

Uso de Large Language Models na Engenharia de Requisitos

Ytalo Silva¹[0009-0006-6156-8929], Maria Gois¹[0009-0002-9730-8601], Abimael Santos¹[0009-0003-7555-6927], Jaelson Castro¹[0000-0002-4635-7297] e Maria Lencastre²[0000-0002-8032-8801]

¹ Universidade Federal de Pernambuco, Recife, PE - Brasil

² Universidade de Pernambuco, Recife, PE - Brasil

¹{ypws, mcdag, ajfs, jbc}@cin.ufpe.br ²mlpm@ecomppoli.br

Abstract. Os Large Language Models (LLMs), que são amplamente adotados em domínios como processamento de linguagem natural e geração de código, emergem como ferramentas promissoras para otimizar a Engenharia de Requisitos (ER). Por meio de uma Revisão Sistemática da Literatura (RSL), este artigo visa analisar sistematicamente como os LLMs têm sido aplicados na ER, explorando suas contribuições, benefícios, desafios e implicações éticas. Seguindo o protocolo estabelecido, foram analisados 50 estudos de várias bases bibliográficas. Os resultados revelam que os LLMs são empregados em todas as fases da ER (elicitação, análise, documentação e validação), destacando-se em automação de tarefas repetitivas, refinamento de requisitos e melhoria da comunicação entre *stakeholders*. Benefícios incluem ganhos de eficiência (redução de tempo/esforço) e maior precisão na documentação. Entretanto, desafios persistem, como inconsistências na saída dos modelos, dependência de intervenção humana, dificuldades em domínios especializados e riscos de vazamento de dados. Preocupações éticas, como transparência e privacidade, são pouco exploradas na literatura, apontando lacunas críticas. Conclui-se que os LLMs transformam a ER, mas exigem *frameworks* robustos de validação e políticas éticas para equilibrar inovação e responsabilidade.

Keywords: Engenharia de Requisitos, Large Language Models, Revisão Sistemática da Literatura.

1 Introdução

A Engenharia de Requisitos (ER) é uma fase crítica no desenvolvimento de software. Ela é responsável pela definição, documentação e manutenção das necessidades dos sistemas. Na literatura, podem ser encontradas diversas classificações para o processo de Engenharia de Requisitos. Neste trabalho, utilizamos a classificação de Kotonya e Sommerville [1] que inclui quatro atividades de alto nível: elicitação de requisitos, análise e negociação de requisitos, documentação de requisitos e validação de requisitos.

Com o avanço das tecnologias de inteligência artificial, especialmente os Large Language Models (LLMs), que são treinados com grandes volumes de textos para

compreender, gerar e manipular linguagem natural de forma contextualizada [2], surgem novas oportunidades para aprimorar processos na ER. A aplicação dessas tecnologias promete aumentar a eficiência e a precisão, ao mesmo tempo que reduz erros e ambiguidades nos requisitos [3].

Diante desse cenário, este trabalho tem como objetivo identificar e analisar como os LLMs têm sido utilizados na Engenharia de Requisitos. Para isso, realizamos uma Revisão Sistemática da Literatura (RSL) a fim de coletar estudos que atendem a critérios de elegibilidade pré-definidos, buscando responder a cinco questões de pesquisa (vide Tabela 1). A análise considera aspectos como os tipos de contribuição e objetivos dos estudos, as fases da ER abordadas, os benefícios das abordagens, desafios e riscos identificados, bem como preocupações éticas associadas ao uso de LLMs na ER.

Como principal contribuição, este trabalho fornece uma visão abrangente do uso do LLM na engenharia de requisitos, destacando: (i) sua aplicação nas etapas de elicitação, análise, documentação e validação; (ii) benefícios recorrentes, como automação, eficiência, qualidade e colaboração; (iii) mapeamento de desafios e riscos enfrentados relacionados à transparência, variabilidade do modelo e privacidade; e (iv) a lacuna na literatura sobre a avaliação crítica e pouca exploração de questões éticas.

As demais seções deste artigo estão organizadas da seguinte forma: Seção 2 discute os trabalhos relacionados; Seção 3 detalha a metodologia de pesquisa; Seção 4 expõe os resultados e as discussões; e, por fim, Seção 5 traz as conclusões e sugestões para trabalhos futuros.

2 Trabalhos Relacionados

O trabalho [4] relatou uma revisão da literatura que avaliou a aplicação do ChatGPT na engenharia de requisitos de software. Foram identificados 22 estudos coletados de 3 fontes de dados (IEEE, ACM e Google Scholar), publicados entre 2022 e 2024. Este estudo [4] é limitado a aplicações de ChatGPT em ER e foca principalmente em seus benefícios e desafios.

Já os resultados preliminares do trabalho [5], apresentado na trilha do estudante do WER 2024, se refere a uma RSL sobre a Integração do ChatGPT para aprimoramento da definição de requisitos de software. Foram identificados 08 estudos coletados de 05 fontes de dados (IEEE, ACM, Google Scholar, Scopus e Science Direct). Este estudo [5] também é limitado a aplicações somente ChatGPT em ER e se baseia em um reduzido número trabalhos publicados em 2023.

