida como o processo de documentar, de maneira sistemática, os requisitos funcionais e não funcionais de um sistema, descrevendo o que o sistema deve fazer e sob quais restrições deve operar (SOMMERVILLE, 2019). Diferentemente da elicitação, que se concentra na descoberta das necessidades, a especificação visa formalizar essas informações em artefatos compreensíveis e utilizáveis pela equipe técnica.

Segundo Vazquez (2016), a especificação representa a transformação das necessidades dos stakeholders em requisitos documentados, reduzindo o caráter subjetivo das informações coletadas durante a elicitação. Esse processo exige interpretação, refinamento e organização do conhecimento, garantindo que os requisitos sejam claros, completos, consistentes e verificáveis.

Outro objetivo fundamental da especificação é atuar como meio de comunicação entre stakeholders e equipe técnica. De acordo com Pressman e Maxim (2021), o documento de requisitos estabelece uma referência comum, alinhando expectativas e reduzindo mal-entendidos ao longo do desenvolvimento. Estudos apresentados nos anais da SBC reforçam que uma especificação bem estruturada contribui para melhorar a comunicação interdisciplinar, especialmente em equipes distribuídas ou projetos de grande porte (FONSECA *et al.*, 2024).

3.2.3.1 *Formas de especificação de requisitos*

A especificação de requisitos pode ser realizada por meio de diferentes formatos e técnicas, variando conforme o contexto do projeto, o perfil dos stakeholders e o nível de formalidade desejado. Entre as formas mais utilizadas estão a linguagem

natural estruturada, casos de uso, histórias de usuário, modelos e diagramas e os documentos de especificação de requisitos de software (SRS) (VAZQUEZ; SIMÕES, 2016; SOMMERVILLE, 2019).

Linguagem Natural Utiliza textos estruturados e templates predefinidos para descrever requisitos, buscando reduzir ambiguidades e facilitar a compreensão. Entretanto, sua eficácia depende da padronização adotada (PRESSMAN; MAXIM, 2021).

Casos de Uso Descrevem as interações entre usuários e sistema, enfatizando o comportamento funcional em diferentes cenários e facilitando a comunicação com os stakeholders (FONSECA *et al.*, 2024).

Histórias de Usuário Representam requisitos de forma concisa sob a perspectiva do usuário final. São amplamente utilizadas em métodos ágeis e devem ser complementadas por critérios de aceitação para garantir maior qualidade na especificação (COSTA *et al.*, 2024).

Modelos e Diagramas Utilizam representações visuais, como diagramas UML, para complementar a especificação textual, auxiliando na compreensão do sistema e na identificação de inconsistências (SOMMERVILLE, 2019).

Documento SRS Consolida os requisitos em um documento formal, geralmente baseado nos padrões IEEE 830 ou IEEE 29148, servindo como referência para o desenvolvimento e a validação do sistema (ISO/IEC/IEEE. . . , 2018).

Diante do exposto, observa-se que a elicitação e a especificação de requisitos desempenham um papel fundamental na Engenharia de Requisitos. A qualidade dessas atividades influencia diretamente o sucesso do desenvolvimento de software, uma vez que requisitos com baixa qualidade podem comprometer as etapas subsequentes do projeto.

3.3 O método 5W2H

O método 5W2H tem origem nas práticas de gestão e controle da qualidade, sendo amplamente difundido nos contextos da administração e da engenharia de produção

como uma ferramenta simples, versátil e eficaz para o planejamento e a organização de ações (OLIVEIRA, 2015).

O 5W2H é composto por sete perguntas fundamentais que orientam a construção de um plano de ação de forma lógica e estruturada (Figura 1). Sua aplicação pode ocorrer de maneira independente ou em conjunto com outras técnicas de gestão, como a análise SWOT, contribuindo para a formulação de estratégias, a definição de responsabilidades e o planejamento de atividades de forma sistemática, objetiva e organizada (JUNIOR, 2018). .

Figura 1 – Ilustração do Método 5W2H



Fonte: Adaptado de Publi (2021).

O raciocínio é formulado ao responder as questões, conforme apresentado a seguir (JUNIOR, 2018):

What (o que?): o que será feito?

Why (por quê?): por que será feito:

Who (quem?): por quem será feito/Quem vai fazer?

Where (onde?): onde será feito?

When (quando?): quando será feito?

How (como?): como será feito/Com que recursos/Qual metodologia?

How much (quanto?): quanto vai custar?

Em razão de sua facilidade de aplicação e capacidade de estruturar informações, o método passou a ser incorporado também em áreas correlatas, como a Engenharia de Software, especialmente em atividades que demandam a coleta, a organização e o detalhamento de informações. Na literatura recente, observa-se a aplicação do 5W2H como apoio à documentação de requisitos e à definição de escopo de projetos de software, especialmente em ambientes acadêmicos e organizacionais que buscam métodos de fácil compreensão e rápida aplicação (DevMedia, 2025; BARROS, 2023).

Ao adaptar as questões do método para este estudo, de forma que pudessem atender aos critérios de qualidade previstos na ISO/IEC/IEEE... (2018), foi pensado na seguinte estrutura para construção de pensamento:

What (o que?): descreve o que deve ser feito ou o que o requisito representa;

Why (por quê?): justifica a necessidade do requisito, relacionando-o a objetivos de negócio ou necessidades do usuário;

Who (quem?): identifica os responsáveis ou usuários envolvidos;

Where (onde?): indica o contexto ou ambiente de aplicação;

When (quando?): define o momento ou condição temporal associada ao requisito;

How (como?): descreve como o requisito será atendido ou executado;

How much (quanto?): relaciona custos, esforços ou restrições quantitativas associadas.

Segundo Oliveira (2016), a principal contribuição do 5W2H está no foco na organização e no detalhamento da informação, permitindo que diferentes dimensões de uma mesma necessidade sejam exploradas de forma sistemática. Quando aplicado à especificação de requisitos, esse conjunto de perguntas auxilia o analista a estruturar descrições mais completas e coerentes, evitando registros excessivamente genéricos ou incompletos (VAZQUEZ; SIMÕES, 2016).

3.3.1 Potencial do 5W2H na especificação de requisitos

A aplicação do método 5W2H na especificação de requisitos apresenta potencial significativo para apoio à clareza e à completude das descrições. Ao orientar o analista a responder sistematicamente às sete questões, o método reduz a probabilidade de omissões e incentiva uma visão mais abrangente do requisito (PRESSMAN; MAXIM, 2021).

Outro benefício relevante é a redução de ambiguidades, um dos principais problemas associados à especificação em linguagem natural. Segundo Sommerville et al. (2019), a estruturação prévia das informações, mesmo quando utilizadas descrições textuais em linguagem natural, contribui para diminuir interpretações múltiplas e inconsistências semânticas (SOMMERVILLE, 2019).

Além disso, o 5W2H favorece a padronização da descrição dos requisitos, especialmente em contextos nos quais não há adoção formal de padrões como o SRS do IEEE. O uso de métodos estruturantes simples pode melhorar a uniformidade da documentação e facilitar a comunicação entre os membros da equipe, conforme identificado no estudo de Costa et al. (2024) .

Embora o trabalho de Barros (2023) não utilize explicitamente o 5W2H, seus resultados indicam que a utilização de instrumentos estruturados entre a elicitação e a especificação formal apoia a consolidação do conhecimento levantado antes da elaboração dos documentos definitivos de requisitos.

3.3.2 Limitações do método

O método 5W2H apresenta limitações quando utilizado isoladamente na especificação de requisitos. Uma das principais refere-se à possibilidade de gerar respostas superficiais ou genéricas, especialmente quando as questões não são respondidas com o devido aprofundamento. Além disso, a qualidade das informações produzidas depende

diretamente da experiência do analista, que deve interpretar corretamente as necessidades dos stakeholders e explorar todas as dimensões relevantes dos requisitos (OLIVEIRA, 2015; VAZQUEZ; SIMÕES, 2016).

Outra limitação é que o 5W2H não substitui técnicas formais de especificação, sendo necessário complementá-lo com atividades de refinamento, validação e outras abordagens de modelagem. Dessa forma, seu uso isolado pode ser insuficiente para capturar requisitos complexos, especialmente os não funcionais, reforçando a necessidade de integração com métodos e ferramentas que ampliem a qualidade e a completude da especificação (SOMMERVILLE, 2019; PRESSMAN; MAXIM, 2021).

Dessa forma, para superar essas limitações, torna-se relevante integrar o método 5W2H a abordagens que ampliem e qualifiquem as informações produzidas, fortalecendo a qualidade da especificação de requisitos.

3.4 Inteligência Artificial Generativa

A Inteligência Artificial (IA) pode ser definida como um campo da Ciência da Computação voltado ao desenvolvimento de sistemas capazes de realizar atividades que tradicionalmente requerem capacidades cognitivas humanas, como raciocínio, aprendizado e tomada de decisão (RUSSELL; NORVIG, 2022).

Em específico, a Inteligência Artificial generativa representa uma evolução significativa desse paradigma, ao empregar técnicas de aprendizado profundo para identificar padrões complexos em grandes volumes de dados e, a partir disso, gerar novos conteúdos de maneira probabilística. Esses conteúdos podem assumir diferentes formas, como identificado por Goodfellow et al. (2016) .

Nesse contexto, destacam-se os Modelos de Linguagem de Grande Porte (Large Language Models – LLMs), projetados para compreender e produzir linguagem natural com elevado grau de fluidez e coerência. Esses modelos demonstram capacidade de adaptação a diferentes contextos e tarefas, mesmo quando fornecidos com poucos

exemplos, característica que impulsionou sua adoção em diversas áreas do conhecimento (BROWN *et al.*, 2020).

Estudos recentes indicam que os LLMs diferenciam-se das abordagens tradicionais por sua habilidade de considerar o contexto global da interação, permitindo respostas variadas, refinadas e semanticamente adequadas, o que amplia significativamente seu potencial de aplicação em atividades que dependem intensamente da linguagem natural, como a Engenharia de Requisitos (ZADENOORI *et al.*, 2025).

Os Modelos de Linguagem de Grande Escala (*Large Language Models* – LLMs) são desenvolvidos a partir de técnicas de aprendizado profundo e podem ser compreendidos em três etapas principais:

Treinamento Os modelos são treinados com grandes volumes de dados textuais utilizando arquiteturas baseadas em redes neurais do tipo *transformer*. Essa arquitetura, proposta por Ashish *et al.* (2017), emprega mecanismos de atenção capazes de capturar relações entre palavras e sentenças, mesmo em contextos extensos.

Aprendizado Durante o treinamento, milhões ou bilhões de parâmetros são ajustados para aprender relações estatísticas entre unidades linguísticas e prever o próximo token de uma sequência textual. Esse processo permite que o modelo aprenda padrões sintáticos, semânticos e pragmáticos da linguagem natural (BROWN *et al.*, 2020).

Geração de Respostas Após o treinamento, o modelo é capaz de produzir respostas contextualizadas a partir de um texto de entrada, denominado *prompt*. A qualidade desse *prompt* influencia diretamente a precisão, relevância e utilidade das respostas geradas, possibilitando aplicações como sumarização, reformulação textual, expansão de descrições e geração de alternativas descritivas (OpenAI, 2023).

No âmbito da Engenharia de Software, essa competência de manipulação linguística torna os modelos generativos especialmente relevantes, uma vez que atividades como elicitación, especificação e documentação de requisitos dependem fortemente

da linguagem natural para a comunicação entre analistas e stakeholders (RUSSELL; NORVIG, 2022).

3.4.1 Aplicações da IA generativa na Engenharia de Software

Há diversas aplicações da IA generativa no apoio às atividades da Engenharia de Software, com ênfase nas fases iniciais do desenvolvimento, onde tais modelos podem auxiliar profissionais na compreensão de problemas, na organização de informações e na proposição de alternativas durante a definição de requisitos (EBERT; LOURIDAS, 2023).

No contexto específico da Engenharia de Requisitos, revisões sistemáticas evidenciam o potencial dos LLMs para apoiar atividades como elicitação, análise e especificação, por meio da geração de perguntas orientadoras, reformulação de requisitos e identificação de inconsistências ou lacunas em descrições textuais (ZADENOORI *et al.*, 2025). Essas funcionalidades contribuem para o aumento da clareza e da completude dos requisitos documentados.

Além disso, técnicas baseadas em Processamento de Linguagem Natural (PLN) e aprendizado de máquina têm sido empregadas para apoiar a formalização de requisitos e a redução de ambiguidades, aspectos críticos para a qualidade dos artefatos produzidos (KOLAHDOUZ-RAHIMI *et al.*, 2023). Trabalhos recentes também destacam a aplicação desses modelos na organização de informações dispersas e na geração de documentação técnica estruturada (PACHECO *et al.*, 2025).

No cenário nacional, pesquisas como a de Pilone *et al.* (2024) indicam ganhos de produtividade e suporte cognitivo ao analista de requisitos quando LLMs são utilizados como ferramentas complementares às práticas tradicionais, especialmente em ambientes colaborativos e multilíngues.

4 TRABALHOS RELACIONADOS

Esta seção apresenta uma revisão da literatura sobre os três pilares que fundamentam esta pesquisa: Engenharia de Requisitos, método 5W2H e LLMs, discutindo os principais conceitos, abordagens e contribuições de estudos relevantes na área.

No contexto da Engenharia de Requisitos aliada à Inteligência Artificial, Dalpiaz e Niu (2020) discutem como o avanço da IA tem transformado práticas tradicionais da área, apontando oportunidades para automação, apoio à tomada de decisão e melhoria da qualidade dos artefatos de requisitos. Os autores destacam que técnicas baseadas em IA podem auxiliar na análise, validação e gerenciamento de requisitos, especialmente em cenários complexos e dinâmicos. Essa visão é complementada por Vogelsang (2024), que aborda a integração de modelos de LLMs ao processo de Engenharia de Requisitos, enfatizando o papel do prompting como elemento central para a obtenção de resultados relevantes e alinhados ao contexto do sistema.

Diversos trabalhos recentes aprofundam o uso de LLMs na elicitación e análise de requisitos. Ronanki et al. (2023) investigam o potencial do ChatGPT como ferramenta de apoio à elicitación, evidenciando ganhos em produtividade e suporte à análise inicial, embora ressaltem limitações relacionadas à compreensão contextual e à dependência da qualidade dos prompts. De forma semelhante, Fotrousi e Tavantzis (2024) apresentam o ReqGenie, uma abordagem conversacional baseada em GPT para elicitación de requisitos, demonstrando que a interação humano-IA pode enriquecer o processo, mas ainda carece de mecanismos estruturados que garantam consistência e completude dos requisitos gerados.

No âmbito da qualidade de requisitos, Misaki et al. (2024) exploram o uso de IA generativa para identificação de *requirements smells*, evidenciando o potencial dessas tecnologias para apoiar a melhoria da qualidade textual e semântica dos requisitos. Embora essa abordagem contribua significativamente para a detecção de problemas

comuns, como ambiguidades e inconsistências, o foco permanece na análise posterior dos requisitos, e não em sua estruturação desde a fase de especificação. Almeida et al. (2025) , por sua vez, avaliam o desempenho de LLMs na transformação de entrevistas de elicitação em requisitos de software, apontando resultados promissores, mas também destacando a necessidade de estratégias mais sistemáticas para orientar a geração dos requisitos.

Outros estudos demonstram contribuições indiretas relevantes ao discutir aspectos de engenharia de prompts e análise de conhecimento em contextos de desenvolvimento de software. Ronanki et al. (2024) apresentam padrões de prompts aplicados à Engenharia de Requisitos, reforçando a importância de estruturas bem definidas para guiar a interação com modelos generativos. Ramalho et al. (2025) analisam o uso do GPT para identificar questões recorrentes em discussões de desenvolvedores, evidenciando a capacidade dos LLMs de sintetizar e estruturar informações complexas, ainda que fora do escopo específico da especificação de requisitos. Trabalhos como o de Tang et al. (2024) , embora situados em outro domínio de aplicação, reforçam o avanço das técnicas de aprendizado profundo e sua capacidade de compreender cenários complexos, contribuindo indiretamente para o amadurecimento das soluções baseadas em IA utilizadas na Engenharia de Software.

Apesar das inúmeras contribuições relacionadas à Engenharia de Requisitos e à IA generativa, observa-se uma escassez significativa de trabalhos que integrem explicitamente os três pilares: Engenharia de Requisitos, método 5W2H e LLMs. A literatura analisada não apresenta estudos que utilizem o 5W2H como um framework estruturante para guiar a geração ou a especificação de requisitos com apoio de modelos de linguagem. Quando o método 5W2H é mencionado em outros contextos da Engenharia de Software ou da gestão, ele aparece de forma isolada, sem articulação direta com técnicas de IA ou com processos formais de Engenharia de Requisitos.

Essa limitação é ainda mais evidente ao se analisar a relação entre o método

5W2H e os demais pilares. Não foram identificados trabalhos que explorem o 5W2H como um mecanismo sistemático para orientar prompts, estruturar entradas para LLMs ou organizar a saída gerada em especificações de requisitos mais completas, claras e verificáveis. Dessa forma, o método permanece subexplorado no contexto de soluções baseadas em IA para Engenharia de Requisitos, apesar de seu potencial reconhecido para organização e clareza da informação.

Diante desse cenário, as contribuições desta pesquisa se distinguem em dois níveis. As contribuições diretas consistem na proposição e avaliação de uma abordagem que integra explicitamente o método 5W2H com IA generativa para a especificação aprimorada de requisitos de software, buscando maior completude, rastreabilidade e redução de ambiguidades desde as fases iniciais. Já as contribuições indiretas incluem o aproveitamento de fundamentos teóricos e evidências empíricas de estudos anteriores sobre LLMs, engenharia de prompts e qualidade de requisitos, que subsidiam a construção da abordagem proposta, ao mesmo tempo em que evidenciam as lacunas ainda não exploradas pela literatura. Assim, este trabalho se posiciona como uma extensão e aprofundamento das pesquisas existentes, oferecendo uma nova perspectiva ao integrar métodos clássicos de estruturação com tecnologias emergentes de Inteligência Artificial.

Um dos principais desafios identificados durante a revisão bibliográfica foi a dificuldade em encontrar trabalhos que integrassem, de forma conjunta e sistemática, esses três pilares, o que evidencia uma lacuna científica e reforça o potencial inovador desta pesquisa. Embora existam estudos consolidados que exploram a aplicação de Inteligência Artificial na Engenharia de Requisitos, como os de Dalpiaz e Niu (2020) e Vogelsang (2024) , e investigações empíricas que analisam o uso de modelos de linguagem para apoio à elicitación e análise de requisitos, como Almeida et al. (2025) e Ronanki et al. (2023) , observa-se que tais pesquisas não contemplam explicitamente a utilização do método 5W2H como estratégia estruturante no processo de especificação de requisitos. Enquanto a maioria dos estudos concentra-se na automação da elicitación,

na redução de ambiguidades ou na melhoria da qualidade textual dos requisitos, esta pesquisa propõe investigar a especificação aprimorada de requisitos de software por meio da integração do método 5W2H com IA generativa, oferecendo uma nova perspectiva sobre a estruturação, clareza e completude dos requisitos.

Um resumo das contribuições destes trabalhos para a comunidade científica e um comparativo com relação ao estudo proposto neste trabalho, podem ser visualizadas no quadro a seguir (Quadro 1):

Quadro 1 – Principais diferenças entre os trabalhos relacionados e a pesquisa proposta

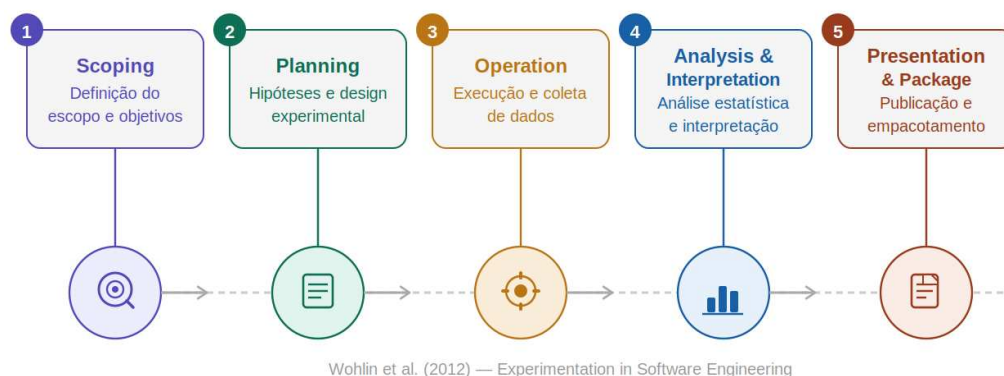
Trabalho	Principal contribuição	Diferença em relação ao presente trabalho
Alicia et al. (2024)	Investiga o uso de LLMs para apoiar atividades da Engenharia de Requisitos.	Não propõe um processo estruturado baseado em 5W2H nem realiza uma avaliação experimental da qualidade dos requisitos produzidos.
Maalej et al. (2023)	Apresenta aplicações de IA Generativa na Engenharia de Requisitos.	Discute potencialidades e desafios, porém não propõe uma metodologia integrada nem valida seus efeitos sobre atributos de qualidade dos requisitos.
Ferrari et al. (2024)	Avalia o uso de IA em atividades específicas de requisitos.	A avaliação concentra-se no desempenho da IA, sem integrar uma técnica sistemática de elicitação ou analisar qualitativamente a percepção dos participantes.
Este Trabalho	Integra o método 5W2H com LLMs por meio de prompts estruturados e valida a abordagem experimentalmente.	Avalia quantitativamente critérios de qualidade (clareza, completude, corretude, verificabilidade, testabilidade, rastreabilidade, organização e padronização), complementando os resultados com análise qualitativa das percepções dos participantes.

