

Caracterização de Vídeos de Ensino de Programação após a popularização das IAs Generativas no YouTube

Bruno Urbano¹, Guilherme Cantoni¹, Henrique Franco¹, Pedro de Barros¹

¹Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
Caixa Postal 1.686 – 30.535-901 – Belo Horizonte – MG – Brasil

Abstract. *With the rise of generative AI tools such as ChatGPT, changes have emerged in how programming education content is created and consumed on YouTube. This study aims to characterize these transformations by analyzing videos published after the public release of ChatGPT. We compare videos that mention AI to traditional ones, focusing on three aspects: audience engagement, thematic distribution, and the nature of comment interactions. Using the YouTube Data API, we collected data from thousands of videos and comments. The results show that traditional videos maintain higher average engagement and more technical questions in comments, while AI-related videos concentrate more mentions of generative tools and have shown recent growth in “Tips” and “Tutorial” categories.*

Resumo. *Com a popularização das IAs generativas, como o ChatGPT, mudanças vêm sendo observadas na forma como conteúdos de ensino de programação são produzidos e consumidos no YouTube. Este estudo visa caracterizar essas transformações, analisando vídeos publicados após o lançamento do ChatGPT. Foram comparados vídeos que mencionam IA com vídeos tradicionais, considerando três aspectos: engajamento do público, distribuição temática e natureza das interações nos comentários. Utilizando a YouTube Data API, foram extraídos dados de milhares de vídeos e comentários. Os resultados indicam que vídeos tradicionais ainda mantêm maior engajamento médio e proporção de comentários técnicos, enquanto vídeos com IA concentram mais menções à tecnologia e apresentam crescimento recente em publicações nas categorias “Dicas” e “Tutoriais”.*

1. Introdução

Nos últimos anos, o desenvolvimento de Inteligências Artificiais (IAs) generativas, como o ChatGPT e o GitHub Copilot, tem promovido mudanças significativas nos processos de ensino e aprendizagem em programação. Essas ferramentas ampliam as possibilidades de geração automatizada de código, personalização de conteúdos e resolução autônoma de dúvidas, o que tem impactado a prática pedagógica tanto em ambientes formais quanto informais de educação.

Paralelamente, o YouTube consolidou-se como uma das principais plataformas de ensino não formal de programação, oferecendo uma ampla gama de conteúdos, tais como tutoriais, cursos, dicas e projetos. A popularização das IAs generativas intensificou a presença de vídeos que as mencionam ou as utilizam como recurso didático, o que levanta questões relevantes acerca das mudanças no comportamento de consumo e na produção de conteúdos educacionais nesse ambiente.

Neste contexto, torna-se necessário compreender como a introdução e a popularização das IAs generativas afetam aspectos centrais do conteúdo educacional de programação no YouTube, especialmente no que se refere ao engajamento do público, às temáticas predominantes e à natureza das interações nos comentários.

Diante desse cenário, este artigo tem como objetivo caracterizar as transformações ocorridas no conteúdo educacional de programação no YouTube após a popularização das IAs generativas. A investigação concentra-se em três aspectos: (i) as diferenças de engajamento entre vídeos que mencionam ou não a IA; (ii) a distribuição temática desses vídeos; e (iii) a natureza das interações nos comentários.

Para atingir esse objetivo, o artigo está estruturado da seguinte forma: a Seção 2 apresenta a fundamentação teórica, discutindo conceitos relacionados às IAs generativas e ao ensino de programação em plataformas digitais; a Seção 3 descreve a metodologia adotada na pesquisa, com detalhamento dos critérios de seleção de vídeos, das métricas utilizadas e dos procedimentos de análise; a Seção 4 expõe e discute os resultados obtidos.