3 Metodologia de Pesquisa

Para a execução desta RSL, as diretrizes e o modelo de protocolo proposto por Kitchenham e Charters [6] foram utilizados, os quais se desdobram em três fases principais: planejamento, execução e relato.

A RSI foi realizada entre junho e dezembro de 2024 e todo o processo foi conduzido por 2 pesquisadores que participaram ativamente da leitura e análise por pares dos

estudos enquanto outros 3 pesquisadores participaram na revisão e escrita do artigo, assegurando a consistência e a qualidade metodológica.

Para facilitar a execução das etapas da revisão, a ferramenta web Parsifal [7] foi escolhida, já que foi desenvolvida com base no protocolo estabelecido por Kitchenham e Charters [6].

3.1 Perguntas de Pesquisa e Motivações

A revisão proposta busca explorar a integração dos Large Language Models na Engenharia de Requisitos, delineando cinco perguntas centrais que orientam o estudo, conforme Tabela 1.

Tabela 1. Perguntas de Pesquisa e Motivações.

Pergunta de Pesquisa	Motivação
RQ1. Que tipo de contribuição e objetivo de pesquisa estão sendo abordados no estudo?	Permite uma compreensão das áreas de interesse exploradas e suas implicações para o avanço teórico e prático do campo.
RQ2. Em qual fase da Engenharia de Requisitos os LLMs estão sendo abordados no estudo?	A análise destaca o impacto potencial dos LLMs nas diversas etapas do processo da ER.
RQ3. Quais são os benefícios das abordagens utilizadas nos estudos?	Busca evidenciar como os LLMs podem aprimorar a eficiência e a qualidade dos processos de requisitos.
RQ4. Quais são os principais desafios e riscos ao utilizar LLMs na Engenharia de Requisitos?	Identifica limitações e obstáculos que podem comprometer sua utilização dos LLMs na ER.
RQ5: De que forma as preocupações éticas sobre a geração de texto automatizada pelos LLMs influenciam sua adoção e uso na Engenharia de Requisitos?	Busca compreender como as preocupações éticas influenciam a aceitação e a adoção dos LLMs na prática da ER, destacando que lidar com essas implicações é imperativo para um desenvolvimento seguro e responsável.

3.2 Estratégias de Busca e Seleção

A partir de uma análise das perguntas de pesquisa, foram obtidas as seguintes palavras-chave: “Engenharia de Requisitos” e “LLM”. Após alguns testes em repositórios acadêmicos, chegamos à elaboração da *string* de busca a seguir, que visa obter artigos que abordassem sobre diferentes fases da Engenharia de Requisitos.

("large language model" OR "LLM" OR "GPT" OR "BERT" OR "generative AI") AND ("requirements elicitation" OR "requirements validation" OR "requirements analysis" OR "requirements specification" OR "requirements management" OR "requirements engineering"))

Na condução de uma revisão, é crucial identificar os trabalhos relevantes para abordar as questões de pesquisa. Para isso, foi realizada uma busca automática nas principais bases de dados eletrônicas que contêm periódicos de engenharia de software e

anais de conferências de alta qualidade, foram selecionadas entre elas: ACM Digital Library, IEEE Digital Library, Science@Direct e SpringerLink.

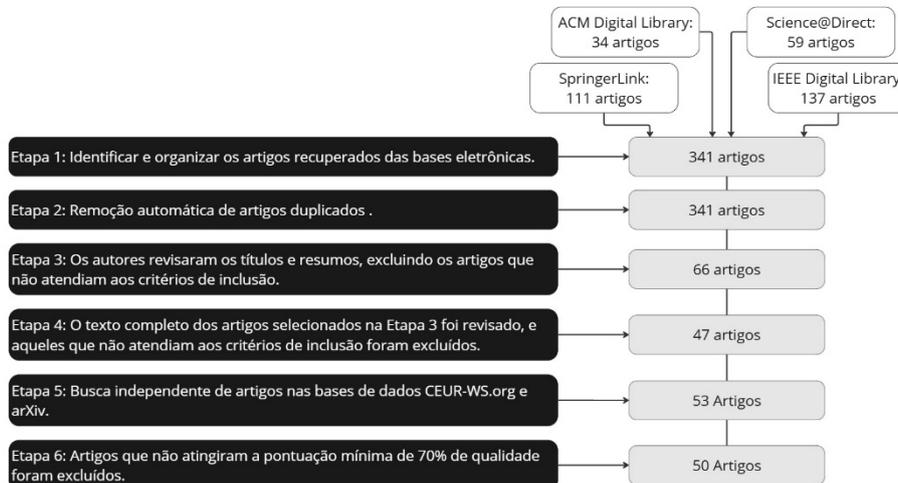
3.3 Processo de Condução e Seleção dos Estudos

Nosso procedimento de seleção dos estudos consistiu em seis etapas principais, conforme mostrado na Figura 1. Na etapa 1, identificamos e organizamos os estudos oriundos das bases de dados, ao total, foram localizados 341 estudos. Foi feita a verificação de estudos duplicatas na etapa 2, no qual não foi encontrado nenhum estudo duplicado.

Na etapa 3, revisamos os títulos e resumos dos 341 artigos identificados, resultando em 66 artigos após a exclusão dos que não atendiam aos critérios de inclusão. Na etapa 4, realizamos a leitura dos 66 artigos na íntegra, excluindo aqueles que não atendiam aos critérios, resultando em 47 artigos.

