

Explorando uma abordagem pedagógica mediada por LLMs para o ensino de Engenharia de Requisitos: um relato de experiência

Gabriela H. Ciudrowski¹, Jessica I. Pegorini², Francisco C. M. Souza³, Alinne C. C. Souza³

¹Núcleo de Excelência em Tecnologias Sociais (NEES) – Maceió, AL – Brasil

²Universidade Federal do Paraná – Curitiba, PR – Brasil

³Universidade Tecnológica Federal do Paraná – Dois Vizinhos, PR – Brasil

gabriela.ciudrowski@nees.ufal.br, pegorini@ufpr.br

{franciscosouza, alinnesouza}@utfpr.edu.br

Abstract. *This study investigates the use of Large Language Models as support for teaching Requirements Engineering through an experience report on the application of an LLM-mediated instructional approach in a practical workshop. Requirements and MVPs of a real project, previously generated with the assistance of AI, were evaluated by graduate students with professional experience. The results indicate that this pedagogical mediation fosters the development of critical thinking, quality analysis, and the understanding of AI as a supportive tool, highlighting its potential to strengthen pedagogical practices aligned with professional contexts.*

Resumo. *Este estudo investiga o uso de Large Language Models como apoio ao ensino da Engenharia de Requisitos, por meio de um relato de experiência sobre a aplicação de uma abordagem didática mediada por LLMs em uma oficina prática. Requisitos e MVPs de um projeto real, previamente gerados com auxílio de IA, foram avaliados por estudantes de pós-graduação com experiência profissional e os resultados indicam que essa mediação pedagógica favorece o desenvolvimento do pensamento crítico, a análise da qualidade e a compreensão da IA como ferramenta de apoio, evidenciando seu potencial no fortalecimento de práticas pedagógicas alinhadas ao contexto profissional.*

1. Introdução

A Engenharia de Requisitos (ER) é um dos pilares da Engenharia de Software (ES), responsável por compreender, definir e documentar os requisitos de um sistema. A elicitación e a análise constituem um processo iterativo com feedback contínuo entre atividades, impactando diretamente a qualidade do produto final [Sommerville 2019]. Quando essas atividades são conduzidas de forma inadequada, podem resultar em requisitos incompletos, ambíguos ou inconsistentes, comprometendo a compreensão do problema e a eficácia da solução proposta [Pacheco et al. 2018].

Para enfrentar esses desafios, a adoção da Inteligência Artificial (IA) generativa tem se expandido na ER, sobretudo com a utilização de *Large Language Models (LLMs)*. Esses modelos, fundamentados em *deep learning*, possuem a capacidade de processar

e gerar linguagem natural [Hou et al. 2023], permitindo sua aplicação na automação de tarefas como a geração de código, documentação técnica e outros artefatos essenciais de software. É importante ressaltar que muitos desses desafios não se restringem apenas ao contexto profissional, uma vez que no âmbito educacional, também é possível encontrar desafios acerca do ensino de ER, como a falta de integrações práticas alinhadas às atividades reais do mercado [Macedo et al. 2024]. Além disso, o uso recorrente de IAs no cenário educacional também traz questionamentos a respeito de sua eficiência como ferramenta de apoio educacional e possibilidades de uso [Maia and Sarkis 2025] [Murad et al. 2025].

Este trabalho apresenta um relato de experiência da aplicação de uma abordagem de ensino mediada por *LLMs* em uma oficina de ER. A experiência consistiu na avaliação da qualidade de requisitos e *MVPs* de um projeto real, previamente gerados com auxílio dos modelos *GPT-5* e *Sonnet 4.5*. A avaliação foi realizada por 11 estudantes de pós graduação na área de computação que também atuam no mercado de trabalho, os quais assumiram o papel de avaliadores técnicos dos artefatos gerados e realizaram análises com base em critérios de qualidade pré-definidos. Essa abordagem visa apoiar a reflexão estruturada e o julgamento técnico em ER, em consonância com as práticas do mercado de trabalho.

Os resultados deste estudo permitem analisar não apenas a qualidade dos artefatos produzidos com o apoio de *LLMs*, mas também as evidências pedagógicas decorrentes do processo avaliativo. A análise realizada pelos participantes revelou manifestações de práticas associadas à ER, como a identificação de ambiguidades, a verificação de aderência ao escopo e a revisão técnica dos artefatos. Como contribuições, o estudo: (i) apresenta o uso de *LLMs* como recurso didático no ensino-aprendizagem de ER; (ii) propõe uma abordagem baseada na avaliação de artefatos previamente gerados; (iii) indica que a qualidade da estruturação dos requisitos influencia os resultados produzidos pelos modelos; e (iv) discute implicações pedagógicas do uso crítico de IA no contexto educacional e sua aproximação com práticas do mercado de trabalho.

