

¹ChatGPT versus Stack Overflow: Um comparativo no contexto de tópicos em estrutura de dados.

Igor Felipe, Rafael Magalhães
{igor.felipe@dcx.ufpb.br, rafael@dcx.ufpb.br}

Resumo: O Relatório Técnico tem como objetivo comparar as respostas fornecidas pelo ChatGPT e pelo Stack Overflow para questões de programação de estruturas de dados e explorar as características de cada uma dessas ferramentas em relação à parte descritiva e a parte objetiva de códigos. O trabalho emprega uma metodologia de análise comparativa de perguntas e respostas em ambos os sistemas, onde foram coletadas questões sobre estruturas de dados e as respostas foram analisadas e comparadas quanto aos critérios subjetivos por parte de avaliadores especialistas e objetivos quantitativos a partir de métricas calculadas computacionalmente. Os resultados obtidos podem fornecer subsídios para futuras pesquisas e para aprimorar o uso de tecnologias educacionais no processo de ensino e aprendizagem.

Palavras-chave: ChatGPT, Estrutura de Dados, Stack Overflow, Tecnologias educacionais, Ensino e aprendizagem

Sumário de Conteúdo do Relatório

1. Introdução	2
2. Fundamentação e Contextualização	3
2.1. ChatGPT	3
2.2 Stack Overflow	4
2.3 Motivação	4
3. Metodologia	5
4. Experimentos	6
4.1 Conjunto de Experimentos - Avaliação por especialistas	9
4.2 Análise das Resultados do Formulário com os Especialistas	9
4.3 Conjunto de Experimentos Quantitativos - Ferramentas Computacionais	11
4.3.1 Análise de Métricas de qualidade de código	11
4.3.1 Análise de linguagem a partir de NLP	13
4.3.1 Análise de texto e código automática por modelo de linguagem GPT	14
5. Considerações Finais	15
Referências	16

¹ “Trabalho de Conclusão de curso, sob orientação do professor < Rafael Marrocos Magalhães> submetido ao curso de Licenciatura Em Ciência Da Computação do Centro De Ciências Aplicadas e Educação (CCAE) da Universidade Federal Da Paraíba, como parte dos requisitos necessários para obtenção do grau de LICENCIADO EM CIÊNCIA DA COMPUTAÇÃO”

1. Introdução

A programação é uma habilidade cada vez mais valorizada em um mundo cada vez mais digital. Com o aumento da demanda por profissionais qualificados em programação, é crucial que os recursos educacionais estejam disponíveis para ajudar os alunos a compreender e aplicar conceitos complexos de programação, como as estruturas de dados.

São inúmeros os recursos disponíveis para auxiliar a aprendizagem de programação e a execução diária de sua prática. Recursos com intuito informativo, de exemplos mas também de discussão e aprendizagem mútua. Uma das fontes mais populares para essas informações é o site Stack Overflow (S.O.), que fornece um fórum para programadores compartilharem conhecimentos e soluções para problemas de programação.

Dentre outros recursos educacionais existentes, percebe-se grande incremento nos que incorporam ou utilizam tecnologias de inteligência artificial e tecnologias correlatas. Conforme um artigo publicado na revista científica "Education Technology Insights" (2021), a inteligência artificial é uma das principais tecnologias que têm transformado a educação nos últimos anos. Dentre os principais exemplos, destacam-se: sistemas de tutores inteligentes, robótica educacional, avaliadores inteligentes, codificação mediada por bots, dentre outros.

Dentre as contribuições neste segmento, um recurso relativamente novo e de maior proeminência destaca-se. Conhecido como ChatGPT, a tecnologia de I.A. é um modelo de processamento de linguagem natural que pode responder a questões e iniciar conversas sobre diversos tópicos, incluindo programação. Enquanto alguns autores questionam sobre seu real potencial como ferramenta (Pimentel, et al. 2023) outros autores (Nascimento, 2023) já consideram que "o ChatGPT pode impulsionar a transformação da educação e trazer para o centro do aprendizado o estímulo à criatividade e o desenvolvimento socioemocional dos alunos".

