

# Investigando o Uso de Inteligência Artificial em Projetos Brasileiros do GitHub

Luiz Andre do Nascimento Ubaldo, Jailton Coelho  
Instituto Federal de Educação, Ciência e Tecnologia do Paraná (IFPR)  
Campus Telêmaco Borba – Telêmaco Borba, PR – Brasil  
20213017755@estudantes.ifpr.edu.br, jailton.coelho@ifpr.edu.br

**Abstract**—Artificial Intelligence (AI) has played an important role in software development, particularly in open-source projects. This study investigates the use of AI in Brazilian GitHub repositories, aiming to identify the most used programming languages, the lifespan of these repositories, and the most frequent areas of AI application. Using a dataset composed of 30 Brazilian open-source projects with the highest number of stars on GitHub, three key areas of AI application were found: machine learning, deep learning, and computer vision. Python leads as the most used programming language, followed by JavaScript, with C++ and C# having the same usage, and Shell and TypeScript also tied after them. The median lifespan of these repositories is approximately 66 months, showcasing the recent maturity of these initiatives in Brazil.

**Keywords**—GitHub; Artificial Intelligence; Brazilian OSS Community.

**Resumo**—A Inteligência Artificial (IA) tem desempenhado um papel importante no desenvolvimento de software, especialmente em projetos de código aberto. Este estudo investiga o uso de IA em repositórios brasileiros do GitHub, com o objetivo de identificar as linguagens de programação mais utilizadas, o tempo de vida desses repositórios e as áreas de aplicação mais frequentes da IA. Utilizando um conjunto de dados composto por 30 projetos de código aberto brasileiros com maior número de estrelas no GitHub, foram encontradas três áreas principais de aplicação: aprendizado de máquina, aprendizagem profunda e visão computacional. Python lidera como a linguagem de programação mais utilizada, seguida por JavaScript, C++ e C# em igual proporção, com Shell e TypeScript também empatados em seguida. A mediana de vida desses repositórios é de aproximadamente 66 meses, demonstrando a recente maturidade dessas iniciativas no Brasil.

**Palavras-chave**—GitHub; Inteligência Artificial; Comunidade Brasileira de Código Aberto.

## I. INTRODUÇÃO

A Inteligência Artificial (IA) tem influenciado em diversos setores do mercado mundial, e seu impacto é perceptível em projetos de código aberto [1]. Ferramentas de geração de código assistido por IA estão se tornando cada vez mais comuns na área de computação, oferecendo a capacidade de criar código a partir de comandos em linguagem natural ou entradas de código parciais. Exemplos notáveis dessas ferramentas incluem

o GitHub Copilot<sup>1</sup> e o ChatGPT da OpenAI<sup>2</sup>. Essas ferramentas contribuem para reduzir a necessidade de escrita manual de código e tornam o processo de desenvolvimento mais eficiente [2].

No Brasil, a Estratégia Nacional de IA, lançada em 2021, tem impulsionado o setor, com investimentos crescentes em *startups* de IA, embora ainda abaixo de países como EUA e China [3]. Em 2019, por exemplo, o Brasil investiu cerca de US\$ 1 milhão em *startups* de IA, comparado a US\$ 224 milhões nos EUA e US\$ 45 milhões na China [4].

Apesar da crescente popularidade da Inteligência Artificial nos últimos anos, será que ela também começou a ser usada no desenvolvimento de projetos brasileiros baseados em código-objeto? Este estudo pretende explorar essa questão e responder às seguintes questões de pesquisa:

**QP1.** Quais as linguagens de programação mais utilizadas pelos repositórios brasileiros que usam Inteligência Artificial?

**QP2.** Qual o tempo de vida dos repositórios brasileiros do GitHub que utilizam Inteligência Artificial?

**QP3.** Em quais áreas da Inteligência Artificial os repositórios brasileiros do GitHub estão sendo utilizados?

Entender quais linguagens, áreas e características de tempo de vida dominam os repositórios brasileiros de IA no GitHub é essencial para mapear o desenvolvimento de tecnologias locais e identificar possíveis lacunas no uso de ferramentas e metodologias de Inteligência Artificial. Os resultados deste estudo podem ser aplicados em diversos contextos, como em iniciativas de educação e treinamento em IA, no desenvolvimento de políticas públicas para fomento à tecnologia, e na definição de estratégias de investimento por *startups* e empresas que desejam se alinhar às práticas mais adotadas no país. Essa análise oferece uma visão panorâmica da adoção de IA no Brasil, indicando tendências, desafios e oportunidades para a comunidade de código aberto nacional. Este estudo auxilia desenvolvedores, pesquisadores e formuladores de políticas a

