



Um Estudo Experimental sobre o Uso de Realidade Aumentada para Apoiar o Ensino de *Model Needs* na Engenharia de Requisitos para Sistemas de IA

Giovanna Bonilha
Instituto Federal do Amazonas
Manaus, Brasil
2024000350@ifam.edu.br

Fabiann Matthaues
Universidade Federal do Amazonas
Manaus, Brasil
fabianm.barbosa@icomp.ufam.edu.br

Bruna Santos
Universidade Federal do Amazonas
Manaus, Brasil
bruna.santos@icomp.ufam.edu.br

Ayumi Aoki
Universidade Federal do Amazonas
Manaus, Brasil
ayumi.santana@icomp.ufam.edu.br

Bianca Trinkenreich
Colorado State University
Fort Collins, EUA
bianca.trinkenreich@colostate.edu

Tayana Conte
Universidade Federal do Amazonas
Manaus, Brasil
tayana@icomp.ufam.edu.br

Resumo

Contexto: A Engenharia de Requisitos para Sistemas de Inteligência Artificial (RE4AI) apresenta desafios, pois conceitos como *Model Needs* envolvem aspectos de treinamento, aprendizagem e impacto de novas entradas nas previsões, nem sempre explicitados nos requisitos tradicionais. Isso dificulta sua compreensão e aplicação pelos estudantes. Nesse contexto, a Realidade Aumentada (RA) tem sido explorada para apoiar experiências de aprendizagem mais interativas e contextualizadas. **Objetivo:** Este artigo tem como objetivo analisar o impacto de uma abordagem baseada em RA no desempenho dos estudantes na compreensão de *Model Needs* no contexto da Engenharia de Requisitos para IA. **Método:** Foi conduzido um experimento controlado com 48 estudantes de graduação, divididos em dois grupos: um grupo experimental, que utilizou a aplicação RE4AI-AR, e um grupo de controle, que realizou uma atividade equivalente em formato tradicional, utilizando papel. O estudo combinou análises quantitativas e qualitativas. O desempenho foi medido por meio da acurácia, enquanto os dados qualitativos foram analisados com base na Teoria da Carga Cognitiva, a fim de compreender as percepções dos participantes em relação à complexidade, ao esforço e à aprendizagem. **Resultados:** O grupo experimental obteve maior acurácia do que o grupo de controle, com diferenças estatisticamente significativas. Os resultados qualitativos indicam que a abordagem baseada em RA melhorou o engajamento e a compreensão, embora tenham sido relatados alguns problemas de usabilidade e navegação. **Conclusão:** Os resultados sugerem que o uso da Realidade Aumentada pode apoiar efetivamente o ensino de *Model Needs* na Engenharia de Requisitos para IA, contribuindo para reduzir a lacuna entre os conceitos teóricos e sua aplicação prática.

Palavras-chave

RE4AI, *Model Needs*, Realidade Aumentada, Educação em Engenharia de Software, Teoria da Carga Cognitiva

1 Introdução

A crescente adoção de sistemas baseados em Inteligência Artificial (IA) tem transformado o desenvolvimento de software e impactado diretamente práticas tradicionais da Engenharia de Software [7, 21, 44]. Diferentemente dos sistemas convencionais, aplicações

com IA apresentam características como comportamento não determinístico, dependência de dados e adaptação contínua [14, 32], impondo novos desafios técnicos e conceituais ao longo do ciclo de vida do software [9, 45]. Nesse contexto, a Engenharia de Requisitos (ER) torna-se ainda mais relevante, pois decisões inadequadas nessa etapa podem comprometer a confiabilidade, a transparência e o comportamento do sistema [10, 12].

Em sistemas baseados em IA, a ER precisa ir além da especificação tradicional de requisitos funcionais e não funcionais. Ela deve incorporar aspectos relacionados ao comportamento dos modelos, às suas limitações, suposições e aos impactos das decisões automatizadas [3, 23]. Estudos recentes indicam que práticas convencionais de ER apresentam limitações para lidar com a complexidade desses sistemas, especialmente na especificação de requisitos de adaptabilidade [22], explicabilidade [24] e uso de dados [3].

Como resposta às limitações das abordagens tradicionais de ER para sistemas de IA, o *framework* RE4HCAI [5] propõe uma abordagem estruturada para apoiar a eliciação de requisitos em sistemas de IA centrados no ser humano [2, 40]. Entre suas diretrizes, destaca-se a identificação das necessidades dos modelos de IA (*Model Needs*), que contempla características operacionais e requisitos relacionados ao comportamento dos modelos [2].

No entanto, apesar desses avanços técnicos, o ensino desses conceitos ainda apresenta limitações [17, 24]. Em cursos de computação, a ER ainda enfrenta desafios para integrar teoria e prática, especialmente por meio de cenários autênticos que reflitam a complexidade de sistemas reais [19]. Essa limitação dificulta o desenvolvimento de competências para analisar o comportamento dos modelos de IA e compreender as implicações de suas decisões para os usuários e para o sistema [10, 11, 40].

Diante desse cenário, é necessário investigar estratégias pedagógicas que aproximem teoria e prática no ensino de ER para sistemas baseados em IA. Tecnologias imersivas, como a Realidade Aumentada (RA), têm sido exploradas na educação por tornarem conteúdos abstratos mais concretos e interativos, além de favorecerem o engajamento dos estudantes [27, 38]. Associadas à gamificação, essas tecnologias podem proporcionar experiências de aprendizagem mais envolventes e facilitar a compreensão de conceitos complexos por meio de simulações próximas a cenários reais [30, 31].

Neste contexto, este trabalho propõe o uso de uma abordagem baseada em Realidade Aumentada e gamificação para apoiar o ensino de Engenharia de Requisitos para sistemas baseados em IA. Para isso, foi desenvolvido o *RE4AI-AR*, um aplicativo que insere o estudante em cenários interativos nos quais ele assume o papel de engenheiro de requisitos em sistemas inteligentes, sendo desafiado a tomar decisões relacionadas ao comportamento dos modelos.

Como etapa inicial de avaliação da abordagem proposta, este estudo concentra-se na diretriz relacionada às necessidades dos modelos de IA (*Model Needs*). O objetivo é investigar os efeitos do uso do *RE4AI-AR* na compreensão dessa diretriz e na carga cognitiva percebida pelos estudantes durante a atividade. Para isso, foi conduzido um experimento controlado com estudantes do curso de Ciência da Computação, comparando o uso do aplicativo com uma atividade equivalente em formato tradicional.

As principais contribuições deste trabalho são: (i) uma abordagem educacional baseada em Realidade Aumentada e gamificação para o ensino de RE4AI; (ii) o aplicativo *RE4AI-AR* como ferramenta de apoio didático; e (iii) evidências experimentais iniciais, obtidas em um estudo focado na diretriz *Model Needs*, sobre os efeitos da abordagem na aprendizagem e na carga cognitiva percebida pelos estudantes.

2 Fundamentação Teórica

2.1 ER para Sistemas de IA

A Engenharia de Requisitos para sistemas baseados em Inteligência Artificial (RE4AI) difere da ER tradicional devido ao comportamento probabilístico desses sistemas, à dependência da qualidade dos dados [22] e à limitada transparência de seus mecanismos de decisão [5]. Diferentemente dos sistemas determinísticos, soluções baseadas em IA variam conforme o treinamento, o contexto de uso e a evolução dos dados, dificultando atividades como elicitação, análise, modelagem, especificação, validação e evolução de requisitos [6].

Nesse contexto, a RE4AI exige considerar aspectos como disponibilidade e governança dos dados, supervisão humana e robustez diante de erros e falhas [4, 33]. Requisitos não funcionais, como explicabilidade, transparência, justiça, ética e confiança, também se tornam especialmente relevantes [33]. Ahmad et al. [3] destacam que esses aspectos devem ser acompanhados ao longo de todo o ciclo de vida do sistema, pois mudanças nos dados, no contexto de uso ou no comportamento do modelo podem alterar requisitos e expectativas previamente definidos.

Para lidar com essa complexidade, o RE4HCAI, proposto por Ahmad et al. [5], oferece uma estrutura para apoiar a elicitação e a organização de requisitos em sistemas de IA. O *framework* agrupa os requisitos em seis áreas: necessidades do usuário, do modelo e dos dados, *feedback* e controle do usuário, explicabilidade e confiança, e erros e falhas. Essa organização evidencia que a RE4AI envolve desafios técnicos, humanos e sociotécnicos, exigindo abordagens que articulem desempenho algorítmico, necessidades dos usuários e mecanismos de governança e responsabilização [1, 4]. Esses desafios também reforçam a necessidade de estratégias educacionais que preparem estudantes e profissionais para lidar com as especificidades da RE4AI [11, 40].