O restante deste artigo está organizado da seguinte forma: a Seção 2 discute os principais trabalhos relacionados ao tema; a Seção 3 descreve o desenvolvimento da abordagem proposta; a Seção 4 apresenta o processo de avaliação da abordagem; a Seção 5 apresenta os resultados obtidos e as lições aprendidas; e a Seção 6 apresenta a conclusão e os trabalhos futuros.

2. Trabalhos relacionados

A utilização de tecnologias emergentes e a discussão sobre qualidade na ER têm sido exploradas sob diferentes perspectivas, especialmente no ensino e na aproximação entre teoria e prática. O estudo [Macedo et al. 2024] analisa o ensino de ER em cursos de TI no Brasil, identificando desafios como a dificuldade de implementar práticas realistas e integrar teoria às demandas da indústria. De forma complementar, [Portela and Vieira 2025] destaca a importância de abordagens práticas no ensino de Engenharia de Software (ES), evidenciando que projetos aplicados e situações próximas ao contexto profissional favorecem maior engajamento e aprendizagem significativa.

No contexto do uso de IA generativa, [Maia and Sarkis 2025] investigam a aplicação de *LLMs* (ChatGPT e DeepSeek) no ensino de programação Python, apon-

tando ganhos em autonomia, engajamento e compreensão conceitual, além de benefícios na personalização didática, embora ressaltem riscos de dependência e necessidade de mediação docente. De maneira semelhante, [Murad et al. 2025] apresentam evidências experimentais de que o uso do ChatGPT pode melhorar a qualidade inicial de modelos de software, especialmente quanto à organização e completude, ainda que com limitações relacionadas à precisão técnica e ao uso acrítico da ferramenta.

De modo geral, esses estudos indicam que tecnologias emergentes podem apoiar o ensino de ES e ER, seja por meio de metodologias práticas, seja pelo uso de IA como ferramenta de suporte. Contudo, a maioria das pesquisas concentra-se na adoção de tecnologias como apoio instrucional ou na avaliação de seus impactos imediatos nos artefatos produzidos. Diferentemente dessas abordagens, este trabalho emprega artefatos gerados por *LLMs* como objeto de análise crítica, posicionando a IA não apenas como suporte, mas como elemento indutor do desenvolvimento de competências analíticas. Assim, busca-se promover a reflexão estruturada sobre qualidade em ER, estimulando a avaliação, identificação de inconsistências e julgamento técnico, em alinhamento com práticas profissionais da área.

3. Abordagem mediada por *LLMs* para o ensino de Engenharia de Requisitos

Esta seção descreve a abordagem de ensino mediada por *LLMs*. A Figura 1 apresenta as quatro fases estruturantes da abordagem: (i) preparação dos documentos; (ii) geração prévia dos artefatos com *LLMs*; (iii) avaliação pelos participantes; e (iv) análise das evidências pedagógicas observadas a partir da aplicação e avaliação dos participantes.

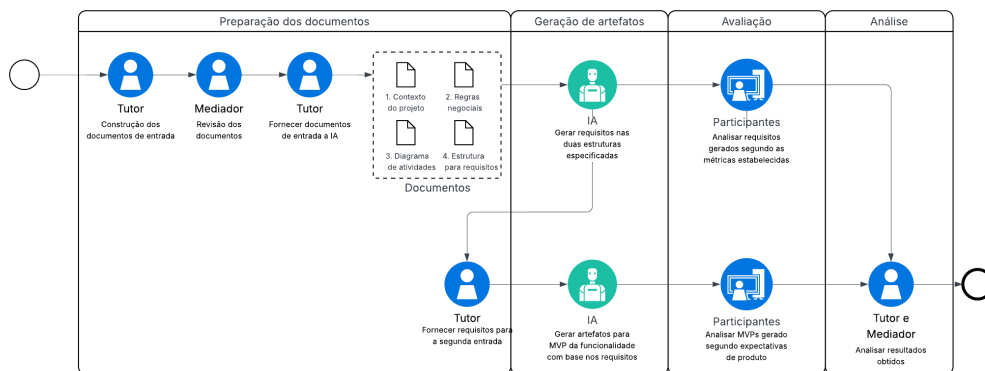


Figura 1. Visão da abordagem mediada por *LLMs*

Fase 1 – Preparação dos documentos: nesta etapa, quatro documentos base ($D = \{D_1, D_2, D_3, D_4\}$) foram elaborados e revisados para subsidiar a geração dos artefatos com *LLMs*: (D_1) o contexto do projeto; (D_2) as regras de negócio; (D_3) um diagrama de atividades (modelado em BPMN¹); e (D_4) duas estruturas distintas de especificação de requisitos ($E = \{E_1, E_2\}$). Por questões éticas e de proteção de dados, os documentos utilizados nesta pesquisa não podem ser divulgados, pois contêm informações confidenciais.