Na etapa 5, foram incluídos 06 novos artigos, totalizando 53 artigos, sendo 03 provenientes de workshops do REFSQ publicados no CEUR-WS.org e 03 artigos publicados no arXiv citados por outros estudos. Na etapa final (etapa 6), procedeu-se à análise de qualidade dos 53 artigos. Desses, 50 atingiram a pontuação praxe mínima de 70%, critério que permitiu sua inclusão na análise final. As avaliações detalhadas dos critérios de qualidade, assim como as referências completas de todos os estudos incluídos nesta revisão sistemática, estão disponíveis em <https://shre.ink/MXdq>.

Figura 1. Processo de revisão sistemática e seleção por bases eletrônicas. Adaptado de [8].



Critérios de inclusão e exclusão. Em virtude de os LLMs terem surgido em 2018, optamos por analisar artigos primários, publicados entre 2019 e 2024, que relacione ER e LLMs. Estão fora do nosso escopo de análise artigos redundantes (da mesma autoria), idioma diferente do inglês, artigos duplicados, artigos secundários, literatura cinzenta, artigos que não estão disponíveis através dos mecanismos de buscas ou entrando em

contato com os autores, artigos com menos de 5 páginas e artigos cujo foco principal não seja a relação entre ER e LLMs.

3.4 Ameaças à Validade

As ameaças à validade representam um desafio crítico em revisões sistemáticas da literatura, especialmente no campo da engenharia de software. Para mitigar isso, as ameaças à validade nesta revisão foram analisadas em quatro dimensões de acordo com [9], com estratégias de mitigação adotadas para garantir a robustez do estudo:

- **Validade de Construto:** O viés na seleção de estudos e a imprecisão na extração de dados são riscos à validade de RSLs, por isso adotamos diretrizes específicas para criar um protocolo de pesquisa confiável, validando-o por comparação com protocolos publicados. A string de busca foi revisada para incluir sinônimos e evitar a omissão de estudos relevantes. No entanto, não incluímos termos explícitos relacionados à ética, o que pode ter limitado a identificação de estudos diretamente ligados à RQ5. Ainda assim, aspectos éticos foram considerados indiretamente com base em menções implícitas nos estudos analisados. Além disso, a estruturação da string de busca poderia ser mais bem explicada.
- **Validade Interna:** Tentamos reduzir ameaças relacionadas ao viés pessoal na análise dos estudos. Dois dos autores têm larga experiência em Engenharia de Requisitos. Porém, pode haver decisões subjetivas na coleta de dados, pois alguns artigos podem não ter descrito seus objetivos e resultados de forma clara.
- **Validade Externa:** Estávamos preocupados com a generalização dos resultados da SLR, que está relacionada ao grau em que os estudos primários representam o tema da revisão. Para mitigar, a busca foi definida após várias buscas experimentais e validada. A análise realizada também melhora a validade externa ao melhorar incrementalmente a qualidade do conjunto de dados utilizado para tirar conclusões gerais.
- **Validade de Conclusão:** O processo de seleção e os critérios de inclusão e exclusão foram cuidadosamente projetados e discutidos pelos autores para minimizar o risco de exclusão de estudos relevantes.

4 Resultados e Discussões

Nesta seção, são apresentados os resultados obtidos a partir da análise dos 50 estudos selecionados, publicados entre 2022 e 2024 (ver estudos incluídos em: <https://shre.ink/MXdq>). Os resultados indicaram que o ponto de maior concentração ocorreu em 2024 (25 artigos), seguido por 2023 (24 artigos) e 2022 (1 artigo). Embora no nosso critério de inclusão tenhamos considerados cinco anos (2019 a 2024), a seleção de estudo final abrange apenas o período de 2022 a 2024, pois nos demais anos os critérios não foram satisfeitos. A seguir, são detalhadas as respostas obtidas para cada uma das questões de pesquisa.

RQ1. Que tipo de contribuição e objetivo de pesquisa estão sendo abordados no estudo?

Esta questão busca identificar os principais focos de pesquisa recente sobre o uso de LLMs na ER. O objetivo é entender como essas tecnologias estão sendo aplicadas para melhorar a qualidade e eficiência do processo de desenvolvimento de requisitos em projetos de software.

Os estudos [E2, E33, E48, E50] abordam, respectivamente, a detecção de contradições [E2], inconsistência [E33], incompletude [E48] e ambiguidade [E50] estabelecendo benchmarks, métricas de qualidade e métodos automatizados para avaliação documental. Enquanto [E43] aborda a identificação de variabilidades, [E53] técnicas de demarcação e [E14] avaliação automática de qualidade.

Os trabalhos de [E11, E15, E18, E23, E36, E46, E47, E51] focam na classificação de requisitos, abrangendo técnicas como modelos de transfer learning [E11], arquiteturas híbridas com Graph Attention Networks e BERT (DBGAT) [E18], comparação de SVM/LSTM/LLMs [E15], e análise de trade-offs em LLMs [E51]. Esses estudos aplicam-se a contextos específicos como requisitos não funcionais [E11, E18], documentos ferroviários [E46], e classificação de requisitos em espanhol [E47], contribuindo para a recuperação de links de rastreabilidade [E23], extração de preocupações com explicabilidade [E36], pipelines automatizados [E46], e aprimoramento da generalização de modelos via integração de informações estruturais e sintáticas [E18], estabelecendo bases para classificação precisa, eficiente e adaptável a diferentes domínios.