2.2 Ensino de RE4AI

Os currículos de Engenharia de Software ainda priorizam práticas convencionais de Engenharia de Requisitos, com atenção limitada aos aspectos éticos, humanos e sociotécnicos dos sistemas de IA [1]. Como consequência, estudantes podem ter dificuldade para traduzir preocupações como viés, justiça, responsabilidade, explicabilidade e confiança em requisitos verificáveis [8]. Além disso, é necessário articular conceitos de ER com noções de IA, dados e comportamento de modelos, o que torna mais desafiador o refinamento desses aspectos em artefatos precisos [11, 40].

Diante desses desafios, estudos recentes têm explorado estratégias para apoiar o ensino de RE4AI. Santos et al. [40] relatam que o uso do *framework* RE4HCAI pode estimular a reflexão dos estudantes sobre viés, responsabilidade e explicabilidade em sistemas de IA. No entanto, o estudo também aponta a necessidade de apoio instrucional mais estruturado para facilitar a compreensão desses conceitos. Batista et al. [11] indicam que abordagens baseadas em GORE, como o método KAOS, podem apoiar a elicitação de requisitos para sistemas de IA. Ainda assim, os autores observam dificuldades no refinamento de requisitos detalhados.

Apesar dos avanços iniciais no ensino de RE4AI, persistem desafios na adoção de abordagens pedagógicas que tornem esse conteúdo mais concreto e acessível aos estudantes [11, 40]. Esse cenário reforça a necessidade de tecnologias educacionais que apoiem a visualização, a interação e a contextualização de problemas complexos. Entre essas possibilidades, destacam-se abordagens imersivas baseadas em AR/XR.

2.3 Tecnologias de XR no Ensino de Engenharia de Software e RE4AI

Tecnologias de Realidade Estendida (XR), como Realidade Aumentada (AR), Realidade Virtual (VR) e Realidade Mista (MR), têm sido exploradas na educação por promoverem experiências de aprendizagem mais interativas [36]. Estudos associam essas tecnologias ao aumento do engajamento, da motivação e da compreensão de conceitos complexos [29, 31]. Entretanto, desafios de infraestrutura, capacitação docente e integração curricular ainda limitam sua adoção em larga escala [36].

Na Educação em Engenharia de Software, tecnologias XR têm sido exploradas para apoiar a compreensão de conceitos e o desenvolvimento de habilidades práticas [18]. Na Engenharia de Requisitos, Hua et al. [25] propõem o XR4PD, uma ferramenta baseada em RA para visualizar diagramas de *Problem Frames* em 3D e tornar a análise de requisitos mais intuitiva. Em contextos educacionais, Rodrigues et al. [39] apresentam o ReqJourney, um jogo imersivo para o ensino de ER com resultados positivos em engajamento e desempenho. De modo semelhante, Otemaier et al. [35] exploram a interpretação imersiva de papéis no ensino de entrevistas de elicitação, aproximando os estudantes da prática profissional.

Embora essas iniciativas representem avanços no uso de XR para o ensino de ER, seu foco ainda recai sobre conceitos tradicionais da área. A RE4AI introduz desafios relacionados à qualidade dos dados, explicabilidade, justiça e comportamento adaptativo de sistemas baseados em IA [33], exigindo que os estudantes lidem com questões técnicas, humanas e éticas nem sempre contempladas no ensino convencional de ER [1]. Além disso, apesar do avanço na

integração entre IA e XR, *frameworks* centrados no humano ainda não contemplam plenamente características de ambientes imersivos [15], como presença social, interação multimodal e *embodied interaction* [20]. Nesse contexto, o uso de tecnologias XR voltado especificamente ao ensino de RE4AI permanece pouco explorado.

3 Abordagem de Ensino de RE4AI Adotada

A proposta educacional deste trabalho baseia-se na experiência apresentada por [40], que investigou o uso do *framework* RE4HCAI no ensino de RE4AI. Nesse estudo, os autores adaptaram o RE4HCAI [5] ao contexto educacional por meio de materiais didáticos, enunciados e um *checklist*. O objetivo foi apoiar estudantes de graduação na elicitação de requisitos técnicos, humanos e sociais de sistemas baseados em IA.

Com base nessa experiência, este trabalho amplia a proposta ao explorar a diretriz *Model Needs* em um ambiente apoiado por Realidade Aumentada. Essa escolha foi motivada pelos resultados do estudo anterior, no qual, apesar dos bons índices de compreensão, os estudantes relataram dificuldades com o vocabulário técnico, os tipos de aprendizado, o ajuste do modelo e o uso de *feedback* [40]. Assim, o RE4AI-AR foi desenvolvido para mitigar essas dificuldades por meio de uma experiência interativa e imersiva, aproximando os estudantes da prática da Engenharia de Requisitos para IA.

Durante a atividade, os estudantes interagem com cenários simulados e analisam aspectos centrais do modelo de IA. Um deles é o tipo de treinamento adotado, que distingue modelos estáticos, com atualizações pontuais, de modelos dinâmicos, capazes de se adaptar durante o uso. Essa característica afeta requisitos de atualização, validação, monitoramento e controle da evolução do sistema [5], pois sistemas baseados em IA podem alterar seu comportamento diante de novos dados, interações ou versões do modelo [41, 45].

A proposta aborda diferentes tipos de aprendizado, como supervisionado, não supervisionado e por reforço. A escolha depende do problema, dos dados disponíveis e dos resultados esperados. Além disso, influencia requisitos de dados rotulados, critérios de avaliação e restrições operacionais do modelo [5, 32].

Outro aspecto explorado é a adaptação do modelo com base em *feedback*, no comportamento dos usuários e em novas entradas recebidas pelo sistema. Esses elementos são tratados como requisitos por afetarem a adaptação, a previsibilidade e a confiabilidade do sistema [5]. Nesse sentido, Amershi et al. [9] e Habiba et al. [22] destacam a importância desses mecanismos ao longo do ciclo de vida de sistemas baseados em IA.

Para aproximar esses conceitos da prática, a abordagem organiza essas decisões em um cenário simulado de elicitação de requisitos. Nele, o estudante interage com personagens, analisa situações relacionadas ao comportamento do modelo de IA e toma decisões associadas às suas necessidades. Com base nessa proposta, foi desenvolvido o RE4AI-AR, um aplicativo educacional em Realidade Aumentada que concretiza a diretriz *Model Needs*.

4 RE4AI-AR

O RE4AI-AR é um aplicativo para dispositivos Android desenvolvido para apoiar o ensino de conceitos de ER para sistemas baseados em Inteligência Artificial, com ênfase inicial na diretriz *Model Needs*. O aplicativo utiliza *Unity 2022.3 LTS*, *AR Foundation 5.2.2* e o *plugin*

ARCore XR, que permitem sobrepor elementos virtuais ao ambiente físico por meio da câmera do dispositivo móvel e proporcionar uma experiência imersiva e interativa [28]. Sua proposta central é aproximar o estudante de um contexto simulado de trabalho em ER para sistemas de IA, tornando a aprendizagem de conceitos abstratos mais concreta, experiencial e contextualizada.

4.1 Arquitetura do Sistema

O sistema adota uma arquitetura em camadas que separa responsabilidades de interação, processamento lógico, persistência de dados e suporte pedagógico, conforme ilustrado na Figura 1. Essa organização favorece a modularidade e permite sua evolução para contemplar outras diretrizes do RE4HCAI.



Figura 1: Arquitetura do RE4AI-AR.

A arquitetura é composta por quatro camadas principais. A primeira corresponde à **interface do estudante**, responsável pela interação com o usuário e pela execução da experiência em Realidade Aumentada. Essa camada reúne componentes como *login*, seleção de missões, visualização do *ranking*, ambiente imersivo em RA, interface de questões e exibição de *feedback*. A segunda é a **camada de lógica da aplicação**, que concentra os principais mecanismos do sistema. Ela é formada por três módulos: (i) o *motor de interação*, que gerencia o fluxo das missões, controla a progressão do usuário e coordena os diálogos entre os personagens; (ii) o *módulo pedagógico*, que implementa as diretrizes do RE4HCAI, gerencia as questões, avalia as respostas e gera *feedback* pedagógico; e (iii) o *módulo de gamificação*, responsável pelo cálculo da pontuação, pela aplicação das regras de combo e pela manutenção dos *rankings*.