5 METODOLOGIA

Este estudo possui caráter aplicado e experimental, com abordagem mista, utilizando métodos quantitativos e qualitativos, pois propõe o desenvolvimento de uma metodologia para a elicitac  o e especifica  o de requisitos de software e avalia empiricamente sua efetividade por meio de um experimento controlado. A pesquisa tamb  m apresenta car  ter explorat  rio e descritivo, por investigar uma abordagem ainda pouco explorada na literatura: a integra  o de um m  todo estruturado de planejamento com IA generativa, a fim de compreender seus impactos na qualidade dos requisitos produzidos.

A metodologia foi desenvolvida seguindo as etapas propostas por Wohlin *et al.* (2012) em seu livro *Experimentation in Software Engineering*, conforme ilustrado na Figura 2.

Figura 2 – M  todo de estudo experimental segundo Wohlin et al. (2012)



Fonte: Elaborado a partir de dados fornecidos para Claude AI (2026).

5.1 Defini  o do Escopo (*Scoping*)

Essa primeira fase corresponde    identifica  o e    elabora  o estruturada do objetivo do estudo, na qual    utilizado o modelo GQM (*Goal-Question-Metric*), para estruturar o objetivo em termos de objeto de estudo, prop  sito, foco de qualidade,

perspectiva e contexto. Dessa forma, o **objetivo** é apresentado conforme descrito na Quadro 2. A partir do objetivo estruturado, foram elaboradas questões que auxiliaram na construção dos artefatos de coleta de dados. Com base nessas questões, foram definidas as métricas utilizadas para avaliar os dados coletados durante a execução do experimento. A construção desse conjunto foi realizada de forma que cada questão pudesse ser avaliada por uma métrica.

A visualização completa do modelo GQM elaborado para este estudo, pode ser verificado no Apêndice A.

Quadro 2 – Definição do objetivo do estudo com base no paradigma GQM.

Analisar	A metodologia proposta, que integra o método 5W2H com IA generativa
Com o propósito de	Avaliar em que medida a metodologia melhora a qualidade da elicitação e documentação de requisitos de software
Em relação a	Requisitos funcionais e não funcionais, considerando clareza, completude, correção, verificabilidade, testabilidade e rastreabilidade
Do ponto de vista da	Pesquisadora
No contexto de	Um ambiente acadêmico, em comparação com a elicitação tradicional

Fonte: Elaborado pela autora (2026).

5.2 Planejamento (Planning)

Nessa fase, foi planejado de forma detalhada como seria o experimento deste estudo, incluindo: i) formulação das hipóteses; ii) identificação e seleção das variáveis; iii) escolha do desenho experimental; e iv) definição dos critérios de validade.

5.2.1 Hipóteses

Para esta pesquisa, foram definidas as seguintes hipóteses (Quadro 3):

Quadro 3 – Hipóteses da pesquisa

Hipótese	Descrição
H₀	Não há diferença estatisticamente significativa entre a abordagem tradicional e a metodologia 5W2H + IA na qualidade da elicitação e documentação de requisitos.
H₁	A metodologia 5W2H + IA melhora significativamente a qualidade da elicitação e documentação de requisitos em comparação com a abordagem tradicional.

Fonte: Elaborado pela autora (2026).

5.2.2 Tipo do Experimento

Este estudo possui caráter experimental, sendo conduzido por meio de um experimento controlado realizado com uma turma do curso de Engenharia de Software da Universidade Federal do Ceará, Campus Russas. A pesquisa apresenta natureza quantitativa, com apoio qualitativo, a fim de comparar empiricamente a elicitação tradicional com a integração do método 5W2H e de ferramentas de inteligência artificial generativa. Os participantes foram divididos em dois grupos distintos, conforme apresentado no Quadro 4:

Quadro 4 – Grupos do estudo

Grupo	Tratamento
Grupo A: grupo de controle	Elicitação tradicional
Grupo B: grupo experimental	Método 5W2H + IA

Fonte: Elaborado pela autora (2026).

5.2.3 Variáveis Identificadas

A definição das variáveis experimentais é uma das etapas do planejamento. Nela, são estabelecidos os fatores que são manipulados, observados e mantidos constantes ao longo da pesquisa. Neste estudo, as variáveis foram classificadas em independentes, dependentes e de controle, conforme descrito a seguir:

5.2.3.1 Variáveis Independentes

As variáveis independentes correspondem aos fatores manipulados durante o experimento, sendo responsáveis por diferenciar os tratamentos aplicados aos participantes. Neste estudo, foram consideradas as seguintes variáveis:

- utilização de Inteligência Artificial generativa;
- aplicação do método 5W2H.

5.2.3.2 Variáveis Dependentes

Correspondem aos resultados observados em função da manipulação das variáveis independentes. Neste experimento, as variáveis dependentes estão diretamente relacionadas às métricas definidas no modelo GQM, sendo agrupadas conforme apresentado no Quadro 5.

Quadro 5 – Variáveis dependentes do experimento

Aspecto observado	Variáveis avaliadas
Qualidade dos requisitos	– Clareza – Completude – Verificabilidade – Corretude – Rastreabilidade – Testabilidade
Qualidade dos artefatos	– Organização – Padronização – Completude dos artefatos – Ausência de ambiguidades
Experiência do participante	– Usabilidade – Confiança – Suporte cognitivo

Fonte: Elaborado pela autora.

5.2.3.3 *Variáveis de Controle*

As variáveis de controle correspondem aos fatores mantidos constantes durante toda a execução do experimento, com o objetivo de minimizar ameaças à validade interna e garantir que os efeitos observados sejam decorrentes exclusivamente da manipulação das variáveis independentes. Para isso, foram controlados os seguintes aspectos:

- mesmo cenário de elicitación de requisitos;
- mesmo tempo disponível para a realização das atividades;
- mesmo ambiente de execução;
- mesmas instruções fornecidas aos participantes;
- mesma ferramenta de Inteligência Artificial generativa;
- mesmos critérios de avaliação dos artefatos produzidos;
- mesmo número de participantes.

5.2.4 *Participantes*

Os participantes foram convidados para o experimento pela orientadora deste trabalho e eram, à época, graduandos do segundo semestre do curso de Engenharia de Software, matriculados na disciplina de Introdução a Processos e Requisitos de Software (IPRS). A participação ocorreu de forma voluntária, mediante a anuência dos estudantes por meio do Termo de Consentimento Livre e Esclarecido (TCLE). O termo, bem como o detalhamento dos procedimentos da pesquisa, foi disponibilizado nos formulários de Perfil e Feedback (Apêndices B e C).

A escolha desse público teve como objetivo reunir participantes com um nível de conhecimento semelhante em Engenharia de Requisitos, uma vez que estavam cursando a disciplina responsável por introduzir os conceitos de elicitación e especificação de requisitos. Dessa forma, buscou-se reduzir a influência de diferenças relacionadas à

formação acadêmica e ao conhecimento prévio sobre o tema, contribuindo para maior homogeneidade da amostra.

Ao todo, participaram dezoito estudantes, distribuídos entre os dois grupos definidos para o estudo: i) Grupo A (Controle), composto por oito participantes; e ii) Grupo B (Experimental), composto por dez participantes.

Para o estudo, foram definidos critérios de inclusão e exclusão, que validam a participação e podem influenciar a validade do experimento (Quadro 6).

Quadro 6 – Critérios de inclusão e exclusão dos participantes

Categoria	Critérios
Critérios de Inclusão	– possuir conhecimento básico em requisitos; – concordar com o termo de consentimento; – participar integralmente da atividade.
Critérios de Exclusão	– não concluir a atividade; – possuir experiência avançada prévia com a metodologia proposta.

Fonte: Elaborado pela autora (2026).

5.2.5 *Contexto Experimental*

O experimento foi realizado no dia 08 de junho de 2026, das 13h30 às 15h30, com a participação de dezoito estudantes da disciplina de IPRS . O estudo foi conduzido em ambiente acadêmico controlado, durante uma sessão supervisionada no laboratório de informática do campus.

Os participantes receberam um cenário fictício contendo um problema de software e foram convidados a elicitar e documentar os requisitos de acordo com o tratamento atribuído ao grupo ao qual pertenciam.

5.2.6 Instrumentação

Os seguintes instrumentos utilizados no experimento são apresentados no Quadro 7:

Quadro 7 – Instrumentos utilizados no experimento

Instrumento	Finalidade
Cenário do sistema	Contextualização do problema, dos atores e das regras de negócio (Apêndice D).
Documento de Requisitos - Grupo A	Apresenta instruções de uso e estrutura para elicitação de requisitos (Figura 3 e Figura 4).
Documento de Requisitos - Grupo B	Apresenta instruções de uso, estrutura para elicitação de requisitos e sugestões de <i>prompts</i> para utilização da IA generativa (Figura 3, Figura 4 e Figura 5).
IA generativa (ChatGPT ou DeepSeek)	Ferramenta utilizada como apoio à elicitação e especificação de requisitos pelos participantes do Grupo B.
Matriz 5W2H	Material de apoio elaborado para orientar a elicitação de requisitos do Grupo B (Figura 6).
Formulário de perfil do participante	Instrumento utilizado para coleta de informações de caracterização dos participantes (Apêndice B).
Formulário de feedback do experimento	Instrumento utilizado para coletar a percepção dos participantes sobre o experimento e a metodologia aplicada (Apêndice C).

Fonte: Elaborado pela autora (2026).

Documento de Requisitos: todos os participantes receberam uma planilha com características similares. Na página inicial havia instruções de como proceder com o experimento, caso o participante tivesse alguma dúvida (Figura 3).

Figura 3 – Página inicial do documento

Instruções do Experimento

Faça uma cópia dessa planilha antes de iniciar, selecionando a opção "Compartilhar com as mesmas pessoas"!

Dado o cenário proposto, identifique e escreva 5 Requisitos Funcionais e 5 Requisitos Não-Funcionais do sistema StudyConnect.

Preenchimento da Planilha

Matrícula: Insira a sua matrícula. Não digite o seu nome!

Horário de Início: Quando **iniciar** o preenchimento dessa planilha, selecione o botão " " ou em "...", que aparecem ao lado do menu "Ajuda", para registrar o momento do início do seu experiemnto

Horário de Término: Quando **finalizar** o preenchimento dessa planilha, selecione o botão " " ou em "...", que aparecem ao lado do menu "Ajuda", para registrar o momento em que finalizou a elicitacão dos requisitos.

Grupo: identifique a qual grupo você pertence: Grupo A, grupo de controle, sem o uso do método e da IA; Grupo B, grupo experimental, que usará o método e a IA durante a elicitacão.

Rastreabilidade: Identifique a qual RF ou RNF o requisito que você escreveu faz referência.

Fonte: Elaborado pela autora (2026).

Além disso, havia uma segunda página com uma formatação padronizada para a especificação de cinco RF e RNF, além dos critérios de aceitação (Figura 4). O tempo foi coletado de forma automatizada por meio de uma extensão do Google Sheets que permite a automação das atividades, a partir do código escrito. Nessa página, a única diferença entre os grupos era a presença da opção de coletar o link do prompt.

Figura 4 – Página de especificação dos requisitos

[illegible]

Fonte: Elaborado pela autora (2026).

Para o grupo experimental, foi definido o DeepSeek como ferramenta de apoio à execução da atividade e ao preenchimento do documento de requisitos, devido à possibilidade de compartilhamento do link da conversa e à utilização gratuita, sem limitações significativas de uso. Entretanto, por dificuldades relatadas pelos participantes, foi liberado, antes do início do experimento, o uso do ChatGPT como ferramenta alternativa.

No documento de requisitos destinado ao Grupo B, havia ainda uma terceira página

contendo três modelos de prompt (Figura 5), disponibilizados como sugestão e previamente testados pela autora deste trabalho. Os prompts foram elaborados de modo que a IA pudesse interpretar as informações fornecidas e gerar respostas compatíveis com a estrutura do método 5W2H, auxiliando os participantes na elicitación e especificação dos requisitos.

Figura 5 – Modelos de prompt

Modelo 1

Dado o [ESPECÍFICO] enviado, quero que elicie [ESPECÍFICO], tendo como base o modelo apresentado no pdf "modelo 5W2H", que se estrutura a partir do método 5w2h.

Modelo 2

Dado o [ESPECÍFICO] enviado, quero que elicie [ESPECÍFICO], tendo como base o modelo apresentado abaixo, que se estrutura a partir do método 5w2h. Para cada requisito, considere identificar as sete questões referentes a funcionalidade:
O quê? | Quem? | Como? | Onde? | Quando? | Quanto? | Por quê?

Modelo 3

Dado o [ESPECÍFICO] enviado contém tudo o que um requisito deve ter. Elabore [ESPECÍFICO], quero que o 5w2h seja o guia para escrever cada requisito.
Ex.: RF01: o sistema deve (complete com as respostas do 5w2h)

Fonte: Elaborado pela autora (2026).

Matriz 5W2H: a proposta da matriz era identificar pontos necessários no requisito e cruzá-los com cada pergunta do método, de modo que, quando enviada como insumo para a IA, ela consiga interpretar e retornar respostas conforme a estrutura proposta.

Figura 6 – Matriz do 5W2H

informação \ pergunta	O que?	Por quê?	Quem?	Quando?	Onde?	Como?	Quanto?
Funcionalidade	o que essa funcionalidade faz?	por que ela faz o que faz?	quem atua nessa função?	quando essa função deve ser ativada?	onde essa função deve ficar?	como essa função vai funcionar?	
Dados	1. quais dados devem ser coletados para essa função? 2. quais entradas são permitidas para cada um desses dados?	por que esse(s) dado(s) deve(em) ser coletado(s)?	quem manipulará esse(s) dado(s)?	1. quando esse dado deve ser coletado? 2. quando esse dado será excluído?	1. onde esse dado será coletado? 2. onde esse dado será armazenado?	1. como esse dado será coletado? 2. como esse dado será armazenado? 3. como esse dado será excluído?	quanto tempo/volume/caracteres?
Restrições	Quais são as restrições aplicadas a essa funcionalidade?	por que essas restrições devem ser aplicadas?	quem receberá essa restrição (ator/dado/tela/funcionalidade)?	quando essas restrições devem ser aplicadas?	onde essas restrições devem ser aplicadas?	como essas restrições devem ser aplicadas para essa funcionalidade?	

Fonte: Elaborado pela autora (2026).

5.2.7 *Análise de Dados*

Por se tratar de um estudo quantitativo experimental com apoio qualitativo, a análise de dados foi dividida de forma a contemplar as duas categorias analíticas:

Quantitativa: realizada utilizando o software JASP 0.98.0, a partir dos dados coletados nos formulários enviados aos participantes e do formulário de análise dos artefatos produzidos por eles.

Qualitativa: realizada por meio de uma análise interpretativa das respostas abertas dos formulários de Perfil e Feedback.

5.2.8 *Ameaças à Validade*

A identificação das ameaças à validade permite reconhecer fatores que podem influenciar os resultados do experimento e comprometer sua confiabilidade. Conforme proposto por Wohlin *et al.* (2012), foram consideradas as ameaças relacionadas à validade interna, externa, de construção e de conclusão, bem como as estratégias adotadas para mitigá-las.

5.2.8.1 *Validade Interna*

A validade interna está relacionada aos fatores que podem influenciar os resultados do experimento de forma não intencional. Neste estudo, foram identificadas as seguintes ameaças:

- diferença de conhecimento entre os participantes;
- influência do pesquisador;
- familiaridade com IA.

Para minimizar essas ameaças, foram adotadas as seguintes estratégias:

- instruções padronizadas;
- divisão equilibrada dos grupos.

5.2.8.2 *Validade Externa*

A validade externa refere-se à possibilidade de generalização dos resultados obtidos. As seguintes ameaças foram identificadas:

- instabilidade da internet no laboratório;
- dificuldades de acesso às ferramentas de IA;
- dificuldade de manipulação da planilha do documento de requisitos.

Neste estudo, não foi adotada uma estratégia específica de mitigação para essa ameaça.

5.2.8.3 *Validade de Construção*

A validade de construção está relacionada à adequação das métricas utilizadas para representar os conceitos investigados. A ameaça identificada foi:

- as métricas podem não representar adequadamente a qualidade dos requisitos.

Para reduzir essa ameaça, foi adotada a estratégia de utilizar métricas baseadas na literatura.

5.2.8.4 *Validade de Conclusão*

A validade de conclusão está relacionada à confiabilidade das inferências obtidas a partir dos resultados do experimento. A ameaça identificada foi:

- amostra reduzida.

Como estratégia de mitigação, foram utilizados testes estatísticos adequados.

5.2.9 *Resultado Esperado*

Com este experimento, espera-se que a integração do método 5W2H + IA:

- produza requisitos mais claros, completos e corretos;

- reduza ambiguidades e redundâncias;
- aumente a rastreabilidade dos requisitos;
- melhore a documentação quanto à padronização e à especificação;
- reduza o tempo de elicitação;
- aumente a confiança dos participantes;
- forneça maior suporte cognitivo nos processos iniciais de Engenharia de Requisitos.

5.3 Operação (Operation)

A operação é a fase de execução do experimento propriamente dita, sendo subdividida em:

Preparação: momento em que foram preparados todos os instrumentos utilizados no experimento;

Execução: coleta de dados com os participantes;

Validação dos dados: verificação de consistência dos dados coletados.

Os dados produzidos a partir destas etapas foram analisados, interpretados e apresentados nas etapas de Análise e Interpretação (etapa 4) e de Apresentação e Empacotamento (etapa 5). Os resultados podem ser verificados no Capítulo 6.

5.3.1 Apresentação & Empacotamento (*Presentation & Package*)

A etapa de apresentação e empacotamento tem como objetivo consolidar e documentar os resultados obtidos durante o experimento, tornando o processo transparente, reproduzível e passível de reutilização em estudos futuros. Conforme proposto por Wohlin *et al.* (2012), essa etapa consiste na organização dos artefatos produzidos, na documentação dos procedimentos adotados e na disponibilização das evidências necessárias para a compreensão e replicação da pesquisa, onde todos os artefatos produzidos para este estudo foram disponibilizados ao longo dos capítulos descritos nesse documento.

5.3.2 *Cr terios de avalia  o dos requisitos*

A qualidade dos documentos de requisitos produzidos pelos participantes foi avaliada a partir de um conjunto de crit rios amplamente utilizados na Engenharia de Requisitos e fundamentados nas caracter sticas de qualidade descritas pela norma ISO/IEC/IEEE 29148:2018, que estabelece recomenda  es para especifica  o de requisitos de sistemas e software. Essa norma define que requisitos de qualidade devem apresentar caracter sticas que favore am sua correta compreens o, implementa  o, valida  o e manuten  o ao longo do ciclo de desenvolvimento.

Com base nessas recomenda  es, foram definidos os seguintes crit rios de avalia  o: **clareza, completude, corretude, verificabilidade, testabilidade, rastreabilidade, organiza  o e padroniza  o**. Cada crit rio foi operacionalizado por meio de tr s afirma  es avaliadas em escala Likert de cinco pontos, permitindo mensurar a percep  o dos avaliadores quanto   qualidade dos documentos produzidos pelos participantes. O Quadro 8 apresenta a finalidade de cada crit rio e os principais aspectos considerados durante a avalia  o.

