Generative AI Dependency and its Effects on Introductory Programming Learning: An Experimental Study

Vinícius Castro Coutinho, Diana Maria Da Câmara Gorayeb, Osvaldo Tavares Viana Junior

Escola Superior de Tecnologia – Universidade do Estado do Amazonas (UEA) Manaus - AM – Brazil

{vcc.snf23, dgorayeb, ojunior}@uea.edu.br

Abstract. *This study investigated the impact of the dependence on Generative Artificial Intelligence, specifically ChatGPT, on learning retention in Introduction to Programming students. A controlled experiment was conducted with 18 students, divided into an experimental group and a control group. The experimental group used ChatGPT to solve exercises and was subsequently deprived of the tool, while the control group had no access to the AI. The results showed that, although the use of AI increased initial performance, students who depended on it showed a drop in performance when deprived of the resource. In contrast, the control group demonstrated improvement in knowledge retention and application. The findings suggest that excessive use of GenAI can foster dependence and hinder learning consolidation, reinforcing the need for its balanced pedagogical use.*

Resumo. *Este estudo investigou o impacto da dependência de Inteligência Artificial Generativa, especificamente o ChatGPT, na retenção do aprendizado de estudantes de Introdução à Programação. Foi realizado um experimento controlado com 18 alunos, divididos em grupo experimental e grupo controle. O grupo experimental utilizou o ChatGPT na resolução de exercícios e posteriormente foi privado da ferramenta, enquanto o grupo controle não teve acesso à IA. Os resultados mostraram que, embora o uso da IA tenha aumentado o desempenho inicial, os alunos que dependeram dela apresentaram queda de desempenho quando privados do recurso. Em contraste, o grupo controle demonstrou melhora na retenção e aplicação do conhecimento. Os achados sugerem que o uso excessivo de GenAI pode favorecer a dependência e prejudicar a consolidação da aprendizagem, reforçando a necessidade de seu uso pedagógico equilibrado.*

1. Introdução

Com o lançamento do ChatGPT em 2022, houve um grande crescimento do uso e do interesse em ferramentas de IA generativa (GenAI) para a resolução de problemas de diversas áreas do conhecimento. Uma dessas áreas é a de computação, especificamente no aprendizado de programação [Becker et al. 2023]. Diversos estudos demonstram que o uso de GenAI aumenta a eficiência dos alunos na resolução de problemas de programação [Kohen-Vacs et al. 2025; Yılmaz, Karaoğlan 2023] e oferece explicações e feedback [Hou et al. 2024].

Entretanto, além dos benefícios esperados que essa ferramenta pode trazer na área, há uma grande preocupação com os impactos negativos que ela pode ter no

aprendizado dos estudantes no que diz respeito ao excesso de dependência, que pode resultar em um aprendizado superficial [Stone 2024; Kohen-Vacs et al. 2025], especialmente em alunos de introdução à programação [Sheard et al. 2024], comprometendo o desenvolvimento de competências básicas e capacidade de análise crítica. [Denny et al. 2024; Mogavi et al. 2024]

O presente trabalho busca avaliar quantitativa e qualitativamente a extensão da dependência de GenAI e seu impacto na retenção do aprendizado de alunos de programação introdutória por meio de um experimento controlado com 18 alunos de Introdução à Programação de Computadores (IPC) da Universidade do Estado do Amazonas (UEA).

2. Revisão da Literatura

O ChatGPT é um Grande Modelo de Linguagem (LLM), um subconjunto de GenAI, e é baseado na arquitetura GPT (*Generative Pre-training Transformer*) e treinado em grandes quantidades de dados textuais da internet. Foi projetado para gerar texto semelhante ao padrão de escrita humana e pode ser usado para tarefas como responder perguntas, corrigir código e sumarizar texto, a depender da instrução do usuário [OpenAI, 2022].

Modelos como o ChatGPT são usados por estudantes de IPC para obter ajuda na programação, depurar código, fornecer retorno e auxiliar na geração de ideias [Stone, 2024]. Denny et al. (2024) afirmam que ferramentas de IA transformam a criação e customização de recursos educacionais, como exercícios de programação, permitindo a geração eficiente de um material educacional personalizado ao aluno. Ao mesmo tempo, reconhece que modelos de GenAI também são altamente capazes de gerar soluções prontas para problemas típicos de IPC, levantando questões sobre excesso de dependência e mau uso da ferramenta.