As pesquisas [E21, E38, E40, E41, E52] utilizam LLMs para transformar requisitos em linguagem natural para modelos visuais, incluindo diagramas UML de sequência [E21], DFD automatizados [E52], extensão de Problem Frames [E38], modelagem GRL iterativa com GPT-4 [E41], e integração com ciclos MAPE-K [E40]. Paralelamente, [E1, E20, E25] abordam modelagem de domínio específico: geração ágil de modelos de domínio [E1], comparação entre métodos tradicionais e baseados em LLMs para fluxos de trabalho [E25], e framework para modelos de dispositivos IoT [E20], demonstrando a adaptabilidade dos LLMs na transição entre requisitos naturais e representações técnicas especializadas.

Adicionalmente, as publicações [E5, E9, E17, E29] automatizam a geração de códigos e testes a partir de requisitos, utilizando engenharia de prompts com padrões estruturados para ChatGPT visando qualidade do código e refatoração [E5], combinados com prompting progressivo para colaboração humano-IA iterativa [E29]. Além disso, o framework ChatCoder é usado para aprimorar requisitos e código via diálogos com LLMs [E17] e a validação de software é feita através da geração de testes de aceitação por extração de condicionais dos requisitos [E9].

Estudos recentes [E27, E34, E44] exploram a criação de roteiros de entrevistas para elicitación de requisitos utilizando três pilares complementares: 1) agentes GPT com encadeamento de prompts para scripts complexos [E27]; 2) visualizações interativas que mapeiam erros comuns durante a elicitação [E34]; e 3) técnicas de decomposição de tarefas adaptadas a contextos empresariais específicos via prompt engineering [E44], integrando geração automatizada com mecanismos de validação e contextualização.

Diversos estudos [E3, E4, E8, E19, E22] têm explorado o uso de LLMs para automatizar a geração e complementação de especificações de requisitos. As abordagens

incluem a geração estruturada de requisitos, como um framework para sistemas IoT [E3], além de ferramentas de assistência, como o autocompletar do ReqCompletion [E22] e uma avaliação empírica em especificação de requisitos de software [E4]. Adicionalmente, foram propostos um modelo automatizado visando otimizar a especificação de requisitos [E8] e a metodologia Quest-RE, que enriquece os requisitos por meio de questionamentos [E19].

Pesquisas [E32, E37, E39, E42] demonstram o potencial dos LLMs na automação de tarefas complexas de ER: desde a geração estruturada de modelos via zero-shot learning [E37] e rastreabilidade [E42], até à análise jurídica de contratos com identificação de requisitos de segurança [E39] e sumarização contextual de contratos legais [E32] para analistas de requisitos. O estudo [E24] avança nesse sentido ao utilizar aprendizado em contexto e design de prompts para validar sistemas, gerando e avaliando argumentos com modelos como GPT-4, propondo melhorias na precisão da verificação. Para elicitación criativa, destacam-se frameworks que integram valores humanos em histórias de usuário [E35] e transformação de requisitos técnicos em narrativas acessíveis [E7]. Abordagens complementares incluem a mineração de requisitos a partir de vídeos em redes sociais [E31] e automação de verificação de completude [E6], reduzindo erros e aumentando a eficiência na coleta de requisitos.

Por fim, os estudos [E10, E12, E13, E16, E26, E28] focaram em aplicar LLMs e IA generativa para resolver desafios como explicabilidade no rastreamento de requisitos, garantindo a transparência no desenvolvimento assistido por IA [E26], otimização na identificação e gestão de requisitos não funcionais [E10], automação de tarefas de ER para otimizar ciclos de desenvolvimento [E16] e o potencial do ChatGPT para auxiliar nos processos de elicitación de requisitos, propondo diretrizes que podem ser aplicadas na prática [E13]. Além disso, [E12] investigou como a IA generativa pode avançar a ER, avaliando o papel dos LLMs na elicitación e análise de requisitos para melhorar a comunicação entre as partes interessadas, enquanto [E28] desenvolveu um sistema de validação de protótipo GUI baseado em LLMs para identificar discrepâncias entre requisitos e implementação, sugerindo componentes de interface adequados.

A análise dos estudos mostra que os LLMs estão transformando a ER ao automatizar tarefas, aprimorar a documentação e integrar-se com métodos ágeis. Eles simplificam os processos tradicionais e indicam potencial para abordagens mais eficientes para rastreamento, elicitación e especificação de requisitos [E21, E23, E34]. Isso leva a fluxos de trabalho mais adaptáveis e marca um avanço significativo no desenvolvimento de software, reduzindo erros e aprimorando a comunicação com as partes interessadas.

RQ2. Em qual fase da Engenharia de Requisitos os LLMs estão sendo abordados no estudo?

