

Uma Abordagem Baseada em Inteligência Artificial Generativa, PBL e Metodologias Ágeis para o Ensino de Modelagem de Sistemas

Cristina Becker M. Nabarro ^{1,2}, André Gomes Regino^{3,4,5}, Rodrigo Bonacin^{3,4,5},
Júlio Cesar Dos Reis ⁶

¹ Centro Universitário Campo Limpo Paulista (Unifaccamp), Campo Limpo Paulista, SP, Brasil

² Faculdade de Tecnologia do Estado de São Paulo (FATEC), Atibaia, SP, Brasil

³ Centro de Tecnologia da Informação Renato Archer (CTI), Campinas, SP, Brasil

⁴ Centro Universitário de Jaguariúna (UniFAJ) – Jaguariúna, SP – Brasil

⁵ Centro Universitário Max Planck (UniMAX) – Indaiatuba, SP – Brasil

⁶ IC - Universidade Estadual de Campinas (Unicamp), Campinas, SP, Brasil

crisbeckermatos@hotmail.com; {aregino, rbonacin}@cti.gov.br,
jreis@ic.unicamp.br

Abstract. *Teaching Unified Modeling Language (UML) modeling requires students to operate at a high level of abstraction, which often results in artifacts with semantic inconsistencies. This article analyzes the guided use of Generative Artificial Intelligence in UML teaching within an undergraduate course, combining Project-Based Learning with agile-inspired practices. The study was conducted with 22 Software Engineering students over a series of iterative sprints. The evaluation considered both the quality of the artifacts produced, as assessed by rubrics, and the students' perceptions, as assessed by a questionnaire. The results show good performance in functional modeling activities, especially requirements elicitation, and greater difficulty with structural modeling, particularly class diagrams.*

Resumo. *O ensino de modelagem na Linguagem de Modelagem Unificada (UML) exige dos estudantes um alto nível de abstração, o que frequentemente resulta em artefatos com inconsistências semânticas. Este artigo apresenta uma proposta de uso orientado de Inteligência Artificial generativa no ensino de UML em uma disciplina de graduação, articulando Aprendizagem Baseada em Projetos e práticas inspiradas em metodologias ágeis. Um estudo foi realizado com 22 estudantes de Engenharia de Software, ao longo de sprints iterativos. A avaliação considerou a qualidade dos artefatos produzidos por meio de rubricas e a percepção dos estudantes por meio de um questionário. Os resultados mostram bom desempenho nas atividades de modelagem funcional, especialmente na elicitação de requisitos, e maiores dificuldades na modelagem estrutural, com destaque para o diagrama de classes.*

1. Introdução

O ensino de Engenharia de Software envolve desafios, especialmente no desenvolvimento de habilidades de abstração necessárias à modelagem de sistemas com a Linguagem de Modelagem Unificada (UML - *Unified Modeling Language*) (Canedo, Oliveira Jr. e Tives, 2025). Frequentemente, estudantes de graduação apresentam dificuldades na transição entre requisitos descritos em linguagem natural e sua representação formal, o que resulta em artefatos com inconsistências semânticas e baixo alinhamento ao domínio do problema (Cámara *et al.*, 2023; Di Rocco *et al.*, 2025). Usualmente, o uso de metodologias ativas, como a Aprendizagem Baseada em Projetos (PBL - *Project-Based Learning*), tem sido uma estratégia eficaz para reduzir esse distanciamento entre teoria e prática (Hayashi e Ferreira, 2025). Podendo PBL ser associada a metodologias ágeis para ensinar habilidades ligadas às demandas do mercado de engenharia de software (Meireles e Bonifácio, 2015).

Nos últimos anos, ferramentas de Inteligência Artificial (IA) generativa têm sido utilizadas com mais frequência em atividades educacionais, ampliando as possibilidades de apoio ao ensino, mas também trazendo desafios para o seu uso pedagógico. Ferramentas como ChatGPT, Claude, DeepSeek e Gemini têm demonstrado capacidade de apoiar a produção de modelos, atuando como assistentes que fornecem feedback imediato e sugestões de design. No entanto, a literatura alerta para riscos, como a aceitação acrítica de modelos incorretos e a dependência tecnológica, o que reforça a necessidade de um uso orientado e estruturado pedagogicamente (Cámara *et al.*, 2023).

Embora o tema venha recebendo mais atenção, ainda são poucos os estudos que observam esse uso ao longo do tempo, em situações reais, no ensino de disciplinas de graduação. Conforme destacado por Farias *et al.* (2024), é importante realizar uma análise aprofundada do uso de modelos de linguagem de grande escala (LLM - *Large Language Model*), visando orientar práticas pedagógicas e promover uma integração mais eficaz. No entanto, as pesquisas atuais focam majoritariamente em avaliações técnicas da capacidade dos modelos, deixando de aprofundar na compreensão da aprendizagem e na evolução da qualidade dos artefatos ao longo de ciclos iterativos de desenvolvimento.

Nesse contexto, este artigo apresenta uma proposta de uso orientado da IA generativa, integrada a um arranjo pedagógico baseado em PBL e em metodologias ágeis, bem como um estudo com alunos na prática. O objetivo é avaliar como essa integração influencia a qualidade dos artefatos de modelagem produzidos e a percepção de aprendizagem dos estudantes.