Kohen-Vacs et al. (2025) realizaram estudos quantitativos com alunos universitários durante um período para capturar sua percepção sobre ferramentas de GenAI. Os alunos foram em geral receptivos ao uso de GenAI e, apesar de alguns participantes reconhecerem que a dependência em excesso dessa ferramenta pode impedir o desenvolvimento de habilidades críticas essenciais, essa preocupação não foi particularmente acentuada ou muito difundida entre os alunos. Os pesquisadores concluíram que essa opinião mista reflete os debates na literatura sobre a dicotomia da inteligência artificial, fomentando a produtividade e a criatividade ao mesmo tempo que levanta questões sobre excesso de dependência e menor envolvimento cognitivo.

Yılmaz e Karaoğlan (2023) realizaram um estudo experimental para investigar o efeito do uso de ChatGPT no que diz respeito a habilidades de pensamento computacional, auto-eficácia de programação e motivação dos alunos por meio de escalas Likert. Estudantes universitários foram divididos em grupo experimental e grupo controle. Ambos os grupos receberam treinamento durante 5 semanas com uma abordagem *hands-on*, isto é, aplicando o conhecimento teórico em exercícios práticos,

mas somente o grupo experimental teve autorização para utilizar ChatGPT nesse treinamento. Após esse período, verificou-se que o uso de ChatGPT aumentou as métricas avaliadas nos estudantes do grupo experimental, incluindo pensamento crítico, resolução de problemas, tarefas simples e complexas de programação.

Silva Junior et al. (2023) descreveram um estudo de caso com questionário usando uma abordagem quali-quantitativa, com alunos de ensino médio integrado ao técnico, considerando o nível dos alunos na resolução de questões de programação e respostas sobre a percepção do uso do ChatGPT. Os pesquisadores observaram que, ao usarem o ChatGPT, houve um aumento na quantidade de alunos que conseguiram resolver integralmente ou parcialmente as questões propostas, principalmente as de maior nível de dificuldade. Perceberam que grande parte dos alunos respondeu ser vantajosa a utilização do ChatGPT para auxiliar na resolução das questões de programação e consideraram útil e eficaz o uso da ferramenta. Ao final, os autores alertaram sobre a possibilidade de a dependência excessiva da ferramenta limitar o desenvolvimento de habilidades críticas e independentes nos alunos.

Analisando a literatura, não foram encontrados trabalhos que investigassem a diferença de desempenho dos grupos que utilizaram GenAI caso fossem privados de seu uso, a fim de investigar em maior detalhe os efeitos desse excesso de dependência sobre o aprendizado de alunos de programação introdutória. O presente estudo buscou analisar esse efeito por meio da resolução de questões de programação por alunos de IPC durante uma semana. Smith e Karpicke (2014) sugerem que a resolução de exercícios é um meio eficiente de retenção de conhecimento, logo o desempenho dos alunos pode ser usado para indicar se o aprendizado foi retido ou não.

3. Metodologia

Nesta pesquisa, foi aplicado um experimento em 18 alunos de uma mesma turma de IPC na UEA, comparando dois grupos dentro da turma: 9 alunos do Grupo 1 (experimental) e 9 do Grupo 2 (controle), divididos aleatoriamente antes da data de início dos experimentos. O experimento foi dividido em duas partes: Dia 1 e Dia 2. A Tabela 1 representa visualmente a divisão do experimento.

Os alunos da turma haviam completado 56 das 60 horas da disciplina antes da aplicação do experimento, tendo aprendido sobre os comandos e estruturas básicas de controle e dados em Python. Isso foi utilizado como medida do conhecimento dos alunos no lugar de um pré-teste por limitações de tempo.

Tabela 1. Divisão do experimento

	Dia 1	**Dia 2**
Grupo 1 (Experimental)	Com ChatGPT	Sem ChatGPT
Grupo 2 (Controle)	Sem ChatGPT	Sem ChatGPT

O experimento consistiu na realização de uma lista de 5 questões em Python e em seguida um questionário de autoavaliação a cada dia. No Dia 1, o Grupo 1 teve acesso ao ChatGPT para resolver as questões, e no Dia 2, foi privado desse acesso, a fim de investigar o desempenho dos participantes quando impedidos de depender de GenAI. O Grupo 2 não teve acesso ao ChatGPT em nenhum dia.