Esta seção tem como objetivo investigar em quais etapas da Engenharia de Requisitos os LLMs são aplicados, identificando em quais fases seu impacto é mais significativo. Conforme indicado na Tabela 2, as pesquisas identificaram o uso predominante em quatro etapas: elicitación, análise, documentação, e validação, com variações consideráveis nos resultados e limitações dependendo do contexto de aplicação.

Tabela 2. Aplicação de LLMs nas fases da Engenharia de Requisitos.

Fases	Estudos
Elicitação de Requisitos	[E5, E8, E12, E13, E16, E25, E27, E31, E32, E34, E35, E39, E44, E46, E53]
Análise e Negociação de Requisitos	[E2, E6, E8, E10, E11, E14, E15, E16, E17, E18, E19, E29, E33, E36, E40, E41, E47, E48, E50, E51]
Documentação de Requisitos	[E1, E3, E4, E7, E21, E22, E29, E37, E38, E52]
Validação de Requisitos	[E9, E17, E20, E23, E24, E25, E26, E28, E42, E43]

Dentre os 50 artigos analisados, 20 exploram a aplicação de LLMs na análise de requisitos, abrangendo tarefas como: (i) classificação automatizada – incluindo a diferenciação entre requisitos funcionais e não-funcionais, com destaque para o estudo de [E10], que categoriza requisitos não-funcionais em dimensões de sustentabilidade e TI Verde; (ii) verificação da qualidade dos requisitos, focada na identificação de anomalias (contradições, inconsistências, ambiguidades e lacunas de completude) [E2, E6, E14, E19, E33, E48, E50]; e (iii) processos de extração, refinamento e especificação de requisitos, visando aprimorar a precisão, completude e aderência às expectativas dos stakeholders [E16, E17, E29, E40, E41].

Além da análise, observou-se em 15 estudos o emprego de LLMs na fase de elicitação de requisitos. Dentre as aplicações, destacam-se: (i) o desenvolvimento de padrões de prompt específicos para aprimorar a extração de requisitos [E5, E12, E13]; (ii) a integração de LLMs na identificação de requisitos funcionais e não funcionais em documentos como contratos e licitações [E8, E16, E32, E39, E44, E46, E53]; e (iii) a geração automatizada de scripts para entrevistas, user stories e outros artefatos [E25, E27, E31, E34, E35].

Além dessas fases, a aplicação de LLMs se estende para a documentação e validação dos requisitos. Em 10 estudos os autores abordam o uso de LLMs na fase de documentação de requisitos, com aplicações em: (i) geração automatizada de artefatos, como modelos de domínio, diagramas UML e histórias de usuários [E4, E21, E52]; (ii) transformação e refinamento de requisitos, incluindo conversão em histórias de usuários, especificação de sistemas IoT e sugestão de prefixos para descrição de requisitos [E3, E7, E22, E29, E37, E38]; e (iii) validação e correção, com foco na construção de modelos de alta qualidade e identificação de inconsistências [E1].

De forma complementar, 9 estudos abordam o uso de LLMs na validação de requisitos, com tarefas envolvendo: (i) automação de testes, como criação de casos de teste de aceitação e validação de protótipos GUI [E9, E28]; (ii) geração e rastreabilidade de código, incluindo a produção de código-fonte a partir de requisitos e verificação de quais requisitos foram atendidos ou ignorados [E17, E24]; (iii) gestão de rastreabilidade, por meio de links entre requisitos, filtragem de informações irrelevantes e identificação de variabilidade em documentos [E23, E42]; e (iv) validação de modelos, como transformação de modelos gráficos em descrições textuais e análise de consistência em linguagem natural [E25, E26, E43].

Os estudos analisados demonstram uma crescente aplicação de LLMs nas fases de elicitação, análise e validação de requisitos, com destaque para tarefas como extração automática, classificação e refinamento de especificações. Essa tendência sugere oportunidades futuras para a integração desses modelos em ferramentas de suporte à Engenharia de Requisitos.

RQ3. Quais são os benefícios das abordagens utilizadas nos estudos?

As abordagens que empregam LLMs oferecem benefícios significativos nas fases da ER, conforme evidenciado pelos estudos analisados. Esses benefícios vão desde a automação de tarefas até à melhoria da qualidade e rastreabilidade dos requisitos, conforme indicado na Tabela 3.

Tabela 3. Benefícios das abordagens de LLMs na ER.

#	Benefícios	Estudos
B1	Automação de tarefas manuais na Engenharia de Requisitos	[E1, E3, E4, E7, E8, E9, E12, E16, E17, E21, E22, E26, E27, E29, E31, E32, E34, E35, E37, E39, E40, E41, E42, E44, E46, E52, E53]
B2	Redução de tempo e esforço no ciclo de desenvolvimento	[E1, E3, E4, E7, E8, E9, E11, E14, E17, E20, E21, E22, E26, E29, E37, E52]
B3	Melhoria na qualidade dos requisitos	[E2, E3, E6, E13, E14, E17, E19, E22, E24, E25, E33, E43, E48, E50]
B4	Classificação e organização eficiente de requisitos	[E6, E10, E11, E18, E23, E25, E36, E39, E44, E47, E51]
B5	Facilitação da comunicação entre stakeholders e equipes técnicas	[E1, E5, E7, E12, E13, E19, E21, E28, E35]
B6	Suporte à tomada de decisão e análise de requisitos	[E10, E11, E14, E15, E39, E40, E41, E44, E46]
B7	Geração de artefatos auxiliares	[E9, E16, E21, E28, E29, E38, E40, E41, E52]
B8	Escalonamento e redução de custos em processos complexos	[E12, E20, E29, E39, E46]
B9	Aprimoramento da rastreabilidade e manutenção de requisitos	[E23, E26, E42, E53]
B10	Capacitação e treinamento em Engenharia de Requisitos	[E27, E34, E35]