As principais contribuições deste trabalho podem ser sintetizadas em três elementos centrais: (1) este estudo apresenta um arranjo pedagógico para o ensino de modelagem UML que articula PBL, ciclos iterativos inspirados em metodologias ágeis e o uso orientado de ferramentas de IA generativa, buscando apoiar os estudantes em uma etapa reconhecidamente desafiadora da formação inicial em Engenharia de Software, a passagem dos requisitos textuais para sua representação em modelos; (2) o artigo apresenta evidências empíricas obtidas em contexto real de sala de aula, a partir de sua aplicação com estudantes de graduação ao longo de diferentes etapas do processo de modelagem; (3) combina a avaliação dos artefatos produzidos com a percepção discente sobre a experiência de aprendizagem, permitindo examinar, de forma articulada, tanto a

qualidade dos modelos quanto a experiência formativa associados ao uso orientado da IA generativa.

O restante deste artigo está organizado da seguinte maneira: a Seção 2 apresenta a fundamentação do artigo e os principais trabalhos relacionados, bem como o diferencial e a contribuição desta pesquisa; a Seção 3 detalha a metodologia aplicada nesta pesquisa; a Seção 4 apresenta os resultados e a discussão desses resultados; e, por fim, a Seção 5 apresenta as considerações finais, as limitações e os trabalhos futuros.

2. Fundamentação e Trabalhos Relacionados

Estudos recentes têm apontado que modelos de linguagem podem apoiar atividades de modelagem de software, especialmente na geração inicial de artefatos UML a partir de descrições em linguagem natural. No campo mais técnico, Cámara *et al.* (2023) estabeleceram uma linha de base para a análise do uso de LLMs em tarefas de modelagem UML e OCL, evidenciando tanto o potencial de apoio quanto limitações importantes, como a variabilidade entre execuções, a sensibilidade ao domínio e a recorrência de inconsistências estruturais.

Combemale, Gray e Rumpe (2023) discutem diferentes formas de incorporar ferramentas de IA à modelagem de software. Os autores abordam desde a automação até o papel de assistente ao modelador, ressaltando que sua utilidade depende do contexto de uso e da forma como são integradas ao processo.

Em estudos voltados à qualidade dos artefatos, De Bari *et al.* (2024) compararam diagramas de classes produzidos por estudantes e por LLMs, observando desempenho competitivo em aspectos sintáticos, mas persistentes limitações semânticas. Garaccione *et al.* (2025), ao investigarem diagramas de casos de uso, também identificaram resultados promissores nas tarefas iniciais de modelagem, embora com fragilidades em relações mais específicas.

Ferrari, Abualhaija e Arora (2024) mostraram que a geração automática de diagramas depende fortemente da qualidade dos requisitos de entrada, reforçando a necessidade de revisão e refinamento ao longo do processo. De modo geral, esses estudos mostram resultados mais positivos nas etapas iniciais da modelagem. Já em artefatos mais complexos, persistem importantes limitações conceituais.

No campo educacional, os resultados apresentados também apontam para o potencial da IA generativa, mas com ressalvas. Jackson, Liu e Van der Hoek (2025) relatam benefícios na ideação e no apoio à produção. Já Daun e Brings (2023) destacam que os efeitos educacionais dessas ferramentas não se restringem à sua capacidade técnica, mas também estão relacionados à forma como são integradas pedagogicamente, contextualizadas didaticamente e acompanhadas de critérios estruturados que subsidiem a avaliação docente. Albonico e Varela (2023), em contexto brasileiro, também observam contribuições no apoio inicial à modelagem e no esclarecimento de dúvidas, ao mesmo tempo em que alertam para riscos de dependência e de redução do esforço cognitivo.

O estudo de Murad *et al.* (2025) contribui ao apresentar evidências empíricas, obtidas por meio de um experimento controlado, de que o uso do ChatGPT pode favorecer o ensino de modelagem em UML. O trabalho avança em relação a estudos mais técnicos ou exploratórios, pois não se limita a avaliar a capacidade da ferramenta de gerar modelos, mas também investiga seus efeitos sobre o desempenho dos

estudantes, considerando eficácia, eficiência e ganho de aprendizagem. Assim, o artigo reforça a relevância de examinar o uso de LLMs em contextos educacionais reais, especialmente no ensino de Engenharia de Software.

Apesar dos avanços recentes, a literatura ainda apresenta limitações importantes. Predominam estudos com foco técnico, relatos de experiência ou avaliações pontuais de artefatos, com menor incidência de investigações educacionais sistemáticas e metodologicamente estruturadas. Também permanecem mais escassos os trabalhos que articulam, no mesmo delineamento, o ensino de UML, a organização didática em PBL, os ciclos iterativos inspirados em metodologias ágeis e a análise conjunta do desempenho técnico e da percepção do estudante. A presente pesquisa visa contribuir para preencher esta lacuna, examinando a aprendizagem de modelagem aplicada em uma disciplina de graduação, considerando não apenas os artefatos gerados, mas também a dinâmica formativa na qual esses artefatos são construídos, revistos e consolidados.

3. Metodologia

Trata-se de um estudo aplicado, descritivo-exploratório, conduzido em contexto real de sala de aula, com predominância quantitativa e incorporação complementar de evidências qualitativas. A intervenção ocorreu em uma unidade da FATEC do Estado de São Paulo, na disciplina de Engenharia de Software I do curso de Tecnologia em Desenvolvimento de Software.