A escolha de um conjunto de questões de programação ocorreu pela facilidade de elaboração e medição dos resultados por aluno. Os alunos tiveram uma hora para realizar os exercícios. Após o término dos exercícios, os alunos responderam ao questionário do dia. O intervalo de tempo entre o Dia 1 e o Dia 2 foi de 1 semana.

3.1. Questões

Os assuntos das questões e seus enunciados estão presentes na Tabela 2. A Lista 2 foi elaborada por meio de pequenas alterações, recomendadas pela professora da disciplina, das questões da Lista 1, com o intuito de manter a dificuldade similar entre as listas, e de modo que os alunos possam aplicar no Dia 2 o conhecimento adquirido no Dia 1, obtendo um desempenho maior. Os assuntos escolhidos foram adaptados dos assuntos abordados na disciplina até o momento do experimento.

Tabela 2. Questões das listas

Assunto	Lista 1 (Dia 1)	Lista 2 (Dia 2)
Produto Aritmético e Lógica Condicional	Escreva uma função que recebe dois inteiros. Se o produto dos dois números for menor ou igual a 1000, retorne seu produto; caso contrário, retorne sua soma.	Escreva uma função que recebe três inteiros. Se o produto dos dois primeiros números for divisível pelo terceiro número, retorne o produto; caso contrário, retorne o produto dos três números.
Múltiplos/Divisores	Faça um programa que receba um número n e mostre todos os múltiplos de 3 entre 1 e n, exceto os que terminam em 7 ou 9.	Faça um programa que receba um número n e mostre todos os seus divisores entre 1 e n.
Maior/Menor Número	Peça 5 números ao usuário e exiba o maior e o menor entre eles, sem usar funções prontas como max() ou min().	Peça 5 números ao usuário e exiba o segundo maior e o segundo menor entre eles.
Números Primos	Faça um programa que verifique se um número é primo (divisível apenas por 1 e por ele mesmo). Se o número não for primo, mostre o número primo mais próximo.	Faça um programa que verifique se um número é primo (divisível apenas por 1 e por ele mesmo). Se o número não for primo, mostre todos seus divisores.

Fibonacci	Escreva um código recursivo que gere os n primeiros números da sequência de Fibonacci.	Escreva um código não recursivo que gere os n primeiros números da sequência de Fibonacci.

3.2. Questionário

Os questionários de ambos os dias continham as seguintes perguntas:

- "Como você avalia a dificuldade nas questões?" (Percepção de Dificuldade)
- "Você conseguiria explicar como suas soluções funcionam?" (Capacidade de Explicação)
- "Você sentiu que aprendeu durante a resolução das questões?" (Percepção de Aprendizado)

No questionário do Dia 1, houve uma pergunta específica para o Grupo 1:

- "Você sentiu que conseguiria resolver as questões sem a IA?" (Confiança do Grupo 1).

No questionário havia um termo de consentimento explicando a natureza voluntária e anônima do experimento e um campo para os participantes enviarem as respostas às questões para correção. Ao todo, participaram 18 alunos da turma de IPC.

4. Resultados da Lista

As métricas avaliadas na lista foram o tempo total de resolução das questões e a quantidade de questões respondidas corretamente. As questões foram enviadas pelos questionários e em seguida corrigidas manualmente pela equipe de pesquisa. Os avaliadores consideraram corretas as questões que, dada uma entrada, produziram uma saída de acordo com o que a questão pedia, incluindo *edge cases*. Questões não respondidas ou respondidas parcialmente (a saída não abordou tudo que o enunciado pedia) foram consideradas como incorretas. Não houve discordâncias nas correções entre os avaliadores.

A Tabela 3 a seguir mostra as médias dos resultados por dia e por grupo. A distribuição dos dados é normal segundo o teste de Shapiro-Wilk [Shapiro, Wilk 1965], e nenhum *outlier* foi encontrado usando amplitude interquartil [Figueiredo Filho et al. 2023].

