

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ – UTFPR**

**Câmpus de Ponta Grossa**

**Curso de Ciência da Computação**

**Departamento Acadêmico de Informática**

**Disciplina de Comunicação Linguística**

**ANÁLISE TÉCNICA SOBRE O DESEMPENHO DA TECNOLOGIA GPT-3.5 NA  
RESOLUÇÃO DE PROBLEMAS LIGADOS A ESTRUTURAS DE DADOS E A  
PROGRAMAÇÃO DE ALGORITMOS**

ARTHUR BALDOQUI BERGAMO

NICOLAS FELIX HRESCAK

TIAGO PANIZIO GOTTARDO

## **RESUMO**

Foi feita uma breve revisão introdutória sobre o assunto de *Large Languages Models* (LLMs) que é a base tecnológica da tecnologia de *Generative Pre-Trained Transformer* (GPT), sendo a tecnologia testada nesta pesquisa. Utilizou-se uma inteligência artificial chamada GPT-3.5 - que é uma GPT - para resolver exercícios e questões de programação de algoritmos e estrutura de dados, a fim de medir suas capacidades de resolução e quantificar através de dados o quão confiável a ferramenta é para esta tarefa. Mediu-se portanto a qualidade das respostas das questões levando em consideração sua dificuldade, assunto e data de publicação da questão. O ano da questão é relevante pois a tecnologia é treinada com dados de até 2021, e assim questões posteriores são “inéditas” para a inteligência artificial.

Palavras-Chave: Algoritmo. GPT-3.5. Tarefas. Exercícios. Dados.

## **1 - INTRODUÇÃO**

Problemas de programação são desafios frequentemente propostos em competições e maratonas de programação para medir a capacidade dos participantes em relação a conceitos de estrutura de dados, programação dinâmica e criação de algoritmos eficientes. Essas competições, como a Olimpíada Brasileira de Informática (OBI) e a Maratona SBC de Programação, são conhecidas por apresentarem problemas complexos que requerem um amplo conhecimento teórico e habilidades práticas de resolução de problemas.

Nos últimos anos, os avanços em inteligência artificial têm despertado o interesse na aplicação dessas tecnologias para a resolução de problemas desafiadores.

Uma das tecnologias mais promissoras nesse sentido é o modelo GPT-3.5, que utiliza uma rede neural conhecida como Transformadores Gerativos Pré-Treinados (GPT). “Os (GPTs) geram profundas transformações, potencializando o crescimento tecnológico, permeando tarefas, impactando profundamente as profissões.”(ELOUNDOU et al, 2023, tradução feita pelos autores).-

Portanto, é importante avaliar o desempenho da tecnologia GPT-3.5 especificamente em relação à resolução de problemas ligados a estruturas de dados e programação de algoritmos. Uma questão fundamental é se a tecnologia é capaz de compreender os conceitos complexos dessas áreas e fornecer soluções precisas e eficientes, assim como um programador experiente seria capaz de fazer.

O objetivo deste trabalho é avaliar a capacidade da tecnologia GPT-3.5 na resolução de problemas específicos dessas áreas. Pretendemos investigar se a tecnologia é mais eficaz na resolução de problemas antigos, cujo treinamento da rede neural ocorreu com dados até 2021, ou se também é capaz de lidar com problemas mais recentes, considerando que sua base de treinamento é anterior a esse período.

Esta análise técnica é relevante e oportuna, pois contribui para o avanço do conhecimento sobre a aplicação da inteligência artificial na resolução de problemas desafiadores da área de programação. Além disso, pode demonstrar se há o potencial de substituição de parte do trabalho humano - no mercado de trabalho relacionado a programação - por sistemas automatizados em um contexto cada vez mais dependente de tecnologias avançadas.

## **2 - Desenvolvimento**

### **2.1 METODOLOGIA**

Para atingir o objetivo proposto, foram selecionados 220 exercícios de concursos e competições de programação que ocorreram no período de 2004 a 2023. Esses exercícios foram escolhidos de forma aleatória, com o intuito de abranger diferentes níveis de dificuldade e tópicos relacionados a estruturas de dados e programação de algoritmos.

Os exercícios foram então submetidos a uma inteligência artificial chamada GPT-3.5. Cada exercício foi fornecido como entrada para o modelo GPT-3.5, e a saída gerada pelo modelo foi avaliada utilizando uma ferramenta de correção para problemas de programação.