Quadro 8 – Critérios utilizados na avaliação dos documentos de requisitos

Critério	Aspectos verificados
Clareza	Verificou se os requisitos estavam escritos de forma objetiva, compreensível e sem ambiguidades.
Compleitude	Avaliou se todas as informações necessárias para compreender o requisito estavam presentes, sem omissões relevantes.
Corretude	Verificou se os requisitos representavam corretamente as necessidades descritas no cenário proposto, sem inconsistências ou informações incorretas.
Verificabilidade	Avaliou se era possível confirmar objetivamente o atendimento de cada requisito por meio de inspeção, análise ou testes.
Testabilidade	Verificou se os requisitos permitiam a definição de casos de teste capazes de validar seu correto funcionamento.
Rastreabilidade	Avaliou a possibilidade de relacionar cada requisito à sua origem e acompanhar sua evolução ao longo do desenvolvimento.
Organização	Verificou se o documento apresentava estrutura lógica, facilitando a leitura e a localização das informações.
Padronização	Avaliou a consistência na utilização da linguagem, terminologia e formato de especificação dos requisitos.

Fonte: Elaborado pela autora (2026)

6 ANÁLISE E INTERPRETAÇÃO DOS RESULTADOS

Neste capítulo são apresentados os resultados do experimento controlado, organizados por categoria: i) quantitativo e ii) qualitativo.

6.1 Dados Quantitativos

Nesta seção, apresentam-se os dados quantitativos provenientes dos três conjuntos de dados do experimento, devidamente tratados e analisados.

6.1.1 *Limpeza dos Dados Quantitativos*

Antes do início da análise dos dados, foi realizada a etapa de limpeza, fundamental para a consistência dos resultados. O experimento gerou três conjuntos de dados:

1. **Perfil dos participantes:** formulário aplicado antes do experimento, utilizado para mapear o perfil dos estudantes que atuaram no experimento, coletando dados acadêmicos e informações sobre experiência em Engenharia de Requisitos e familiaridade com IA;
2. **Qualidade dos documentos produzidos:** métricas extraídas dos artefatos gerados durante o experimento;
3. **Feedback do participante:** respostas obtidas no questionário pós-experimento, relacionadas à usabilidade, confiança, suporte cognitivo e satisfação com a metodologia utilizada.

6.1.1.1 *Etapa 1 - Cruzamento dos dados - Formulários de Perfil e Feedback*

Inicialmente, realizou-se o cruzamento dos dados para identificar os participantes que responderam tanto ao formulário de Perfil quanto ao de Feedback. Foram

excluídas da análise as respostas dos indivíduos que não completaram ambos os instrumentos. Esse procedimento de filtragem foi fundamental para mitigar vieses e eliminar inconsistências na amostragem, assegurando a homogeneidade e a integridade do conjunto de dados submetido às análises subsequentes (Quadro 9).

Quadro 9 – Conjunto de dados após a etapa de limpeza

Conjunto de dados	Coletados	Válidos
Perfil dos participantes	24 respondentes	17 válidos
Feedback dos participantes	18 respondentes	14 válidos

Fonte: Elaborado pela autora (2026).

Nota: Um participante respondeu ao formulário de *Feedback do Participante*, mas não respondeu ao formulário de *Perfil do Participante*. Por esse motivo, seus dados foram excluídos da análise.

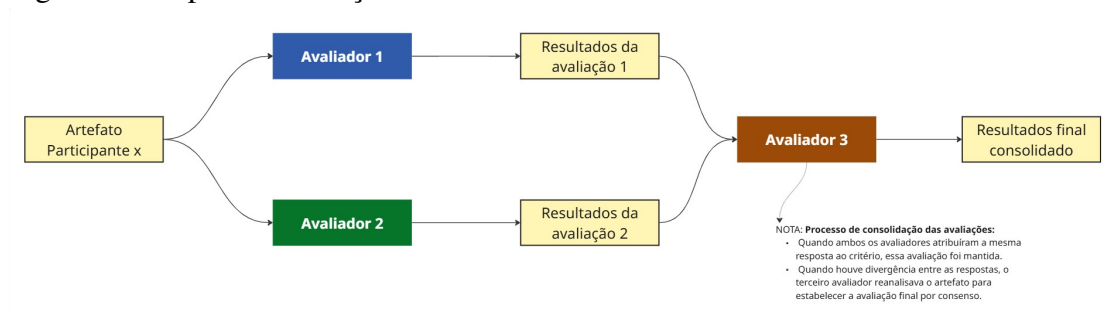
6.1.1.2 Etapa 2 - Análise dos artefatos - Grupos A e B

Após a etapa de cruzamento dos dados, iniciou-se a análise dos artefatos produzidos pelos participantes de ambos os grupos. Para essa atividade, contou-se com o apoio de outros quatro pesquisadores, responsáveis pela avaliação dos documentos. Com o objetivo de garantir maior consistência e padronização, foi elaborado um formulário específico para a avaliação dos artefatos, apresentado no Apêndice E.

Durante a avaliação, verificou-se que dois participantes do Grupo B não compartilharam o link do prompt, o que impossibilitou a análise de parte do formulário. Além disso, um participante do mesmo grupo compartilhou um link inválido. Em razão dessas inconsistências, esses casos foram excluídos das análises subsequentes.

Cada artefato foi avaliado de forma independente por dois pesquisadores, resultando em dois pareceres para cada documento. Posteriormente, foi realizada uma terceira etapa de consolidação, com o objetivo de validar e sistematizar os resultados obtidos para não enviesá-los (Figura 7).

Figura 7 – Etapas da avaliação de cada artefato



Fonte: Elaborado pela autora (2026)

Todos os avaliadores envolvidos na análise dos artefatos possuíam conhecimentos em Engenharia de Requisitos, adquiridos por meio da disciplina de Requisitos de Software. Além disso, dois dos avaliadores possuíam experiência prática, equivalente ao nível de analista de requisitos júnior, decorrente de sua participação ativa em um projeto de desenvolvimento de software da universidade. A avaliação de desempate e a consolidação dos resultados foram conduzidas pela pesquisadora deste estudo, que reavaliava os critérios em que havia divergência entre os dois avaliadores iniciais, estabelecendo a avaliação final por consenso.

6.1.1.3 Etapa 3 - Preparação dos dados

Concluída a etapa de avaliação, os dados foram preparados para a análise estatística. Inicialmente, as respostas obtidas por meio da escala Likert de cinco pontos foram convertidas em valores numéricos (Quadro 10).

Quadro 10 – Conversão da escala Likert em valores numéricos

Resposta na escala Likert	Valor numérico
Discordo totalmente	1
Discordo parcialmente	2
Nem concordo nem discordo	3
Concordo parcialmente	4
Concordo totalmente	5

Em seguida, os dados obtidos por meio da escala Likert de cinco pontos foram previamente submetidos ao procedimento de *reverse scoring* (recodificação reversa) para as questões redigidas de forma negativa. Esse procedimento consiste em inverter a pontuação originalmente atribuída às respostas, preservando a ordem das categorias, porém alterando sua direção de interpretação. Dessa forma, respostas que originalmente indicavam menor qualidade passam a receber maiores escores, enquanto respostas que indicavam maior ocorrência de problemas recebem menores escores.

A recodificação reversa foi adotada com o objetivo de padronizar a interpretação dos resultados, garantindo que, em todos os critérios avaliados, valores mais elevados representassem maior qualidade dos requisitos especificados. Essa padronização facilita a interpretação das estatísticas descritivas e inferenciais, bem como a comparação entre diferentes critérios avaliados.

O procedimento foi aplicado inclusive aos critérios cujas questões eram integralmente formuladas de maneira negativa, como os critérios relacionados à presença de ambiguidades, redundâncias, contradições, vieses e informações não justificadas introduzidas pela IA. Nesses casos, a concordância com as afirmações indicava a existência de defeitos no documento de requisitos, ou seja, menor qualidade do artefato. Assim, embora todas as questões de determinado critério possuíssem a mesma orientação negativa, a recodificação foi necessária para manter a consistência interpretativa entre todos os critérios do instrumento, fazendo com que maiores escores correspondessem, uniformemente, a melhores resultados.

A recodificação foi realizada utilizando a equação abaixo, considerando uma escala Likert de cinco pontos.

$$x' = 6 - x$$

em que x representa a resposta original e x' corresponde ao valor recodificado.

Quadro 11 – Exemplo de aplicação do *reverse scoring*

Questão (redação negativa)	Resposta original	Após recodificação
Foram identificadas ambiguidades nos requisitos?	5 (Concordo totalmente)	1
Foram identificadas ambiguidades nos requisitos?	4 (Concordo parcialmente)	2
Foram identificadas ambiguidades nos requisitos?	3 (Nem concordo nem discordo)	3
Foram identificadas ambiguidades nos requisitos?	2 (Discordo parcialmente)	4
Foram identificadas ambiguidades nos requisitos?	1 (Discordo totalmente)	5

Interpretação: antes da recodificação, valores mais elevados indicavam maior ocorrência de defeitos (ambiguidades, redundâncias, contradições, vieses etc.). Após a aplicação do *reverse scoring*, valores mais elevados passaram a representar maior qualidade dos requisitos, mantendo a mesma direção interpretativa para todos os critérios avaliados.

Por fim, considerando que cada critério era composto por três itens, foi calculado a média aritmética das respostas para obtenção do escore final por critério.

6.1.2 Resultados das análises quantitativas

6.1.2.1 Perfil dos participantes

Mapeando o perfil acadêmico, a experiência com Engenharia de Requisitos (ER) e familiaridade com ferramentas de IA, foi encontrado que todos os participantes

do experimento estão cursando o segundo semestre do curso de Engenharia de Software e utilizam a Inteligência artificial para: *"Estudo e esclarecimento de dúvidas conceituais., Geração ou correção de código-fonte (programação)., Geração de criativos (ex.: imagem e vídeo), Escrita de textos, relatórios ou documentação."*

Analisando estatisticamente a experiência com Engenharia de Requisitos (ER) e familiaridade com ferramentas de IA, obtivemos o seguinte resultados (Tabela 1):

Tabela 1 – Estatística Descritiva - Perfil do Participante - Parte 1

Métrica	Conhecimento em Engenharia de Requisitos		Conhecimento sobre requisitos de software		Desempenho elicitando requisitos de software	
	Grupo A	Grupo B	Grupo A	Grupo B	Grupo A	Grupo B
Median	1,000	2,000	2,000	2,500	2,000	2,000
Mean (arithmetic)	1,750	2,333	2,375	2,333	2,125	2,167
Std. Deviation	1,165	0,5164	1,188	0,8165	1,126	0,7528
Minimum	1,000	2,000	1,000	1,000	1,000	1,000
Maximum	4,000	3,000	4,000	3,000	4,000	3,000

A Tabela 1 apresenta a estatística descritiva do perfil dos participantes dos Grupos A e B, considerando cinco variáveis relacionadas à experiência prévia em Engenharia de Requisitos e ao uso de IA Generativa: conhecimento em Engenharia de Requisitos, conhecimento sobre os tipos de requisitos de software, desempenho na elicitación de requisitos de software, familiaridade com IA generativa e frequência de uso dessa tecnologia.

De modo geral, observa-se que os dois grupos apresentaram perfis semelhantes antes da realização do experimento. Em relação ao **conhecimento em Engenharia de Requisitos**, o Grupo B apresentou mediana (2) e média (2,333) ligeiramente superiores às do Grupo A (mediana = 1; média = 1,750), sugerindo um conhecimento inicial discretamente maior. Entretanto, o Grupo A apresentou maior variabilidade nas respostas, evidenciada pelo maior desvio-padrão (1,165), indicando uma composição mais heterogênea em relação ao nível de experiência.

Quanto ao **conhecimento sobre os tipos de requisitos de software**, ambos os grupos apresentaram resultados bastante próximos. Embora o Grupo B tenha obtido uma mediana ligeiramente superior (2,5 contra 2,0), as médias foram praticamente idênticas (2,375 no Grupo A e 2,333 no Grupo B), indicando um nível semelhante de conhecimento entre os participantes.

Para a variável **desempenho na elicitação de requisitos de software**, verifica-se que os grupos também apresentaram resultados praticamente equivalentes. As medianas foram iguais (2) e as médias diferiram em apenas 0,042 pontos (2,125 no Grupo A e 2,167 no Grupo B), sugerindo que ambos iniciaram o experimento com níveis muito próximos de experiência prática em atividades de elicitação.

Em relação à **familiaridade com IA generativa**(Tabela 2), observa-se uma tendência de maior experiência por parte do Grupo B, cuja mediana foi igual a 4 e média igual a 4,000, enquanto o Grupo A apresentou mediana de 3,5 e média de 3,375. Apesar dessa diferença, o Grupo A apresentou maior dispersão dos dados (desvio-padrão = 1,408), indicando que seus participantes possuíam níveis bastante variados de familiaridade com esse tipo de tecnologia.

Resultados semelhantes foram observados para a **frequência de uso da IA generativa** (Tabela 2). Ambos os grupos relataram utilizar ferramentas de IA com frequência relativamente elevada, apresentando médias de 3,875 (Grupo A) e 4,000 (Grupo B), bem como medianas de 3,5 e 4, respectivamente. Os desvios-padrão inferiores a 1 em ambos os grupos sugerem respostas relativamente homogêneas.

Tabela 2 – Estatística Descritiva - Perfil do Participante - Parte 2

Métrica	Familiaridade IA generativa		Frequência no uso de IA Generativa	
	Grupo A	Grupo B	Grupo A	Grupo B
Median	3,500	4,000	3,500	4,000
Mean (arithmetic)	3,375	4,000	3,875	4,000
Std. Deviation	1,408	0,8944	0,9910	0,8944
Minimum	1,000	3,000	3,000	3,000
Maximum	5,000	5,000	5,000	5,000

A análise de normalidade, realizada por meio do teste de Shapiro–Wilk (Tabela 3), mostrou que as variáveis conhecimento em Engenharia de Requisitos ($p = 0,007$) e frequência de uso da IA generativa ($p = 0,006$) não apresentaram distribuição normal, enquanto as demais variáveis atenderam ao pressuposto de normalidade ($p > 0,05$). Considerando que parte das variáveis violou esse pressuposto, além do fato de os dados serem provenientes de escalas do tipo Likert, a utilização de um teste não paramétrico mostrou-se adequada para a comparação entre os grupos.

Tabela 3 – Teste de normalidade - Perfil dos Participantes

Resíduos	W	p
Conhecimento em Engenharia de Requisitos	0,8124	,007
Conhecimento sobre os tipos de requisitos de software	0,9023	,122
Desempenho elicitando requisitos de software	0,8930	,089
Familiaridade IA generativa	0,9612	,742
Frequência no uso de IA Generativa	0,8036	,006

Nota. Resultados significativos sugerem um desvio da normalidade.

Para verificar a equivalência entre os grupos antes da intervenção, foi aplicado o teste de Mann-Whitney (Tabela 4), cuja escolha foi motivada pela verificação prévia do pressuposto de normalidade por meio do teste de Shapiro–Wilk. Os resultados indicaram ausência de diferenças estatisticamente significativas entre os grupos para

todas as variáveis avaliadas, uma vez que todos os valores de p foram superiores ao nível de significância de 5% ($p > 0,05$). Especificamente, os valores obtidos foram: conhecimento em Engenharia de Requisitos ($p = 0,154$), conhecimento sobre os tipos de requisitos ($p = 1,000$), desempenho na elicitação de requisitos ($p = 0,892$), familiaridade com IA generativa ($p = 0,463$) e frequência de uso de IA generativa ($p = 0,836$).

Tabela 4 – Teste t de Student para amostras independentes - Perfil do Participante

Variável	U	gl	p
Conhecimento em Engenharia de Requisitos	13,00		,154
Conhecimento sobre os tipos de requisitos de software	23,50		1,000
Desempenho elicitando requisitos de software	22,50		,892
Familiaridade IA generativa	18,00		,463
Frequência no uso de IA Generativa	22,00		,836

Nota. Teste-U de Mann-Whitney.

De maneira geral, os resultados indicam que os participantes dos Grupos A e B possuíam características iniciais bastante semelhantes quanto à experiência em Engenharia de Requisitos e ao uso de Inteligência Artificial Generativa. A inexistência de diferenças estatisticamente significativas entre os grupos fortalece a validade interna do experimento, uma vez que reduz a influência de possíveis diferenças prévias entre os participantes sobre os resultados obtidos após a aplicação dos tratamentos experimentais.

6.1.2.2 *Feedback dos Participantes*

O formulário do feedback foi dividido em duas partes: i) na primeira, os participantes dos dois grupos forneciam sua percepção acerca dos critérios de qualidade e da qualidade documental como um todo; ii) já na segunda, os participantes do Grupo B avaliavam sua percepção sobre a integração do método 5W2H com IA para elicitação e

especificação dos requisitos.

Na primeira parte, inicialmente foi obtido a estatística descritiva dos dados compartilhados por ambos os grupos, conforme mostrado na Tabela 5.

Tabela 5 – Estatística Descritiva - Feedback dos Participantes - Grupos A e B

	Facilidade de Execução da Atividade		Percepção sobre a Qualidade dos Requisitos		Comparação indireta entre os métodos	
	Grupo A	Grupo B	Grupo A	Grupo B	Grupo A	Grupo B
Median	4,100	4,600	3,700	4,300	4,000	4,500
Mean (arithmetic)	4,050	4,467	3,650	4,133	3,750	4,417
Std. Deviation	0,7540	0,6532	0,7982	0,8824	1,000	0,6646
Minimum	3,000	3,400	2,400	3,000	2,000	3,500
Maximum	5,000	5,000	5,000	5,000	5,000	5,000

Quando questionados a respeito da **facilidade sobre a execução da atividade**, o Grupo B obteve uma média e mediana maior em relação ao Grupo A, além de um desvio padrão ligeiramente menor, que indica certa homogeneidade nas respostas. Com relação à percepção dos participantes em relação à **qualidade dos requisitos elicitados**, o Grupo B percebeu requisitos com maior qualidade (média 4,13 contra 3,65), com uma dispersão um pouco maior (DP = 0,882), indicando respostas mais variadas. Quanto a **comparação indireta entre os métodos**, que avaliava a simplicidade da execução da atividade e a possibilidade de avançar as etapas sem muita dificuldade, o Grupo B obteve médias maiores e o Grupo A apresentou maior variabilidade nas respostas (DP = 1,00).

Após a conclusão da análise descritiva, foi realizado o Teste de Normalidade (Shapiro-Wilk), a fim de verificar se os dados seguem uma distribuição aproximadamente normal. A partir da Tabela 6, foi possível observar que todos os valores de p são maiores que 0,05, mostrando que não há evidências de violação da normalidade.

Tabela 6 – Teste de normalidade - Feedback do Participante - Grupos A e B

Resíduos	W	p
Facilidade de Execução da Atividade	0,9335	0,342
Percepção sobre a Qualidade dos Requisitos	0,9664	0,326
Comparação indireta entre os métodos	0,9350	0,358

Após a constatação da normalidade dos dados, foi verificado se as diferenças entre os grupos são estatisticamente significativas, através Teste de No(Tabela 7).

Tabela 7 – Teste *t* de Student para amostras independentes – Feedback dos participantes - Grupo A e B

Variável	t	gl	p	d de Cohen	EP d de Cohen	95% IC para d de Cohen	
						Inferior	Superior
Facilidade de Execução da Atividade	-1,081	12	,301	-0,5838	0,5594	-1,656	0,5111
Percepção sobre a Qualidade dos Requisitos	-1,073	12	,305	-0,5793	0,5591	-1,651	0,5151
Comparação indireta entre os métodos	-1,409	12	,184	-0,7610	0,5726	-1,847	0,3540

O teste mostrou que não há nenhuma diferença foi estatisticamente significativa entre as três variáveis avaliadas, isso porque todos os valores de *p* são maiores que 0,05. Embora o Grupo B tenha apresentado médias maiores em todas as medidas, não há evidência estatística suficiente para afirmar que os grupos diferem de forma significativa na população. Já os valores de *d* (Tamanho do efeito (Cohen's *d*)), sugerem que pode existir um efeito prático relevante, principalmente na comparação entre métodos, mas a amostra parece pequena para alcançar significância estatística, isso é importante porque, mesmo sem significância estatística, os tamanhos de efeito indicam uma tendência prática favorável ao Grupo B.

6.1.2.2.1 Feedback dos Participantes - Grupo B

Dado que os grupos possuíam atividades diferentes e o foco da pesquisa é na integração do método, foram incluídas, para os participantes do Grupo B, perguntas mais relacionadas ao estudo em questão e foi realizada uma análise descritiva dos dados obtidos. Para eles foram consideradas as seguintes questões: "Facilidade percebida", "Suporte cognitivo", "Confiança nos resultados", "Viés e confiabilidade, Supervisão humana", "Dependência da IA" e "Ajustes necessários", que nos auxiliaram a comparar a percepção que eles tiveram sobre a integração do método com IA para auxiliar na elicitación e especificação de requisitos, com relação ao que foi avaliado.

Conforme apresentados na Tabela 8, os resultados indicam que os participantes do Grupo B avaliaram positivamente a maior parte dos critérios analisados, especialmente aqueles relacionados à facilidade de uso, suporte cognitivo, confiança nos resultados e necessidade de ajustes. Entretanto, aspectos ligados ao viés e confiabilidade, supervisão humana e dependência da IA apresentaram avaliações mais baixas e maior dispersão entre os participantes.