Tabela 3. Resultados (Médias)

	Grupo 1		**Grupo 2**	
	Tempo	**Acertos**	**Tempo**	**Acertos**
Dia 1	00:38	4,0	00:56	2,1
Dia 2	00:57	2,2	00:52	3,3

4.1 Grupo 1

No Dia 1, o Grupo 1 obteve um bom desempenho, acertando 80% das questões em 38 minutos em média. No Dia 2, o desempenho foi menor. O grupo acertou em média 2,2 questões em um tempo de 57 minutos. Isso significa um aumento de 50% no tempo, e uma redução de 45% nos acertos em comparação com o primeiro dia

4.2 Grupo 2

No Dia 1, o Grupo 2 obteve um desempenho baixo, acertando 2,1 questões em 56 minutos em média. No Dia 2, o desempenho foi melhor. O grupo acertou em média 3,3 em um tempo de 52 minutos. Isso significa uma redução de apenas 7,1% no tempo, mas um aumento de 57,1% nos acertos em comparação com o primeiro dia.

4.3. Análise Entre Grupos

No Dia 1, o Grupo 1 foi melhor que o Grupo 2 em todas as métricas. Do Dia 1 para o Dia 2, o Grupo 1 apresentou uma piora em todas as métricas, enquanto o Grupo 2 apresentou uma melhora em todas as métricas, resultando no Grupo 1 sendo superado pelo Grupo 2.

5. Resultados do Questionário

Esta seção apresenta os resultados do questionário descritos em gráficos de coluna.

5.1 Dia 1

A Figura 1 mostra a percepção de dificuldade e a capacidade de explicação dos participantes no Dia 1. O Grupo 1 percebeu as questões como sendo mais fáceis em relação ao Grupo 2, sendo os únicos a considerá-las fáceis ou muito fáceis, e nenhum participante desse grupo as considerou difíceis ou muito difíceis. O Grupo 2 quase em sua totalidade as considerou de dificuldade média, e um participante as considerou difíceis.

Em relação à capacidade de explicação, o Grupo 1 foi o único a ter integrantes que afirmaram não conseguir explicar quase nada da resolução (2 participantes), enquanto o restante (7 participantes) afirmou conseguir explicar tudo ou quase tudo. O Grupo 2 teve menos integrantes nas categorias “Tudo” e “Quase tudo” (6 participantes ao total), mas o restante afirmou conseguir explicar em partes a resolução (3 participantes).

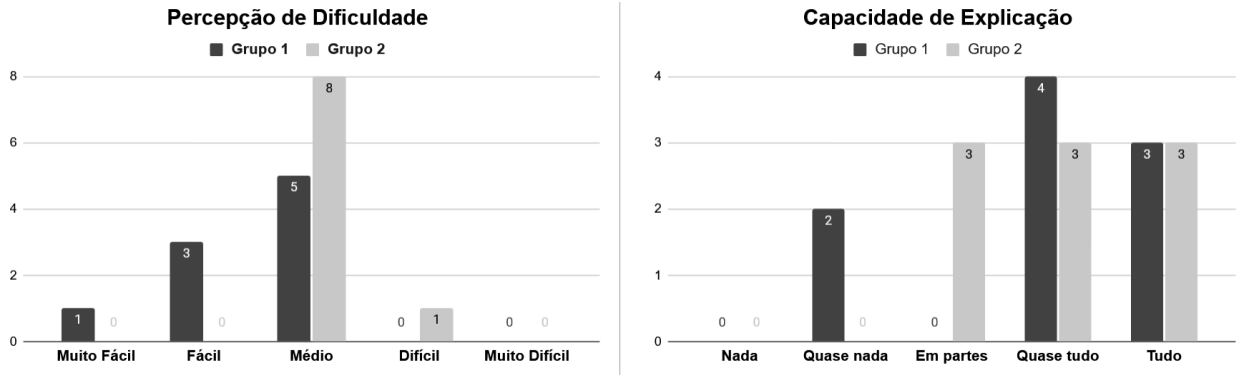


Figura 1. Percepção de Dificuldade e Capacidade de Explicação no Dia 1

A Figura 2 mostra a percepção de aprendizado dos participantes no Dia 1 (“Você sentiu que aprendeu durante a resolução das questões?”). Ambos os grupos tiveram a mesma distribuição de respostas. A maioria dos participantes respondeu que aprendeu moderadamente ou muito.