Durante o processo de teste, foram coletados dados sobre os acertos - que podem ser parciais (5% a 100% da questão) - e erros da tecnologia GPT-3.5 para cada exercício. Essas informações foram registradas, permitindo a análise posterior do desempenho da tecnologia.

A análise do trabalho leva em consideração três dados intrínsecos de cada questão: o ano de publicação da questão, a dificuldade atribuída a ela, categorizada de 1 a 10, e o assunto da questão. Tanto o nível quanto o assunto da questão foram catalogadas por uma plataforma virtual chamada “beecrowd”.

O assunto da questão é categorizado em 8 tipos, sendo eles termos em inglês que descrevem o campo de conhecimento necessário para resolução são eles: *Data Structures and Libraries*, *Strings*, *Ad-Hoc*, *Beginner*, *Mathematics*, *Computational Geometry*, *Graph*, *Paradigms*. Os termos estão em inglês como encontrados na plataforma, não foi analisado o que cada termo se refere, e sim como a inteligência artificial lida com os diferentes tipos de questões.

A fim de verificar possíveis diferenças no desempenho da tecnologia GPT-3.5, os resultados obtidos serão apresentados em tabelas e gráficos. Será realizada uma análise comparativa entre os acertos de acordo com as datas, dificuldades e assuntos de cada questão, com o objetivo de identificar padrões ou tendências. Essa análise permitirá avaliar a possibilidade do uso da tecnologia na resolução de questões inéditas de eventuais provas, bem como explorar a expertise da tecnologia na área de algoritmos, o que pode ter impacto relevante nesse campo.

## 2.2 REFERENCIAL TEÓRICO

A tecnologia GPT-3.5 é uma tecnologia baseada em Modelos de Linguagens Grandes (LLMs), “LLMs treinados em corpos maciços de textos mostraram sua capacidade de realizar novas tarefas a partir de instruções textuais ou de alguns exemplos” (BROWN, 2020, tradução feita pelos autores).

Estudos recentes têm explorado a capacidade do GPT-3.5 na resolução de problemas, Eloundou et al. (2023) investigaram o impacto do desenvolvimento de

modelos GPT no mercado de trabalho e destacaram que 15% dos trabalhos nos Estados Unidos podem ser potencialmente afetados positivamente por modelos GPT. Essa pesquisa propôs tarefas anteriormente realizadas por trabalhadores humanos para serem executadas pela inteligência artificial.

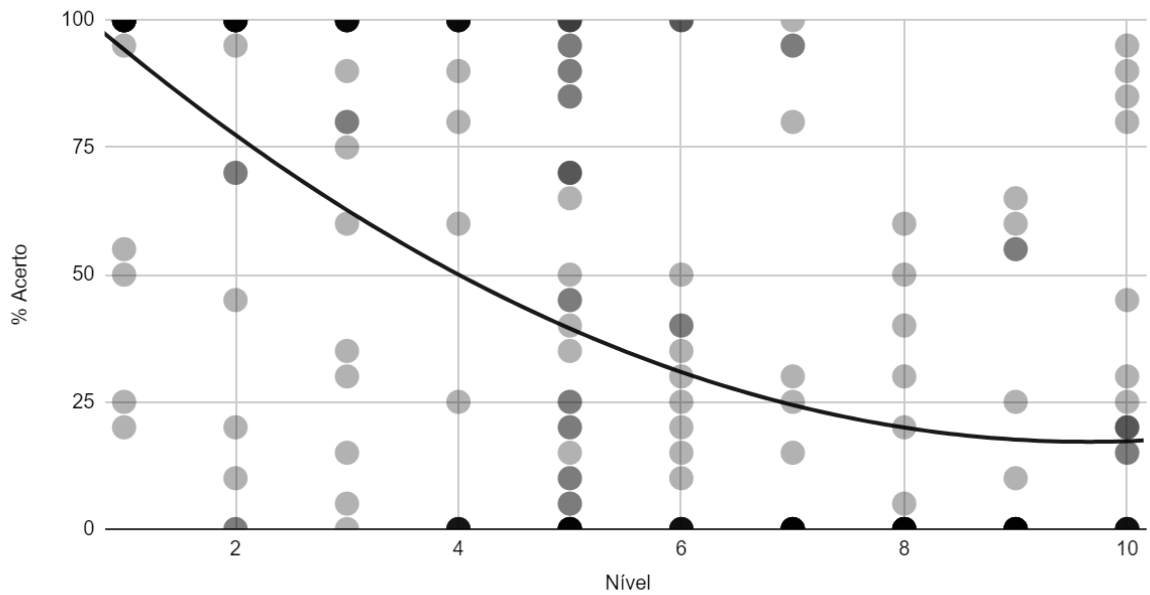
No contexto específico da resolução de problemas de programação, é fundamental avaliar a capacidade do GPT-3.5 em lidar com desafios relacionados a estruturas de dados, algoritmos e programação dinâmica. A análise dos resultados obtidos com a aplicação do GPT-3.5 nesse contexto pode fornecer insights valiosos sobre o potencial dessa tecnologia em substituir parte do trabalho humano na área de programação.