Os LLMs têm a capacidade de automatizar tarefas repetitivas (B1), incluindo a elaboração de documentos como, Especificações de Requisitos de Software (Software Requirements Specification), histórias de usuários, scripts de entrevistas, modelos UML e diagramas de fluxo de dados (DFDs). Essa automação reduz a carga cognitiva dos engenheiros e acelera processos como a elicitação, especificação e validação de

requisitos. Estudos como [E1, E3, E7, E21, E29] destacam a geração automática de artefatos, enquanto [E27, E34] mostram a criação de materiais didáticos para treinamento. Além disso, outro benefício importante (B2) contribui para que os requisitos sejam mais claros, completos e consistentes por meio de ferramentas como o BERT e ChatGPT que identificam termos faltantes, ambiguidade e até mesmo inconsistência em requisitos [6E, E14, E50].

A automação de tarefas, juntamente com o processamento rápido de textos em linguagem natural oferecido pelos LLMs permite ter ganhos de eficiência, contribuindo para o benefício (B3). Por exemplo, a aceleração de requisitos em histórias de usuários [E7] e geração de código [E29] que mostram redução de tempo e esforço nessas tarefas, conforme demonstrado nos estudos [E4, E26].

Os LLMs podem melhorar a interação entre as equipes e stakeholders (B4), como, por exemplo, a geração de perguntas direcionadas para promover a esclarecimento de requisitos a fim de reduzir mal-entendidos [E19]. E para a tomada de decisão e análise estratégicas (B5), como o estudo [E41] que auxilia na identificação de trade-offs entre metas e [E39] que faz análise de conformidade com regulamentos.

Outro benefício importante é a criação de artefatos (B6) a partir de requisitos, como diagramas de sequências UML [E21]. Além disso, os LLMs auxiliam na geração de links entre requisitos, código e outros artefatos (B7), como relatado nos estudos [E23, E42] que utilizam LLMs para recuperar links de rastreabilidade.

A geração de materiais didáticos e de simulações para capacitação (B8) como os estudos [E27, E34] que proporciona materiais didáticos ricos e variados para treinamento de futuros analistas de requisitos com a automação da geração de scripts de entrevistas para elicitación. Outro aspecto relevante é a redução de custos operacionais e escalabilidade de soluções (B9), que oferece uma solução escalável e com redução de custos para testar o impacto dos requisitos antes de ser implementado [E20].

Além disso, os LLMs podem ser utilizados para organizar e classificar requisitos (B10), diferenciando em funcionais e não-funcionais [E11], como também requisitos obrigatórios de segurança e privacidade [E44] e até mesmo mapeando requisitos para dimensões de sustentabilidade [E10].

RQ4. Quais são os principais desafios e riscos ao utilizar LLMs na Engenharia de Requisitos?

A análise dos 50 estudos revelou 12 limitações (L) que indicam os desafios e riscos ao aplicar LLMs na ER. Esses desafios abrangem desde limitações técnicas até dependências metodológicas, conforme sintetizado na Tabela 4 e detalhado a seguir.

A qualidade dos resultados pode estar naturalmente ligada à intervenção humana e à solidez dos dados (L1). É o caso do estudo [E17] que aborda a necessidade de interação contínua entre humanos e LLMs para validar e ajustar os requisitos refinados e o estudo [E18] que fala que o BERT precisa de dados de treinamento mais robustos.

Além disso, a análise dos estudos indica que o desempenho pode variar significativamente entre os LLMs e execuções (L2), como, por exemplo, no estudo [E15] notou desequilíbrios em conjuntos de dados, enquanto no estudo [E53] fizeram um alerta para a necessidade de ajustes contínuos nos hiperparâmetros.

Tabela 4. Desafios e riscos no uso de LLMs para ER.

#	Desafios e Riscos	Estudos
L1	Dependência de validação humana e dados	[E6, E13, E14, E17, E18, E19, E20, E22, E28, E29, E33, E35, E48, E52]
L2	Variabilidade entre modelos e respostas	[E4, E9, E15, E23, E28, E31, E38, E41, E42, E43, E46, E50, E53]
L3	Limitações em dados/treinamento	[E6, E10, E11, E12, E15, E18, E22, E36, E47, E48, E51, E53]
L4	Dependência da qualidade dos prompts	[E3, E5, E24, E26, E29, E32, E37, E40, E42, E43]
L5	Interpretações ambíguas ou errôneas	[E17, E20, E24, E26, E33, E39, E44, E51, E52]
L6	Limitações de contexto e domínio	[E8, E11, E20, E22, E35, E37, E47, E51]
L7	Ruídos e falsos positivos	[E1, E6, E9, E23, E31, E46, E48]
L8	Complexidade na tradução de requisitos	[E2, E16, E34, E38, E39, E44]
L9	Perda de informações ou nuances	[E7, E21, E25, E27, E32, E35]
L10	Inconsistência em validação e correção automática	[E1, E7, E4, E6, E50]
L11	Opacidade (natureza "caixa-preta")	[E19, E26, E36, E50]
L12	Saídas genéricas ou alucinações	[E25, E35, E40, E41]

A qualidade e generalização dos resultados que dependem, geralmente, dos dados de treinamento e de ajustes finos (L3), conforme discutido em [E6] que destaca a necessidade de otimizar as configurações do modelo para contextos específicos, [E11] que alerta sobre desequilíbrios de dados em novos domínios, e [E51] que enfatiza a escassez de dados específicos de ER que limitam a adaptação de modelos gerais.