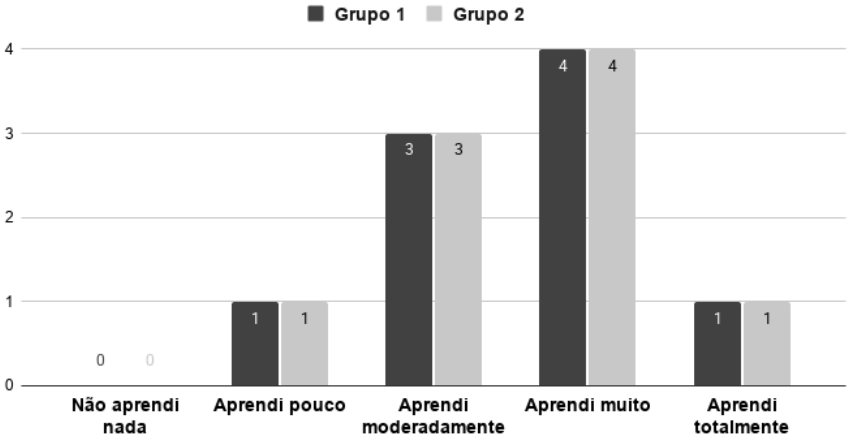


Figura 2. Percepção de Aprendizado no Dia 1

A Figura 3 mostra a confiança dos participantes do Grupo 1 (“Você sentiu que conseguiria resolver as questões sem a IA?”) no Dia 1. A maior parte (7) dos participantes teve uma confiança alta, entre os intervalos de “provavelmente” e “definitivamente”, e apenas 2 participantes responderam “talvez”. Nenhum participante demonstrou a confiança no intervalo de “provavelmente não” a “definitivamente não”.

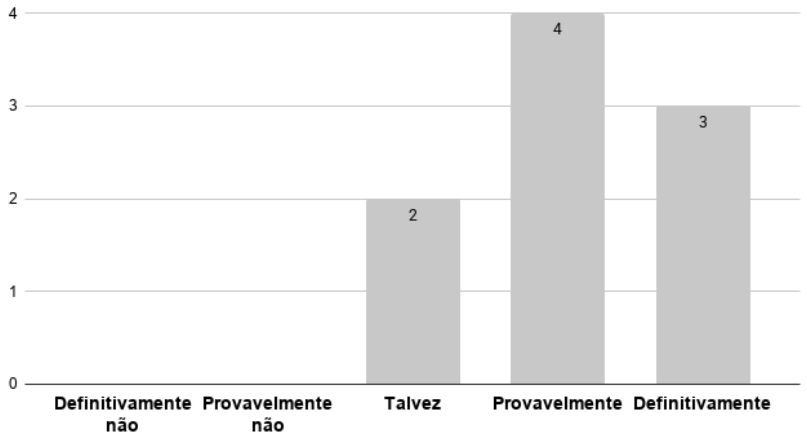


Figura 3. Confiança do Grupo 1 no Dia 1

5.2 Dia 2

A Figura 4 mostra a percepção de dificuldade e capacidade de explicação dos participantes no Dia 2. O Grupo 1 teve 3 participantes que consideraram as questões difíceis, 4 que perceberam como médias e 2 como fáceis. Já o Grupo 2 teve apenas 1 participante que as considerou difíceis, e a maioria (7) as considerou médias, com apenas um participante percebendo-as como fáceis.

Em relação à capacidade de explicar a resolução das questões, ambos os grupos tiveram a mesma distribuição. Em cada grupo, 7 participantes afirmaram conseguir explicar tudo ou quase tudo, 2 participantes afirmaram conseguir explicar em partes, e nenhum participante afirmou não conseguir explicar nada ou quase nada.