¹Modelagem e notação de processos de negócios: notação gráfica para documentar processos de negócios (<https://www.ibm.com/br-pt/think/topics/bpmn>)

Os três primeiros documentos (D_1 , D_2 e D_3) foram criados a partir de um projeto real utilizado como cenário desta experiência. Os documentos D_1 e D_2 foram criados utilizando *Google Docs* e o D_3 com o auxílio do *Draw.io*². O projeto consiste em uma ferramenta *online* de gestão de processos escolares, desenvolvida para otimizar as atividades administrativas e pedagógicas de redes públicas de ensino. O sistema atende a diversos perfis, como gestores municipais e profissionais de escolas, oferecendo funcionalidades que abrangem desde a matrícula de estudantes e o diário de classe digital até o controle de matrizes curriculares e calendários letivos. Seu propósito central é automatizar rotinas escolares, promovendo a modernização administrativa e a eficiência educacional.

No que tange ao documento D_4 , as estruturas de especificação definidas foram: i) **Estrutura 1 - E_1** : Épico + História de Usuário³ + Critérios de Aceite (BDD)⁴; e ii) **Estrutura 2 - E_2** : Épico + Regra de Negócio + História de Usuário + Descrição de Fluxo + Critérios de Aceite (BDD). Essas estruturas foram fundamentadas em técnicas de especificação reconhecidas no mercado e validadas pela literatura [Pasuksmit et al. 2021] [Frattini and Frattini 2025] [Kuhail and Lauesen 2022]. A definição de duas estruturas distintas visou estimular a análise crítica dos participantes sobre como diferentes níveis de detalhamento influenciam a qualidade dos artefatos gerados por *LLMs*.

Fase 2 – Geração dos artefatos: esta fase foi dividida em duas subfases: ($F2.1$) geração dos requisitos para cada estrutura definida utilizando o modelo *GPT-5*, a partir dos quatro documentos elaborados na Fase 1; e ($F2.2$) geração dos respectivos *MVPs* utilizando o modelo *Sonnet 4.5*, com base nos requisitos gerados na subfase ($F2.1$). Devido à confidencialidade dos dados do projeto, os requisitos e artefatos gerados não podem ser disponibilizados, porém, para fins de replicação, os *prompts* utilizados juntamente com o passo a passo para replicação e geração de artefatos está disponível no link ⁵.

Fase 3 – Avaliação da abordagem: nesta etapa, os participantes realizaram a avaliação da abordagem. A avaliação consistiu na análise técnica dos requisitos e dos *MVPs* gerados, utilizando critérios de qualidade predefinidos. A Tabela 1 detalha os atributos adotados para a avaliação dos requisitos, enquanto que a Tabela 2 consolida os critérios aplicados à análise dos *MVPs*.

Tabela 1. Critérios de qualidade para avaliação dos requisitos gerados pelo modelo *GPT-5*

Critério	Descrição
Corretude	O requisito representa fielmente a necessidade real do usuário ou do sistema.
Não ambiguidade	O requisito está escrito de maneira clara e única, sem permitir múltiplas interpretações.
Compleitude	O requisito contém todas as informações necessárias, sem lacunas que precisem ser inferidas.
Consistência	O requisito está em harmonia com os demais, sem apresentar conflitos entre eles.
Verificabilidade	É possível testar ou inspecionar objetivamente se o requisito foi atendido.

Fase 4 – Análise dos resultados: nesta fase, o mediador e o tutor conduziram uma análise integrada dos dados coletados, consolidando as percepções e avaliações técnicas fornecidas pelos participantes.

²<https://www.drawio.com/>

³Técnica para representar requisitos de forma simples e orientada ao valor para o usuário [Cohn 2005]

⁴*Behavior Driven Development* é uma técnica de desenvolvimento de software orientada ao comportamento do sistema, que busca aproximar as regras de negócio da implementação [Smart 2014].

⁵Link: <https://acesse.one/693p2w9>

Tabela 2. Critérios de qualidade para avaliação dos MVPs gerados pelo modelo Sonnet 4.5

Critério	Descrição
Aderência do produto final	O artefato gerado pela IA apresentou um resultado próximo do esperado.
Correção funcional	O artefato gerado apresentou erros ou falhas funcionais.
Esforço de retrabalho	O artefato gerado necessitaria de correção humana para implementação da funcionalidade no produto final.
Compreensão da IA	O artefato gerado apresenta interpretações equivocadas ou distorcidas pela IA.