2. Trabalhos Relacionados

Já há alguns anos, o YouTube vem sendo utilizado como ferramenta para o ensino de ciências [C. Aranha and Silva 2019]. Além de ciências, surgiu a necessidade de avaliar a viabilidade do uso da plataforma para o ensino de programação, para desenvolvedores de *software* [Paula et al. 2025]. Ambos os estudos, após análises e ponderações, demonstram um valor pedagógico e educacional relevante para a comunidade científica e de desenvolvimento de *software*. Ambos os estudos serviram como justificativa sobre a capacidade do YouTube de fornecer conteúdo relevante e educativo que contribui para o ensino de diversas áreas do conhecimento.

Com relação às mudanças impostas à dinâmica da comunidade pela popularização das IAs generativas baseadas em LLMs, existem pesquisas que analisam: conteúdo produzido utilizando IAs generativas [Li B. 2024], as vantagens das IAs generativas para a indústria acadêmica [Kshetri 2023], IAs generativas para *feedbacks* escaláveis no contexto de educação *online* [Jürgensmeier and Skiera 2024] e o impacto das IAs generativas na contribuição de conhecimento em comunidades como o *StackOverflow* [Kabir et al. 2024]. Estes estudos revelam consequências das IAs generativas na dinâmica de usuários dentro de comunidades de ensino *online*, e justificam a necessidade de se realizar um estudo considerando o YouTube como uma plataforma de ensino *online*.

Dentro do escopo específico deste estudo, existe um artigo que visa compreender a percepção de usuários a *thumbnails* criadas utilizando IAs generativas [J. 2025]. No estudo em questão, notou-se uma diferença nas reações dos usuários com relação ao conteúdo produzido por IA àquele considerado tradicional. Isso justifica este estudo, onde se pode criar uma hipótese de que isso possa estar acontecendo também com os títulos e menções às IAs em vídeos de ensino de programação.

3. Metodologia

A metodologia deste trabalho inclui quatro grandes etapas: (i) definição da data de corte, (ii) definição dos termos de busca, (iii) coleta de dados e (iv) análise dos dados. Nas subseções a seguir, detalha-se cada uma dessas etapas.

3.1. Definição da Data de corte

Visando caracterizar os vídeos de ensino de programação após a popularização das IAs generativas no YouTube, a primeira etapa deste trabalho consistiu em definir como marco temporal inicial o dia 30 de novembro de 2022, data em que o ChatGPT foi disponibilizado ao público. A partir desse ponto, considerou-se que a presença de ferramentas baseadas em inteligência artificial generativa se tornou significativamente mais visível e acessível. Dessa forma, apenas vídeos publicados a partir dessa data foram incluídos na amostra analisada.

3.2. Definição dos Termos de busca

Na segunda etapa buscou-se definir os termos de busca para mineração dos vídeos e estruturação dos mesmos. Primeiramente, foram definidas três linguagens de programação presentes no *ranking* definido pelo GitHub [GitHub 2024], das linguagens de programação mais populares, sendo elas: Python, JavaScript e Java. Essas linguagens foram utilizadas para uma maior especificação do objeto de estudo desse trabalho, assim filtrando o amplo universo de vídeos de ensino de programação e possibilitando uma melhor análise dos dados coletados.

Posteriormente, foram definidos dois grupos para organização dos vídeos, sendo eles: Grupo IA e Grupo Tradicional. O Grupo IA é composto por vídeos que mencionam inteligência artificial ou ferramentas de IA generativa, como ChatGPT, Copilot, Claude, entre outras. Já o Grupo Tradicional é composto por vídeos que não apresentam qualquer referência a essas tecnologias. Visando garantir essa separação, todos os vídeos inicialmente classificados como pertencentes ao Grupo Tradicional, foram submetidos a uma filtragem nos títulos para assegurar a ausência de termos relacionados à inteligência artificial. Da mesma forma, os vídeos classificados como pertencentes ao Grupo IA, foram submetidos a uma filtragem nos títulos para garantir a presença de termos relacionados à inteligência artificial.