A terceira camada corresponde à **camada de persistência**, implementada com o *Firebase Firestore*. Nela são armazenados dados dos usuários, progresso nas missões, respostas, pontuações e histórico de tentativas, além das informações de autenticação e das regras de segurança para controle de acesso. A quarta camada corresponde à **interface do docente**, representada por um painel administrativo que permite o acompanhamento do desempenho dos estudantes, a visualização de métricas por questão e o gerenciamento de turmas e conteúdos. Embora também constitua uma interface de apresentação, essa camada é separada da *interface do estudante* por atender a usuários, objetivos e funcionalidades distintos.

O funcionamento do sistema ocorre de forma integrada entre as quatro camadas. Inicialmente, o estudante realiza o *login* e seleciona

uma missão na **interface do estudante**. A partir dessa ação, o sistema aciona o módulo de Realidade Aumentada, que detecta o marcador físico e renderiza o ambiente virtual com os personagens. Em seguida, o motor de interação controla o fluxo da experiência, dispara eventos e conduz a narrativa. Quando necessário, o módulo pedagógico apresenta questões e avalia as respostas do estudante. Os resultados são processados pelo módulo de gamificação, que calcula a pontuação e atualiza o *ranking*. Todas as interações e os dados gerados durante a execução são armazenados na **camada de persistência** e posteriormente utilizados pela **interface docente** para análise e acompanhamento do desempenho.

4.2 Interação do Aluno no jogo

Ao acessar o aplicativo, o estudante informa seu nome e seleciona a turma à qual pertence. Em seguida, é direcionado à tela de missões disponíveis. Ao selecionar a primeira missão, recebe um *briefing* com o seguinte contexto narrativo: ele acaba de ser contratado como engenheiro de requisitos de uma *startup* que desenvolve, com Inteligência Artificial, um sistema inteligente de planejamento de refeições sustentáveis. Seu objetivo é participar da reunião de definição do modelo.

A experiência imersiva começa quando o estudante aponta a câmera do dispositivo para um marcador impresso de 20 × 20 cm, usado como âncora para projetar o ambiente virtual. A partir desse momento, uma sala de reunião tridimensional é renderizada sobre o ambiente físico, permitindo a interação com três personagens que representam diferentes papéis no processo de Engenharia de Requisitos: o gerente, responsável por apresentar o projeto e conduzir a reunião; o cliente, que expõe as necessidades do sistema e os requisitos esperados; e o robô, que representa o componente de Inteligência Artificial do sistema em desenvolvimento. Essa dinâmica insere o estudante em um cenário próximo ao contexto real de elicitação de requisitos. Exemplos dessa interação são apresentados na Figura 2, que ilustra o ambiente virtual e a interação com os personagens durante a atividade.

Após a interação com os personagens, o estudante responde cinco questões de múltipla escolha elaboradas com base nas diretrizes do *framework* RE4HCAI [5], abordando o conceito central da missão. As questões tratam de decisões relevantes para a especificação de modelos de Inteligência Artificial em sistemas reais, como a definição da abordagem de aprendizado, a forma de utilização do *feedback* do usuário, o mecanismo de adaptação do modelo e o impacto das novas entradas no comportamento do sistema. Para cada resposta selecionada, os personagens reagem de forma diferenciada conforme o acerto ou erro do estudante, reforçando o caráter interativo e contextualizado da experiência.

Ao término da missão, o aplicativo exibe a pontuação obtida pelo estudante, calculada com base no número de acertos em uma escala de zero a cinco pontos, com bônus extra para respostas corretas consecutivas e *feedback* personalizado de acordo com o desempenho alcançado. O modo de revisão permite que o estudante visualize suas respostas em comparação com as corretas, favorecendo a reflexão sobre o conteúdo abordado. O aplicativo também disponibiliza um sistema de *ranking* geral e por turma, incentivando o engajamento por meio de elementos de gamificação [37].



(a) Ambiente virtual em RA (b) Interação com personagens

Figura 2: Exemplo da interação do usuário com o RE4AI-AR em ambiente de Realidade Aumentada.

4.3 Suporte ao Docente e Tecnologias

Para apoiar a atuação docente, o sistema oferece um painel administrativo protegido por autenticação, que permite visualizar o desempenho dos estudantes por questão, missão e turma. O painel também possibilita gerenciar turmas, missões e conteúdos, além de exportar os dados para análise.

Os dados gerados durante o uso do aplicativo são armazenados em nuvem por meio do *Firebase Firestore*, garantindo sua persistência e disponibilidade. O desenvolvimento do aplicativo utilizou o **Blender** para modelagem e animação dos personagens tridimensionais, o **Figma** para prototipação e design das interfaces e o **Visual Studio Code** para desenvolvimento e edição do código-fonte.

5 Experimento Controlado

Este estudo foi conduzido por experimento controlado para investigar os efeitos do aplicativo *RE4AI-AR* no ensino de Engenharia de Requisitos para sistemas baseados em Inteligência Artificial. Além do desempenho dos estudantes, analisou-se o impacto da abordagem na carga cognitiva percebida, com base na *Cognitive Load Theory (CLT)*, proposta por Sweller et al. [43], que considera as limitações da memória de trabalho durante a aprendizagem.

O planejamento e a execução do experimento seguiram as diretrizes de Wohlin et al. [46], amplamente adotadas em Engenharia de Software Experimental. O objetivo do estudo foi estruturado segundo o modelo GQM (*Goal Question Metric*), conforme apresentado a seguir:

Analisar <o uso do aplicativo *RE4AI-AR* em comparação com uma atividade equivalente realizada em papel>

Com o propósito de <avaliar o potencial de uma abordagem baseada em Realidade Aumentada para apoiar o ensino de RE4AI e reduzir dificuldades na compreensão da diretriz *Model Needs*>

Com relação ao <desempenho dos estudantes na compreensão de *Model Needs* e à carga cognitiva percebida durante a atividade>

Do ponto de vista <dos pesquisadores em Engenharia de Software>

No contexto <de uma disciplina de Engenharia de Software cursada por estudantes de graduação em Ciência da Computação>.

5.1 Hipóteses

Com base na questão de pesquisa, foram definidas as seguintes hipóteses, considerando a acurácia como a proporção de respostas corretas em relação ao total de questões:

- **H0:** Não há diferença na acurácia da compreensão de *Model Needs* entre os estudantes que utilizaram o *RE4AI-AR* e aqueles que realizaram a atividade equivalente em papel.
- **H1:** Há diferença na acurácia da compreensão de *Model Needs* entre os estudantes que utilizaram o *RE4AI-AR* e aqueles que realizaram a atividade equivalente em papel.

5.2 Contexto

O experimento foi conduzido em contexto acadêmico, no âmbito da disciplina de Engenharia de Software I de um curso de graduação em Ciência da Computação. Segundo Shull et al. [42], o primeiro estudo destinado a avaliar e aprimorar uma nova tecnologia deve ser um estudo de viabilidade. Seu objetivo é verificar se a tecnologia atende aos requisitos gerais para os quais foi concebida, além de gerar hipóteses iniciais sobre sua utilidade e seu potencial.

A abordagem investigada tem natureza educacional e busca apoiar o processo de ensino-aprendizagem [26]. Para isso, utiliza a Realidade Aumentada para inserir os estudantes em situações próximas a contextos reais de tomada de decisão, reduzindo a abstração dos conceitos e favorecendo uma experiência interativa, contextualizada e imersiva.

Assim, ambientes de sala de aula são adequados à condução desse tipo de estudo, pois permitem observar a interação dos participantes com a tecnologia em um cenário autêntico de aprendizagem [47]. A integração da ferramenta à dinâmica da disciplina possibilita avaliar seu potencial como recurso pedagógico complementar, alinhado aos objetivos instrucionais [13].

5.3 Seleção de Variáveis

A variável independente foi a abordagem utilizada na atividade avaliativa, definida em dois tratamentos: uso do aplicativo *RE4AI-AR* e realização de uma atividade tradicional em papel. A variável dependente foi o desempenho dos estudantes na compreensão de *Model Needs*, mensurado pela acurácia das respostas ao questionário. A acurácia de cada participante foi calculada da seguinte forma:

$$Acurácia = \frac{\sum_{n=1}^X}{\sum_{n=1}^Y} \times 100$$

Onde:

- X = soma total de acertos do participante;
- Y = número total de questões (5).

As respostas corretas foram definidas com base no conteúdo do *framework* RE4HCAI, que constituiu o oráculo do estudo. Cada questão recebeu pontuação binária: 1,0 para resposta correta e 0,0 para incorreta, pois todas eram de múltipla escolha e apresentavam uma única alternativa correta.