A amostra foi composta por 22 estudantes, com idade média de 26 anos, que não possuíam formação prévia em modelagem UML, o que permitiu observar o processo de aprendizagem conceitual desde os estágios iniciais.

O docente foi responsável pelo planejamento da atividade, pela definição dos objetivos de aprendizagem, pela apresentação dos conceitos de UML, pela orientação sobre o uso da IA, pelo estabelecimento dos critérios das rubricas e pela realização do feedback formativo. Também competiu ao docente intervir diante de erros conceituais, apoiar a discussão das alternativas e orientar a validação semântica dos artefatos.

A intervenção pedagógica foi organizada com base na Aprendizagem Baseada em Projetos, articulada em três ciclos iterativos inspirados em práticas ágeis. Ocorreu ao longo de cinco semanas, com quatro horas-aula cada. Cada sprint compreendeu quatro momentos: a apresentação do problema, a elaboração inicial dos artefatos com apoio orientado das ferramentas de IA generativa, a revisão crítica das sugestões geradas e o refinamento do artefato a partir do feedback docente e dos critérios estabelecidos nas rubricas.

Nesse processo, as ferramentas ChatGPT e Gemini foram usadas como apoio à interpretação do problema e à ideação inicial, enquanto os estudantes precisavam revisar criticamente as sugestões recebidas e justificar suas decisões de design. Conforme apresenta a Figura 1, a dinâmica foi estruturada em três *sprints*:

- *Sprint 0* - preparação e diagnóstico, os estudantes receberam um cenário simulado e revisaram conceitos introdutórios relacionados à elicitación de requisitos e à modelagem UML. Nessa etapa, foram apresentadas as ferramentas de IA generativa ChatGPT e Gemini, explorando seu uso como apoio à construção de artefatos UML. De forma orientada e colaborativa, a turma elaborou requisitos funcionais, diagrama de casos de uso, protótipos

de baixa fidelidade e diagrama de classes, analisando a coerência das sugestões geradas, a fundamentação empregada na elaboração dos prompts e a adequação dos resultados gerados no PlantUML. O *sprint* não recebeu atribuição de nota e teve como finalidade familiarizar os estudantes com o processo de elaboração, análise crítica, revisão e refinamento dos artefatos.

- *Sprint 1* - concentrou-se na modelagem funcional, incluindo a elicitação de requisitos. Artefatos de um cenário do PBL intitulado "IT Solutions – Controle de Eventos" foram elaborados individualmente pelos estudantes, com a IA como parte do processo de desenvolvimento. O *Sprint 1* foi inicialmente realizado com o levantamento de requisitos funcionais e a associação a regras de negócio; posteriormente, desenvolveu-se o Diagrama de Casos de Uso.
- *Sprint 2* - focou na modelagem estrutural e de interface. Os estudantes produziram o Diagrama de Classes e os artefatos dos protótipos de baixa fidelidade, mantendo o alinhamento com os requisitos do *Sprint 1*.

O desenvolvimento dos artefatos com o uso da IA generativa não apenas foi permitido, mas também foi metodologicamente orientado. A instrução reforça que a IA é como um "par colaborativo", uma vez que estudos mostram bons resultados no aprendizado colaborativo de UML entre pares de estudantes (Azevedo *et al.*, 2019). Assim, propõe-se que a IA seja utilizada como apoio à interpretação dos enunciados, à proposição de cenários e à validação da lógica dos artefatos, desencorajando a reprodução automática de respostas e incentivando a análise e a revisão crítica. Essa abordagem foi adotada desde o *Sprint 0*. A IA ampliou as possibilidades de interação e de experimentação durante a atividade, mas não substituiu a mediação pedagógica, a avaliação docente nem a construção conceitual realizada pelos estudantes.