Em seguida, foi realizada a classificação temática dos vídeos com base em três categorias recorrentes no ensino de programação online, sendo elas: Cursos e Tutoriais, Projetos, Dicas. A definição dessas categorias foi fundamentada na análise de títulos, *playlists* e termos mais utilizados em canais de referência no ensino de programação na língua inglesa, como [BroCode 2019], [ThePrimeTime 2021], [GiraffeAcademy 2017], [CodingwithLewis 2020], [TheCodingSloth 2023] e [ForrestKnight 2016]. Além disso, a partir dessa análise foram definidos conjuntos de termos que representam cada categoria, tanto para o Grupo Tradicional quanto para o Grupo IA.

Nesse viés, para as categorias no Grupo Tradicional, a busca envolveu termos como “python tutorial”, “java course”, “javascript project” e “python tips”. Já para o Grupo IA, foram envolvidos termos como “Learn Python with AI”, “Project in Java using AI”, “Java programming tips using AI”, entre outros. Ou seja, os termos foram baseados nos utilizados para o Grupo Tradicional, porém adicionados de termos amplamente utilizados para se referir à inteligência artificial.

Finalmente, para a análise de comentários, necessária para a observação proposta na terceira questão problema, foram definidos dois conjuntos de termos para a análise da natureza dos comentários. O primeiro conjunto engloba termos amplamente usados por usuários em fóruns de computação, como Stack Overflow, para se referirem a assuntos

relacionados a dúvidas técnicas, como “bug”, “error”, “problem”, “I’m confused about”, entre outros. Já o segundo conjunto engloba expressões que identificam menções a ferramentas de IA, incluindo termos como “AI”, “ChatGPT”, “GPT”, “Copilot” e “Bard”.

Além disso, foi feita uma normalização para tornar comparável o volume de comentários entre vídeos com diferentes quantidades de visualizações. Para isso, foi calculada a proporção de comentários por visualização em cada vídeo e, em seguida, foi obtida a média dessa proporção para cada grupo (IA e Tradicional). O valor final, multiplicado por 1.000, representa a média de comentários a cada mil visualizações, assim permitindo comparar o engajamento proporcional entre os grupos.

3.3. Coleta de Dados

Nesta etapa, foi realizada a coleta dos dados necessários para o estudo. Os dados foram minerados utilizando a *Application Programming Interface* (API) YouTubeData, com acesso apenas a vídeos públicos. Ela foi utilizada para obter as seguintes informações de cada vídeo: título, descrição, data de publicação, número de visualizações, número de likes e número de comentários. Para a análise dos comentários, foram coletados os textos completos dos comentários disponíveis publicamente, respeitando os limites de requisição da *Application Programming Interface* (API).

Com relação à quantidade de vídeos coletados para análise, para a primeira pergunta do estudo, foram coletados 500 vídeos para cada termo definido. Para a segunda pergunta, foram coletados 250 vídeos por termo definido. Já para a terceira pergunta, foram utilizados os mesmos vídeos que foram coletados na primeira pergunta, a partir dos quais foram extraídos até 150 comentários por vídeo. Esses números foram definidos, após serem verificadas as quantidades totais de vídeos e comentários que foram retornadas pela mineração inicial dos dados. Portanto, visando uma comparação justa entre os grupos e categorias, foram definidos o mesmo número de termos e consequentemente o mesmo número de vídeos para cada grupo e categoria.

3.4. Análise dos Dados

A quarta etapa consistiu na análise dos dados obtidos através da mineração em larga escala através da *Application Programming Interface* (API) YouTubeData. Essa análise foi conduzida por meio de métricas descritivas e comparativas, com o propósito de responder às três perguntas propostas no estudo. Inicialmente, foram calculadas médias e proporções de visualizações, curtidas e comentários por grupo. Logo em seguida, foi analisada a distribuição do público entre as categorias temáticas para visualizar o volume total e a proporção de visualizações por categoria dentro de cada grupo.

Nesse viés, foi investigada a distribuição temporal dos vídeos por semestre, visando identificar os períodos com maior concentração de publicações de vídeos dentro de cada categoria temática. Finalmente, a partir da frequência de termos que expressam dúvidas técnicas e de menções a IA, foi analisada a natureza dos comentários, visando inferir possíveis mudanças no objetivo de interação da audiência dos vídeos nos dois grupos.