A formulação de prompts também influencia diretamente a precisão das saídas (L4). Isso é evidenciado por [E5], que mostra como variações nos prompts geram inconsistências, e por [E42], que destaca que a estratégia de prompt ideal pode variar conforme o contexto, exigindo experimentação e ajustes contínuos para otimizar os resultados.

Outro risco e desafios associado ao uso de LLMs na ER é a ambiguidade na linguagem natural que pode levar a erros de interpretação (L5), conforme colocado pelo estudo [E52] que identificou diagramas incompletos devidos a falhas semânticas.

Um dos grandes desafios enfrentado pelos LLMs é a dificuldade de contexto e domínio especializados (L6). O estudo [E8] discute a dificuldade em garantir a exatidão e relevância das respostas em contextos complexos. Já [E22] alerta sobre os riscos de alinhamento impreciso com as intenções dos engenheiros, caso o modelo não capture nuances específicas do domínio. Além disso, a presença de informações irrelevantes ou inconsistentes nos requisitos (L7) podem afetar a precisão, por exemplo, dos links de rastreabilidade [E23].

A conversão de requisitos em linguagem natural para estruturas lógicas ou código também se mostra um desafio significativo (L8), conforme evidenciado pelos estudos [E2, E16], pois traduzir linguagem natural é complexo e apresenta diferentes nuances, incluindo a conversão de requisitos em fórmulas lógicas [E2] e a tradução em semântica de programação [E16].

Os LLMs tendem, ainda, a simplificar excessivamente ou omitir detalhes críticos (L9), conforme relatado em [E25], ocasionando a perda de nuances processuais em descrições automatizadas. Alguns estudos também apontam variações de inconsistência (L10) na identificação de relação entre elementos de modelos [E1] e na correção automática de requisitos [E4].

Outro aspecto relevante é a falta de transparência (L11) nas decisões dos LLMs, que pode dificultar a auditoria de requisitos, conforme evidenciado por [E19] que indica a falta de transparência nos resultados. Enquanto [E26] destaca a natureza não interpretável dos modelos o que compromete a rastreabilidade de requisitos. Por fim, os LLMs tendem a gerar respostas superficiais ou até mesmo incorretas (L12) em cenários ambíguos, conforme relatado em [E25] que indica a perda de informações críticas e [E41] que adverte sobre a ocorrência de alucinações em requisitos não estruturados.

Portanto, embora os LLMs sejam uma ferramenta de grande potencial para uso em tarefas da ER, os desafios e riscos evidenciados pelos estudos exigem cuidados e estratégias complementares como, integração com ferramentas tradicionais e frameworks de validação robustos, para garantir que os resultados sejam precisos e úteis no contexto de desenvolvimento de software.

RQ5: De que forma as preocupações éticas sobre a geração de texto automatizada pelos LLMs influenciam sua adoção e uso na Engenharia de Requisitos?

A análise dos estudos nos revelou que as preocupações éticas que estão relacionadas à geração automática de textos pelos LLMs não são muito discutidas na literatura atual na área de Engenharia de Requisitos. Dos 50 estudos analisados, apenas alguns mencionam indiretamente questões relativas à ética, como a transparência, confidencialidade de dados e necessidade de supervisão humana. Embora a falta de discussões explícitas, os tópicos citados oferecem ideias sobre potenciais impactos éticos na adoção de LLMs.

O estudo [E14] fala sobre a importância da transparência nas decisões geradas por LLMs e reforça a necessidade de intervenção humana para validar resultados. A falta de explicabilidade nos processos de LLMs pode comprometer a confiança dos profissionais nas saídas geradas, principalmente em contextos críticos de ER, onde decisões vagas ou injustificadas podem comprometer projetos.

Além disso, os estudos [E32, E44] destacam o risco de vazamento de informações sensíveis ao utilizar LLMs. Por exemplo, [E32] optou por ocultar os dados confidenciais de contratos ao treinar modelos, já [E44] evitou o uso de LLMs mais atuais, que têm acesso a informações, como o GPT-4, a fim de evitar exposições de dados. Essas abordagens refletem preocupações éticas com privacidade e segurança que podem restringir o uso de LLMs em domínios regulamentados.

Além disso, os estudos [E14, E20, E29, E52] reforçam a necessidade de validação humana para garantir a qualidade das saídas geradas pelos LLMs. O estudo [E29] defende a supervisão humana para reduzir erros e vieses não detectáveis por LLMs. Essa

dependência de revisão humana, conforme discutido na resposta de pesquisa anterior, é um desafio relacionado ao uso de LLMs em tarefas de ER, tendo em vista que pode aumentar custo e complexidade, mas pode ser benéfico, podendo um humano validar e interagir com os modelos para chegar a uma saída melhor.