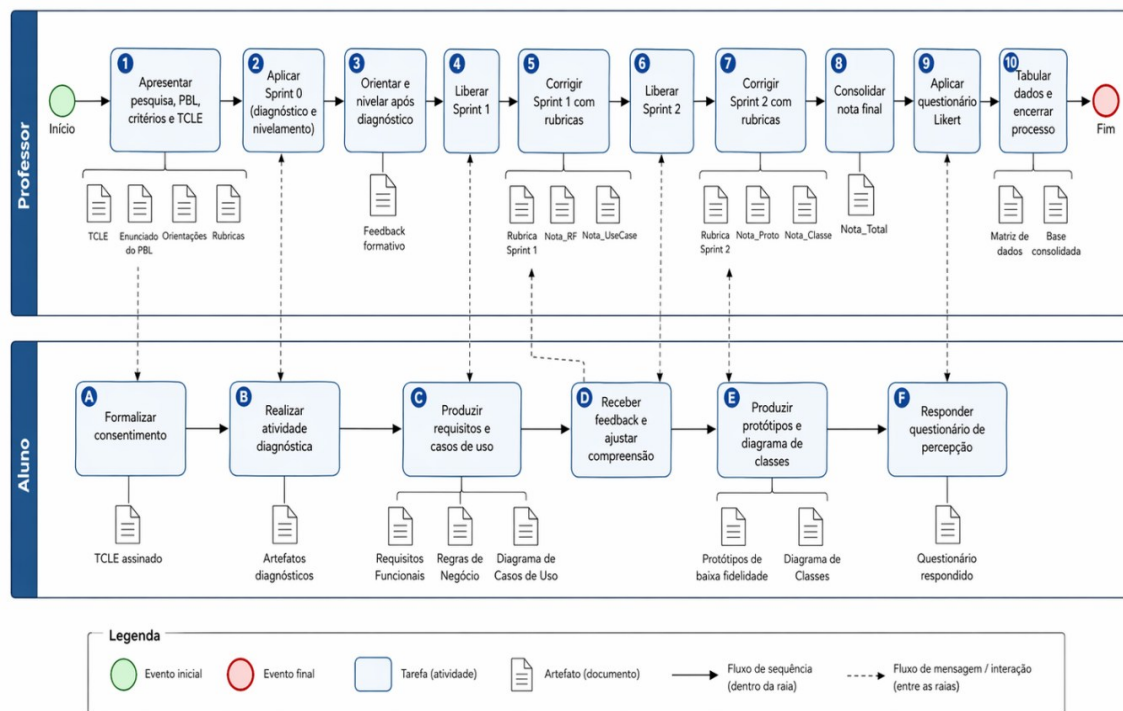


Figura 1 – Processo de Aplicação do PBL. Fonte: Elaborado pelos autores (2026)

A avaliação dos resultados foi realizada por meio da triangulação de dois instrumentos principais descritos nos próximos parágrafos.

A dimensão técnica foi quantificada por meio da aplicação de rubricas estruturadas para avaliar cada artefato produzido pelos estudantes, elaboradas conforme os objetivos de cada sprint. Essa definição buscou tornar mais explícitos os critérios de correção e reduzir a subjetividade da avaliação, considerando tanto os aspectos técnicos da modelagem quanto a aderência ao cenário proposto e às regras de negócio apresentadas no problema.

No *Sprint 1*, a rubrica foi organizada em torno de dois artefatos centrais, a identificação dos requisitos funcionais e o diagrama de casos de uso. Na avaliação dos requisitos funcionais, foram observados aspectos relacionados à abrangência das funcionalidades identificadas, à clareza e precisão de sua formulação e à associação com as regras de negócio do domínio. No caso do diagrama de casos de uso, consideraram-se a identificação dos atores principais, a definição dos casos de uso mais relevantes, a coerência dos relacionamentos estabelecidos e a adequação à notação UML.

No *Sprint 2*, a avaliação incluiu o protótipo da funcionalidade e o diagrama de classes. Para o protótipo, foram analisados elementos relacionados ao fluxo de informações, à organização dos formulários, à navegação entre telas e à aderência aos requisitos e às regras de negócio do cenário. Já no diagrama de classes, observaram-se a identificação das classes e de seus atributos, as associações e suas multiplicidades, a atribuição de operações coerentes às classes e a representação das principais regras de negócio na modelagem.

De modo geral, as rubricas foram estruturadas em níveis de desempenho, com pontuação atribuída conforme o grau de adequação do artefato aos critérios definidos. Essa organização permitiu avaliar, de forma mais consistente, a qualidade dos modelos produzidos pelos estudantes ao longo da atividade.

A dimensão subjetiva do processo foi capturada por meio de um questionário em escala Likert de 5 pontos, respondido ao término da intervenção. Este instrumento foi concebido para avaliar construtos como a percepção de aprendizagem, a utilidade da IA em tarefas específicas e o nível de reflexão crítica dos alunos. Os dados obtidos foram submetidos à análise estatística descritiva e correlacional (coeficiente de correlação de Pearson), com o objetivo de investigar a convergência entre o desempenho técnico real e a experiência formativa relatada.

Os itens foram agrupados em dimensões que dialogam com as etapas da intervenção, incluindo:

- *Pré-Conhecimento (D0 – PRE_UML)*: Mapeia a autoconfiança inicial do aluno em relação à UML antes do desenvolvimento da atividade avaliativa continuada.
- *Qualidade Percebida (D1 – APR_UML)*: Mensura o quanto o estudante acredita que seus artefatos ficaram adequados e o quanto sente que aprendeu.
- *Apoio da IA (D2 – IA_S1 e D3 – IA_S2)*: Avalia a utilidade percebida da IA em tarefas específicas de cada *Sprint*.

- *Criticidade (D4 – CRIT_IA)*: Investiga a postura do aluno, se houve revisão crítica das respostas da IA e se houve ganho de autonomia.
- *Visão Sistêmica (D5 – GLOB_IA_PBL)*: Avalia a percepção sobre a eficácia da metodologia combinada (PBL + Ágil + IA).

A condução desta pesquisa pautou-se pelo rigor ético aplicável a pesquisas envolvendo seres humanos, em estrita observância às diretrizes do Sistema CEP/Conep (Resoluções CNS nº 466/2012 e nº 510/2016), bem como aos preceitos de proteção de dados pessoais estabelecidos pela Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709/2018). Dado que o estudo ocorreu em ambiente educacional, com a coleta de dados de desempenho e de percepção discente, as ações foram orientadas para mitigar riscos, salvaguardar a autonomia dos participantes e assegurar o sigilo das informações em todas as etapas do ciclo de pesquisa. A adesão dos participantes foi precedida por uma explicação presencial sobre os propósitos, os protocolos e as implicações da investigação. Na ocasião, procedeu-se à leitura conjunta e comentada do Termo de Consentimento Livre e Esclarecido (TCLE).