4. Resultados

4.1. (QP1) Quais são as diferenças de engajamento entre vídeos com e sem menção à IA no ensino de programação no YouTube?

A Figura 1 apresenta os resultados obtidos a partir da análise do engajamento médio dos vídeos dos grupos IA e Tradicional, segmentados pelas linguagens de programação Python, Java e JavaScript. A análise considera três métricas de engajamento: a média de visualizações por vídeo, a média de curtidas e a média de comentários.

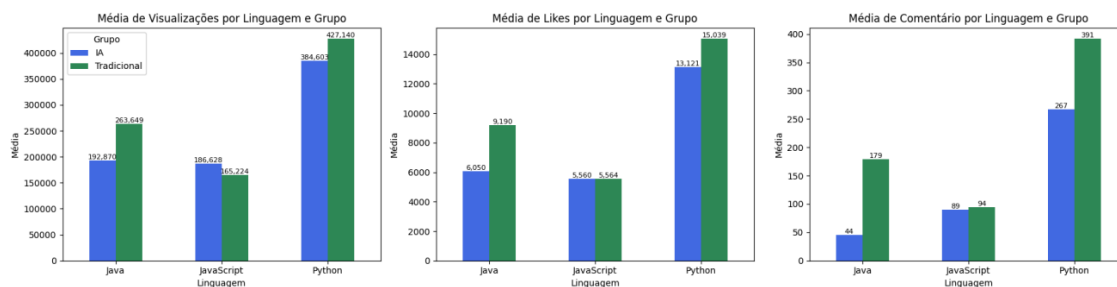


Figura 1. Engajamento Médio por Linguagem e Grupo

Na análise da média de visualizações, observa-se que, para todas as linguagens analisadas, os vídeos do Grupo Tradicional superam os do Grupo IA. A diferença é particularmente expressiva em Python, com média de 427.140 visualizações nos vídeos tradicionais, contra 384.603 no Grupo IA. Para Java, os vídeos tradicionais também apresentam média superior, sendo 263.649 contra 192.870 visualizações. A menor diferença ocorre em JavaScript, com valores mais próximos entre os grupos, sendo uma média de 186.628 visualizações no Grupo Tradicional e 165.224 visualizações no Grupo IA.

Em relação à média de curtidas, observa-se que os vídeos tradicionais obtêm maior média em todas as linguagens. No caso de Python, a média de likes nos vídeos tradicionais é de 15.039, enquanto os vídeos com IA alcançam 13.121. Para Java, a diferença também é notável, sendo uma média de 9.190 comentários no Grupo Tradicional contra uma média 6.050 no Grupo IA. Já para JavaScript os valores se aproximam, sendo uma média de 5.564 no Grupo Tradicional contra 5.560 no Grupo IA, sugerindo uma menor discrepância de engajamento nessa linguagem.

Quanto à média de comentários por vídeo, os vídeos tradicionais novamente apresentam superioridade em todas as linguagens. Em Python, a diferença é mais acentuada, sendo 391 comentários no Grupo Tradicional contra 267 no Grupo IA. Em Java, a diferença é ainda mais notável, em que são 179 comentários no Grupo Tradicional contra apenas 44 comentários no Grupo IA. Somente em JavaScript os grupos apresentam médias similares, sendo 94 no Grupo Tradicional contra 89 no Grupo IA.

4.2. (QP2) Dentro dos grupos Tradicional e IA, como está distribuído o público entre as principais categorias temáticas de vídeos de ensino de programação publicados após a popularização das IAs?

As Figuras 2, 3, 4 e 5 apresentam os dados resultantes da análise de distribuição do público entre as categorias temáticas — Cursos e Tutoriais, Projetos e Dicas — nos grupos IA e Tradicional. Para isso, foram consideradas três métricas principais: o número total de

visualizações por categoria temática em cada grupo, a proporção de visualizações de cada categoria dentro de cada grupo, e a proporção de vídeos publicados por semestre em cada categoria e grupo.