5.4 Seleção de Participantes

O estudo contou com a participação voluntária de 48 estudantes do curso de graduação em Ciência da Computação, matriculados na disciplina de Engenharia de Software I. Todos assinaram o Termo de Consentimento Livre e Esclarecido (TCLE), em conformidade com as diretrizes éticas para pesquisas com seres humanos. O estudo foi submetido e aprovado por um Comitê de Ética em Pesquisa.

Os participantes também responderam a um formulário de caracterização para identificar sua experiência prévia em desenvolvimento de software, Inteligência Artificial e Engenharia de Requisitos para sistemas baseados em IA. A Tabela 1 resume o perfil dos participantes.

Tabela 1: Perfil dos participantes

Característica	N
Nenhuma experiência com DS ou IA	39
Experiência com desenvolvimento de software	5
Experiência com projetos de IA	1
Experiência com ambos	0
Contato prévio com requisitos para sistemas de IA	4

A escolha por estudantes de graduação como participantes permitiu conduzir a avaliação no público-alvo da tecnologia, uma vez que, além de representarem esse perfil, podem apresentar habilidades semelhantes às de profissionais com menor experiência, sendo, portanto, participantes válidos para estudos dessa natureza [16].

5.5 Design Experimental

O design aplicado neste experimento foi de um fator com dois tratamentos. O fator corresponde ao tipo de atividade realizada pelo estudante, e os dois tratamentos são: o uso do aplicativo *RE4AI-AR* pelo grupo experimental e a realização de uma atividade tradicional em papel pelo grupo controle.

A divisão dos participantes nos grupos foi realizada em duas etapas. Para os estudantes que declararam possuir algum nível de experiência prévia com desenvolvimento de software ou IA, foi aplicado o princípio de *blocking*, distribuindo-os igualmente entre os dois grupos com o objetivo de eliminar sistematicamente o efeito indesejado da experiência prévia na comparação entre os tratamentos [34, 46]. Para os demais participantes, sem experiência prévia declarada, a distribuição foi realizada de forma aleatória. Cada grupo foi composto por 24 participantes, e cada participante foi submetido a apenas um tratamento.

5.6 Instrumentação

Os seguintes artefatos foram utilizados na condução do experimento:

- **Termo de Consentimento Livre e Esclarecido (TCLE):** documento que apresentava os objetivos do estudo, a natureza voluntária da participação e o uso dos dados para fins acadêmicos e científicos;
- **Formulário de caracterização:** questionário aplicado antes da atividade com o objetivo de coletar dados sobre o perfil

dos participantes, incluindo experiência prévia com desenvolvimento de software, Inteligência Artificial e requisitos para sistemas de IA;

- **Aplicativo RE4AI-AR:** instrumento utilizado pelo grupo experimental, contendo o cenário narrativo interativo e cinco questões de múltipla escolha sobre *Model Needs*;
- **Atividade tradicional em papel:** instrumento utilizado pelo grupo de controle, contendo o mesmo cenário narrativo em formato textual e as mesmas cinco questões de múltipla escolha;
- **Questionário qualitativo:** instrumento aplicado aos dois grupos após a atividade para coletar percepções sobre a experiência, incluindo compreensão do conteúdo, interação com a atividade e carga cognitiva percebida.

5.7 Preparação

Todos os participantes receberam uma aula teórica de aproximadamente 45 minutos sobre *Model Needs* e suas diretrizes no *framework* RE4HCAI. Ao final, foi realizado um exercício prático, seguido da correção e discussão coletiva das respostas, para consolidar a compreensão do conteúdo antes da atividade experimental.

Antes do estudo, foram conduzidos testes piloto com dois estudantes concluintes do curso de Ciência da Computação para refinar os materiais experimentais e validar a duração estimada das atividades, incluindo a entrevista.

5.8 Execução

A execução do experimento contou com seis pesquisadores como moderadores, três em cada sala. Eles apresentaram as instruções gerais da atividade e esclareceram dúvidas, sem orientar os participantes sobre o conteúdo das questões ou as respostas esperadas.

Em seguida, os participantes foram distribuídos entre os dois tratamentos. As atividades foram realizadas individual e simultaneamente, sem limite máximo de tempo.

O grupo experimental utilizou o aplicativo *RE4AI-AR* em dispositivos Android. Para isso, foi disponibilizado um *QR code* para instalação, permitindo que cada participante utilizasse o próprio smartphone. Quando necessário, a equipe de pesquisa forneceu dispositivos adicionais. O grupo de controle realizou a atividade em formato tradicional, com o mesmo cenário narrativo apresentado em texto. Para garantir a comparabilidade entre os tratamentos, ambos os grupos responderam ao mesmo conjunto de questões de múltipla escolha.

Durante a execução, os moderadores ofereceram apenas suporte operacional, sem interferir nas tarefas. Ao final, todos os participantes responderam a um formulário de pós-avaliação sobre a utilidade das abordagens, a intenção de uso futuro, os aspectos positivos e negativos e as sugestões de melhoria.

5.9 Pré-processamento

Após o experimento, os dados dos dois grupos foram organizados em uma planilha eletrônica. As respostas do grupo experimental foram extraídas do *Firebase Firestore*¹, enquanto as do grupo de controle, coletadas em papel, foram corrigidas e registradas manualmente na mesma planilha.

¹<https://firebase.google.com/>

Durante a análise, identificou-se que alguns participantes do grupo experimental iniciaram a atividade mais de uma vez sem concluí-la, gerando registros incompletos no banco de dados. Nesses casos, apenas a tentativa concluída foi considerada, e os demais registros foram descartados. Com base nos dados coletados, foram realizadas análises quantitativas e qualitativas.

6 Resultados e Discussões

6.1 Desempenho dos Participantes

A Tabela 2 resume a caracterização dos participantes e a acurácia obtida em cada grupo. Observou-se que a maioria possuía pouca ou nenhuma experiência prévia em desenvolvimento de software ou em projetos com IA. Além disso, grande parte informou não ter contato anterior com atividades de definição ou análise de requisitos para sistemas baseados em IA.

Tabela 2: Caracterização dos participantes

Característica	Nº (%)
Sem experiência em software/IA	42 (87,5%)
Com experiência em software/IA	6 (12,5%)
Sem contato prévio com RE4AI	42 (87,5%)
Com contato prévio com RE4AI	6 (12,5%)
Total	48 (100%)

Cada participante respondeu a cinco questões de múltipla escolha sobre a diretriz *Model Needs*, com pontuação 1 para respostas corretas e 0 para incorretas. Assim, a acurácia individual variou de 0% a 100%. A Figura 3 apresenta a distribuição da acurácia em cada grupo.

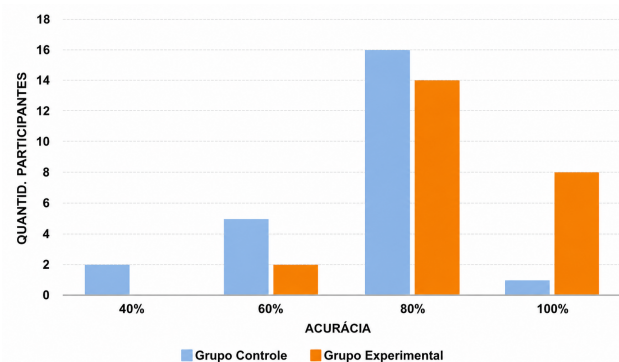


Figura 3: Frequência da acurácia dos participantes por grupo

A partir da distribuição apresentada, observa-se que ambos os grupos apresentaram desempenho predominantemente elevado, com maior concentração de participantes na faixa de 80% de acerto. No entanto, o grupo experimental (*RE4AI-AR*) apresentou maior frequência de desempenhos máximos (100%) e menor ocorrência de valores intermediários (60%) em comparação com o grupo de controle (papel).

O grupo de controle apresentou maior dispersão nos resultados, com maior presença de desempenhos intermediários e baixos, enquanto o grupo experimental apresentou uma distribuição mais concentrada nas faixas superiores de acurácia. Ao calcular a acurácia geral de cada grupo, verificou-se que o grupo experimental obteve média de 85%, enquanto o grupo de controle apresentou média de 74,17%, resultando em uma diferença aproximada de 10,83 pontos percentuais.