No Stack Overflow, os usuários podem fazer perguntas relacionadas a um tópico específico e receber respostas de outros membros da comunidade. As respostas são avaliadas pelos próprios usuários, o que significa que as respostas mais úteis e precisas sobem ao topo da lista. Deste modo, a sua grande quantidade de conteúdo útil para a comunidade de programadores, com milhões de perguntas e respostas em seu banco de dados é, de certo modo, validado e referendado. No entanto, o Stack Overflow pode ter algumas limitações. As respostas fornecidas pelos usuários nem sempre são precisas ou atualizadas, e pode haver vários métodos diferentes para resolver um problema. Além disso, as respostas podem ser difíceis de entender para aqueles que não são proficientes em inglês ou que têm pouca experiência em programação.

O ChatGPT, por outro lado, é uma nova tecnologia que pode fornecer respostas personalizadas e precisas para perguntas de programação. O modelo de processamento de linguagem natural foi treinado em uma grande quantidade de dados para entender a linguagem natural e gerar respostas precisas para uma variedade de perguntas. Isso significa que os usuários podem fazer perguntas em sua própria língua materna e obter respostas precisas e personalizadas. Porém, assim como o Stack Overflow, o ChatGPT também pode ter algumas limitações. Embora seja capaz de fornecer respostas precisas, não é capaz de avaliar a qualidade das respostas ou a melhor maneira de resolver um problema.

Neste sentido, este trabalho visa comparar as respostas fornecidas pelo ChatGPT e pelo Stackoverflow para questões de programação de estruturas de dados e explorar o potencial de cada uma dessas ferramentas como auxiliares de aprendizagem na disciplina de programação. Os objetivos deste trabalho são os de comparar e contrastar as respostas fornecidas pelo ChatGPT e pelo Stackoverflow para questões de estruturas de dados, avaliar a qualidade das respostas fornecidas e determinar se uma ferramenta é mais adequada do que a outra para determinados tipos de perguntas.

Para atingir esses objetivos, empregou-se uma metodologia de análise comparativa de perguntas e respostas em ambos os sistemas Stack Overflow e ChatGPT. Para tanto um experimento de coletas de questões sobre estruturas de dados em ambos os sistemas e as respostas foram analisadas e comparadas quanto aos critérios específicos.

A partir dessa análise, pretende-se identificar as principais vantagens e desvantagens do uso do ChatGPT em relação ao Stack Overflow na disciplina de Estrutura de Dados. Os resultados obtidos podem fornecer subsídios para futuras pesquisas e para aprimorar o uso de tecnologias educacionais no processo de ensino e aprendizagem.

2. Fundamentação e Contextualização

Esta seção apresenta as duas tecnologias avaliadas, ChatGPT e Stack Overflow, e a motivação para realização dos experimentos do trabalho.

2.1. ChatGPT

O ChatGPT é uma tecnologia avançada que facilita a forma como as pessoas interagem com a inteligência artificial. Segundo sua própria desenvolvedora, a OpenAI (OPENAI, 2023), ele é "um modelo de processamento de linguagem natural (NLP) que utiliza um algoritmo avançado para responder perguntas e iniciar conversas em diversos temas". Segundo Pimentel (2023), ele é treinado com base em grandes volumes de dados de texto, e é capaz de compreender e processar informações de uma forma que imita a linguagem natural humana, o que o torna uma ferramenta poderosa para a aprendizagem, resolução de problemas e tomada de decisões.

O ChatGPT pode ser usado em diversas áreas, como educação, medicina, direito e atendimento ao cliente. Na educação (NASCIMENTO, 2023), ele tem sido adotado como uma ferramenta de apoio para a aprendizagem dos conteúdos vistos em sala de aula, ajudando os alunos a tirar dúvidas, resolver problemas matemáticos, entre outras funções. Na medicina (ESTADÃO, 2023), o ChatGPT tem sido utilizado para contribuir com o debate sobre validação de informações na área de Saúde, avaliar protocolos e resultados clínicos e experimentais. Já no campo jurídico (FÉRLIN, 2023), ele pode ser usado como uma ferramenta para responder perguntas sobre um corpus jurídico, desenvolver minutas, analisar argumentos, dentre outras.