Além disso, é importante considerar as limitações e desafios associados à aplicação do GPT-3.5 na resolução de problemas de programação. Dados de treinamento específicos para problemas dessa natureza podem ser escassos, o que pode impactar o desempenho da tecnologia.

## **2.2 RESULTADOS**

Foi coletado os dados de acertos e erros pela tecnologia, das 220 questões. A partir daí nota-se no gráfico a seguir, como esperado, uma maior taxa de acertos em questões de níveis mais baixos. O gráfico mostra a dispersão dos acertos parciais, sendo os pontos mais escuros uma sobreposição de acertos, e a curva de tendência demonstra que a tecnologia tende a acertar questões de níveis mais baixos.

Gráfico 1 - Dispersão geral dos acertos.

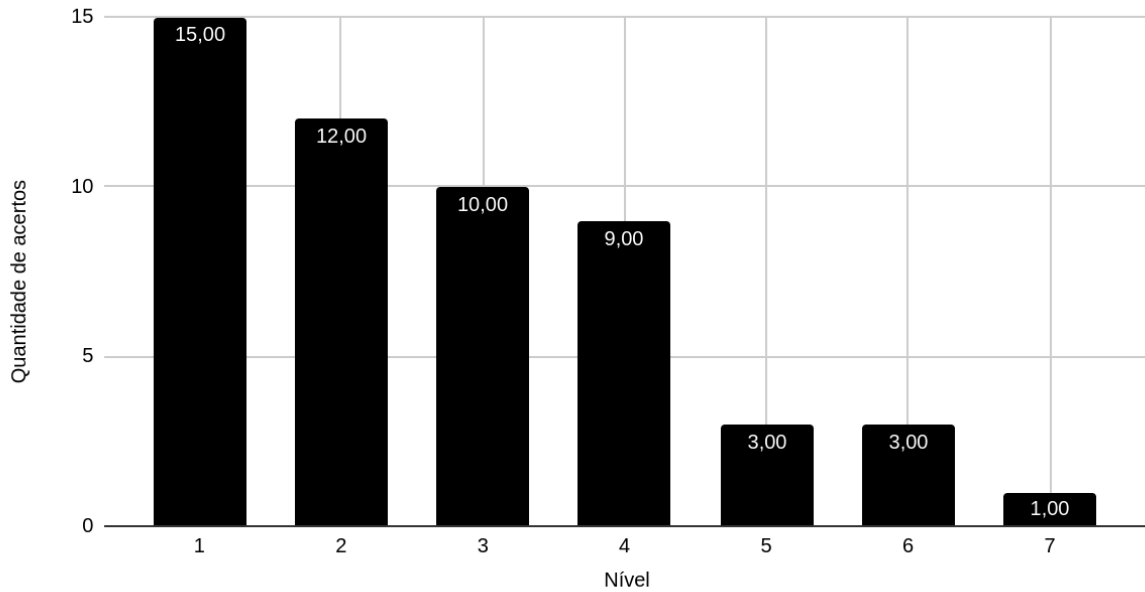


Fonte: Elaborado pelos autores (2023)

O gráfico demonstra que, em questões de nível mais baixo, a tecnologia consegue gerar soluções para o problema de modo a resolvê-lo completamente, e conforme o nível aumenta, mais acertos parciais de valor baixo são observados.