Tabela 8 – Estatística Descritiva - Feedback dos Participantes - Grupo B - Parte 1

	Facilidade percebida	Suporte cognitivo	Confiança nos resultados
Median	5,000	4,500	4,250
Mean (arithmetic)	4,750	4,367	4,333
95% CI A. Mean Lower	4,311	3,623	3,698
95% CI A. Mean Upper	5,189	5,111	4,969
Std. Deviation	0,4183	0,7090	0,6055
Minimum	4,000	3,200	3,500
Maximum	5,000	5,000	5,000

O critério **Facilidade percebida** (Tabela 8) apresentou a maior média entre todos os fatores avaliados ($M = 4,75$; $DP = 0,42$), com mediana igual a 5, indicando que mais da metade dos participantes atribuiu a pontuação máxima. Além disso, o baixo

desvio-padrão demonstra elevada concordância entre os respondentes, reforçando que o sistema foi considerado fácil de utilizar. O intervalo de confiança de 95% para a média (4,31–5,19) confirma que essa percepção positiva foi consistente na amostra.

O critério **Suporte cognitivo** (Tabela 8) também recebeu avaliação elevada ($M = 4,37$; $DP = 0,71$), com mediana de 4,5, indicando que os participantes perceberam que a ferramenta contribuiu para apoiar seu raciocínio durante a realização da tarefa. A variabilidade foi moderada, sugerindo pequenas diferenças entre as avaliações individuais.

De forma semelhante, **Confiança nos resultados** (Tabela 8) apresentou média de 4,33 e mediana de 4,25, evidenciando que os participantes confiaram, em geral, nas respostas fornecidas pela ferramenta. O desvio-padrão relativamente baixo ($DP = 0,61$) demonstra boa consistência nas respostas.

Olhando para o critério **Viés e confiabilidade** (Tabela 9) percebemos uma média de 2,83 ($DP = 1,37$) e a mediana de 2,75. Como esse fator foi mensurado por meio de afirmações negativas, valores menores representam uma percepção mais favorável. Dessa forma, a média abaixo do ponto central da escala sugere uma tendência dos participantes a discordarem da existência de vieses ou problemas de confiabilidade, indicando que, em geral, a ferramenta não foi percebida como tendenciosa ou pouco confiável. Entretanto, o elevado desvio-padrão e a amplitude das respostas (mínimo = 1; máximo = 5) revelam que houve divergência entre os participantes. Enquanto parte deles discordou fortemente das afirmações negativas, demonstrando elevada confiança na ferramenta, outros atribuíram notas mais altas, sugerindo que perceberam algum nível de viés ou limitações quanto à confiabilidade. Assim, embora o resultado geral seja favorável, observa-se uma percepção heterogênea sobre esse aspecto.

Em relação à **Supervisão humana** (Tabela 9), a média foi de 3,33 e a mediana de 3,5, sugerindo uma posição intermediária dos participantes. O desvio-padrão relativamente elevado ($DP = 1,21$) indica que não houve consenso sobre a necessidade

de supervisão humana durante o uso da ferramenta.

Tabela 9 – Estatística Descritiva - Feedback dos Participantes - Grupo B - Parte 2

	Viés e confiabilidade	Supervisão humana
Median	2,750	3,500
Mean (arithmetic)	2,833	3,333
95% CI A. Mean Lower	1,400	2,062
95% CI A. Mean Upper	4,267	4,604
Std. Deviation	1,366	1,211
Minimum	1,000	2,000
Maximum	5,000	5,000

A **Dependência da IA** (Tabela 10) apresentou média de 2,67 e mediana de 2,5, representando um dos menores valores observados. Isso sugere que, de maneira geral, os participantes não perceberam uma elevada dependência da inteligência artificial para elicitar e especificar requisitos utilizando o método. Entretanto, o desvio-padrão de 0,82 demonstra que essa percepção variou moderadamente entre os indivíduos.

Por fim, o critério **Ajustes necessários** (Tabela 10) apresentou média de 4,17 e mediana de 4,5, indicando que os participantes reconheceram que, apesar da qualidade das respostas da IA, ainda foi necessário realizar alguns ajustes. Esse foi o critério com maior dispersão ($DP = 1,47$), refletindo diferenças consideráveis nas opiniões dos participantes quanto ao nível de refinamento necessário.

Tabela 10 – Estatística Descritiva - Feedback dos Participantes - Grupo B - Parte 3

	Dependência da IA	Ajustes necessários
Median	2,500	4,500
Mean (arithmetic)	2,667	4,167
95% CI A. Mean Lower	1,810	2,622
95% CI A. Mean Upper	3,524	5,711
Std. Deviation	0,8165	1,472
Minimum	2,000	2,000
Maximum	4,000	6,000

De maneira geral, os participantes avaliaram bem a integração do método 5W2H com Inteligência Artificial Generativa para auxiliar nos processos iniciais da Engenharia de Requisitos.

6.1.3 Avaliação da Qualidade dos Artefatos

6.1.3.1 Avaliação dos artefatos - Grupos A e B

Para cada critério, inicialmente foram calculadas medidas descritivas (média e desvio-padrão) a partir das respostas dos avaliadores. Em seguida, verificou-se a distribuição dos dados utilizando o teste de Shapiro–Wilk. Como os dados não atenderam aos pressupostos de normalidade, adotou-se o teste não paramétrico de Mann–Whitney para comparar os resultados obtidos pelos grupos A e B.

A análise de cada critério consistiu em verificar se existia diferença estatisticamente significativa entre os documentos produzidos utilizando a abordagem tradicional e aqueles elaborados com o apoio da integração entre o método 5W2H e a Inteligência Artificial Generativa. Para isso, foram interpretados conjuntamente os valores das médias, que indicam a percepção geral dos avaliadores sobre a qualidade dos documentos, e o valor de significância (p), utilizado para verificar se as diferenças observadas entre os grupos poderiam ser atribuídas ao tratamento experimental.

A Tabela 11 apresenta os resultados das estatísticas dos critérios que observam a clareza, a completude e a corretude dos documentos produzidos, onde foi evidenciado uma vantagem consistente do Grupo B em relação ao Grupo A.

Tabela 11 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 1

	Clareza		Compleitude		Corretude	
	Grupo A	Grupo B	Grupo A	Grupo B	Grupo A	Grupo B
Median	2,333	4,167	3,500	3,833	3,333	5,000
Mean (arithmetic)	2,333	4,000	3,167	3,833	3,250	4,611
95% CI A. Mean Lower	1,531	3,088	2,651	–	2,311	3,969
95% CI A. Mean Upper	3,135	4,912	3,683	–	4,189	5,253
Std. Deviation	0,9595	0,8692	0,6172	0,1826	1,123	0,6116
Minimum	1,333	2,667	2,333	3,667	1,667	3,667
Maximum	3,667	5,000	3,667	4,000	4,667	5,000

Nota: Os valores do intervalo de confiança (CI) para o Grupo B não foram computados pelo programa estatístico em decorrência da baixa variabilidade dos dados observada na amostra, combinada com as restrições matemáticas do Modelo T (distribuição *t* de Student). Como o desvio padrão deste grupo apresentou-se extremamente reduzido (0,1826) e a amplitude entre os valores mínimo e máximo foi estreita, a variância amostral revelou-se insuficiente para a estimativa precisa dos limites probabilísticos bicaudais exigidos pelo referido modelo.

Para **clareza**, o Grupo B apresentou média de 4,00, enquanto o Grupo A obteve 2,33, indicando que os documentos produzidos pelo Grupo B foram considerados significativamente mais claros pelos avaliadores. Em **compleitude**, o Grupo B também obteve desempenho superior, com média de 3,83 frente a 3,17 do Grupo A, além de apresentar baixa variabilidade nas avaliações, refletindo elevado consenso entre os avaliadores. Da mesma forma, no critério que avalia a cobertura e adequação ao cenário disponibilizado, o Grupo B alcançou média de 4,61, superando a média de 3,25 do Grupo A, o que demonstra que seus documentos foram percebidos como mais **corretos e consistentes**.

Com relação aos critérios que avaliam a verificabilidade dos requisitos, a testabilidade dos requisitos e critérios de aceitação e a rastreabilidade, os resultados das análises também demonstram um desempenho superior do Grupo B em comparação ao Grupo A (Tabela 12).

Tabela 12 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 2

	Verificabilidade		Testabilidade		Rastreabilidade	
	Grupo A	Grupo B	Grupo A	Grupo B	Grupo A	Grupo B
Median	1,500	4,000	1,167	4,167	4,375	5,000
Mean (arithmetic)	1,875	3,667	1,333	4,167	4,344	4,875
95% CI A. Mean Lower	1,033	2,321	0,8623	3,512	3,942	4,655
95% CI A. Mean Upper	2,717	5,012	1,804	4,821	4,746	5,095
Std. Deviation	1,007	1,282	0,5634	0,6236	0,4807	0,2092
Minimum	1,000	1,667	1,000	3,333	3,750	4,500
Maximum	3,333	5,000	2,667	5,000	5,000	5,000

Em **verificabilidade** o Grupo B apresentou média de 3,67 enquanto o Grupo A obteve 1,88 indicando que os requisitos produzidos pelo Grupo B foram considerados mais fáceis de verificar e validar. A diferença foi ainda mais expressiva em **testabilidade** na qual o Grupo B alcançou média de 4,17 frente a 1,33 do Grupo A, sugerindo que seus requisitos favorecem significativamente a definição e execução de critérios de aceitação. Já em relação à **rastreabilidade**, embora ambos os grupos tenham obtido avaliações elevadas, o Grupo B apresentou média ligeiramente superior comparando-se ao Grupo A, indicando uma melhor capacidade de relacionar os requisitos às suas respectivas fontes, objetivos e artefatos do processo de desenvolvimento. Além das médias mais altas, o Grupo B apresentou menor variabilidade nas avaliações, especialmente em rastreabilidade, evidenciando maior consistência na percepção dos especialistas quanto à qualidade dos requisitos produzidos.

Na Tabela 13, apresentamos os resultados das análises estatísticas dos critérios que cobrem aspectos de organização do documento de requisitos no momento da especificação, a padronização da escrita desses requisitos e verificação com relação a requisitos ambíguos, redundantes e contraditórios.

Tabela 13 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 3

	Organização		Padronização		Ambiguidade	
	Grupo A	Grupo B	Grupo A	Grupo B	Grupo A	Grupo B
Median	3,167	4,500	4,333	5,000	4,000	5,000
Mean (arithmetic)	3,167	4,389	4,125	4,833	4,125	4,556
95% CI A. Mean Lower	2,407	3,710	3,412	4,541	3,530	–
95% CI A. Mean Upper	3,926	5,068	4,838	5,126	4,720	–
Std. Deviation	0,9085	0,6469	0,8533	0,2789	0,7113	0,6885
Minimum	2,000	3,333	2,333	4,333	3,000	3,667
Maximum	4,333	5,000	5,000	5,000	5,000	5,000

Nos critérios de **organização e padronização**, o Grupo B obteve desempenho superior, com médias de 4,39 e 4,83, respectivamente, enquanto o Grupo A alcançou médias de 3,17 e 4,13. Esses resultados indicam que os documentos produzidos pelo Grupo B foram percebidos como mais bem estruturados, organizados e coerente a um padrão de escrita dos requisitos. Além disso, o Grupo B apresentou menor variabilidade nas avaliações, evidenciando maior consenso entre os avaliadores quanto à qualidade desses atributos. Em contrapartida, no quando olhamos para a **ambiguidade**, que representa um aspecto negativo da qualidade documental e no qual valores menores indicam melhor desempenho, o Grupo A apresentou média ligeiramente inferior (4,13) à do Grupo B (4,56), sugerindo uma menor presença de ambiguidades em seus documentos. Entretanto, a diferença entre os grupos foi relativamente pequena, indicando que ambos ainda apresentaram níveis elevados de ambiguidade segundo a avaliação dos especialistas.

Finalizando a análise descritiva com os critérios que avaliam a qualidade do documento de um modo geral, com relação ao seu uso no desenvolvimento e facilidade de interpretação por outros profissionais que terão acesso a ele, além do critério que identifica a presença de erros ou omissões que podem impossibilitar o uso do documento (Tabela 14).

Tabela 14 – Estatística descritiva – Qualidade documental - Grupos A e B - Parte 4

	Qualidade Documental		Erros	
	Grupo A	Grupo B	Grupo A	Grupo B
Median	1,000	4,500	1,000	3,500
Mean (arithmetic)	1,292	4,222	1,375	3,389
95% CI A. Mean Lower	0,9443	3,266	0,8497	2,189
95% CI A. Mean Upper	1,639	5,178	1,900	4,589
Std. Deviation	0,4155	0,9108	0,6284	1,143
Minimum	1,000	2,667	1,000	1,667
Maximum	2,000	5,000	2,667	5,000

Para **qualidade documental**, o Grupo B obteve desempenho significativamente superior, com média de 4,22, enquanto o Grupo A apresentou média de 1,29, indicando que os documentos produzidos utilizando o método integrado, foram avaliados como de maior qualidade documental. Em contrapartida, no critério de **erros**, que constitui um aspecto negativo da qualidade dos documentos, o Grupo A apresentou média de 1,38, inferior à do Grupo B (3,39), sugerindo que seus documentos continham menor quantidade de erros segundo a avaliação dos especialistas. Além disso, observa-se que o Grupo B apresentou menor variabilidade nas avaliações de qualidade documental, enquanto, para o critério erros, ambos os grupos apresentaram maior dispersão das respostas, especialmente o Grupo B, indicando menor consenso entre os avaliadores quanto à ocorrência de erros.

Para verificar a normalidade dos dados, aplicamos o teste de normalidade de Shapiro-Wilk, que indicou que os critério de Completude ($p = 0,049$) e Padronização ($p = 0,034$) violaram a hipótese de normalidade, justificando o uso de um teste não paramétrico para comparação entre os grupos. Nos demais critérios, não houve evidências suficientes para rejeitar a normalidade ($p > 0,05$) (Tabela 15).

Tabela 15 – Teste de normalidade - Qualidade Documental - Grupos A e B

Resíduos	W	p
Clareza	0,9267	,275
Compleitude	0,8746	,049
Corretude	0,9289	,294
Verificabilidade	0,9335	,341
Testabilidade	0,9077	,146
Rastreabilidade	0,9451	,488
Organização	0,9443	,476
Padronização	0,8636	,034
Ambiguidade	0,9081	,148
Qualidade Documental	0,8930	,089
Erros	0,9389	,405

Apesar da constatação de não normalidade para os critérios de Compleitude e Padronização, seguimos com a aplicação do teste não paramétrico de Mann-Whitney, onde houve uma diferença estatisticamente significativa entre os dois grupos na grande maioria dos critérios de qualidade avaliados. Os critérios Clareza, Compleitude, Corretude, Verificabilidade, Testabilidade, Rastreabilidade, Organização, Padronização, Qualidade Documental, e Erros apresentaram valores de $p < 0,05$, indicando que as distribuições desses resíduos diferem de forma relevante entre os grupos avaliados. Apenas o critério Ambiguidade não demonstrou diferença estatística significativa ($p = 0,307$), sugerindo um comportamento similar entre os grupos analisados para essa variável específica.

Complementando a análise, o tamanho do efeito medido pela correlação ranque-bisserial revela que o impacto dessas diferenças é de magnitude muito forte na maioria dos casos. Destacam-se os critérios Testabilidade e Qualidade Documental que alcançaram o valor máximo de correlação (1,000), seguidos de perto por Erros (0,8958) e Clareza (0,7917). Todos os intervalos de confiança de 95% para esses efeitos significativos mantiveram-se estritamente positivos e distantes do zero, o que reforça a robustez e a precisão da rejeição da hipótese nula para estas variáveis (Tabela 16).

Tabela 16 – Teste U de Mann–Whitney – Qualidade documental - Grupos A e B

Variável	U	gl	p
Clareza	5,000		,016
Compleitude	6,000		,015
Corretude	7,000		,030
Verificabilidade	5,500		,019
Testabilidade	0,000		,002
Rastreabilidade	7,500		,033
Organização	6,500		,027
Padronização	8,500		,045
Ambiguidade	16,00		,307
Qualidade Documental	0,000		,002
Erros	2,500		,006

Além disso, a correlação ranque-bisserial, utilizada como medida de tamanho de efeito, foi elevada na maioria dos critérios (entre aproximadamente 0,65 e 1,00), indicando que as diferenças observadas possuem não apenas significância estatística, mas também grande magnitude prática. Destacam-se os critérios de Testabilidade e Qualidade Documental, ambas com tamanho de efeito igual a 1,00, representando separação completa entre os grupos nesse conjunto de dados, conforme apresentado na Tabela 17

Tabela 17 – Tamanho de Efeito - Correlação Ranque-Bisserial

Variável	Correlação ranque-bisserial	EP Correlação ranque-bisserial	95% IC para Correlação ranque-bisserial	
			Inferior	Superior
Clareza	0,7917	0,3120	0,4164	0,9365
Compleitude	0,7500	0,3120	0,3278	0,9225
Corretude	0,7083	0,3120	0,2461	0,9081
Verificabilidade	0,7708	0,3120	0,3712	0,9295
Testabilidade	1,000	0,3120	1,000	1,000
Rastreabilidade	0,6875	0,3120	0,2076	0,9007
Organização	0,7292	0,3120	0,2862	0,9153
Padronização	0,6458	0,3120	0,1347	0,8855
Ambiguidade	0,3333	0,3120	-0,2785	0,7527
Qualidade	1,000	0,3120	1,000	1,000
Documental				
Erros	0,8958	0,3120	0,6741	0,9695

6.1.3.2 Avaliação dos artefatos - Grupo B

Avaliando apenas os critérios que foram destinados apenas para o Grupo B, identificamos que a maioria dos critérios relacionados à qualidade dos requisitos produzidos com apoio da LLM, receberam avaliações positivas. Os critérios positivos apresentaram médias entre aproximadamente 3,6 e 4,6, indicando concordância parcial ou total com as afirmações. Por outro lado, os critérios negativos apresentaram médias baixas (entre 1,5 e 2,5), sugerindo que os participantes, em geral, discordaram da presença de problemas no documento, o que representa um resultado favorável. Além disso, os desvios-padrão mostram que alguns critérios apresentaram maior consenso entre os participantes, enquanto outros revelaram maior variabilidade nas respostas.

Os critérios de Fidelidade da Transformação, Precisão Textual e Detalhamento receberam avaliações predominantemente positivas pelos avaliadores (Tabela 18). A **Fidelidade da Transformação** apresentou uma avaliação favorável ($M = 4,333$; $Md = 4,333$; $DP = 0,596$), sugerindo que os requisitos refletiam corretamente as perguntas do

método, sem distorções na interpretação e preservando o contexto original do cenário, com baixa dispersão das respostas. A **Precisão Textual** obteve a maior média entre os três critérios ($M = 4,444$; $Md = 4,5$; $DP = 0,544$), indicando elevada concordância de que os requisitos foram redigidos de forma clara, consistente e reduzindo ambiguidades, além de apresentar o menor desvio-padrão, evidenciando forte consenso entre os avaliadores. Já o critério de **Detalhamento** apresentou média positiva ($M = 3,944$; $Md = 4,667$), porém com maior variabilidade nas avaliações ($DP = 1,389$), indicando que, embora a percepção geral seja de que os requisitos continham um nível adequado de detalhamento, houve diferenças mais acentuadas entre os avaliadores quanto a esse aspecto.

Tabela 18 – Estatística descritiva – Qualidade documental - Grupo B - Parte 1

	Fidelidade da Transformação	Precisão Textual	Detalhamento
Median	4,333	4,500	4,667
Mean (arithmetic)	4,333	4,444	3,944
95% CI A. Mean Lower	3,708	3,873	2,487
95% CI A. Mean Upper	4,959	5,016	5,402
Std. Deviation	0,5963	0,5443	1,389
Minimum	3,667	3,667	2,000
Maximum	5,000	5,000	5,000

Quando olhamos para o critério que avalia a **relevância das informações fornecidas** nos artefatos provenientes do método 5W2H, notamos uma média de 4,111 ($Md = 4,833$; $DP = 1,573$), indicando que, na percepção dos avaliadores, os artefatos continham informações pertinentes aos requisitos especificados. Entretanto, o elevado desvio-padrão demonstra que houve maior divergência entre os avaliadores quanto à relevância das informações presentes em alguns documentos. Observando critério de **Cobertura do Cenário** observou-se uma média de 3,611 ($Md = 3,5$; $DP = 1,182$), sugerindo que, de modo geral, os artefatos contemplaram os principais elementos dos cenários analisados, embora esse tenha sido o critério positivo com menor média, indi-

cando que alguns documentos deixaram de abranger aspectos considerados importantes pelos avaliadores. Por sua vez, o critério **Impacto na Qualidade Final** obteve média de 3,889 (Md = 4,5; DP = 1,544), evidenciando que os avaliadores perceberam uma contribuição positiva da integração do método 5W2H com IA para processos iniciais da Engenharia de Requisitos, no que tange a qualidade final dos requisitos. Contudo, assim como observado em Relevância das Informações, a elevada dispersão das respostas indica que essa percepção variou entre os documentos avaliados, refletindo diferenças na qualidade alcançada pelos participantes (Tabela 19).