4. Processo de avaliação da abordagem

A avaliação da abordagem foi conduzida por meio de uma oficina prática da disciplina de Engenharia de Requisitos de um curso de pós-graduação em Computação, com o objetivo de estimular a análise crítica de artefatos gerados por IA em um contexto próximo ao mercado de trabalho. A oficina, realizada em novembro de 2025 e com duração de três horas, contou com a participação de 11 estudantes com atuação profissional em áreas relacionadas à Engenharia de Software. Empregando os princípios da ABP (Aprendizagem Baseada em Problemas), a atividade foi estruturada como uma situação-problema, na qual os participantes assumiram o papel de avaliadores técnicos de artefatos gerados por IA. O foco da avaliação consistiu na análise crítica e na reflexão sobre a qualidade dos requisitos gerados por *LLMs*, bem como na investigação da influência das estruturas definidas tanto na qualidade desses requisitos quanto na qualidade percebida dos MVPs gerados por *LLMs*. Para alcançar o objetivo da avaliação, são derivadas três seguintes Questões de análise (QA) para serem respondidas, conforme apresentada na Tabela 3.

Tabela 3. Questões de análise para apoiar a avaliação da abordagem

Questões de Análise	Descrição
<i>QA</i> ₁ : Como os participantes avaliam a qualidade técnica de requisitos gerados por <i>LLMs</i> ?	Esta questão investiga a avaliação dos participantes quanto a critérios formais de qualidade, como correteude, clareza, completude, consistência e verificabilidade.
<i>QA</i> ₂ : Como diferentes estruturas de especificação de requisitos influenciam a qualidade percebida dos requisitos gerados por <i>LLMs</i> ?	Esta questão busca analisar como o nível de detalhamento e organização estrutural impacta a avaliação realizada pelos participantes.
<i>QA</i> ₃ : Como a estrutura dos requisitos influencia a qualidade percebida dos MVPs gerados por <i>LLMs</i> ?	Esta questão examina a relação entre estruturação dos requisitos e critérios como aderência, funcionalidade e necessidade de retrabalho.
<i>QA</i> ₄ : Como o processo de avaliação de artefatos gerados por <i>LLMs</i> contribui para o desenvolvimento do pensamento crítico em Engenharia de Requisitos?	Esta questão investiga de que forma a identificação de limitações, inconsistências e ambiguidades nos artefatos evidencia competências analíticas e reflexivas dos participantes.

4.1. Preparação dos instrumentos de coleta

Para avaliar a percepção dos participantes sobre a qualidade dos requisitos e dos MVPs gerados pelas *LLMs*, foram elaborados dois questionários por meio da ferramenta online *Google Forms*. Os instrumentos foram classificados como semiestruturados e compostos por questões fechadas e abertas. Os questionários incluíram: (i) questões de múltipla escolha para caracterização do perfil dos participantes; (ii) questões em escala de concordância e satisfação relacionadas à qualidade dos requisitos gerados; (iii) questões em escala voltadas à avaliação da qualidade percebida dos MVPs; e (iv) questões abertas para coleta das percepções dos participantes acerca de limitações, benefícios e do uso de *LLMs*.

4.2. Condução da oficina

A condução na oficina foi realizada em um encontro síncrono por meio da plataforma de videoconferência *Google Meet*. O processo de condução da oficina foi dividida em quatro etapas: **1. Explicação da oficina:** foi apresentado o objetivo da oficina e da atividade, bem como dos artefatos gerados e dos instrumentos de coletas; **2. Disponibilização dos formulários:** os formulários foram disponibilizados eletronicamente via *WhatsApp* aos participantes; **3. Execução da atividade:** o mediador e o tutor ficaram online durante a execução da atividade na qual os participantes avaliaram a abordagem por meio dos formulários disponibilizados; **4. Coleta e análise dos dados:** após a execução da atividade, os dados foram coletados e em seguida, foi realizada uma análise quantitativa utilizando estatística descritiva para caracterização das respostas dos estudantes e qualitativa aplicando a técnica de sumarização de conteúdo [Bardin 2011] para as respostas das questões subjetivas. Para atingir as médias ponderadas utilizadas como critérios nas QA_2 e QA_3 , foi utilizada a equação 1, em que cada opção de avaliação i recebe um valor numérico $v_i \in 1, 2, 3, 4, 5$ e é associada a uma frequência de respostas f_i . O total de respostas por critério e estrutura é representado por $N = 28$.

$$\bar{Q} = \frac{\sum_{i=1}^5 (v_i \cdot f_i)}{N} \quad (1)$$

5. Resultados

Nesta seção são discutidos os resultados da avaliação da abordagem, a partir da percepção dos participantes referente à qualidade dos requisitos gerados por *LLMs*, influência da estrutura na qualidade dos requisitos gerados por *LLMs* e influência da estrutura dos requisitos na qualidade percebida dos *MVPs* gerados por *LLMs*.