O mesmo ocorre quando é observado a quantidade de acertos - 100% da questão - desprezando os acertos parciais, temos que em questões de nível baixo o modelo GPT teve mais sucesso do que de níveis altos, de modo que em questões a partir do nível 7 não houve acertos.

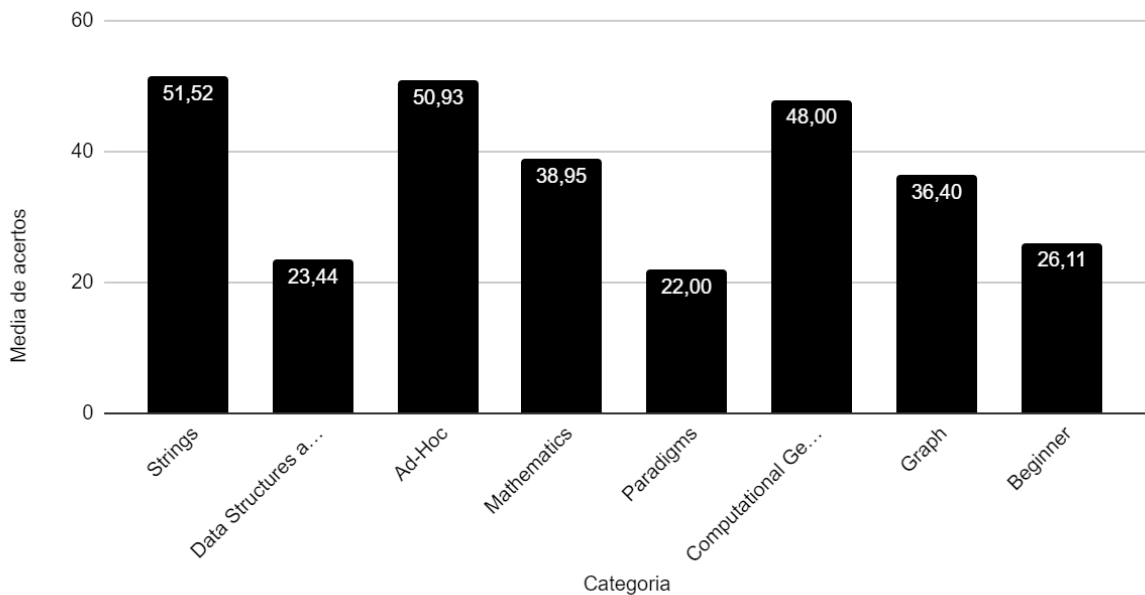
Gráfico 4 - Acertos por nível de questão.



Fonte: Elaborado pelos autores (2023)

Outrossim, notou-se que a média de acertos variou de acordo com a categoria da questão, como é possível observar no gráfico abaixo que apresenta a média simples dos acertos - calculada sobre a porcentagem do acerto parcial das questões - em comparação com cada categoria testada.

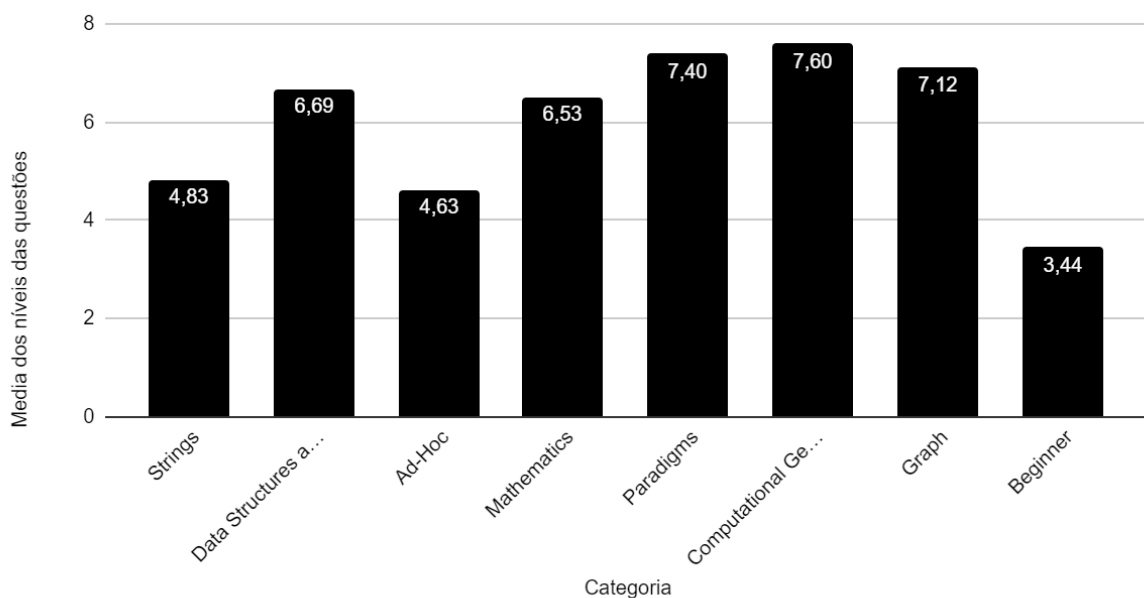
Gráfico 2 - Média de acertos por categoria.