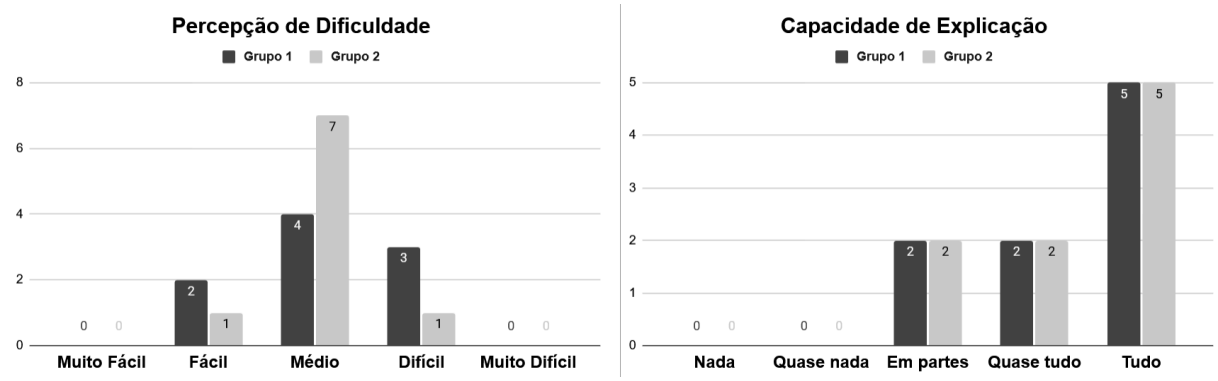


Figura 4. Percepção de Dificuldade e Capacidade de Explicação no Dia 2

Em relação à percepção de aprendizado, apresentada na Figura 5, nenhum participante afirmou não ter aprendido nada ou ter aprendido pouco no segundo dia. No Grupo 1, a maior parte dos participantes (4) afirmou ter aprendido totalmente com os exercícios. Já no Grupo 2, a maioria (5) dos participantes afirmou apenas ter aprendido muito, com somente 2 participantes afirmando ter aprendido totalmente.

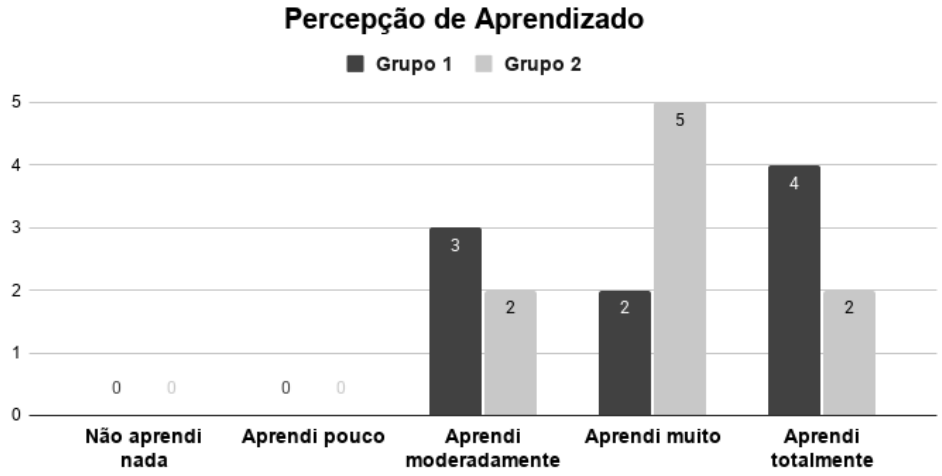


Figura 5. Percepção de Aprendizado no Dia 2

6. Discussões

Os resultados apoiam a hipótese de que o uso de IA pode provocar dependência excessiva e que, neste estudo, o uso do ChatGPT esteve associado a uma redução do desempenho após sua retirada.

O Grupo 2 conseguiu aumentar seu desempenho do primeiro para o segundo dia, enquanto o Grupo 1 apresentou redução após a retirada do ChatGPT. Esse padrão é consistente com a hipótese de que os participantes que utilizaram a ferramenta no Dia 1 tiveram maior dificuldade em manter o desempenho quando privados de seu uso no segundo dia.

O Grupo 1 foi o único a ter participantes que afirmaram conseguir explicar quase nada da própria resolução no Dia 1, diferentemente do Grupo 2, cujos participantes conseguiam explicar no mínimo em partes a sua resolução. No segundo dia, o Grupo 1 afirmou conseguir explicar muito melhor sua resolução, como se percebe na Figura 6, sugerindo que o Grupo 1, ao ser privado de GenAI, relatou uma maior percepção de aprendizado com as questões, ao mesmo tempo que teve maior confiança na sua capacidade de explicar as respostas. O Grupo 1 também apresentou um aumento de percepção de aprendizado do primeiro para o segundo dia, como mostra a Figura 7.