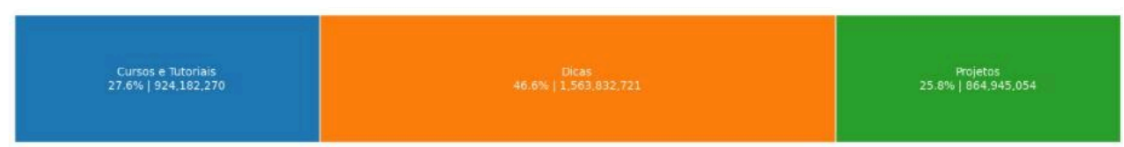


Figura 2. Distribuição Proporcional de Visualizações - Grupo IA



Figura 3. Distribuição Proporcional de Visualizações - Grupo Tradicional

As Figuras 2 e 3 apresentam a distribuição proporcional de visualizações por categoria temática, respectivamente dentro dos grupos IA e Tradicional. Observa-se que, em ambos os grupos, a categoria Dicas concentra o maior volume de visualizações: 46,6% no grupo IA e 43,3% no grupo Tradicional. Em seguida, a categoria Cursos e Tutoriais representa 27,6% das visualizações no grupo IA e 32,9% no grupo Tradicional. Por fim, a categoria Projetos aparece com 25,8% das visualizações no grupo IA e 23,7% no grupo Tradicional.

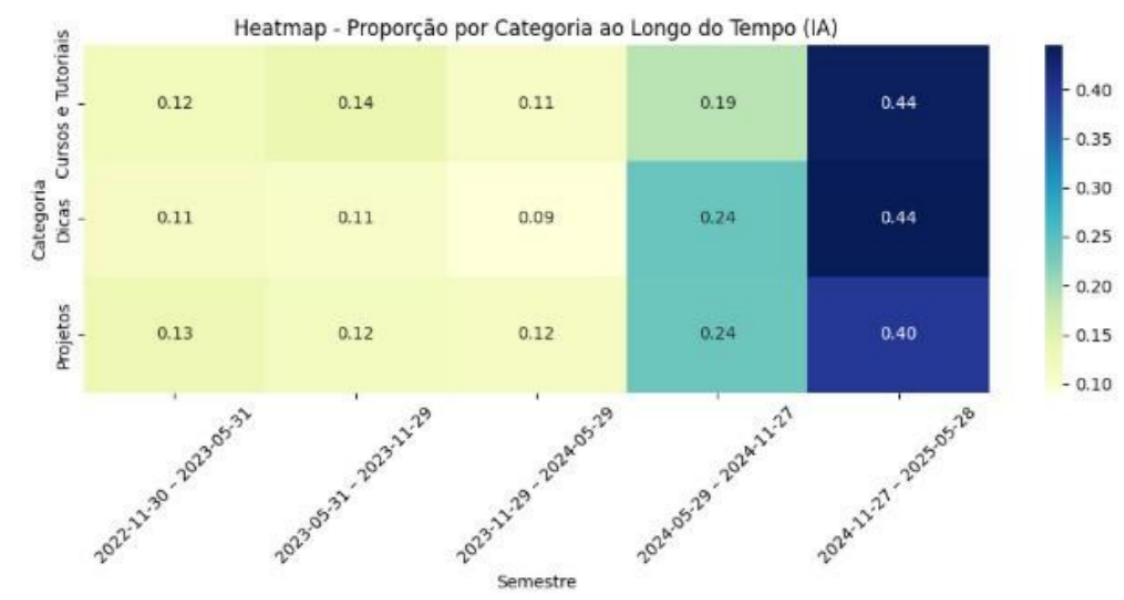


Figura 4. Proporção por categoria ao longo do tempo - Grupo IA

As Figuras 4 e 5 apresentam a proporção de vídeos publicados por categoria ao longo do tempo, respectivamente para os grupos IA e Tradicional. Os valores foram organizados por semestre a partir da data de corte adotada no estudo (30/11/2022).