<sup>1</sup><https://github.com/features/copilot>

<sup>2</sup><https://openai.com/chatgpt/>

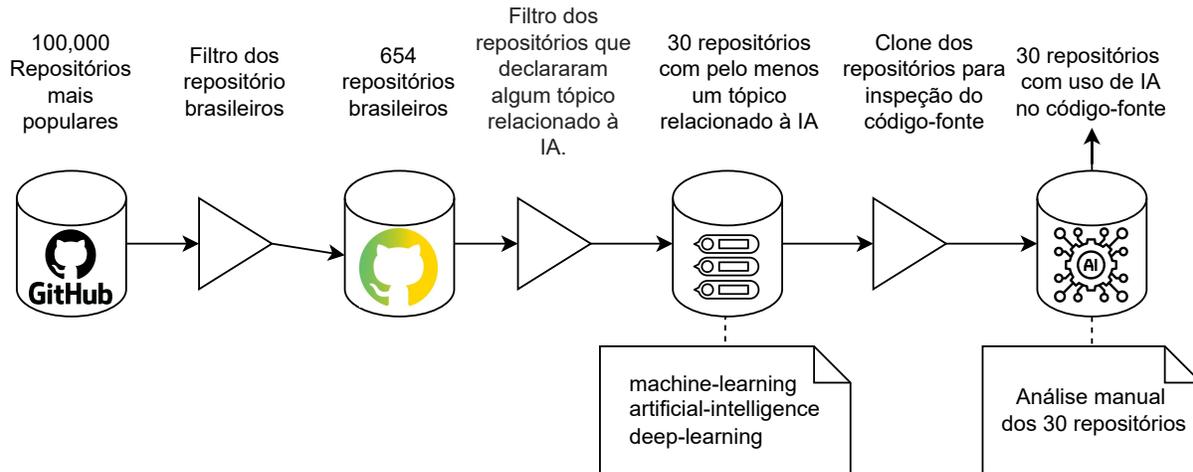


Fig. 1: Etapas realizadas para a coleta das informações dos repositórios brasileiros do GitHub.

reconhecer as áreas de destaque em IA, facilitando investimentos mais direcionados e promovendo avanços tecnológicos adaptados às necessidades do Brasil.

O restante deste artigo está organizado conforme descrito a seguir. A Seção II descreve como foi realizada a coleta das informações. A Seção III apresenta os resultados obtidos através da análise dos dados. A Seção IV aborda as ameaças à validade deste estudo. A Seção V apresenta as conclusões finais.

## II. PROCEDIMENTOS METODOLÓGICOS

A seguir, estão descritos os passos realizados para coleta de informações dos repositórios brasileiros e validação dos dados coletados, resumidos na Figura 1:

**Top 100.000 repositórios do GitHub.** O primeiro passo foi criar uma lista dos 100.000 repositórios mais populares do GitHub, critério similar utilizado em outros estudos [5, 6].

**Repositórios brasileiros.** No segundo passo, foram extraídos apenas os repositórios cuja localização do proprietário do repositório fizesse parte de alguma região do Brasil. Neste passo, foi gerada uma lista com 654 repositórios brasileiros.

**Repositórios com algum tópico relacionado à IA.** No terceiro passo, foram filtrados os repositórios que declararam algum tópico relacionado à IA. Foi utilizado um total de 10 tópicos, que são: *artificial intelligence*, *large language models*, *image generation*, *deep neural networks*, *machine learning*, *deep learning*, *neural network*, *computer vision*, *object detection* e *generative ai*. Os tópicos de um repositório do GitHub servem para categorizar e descrever o conteúdo do repositório [7].

Ao final deste passo, foi obtida uma lista com 30 repositórios brasileiros relacionados à Inteligência Artificial.

**Inspeção manual do código-fonte dos repositórios.** Todos os repositórios do passo anterior foram manualmente inspecionados para confirmar se realmente tinham relação com algum tópico de IA. E, de fato, foi confirmado para todos eles.

## III. RESULTADOS E DISCUSSÕES

### QP1. Quais as linguagens de programação mais utilizadas pelos repositórios brasileiros que usam Inteligência Artificial?

O gráfico da Figura 2 mostra os resultados obtidos com a análise dos 30 repositórios brasileiros do GitHub relacionados à Inteligência Artificial. Pode-se observar que Python é a linguagem de programação mais utilizada, com 17 repositórios, seguida por JavaScript, com 3 repositórios. C++ e C# aparecem empatados, cada um com 2 repositórios e, por fim, Shell e TypeScript ocupam a última posição, ambos com 1 repositório cada. A predominância de Python nos repositórios analisados evidencia sua popularidade na área de Inteligência Artificial, impulsionada por bibliotecas robustas como TensorFlow e PyTorch. A presença de outras linguagens, como JavaScript, C++, C#, Shell e TypeScript, embora menos frequente, sugere que os desenvolvedores adaptam a escolha da tecnologia às necessidades específicas de cada projeto.