Fonte: Elaborado pelos autores (2023)

É possível observar no gráfico que o desempenho foi melhor nas categorias classificadas como, *Strings*, *Ad-Hoc* e *Computational Geometry*, e que o desempenho foi pior nas questões classificadas como, *Data Structures and Libraries*, *Paradigms* e *Beginner*. Uma possível explicação para a diferença de acertos entre as categorias, poderia ser um desbalanceamento entre os níveis das questões para cada categoria, e portanto deveríamos observar um maior nível nas questões de *Paradigms* do que de *Ad-Hoc*. Podemos fazer essa comparação através do gráfico abaixo.

Gráfico 3 - Média dos níveis das questões por categoria.



Fonte: Elaborado pelos autores (2023)

É observado no gráfico o esperado, com exceção da categoria *Computational Geometry*, que apesar do alto nível médio, a inteligência artificial conseguiu ter uma boa média de acertos.

Observou-se ainda que a tecnologia teve um desempenho em 20 questões de 2022 (únicas questões posteriores a 2021 na base de testes) de 25,2% de média dos acertos parciais. No entanto, a tecnologia performou muito melhor para questões de mesmo nível mais antigas, com uma média de 53% calculada sobre os acertos parciais. Ou seja, o modelo resolve de forma mais assertiva questões antigas.

### 3 - CONCLUSÃO

Os testes demonstraram que o modelo é muito bom respondendo as questões mais fáceis, gabaritando quase todas, o que demonstra uma boa capacidade da tecnologia em interpretar, entender e resolver questões postas.

Este trabalho se limita a considerar a capacidade do modelo GPT-3.5 de resolver problemas específicos. Delimitar quais tipo tarefas podem ser feitas pelo modelo e que hoje são feitas por programadores, não faz parte do escopo desta pesquisa. Tal perspectiva deve ser abordada por trabalhos seguintes.

Sobre as limitações da pesquisa, a base de questões utilizadas é pequena, em especial quando olhamos para as questões posteriores a 2021, apenas 20 questões foram testadas, e para resultados mais conclusivos seria necessário novas pesquisas com uma base de testes maior. Outrossim como a classificação de nível da questão é feita por terceiros, não há como verificar a metodologia usada para esta classificação, nem portanto sua relevância para a pesquisa.

## **BIBLIOGRAFIA**

BROWN, T. B. et al. Language Models are Few-Shot Learners. arxiv.org, 28 maio 2020.

ELOUNDOU, T. et al. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. arXiv:2303.10130, 17 mar. 2023. Disponível em: <https://arxiv.org/abs/2303.10130>. Acesso em: 18 jun. 2023.

## **ABSTRACT**

A brief introductory review was made on the subject of Large Languages Models (LLMs) which is the technological basis of the Generative Pre-Trained Transformer (GPT) technology, being the technology tested in this research. An artificial intelligence called GPT-3.5 - which is a GPT - was used to solve exercises and questions of algorithm programming and data structure, in order to measure its resolution capabilities and quantify through data how reliable the tool is for This task. Therefore, the quality of the answers to the questions was measured taking into account their difficulty, subject and date of publication of the question. The year of the question is relevant as the technology is trained on data as far back as 2021, and so later questions are "unprecedented for artificial intelligence".

**KEYWORDS:** Algorithm. GPT-3.5. Activity. Exercise. Data.