4. Apresentação e Discussão dos Resultados

A análise quantitativa fundamentou-se na articulação de dois eixos de evidências: indicadores objetivos de desempenho, aferidos por meio das rubricas estruturadas detalhadas anteriormente, e dimensões perceptivas extraídas do instrumento aplicado aos participantes.

Essas dimensões, apresentadas na Figura 2, constituem escores consolidados, resultantes do agrupamento de itens por afinidade conceitual (Dimensões A–F), abrangendo: (A) aprendizagem percebida; (B) PBL e apoio da IA aos estudantes; (C) esforço e retrabalho; (D) confiança na modelagem; (E) qualidade percebida dos artefatos; e (F) reflexão crítica.

Conforme apresentado na Figura 2, as médias das dimensões de percepção avaliadas situaram-se consistentemente acima do ponto médio da escala (3,0). Destacam-se os indicadores de aprendizagem percebida em UML (3,43), o nível de reflexão promovido pela ferramenta (3,61) e a percepção global da metodologia (3,63). A consistência interna do instrumento mostrou-se elevada (α geral = 0,955), o que confere robustez à interpretação dos escores agregados e reduz a probabilidade de variabilidade aleatória, corroborando a aceitação positiva da dinâmica pedagógica proposta.

A análise do desempenho dos estudantes mostrou um resultado geral satisfatório, com média final de 7,08 (desvio padrão = 1,37). No entanto, os resultados variaram conforme o tipo de artefato produzido.

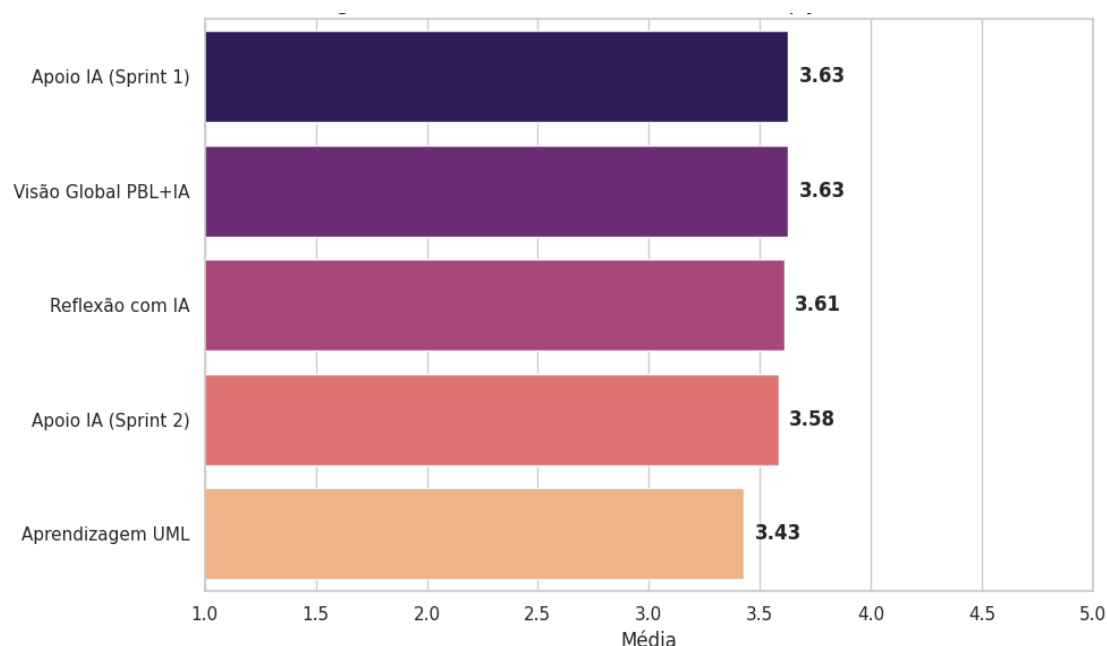


Figura 2 - Média das Dimensões de Percepção. Fonte: Elaborado pelos autores (2026).

Os requisitos funcionais apresentaram o melhor aproveitamento relativo (83,3%), como mostra a Figura 3, e a menor dispersão (Figura 4) entre os estudantes (desvio padrão = 0,32), indicando que a IA foi mais útil nas tarefas iniciais de organização e formulação textual.