Foi aplicado o teste de Shapiro-Wilk para verificar a normalidade dos dados de acurácia. Os resultados indicaram violação do pressuposto de normalidade ($W = 0,902$, $p < 0,001$), evidenciando que os dados não seguem uma distribuição normal. Diante disso, optou-se pelo uso do teste não paramétrico de Mann-Whitney U para comparar os grupos. A análise estatística foi conduzida com as ferramentas *JASP 0.96*² e *RStudio 2026.04.0+526*³, considerando um nível de significância de $\alpha = 0,05$.

Os dados do experimento são apresentados na Tabela 3, enquanto a Figura 4 apresenta o *boxplot* com a distribuição da acurácia nos grupos de controle e experimental.

Tabela 3: Estatísticas descritivas da acurácia por grupo

Grupo	N.	Mediana	25 Percentil	75 Percentil
Controle	24	0,80	0,60	0,80
Experimental	24	0,80	0,80	1,00

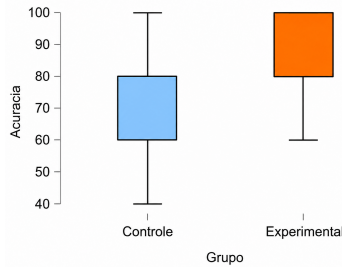


Figura 4: Boxplot da distribuição da acurácia

A partir do *boxplot*, observa-se que ambos os grupos apresentam medianas semelhantes, indicando desempenho típico equivalente. No entanto, o grupo experimental concentra mais valores nas faixas superiores de acurácia e apresenta menor ocorrência de desempenhos intermediários e baixos que o grupo de controle.

A comparação das duas amostras pelo teste de Mann-Whitney indicou diferença estatisticamente significativa entre os grupos ($U = 160,0$, $p = 0,003$). Esse resultado leva à rejeição da hipótese nula (H_0), indicando diferença significativa no desempenho dos participantes. Além da significância estatística, calculou-se o tamanho de efeito por meio da correlação bisserial de postos (*rank-biserial correlation*), apropriada ao teste de Mann-Whitney. O valor obtido foi $r = 0,444$, indicando efeito de magnitude moderada.

²<https://jasp-stats.org/>

³<https://dailies.rstudio.com/version/2026.04.0+526/>

6.2 Percepções dos Participantes

Além da análise de desempenho apresentada anteriormente, os formulários aplicados ao final da atividade foram analisados para compreender a experiência de uso da abordagem proposta, com ênfase nos aspectos cognitivos envolvidos na realização da tarefa. A análise fundamentou-se na Teoria da Carga Cognitiva (*Cognitive Load Theory* – CLT) [43], apresentada na Seção 5. As respostas foram interpretadas considerando as três dimensões da carga cognitiva: intrínseca, extrínseca e relevante [43].

Para isso, os participantes responderam a um questionário composto por questões em escala *Likert* de cinco pontos e questões abertas destinadas à coleta de opiniões, dificuldades e sugestões. As questões em escala *Likert* apoiaram a análise de aspectos relacionados à complexidade percebida, ao esforço cognitivo, à dificuldade de compreensão e ao estímulo ao raciocínio, conforme apresentado na Tabela 4. As questões abertas, por sua vez, permitiram aprofundar a compreensão das experiências relatadas pelos participantes.

Tabela 4: Questões utilizadas na avaliação das percepções dos participantes

Q	Questão
Q1	Quão complexo você considerou o problema apresentado nesta atividade?
Q2	Quanto seu conhecimento prévio em Engenharia de Software contribuiu para a resolução da atividade?
Q3	O quanto a forma de apresentação do cenário dificultou sua compreensão do problema?
Q4	Quão difícil foi compreender os elementos e informações apresentados no cenário?
Q5	O quanto a atividade contribuiu para seu raciocínio na identificação de necessidades em um contexto prático?
Q6	O quanto o esforço mental que você teve durante a atividade contribuiu para seu aprendizado?
Q7	Avalie seu conhecimento percebido sobre <i>Model Needs</i> após a atividade.

Inicialmente, os comentários das respostas abertas foram transcritos e organizados para análise. Em seguida, as respostas foram interpretadas com base nas dimensões da carga cognitiva, considerando aspectos relacionados à percepção de complexidade e compreensão, ao esforço cognitivo percebido e à percepção de aprendizagem. Também foram analisados relatos sobre dificuldades, limitações e sugestões de melhoria da abordagem proposta.

A percepção de complexidade e compreensão foi analisada em relação à carga cognitiva intrínseca, refletindo o nível de dificuldade inerente ao conteúdo. O esforço cognitivo percebido foi discutido sob a perspectiva da carga cognitiva extrínseca, evidenciando possíveis dificuldades associadas à forma de apresentação da atividade. Já a percepção de aprendizagem foi interpretada com base na carga cognitiva relevante, indicando o potencial da atividade para apoiar a construção do conhecimento.

6.2.1 Percepção de Complexidade e Compreensão. A percepção de complexidade da atividade (Q1) indicou que, de forma geral, os participantes não consideraram o problema excessivamente complexo.

Já as respostas relacionadas à contribuição do conhecimento prévio em Engenharia de Software (Q2) evidenciaram que os participantes reconheceram a importância desse conhecimento para apoiar a resolução da atividade. Em ambos os grupos, a maior parte das respostas concentrou-se em baixos níveis de complexidade percebida e altos níveis de concordância quanto à contribuição do conhecimento prévio, conforme apresentado na Figura 5.

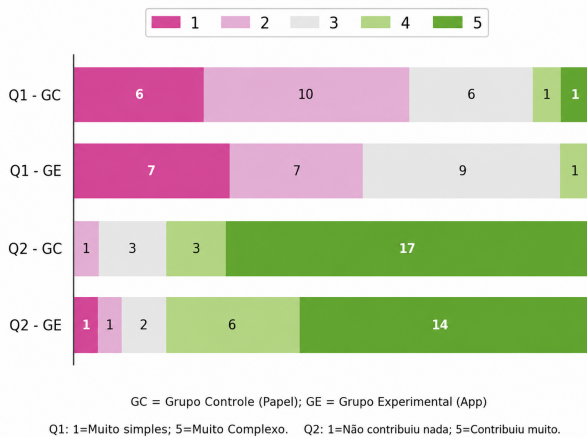


Figura 5: Frequência das respostas relacionadas à carga cognitiva intrínseca por grupo

Os resultados relacionados à percepção de complexidade indicam que a carga cognitiva intrínseca associada ao conteúdo não representou um obstáculo significativo ao desempenho dos participantes. Ainda assim, foram observados relatos pontuais de dificuldades relacionadas à distinção entre tipos de modelos e formas de aprendizado. No grupo controle, por exemplo, o participante P8 afirmou: “ainda tenho um pouco de dificuldade nas diferenças entre *feedbacks* e *tipos de modelo*”. Esses resultados reforçam que, apesar da atividade ter sido percebida como acessível, determinados aspectos da diretriz *Model Needs* ainda demandaram maior compreensão conceitual.

6.2.2 Esforço Cognitivo Percebido. A análise das questões relacionadas à dificuldade de compreensão do cenário (Q3 e Q4) e ao esforço mental (Q6), apresentadas na Figura 6, demonstra que, de maneira geral, os participantes não perceberam a atividade como mentalmente exigente. A maioria das respostas indicou baixos níveis de dificuldade e esforço, sugerindo baixa incidência de carga cognitiva extrínseca durante a realização da atividade.

Entretanto, alguns participantes relataram dificuldades específicas relacionadas à forma de apresentação da atividade, especialmente no grupo que utilizou RA. Foram identificados problemas como limitações técnicas e dificuldades de navegação, conforme evidenciado em relatos como o participante P25: “o AR não funcionou direito” e o participante P38: “o cenário virtual é confuso”. Por outro lado, no grupo que realizou a atividade em formato tradicional (Papel), alguns participantes relataram dificuldades relacionadas à compreensão e abstração do cenário quando apresentado exclusivamente em texto. Como exemplo, o participante P12 afirmou: “Achei que, por ser apenas em texto, ficou mais difícil imaginar o cenário e entender melhor o problema”.