Tabela 4. Perfil dos participantes

Categoria	Descrição	N	%
Faixa etária	Entre 22 e 26 anos	6	54,5%
	Entre 26 e 30 anos	4	36,4%
	Mais de 30 anos	1	9,1%
Área de atuação profissional	Requisitos	8	72,7%
	Product Manager	2	18,2%
	Desenvolvimento	1	9,1%
Tempo de experiência em Computação	Entre 1 e 3 anos	5	45,5%
	Entre 3 e 6 anos	4	36,4%
	Mais de 6 anos	2	18,1%
Familiaridade com ER	Alta familiaridade	2	28,6%
	Familiaridade total	5	71,4%
Conhecimento sobre técnicas de especificação	Bom conhecimento	1	14,2%
	Excelente conhecimento	6	85,8%

Conforme apresentado na Tabela 4, observa-se que a maioria dos participantes possui idade entre 22 e 30 anos (90,9%), caracterizando um grupo predominantemente jovem. Em relação à área de atuação profissional, 72,7% dos estudantes desempenham funções diretamente relacionadas a requisitos, enquanto 18,2% atuam como *Product Managers* e 9,1% na área de desenvolvimento. Quanto ao tempo de experiência na área de Computação, 81,9% possuem até seis anos de atuação profissional, indicando perfil em fase de consolidação de carreira. Entre os estudantes que responderam às questões específicas sobre ER (N=7), 100% declararam possuir alta ou total familiaridade com

atividades de elicitación, sendo que 85,8% afirmaram ter excelente conhecimento sobre técnicas de especificação. De modo geral, o perfil identificado demonstra que os participantes possuem base técnica consistente para realizar avaliação crítica dos artefatos, contribuindo para a validade das análises realizadas no contexto da atividade.

5.1. Qualidade de requisitos gerados por LLMs

Os resultados indicam que, na percepção dos estudantes, os requisitos gerados pelas LLMs apresentaram qualidade predominantemente satisfatória, com maior concentração de avaliações positivas nos critérios de corretude (85,7%) e consistência (82,1%). A verificabilidade também apresentou avaliação elevada (71,4%), enquanto a completude demonstrou maior variação nas respostas, evidenciando percepção de lacunas pontuais de detalhamento. Os resultados consolidados para cada um dos critérios, estão apresentados na Tabela 5.

Tabela 5. Avaliação dos requisitos gerados pela LLM por critério e estrutura

Critério	Opções de Avaliação	Estrutura 1	Estrutura 2
Corretude	Totalmente incorreto	0	0
	Parcialmente incorreto	2	0
	Neutro	2	2
	Parcialmente correto	19	9
	Totalmente correto	5	17
Ausência de ambiguidades	Muito ambíguo	0	0
	Ambíguo	1	0
	Neutro	8	3
	Pouco ambíguo	13	9
	Não ambíguo	6	16
Completude	Muito incompleto	0	0
	Incompleto	4	0
	Parcialmente completo	11	5
	Quase completo	10	13
	Totalmente completo	3	10
Consistência	Totalmente inconsistente	0	0
	Parcialmente inconsistente	0	0
	Neutro	5	4
	Parcialmente consistente	17	5
	Totalmente consistente	6	19
Verificabilidade	Não verificável	0	0
	Pouco verificável	2	0
	Parcialmente verificável	6	0
	Quase verificável	10	10
	Totalmente verificável	10	18

As respostas abertas reforçam que, embora os requisitos tenham sido considerados tecnicamente adequados, alguns estudantes identificaram ausência de maior especificidade e possíveis ambiguidades residuais, especialmente quando a estrutura utilizada era mais concisa. Do ponto de vista educacional, o processo avaliativo exigiu aplicação consciente de critérios formais da ER, promovendo análise criteriosa e julgamento técnico estruturado sobre artefatos produzidos por IA. Para isso, foram adotados como referência os atributos de qualidade definidos pela IEEE 830 (1998) [IEEE 1998] e pela ISO/IEC/IEEE 29148 (2018) [ISO/IEC/IEEE 2018], já apresentados na Tabela 1. A utilização dessas normas como métricas de avaliação permitiu que os estudantes aplicassem padrões reconhecidos internacionalmente na análise prática dos requisitos gerados, fortalecendo a compreensão dos critérios de qualidade e consolidando o aprendizado por meio da experiência analítica estruturada.