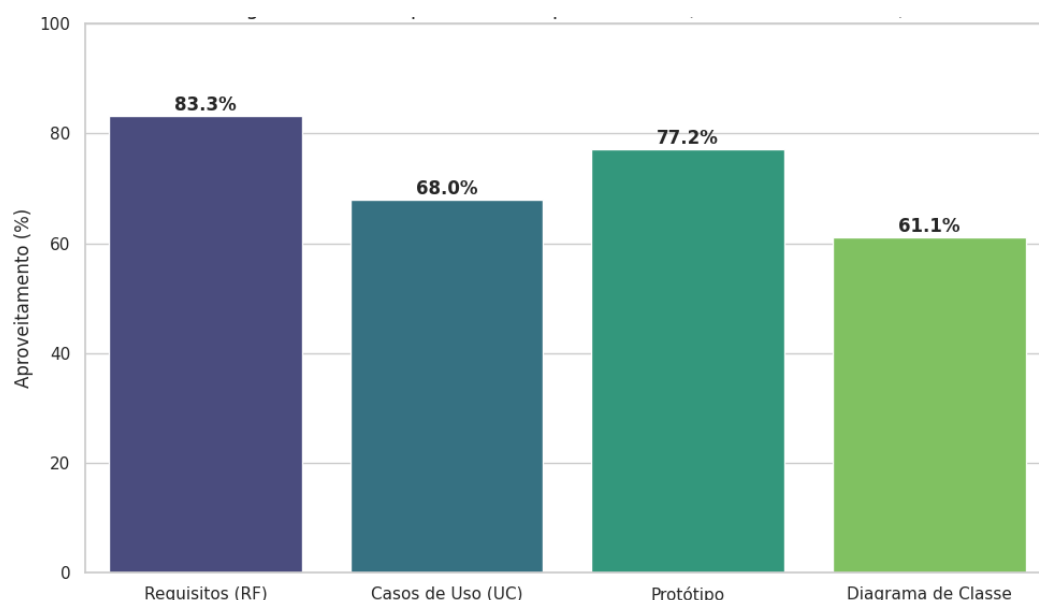


Figura 3. Desempenho médio por artefato. Fonte: Elaborado pelos autores (2026).

Em contrapartida, o Diagrama de Classes apresentou o menor aproveitamento (61,1%) e a maior variabilidade de notas (desvio padrão = 0,60). Esse contraste indica

que, embora os modelos de linguagem auxiliem substancialmente na etapa descritiva, evidenciam limitações diante de uma lacuna semântica na transposição estrutural. A definição de classes, a atribuição de responsabilidades e a configuração de relacionamentos exigiram dos estudantes um nível de integração conceitual e de abstração orientada a objetos que a ferramenta não é capaz de proporcionar de forma autônoma.

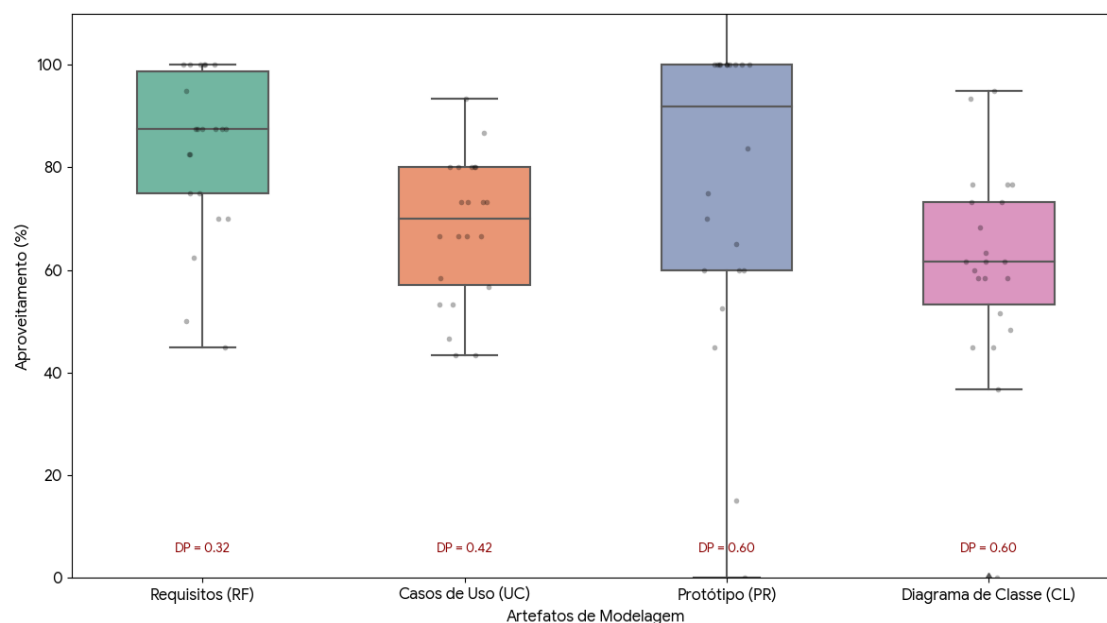


Figura 4. Dispersão de desempenho por artefato. Fonte: Elaborado pela autora (2026).

Na percepção dos estudantes, a experiência foi positiva e os resultados indicam que a proposta teve valor pedagógico. O instrumento de avaliação apresentou elevada consistência interna ($\alpha = 0,955$), com médias superiores a 3,5 (em uma escala de 5 pontos) nas dimensões de percepção global da metodologia e de nível de reflexão promovido pela interação com a ferramenta. A análise correlacional revelou uma associação positiva (r entre 0,51 e 0,57) entre o engajamento reflexivo com a IA e a consolidação conceitual percebida na UML. Os dados sugerem que a necessidade de revisar, corrigir e refinar as sugestões geradas levou os estudantes a analisar com mais cuidado os conceitos trabalhados, fortalecendo sua compreensão técnica.

Esses resultados são relevantes para o debate sobre IA na Educação em Computação. Eles sugerem que, quando há orientação, o uso da ferramenta não implica necessariamente a aceitação automática das respostas geradas. O estudo identificou uma correlação moderada ($r = 0,51$) entre o autorreconhecimento do domínio estrutural e a nota efetiva obtida no diagrama de classes, sugerindo convergência entre a percepção de aprendizagem e a qualidade do modelo produzido. A ausência de correlações de maior magnitude entre os artefatos de requisitos e de casos de uso é explicada, numericamente, pela restrição de variância observada nesses entregáveis, decorrente da elevada padronização desses artefatos.