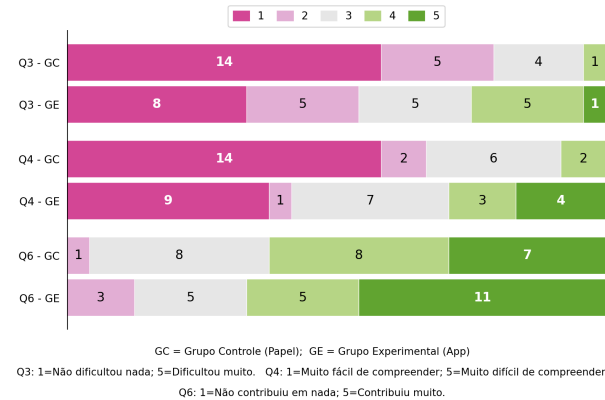


Figura 6: Frequência das respostas relacionadas à carga cognitiva extrínseca por grupo

6.2.3 Percepção de Aprendizagem. As respostas relacionadas ao estímulo ao raciocínio (Q5), ao esforço mental voltado ao aprendizado (Q6) e ao conhecimento percebido após a atividade (Q7), apresentadas na Figura 7, indicam resultados consistentes de aprendizagem significativa.

Diversos participantes que utilizaram a abordagem baseada em RA relataram que a atividade contribuiu diretamente para a compreensão dos conceitos. Relatos dos participantes P26, P27 e P37 ilustram essa percepção ao indicar maior compreensão do problema, engajamento com a atividade e aproximação com contextos mais realistas. Os participantes afirmaram, respectivamente, que a abordagem “ajudou muito a identificar as necessidades”, “ajudou muito por simular uma situação real de reunião com cliente” e que “a interação com um sistema ajudou a entender situações mais práticas”.

Embora participantes do grupo Papel também tenham apontado contribuições para o aprendizado, seus relatos enfatizaram principalmente a clareza textual e a organização das informações, destacando que a atividade “tornou fácil a identificação de requisitos” e que “o texto é bem claro para identificar a resposta”. Em contraste, os relatos do grupo baseado em RA enfatizaram com maior frequência aspectos relacionados à contextualização prática, interação com o cenário e aproximação com situações mais próximas da prática profissional.

6.2.4 Percepções sobre RA e Papel. A análise das respostas abertas relacionadas à influência da forma de apresentação da atividade buscou compreender como os participantes perceberam a abordagem adotada durante a identificação das necessidades no cenário.

No grupo que utilizou RA, foram observados relatos positivos associados à interação com a atividade e à contextualização prática dos conceitos. Conforme destacado pelo participante P37, “a interação com um sistema ajudou a entender situações mais práticas”, enquanto o participante P40 afirmou que “a abordagem interativa foi muito interessante e deu vontade de responder e entender o assunto”. Esses relatos sugerem que a abordagem contribuiu para aproximar os conceitos discutidos em sala de situações mais práticas e contextualizadas.

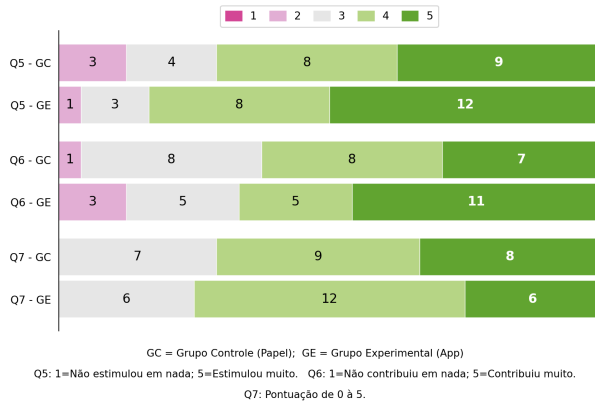


Figura 7: Frequência das respostas relacionadas à carga cognitiva relativa por grupo

No grupo controle (Papel), os participantes destacaram principalmente a clareza textual e a organização das informações apresentadas no cenário. O participante P8 mencionou que “*o texto é bem claro para identificar a resposta*”, enquanto o participante P10 afirmou que “*a linguagem clara facilitou a identificação das necessidades*”. Esses relatos sugerem que a descrição textual utilizada na atividade contribuiu para a compreensão do conteúdo e para a identificação dos elementos presentes no cenário.

6.2.5 Limitações e Sugestões de Melhoria. A análise das respostas abertas buscou identificar dificuldades de compreensão e sugestões de melhoria relacionadas à forma de apresentação da atividade. Os relatos foram organizados em dois eixos principais: conceituais e operacionais.

No eixo conceitual, foram identificados desafios na compreensão de conceitos específicos em ambos os grupos, especialmente quanto aos tipos de aprendizado e ao comportamento dos modelos. Esse aspecto é evidenciado no relato do participante P19: “*tive dificuldade em diferenciar os tipos de aprendizado*”. Essa dificuldade sugere a necessidade de maior aprofundamento ou apoio didático na apresentação desses conceitos.

No grupo experimental, foram identificadas limitações operacionais relacionadas à interface e à navegação do *RE4AI-AR*, particularmente no contexto da utilização da RA. Entre os principais pontos, destacam-se a ausência de mecanismos que permitam revisar conteúdos anteriores e dificuldades na visualização do cenário. Tais limitações são ilustradas nos relatos dos participantes P29, que sugere que “*um botão para voltar seria melhor*”, e P46, que menciona “*não consegui visualizar partes do texto*”.

As sugestões dos participantes concentraram-se em melhorias relacionadas à interface do *RE4AI-AR*, maior interatividade e aprimoramento da navegação. Entre os aspectos mencionados, destacam-se melhorias na visualização dos elementos, uso de cenários 3D alternativos à RA e maior controle da navegação, como revisar perguntas e retornar a informações anteriores.

7 Ameaças à Validade

7.1 Validade Interna

Foram identificadas três ameaças internas: (1) efeito do treinamento, (2) classificação de experiência e (3) familiaridade com tecnologia móvel. Para mitigar a ameaça (1), todos os participantes receberam a mesma aula teórica e realizaram o mesmo exercício prático antes da divisão dos grupos. Em relação à ameaça (2), a experiência prévia foi identificada por autoavaliação no formulário de caracterização e balanceada entre os grupos por meio de *blocking*. Já a ameaça (3) refere-se à familiaridade dos participantes com dispositivos Android, o que pode ter influenciado os resultados do grupo experimental independentemente do conteúdo abordado, sendo reconhecida como limitação do estudo.

7.2 Validade Externa

Em relação à validade externa, foram identificadas duas ameaças: (1) representatividade da amostra e (2) representatividade do cenário. Quanto à ameaça (1), o estudo foi conduzido com uma única turma de graduação em Engenharia de Software, limitando a generalização dos resultados para outros contextos acadêmicos e perfis de estudantes. Ainda assim, ambientes acadêmicos são adequados para investigações voltadas à aprendizagem em contexto real de ensino [42]. Em relação à ameaça (2), o cenário utilizado é fictício, embora inspirado em contextos reais de desenvolvimento de sistemas inteligentes. Dessa forma, os resultados devem ser interpretados como evidências iniciais de viabilidade.

7.3 Validade de Conclusão

Neste experimento, foi considerada a seguinte questão como ameaça à validade de conclusão: (1) Tamanho da amostra. O número de participantes, embora adequado para um estudo de viabilidade, pode não ser suficiente do ponto de vista estatístico para permitir a generalização dos resultados. Por esse motivo, os resultados obtidos são interpretados como indícios de viabilidade, e não como conclusões definitivas, sendo esperada a replicação do estudo em contextos mais amplos para consolidação das evidências.

7.4 Validade de Constructo

Referente à validade de constructo, foram consideradas duas ameaças: (1) uso de uma única métrica quantitativa para avaliação do desempenho e (2) exposição de cada participante a apenas um dos tratamentos experimentais. Em relação à ameaça (1), a acurácia foi utilizada como principal métrica para avaliar a compreensão da diretriz *Model Needs*. Embora adequada para análise de desempenho, essa métrica não contempla aspectos como confiança e tempo de resposta. Para complementar essa limitação, também foram utilizados questionários em escala *Likert* e questões abertas. Em relação à ameaça (2), cada participante foi submetido a apenas um dos tratamentos experimentais, o que reduz efeitos de contaminação entre grupos, mas impede comparações intraindividuais.

8 Considerações Finais

Este artigo apresentou um experimento controlado com o objetivo de avaliar o uso do aplicativo *RE4AI-AR*, baseado em Realidade Aumentada e gamificação, como recurso de apoio ao ensino do

conceito de *Model Needs* em uma disciplina de Engenharia de Requisitos para sistemas de Inteligência Artificial, em comparação a uma atividade equivalente realizada em papel.

A análise de acurácia indicou que o grupo experimental apresentou acurácia média superior à do grupo controle, com uma diferença de aproximadamente 10,83 pontos percentuais entre os grupos. A comparação entre as amostras por meio do teste de Mann-Whitney revelou diferenças estatisticamente significativas, levando à rejeição da hipótese nula H_0 e indicando que o uso do *RE4AI-AR* influenciou positivamente o desempenho dos estudantes na compreensão do conceito abordado. Além disso, o tamanho de efeito apresentou magnitude moderada, reforçando a relevância prática da diferença observada.