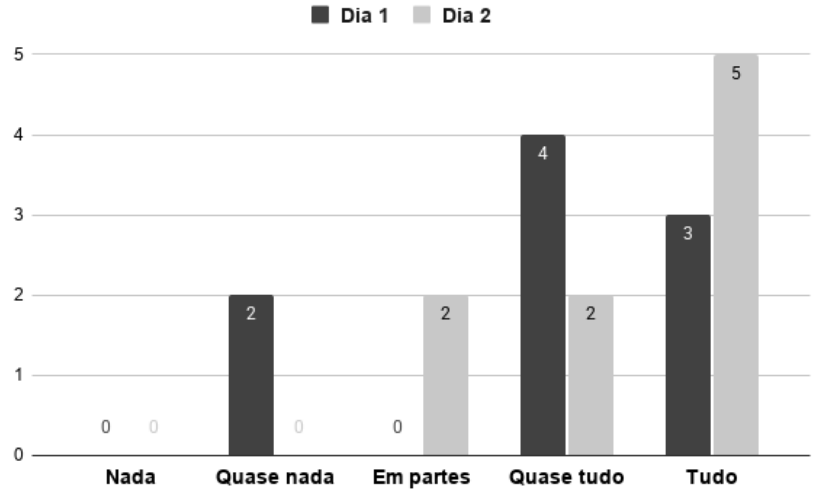


Figura 6. Comparação de capacidade de explicação do Grupo 1

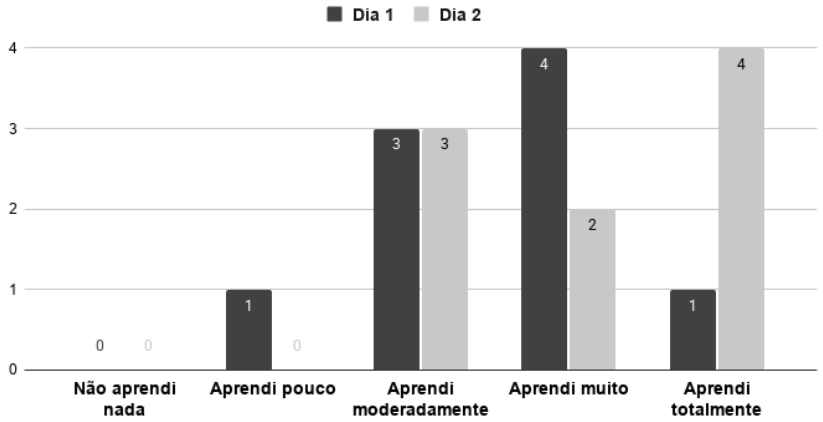


Figura 7. Comparação de aprendizado do Grupo 1

Os resultados ratificam os problemas de dependência levantados pelos trabalhos relacionados, mas evidenciam um distanciamento entre a dependência percebida pelos estudantes e a dependência evidenciada pelo desempenho observado. Neste trabalho, assim como no de Kohen-Vacs et al. (2025), os estudantes não exprimiram uma preocupação forte em relação ao excesso de dependência, evidenciando confiança nas suas habilidades, como mostrado na Figura 3. Todavia, quando privados de GenIA, o Grupo 1 teve um desempenho pior que o grupo controle, revelando uma discrepância entre a percepção dos alunos e a realidade.

Os resultados também foram diferentes dos encontrados por Yılmaz e Karaoğlan (2023) no que diz respeito à capacidade de resolução de problemas de programação. O presente artigo verificou que os estudantes apresentaram maior desempenho e maior percepção de aprendizado quando não utilizaram IA, em contraste aos achados desses pesquisadores. Vale ressaltar que o experimento deles foi aplicado durante 5 semanas, em oposição ao deste de 1 semana, e a forma de coleta de dados foi um questionário

diferente. Entretanto, Yılmaz e Karaoğlan (2023) se basearam inteiramente na autoavaliação dos participantes, sem meio de avaliar seu desempenho prático.

7. Limitações

O tamanho da amostra (N=18) é pequeno, então os resultados devem ser interpretados com cautela quanto à sua generalização para populações maiores. Apesar de termos inferido o conhecimento dos alunos pelo plano de ensino da disciplina, isso não é tão robusto quando a realização de um pré-teste. Também é possível que os resultados tenham sido influenciados por fatores externos, como fadiga, ansiedade, engajamento ou familiarização com o formato do experimento pelos alunos. Outra limitação é que nenhum teste estatístico de significância foi aplicado.

O intervalo de uma semana entre os experimentos foi adotado com o objetivo de reduzir a possibilidade de aquisição significativa de novos conteúdos pelos alunos no período entre as atividades, de modo que os participantes realizassem as atividades com base principalmente nas competências previamente desenvolvidas.