Quatro repositórios brasileiros não possuíam uma linguagem de programação definida no GitHub: THE-AKIRA/COMPUTER-SCIENCE-RESOURCES (coleta de recursos educacionais de ciência da computação),

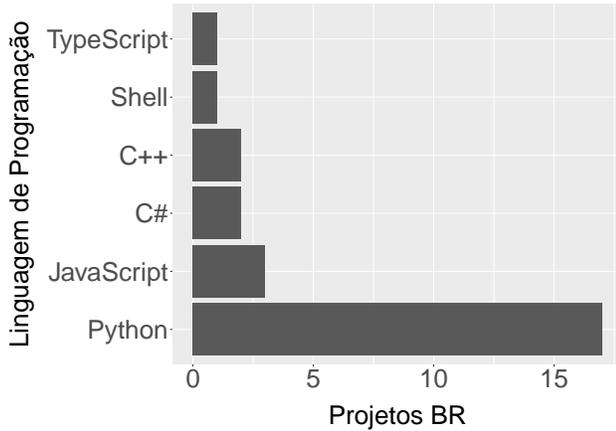


Fig. 2: Linguagens de programação dos repositórios brasileiros do GitHub relacionados à IA.

PIZZADEDADOS/DATASCIENCE-PIZZA (materiais sobre ciência de dados), FILIPECALEGARIO/AWESOME-GENERATIVE-AI (curadoria de ferramentas e projetos de IA generativa) e LEOMAURODESENV/GAME-DATASETS (conjunto de dados para desenvolvimento de jogos).

### QP2. Qual o tempo de vida dos repositórios brasileiros do GitHub que utilizam Inteligência Artificial?

A Figura 3 mostra a distribuição dos tempos de vida dos repositórios brasileiros do GitHub que utilizam Inteligência Artificial, calculados como a diferença entre a data da última atualização e a data de criação do repositório. O tempo mínimo de existência registrado é de quase 2 meses. O primeiro quartil (25%) demonstra que um quarto dos repositórios possuem tempos de vida abaixo de 48 meses. A mediana dos tempos de vida é de aproximadamente 66 meses, indicando que há um histórico de atualizações consistentes. O terceiro quartil (75%) aponta que 25% dos repositórios têm mais de 87 meses de atividade, demonstrando que um número expressivo desses projetos supera os sete anos de atividade. O repositório com o maior tempo de vida, 146 meses, refere-se ao ACCORD-NET/Framework, que possui mais de 4,5 mil estrelas, 2 mil forks, mas que atualmente encontra-se *archived*, ou seja, sem manutenção.

### QP3. Em quais áreas da Inteligência Artificial os repositórios brasileiros do GitHub estão sendo utilizados?

A Figura 4 mostra uma nuvem de palavra com os termos mais citados nos *topics* dos repositórios. Pode-se observar

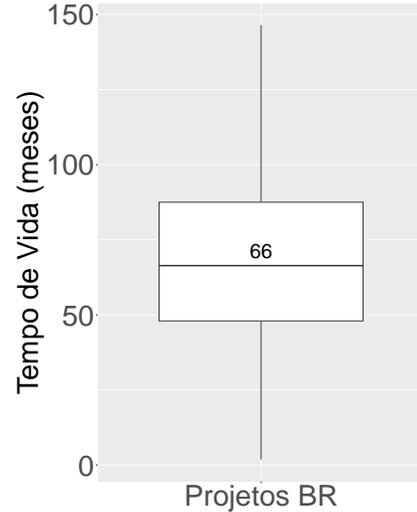


Fig. 3: Distribuição do tempo de vida (em meses) dos repositórios brasileiros do GitHub relacionados à IA.

que destacam-se três conceitos-chave: *machine learning*, *deep learning* e *computer vision*. *Machine learning* refere-se a sistemas que aprendem e melhoram a partir da experiência, sem serem explicitamente programados para tal, utilizando algoritmos que analisam dados, reconhecem padrões e tomam decisões com base em informações históricas [8]. *Deep learning*, um subcampo do machine learning, utiliza redes neurais artificiais com múltiplas camadas para modelar e entender dados complexos, sendo eficaz em tarefas como reconhecimento de fala, tradução de idiomas e processamento de imagens [9]. Por fim, *Computer vision* é um campo da Inteligência Artificial que busca capacitar máquinas a interpretar e entender o mundo visual, emulando a capacidade humana de percepção visual. Ela envolve a aquisição, processamento e análise de imagens e vídeos, permitindo a identificação de objetos, reconhecimento facial e segmentação de imagens.

### IV. AMEAÇAS À VALIDADE

Nesta seção, são apresentadas as ameaças à validade deste estudo, bem como as estratégias adotadas para mitigá-las.

Primeiramente, em relação à validade de construção, o tamanho total da população de repositórios brasileiros no GitHub que utilizam Inteligência Artificial não é conhecido. Para mitigar essa ameaça, foi utilizada uma amostra representativa de 30 repositórios com maior número de estrelas, garantindo uma quantidade relevante para a análise.

Quanto à validade interna, a presença de tópicos relacionados à Inteligência Artificial nos repositórios não garante que o projeto esteja efetivamente utilizando essas tecnologias. Isso