As percepções dos participantes do grupo experimental sugerem que o *RE4AI-AR* contribuiu para a compreensão do conteúdo ao aproximar os conceitos de situações práticas e contextualizadas. Os relatos destacaram aspectos relacionados à interação com a atividade, engajamento e aproximação com cenários mais próximos da prática. No entanto, também foram identificadas limitações relacionadas à interface e à navegação do aplicativo, como dificuldades na visualização do cenário em Realidade Aumentada e ausência de mecanismos para revisar conteúdos anteriores.

Em relação às percepções dos participantes do grupo controle, verificou-se que a atividade em papel foi avaliada positivamente no que diz respeito à clareza textual e à organização das informações. Os participantes destacaram que a linguagem clara facilitou a identificação das necessidades do sistema. No entanto, alguns participantes relataram dificuldade em abstrair o problema a partir de uma apresentação exclusivamente textual, o que pode ter limitado a contextualização prática do conteúdo. Um dos diferenciais do *RE4AI-AR* está justamente na proposta de apresentar os cenários de forma mais interativa e contextualizada, buscando aproximar os conceitos discutidos em sala de situações mais próximas da prática.

Quanto ao valor do *RE4AI-AR* como recurso educacional, a hipótese dos pesquisadores é que o aplicativo pode ser especialmente útil em duas situações: (1) para estudantes em fase inicial de contato com Engenharia de Requisitos para sistemas de IA, que ainda não desenvolveram familiaridade com conceitos abstratos como *Model Needs*; e (2) em disciplinas que buscam aproximar o ensino teórico de contextos práticos, oferecendo uma experiência simulada que complementa as abordagens tradicionais.

Como trabalhos futuros, pretende-se conduzir um plano incremental de avaliação do *RE4AI-AR*, contemplando: (i) a replicação do experimento controlado com uma amostra mais ampla e diversificada, além de estudos longitudinais; (ii) o desenvolvimento de novas missões voltadas a outras diretrizes do framework *RE4HCAI*, a fim de avaliar a generalização dos resultados obtidos com *Model Needs*; (iii) avaliações específicas de usabilidade e experiência do usuário; e (iv) avaliações com docentes sobre a utilidade do painel administrativo no acompanhamento do desempenho e na identificação de dificuldades de aprendizagem.

Dessa maneira, os resultados deste artigo contribuem para pesquisas sobre o uso de tecnologias imersivas no ensino de Engenharia de Requisitos para sistemas de Inteligência Artificial. Além disso, o *RE4AI-AR* apresenta potencial para apoiar o ensino de conceitos do framework *RE4HCAI* no contexto da formação de engenheiros de software para sistemas inteligentes.

Disponibilidade de Artefatos

Organizamos os conjuntos de dados, a ferramenta *RE4AI-AR* e os resultados deste estudo no material suplementar, disponível em: <https://doi.org/10.5281/zenodo.22151219>

Declaração Sobre Uso de Inteligência Artificial

Os autores utilizaram o ChatGPT (versão GPT-5.5) como ferramenta de apoio na organização textual, revisão gramatical, aprimoramento da clareza da escrita e revisão estrutural do artigo. Todo o conteúdo produzido com auxílio da ferramenta foi revisado, validado e editado pelos autores, que se responsabilizam integralmente pelas informações, interpretações e conclusões apresentadas no trabalho.

Agradecimentos

Agradecemos a todos os participantes do estudo experimental e aos membros do Grupo de Pesquisa USES pelo apoio. Gostaríamos de agradecer pelo suporte financeiro concedido pelo CNPq 314797/2023-8; CNPq 443934/2023-1; CNPq 445029/2024-2; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (AUXPE-CAPES-PROEX) - Código de Financiamento 001; Fundação de Amparo à Pesquisa do Estado do Amazonas - FAPEAM - por meio dos projetos PDPG-CAPES e POSGRAD 2025-2026 e POSGRAD 2026-2027.

Referências

- [1] Silvia Abrahão, John Grundy, Mauro Pezzè, Margaret-Anne Storey, and Damian A Tamburri. 2025. Software Engineering by and for Humans in an AI Era. *ACM Transactions on Software Engineering and Methodology* 34, 5 (2025), 1–46.
- [2] Khlood Ahmad, Mohamed Abdelrazek, Chetan Arora, Arbind Agrahari Baniya, Muneera Bano, and John Grundy. 2023. Requirements engineering framework for human-centered artificial intelligence software systems. *Applied Soft Computing* 143 (2023), 110455.
- [3] Khlood Ahmad, Mohamed Abdelrazek, Chetan Arora, Muneera Bano, and John Grundy. 2023. Requirements engineering for artificial intelligence systems: A systematic mapping study. *Information and Software Technology* 158 (2023), 107176.
- [4] Khlood Ahmad, Mohamed Abdelrazek, Chetan Arora, Muneera Bano, and John Grundy. 2023. Requirements practices and gaps when engineering human-centered Artificial Intelligence systems. *Applied Soft Computing* 143 (2023), 110421.
- [5] Khlood Ahmad, Mohamed Abdelrazek, Chetan Arora, John Grundy, and Muneera Bano. 2023. Requirements Elicitation and Modelling of Artificial Intelligence Systems: An Empirical Study. (2023).
- [6] Khlood Ahmad, Muneera Bano, Mohamed Abdelrazek, Chetan Arora, and John Grundy. 2021. What's up with requirements engineering for artificial intelligence systems?. In *2021 IEEE 29th International Requirements Engineering Conference (RE)*. IEEE, 1–12.
- [7] Iftekhar Ahmed, Aldeida Aleti, Haipeng Cai, Alexander Chatzigeorgiou, Pinjia He, Xing Hu, Mauro Pezzè, Denys Poshyvanyk, and Xin Xia. 2025. Artificial intelligence for software engineering: The journey so far and the road ahead. *ACM Transactions on Software Engineering and Methodology* 34, 5 (2025), 1–27.
- [8] Mahir Akgun and Hadi Hosseini. 2025. AI education in a mirror: Challenges faced by academic and industry experts. In *International Conference on Artificial Intelligence in Education*. Springer, 336–348.
- [9] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 291–300.
- [10] Nagadiyya Balasubramaniam, Marjo Kauppinen, Antti Rannisto, Kari Hiekkänen, and Sari Kujala. 2023. Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology* 159 (2023), 107197.
- [11] Beatriz Braga Batista, Márcia Sampaio Lima, and Tayana Uchôa Conte. 2024. Teaching requirements engineering for AI: A goal-oriented approach in software engineering courses. In *Proceedings of the XXIII Brazilian Symposium on Software Quality*. 613–623.