Tabela 19 – Estatística descritiva – Qualidade documental - Grupo B - Parte 2

	Relevância das Informações	Cobertura do Cenário	Impacto na Qualidade Final
Median	4,833	3,500	4,500
Mean (arithmetic)	4,111	3,611	3,889
95% CI A. Mean Lower	2,460	2,371	2,268
95% CI A. Mean Upper	5,762	4,851	5,510
Std. Deviation	1,573	1,182	1,544
Minimum	1,000	1,667	1,000
Maximum	5,000	5,000	5,000

Os critérios apresentados na Tabela 20 evidenciam diferentes percepções dos avaliadores em relação aos aspectos relacionados ao processo de utilização da IA e à qualidade dos artefatos produzidos pelos participantes do Grupo B. O critério de **Correções Necessárias** obteve média de 1,833 (Md = 1,5; DP = 0,931), indicando que, de modo geral, os avaliadores identificaram pouca necessidade de correções nos requisitos produzidos. De forma semelhante, o critério **Dependência da IA** apresentou uma das menores médias entre todos os critérios (M = 1,500; Md = 1,25; DP = 0,775), sugerindo que os artefatos evidenciaram baixa dependência das respostas geradas pela IA, indicando que os participantes realizaram intervenções e refinamentos durante a elaboração dos requisitos. O critério **Ajustes** também apresentou resultado favorável (M

= 2,500; Md = 2,25; DP = 1,095), indicando que, embora alguns documentos tenham exigido refinamentos adicionais, a não foram realizados muitos ajustes nos prompts e nos requisitos gerados.

Tabela 20 – Estatística descritiva – Qualidade documental - Grupo B - Parte 3

	Viés	Correções Necessárias	Dependência da IA	Ajustes
Median	5,000	1,500	1,250	2,250
Mean (arithmetic)	4,583	1,833	1,500	2,500
95% CI A. Mean Lower	3,743	0,8564	0,6871	1,350
95% CI A. Mean Upper	5,424	2,810	2,313	3,650
Std. Deviation	0,8010	0,9309	0,7746	1,095
Minimum	3,000	1,000	1,000	1,000
Maximum	5,000	3,000	3,000	4,000

No critério **Viés**, as respostas foram previamente invertidas para que todos os critérios apresentassem a mesma direção de interpretação, ou seja, médias mais elevadas representassem resultados mais favoráveis. Dessa forma, a média de 4,583 (Md = 5; DP = 0,801) indica que os avaliadores perceberam baixa presença de vieses nas respostas geradas pela IA. Além disso, o baixo desvio-padrão demonstra que essa percepção foi consistente entre os avaliadores, evidenciando consenso quanto à reduzida influência de vieses nos artefatos produzidos pelo Grupo B.

De modo geral, os resultados descritivos indicam que a aplicação da integração proposta foi bem avaliada pelos especialistas. Os critérios de qualidade apresentaram médias predominantemente elevadas, enquanto os critérios relacionados ao uso da IA evidenciaram baixa necessidade de correções, ajustes e dependência da ferramenta. Além disso, o critério Viés, após a inversão das respostas negativas, também apresentou média elevada, indicando baixa percepção de vieses. Em conjunto, os resultados sugerem que os artefatos produzidos alcançaram um nível satisfatório de qualidade.

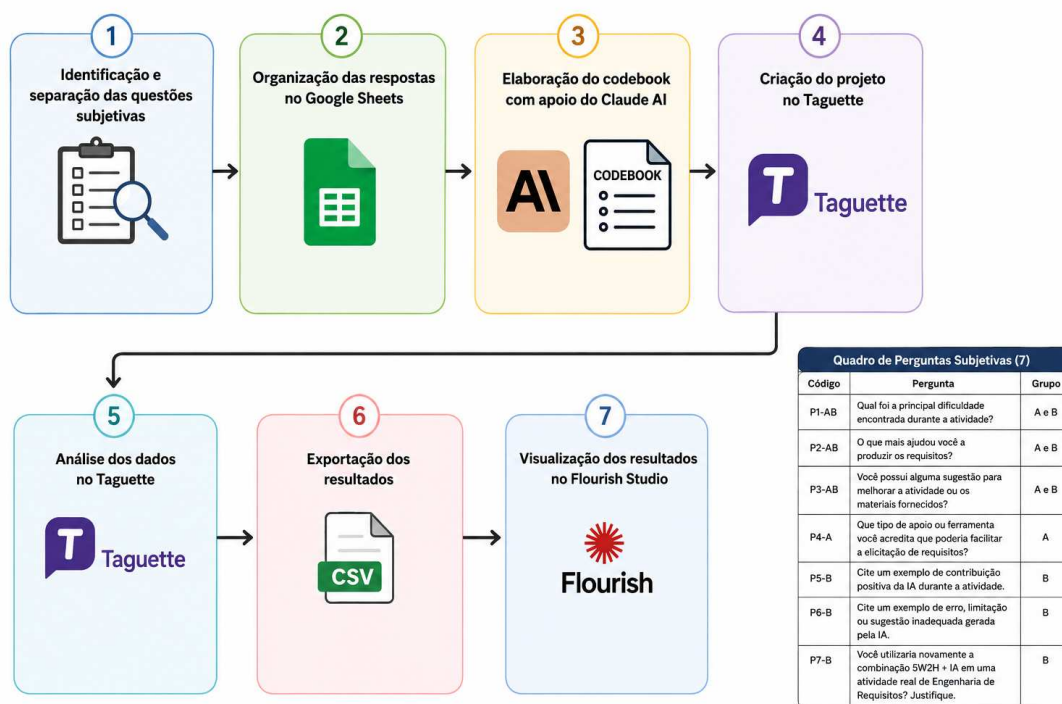
6.2 Dados Qualitativos

Nessa seção, é apresentado o resultado das análises qualitativas dos documentos gerados, a partir das respostas do formulário de Feedback do Participante.

6.2.1 Preparação dos dados

A preparação dos dados qualitativos seguiram um fluxo de sete etapas, conforme ilustrado na Figura 8:

Figura 8 – Etapas da preparação dos dados qualitativos



Fonte: Elaborado pelo ChatGPT (2026)

Etapa 1 : Antes de iniciar a análise dos dados, foi necessário realizar uma etapa de tratamento, a fim de garantir que eles estivessem adequados para serem analisados corretamente. Para isso, inicialmente foi feita a identificação e a separação dos

questões subjetivas presentes no formulário de feedback.

Etapa 2 : Em seguida, as respostas foram organizadas em uma planilha no Google Sheets, sendo criada uma aba para cada pergunta e suas respectivas respostas, o que facilitou a identificação dos códigos durante o processo de análise. Ao todo, o formulário possuía sete perguntas subjetivas.

Etapa 3 : A análise qualitativa dos dados foi realizada seguindo a técnica de **Codificação Fechada**, onde os códigos são organizados e definidos previamente e a codificação é orientada, a partir deles. Portanto, com o auxílio do Claude AI (Apêndice F), foi elaborado o *codebook*, contendo as tags e as descrições de cada uma delas (Apêndice G).

Etapa 4 : Após a etapa anterior, foi criado no Taguette, um projeto no qual foram importados os sete documentos de cada pergunta (Figura 9) e o *codebook* elaborado pelo Claude AI

Figura 9 – Organização dos documentos no Taguette



Fonte: Print da organização no Taguette (2026)

A organização dos dados foi realizada dessa forma com o objetivo de facilitar a análise na ferramenta.

Etapa 5 : Nesta etapa, foi realizada a codificação de cada documento com base no codebook previamente elaborado. Durante esse processo, foi analisada a natureza de cada pergunta, assegurando que as respostas fossem interpretadas de forma adequada e reduzindo o risco de classificações incorretas decorrentes de ambiguidades ou má interpretação.

Etapa 6 : Após a análise dos dados, os resultados foram exportados, como um novo caderno de códigos com a contagem de quantas vezes cada código foi identificado nas respostas dos participantes, durante a análise.

Etapa 7 : Os dados exportados foram posteriormente importados para a ferramenta Flourish Studio, na qual foi elaborada uma visualização do tipo nuvem de bolhas. Essa representação gráfica permitiu evidenciar a frequência e a distribuição das categorias identificadas durante a codificação, contribuindo para uma análise mais clara e intuitiva dos resultados.

As etapas descritas proporcionaram uma avaliação sistemática e criteriosa das respostas coletadas. No entanto, devido ao número reduzido de dados obtidos, optou-se por apresentar os resultados de forma agregada, de modo a evidenciar as principais percepções dos participantes em relação ao experimento e à metodologia proposta.

6.2.2 Resultados da análise qualitativa

Após a conclusão da fase da limpeza dos dados, foi iniciada a interpretação da Figura 10, que apresenta a frequência de ocorrência dos códigos atribuídos às respostas qualitativas dos participantes. O tamanho das bolhas representa a quantidade de trechos codificados em cada categoria, permitindo identificar os temas mais recorrentes durante a análise dos feedbacks.

Figura 10 – Resultado da análise qualitativa



Fonte: Elaborado no Flourish Studio (2026).

De modo geral, os resultados evidenciam uma percepção predominantemente positiva em relação à abordagem proposta. O código de maior frequência foi **IA facilita a compreensão do requisito**, com 12 ocorrências, indicando que a principal contribuição percebida pelos participantes foi o auxílio da inteligência artificial na interpretação e compreensão dos requisitos. Esse resultado sugere que a IA favoreceu uma melhor assimilação das informações, reduzindo dificuldades relacionadas à compreensão do problema e contribuindo para a elaboração dos artefatos.

O segundo código mais frequente foi **Aplicabilidade prática da abordagem** (8 ocorrências), demonstrando que os participantes reconheceram potencial para utilização da integração do método 5W2H integrada à Inteligência Artificial para auxiliar nos processos iniciais da Engenharia de Requisitos, em contextos reais de desenvolvimento de software. Esse achado indica que a abordagem foi percebida como útil não apenas no ambiente experimental, mas também em atividades práticas de Engenharia de Requisitos.

Outro aspecto amplamente citado, com 7 ocorrências identificadas, foi **IA auxilia na geração de ideias**, evidenciando que a inteligência artificial foi utilizada como ferramenta de apoio à criatividade, auxiliando na identificação de requisitos, levantamento de possibilidades e exploração de soluções durante a atividade.

Com 5 ocorrências, também se destacam os códigos **Motivação e engajamento** e **Contribuição para a elicitação de requisitos**, sugerindo que a utilização da IA tornou a atividade mais atrativa e colaborou para o processo de levantamento das necessidades do sistema.

Na etapa de especificação, observou-se o código **Contribuição para a especificação de requisitos** (4 ocorrências), indicando que parte dos participantes percebeu melhorias na documentação dos requisitos produzidos. Esse resultado é reforçado pelos códigos **Facilita a estruturação do raciocínio** (4 ocorrências) e **Carga cognitiva baixa** (4 ocorrências), que demonstram que a abordagem contribuiu para organizar o pensamento e reduzir o esforço cognitivo necessário para execução da tarefa.

Outros benefícios identificados foram **IA agiliza o processo** e **Facilita a comunicação da necessidade**, com 4 e 3 ocorrências, respectivamente, indicando ganhos relacionados à produtividade e à comunicação dos requisitos entre os participantes.

Em relação à experiência de uso, o código **Satisfação com a experiência** (3 ocorrências) demonstra que parte dos participantes avaliou positivamente a utilização da abordagem durante o experimento.

Apesar da predominância de percepções favoráveis, alguns participantes

relataram limitações da abordagem. Os códigos **Sem impacto percebido da abordagem** (3 ocorrências) e **Sem impacto percebido da IA** (3 ocorrências) indicam que nem todos identificaram benefícios significativos decorrentes da utilização da técnica ou da inteligência artificial durante a atividade.

Outra limitação frequentemente mencionada foi a **Necessidade de validação humana da IA** (3 ocorrências), evidenciando que os participantes reconhecem que as respostas produzidas pela inteligência artificial devem ser analisadas criticamente antes de serem incorporadas aos artefatos de requisitos.

Além disso, o código **IA gera informação incorreta** (2 ocorrências) demonstra que alguns participantes identificaram inconsistências ou respostas equivocadas produzidas pela ferramenta, reforçando a necessidade de supervisão humana durante sua utilização.

Também foram identificados os códigos **Facilidade geral na atividade** (2 ocorrências), **Sugestão de melhoria para o processo geral** (2 ocorrências), **Aprendizado percebido** (2 ocorrências) e **Dificuldade não relacionada ao experimento** (2 ocorrências), indicando que alguns participantes relataram benefícios adicionais da experiência, bem como oportunidades de aperfeiçoamento da abordagem.

Por fim, alguns temas apareceram de forma pontual, como **Frustração/Insatisfação** (1 ocorrência), **Comparação explícita entre abordagens** (1 ocorrência), **IA sem contexto** (1 ocorrência), **Interpretação** (1 ocorrência) e **Indicação de vieses** (1 ocorrência). Embora pouco frequentes, esses registros evidenciam reflexões específicas sobre o comportamento da inteligência artificial e comparações entre diferentes estratégias de utilização da abordagem.

De maneira geral, a análise qualitativa demonstra que os participantes perceberam a abordagem como útil para apoiar as atividades de Engenharia de Requisitos (ER). Em contrapartida, os resultados também evidenciam que os participantes reconhecem limitações inerentes ao uso da inteligência artificial generativa. Esses achados

reforçam que a IA é percebida como uma ferramenta de apoio na ER, potencializando suas atividades, mas não substituindo sua capacidade de análise crítica e tomada de decisão.

6.3 Discussão dos resultados

Os resultados obtidos permitem responder aos objetivos propostos por este estudo e às questões definidas no modelo GQM. De modo geral, a integração da metodologia 5W2H com a IA Generativa contribuiu para a produção de requisitos de maior qualidade quando comparada à abordagem tradicional de elicitação.

Sob a perspectiva do objetivo geral, verificou-se que a metodologia proposta foi capaz de apoiar o processo de elicitação e especificação de requisitos, produzindo artefatos mais claros, completos, verificáveis, rastreáveis, testáveis e organizados. As análises estatísticas evidenciaram diferenças significativas entre os grupos para a maioria dos critérios avaliados, indicando que a utilização da metodologia integrada exerceu influência positiva sobre a qualidade dos requisitos produzidos.

Em relação aos objetivos específicos, observou-se que os participantes do Grupo B produziram documentos com maior qualidade documental e menor necessidade de correções, além de apresentarem baixa dependência das respostas geradas pela IA. Os resultados também indicaram que a utilização da IA, quando associada ao método 5W2H, favoreceu a fidelidade da transformação das informações, a precisão textual e o detalhamento dos requisitos, corroborando os objetivos definidos no modelo GQM.

Embora os participantes tenham obtido ganhos significativos com a abordagem proposta, todos eram estudantes do segundo semestre do curso de Engenharia de Software. Dessa forma, os resultados refletem o comportamento de profissionais em formação, não sendo possível afirmar que o mesmo impacto seria observado entre analistas de requisitos experientes, cuja bagagem técnica e estratégias de elicitação podem diferir substancialmente.

Sob a perspectiva qualitativa, os relatos dos participantes reforçam os resultados quantitativos ao apontarem que a IA contribuiu para a organização das ideias, redução de ambiguidades e apoio cognitivo durante a atividade. Ao mesmo tempo, também foram identificadas limitações relacionadas à necessidade de validação das respostas produzidas pela IA, evidenciando que sua utilização deve ocorrer como ferramenta de apoio, e não como substituta da análise humana.

Sob a ótica de replicabilidade e ampliação do experimento, que foi conduzido utilizando um cenário de elicitação com complexidade controlada, adequado ao tempo disponível para a atividade experimental, podemos inferir que, embora os resultados indiquem benefícios da integração entre o método 5W2H e a IA Generativa, ainda não há evidências de que esses resultados sejam diretamente reproduzidos em projetos de grande porte, envolvendo múltiplos stakeholders, elevado número de requisitos e regras de negócio mais complexas.

6.3.1 Ameaças à validade

Embora os resultados obtidos tenham sido consistentes, algumas limitações devem ser consideradas durante sua interpretação. A realização do experimento em contexto acadêmico restringe a generalização dos resultados para ambientes industriais. Além disso, o tamanho reduzido da amostra limita o poder de generalização das conclusões, ainda que os testes estatísticos empregados tenham sido adequados para esse cenário.

Outro aspecto relevante refere-se à experiência prévia dos participantes com Engenharia de Requisitos e com ferramentas de IA Generativa. Apesar da distribuição equilibrada entre os grupos e da padronização das instruções, diferenças individuais podem ter influenciado parcialmente o desempenho observado. Adicionalmente, a avaliação dos artefatos envolveu julgamento humano, o que pode introduzir certo grau de subjetividade, embora esse efeito tenha sido reduzido por meio da utilização de critérios

de avaliação fundamentados na literatura e da participação de múltiplos avaliadores.

Além disso, os participantes foram estudantes do segundo semestre, ou seja, sem formação profissional na área. Embora esse perfil seja frequentemente utilizado em experimentos de Engenharia de Software, os resultados não podem ser generalizados diretamente para profissionais experientes. Estudos futuros podem replicar o experimento com analistas de requisitos atuantes na indústria para verificar se os benefícios observados permanecem em diferentes níveis de experiência.

Levando em consideração os riscos envolvendo o uso de IA Generativa, embora os resultados obtidos sejam promissores, a utilização de Modelos de Linguagem de Grande Escala apresenta limitações inerentes à tecnologia. Esses modelos podem produzir respostas plausíveis, porém incorretas ou inconsistentes com o domínio do problema, fenômeno conhecido como alucinação, mesmo que não tendo sido fortemente identificado pelos participantes do experimentos, houveram algumas observações acerca desse fato. Além disso, a qualidade dos artefatos gerados depende diretamente do modelo empregado, de sua versão e da forma como os prompts são elaborados. Como os provedores atualizam continuamente seus modelos, a reprodução exata dos resultados tende a se tornar mais difícil ao longo do tempo. Dessa forma, a abordagem proposta deve ser compreendida como um mecanismo de apoio ao analista de requisitos, não substituindo a análise crítica e a validação humana dos artefatos produzidos.

7 CONCLUSÕES E TRABALHOS FUTUROS

Com base nas evidências obtidas, conclui-se que a metodologia proposta atingiu os objetivos estabelecidos neste estudo. A integração do método 5W2H com a IA Generativa mostrou-se eficaz para apoiar a elicitación e a especificação de requisitos de software, promovendo melhorias na qualidade dos artefatos produzidos e oferecendo suporte ao processo de elaboração dos requisitos.

Os resultados quantitativos e qualitativos convergem ao demonstrar que a utilização da metodologia integrada favoreceu a produção de requisitos mais consistentes e completos, mantendo baixa ocorrência de vieses, reduzida dependência da IA e menor necessidade de correções posteriores. Dessa forma, as evidências obtidas sustentam a viabilidade da proposta como apoio às atividades iniciais da Engenharia de Requisitos, desde que seu uso seja acompanhado de supervisão humana.

Como contribuição para a área de Engenharia de Requisitos ao propor e avaliar empiricamente uma abordagem que integra o método 5W2H com ferramentas de IA para apoiar os processos de elicitación e especificação de requisitos de software. Como contribuição científica, apresenta evidências experimentais sobre os impactos dessa integração na qualidade dos artefatos produzidos, considerando aspectos como completude, clareza, consistência e tempo de execução. Como contribuição prática, disponibiliza uma metodologia de aplicação, modelos de prompts e instrumentos de avaliação que podem ser utilizados e adaptados por pesquisadores, docentes e profissionais interessados em empregar IA generativa como apoio às atividades de Engenharia de Requisitos.