8. Considerações Finais

Este trabalho investigou o impacto da GenAI, especificamente o ChatGPT, sobre o aprendizado de estudantes de programação introdutória. O trabalho corrobora parte da literatura científica, demonstrando a dependência excessiva dos alunos de ferramentas de GenAI de modo prático, por meio do rendimento apresentado (medido pelo número de questões respondidas), da aprendizagem evidenciada (medida pela diferença de desempenho na resposta de questões em diferentes dias), e da percepção de aprendizado, dificuldade e capacidade de explicação (autoavaliadas pelos alunos). O grupo controle, diferentemente do grupo experimental, apresentou uma evolução no processo de aprendizagem tanto na assimilação do conteúdo quanto no tempo necessário para a resolução das questões.

O esforço na realização deste estudo se justifica pelo pequeno número de estudos de mesma natureza em disciplinas introdutórias de programação, em ambientes reais de sala de aula, o que significou trabalhar com estudantes iniciantes na área da computação no cenário de instituições públicas brasileiras.

Por fim, a influência do ChatGPT no Grupo 1 demonstrou que ele pode agilizar a elaboração de respostas das questões, mas não necessariamente do aprendizado. Porém, a longo prazo, como alguns autores relatam na literatura, poderia haver uma influência positiva na aquisição do conhecimento quando associada a uma prática mais contínua. Isso porque a GenAI é uma ferramenta cada vez mais importante na trajetória dos alunos, que pode proporcionar benefícios tangíveis e, portanto, não deve ser ignorada nem proibida no ambiente acadêmico, mas deve ser adaptada e regulada de modo a complementar o ciclo natural de aprendizado.

Referências

Stone, I. (2024) “Exploring Human-Centered Approaches in Generative AI and Introductory Programming Research: A Scoping Review”, Proceedings of the 2024 Conference on United Kingdom & Ireland Computing Education Research (UKICER '24). Association for Computing Machinery, New York, NY, USA, Article 4, 1–7.

Becker, B., et al. (2023) "Generative ai in introductory programming." Computer Science Curricula: 438-439

Kohen-Vacs, D., Usher, M. e Jansen, M. (2025) “Integrating Generative AI into Programming Education: Student Perceptions and the Challenge of Correcting AI Errors”. International Journal of Artificial Intelligence in Education

Yılmaz, R., e Karaoğlan Yılmaz, F. G. (2023) “The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation”. Computers and Education: Artificial Intelligence. 4.

Smith, M. A. e Karpicke, J. D. (2014) Retrieval practice with short-answer, multiple choice, and hybrid tests, Memory, 22:7, 784-802

Sheard, J. et al. (2024). Instructor Perceptions of AI Code Generation Tools - A Multi-Institutional Interview Study. Em Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1 (Portland, OR, USA) (SIGCSE 2024). ACM, New York, NY, USA, 1223–1229.

Hou, I., Mettille, S., Man, O., Li, Z., Zastudil, C., e MacNeil, S. (2024). The Effects of Generative AI on Computing Students’ Help-Seeking Preferences. Em Proceedings of the 26th Australasian Computing Education Conference (Sydney, NSW, Australia) (ACE ’24). ACM, New York, NY, USA, 39–48

Mogavi, R. H. et al. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters’ utilization and perceptions. Computers in Human Behavior: Artificial Humans, 2(1).

Denny, P. et al. (2024). Computing Education in the Era of Generative AI. Commun. ACM 67, 2 (Jan. 2024), 56–67

Silva Junior, S. M. et al. (2023) "ChatGPT no auxílio da aprendizagem de programação: Um estudo de caso", Em: Anais do XXXIV Simpósio Brasileiro de Informática na Educação (SBIE), Passo Fundo. pp. 1375-1384.

OpenAI (2022). “Conheça o ChatGPT”, https://openai.com/pt-BR/index/chatgpt/

Shapiro, S. S., Wilk, M. B. An analysis of variance test for normality (complete samples), *Biometrika*, Volume 52, Issue 3-4, December 1965, Pages 591–611.

Figueiredo Filho, D. et al. (2023). Living with outliers: How to detect extreme observations in data analysis. *BIB - Revista Brasileira De Informação Bibliográfica Em Ciências Sociais*, (99)