Por fim, a falta de debates éticos explícitos sugere uma lacuna na literatura, mas os poucos estudos analisados que abordam o tema indiretamente, destacam desafios práticos que afetam a confiança do LLMs. Para mitigar esses desafios, é necessário implementar políticas de segurança robustas, garantir maior transparência nos processos de decisão dos modelos e ajustar os algoritmos para reduzir vieses. Somente assim será possível explorar todo o potencial dos LLMs nesse campo, equilibrando inovação com responsabilidade ética.

5 Conclusão

O estudo realizou uma revisão sistemática da literatura, destacando o potencial da integração de Large Language Models na Engenharia de Requisitos. A qualidade dos requisitos é um fator crucial para o sucesso de qualquer projeto de desenvolvimento de software, e a Inteligência Artificial, por meio dos LLMs, oferece uma abordagem inovadora para facilitar esse processo. Esses modelos têm demonstrado grande versatilidade e capacidade de contribuição em diversas etapas da Engenharia de Requisitos, desde o levantamento e análise dos requisitos até a validação deles.

Contudo, há desafios significativos, especialmente as preocupações éticas associadas ao uso de LLMs [10]. Além disso, os LLMs impactam a dinâmica de colaboração entre os stakeholders durante a Engenharia de Requisitos, podendo tanto facilitar a comunicação e o entendimento mútuo quanto introduzir novos desafios relacionados à coordenação e à confiança em LLMs [11]. Assim, embora os LLMs possam contribuir de forma relevante para diversas fases do processo, sua integração deve ser conduzida de maneira cuidadosa e ética, visando assegurar resultados responsáveis e eficazes.

Portanto, é necessária uma continuidade de pesquisas sobre os aspectos éticos relacionados ao uso de LLMs na elicitação de requisitos. É fundamental que organizações e pesquisadores adotem uma postura proativa na análise e implementação de políticas éticas e de segurança ao empregar essas tecnologias. Além disso, promover uma maior transparência e conscientização sobre o uso dos LLMs é essencial para garantir que seus benefícios sejam maximizados enquanto os riscos são mitigados, oferecendo um equilíbrio entre inovação e responsabilidade.

Para trabalhos futuros, é recomendável expandir a busca de artigos em outras línguas, como português e espanhol, além de apenas o inglês, e utilizar outras bibliotecas, como a SOL, mantida pela Sociedade Brasileira de Computação (SBC), a base WER-Papers, Google Scholar, Scopus, EI Compendex, Willey e Web of Science para obter uma visão mais ampla e diversificada sobre o uso dos LLMs na Engenharia de Requisitos.

Referências

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) e Fundação de Amparo a Ciência e Tecnologia do Estado de Pernambuco (FACEPE).

Referências

1. Kotonya, G., Sommerville, I.: *Requirements Engineering: Processes and Techniques*. John Wiley & Sons, Re-Printed, Chichester (2002).
2. Varma, D.M.N.: Advanced Transformer-Based Deep Learning Techniques For Enhancing Contextual Understanding In: *Natural Language Processing*. *AJBR*. 5153–5161 (2024). <https://doi.org/10.53555/AJBR.v27i4S.4562>.
3. Lutze, R., Waldhör, K.: Generating Specifications from Requirements Documents for Smart Devices Using Large Language Models (LLMs). In: Kurosu, M., Hashizume, A. (eds.) *Human-Computer Interaction*. pp. 94–108. Springer Nature Switzerland, Cham (2024).
4. Marques, N., Silva, R.R., Bernardino, J.: Using ChatGPT in Software Requirements Engineering: A Comprehensive Review. In: *Future Internet*. 16, 180 (2024).
5. Silva, I.S., Sousa, A.M.C.: Integração do ChatGPT para Aprimoramento da Definição de Requisitos de Software: Uma Revisão Sistemática da Literatura. *Trilha de Estudantes da 27 Workshop de Engenharia de Requisitos (WER)*, Buenos Aires (2024).
6. Kitchenham, B., Charters, S.M.: *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. EBSE Technical Report (2007). Disponível em: https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering.
7. Perform Systematic Literature Reviews. Available at: <https://parsif.al/>.
8. Ribeiro, Q., et al.: A View of the Technical, Individual, and Social Dimensions of Sustainable Software Systems: A Systematic Literature Review. In: *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, pp. 1169–1177. ACM, (2024).
9. Zhou, X., et al.: A Map of Threats to Validity of Systematic Literature Reviews in Software Engineering. In: *2016 23rd Asia-Pacific Software Engineering Conference*, IEEE (2016).
10. Binns, R.: Fairness in Machine Learning: Lessons from Political Philosophy. In: *Proceedings of Machine Learning Research* 81:1-11 (2018). Disponível em: <http://arxiv.org/abs/1712.03586v2>.
11. Ronanki, K., Berger, C., Horkoff, J.: Investigating ChatGPT’s Potential to Assist in Requirements Elicitation Processes. *arXiv (Cornell University)*, (2023).