Apesar de suas contribuições, o estudo apresenta algumas limitações que devem ser consideradas na interpretação dos resultados. O experimento foi conduzido com uma amostra por conveniência, composta exclusivamente por estudantes do segundo semestre do curso de Engenharia de Software de uma única instituição de ensino, o que pode limitar a generalização dos resultados para outros contextos, como equipes de

desenvolvimento experientes ou projetos reais. Além disso, a avaliação concentrou-se em um único cenário de elicitação e especificação de requisitos, utilizando o método 5W2H e ferramentas de Inteligência Artificial Generativa específicas, não contemplando outros domínios de aplicação, métodos de elicitação ou modelos de IA. Apesar dessas limitações, o estudo foi planejado de forma a controlar as principais variáveis experimentais e fornecer evidências empíricas sobre a viabilidade e os impactos da abordagem proposta.

A partir dos resultados obtidos neste estudo, identificam-se oportunidades para o aprofundamento e a evolução da abordagem proposta. Embora os resultados indiquem o potencial da integração entre o método 5W2H e a IA no apoio à elicitação e à especificação de requisitos, novas investigações podem ampliar sua aplicação e validar sua efetividade em diferentes contextos. Nesse sentido, destacam-se como possibilidades de trabalhos futuros como: i) aperfeiçoar a matriz 5W2H proposta, incorporando novos elementos que auxiliem na elicitação e na especificação de requisitos e reduzam ambiguidades durante o processo; ii) desenvolver uma ferramenta que integre a matriz 5W2H a modelos de IA, permitindo automatizar parte do processo de elicitação e especificação de requisitos e oferecer suporte aos engenheiros de requisitos e outros profissionais da área e iii) replicar o experimento em ambientes industriais, envolvendo profissionais da área de tecnologia e projetos reais, com o objetivo de avaliar a aplicabilidade, a escalabilidade e os impactos da abordagem em contextos distintos do ambiente acadêmico.

REFERÊNCIAS

- AKHTAR, A.; AKHTAR, S. Requirements elicitation in transition: A review of conventional and contemporary approaches. **ICCK Journal of Software Engineering**, Institute of Central Computation and Knowledge, v. 1, n. 1, p. 32–45, 2025.
- AKRAM, F.; AHMAD, T.; SADIQ, M. Recommendation systems-based software requirements elicitation process—a systematic literature review. **Journal of Engineering and Applied Science**, Springer, v. 71, n. 1, p. 29, 2024.
- ALFLEN, N. C.; PRADO, E. P. V. Técnicas de elicitação de requisitos no desenvolvimento de software: uma revisão sistemática da literatura. **AtoZ: novas práticas em informação e conhecimento**, v. 10, n. 1, p. 39–49, dez. 2021. Disponível em: <https://revistas.ufpr.br/atoz/article/view/77393>.
- ALMEIDA, C.; COPQUE, I.; OLIVEIRA, A.; AROUCA, M.; BARBOSA, A.; FREIRE, S.; MENDONÇA, M.; LEITE, J. C. From elicitation interviews to software requirements: Evaluating llm performance in requirement generation. In: **Anais do Workshop em Engenharia de Requisitos - Proceedings of the 28th Workshop on Requirements Engineering (WER2025)**. [S. l.: s. n.], 2025. ISBN 978-65-01-52831-1. ISSN 2675-0066.
- ASHISH, V. Attention is all you need. **Advances in neural information processing systems**, v. 30, p. I, 2017.
- BARROS, G. L. **Melhoria de um processo de engenharia de requisitos de software**. [S. l.], 2023. Apresentado em 21 de dezembro de 2023; publicado em 21 de maio de 2024. Disponível em: <https://bdm.unb.br/handle/10483/38579>.
- BERRY, D. M.; BUCCHIARONE, A.; GNESI, S.; LAMI, G.; TRENTANNI, G. A new quality model for natural language requirements specifications. In: **Proceedings of the international workshop on requirements engineering: foundation of software quality (REFSQ)**. [S. l.: s. n.], 2006.
- BROWN, T.; MANN, B.; RYDER, N.; SUBBIAH, M.; KAPLAN, J. D.; DHARIWAL, P.; NEELAKANTAN, A.; SHYAM, P.; SASTRY, G.; ASKELL, A. *et al.* Language models are few-shot learners. **Advances in neural information processing systems**, v. 33, p. 1877–1901, 2020.
- CARVALHO, E. A.; GOMES, J. O.; JATOBÁ, A.; SILVA, M. F.; CARVALHO, P. V. R. Software requirements elicitation for complex systems with the functional resonance analysis method (fram). In: SBC. **Simpósio Brasileiro de Sistemas de Informação (SBSI)**. [S. l.], 2021.

CHUNG, L.; NIXON, B. A.; YU, E.; MYLOPOULOS, J. **Non-functional requirements in software engineering**. [S. l.]: Springer Science & Business Media, 2012. v. 5.

COSTA, A. F.; FREITAS, L. I.; CAVALCANTE, D. A.; OLIVEIRA, V. T.; LELLI, V.; SANTOS, I. S.; OLIVEIRA, P. A.; NOGUEIRA, T. P.; ANDRADE, R. M. Especificação de requisitos em um projeto de big data no setor público. In: SBC. **Congresso Ibero-Americano em Engenharia de Software (CibSE)**. [S. l.], 2024. p. 417–420.

DALPIAZ, F.; NIU, N. Requirements engineering in the days of artificial intelligence. **IEEE software**, IEEE, v. 37, n. 4, p. 7–10, 2020.

DevMedia. **Ferramenta 5W2H para levantamento de requisitos e elaboração de user stories**. 2025. Acesso em: dd mmm. aaaa. Disponível em: <https://www.devmedia.com.br/ferramenta-5w2h-para-levantameto-de-requisitos-e-elaboracao-de-user-stories/25603>.

EBERT, C.; LOURIDAS, P. Generative ai for software practitioners. **IEEE Software**, v. 40, p. 30–38, 07 2023.

FONSECA, C.; ZAINA, L.; MARQUES, A. Avaliação de métodos para elicitación e especificação de requisitos de usabilidade com histórias de usuário: Um experimento controlado. In: **Anais do XXXVIII Simpósio Brasileiro de Engenharia de Software**. Porto Alegre, RS, Brasil: SBC, 2024. p. 257–268. ISSN 2833-0633.

FOTROUSI, F.; TAVANTZIS, T. Reqgenie: Gpt-powered conversational-ai for requirements elicitation. In: SPRINGER. **International Conference on Product-Focused Software Process Improvement**. [S. l.], 2024. p. 352–359.

GLINZ, M. On non-functional requirements. In: IEEE. **15th IEEE international requirements engineering conference (RE 2007)**. [S. l.], 2007. p. 21–26.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A.; BENGIO, Y. **Deep learning**. [S. l.]: MIT press Cambridge, 2016. v. 1.

HOY, Z.; XU, M. Agile software requirements engineering challenges-solutions—a conceptual framework from systematic literature review. **Information**, MDPI, v. 14, n. 6, p. 322, 2023.

IEEE Recommended Practice for Software Requirements Specifications. **IEEE Std 830-1998**, p. 1–40, 1998.

IEEE Standard Glossary of Software Engineering Terminology. **IEEE Std 610.12-1990**, p. 1–84, 1990.

ISO/IEC/IEEE International Standard - Systems and software engineering – Life cycle processes – Requirements engineering. **ISO/IEC/IEEE 29148:2018(E)**, p. 1–104, 2018.

JUNIOR, C. F. F. **Administração Moderna**. São Paulo: Saraiva Uni, 2018. ISBN 9788553131389.

KOLAHDOUZ-RAHIMI, S.; LANO, K.; LIN, C. **Requirement Formalisation using Natural Language Processing and Machine Learning: A Systematic Review**. 2023. Disponível em: <https://arxiv.org/abs/2303.13365>.

KOTONYA, G.; SOMMERVILLE, I. **Requirements engineering: processes and techniques**. [S. l.]: Wiley Publishing, 1998.

LAMSWEERDE, A. van. **Requirements Engineering: From System Goals to UML Models to Software Specifications**. Chichester, UK: Wiley, 2009.

MISAKI, L. T.; CRUZ, C. G. R.; SILVA, R. R. Aprimorando a qualidade de software: Explorando requirements smells com inteligência artificial generativa: Uma investigação empírica (outubro de 2023). **South American Development Society Journal**, v. 10, n. 28, p. 20–20, 2024.

NECULA, S.-C.; DUMITRIU, F.; GREAVU-ŞERBAN, V. A systematic literature review on using natural language processing in software requirements engineering. **Electronics**, MDPI, v. 13, n. 11, p. 2055, 2024.

NUSEIBEH, B.; EASTERBROOK, S. Requirements engineering: A roadmap. In: **Proceedings of the Conference on The Future of Software Engineering (FOSE)**. [S. l.]: ACM, 2000. p. 35–46.

OLIVEIRA, D. d. P. R. de. **Planejamento estratégico**. [S. l.]: Editora Atlas SA, 2015.

OpenAI. **GPT-4 Technical Report**. 2023. Disponível em: <https://arxiv.org/abs/2303.08774>.

OZKAYA, I. Application of large language models to software engineering tasks: Opportunities, risks, and implications. **IEEE Software**, IEEE, v. 40, n. 3, p. 4–8, 2023.

PACHECO, G. N.; MARTINS, L. E. G.; SILVA, A. E. A. da; LAVESSON, N.; GORSCHKE, T. Natural language processing in software engineering: A systematic literature review. **Journal of Software Engineering Research and Development**, v. 13, n. 2, p. 13:255 – 13:277, Dec. 2025. Disponível em: <https://journals-sol.sbc.org.br/index.php/jserd/article/view/5097>.

PILONE, A.; MEIRELLES, P.; KON, F.; MAALEJ, W. Multilingual crowd-based requirements engineering using large language models. In: **Anais do XXXVIII Simpósio Brasileiro de Engenharia de Software**. Porto Alegre, RS, Brasil: SBC, 2024. p. 679–685. ISSN 2833-0633. Disponível em: <https://sol.sbc.org.br/index.php/sbes/article/view/30412>.

POHL, K. **Requirements Engineering: Fundamentals, Principles, and Techniques**. 1st. ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016. XVII, 813 p. ISBN 978-3-662-51888-5.

PRESSMAN, R. S.; MAXIM, B. R. **Engenharia de software-9**. [S. l.]: McGraw Hill Brasil, 2021.

PUBLI. **5W2H: o que é e como aplicar no seu planejamento**. 2021. Acesso em: 22 de janeiro de 2026. Disponível em: <https://www.publi.com.br/5w2h-o-que-e-e-como-aplicar-no-seu-planejamento/>.

RAMALHO, C. T.; GARCIA, A.; PEREIRA, J. A.; ASSUNÇÃO, W. K. G.; COUTINHO, D.; BARBOSA, C.; LUCENA, C.; ITO, R. On the use of gpt to reveal common questions in developers' discussions. In: **2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)**. [S. l.: s. n.], 2025. p. 1677–1682.

RONANKI, K.; CABRERO-DANIEL, B.; HORKOFF, J.; BERGER, C. Requirements engineering using generative ai: Prompts and prompting patterns. In: **Generative AI for effective software development**. [S. l.]: Springer, 2024. p. 109–127.

RUSSELL, S. J.; NORVIG, P. **Inteligência Artificial: Uma Abordagem Moderna**. 4. ed. Rio de Janeiro: GEN LTC, 2022. E-book. ISBN 9788595159495. Disponível em: <https://integrada.minhabiblioteca.com.br/#/books/9788595159495>.

SANTANA, T.; KUDO, T.; BULCÃO-NETO, R. Requisitos em ação: Uma arquitetura pedagógica para o ensino de engenharia de requisitos de software. In: **Anais do XXXIV Simpósio Brasileiro de Informática na Educação**. Porto Alegre, RS, Brasil: SBC, 2023. p. 210–221. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbie/article/view/26663>.

SOMMERVILLE, I. **Software Engineering**. 10. ed. Harlow, UK: Pearson, 2015.

SOMMERVILLE, I. **Software Engineering**. 10. ed. [S. l.]: Pearson, 2016.

SOMMERVILLE, I. **Engenharia de Software**. 10. ed. São Paulo: Pearson Universidades, 2019. ISBN 978-8543024974.

SULTAN, M.; MIRANSKY, A. Ordering interrogative questions for effective requirements engineering: The w6h pattern. In: **2015 IEEE Fifth International Workshop on Requirements Patterns (RePa)**. IEEE, 2015. p. 1–8. Disponível em: <http://dx.doi.org/10.1109/RePa.2015.7407731>.

TANG, J.; LU, H.; WU, R.; XU, X.; MA, K.; FANG, C.; GUO, B.; LU, J.; CHEN, Q.; CHEN, Y.-C. Hawk: Learning to understand open-world video anomalies. In: GLOBERSON, A.; MACKEY, L.; BELGRAVE, D.; FAN, A.;

PAQUET, U.; TOMCZAK, J.; ZHANG, C. (Ed.). **Advances in Neural Information Processing Systems**. Curran Associates, Inc., 2024. v. 37, p. 139751–139785.

Disponível em: https://proceedings.neurips.cc/paper_files/paper/2024/file/fca83589e85cb061631b7ebc5db5d6bd-Paper-Conference.pdf.

VAZQUEZ, C. E.; SIMÕES, G. S. **Engenharia de Requisitos: software orientado ao negócio**. [S. l.]: Brasport, 2016.

VOGELSANG, A. Prompting the future: Integrating generative llms and requirements engineering. In: **REFSQ Workshops**. [S. l.: s. n.], 2024.

WIEGERS, K.; BEATTY, J. **Software requirements**. [S. l.]: Pearson Education, 2013.

WOHLIN, C.; RUNESON, P.; HÖST, M.; OHLSSON, M. C.; REGNELL, B.; WESSLÉN, A. *et al.* **Experimentation in software engineering**. [S. l.]: Springer, 2012. v. 236.

ZADENOORI, M. A.; DaBROWSKI, J.; ALHOSHAN, W.; ZHAO, L.; FERRARI, A. **Large Language Models (LLMs) for Requirements Engineering (RE): A Systematic Literature Review**. 2025. Disponível em: <https://arxiv.org/abs/2509.11446>.

ZHAO, L.; ALHOSHAN, W.; FERRARI, A.; LETSHOLO, K. J.; AJAGBE, M. A.; CHIOASCA, E.-V.; BATISTA-NAVARRO, R. T. Natural language processing for requirements engineering: A systematic mapping study. **ACM Comput. Surv.**, Association for Computing Machinery, New York, NY, USA, v. 54, n. 3, abr. 2021. ISSN 0360-0300. Disponível em: <https://doi.org/10.1145/3444689>.

APÊNDICE A – *MODELO GQM DO ESTUDO*

Objetivo GQM

Goal
Objetivo Geral Analisar , com a perspectiva de uma pesquisadora em um contexto acadêmico, em que medida a metodologia proposta, que integra o método 5W2H com IA generativa, melhora a qualidade da elicitación e documentação de requisitos de software, considerando requisitos funcionais e não funcionais e avaliando sua clareza, completude, correção, verificabilidade, testabilidade e rastreabilidade, quando comparada à elicitación tradicional.

Questions	
1. Qualidade dos requisitos gerados	
ID	Questões
Q1.1	Em que medida os requisitos gerados pela metodologia 5W2H + IA são mais claros que os gerados pela elicitación tradicional?
Q1.2	A metodologia integrada produz requisitos mais completos (incluindo RF e RNF) do que a abordagem tradicional?
Q1.3	Os requisitos elicitados com 5W2H + IA apresentam maior aderência ao problema e menos interpretações equivocadas do que os produzidos sem IA?
Q1.4	Os requisitos obtidos com a metodologia são mais verificáveis (possíveis de serem validados por RN, por exemplo)?
Q1.5	A metodologia integrada facilita a definição de critérios de aceitação e cenários de teste ?
Q1.6	Em que medida a IA mantém rastreabilidade clara entre cada resposta do 5W2H e os requisitos gerados a partir delas?
2. Qualidade da documentação dos requisitos	

ID	Questões
Q2.1	A integração do 5W2H com IA melhora a organização e estruturação da documentação de requisitos?
Q2.2	A documentação gerada é mais padronizada e consistente em comparação à produzida tradicionalmente?
Q2.3	Em que medida os requisitos gerados pela IA, a partir das respostas do 5W2H, apresentam menor ambiguidade, redundância ou contradições quando comparados aos criados manualmente?
3. Comparação entre métodos (tradicional vs. 5W2H + IA)	
ID	Questões
Q3.1	A metodologia integrada permite elicitar requisitos mais rapidamente do que a abordagem tradicional?
Q3.2	A qualidade geral dos requisitos elicitados pela metodologia proposta supera a da abordagem tradicional?
Q3.3	Há redução de erros, omissões ou requisitos ambíguos quando se utiliza a IA com o 5W2H?
Q3.4	Participantes percebem maior facilidade no processo quando utilizam a metodologia 5W2H + IA?
4. Impacto da IA generativa no processo de elicitação	
	Questões
Q4.1	Em que medida a IA interpreta corretamente cada resposta do 5W2H e transforma essas informações em requisitos estruturados, coerentes e alinhados aos atributos de qualidade avaliados?
Q4.2	Em que medida a transformação feita pela IA, das respostas 5W2H para requisitos, reduz ambiguidades e melhora a precisão do conteúdo gerado?
Q4.3	Em que medida a IA é capaz de identificar requisitos adicionais ou implícitos a partir das respostas fornecidas no 5W2H, preservando coerência e alinhamento ao contexto?
Q4.4	Em que medida a IA gera requisitos mais detalhados ou variantes úteis, baseados exclusivamente nas respostas do 5W2H, sem distorcer o contexto fornecido?

ID	Questões
Q5.1	As perguntas estruturadas do 5W2H guiam os participantes a fornecerem informações mais relevantes para a geração de requisitos?
Q5.2	As respostas estruturadas no 5W2H fornecem à IA dados suficientes para gerar requisitos completos (RF e RNF), sem lacunas significativas?
Q5.3	Em que medida o formato estruturado das respostas (5W2H) facilita o processo de geração de requisitos pela IA, resultando em maior clareza, completude e testabilidade?
6. Usabilidade e experiência dos participantes	
ID	Questões
Q6.1	Os participantes consideram a metodologia fácil ou difícil de aplicar?
Q6.2	O processo integrado aumenta a confiança dos participantes quanto à qualidade dos requisitos gerados?
Q6.3	Os usuários percebem maior suporte cognitivo ao utilizar IA durante a elicitación?
7. Riscos, limitações e confiabilidade	
ID	Questões
Q7.1	Em que medida a metodologia apresenta riscos de viés da IA na formulação ou interpretação dos requisitos?
Q7.2	A supervisão humana é suficiente para mitigar erros provenientes da IA?
Q7.3	O uso da IA compromete a originalidade ou independência da interpretação dos participantes?
Q7.4	Foram necessários fazer muitos ajustes nos requisitos gerados pela IA?

Obs.: essas questões podem acabar mudando, dependendo do design do nosso experimento.

Metrics		
ID.Q	ID	Métrica
Q1	M1.1	Métrica: Grau de clareza dos requisitos. Indicador: Percentual de requisitos considerados claros por avaliadores independentes. Forma de coleta: Avaliação em escala Likert (1–5). Escala: Likert 1–5.
	M1.2	Métrica: Índice de completude. Indicador: Proporção de requisitos essenciais presentes no conjunto. Forma de coleta: Checklist de requisitos esperados. Escala: Percentual (%).
	M1.3	Métrica: Aderência ao contexto do problema. Indicador: Número de requisitos incorretos ou mal interpretados. Forma de coleta: Julgamento de especialistas. Escala: Contagem + Likert (1–5).
	M1.4	Métrica: Grau de verificabilidade. Indicador: Percentual de requisitos com critérios verificáveis definidos. Forma de coleta: Checklist de verificabilidade. Escala: Percentual (%).
	M1.5	Métrica: Facilidade de derivar critérios de aceitação. Indicador: Quantidade e qualidade dos critérios de teste derivados. Forma de coleta: Avaliação de testadores. Escala: Contagem + Likert (1–5).
	M1.6	Métrica: Força da rastreabilidade. Indicador: Proporção de requisitos que podem ser ligados diretamente a uma pergunta específica do 5W2H. Forma de coleta: Matriz de rastreabilidade. Escala: Percentual (%).
Q2	M2.1	Métrica: Índice de organização documental. Indicador: Avaliação da estrutura (seções, padronização, lógica). Forma de coleta: Instrumento Likert aplicado a avaliadores. Escala: Likert (1–5).
	M2.2	Métrica: Grau de padronização documental.