Com o objetivo de complementar os resultados quantitativos, também foram examinadas as respostas discursivas dos estudantes às questões abertas do questionário final. A análise teve caráter exploratório e buscou identificar percepções recorrentes

relacionadas ao apoio oferecido pela IA generativa, às dificuldades encontradas durante a elaboração dos artefatos e às correções realizadas pelos participantes.

Os estudantes relataram a necessidade de corrigir requisitos funcionais duplicados, ajustar elementos dos diagramas de casos de uso e revisar associações e multiplicidades nos diagramas de classes. Um dos participantes afirmou ter corrigido uma multiplicidade inadequada e concluiu que a IA “ajuda, mas não substitui o raciocínio crítico”. Outro estudante observou que a ferramenta sugeriu elementos pouco relevantes para o sistema, o que exigiu revisão e adaptação manuais.

Essas evidências qualitativas complementam os resultados quantitativos ao mostrar que o apoio da IA não ocorreu de maneira automática nem uniforme. A ferramenta foi considerada mais útil para a organização e a geração inicial de alternativas, enquanto sua aplicação em decisões estruturais exigiu maior conhecimento conceitual, análise crítica e validação por parte dos estudantes. Assim, os relatos reforçam a importância da mediação docente e da revisão das respostas produzidas pela IA, especialmente em tarefas que envolvem associações, multiplicidades e responsabilidades das classes.

5. Considerações Finais

O estudo partiu da identificação de estudantes de graduação que enfrentam dificuldades na transição entre requisitos descritos em linguagem natural e sua representação formal em modelos UML, o que frequentemente resulta em artefatos com inconsistências semânticas. Investigou-se como a integração entre PBL, ciclos iterativos inspirados em metodologias ágeis e o uso orientado de ferramentas de IA generativa poderia influenciar a qualidade dos artefatos produzidos e a percepção de aprendizagem dos estudantes. Os resultados do estudo indicam que o uso de IA generativa no ensino de modelagem UML pode ser uma estratégia pedagógica viável quando articulado ao PBL e a ciclos iterativos de trabalho. Os resultados mostram que a combinação entre PBL, ciclos iterativos e uso orientado da IA contribuiu para o aprimoramento progressivo dos artefatos e para a aprendizagem relatada pelos estudantes.

Um dos pontos que este estudo destaca é que o uso da IA no ensino de Engenharia de Software evidencia a necessidade de revisar tanto o papel do professor quanto os critérios de avaliação adotados. No estudo realizado, a IA apresentou melhor desempenho quando usada com critérios claros de qualidade e com verificação das respostas, atuando como apoio ao trabalho dos estudantes, e não como substituição do raciocínio deles. Também é preciso considerar a consistência semântica, a justificativa das decisões de design e a forma como o estudante analisa as sugestões geradas pela ferramenta. A tecnologia deve ser inserida intencionalmente no desenho didático, assegurando que o protagonismo cognitivo permaneça com o estudante.

Os resultados deste estudo indicam que a IA generativa pode ser incorporada ao ensino de modelagem UML de forma produtiva quando utilizada em atividades orientadas e articuladas a metodologias ativas. No contexto analisado, as ferramentas de IA contribuíram de forma mais clara para as etapas iniciais da modelagem, como a organização de requisitos e a ideação funcional. Por outro lado, os estudantes

continuaram a apresentar mais dificuldades em tarefas que exigem maior abstração conceitual, como a construção de diagramas de classes.

Esses achados indicam que o uso da IA no ensino de Engenharia de Software funciona melhor quando associado a estratégias que incentivem a análise, a revisão e a justificativa das decisões dos estudantes. Em vez de ocupar o lugar do aluno, a ferramenta tende a ser mais útil quando apoia a discussão e o refinamento dos modelos.

Não foi foco deste estudo realizar análises comparativas com um grupo de controle. Com essa referência comparativa, será possível confrontar diretamente o arranjo proposto com um modelo de ensino tradicional e, em estudos futuros, isolar os efeitos observados exclusivamente no uso da ferramenta generativa. Em um cenário educacional, fatores como a maturação progressiva ao longo dos *sprints*, o ganho por prática decorrente da repetição dos ciclos de revisão e a mediação docente tendem a atuar simultaneamente. Assim, os achados não perdem valor, mas precisam ser lidos com outra ênfase; as evidências apontam mais para um efeito combinado entre PBL, ciclos ágeis e uso orientado da IA generativa do que para uma relação causal isolada.

Também é necessário considerar o tamanho e a composição da amostra. O estudo concentrou-se em um único contexto institucional, em uma disciplina introdutória, com 22 participantes. Como continuidade desta pesquisa, o processo pedagógico pode ser aplicado em turmas de diferentes tamanhos, faixas etárias, níveis de experiência e perfis acadêmicos e profissionais. Estudos futuros podem contemplar períodos diurnos e noturnos, instituições com diferentes condições de infraestrutura e modalidades presenciais, híbridas e a distância. A comparação entre estudantes iniciantes e estudantes com conhecimento prévio de UML poderá contribuir para verificar se o apoio da IA generativa produz efeitos distintos conforme o nível de formação.