5.2. Influência da estrutura na Qualidade dos Requisitos Gerados por LLMs

A comparação entre as duas estruturas evidenciou que a Estrutura 2 obteve médias superiores em todos os critérios avaliados, conforme apresentado na Tabela 6. Destacam-se os critérios de verificabilidade (4,64), consistência (4,54) e corretude (4,54), indicando desempenho aproximadamente 15% superior em relação à Estrutura 1 no conjunto das métricas analisadas.

Tabela 6. Médias ponderadas da qualidade dos requisitos por critério e estrutura

Critério	Estrutura 1	Estrutura 2
Corretude	3,96	4,54
Ausência de ambiguidade	3,86	4,46
Completude	3,43	4,18
Consistência	4,04	4,54
Verificabilidade	4,00	4,64

Nas respostas abertas, 85,7% dos estudantes indicaram a Estrutura 2 como a mais completa e facilitadora da análise técnica, destacando o maior nível de detalhamento, a melhor contextualização do fluxo funcional e o alinhamento mais claro com as regras de negócio e o diagrama de atividades. Em contrapartida, a Estrutura 1 foi percebida como mais objetiva e concisa, favorecendo a leitura, porém com menor profundidade e menor suporte à validação dos requisitos. Além da percepção sobre qualidade, os estudantes relataram que a Estrutura 2 contribuiu para uma compreensão mais clara da lógica da funcionalidade e reduziu ambiguidades na interpretação. O maior detalhamento foi associado à maior segurança na avaliação técnica, especialmente nos critérios de verificabilidade e consistência. Contudo, alguns participantes apontaram que o excesso de informações pode aumentar a complexidade de leitura, exigindo maior esforço cognitivo. De modo geral, os resultados indicam que estruturas mais detalhadas e contextualizadas não apenas elevam a qualidade percebida dos requisitos gerados por LLMs, mas também favorecem a compreensão do comportamento esperado do sistema. Além disso, a comparação entre as estruturas estimulou reflexão crítica sobre como o nível de especificação influencia diretamente a clareza, a completude e a possibilidade de validação dos requisitos, fortalecendo a percepção dos estudantes acerca da importância da estruturação adequada na ER.

5.3. Influência da Estrutura dos Requisitos na Qualidade Percebida dos MVPs

Os resultados indicam que ambos os MVPs apresentaram níveis satisfatórios de qualidade, especialmente no critério de correção funcional. Contudo, o MVP 2, derivado da Estrutura 2, apresentou médias superiores em todos os critérios avaliados, conforme apresentado na Tabela 7.

Tabela 7. Médias ponderadas dos MVPs gerados por critério de qualidade

Critério de Avaliação	MVP 1	MVP 2
Aderência ao produto final	3,25	4,00
Correção funcional	3,75	4,25
Esforço de retrabalho	3,00	4,00
Compreensão pela IA	4,00	4,50

Destacam-se os critérios de aderência ao produto final (4,00), esforço de retrabalho (4,00) e compreensão pela IA (4,50), evidenciando melhor alinhamento entre especificação e implementação quando comparado ao MVP 1. As respostas qualitativas indicaram que o maior detalhamento presente na Estrutura 2 favoreceu melhor

interpretação pela IA e maior fidelidade entre os requisitos especificados e o *MVP* gerado. Em contraste, a menor contextualização da Estrutura 1 foi associada à necessidade de ajustes adicionais, especialmente no que se refere à aderência ao escopo e à completude funcional. Do ponto de vista pedagógico, a atividade permitiu aos estudantes compreender, de forma prática, a relação entre qualidade da especificação e qualidade da implementação. Ao observar diferenças concretas entre os *MVPs*, os participantes puderam refletir sobre como lacunas, ambiguidades ou ausência de detalhamento impactam diretamente o resultado final. Esse processo reforçou o entendimento de que a estruturação adequada dos requisitos não é apenas uma formalidade documental, mas um fator determinante para reduzir retrabalho e aumentar a precisão da solução. Assim, a experiência contribuiu para consolidar o aprendizado sobre a importância da clareza, completude e consistência na Engenharia de Software, fortalecendo o pensamento crítico acerca do uso de IA como ferramenta de apoio ao desenvolvimento.