		Indicador: Número de inconsistências de formatação, estilo e estrutura. Forma de coleta: Checklist de consistência. Escala: Contagem.
	M2.3	Métrica: Taxa de ambiguidades e redundâncias. Indicador: Número de ambiguidades, contradições ou redundâncias detectadas. Forma de coleta: Avaliação de especialistas. Escala: Contagem + Likert (1–5).
Q3	M3.1	Métrica: Tempo total para eliciação. Indicador: Tempo em minutos para completar a atividade. Forma de coleta: Cronometragem. Escala: Tempo (minutos).
	M3.2	Métrica: Índice composto de qualidade (média de clareza, completude etc.). Indicador: Média ponderada das avaliações. Forma de coleta: Instrumento Likert multiatributo. Escala: Likert (1–5).
	M3.3	Métrica: Taxa de erros. Indicador: Número de erros, omissões ou ambiguidades por pacote de requisitos. Forma de coleta: Avaliação de especialistas. Escala: Contagem.
	M3.4	Métrica: Percepção de facilidade. Indicador: Avaliação subjetiva do participante. Forma de coleta: Questionário pós-atividade. Escala: Likert (1–5).
Q4	M4.1	Métrica: Acurácia da transformação 5W2H → requisito. Indicador: Percentual de requisitos gerados corretamente a partir das respostas. Forma de coleta: Comparação com gabarito do especialista. Escala: Percentual (%).
	M4.2	Métrica: Taxa de precisão textual. Indicador: Redução percentual de ambiguidades identificadas. Forma de coleta: Avaliação comparativa tradicional x IA. Escala: Percentual (%).
	M4.3	Métrica: Cobertura de requisitos emergentes.

		Indicador: Número de requisitos adicionais relevantes gerados. Forma de coleta: Julgamento de especialistas. Escala: Contagem.
	M4.4	Métrica: Grau de detalhamento dos requisitos. Indicador: Média de detalhes adicionados (descrições, condições, pós-condições). Forma de coleta: Avaliação qualitativa. Escala: Likert (1–5).
Q5	M5.1	Métrica: Relevância da informação capturada. Indicador: Percentual de respostas 5W2H julgadas úteis para geração de requisitos. Forma de coleta: Avaliação de especialistas. Escala: Percentual (%).
	M5.2	Métrica: Cobertura informacional do 5W2H. Indicador: Existência ou não de lacunas informacionais. Forma de coleta: Checklist de completude. Escala: Contagem de lacunas.
	M5.3	Métrica: Impacto da estrutura 5W2H na qualidade final. Indicador: Aumento percentual em clareza, completude e testabilidade quando a IA usa o 5W2H. Forma de coleta: Comparação de conjuntos com e sem 5W2H. Escala: Percentual (%).
Q6	M6.1	Métrica: Grau de usabilidade percebida. Indicador: Avaliação do participante. Forma de coleta: Likert. Escala: Likert (1–5).
	M6.2	Métrica: Confiança subjetiva na qualidade dos requisitos. Indicador: Nível de concordância do participante. Forma de coleta: Likert. Escala: Likert (1–5).
	M6.3	Métrica: Percepção de suporte dado pela IA. Indicador: Grau em que a IA facilita o raciocínio e organização das ideias. Forma de coleta: Likert. Escala: Likert (1–5).
Q7	M7.1	Métrica: Incidência de vieses. Indicador: Número de requisitos enviesados ou desalinhados.

		Forma de coleta: Avaliação de especialistas. Escala: Contagem.
	M7.2	Métrica: Correções necessárias atribuídas a falhas da IA. Indicador: Número de correções aplicadas pelo avaliador humano. Forma de coleta: Registro de edição. Escala: Contagem.
	M7.3	Métrica: Grau de dependência da IA. Indicador: Diferença entre respostas iniciais humanas e requisitos finais gerados. Forma de coleta: Comparação manual. Escala: Likert (1–5).

1.

APÊNDICE B – FORMULÁRIO - PERFIL DO PARTICIPANTE

Perfil do Participante - Engenharia de Requisitos com 5W2H + IA

Termo de Consentimento Livre e Esclarecido

Você está sendo convidado(a) a participar de uma pesquisa, a fim de coletar dados pertinentes para o desenvolvimento de um Trabalho de Conclusão de Curso (TCC) e um artigo científico. A seguir, há informações correlatas ao trabalho e à coleta de seus dados pessoais, após a ciência sobre as informações apresentadas e o aceite em contribuir com a pesquisa, confirme os dados solicitados ao final desta página e prossiga. Em caso de recusa, você não será penalizado(a) de forma alguma e o formulário se encerrará.

A seguir, informações importantes relacionadas ao experimento:

Título da pesquisa:

ELICITAÇÃO APRIMORADA DE REQUISITOS DE SOFTWARE POR MEIO DA INTEGRAÇÃO DO MÉTODO 5W2H COM IA GENERATIVA

Objetivo: O experimento tem como objetivo avaliar a eficácia da metodologia proposta que integra o método 5W2H com Inteligência Artificial Generativa no processo de elicitação e documentação de requisitos de software, comparando-a com a abordagem tradicional. O objetivo segue o paradigma GQM definido no estudo.

Pesquisadora responsável: Andreza Honório Magalhães

1.Descrição do Questionário:

Este questionário online tem como objetivo coletar dados, que auxiliem na análise estatística do experimento sobre o uso de combinado do método 5W2H com IA para elicitação e especificação de requisitos de software,

O questionário apresenta um total de 6 perguntas: todas objetivas

Sobre a pesquisa:

Trata-se de uma pesquisa para a escrita de um TCC e um artigo científico, onde será analisado se existe uma maior eficiência na escrita de requisitos utilizando o método estruturado 5W2H com IA e quanto isso impacta na qualidade dos requisitos.

2.Coleta de Dados:

Entendo que as informações fornecidas por mim neste questionário serão coletadas e utilizadas exclusivamente para os fins descritos no item 1. As informações coletadas podem incluir, mas não se limitam a: Respostas às perguntas do questionário e Dados demográficos, como idade e gênero.

3.Anonimato e Confidencialidade:

Tenho a garantia de que minhas **respostas** serão tratadas de forma **anônima e confidencial**, e que nenhum dado pessoal identificável será divulgado.

4. Tempo de Participação:

Compreendo que a minha participação neste questionário é voluntária e que posso interrompê-la a qualquer momento sem ocorrência de penalidades.

5. Uso dos Resultados:

Estou ciente de que os resultados deste questionário poderão ser utilizados para análises, pesquisas ou divulgações acadêmicas e institucionais, mas sempre de forma agregada e anônima.

6. Contato e Dúvidas:

Qualquer dúvida relacionada a este questionário ou à minha participação nele pode ser encaminhada para o e-mail: andrezahm@alu.ufc.br.

7. Consentimento Voluntário:

Declaro que este consentimento é concedido de forma voluntária e que não estou sendo coagido a participar do questionário.

Ademais, é importante ressaltar que:

- Não haverá nenhum tipo de custo a ser cobrado por você participar, bem como nada será pago por sua participação no estudo;
- A participação nesta pesquisa não apresenta riscos conhecidos à saúde física e ao bem estar dos voluntários, ademais, o formulário estará livre de perguntas invasivas e desconfortáveis. Avaliando todas possibilidades, e riscos a integridade dos participantes, é possível que casos específicos, por ser uma tarefa realizada em uma plataforma virtual, você poderá sentir algum desconforto característico o uso de ambiente virtuais como cansaço dos olhos e instabilidade da conexão;
- Para um melhor e mais rápido resultado, é aconselhado que você esteja em um local com um bom acesso a internet e em ambiente agradável para evitar distrações.

* Indica uma pergunta obrigatória

1. Após a leitura do **Termo de Consentimento Livre e Esclarecido**, você **aceita** participar voluntariamente dessa pesquisa? *

Marcar apenas uma oval.

☐ Sim

☐ Não

Seção 1 - Perfil Acadêmico (Contextualização)

Essa seção apresenta 3 perguntas

2. Informe a sua **matrícula:** *

3. Informe seu **curso de graduação:** *

Marcar apenas uma oval.

☐ Engenharia de Software

☐ Ciência da Computação

4. Informe o seu **semestre atual:** *

 Dropdown

Marcar apenas uma oval.

☐ 2º Semestre

☐ 3º Semestre

☐ 4º Semestre

☐ 5º Semestre

☐ 6º Semestre

☐ 7º Semestre

☐ 8º Semestre

☐ 9º Semestre

☐ 10º Semestre

☐ +10º Semestre

Seção 2 - Mapeamento do Conhecimento Prévio

Nessa seção mapearei seus conhecimentos prévios em engenharia de requisitos, ia generativa e no método 5W2H.

Iniciarei mapeando seus conhecimentos sobre **Engenharia de Requisitos**

Essa subseção apresenta 3 perguntas

5. Informe o seu nível de conhecimento sobre **tipos de requisitos de software**: *

Classifique de 0 a 5, onde 0 é "nenhum conhecimento - ainda estou aprendendo" e 5 é "conhecimento elevado - já apliquei em projetos/indústria".

Lembrando que temos: Requisitos Funcionais, Requisitos Não-funcionais e Regras de negócio ou Requisitos de Domínio

Marcar apenas uma oval.

1	2	3	4	5	
nenh	☐	☐	☐	☐	conhecimento elevado - "já apliquei em projetos/indústria"

6. Como você avalia seu **conhecimento teórico/prático em Elicitação e Engenharia de Requisitos**? *

Classifique de 0 a 5, onde 0 é "nenhum conhecimento - ainda estou aprendendo" e 5 é "conhecimento elevado - já apliquei em projetos/indústria".

Marcar apenas uma oval.

1	2	3	4	5	
Nun	☐	☐	☐	☐	Trabalho com conhecimento teórico/prático em elicitação e engenharia de requisitos

7. Como você avalia o seu **desempenho elicitando requisitos de software**? *

No que diz respeito à sua dificuldade e à qualidade dos requisitos

Marcar apenas uma oval.

1 2 3 4 5

Não ☐ ☐ ☐ ☐ ☐ Sinto-me excelente elicitando requisitos de software

Continuando, a partir de agora verificarei sua **familiaridade com IA generativa**

Essa subseção apresenta 3 perguntas

8. Informe a sua **familiaridade com qualquer IA generativa** (ex.: ChatGPT, Gemini, Claude e etcc): *

Classifique de 0 a 5, onde 0 é "nenhum conhecimento - ainda estou aprendendo" e 5 é "conhecimento elevado - já apliquei em projetos/indústria".

Marcar apenas uma oval.

1 2 3 4 5

nenhuma familiaridade ☐ ☐ ☐ ☐ ☐ conhecimento elevado - "aplico muito no meu dia-a-dia"

9. **Com que frequência você utiliza ferramentas de IA Generativa** (ex: ChatGPT, DeepSeek, Claude, Copilot) no seu dia a dia acadêmico ou profissional? *

Classifique de 0 a 5, onde 0 é "nunca utilizo IA no meu dia a dia acadêmico ou profissional" e 5 é "Utilizo diariamente IA no meu dia a dia acadêmico ou profissional".

Marcar apenas uma oval.

1 2 3 4 5

Nunca ☐ ☐ ☐ ☐ ☐ Utilizo diariamente no meu dia a dia acadêmico ou profissional

10. Para quais finalidades você costuma usar IA Generativa? *

Marque todas que se aplicam.

- ☐ Estudo e esclarecimento de dúvidas conceituais.
- ☐ Geração ou correção de código-fonte (programação).
- ☐ Geração de criativos (ex.: imagem e vídeo)
- ☐ Escrita de textos, relatórios ou documentação.
- ☐ Não utilizo ferramentas de IA.
- ☐ Outro: _____

11. Informe seu nível de **familiaridade com o uso do DeepSeek** *

Classifique de 0 a 5, onde 0 é "nenhum conhecimento - ainda estou aprendendo" e 5 é "conhecimento elevado - aplico muito no meu dia-a-dia"

Marcar apenas uma oval.

	1	2	3	4	5	
nenhuma familiaridade	☐	☐	☐	☐	☐	conhecimento elevado - "aplico muito no meu dia-a-dia"

12. Você já utilizou o IA generativa para elicitar requisitos de software? *

Marcar apenas uma oval.

- ☐ Sim
- ☐ Não

Por fim, verificarei sua familiaridade como **método 5W2H**

Essa subseção apresenta 2 perguntas

13. **Você já conhecia ou utilizou a metodologia 5W2H** antes deste experimento?



Marcar apenas uma oval.

- ☐ Não, nunca ouvi falar.
- ☐ Já ouvi falar, mas nunca utilizei.
- ☐ Sim, já utilizei em contextos acadêmicos, profissionais ou pessoais.

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

APÊNDICE C – FORMULÁRIO - FEEDBACK DO PARTICIPANTE

Feedback do Participante - Engenharia de Requisitos com 5W2H + IA

Termo de Consentimento Livre e Esclarecido

Você está sendo convidado(a) a participar de uma pesquisa, a fim de coletar dados pertinentes para o desenvolvimento de um Trabalho de Conclusão de Curso (TCC) e um artigo científico. A seguir, há informações correlatas ao trabalho e à coleta de seus dados pessoais, após a ciência sobre as informações apresentadas e o aceite em contribuir com a pesquisa, confirme os dados solicitados ao final desta página e prossiga. Em caso de recusa, você não será penalizado(a) de forma alguma e o formulário se encerrará.

A seguir, informações importantes relacionadas ao experimento:

Título da pesquisa:

ELICITAÇÃO APRIMORADA DE REQUISITOS DE SOFTWARE POR MEIO DA INTEGRAÇÃO DO MÉTODO 5W2H COM IA GENERATIVA

Objetivo: O experimento tem como objetivo avaliar a eficácia da metodologia proposta que integra o método 5W2H com Inteligência Artificial Generativa no processo de elicitación e documentação de requisitos de software, comparando-a com a abordagem tradicional. O objetivo segue o paradigma GQM definido no estudo.

Pesquisadora responsável: Andreza Honório Magalhães

1.Descrição do Questionário:

Este questionário online tem como objetivo coletar dados, que auxiliem na análise estatística do experimento sobre o uso de combinado do método 5W2H com IA para elicitación e especificação de requisitos de software,

O questionário apresenta um total de 6 perguntas: todas objetivas

Sobre a pesquisa:

Trata-se de uma pesquisa para a escrita de um TCC e um artigo científico, onde será analisado se existe uma maior eficiência na escrita de requisitos utilizando a o método estruturado 5W2H com IA e quanto isso impacta na qualidade dos requisitos.

2.Coleta de Dados:

Entendo que as informações fornecidas por mim neste questionário serão coletadas e utilizadas exclusivamente para os fins descritos no item 1. As informações coletadas podem incluir, mas não se limitam a: Respostas às perguntas do questionário e Dados demográficos, como idade e gênero.

3. Anonimato e Confidencialidade:

Tenho a garantia de que minhas **respostas** serão tratadas de forma **anônima e confidencial**, e que nenhum dado pessoal identificável será divulgado.

4. Tempo de Participação:

Compreendo que a minha participação neste questionário é voluntária e que posso interrompê-la a qualquer momento sem ocorrência de penalidades.

5. Uso dos Resultados:

Estou ciente de que os resultados deste questionário poderão ser utilizados para análises, pesquisas ou divulgações acadêmicas e institucionais, mas sempre de forma agregada e anônima.

6. Contato e Dúvidas:

Qualquer dúvida relacionada a este questionário ou à minha participação nele pode ser encaminhada para o e-mail: andrezahm@alu.ufc.br.

7. Consentimento Voluntário:

Declaro que este consentimento é concedido de forma voluntária e que não estou sendo coagido a participar do questionário.

Ademais, é importante ressaltar que:

- Não haverá nenhum tipo de custo a ser cobrado por você participar, bem como nada será pago por sua participação no estudo;
- A participação nesta pesquisa não apresenta riscos conhecidos à saúde física e ao bem estar dos voluntários, ademais, o formulário estará livre de perguntas invasivas e desconfortáveis. Avaliando todas possibilidades, e riscos a integridade dos participantes, é possível que casos específicos, por ser uma tarefa realizada em uma plataforma virtual, você poderá sentir algum desconforto característico o uso de ambiente virtuais como cansaço dos olhos e instabilidade da conexão;
- Para um melhor e mais rápido resultado, é aconselhado que você esteja em um local com um bom acesso a internet e em ambiente agradável para evitar distrações.

* Indica uma pergunta obrigatória

1. Após a leitura do **Termo de Consentimento Livre e Esclarecido**, você **aceita** participar voluntariamente dessa pesquisa? *

Marcar apenas uma oval.

☐ Sim

☐ Não

Seção 1 - Perguntas gerais sobre o experimento

2. Informe a sua **matrícula**: *

3. Seção 1 – Facilidade de Execução da Atividade *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 - Concordo parcialmente	5 - Concordo totalmente
Conseguí compreender claramente o que deveria ser feito durante a atividade.	☐	☐	☐	☐	☐
As instruções fornecidas foram claras e suficientes para realizar a tarefa.	☐	☐	☐	☐	☐
Conseguí identificar as informações necessárias para elicitar os requisitos.	☐	☐	☐	☐	☐
O processo utilizado durante a atividade foi fácil de seguir.	☐	☐	☐	☐	☐
Conseguí concluir a atividade dentro do tempo disponível.	☐	☐	☐	☐	☐

4. Seção 2 – Percepção sobre a Qualidade dos Requisitos *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos que produzi ficaram claros e fáceis de compreender.	☐	☐	☐	☐	☐
Estou confiante de que os requisitos que produzi representam adequadamente as necessidades do sistema.	☐	☐	☐	☐	☐
Acredito que os requisitos produzidos estão suficientemente completos.	☐	☐	☐	☐	☐
Considereei fácil identificar critérios de aceitação ou formas de validar os requisitos elaborados.	☐	☐	☐	☐	☐
Tenho confiança de que outra pessoa conseguiria compreender os requisitos que produzi	☐	☐	☐	☐	☐

que produzi
que produzi
sem grandes
sem grandes
dificuldades.
dificuldades.

5. Seção 3 – Comparação indireta entre os métodos *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Considereei a atividade simples de executar.	☐	☐	☐	☐	☐
Conseguir avançar na atividade sem grandes dificuldades.	☐	☐	☐	☐	☐

6. Qual foi a **principal dificuldade** encontrada durante a atividade?

7. O que **mais ajudou** você a produzir os requisitos?

8. Você possui alguma **sugestão para melhorar** a atividade ou os materiais fornecidos?

9. Informe a **qual grupo você pertenceu**: *

Marcar apenas uma oval.

☐

Grupo A- Grupo de controle (elicitação tradicional)

Pular para a pergunta 10

☐

Grupo B - Grupo experimental (5W2H + IA)

Pular para a pergunta 13

Seção sobre a Elicitação Tradicional

10. Facilidade do Processo *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
m alguns momentos senti falta de uma estrutura ou guia que me ajudasse a identificar requisitos.	☐	☐	☐	☐	☐
Tive dificuldades para decidir quais informações do cenário eram mais relevantes para transformar em requisitos.	☐	☐	☐	☐	☐

11. Confiança e Qualidade *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Durante a atividade tive dúvidas se todos os critério de qualidade haviam sido identificados.	☐	☐	☐	☐	☐
Acredito que poderia ter produzido requisitos mais completos se tivesse recebido algum apoio adicional.	☐	☐	☐	☐	☐

12. Que tipo de **apoio ou ferramenta** você acredita que poderia facilitar a elicitação de requisitos?

Pular para a pergunta 24

Seção relacionada à aplicação do **método 5W2H + IA**

13. Qual LM você utilizou? *

Marcar apenas uma oval.

☐ ChatGPT

☐ DeepSeek

14. Facilidade percebida *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
O método 5W2H facilitou a identificação das informações necessárias para a elicitação dos requisitos.	☐	☐	☐	☐	☐
A combinação entre 5W2H e IA tornou o processo mais fácil do que eu esperava.	☐	☐	☐	☐	☐
O formato estruturado do 5W2H ajudou a organizar meu raciocínio durante a atividade, para realizar correções no prompt.	☐	☐	☐	☐	☐
As respostas registradas no 5W2H foram	☐	☐	☐	☐	☐

suficientes
suficientes
para gerar
para gerar
os requisitos
os requisitos
necessários
necessários

15. Suporte cognitivo *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 - Concordo parcialmente	5 - Concordo totalmente
A IA interpretou corretamente as informações fornecidas por mim (Cenário, Atores, RN e etc).	☐	☐	☐	☐	☐
Os requisitos gerados pela IA estavam alinhados ao contexto do problema.	☐	☐	☐	☐	☐
A utilização da IA reduziu meu esforço mental durante a atividade.	☐	☐	☐	☐	☐
Os requisitos gerados pela IA apresentaram um nível adequado de detalhamento.	☐	☐	☐	☐	☐
Os requisitos gerados pela IA apresentaram a estruturação	☐	☐	☐	☐	☐

do 5W2H.
estruturação
do 5W2H.