- [12] Hrvoje Belani, Marin Vukovic, and Željka Car. 2019. Requirements engineering challenges in building AI-based complex systems. In *2019 IEEE 27th international requirements engineering conference workshops (REW)*. IEEE, 252–255.
- [13] Angeliki Boltzi, Konstantinos Kalovrektis, Apostolos Xenakis, Periklis Chatzimisios, and Costas Chaikalis. 2024. Digital tools, technologies, and learning methodologies for education 4.0 frameworks: A STEM oriented survey. *Ieee Access* 12 (2024), 12883–12901.
- [14] Jan Bosch, Helena Holmström Olsson, and Ivica Crnkovic. 2021. Engineering ai systems: A research agenda. *Artificial intelligence paradigms for smart cyber-physical systems* (2021), 1–19.
- [15] Andrea Antonio Cantone, Alireza Mortezaipoor, Otino Pio Santosuosso, Aurelio Sepe, Simone Spera, and Giuliana Vitiello. 2025. A Glimpse on HCAI Frameworks: A Scoping Review to Adapt them to Mixed Reality. In *2025 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXR/AINe)*. IEEE, 178–183.
- [16] J. Carver, L. Jaccheri, S. Morasca, and F. Shull. 2004. Issues in Using Students in Empirical Studies in Software Engineering Education. In *Proceedings of the 5th International Workshop on Enterprise Networking and Computing in Healthcare Industry*. IEEE, Sydney, NSW, Australia, 239–249. doi:10.1109/METRIC.2003.1232471
- [17] Marian Daun, Alicia M Grubb, Viktoria Stenkova, and Bastian Tenbergen. 2023. A systematic literature review of requirements engineering education. *Requirements Engineering* 28, 2 (2023), 145–175.
- [18] Gustavo Vargas de Andrade, André Luiz Cordeiro Gomes, Felipe Rohr Hoinoski, Marília Guterres Ferreira, Pablo Schoeffel, and Adilson Vahldick. 2022. Virtual reality applications in software engineering education: A systematic review. *arXiv preprint arXiv:2204.12008* (2022).
- [19] Maria Alice de Souza Macedo, Carla Bezerra, and Emanuel Coutinho. 2024. Uma Pesquisa Qualitativa do Contexto de Ensino em Requisitos de Software no Brasil. In *Workshop sobre Educação em Computação (WEL)*. SBC, 669–679.
- [20] Paul Dourish. 2001. *Where the action is: the foundations of embodied interaction*. MIT press.
- [21] Usman Khan Durrani, Mustafa Akpınar, Muhammed Fatih Adak, Abdullah Talha Kabakus, Muhammed Maruf Öztürk, and Mohammed Saleh. 2024. A decade of progress: A systematic literature review on the integration of AI in software engineering phases and activities (2013-2023). *IEEE Access* 12 (2024), 171185–171204.
- [22] Umm-e Habiba, Markus Haug, Justus Bogner, and Stefan Wagner. 2024. How mature is requirements engineering for AI-based systems? A systematic mapping study on practices, challenges, and future research directions. *Requirements Engineering* 29, 4 (2024), 567–600.
- [23] Hans-Martin Heyn, Eric Knauss, Amna Pir Muhammad, Olof Eriksson, Jennifer Linder, Padmini Subbiah, Shameer Kumar Pradhan, and Sagar Tungal. 2021. Requirement engineering challenges for ai-intense systems development. In *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)*. IEEE, 89–96.
- [24] Jennifer Horkoff. 2019. Non-functional requirements for machine learning: Challenges and new directions. In *2019 IEEE 27th international requirements engineering conference (RE)*. IEEE, 386–391.
- [25] Yang Hua, Hongbin Xiao, Wenhao Wang, and Zhi Li. 2025. XR4PD: Augmented Reality for Visualising Problem Diagrams in Requirements Analysis. In *Proceedings of the 16th International Conference on Internetware*. 147–151.
- [26] Anghelo Josué, MC Bedoya-Flores, EF Mosquera-Quiñonez, Ángel E Mesías-Simisterra, and JV Bautista-Sánchez. 2023. Educational platforms: Digital tools for the teaching-learning process in education. *Ibero-American Journal of Education & Society Research*, 3 (1), 259–263. *Courses in Network Era* (2023).
- [27] Dorota Kamińska, Grzegorz Zwoliński, Anna Laska-Leśniewicz, Rui Raposo, Mário Vairinhos, Elisabeth Pereira, Frane Urem, Martina Ljubić Hinić, Rain Eric Haamer, and Gholamreza Anbarjafari. 2023. Augmented reality: Current and new trends in education. *Electronics* 12, 16 (2023), 3531.
- [28] Victor Klisman, Luan Souza Marques, João Pedro de Lima, Leonardo Marques, Genildo Gomes, Tayana Conte, and Bruno Gadelha. 2021. LogMe: An Application for Generating Logs in Immersive Interactions for UX Evaluation. In *Proceedings of the 23rd International Conference on Enterprise Information Systems (ICEIS)*, Vol. 2. SciTePress, 549–556. doi:10.5220/0010460105490556
- [29] Yiannis Koumpourous. 2024. Revealing the true potential and prospects of augmented reality in education. *Smart Learning Environments* 11, 1 (2024), 2.
- [30] Mohammad Amin Kuhail, Areej ElSayary, Shahbano Farooq, and Ahlam Alghamdi. 2022. Exploring immersive learning experiences: A survey. In *Informatics*, Vol. 9. MDPI, 75.
- [31] Georgios Lampropoulos, Euclid Keramopoulos, Konstantinos Diamantaras, and Georgios Evangelidis. 2022. Augmented reality and gamification in education: A systematic literature review of research, applications, and empirical studies. *Applied sciences* 12, 13 (2022), 6809.
- [32] Silverio Martínez-Fernández, Justus Bogner, Xavier Franch, Marc Oriol, Julien Siebert, Adam Trendowicz, Anna Maria Vollmer, and Stefan Wagner. 2022. Software engineering for AI-based systems: a survey. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 31, 2 (2022), 1–59.
- [33] Mariana Crisostomo Martins, Livia Mancine C Campos, João Lucas R Soares, Taciana Novo Kudo, and Renato F Bulcão-Neto. 2025. Requirements engineering for machine learning-based ai systems: A tertiary study. *Journal of Software Engineering Research and Development* 13, 2 (2025), 13–129.
- [34] Maria Alcimar C. Meireles, Sabrina Rocha, Jose Carlos Duarte, Tayana Conte, and Jose Carlos Maldonado. 2022. Evaluating Approaches to Selecting Design Thinking Techniques: Quantitative and Qualitative Analysis. In *Proceedings of the Brazilian Symposium on Software Quality (SBQS)*. ACM, Curitiba, Brazil. doi:10.1145/3571473.3571482
- [35] Kelly Rafaela Otemaier, Regina Albuquerque, Rodolfo Diniz, Sheila Reinehr, and Andreia Malucelli. 2024. Immersive role-playing: An experience report on a promising approach to learning requirements elicitation. In *Proceedings of the XXIII Brazilian Symposium on Software Quality*. 586–595.
- [36] Rafaela Otemaier, Regina Albuquerque, Ericson Savio Falabretti, Sheila Reinehr, and Andreia Malucelli. 2025. Teaching in the Immersive Era: Professors' Perspectives on XR in Higher Education. In *ACM International Conference on Interactive Media Experiences Workshops (IMXw)*. SBC, 150–154.
- [37] Laís Paiva, Moisés Barbosa, Solano Oliveira, Max Barbosa, Victor Klisman, José Carlos Duarte, Genildo Gomes, Bruno Gadelha, and Tayana Conte. 2022. To the Next Level! an Exploratory Study on the Influence of User Experience on the Acceptance of a Gamified Learning Platform. In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU)*. SciTePress, 281–288. doi:10.5220/0011067900003182
- [38] P Janardhana Kumar Reddy and G Singaravelu. 2020. Augmented reality (AR): the new trend in transforming teaching and learning in education. *International Journal of Analytical and Experimental Modal Analysis* 12, 4 (2020), 620–626.
- [39] Claudia Susie C Rodrigues, Thiago S Nobre, Vitoria Nazareth, and Cláudia Werner. 2026. ReqJourney: Engaging Students with a 2D Metaverse Serious Game for Software Requirements Education. *Journal on Interactive Systems* 17, 1 (2026), 214–224.
- [40] Bruna Santos, Márcia Lima, Márcio Ribeiro, and Tayana Conte. 2025. Teaching Requirements Engineering for Human-Centered AI: A Classroom Experience with the RE4HCAI Framework. In *Anais do XXIV Simpósio Brasileiro de Qualidade de Software (SBQS)*. Sociedade Brasileira de Computação, São José dos Campos, SP, Brazil, 540–550. doi:10.5753/sbqs.2025.15264
- [41] David Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-Francois Crespo, and Dan Dennison. 2015. Hidden technical debt in machine learning systems. *Advances in neural information processing systems* 28 (2015).
- [42] Forrest Shull, Jeffrey Carver, and Guilherme H. Travassos. 2001. An Empirical Methodology for Introducing Software Processes. *ACM SIGSOFT Software Engineering Notes* 26, 5 (2001), 288–296. doi:10.1145/503271.503248
- [43] John Sweller, Jeroen JG Van Merriënboer, and Fred Paas. 2019. Cognitive architecture and instructional design: 20 years later. *Educational psychology review* 31, 2 (2019), 261–292.
- [44] Valerio Terragni, Annie Vella, Partha Roop, and Kelly Blincoe. 2025. The future of ai-driven software engineering. *ACM Transactions on Software Engineering and Methodology* 34, 5 (2025), 1–20.
- [45] Andreas Vogelsang and Markus Borg. 2019. Requirements engineering for machine learning: Perspectives from data scientists. In *2019 IEEE 27th international requirements engineering conference workshops (REW)*. IEEE, 245–251.
- [46] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, Anders Wesslén, et al. 2012. *Experimentation in software engineering*. Vol. 236. Springer.
- [47] Zehui Zhan, Qianyi Wu, Zhihua Lin, and Jiayi Cai. 2021. Smart classroom environments affect teacher-student interaction: Evidence from a behavioural sequence analysis. *Australasian Journal of Educational Technology* 37, 2 (2021), 96–109.