5.4. Análise dos resultados e percepções do Tutor e Mediador

A análise integrada das QA1 a QA3 evidencia que a utilização de *LLMs* na geração de requisitos e *MVPs* produziu artefatos com qualidade predominantemente satisfatória, segundo a percepção dos participantes. De modo geral, os requisitos apresentaram melhor desempenho nos critérios de corretude e consistência, enquanto a completude demonstrou maior variabilidade, indicando que o nível de detalhamento influencia diretamente a robustez dos artefatos gerados. A comparação entre as estruturas evidenciou impacto direto da forma de especificação na qualidade dos artefatos. A Estrutura 2 apresentou médias superiores em todos os critérios, com melhora aproximada entre 12% e 22% em relação à Estrutura 1. Essa superioridade foi também reconhecida qualitativamente: 85,7% dos participantes indicaram a Estrutura 2 como mais completa e melhor alinhada aos artefatos de entrada, destacando maior contextualização, organização e clareza técnica. Esse efeito refletiu-se igualmente nos *MVPs*. O *MVP 2*, derivado da Estrutura 2, apresentou melhora entre 12% e 33% nos critérios avaliados, especialmente quanto ao esforço de retrabalho. Além disso, 75% dos *stakeholders* indicaram o *MVP 2* como o mais aderente ao produto esperado. Tais resultados reforçam a relação direta entre qualidade da especificação e qualidade do artefato implementado.

Entretanto, um dos achados mais relevantes do estudo não se restringe à qualidade técnica dos artefatos, mas ao processo de aprendizagem promovido. Os benefícios e limitações do uso das *LLMs* foram explicitamente identificados pelos próprios participantes durante a análise comparativa. Ao aplicar critérios formais de qualidade e confrontar diferentes estruturas, os participantes precisaram examinar criticamente os requisitos gerados, identificar lacunas, ambiguidades e inconsistências, e refletir sobre os impactos dessas falhas nos *MVPs* produzidos. Nesse sentido, a IA não atuou apenas como ferramenta de geração automática, mas como catalisadora de aprendizagem prática. O exercício de avaliar, comparar e justificar tecnicamente as diferenças entre as estruturas exigiu mobilização ativa de conceitos de ER, promovendo desenvolvimento de pensamento crítico e capacidade analítica. Os estudantes reconheceram, por exemplo, que lacunas de especificação resultaram em maior retrabalho ou menor aderência funcional, evidenciando compreensão sistêmica do ciclo requisito–implementação. Na Tabela 8 são sumarizados os benefícios e limitações percebidos a partir da avaliação dos participantes quanto ao uso de *LLMs*. Como destacado por um dos participantes: “*O trabalho humano*

ainda se faz necessário em revisar os requisitos gerados e pensar além dos cenários ofertados pela LLM...”. Tal observação evidencia que os estudantes compreenderam a IA como ferramenta de apoio, e não como substituta do profissional de ER.

Tabela 8. Benefícios e limitações percebidos a partir da avaliação dos participantes quanto ao uso de LLMs

Benefícios percebidos	Limitações percebidas
Agilidade na elaboração inicial: geração rápida de versões preliminares de requisitos, permitindo foco em análise e refinamento.	Dependência de detalhamento explícito: dificuldade em considerar requisitos implícitos ou conhecimento tácito do domínio.
Melhor visualização estrutural: organização textual que facilitou a compreensão do fluxo funcional.	Cobertura incompleta de exceções: cenários alternativos e casos extremos nem sempre contemplados.
Estímulo à revisão crítica: necessidade de validação incentivando postura investigativa e analítica.	Interpretação limitada de regras complexas: regras específicas exigem intervenção humana.
Aprendizagem prática baseada em evidências: comparação qualitativa e quantitativa evidenciando impacto da qualidade dos requisitos nos resultados técnicos.	Necessidade de validação especializada: supervisão humana indispensável para assegurar aderência ao contexto real.

6. Considerações Finais e Trabalhos Futuros

A principal contribuição do estudo está em sua dimensão pedagógica, ao demonstrar que o uso orientado de *LLMs* no processo avaliativo promove aprendizado ativo e desenvolvimento do pensamento crítico. Ao analisar diferentes estruturas de requisitos e avaliar sua qualidade com base em critérios formais como completude, consistência e verificabilidade, os estudantes exercitaram julgamento técnico e argumentação fundamentada. A IA não foi utilizada apenas como ferramenta de automação, mas como mediadora do aprendizado. A interação com artefatos gerados por *LLMs* exigiu validação, refinamento e tomada de decisão, evidenciando que a qualidade dos resultados depende da clareza das entradas e da supervisão humana. Embora os participantes tenham reconhecido benefícios como agilidade e melhor organização estrutural, também identificaram limitações, como omissões e a necessidade de validação especializada. Assim, o estudo reforça que *LLMs* podem apoiar a ER e seu ensino, desde que utilizadas de forma crítica, orientada e supervisionada. Como limitações do estudo, destacam-se o número reduzido de participantes, o escopo restrito a uma funcionalidade específica do sistema e as limitações operacionais das versões gratuitas das ferramentas utilizadas. Tais fatores impactam a generalização dos resultados e indicam a necessidade de ampliação do escopo investigativo.