16. Confiança nos resultados *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
O uso da IA aumentou minha confiança na qualidade dos requisitos produzidos.	☐	☐	☐	☐	☐
Considero que os requisitos gerados pela IA estavam alinhados com o contexto apresentado.	☐	☐	☐	☐	☐

17. Viés e confiabilidade *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 - Concordo parcialmente	5 - Concordo totalmente
Em alguns momentos percebi que a IA sugeriu requisitos que não estavam claramente presentes no cenário.	☐	☐	☐	☐	☐
IA apresentou interpretações que poderiam influenciar indevidamente minhas decisões.	☐	☐	☐	☐	☐

18. Supervisão humana *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Conseguir identificar facilmente quando a IA gerou informações inadequadas ou incorretas.	☐	☐	☐	☐	☐
Considero que minhas revisões foram suficientes para corrigir possíveis erros gerados pela IA.	☐	☐	☐	☐	☐

19. Dependência da IA *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Eu teria dificuldade para produzir requisitos com a mesma qualidade sem utilizar a IA.	☐	☐	☐	☐	☐
Em vários momentos aceitei sugestões da IA sem realizar muitas alterações.	☐	☐	☐	☐	☐
A IA influenciou significativamente minha interpretação do problema.	☐	☐	☐	☐	☐

20. Ajustes necessários *

Marcar apenas uma oval por linha.

	1 - Discordo Totalmente	2 - Discordo Parcialmente	3 - Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Foi necessário revisar ou corrigir vários requisitos gerados pela IA.	☐	☐	☐	☐	☐
Os requisitos produzidos pela IA precisaram de muitos refinamentos antes de serem considerados adequados.	☐	☐	☐	☐	☐

21. Cite um exemplo de contribuição positiva da IA durante a atividade.

22. Cite um **exemplo de erro, limitação ou sugestão inadequada** gerada pela IA.

23. **Você utilizaria novamente a combinação 5W2H + IA em uma atividade real de Engenharia de Requisitos? Justifique.** *

Pular para a pergunta 24

Envio dos artefatos - Grupo A

Apenas para garantir!

24. Insira abaixo o link de compartilhamento do seu **documento de requisitos:** *

Envio dos artefatos - Grupo B

Apenas para garantir!

25. Insira abaixo o link de compartilhamento do seu **prompt:** *

26. Insira abaixo o link de compartilhamento do seu **documento de requisitos**: *

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

APÊNDICE D – CENÁRIO DO EXPERIMENTO

Aplicativo de Estudos **StudyConnect**

Descrição do Cenário

A rotina de estudantes universitários e de cursos técnicos envolve atividades individuais e colaborativas, como leitura de materiais, resolução de exercícios, preparação para provas e realização de trabalhos em grupo. Entretanto, muitos estudantes enfrentam dificuldades para organizar seus estudos, acompanhar seu progresso e coordenar atividades com colegas.

Para solucionar esse problema, será desenvolvido o *StudyConnect*, um aplicativo móvel e web que permite o gerenciamento de estudos individuais e em grupo. A plataforma deve oferecer funcionalidades para criação de planos de estudo, acompanhamento de metas, compartilhamento de materiais, agendamento de sessões colaborativas e monitoramento do desempenho acadêmico.

O sistema atende estudantes, monitores e professores, permitindo que cada usuário organize seus conteúdos, participe de grupos de estudo e acompanhe sua evolução ao longo do tempo.

Descrição do Aplicativo

O *StudyConnect* é uma plataforma de apoio à aprendizagem que integra ferramentas de organização pessoal e colaboração acadêmica.

Funcionalidades Principais:

- Estudo Individual
- Criação de planos de estudo.
- Definição de metas semanais e mensais.
- Registro de horas estudadas.
- Lista de tarefas e atividades.
- Biblioteca pessoal de materiais.
- Acompanhamento de desempenho por disciplina.
- Notificações de prazos e compromissos.

Estudo em Grupo:

- Criação e gerenciamento de grupos de estudo.
- Compartilhamento de arquivos e materiais.
- Calendário colaborativo.
- Agendamento de reuniões presenciais ou online.
- Chat entre participantes.
- Distribuição de tarefas.
- Controle de participação dos membros.

Recursos Complementares:

- Dashboard de desempenho.
- Relatórios de progresso.
- Sistema de recomendações de conteúdos.
- Integração com plataformas de videoconferência.
- Sistema de conquistas e gamificação.

Atores do Sistema

<i>Ator</i>	<i>Descrição</i>
Estudante	Usuário principal da plataforma, responsável por gerenciar seus estudos individuais e participar de grupos.
Líder do grupo de estudos	Estudante responsável pela administração de um grupo de estudos.
Professor/Monitor	Usuário que pode criar grupos institucionais, compartilhar materiais e acompanhar o desempenho dos participantes.
Administrador	Responsável pela manutenção da plataforma e gerenciamento geral dos usuários.

Regras de Negócio

RN01 – Cadastro de Usuário

Todo usuário deve possuir um cadastro válido contendo:

- Nome completo;
- E-mail único;
- Senha;
- Instituição de ensino (opcional).

RN02 – Autenticação

Somente usuários autenticados podem acessar funcionalidades de estudo e grupos.

RN03 – Criação de Grupo

Apenas usuários cadastrados podem criar grupos de estudo.

RN04 – Limite de Participantes

Cada grupo de estudo poderá possuir no máximo 30 participantes.

RN05 – Participação em Grupos

Um estudante poderá participar simultaneamente de vários grupos de estudo.

RN06 – Convite para Grupos

O ingresso em grupos privados somente poderá ocorrer mediante convite ou aprovação do líder.

RN07 – Compartilhamento de Materiais

Somente membros do grupo podem visualizar ou compartilhar materiais relacionados ao grupo.

RN08 – Exclusão de Materiais

O autor do material ou o líder do grupo poderá remover materiais compartilhados.

RN09 – Registro de Horas de Estudo

O sistema deverá registrar a quantidade de horas estudadas pelo usuário para geração de relatórios de desempenho.

RN10 – Metas de Estudo

Os usuários poderão definir metas pessoais de estudo com data de início e término.

RN11 – Controle de Tarefas

Toda tarefa criada deverá possuir:

- Título;
- Descrição;
- Data de entrega;
- Responsável;
- Status.

RN12 – Distribuição de Atividades

Em grupos de estudo, cada tarefa deverá possuir pelo menos um responsável.

RN13 – Agendamento de Reuniões

Somente líderes e professores poderão agendar reuniões para o grupo.

RN14 – Notificações

O sistema deverá enviar notificações para:

- Novas tarefas;
- Convites para grupos;

- Alterações em reuniões;
- Prazos próximos do vencimento.

RN15 – Controle de Frequência

A presença dos participantes nas reuniões poderá ser registrada pelo líder ou professor.

RN16 – Relatórios de Desempenho

O sistema deverá gerar relatórios contendo:

- Horas estudadas;
- Metas concluídas;
- Tarefas concluídas;
- Participação em grupos.

RN17 – Gamificação

O usuário receberá medalhas virtuais ao atingir marcos predefinidos, como:

- 10 horas estudadas;
- 30 dias consecutivos de acesso;
- Conclusão de metas semanais.

RN18 – Integridade dos Dados

Nenhum grupo poderá permanecer sem líder. Caso o líder saia do grupo, deverá transferir a liderança para outro participante.

RN19 – Exclusão de Conta

Ao solicitar a exclusão da conta, os dados pessoais do usuário deverão ser removidos ou anonimizados conforme a política de privacidade da plataforma.

RN20 – Disponibilidade dos Relatórios

Os relatórios de desempenho deverão estar acessíveis apenas ao próprio usuário, exceto quando este autorizar o compartilhamento com professores ou monitores.

**APÊNDICE E – *FORMULÁRIO - AVALIAÇÃO DA QUALIDADE DOS
DOCUMENTOS***

Formulário de Avaliação dos Documentos de Requisitos

* Indica uma pergunta obrigatória

1. Pesquisador: *

⌵ Dropdown

Marcar apenas uma oval.

- ☐ Alan
- ☐ Andreza
- ☐ Emilly
- ☐ Henry
- ☐ Ylana

2. Matrícula do participante: *

 Dropdown*Marcar apenas uma oval.*☐ 583270☐ 582175☐ 583401☐ 585854☐ 582779☐ 585964☐ 585750☐ 583066☐ 585658☐ 585368☐ 583396☐ 583436☐ 585561☐ 585941☐ 587959☐ 582740☐ 578101

3. Clareza dos requisitos *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos estão descritos de forma clara e fácil de compreender?	☐	☐	☐	☐	☐
Há termos vagos ou expressões subjetivas que dificultam a interpretação?	☐	☐	☐	☐	☐
Os requisitos permitem que diferentes avaliadores tenham a mesma interpretação?	☐	☐	☐	☐	☐

4. Completude *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
O documento apresenta todos os 5 requisitos funcionais pedidos para a atividade?	☐	☐	☐	☐	☐
O documento apresenta os 5 requisitos não funcionais pedidos para a atividade?	☐	☐	☐	☐	☐
Existem lacunas importantes de informação nos requisitos documentados?	☐	☐	☐	☐	☐

5. Aderência ao problema *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos refletem corretamente as necessidades apresentadas no cenário?	☐	☐	☐	☐	☐
Foram identificados requisitos incorretos ou desalinhados com o contexto?	☐	☐	☐	☐	☐
O conjunto de requisitos representa adequadamente os objetivos do sistema?	☐	☐	☐	☐	☐

6. Verificabilidade *

1. Um requisito pode ser **verificado** quando ele atende as especificações técnicas e **validado** quando ele atende suas expectativas de negócio.
2. Um requisito é mensurável quando apresenta dados/informações que podem ser verificadas. Ex.: "O sistema deve ficar disponível **22 horas por semana**"

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos podem ser objetivamente verificados ou validados?	☐	☐	☐	☐	☐
Os requisitos apresentam condições mensuráveis ou observáveis?	☐	☐	☐	☐	☐
Há requisitos cuja implementação não pode ser claramente avaliada?	☐	☐	☐	☐	☐

7. Critérios de aceitação *

2. Para um requisito fornecer informações suficientes para a elaboração de casos de testes, ele deve apresentar a forma correta de como o usuário deve preencher aquele dado. Ex.: "O usuário deve fornecer a senha com no mín. 8 e no máx. 10 caracteres, sendo aceito: letras, números e caracteres especiais (lista dos caracteres aceitos)"

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Foi possível derivar critérios de aceitação a partir dos requisitos?	☐	☐	☐	☐	☐
Os requisitos fornecem informações suficientes para elaboração de casos de teste?	☐	☐	☐	☐	☐
Os critérios de aceitação derivados são consistentes e completos?	☐	☐	☐	☐	☐

8. Rastreabilidade *

Mesmo que eles não tenha identificado a rastreabilidade, você, como avaliador, deve conseguir identificar.

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
O participante documentou a origem de cada requisito?	☐	☐	☐	☐	☐
É possível identificar a origem de cada requisito nas informações fornecidas?	☐	☐	☐	☐	☐
Os requisitos apresentam rastreabilidade adequada com as informações do cenário?	☐	☐	☐	☐	☐
O documento facilita o acompanhamento da origem dos requisitos?	☐	☐	☐	☐	☐

9. Organização documental *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
O documento apresenta estrutura organizada e lógica?	☐	☐	☐	☐	☐
Os requisitos estão agrupados e apresentados de forma consistente?	☐	☐	☐	☐	☐
A documentação facilita a localização de informações?	☐	☐	☐	☐	☐

10. Padronização *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos seguem um padrão uniforme de escrita?	☐	☐	☐	☐	☐
Há consistência na nomenclatura utilizada ao longo do documento?	☐	☐	☐	☐	☐
O formato dos requisitos foi mantido de maneira padronizada?	☐	☐	☐	☐	☐

11. Ambiguidades e redundâncias *

1. Um requisito é **ambíguo** quando ele é aquele que pode ser interpretado de mais de uma maneira
2. Um requisito é **redundante** quando descreve uma função, mas é encontrado outro requisito que descreve a mesma função, porém de forma diferente
3. Um requisito é **contraditório** quando ele descreve uma função, mas há outro requisito que anula a função descrita anteriormente

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Foram identificadas ambiguidades nos requisitos?	☐	☐	☐	☐	☐
Foram identificados requisitos redundantes?	☐	☐	☐	☐	☐
Foram identificadas contradições entre requisitos?	☐	☐	☐	☐	☐

12. Qualidade Geral *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
O conjunto de requisitos apresenta qualidade satisfatória para utilização em um projeto de software?	☐	☐	☐	☐	☐
O documento fornece informações suficientes para apoiar etapas posteriores do desenvolvimento? (prototipação, desenvolvimento, implementação e testes)	☐	☐	☐	☐	☐
O documento apresenta maturidade adequada para servir como especificação inicial?	☐	☐	☐	☐	☐

13. Erros e omissões *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Foram identificados erros relevantes nos requisitos?	☐	☐	☐	☐	☐
Foram identificadas omissões significativas?	☐	☐	☐	☐	☐
O número de problemas encontrados compromete a utilização do documento?	☐	☐	☐	☐	☐

14. Grupo ao qual o participante pertencia? *

Marcar apenas uma oval.

- ☐ Grupo A
- ☐ Grupo B *Pular para a pergunta 15*

Seção voltada para análise do grupo B

15. Fidelidade da transformação *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos refletem corretamente as informações fornecidas nas respostas do 5W2H?	☐	☐	☐	☐	☐
Houve distorções na interpretação das respostas durante a geração dos requisitos?	☐	☐	☐	☐	☐
Os requisitos preservam o contexto original descrito nas respostas?	☐	☐	☐	☐	☐

16. Precisão textual *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos gerados apresentam linguagem objetiva e precisa?	☐	☐	☐	☐	☐
A transformação reduziu ambiguidades presentes nas informações iniciais?	☐	☐	☐	☐	☐
Os requisitos produzidos apresentam boa qualidade textual?	☐	☐	☐	☐	☐

17. Detalhamento *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos apresentam nível adequado de detalhamento?	☐	☐	☐	☐	☐
Os requisitos incluem informações complementares úteis para implementação?	☐	☐	☐	☐	☐
O detalhamento contribui para aumentar a compreensão do sistema?	☐	☐	☐	☐	☐

18. Relevância das informações capturadas *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
As respostas do 5W2H forneceram informações relevantes para a elaboração dos requisitos?	☐	☐	☐	☐	☐
As informações capturadas contribuíram para aumentar a qualidade dos requisitos?	☐	☐	☐	☐	☐
O conteúdo obtido pelo 5W2H foi suficiente para apoiar a documentação?	☐	☐	☐	☐	☐

19. Cobertura das informações do cenário *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Foram identificadas lacunas de informação nas respostas do 5W2H?	☐	☐	☐	☐	☐
As respostas cobriram adequadamente os principais aspectos do problema?	☐	☐	☐	☐	☐
A ausência de informações comprometeu a geração dos requisitos?	☐	☐	☐	☐	☐

20. Impacto da qualidade final *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
O uso do 5W2H contribuiu para aumentar a clareza dos requisitos?	☐	☐	☐	☐	☐
O uso do 5W2H contribuiu para aumentar a completude dos requisitos?	☐	☐	☐	☐	☐
O uso do 5W2H contribuiu para aumentar a testabilidade dos requisitos?	☐	☐	☐	☐	☐

Abaixo inicia-se uma avaliação mais voltada para o prompt em si

21. Viés *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Foram identificados requisitos enviesados ou desalinhados ao contexto?	☐	☐	☐	☐	☐
A IA introduziu informações não justificadas pelas respostas fornecidas?	☐	☐	☐	☐	☐

22. Correções necessárias *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos finais dependem fortemente do conteúdo gerado pela IA?	☐	☐	☐	☐	☐
O documento final apresenta evidências de contribuição humana significativa?	☐	☐	☐	☐	☐

23. Dependência da IA *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os requisitos finais dependem fortemente do conteúdo gerado pela IA?	☐	☐	☐	☐	☐
O documento final apresenta evidências de contribuição humana significativa?	☐	☐	☐	☐	☐

24. Quantos ajustes foram realizados no prompt ? *

Marcar apenas uma oval.

- ☐ 1 ajuste
- ☐ 2 ajustes
- ☐ 3 ajustes
- ☐ 4 ajustes
- ☐ 5 ajustes
- ☐ entre 6 a 10 ajustes
- ☐ 11 ou mais

25. Ajustes realizados *

Marcar apenas uma oval por linha.

	1 – Discordo totalmente	2 – Discordo parcialmente	3 – Nem concordo nem discordo	4 – Concordo parcialmente	5 – Concordo totalmente
Os ajustes realizados foram essenciais para tornar os requisitos adequados ao contexto?	☐	☐	☐	☐	☐

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários

APÊNDICE F – PROMPT PARA GERAÇÃO DO CODEBOOK

Primeira interação:

"Chat, irei começar a análise qualitativa do meu TCC. a partir das respostas subjetivas dos participantes no formulário de Feedback. Tendo em vista que o tema do meu tcc é: Elicitação e especificação aprimorada de requisitos de software, com a integração do método 5W2H com IA Generativa", quero que crie um codebook de tags e descrição, que abranja os dois grupos do experimento, e possa identificar tanto pontos positivos quanto negativo da experiência dos participantes, além dos pontos de pesquisa do TCC. Quero que gere um codebook único que contemple todas as tags para usá-lo no taguette.

O que eu quero avaliar? A percepção dos participantes sobre a integração do método 5W2H + IA para auxiliar nos processos iniciais da Engenharia de requisitos.

No que eu não quero focar? (faça tags genéricas sobre essas observações)
Em variáveis externas Em dificuldades não relacionadas diretamente ao experimento

No que eu quero focar? Na percepção sobre o auxílio do método 5w2h na elicitação Na percepção sobre o auxílio da IA Na experiência percebida durante a atividade Nas limitações do uso da IA"

Segunda interação:

"Não separe por categorias. O grupo A é o grupo de controle: elicitação tradicional e o B o grupo experimental: 5W2H + IA Generativa. Quero as tags por extenso e não em formato de código.

APÊNDICE G – *LIVRO DE CÓDIGOS - ANÁLISE QUALITATIVA*

Tag	Descrição
Comparação explícita entre abordagens	Participante compara diretamente sua condição (Grupo A ou B) com a outra abordagem.
Frustração ou insatisfação	Expressão de frustração, desconforto ou insatisfação com a experiência.
IA sem contexto ou especificidade suficiente	IA não compreendeu bem o domínio ou a intenção real do participante.
Indicação de viés	Participante não entendeu por que a IA sugeriu determinada coisa.
Interpretação errada do requisito pela IA	IA interpretou mal a intenção do participante ao gerar sugestões.
Aprendizado percebido	Participante relata ter aprendido algo durante a atividade.
Dificuldade não relacionada ao experimento	Dificuldade relatada que não se relaciona à técnica, à IA ou ao processo de requisitos (ex: interface do formulário, problema técnico). Tag genérica, não desdobrar.
Facilidade geral na atividade	Relato geral de que a atividade foi fácil, tranquila ou intuitiva.
IA gera informação incorreta	Relato de que a IA sugeriu algo incorreto, contraditório ou tecnicamente inadequado (inclui alucinações).
Sugestão de melhoria para o processo geral	Sugestão de melhoria para a dinâmica geral da atividade/experimento.
Facilita a comunicação da necessidade	Indica que a abordagem ajudou a expressar/comunicar a necessidade de forma mais clara.
Necessidade de validação humana da IA	Indicação de que a sugestão da IA precisava ser revisada antes de aceitar.
Satisfação com a experiência	Expressão de satisfação ou avaliação positiva geral da experiência.
Sem impacto percebido da IA	Participante menciona a IA mas não atribui a ela impacto positivo nem negativo claro.
Sem impacto percebido da técnica	Participante menciona a técnica mas não atribui a ela impacto positivo nem negativo claro.
Carga cognitiva baixa	Relato de esforço mental diminuído pelo uso do método ou da IA durante a tarefa.
Contribuição para a especificação de requisitos	Trecho relaciona a abordagem com a qualidade/clareza da especificação final do requisito.
Facilita a estruturação do raciocínio	Trecho indica que a técnica utilizada (5W2H, no Grupo B) ajudou a organizar ou guiar o raciocínio durante o levantamento do requisito.
IA agiliza o processo	Relato de que a IA tornou a tarefa mais rápida ou reduziu esforço percebido.
Contribuição para a eliciação de requisitos	Trecho relaciona diretamente a abordagem com a qualidade/facilidade da eliciação.

Motivação e engajamento	Relato de interesse, curiosidade ou engajamento durante a tarefa.
IA auxilia na geração de ideias	IA ajudou a gerar sugestões ou pontos que o participante não havia pensado.
Aplicabilidade prática da abordagem	Participante opina se usaria a abordagem em um contexto real de trabalho/projeto.
IA facilita a compreensão do requisito	IA ajudou o participante a entender melhor o próprio requisito ou a técnica 5W2H.