Declaração sobre uso de Inteligência Artificial

Os autores declaram que ferramentas de Inteligência Artificial generativa foram utilizadas de forma limitada, como apoio à revisão da escrita e à organização textual de partes do manuscrito. Essas ferramentas foram empregadas exclusivamente para auxiliar na melhoria da clareza, da fluidez e da revisão gramatical do texto, não tendo sido utilizadas para a elaboração original do conteúdo científico, a definição da metodologia, a análise dos resultados ou a formulação das conclusões apresentadas neste trabalho. Todas as decisões relacionadas ao conteúdo do artigo, incluindo a concepção do estudo, a análise dos dados e a redação final do manuscrito, foram tomadas pelos autores humanos, que assumem a responsabilidade integral pelo conteúdo apresentado. Nenhuma ferramenta de Inteligência Artificial foi utilizada como autora deste trabalho.

Referências

- Azevedo, R., Chourio, P., Castro, A., e Gadelha, B. (2019). Revisão por Pares na Aprendizagem de Modelagem de Sistemas: concepção de uma ferramenta de suporte. In *Anais do XXX Simpósio Brasileiro de Informática na Educação*, (pp. 1975-1984). Porto Alegre: SBC. <https://doi.org/10.5753/cbie.sbie.2019.1975>.
- Albonico, M., Varela, P. J. (2023). A report on the use of ChatGPT in software engineering and systems analysis courses. In *Proceedings of the XXXVII Brazilian Symposium on Software Engineering (SBES 2023)*, pages 303–311. ACM. <https://doi.org/10.1145/3613372.3614189>.
- Cámara, J., Troya, J., Burgueño, L. et al. (2023). On the assessment of generative AI in modeling tasks: an experience report with ChatGPT and UML. *Software and Systems Modeling*, 22: 781–793. <https://doi.org/10.1007/s10270-023-01105-5>
- Canedo, E. D., Oliveira Jr., E. and Tives, H. A. (2025). Professional Use and Teaching of UML: Perceptions of Brasília-DF Region Software Practitioners and Higher Education Instructors. In *of the Workshop sobre Educação em Computação (WEI)*. 24–35. Sociedade Brasileira de Computação. <https://doi.org/10.5753/wei.2025.6933>
- Combemale, B., Gray, J. and Rumpe, B. (2023) ChatGPT in software modeling. *Softw Syst Model* 22, 777–779. <https://doi.org/10.1007/s10270-023-01106-4>
- Bari, D., Garaccione, G., Coppola, R., Torchiano, M. and Ardito, L. (2024). Evaluating large language models in exercises of UML class diagram modeling. In: *Proceedings De of the ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. 393–399. ACM. <https://doi.org/10.1145/3674805.369074>
- Di Rocco, J., Di Ruscio, D., Di Sipio, C. et al. (2025). On the use of large language models in model-driven engineering. *Software and Systems Modeling* 24, 923–948. <https://doi.org/10.1007/s10270-025-01263-8>
- Farias, I., Albuquerque, D., Rodrigues, G., Filho, E., Xavier, K. A. and Silva, J. A. C. (2024). Investigando o uso de ferramentas baseadas em grandes modelos de linguagem no contexto acadêmico. In: *Proceedings of the Workshop sobre Educação em Computação (WEI)*. 489–500. Sociedade Brasileira de Computação. <https://doi.org/10.5753/wei.2024.2092>
- Ferrari, A., Abualhaija, S. and Arora, C. (2024). Model generation with LLMs: From requirements to UML sequence diagrams. In *2024 IEEE 32nd International Requirements Engineering Conference Workshops (REW)* (pp. 291-300). IEEE. <https://doi.org/10.1109/REW61692.2024.00044>.
- Garaccione, G. et al. (2025). Evaluating use case diagram generation with large language models. In: *Proceedings of the IEEE/ACM International Workshop on Natural Language-Based Software Engineering (NLBSE)*. Ottawa, ON, Canada, pp. 41-44. IEEE. <https://doi.org/10.1109/NLBSE66842.2025.00015>.
- Hayashi, V. T. and Ferreira, I. F. G. (2025). Recommendation Systems Education using Project-Based Learning (PBL). In: *Proceedings of the Workshop sobre Educação em*

Computação (WEI). 207–216. Sociedade Brasileira de Computação. <https://doi.org/10.5753/wei.2025.7483>

Jackson, D., Liu, R. and Van der Hoek, A. (2025). Using generative AI in software design education: an experience report. arXiv preprint arXiv:2506.21703. <https://doi.org/10.48550/arXiv.2506.21703>.

Meireles, M. e Bonifácio, B. (2015). Uso de Métodos Ágeis e Aprendizagem Baseada em Problema no Ensino de Engenharia de Software: Um Relato de Experiência. In *Anais do XXVI Simpósio Brasileiro de Informática na Educação*, (pp. 180-189). Porto Alegre: Sociedade Brasileira de Computação. <https://doi.org/10.5753/cbie.sbie.2015.180>.

Murad, S. B. et al. (2025). Evidências sobre o uso do ChatGPT no ensino de modelagem de software: um experimento controlado. In: *Proceedings of the Workshop sobre Educação em Computação* (WEI). 1289–1300. Sociedade Brasileira de Computação. <https://doi.org/10.5753/wei.2025.9261>.

Reinhartz-Berger ISanto, Ali SJ, Bork D (2025) Leveraging LLMs for domain modeling: the impact of granularity and strategy on quality. In: *Krogstie J et al (eds) Advanced information systems engineering*. Springer, Cham, pp 3–19. Springer. https://doi.org/10.1007/978-3-031-94569-4_1.