No que se refere às ameaças à validade desse estudo, destacam-se: (i) o tamanho reduzido da amostra, (ii) a subjetividade inerente às avaliações, e (iii) a limitação do contexto educacional específico. Para mitigar tais fatores, o instrumento foi revisado previamente, garantiu-se anonimato aos participantes e os dados foram interpretados considerando o contexto específico da aplicação, além de que, adotou-se triangulação entre dados quantitativos e qualitativos, fortalecendo a robustez das conclusões. Como trabalhos futuros, sugere-se: (i) expandir a amostra de participantes e diversificar o perfil dos avaliadores; (ii) aplicar o experimento em múltiplas funcionalidades e contextos de sistemas distintos; (iii) realizar análises comparativas entre diferentes modelos e versões de *LLMs*; e (iv) investigar longitudinalmente o impacto do uso sistemático de IA no desenvolvimento das competências em ER. Essas extensões podem aprofundar a compreensão sobre como integrar efetivamente *LLMs* às práticas profissionais e educacionais da área.

7. Declaração de Uso de Inteligência Artificial Generativa

Este trabalho contou com o suporte de ferramentas de Inteligência Artificial (IA) generativa estritamente para fins de aprimoramento linguístico e revisão gramatical. A tecnologia foi aplicada para elevar a clareza do texto e refinar a redação de passagens específicas do manuscrito. Ressalta-se que todo o conteúdo científico, o que inclui a concepção do estudo, o delineamento da metodologia, a análise dos resultados e as conclusões, foi desenvolvido e revisado integralmente pelas pessoas autoras. Não houve uso de IA para a geração de dados primários, resultados experimentais ou referenciamento bibliográfico.

Referências

- Bardin, L. (2011). *Análise de conteúdo*. São Paulo: Edições 70.
- Cohn, M. (2005). *Agile Estimating and Planning*. Prentice Hall.
- Frattini, J. and Frattini, A. (2025). Adopting use case descriptions for requirements specification: an industrial case study. *arXiv preprint arXiv:2506.13303*.
- Hou, X., Zhao, Y., Liu, Y., Yang, Z., Wang, K., Li, L., Luo, X., Lo, D., Grundy, J., and Wang, H. (2023). Large language models for software engineering: A systematic literature review. *arXiv preprint arXiv:2308.10620*. 2308.10620.
- IEEE (1998). IEEE recommended practice for software requirements specifications.
- ISO/IEC/IEEE (2018). ISO/IEC/IEEE 29148:2018 systems and software engineering — life cycle processes — requirements engineering.
- Kuhail, M. A. and Lauesen, S. (2022). User story quality in practice: A case study. *Software*, 1(3):223–243.
- Macedo, M., Bezerra, C., and Coutinho, E. (2024). Uma pesquisa qualitativa do contexto de ensino em requisitos de software no brasil. In *Anais do XXXII Workshop sobre Educação em Computação*, pages 669–679, Porto Alegre, RS, Brasil. SBC.
- Maia, S. and Sarkis, L. (2025). Utilização de llm como ferramenta de apoio no ensino-aprendizagem de programação python para iniciantes: Um relato de experiência. In *Anais do XXXIII Workshop sobre Educação em Computação*, pages 385–396, Porto Alegre, RS, Brasil. SBC.
- Murad, S., Lemos, F., Melo, S., Paschoal, L., and Prates, J. (2025). Evidências sobre o uso do chatgpt no ensino de modelagem de software: um experimento controlado. In *Anais do XXXIII Workshop sobre Educação em Computação*, pages 1289–1300, Porto Alegre, RS, Brasil. SBC.
- Pacheco, C., García, I., and Reyes, M. (2018). Requirements elicitation techniques: a systematic literature review based on the maturity of the techniques. *IET Software*, 12(4):365–378.
- Pasuksmit, J., Thongtanunam, P., and Karunasekera, S. (2021). Towards just-enough documentation for agile effort estimation: What information should be documented? In *2021 IEEE International Conference on Software Maintenance and Evolution (ICSME)*, pages 251–262.

- Portela, A. and Vieira, J. (2025). Um estudo exploratório sobre o uso de IIs como recurso para o aprendizado de fundamentos de programação. In *Anais do XXXIII Workshop sobre Educação em Computação*, pages 527–538, Porto Alegre, RS, Brasil. SBC.
- Smart, J. F. (2014). *BDD in Action: Behavior-Driven Development for the Whole Software Lifecycle*. Manning Publications.
- Sommerville, I. (2019). *Engenharia de Software*. Pearson, São Paulo, 10 edition.