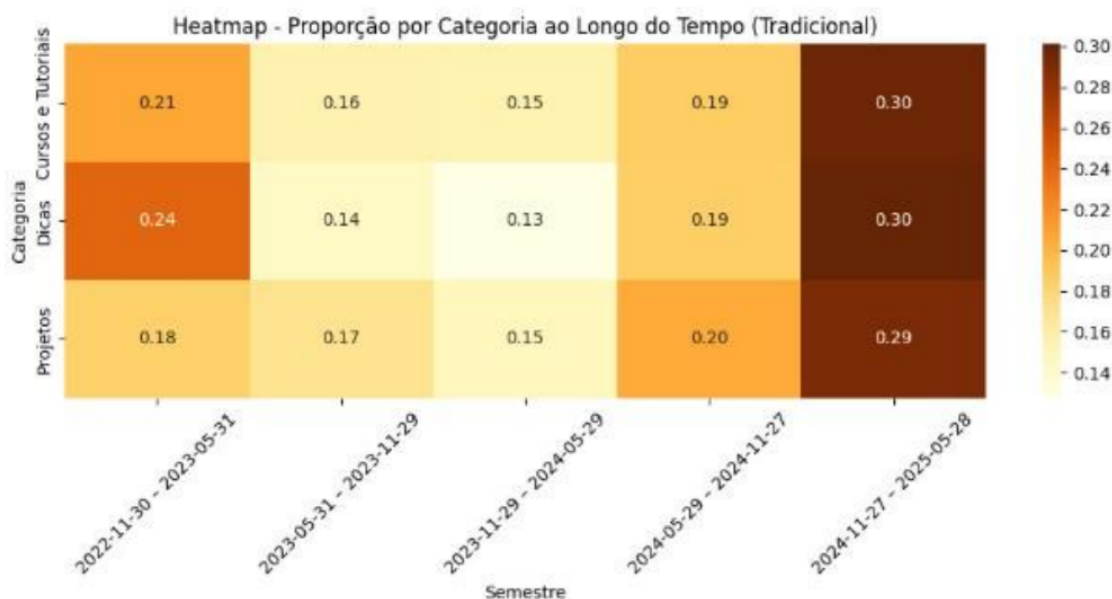


Figura 5. Proporção por categoria ao longo do tempo - Grupo Tradicional

No Grupo IA (Figura 4), observa-se um aumento expressivo na proporção de vídeos publicados nas categorias Dicas e Cursos e Tutoriais no último semestre analisado, 27 de Novembro de 2024 a 28 de Maio de 2025, ambas alcançando 44% das publicações. Já a categoria Projetos mantém uma proporção mais estável ao longo do tempo, com um leve crescimento recente de 40% no último semestre.

Por outro lado, no Grupo Tradicional (Figura 5), a evolução ao longo dos semestres é mais gradual. A partir de 2024, nota-se um crescimento constante das três categorias, resultando em um equilíbrio relativo no último semestre, com Cursos e Tutoriais (30%), Dicas (30%) e Projetos (29%) apresentando proporções bastante próximas.

4.3. (QP3) Qual a diferença na natureza das interações nos comentários entre vídeos de ensino de programação que mencionam IA e vídeos tradicionais?

A Tabela 1 e as Figuras 6 e 7 apresentam os dados obtidos a partir da análise dos comentários extraídos dos vídeos dos grupos IA e Tradicional. A análise buscou compreender a diferença na natureza das interações, considerando três métricas principais: a frequência de comentários que indicam dúvidas técnicas, a frequência de menções a IA e o número médio de comentários a cada mil visualizações.

Grupo	Total de Comentários	Dúvidas técnicas	Menções à IA	Média de comentários por 1000 visualizações
IA	74.490	2.730	10.871	1,34
Tradicional	98.129	3.034	1.623	1,93

Tabela 1. Métricas sobre os comentários nos vídeos dos grupos IA e Tradicional.



Figura 6. Distribuição proporcional dos comentários - Grupo IA



Figura 7. Distribuição proporcional dos comentários - Grupo Tradicional

No total, foram analisados 74.490 comentários de vídeos do grupo IA e 98.129 comentários de vídeos do grupo Tradicional. A frequência de comentários com dúvidas técnicas foi semelhante entre os grupos, sendo 3.034 (3,1%) no Grupo Tradicional e 2.730 (3,7%) no Grupo IA. Entretanto, quando analisadas as menções a ferramentas de IA, a diferença é significativa. No grupo IA, 10.871 comentários (14,6%) incluem termos como “ChatGPT”, “Copilot”, “GPT” ou “IA”. Já no grupo Tradicional, apenas 1.623 comentários (1,7%) contêm esse tipo de menção.

Por fim, ao observar a métrica de comentários por mil visualizações, nota-se que vídeos do grupo Tradicional apresentam uma média de 1,93 comentários a cada 1.000 visualizações, enquanto o grupo IA apresenta uma média de 1,34 comentários. Diante desse resultado observa-se que, proporcionalmente, os vídeos tradicionais geram mais interações nos comentários em relação ao seu volume de visualizações.

Referências

- BroCode (2019). Bro code. Disponível em: <https://www.youtube.com/@BroCodez>. Acesso em: 10 de maio 2025.
- C. Aranha, R. Sousa, J. B. J. R. and Silva, A. (2019). O youtube como ferramenta educativa para o ensino de ciências. *Revista de Ensino da Escola de Educação Básica da Universidade Federal de Uberlândia*.
- CodingwithLewis (2020). Coding with lewis. Disponível em: <https://www.youtube.com/@CodingWithLewis>. Acesso em: 10 de maio 2025.
- ForrestKnight (2016). Forrestknight. Disponível em: <https://www.youtube.com/@fknight>. Acesso em: 10 de maio 2025.
- GiraffeAcademy (2017). Giraffe academy. Disponível em: <https://www.youtube.com/@GiraffeAcademy>. Acesso em: 10 de maio 2025.
- GitHub (2024). Octoverse: Ai leads python to top language as the number of global developers surges. Disponível em: <https://github.blog/news-insights/octoverse/octoverse-2024/>. Acesso em: 10 de maio 2025.
- J., S. (2025). A study on user reactions to youtube thumbnails created using generative ai. *The Journal of the Convergence on Culture Technology*.

- Jürgensmeier, L. and Skiera, B. (2024). Generative ai for scalable feedback to multimodal exercises. *International Journal of Research in Marketing*, 41(3):468–488.
- Kabir, S., Udo-Imeh, D. N., Kou, B., and Zhang, T. (2024). Is stack overflow obsolete? an empirical study of the characteristics of chatgpt answers to stack overflow questions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, CHI '24, New York, NY, USA. Association for Computing Machinery.
- Kshetri, N. (2023). The economics of generative artificial intelligence in the academic industry. *Computer*, 56(8):77–83.
- Li B., Wang C., B. C. J. (2024). Exploring inventions in self-directed language learning with generative ai: Implementations and perspectives of youtube content creators. *TechTrends*.
- Paula, G., Guerra, J., Souza, L., Soares, L., Machado, P., Brito, A., and Xavier, L. (2025). Explorando o youtube como fonte de aprendizado para desenvolvedores: Uma análise em larga escala sobre vídeos de tecnologia. In *Anais do V Simpósio Brasileiro de Educação em Computação*, pages 72–84, Porto Alegre, RS, Brasil. SBC.
- TheCodingSloth (2023). The coding sloth. Disponível em: <https://www.youtube.com/@TheCodingSloth>. Acesso em: 10 de maio 2025.
- ThePrimeTime (2021). Theprimetime. Disponível em: <https://www.youtube.com/@ThePrimeTimeagen>. Acesso em: 10 de maio 2025.