Além das aplicações mencionadas, o ChatGPT tem um grande potencial em diversos setores e pode contribuir para o avanço tecnológico e melhorar a eficiência e qualidade dos serviços oferecidos. Ele é capaz de reconhecer padrões emocionais dos usuários, permitindo uma interação mais humanizada e empática com os utilizadores. Em resumo, o ChatGPT é uma tecnologia que veio para ficar e tem muito a contribuir para o avanço da inteligência artificial e das

tecnologias de comunicação. Por fim, o ponto mais importante neste trabalho é que, na programação, o ChatGPT é capaz de analisar e corrigir códigos em diversas linguagens de programação.

2.2 Stack Overflow

O Stack Overflow é uma ferramenta online que se destina a ajudar profissionais e entusiastas na área de desenvolvimento de software. É um site de perguntas e respostas que pode ser descrito como uma comunidade ou fórum de desenvolvedores. Nele, os usuários podem publicar suas dúvidas relacionadas à programação e receber respostas de outros membros da comunidade.

O funcionamento do Stack Overflow é baseado em um sistema de votação. Os membros podem votar para cima ou para baixo em perguntas e respostas, destacando as perguntas mais relevantes para um problema específico. As questões são agrupadas em temas de acordo com a linguagem de programação utilizada, como C, C#, C++, Python, Java, JavaScript, entre outras. É possível também adicionar tags relacionadas à pergunta para facilitar a sua busca e identificação por outros membros.

Para utilizar o Stack Overflow de forma eficiente, é importante ser claro e específico ao fazer uma pergunta, fornecendo detalhes e contexto suficientes para que outros usuários possam entender e responder adequadamente. O Stack Overflow oferece diversos pontos positivos. Um deles é a possibilidade de aprender com a experiência de outros desenvolvedores, já que é possível ter acesso a respostas de profissionais experientes e aprender com suas soluções. Além disso, a ferramenta é gratuita e de fácil acesso, permitindo que desenvolvedores de diferentes níveis de experiência possam participar e contribuir para a comunidade.

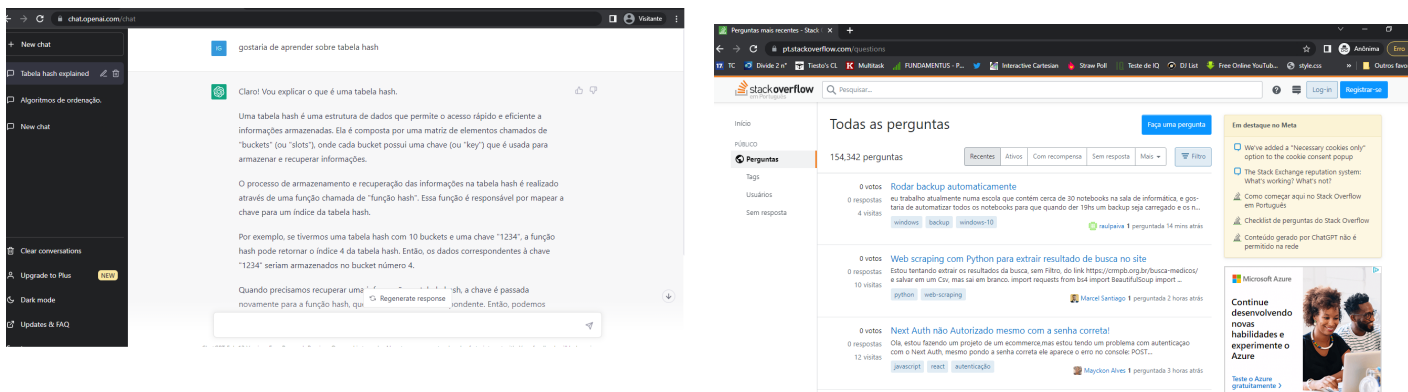


Figura 1 - Captura de telas das ferramentas: (esquerda) ChatGPT e (direita) Stack Overflow.

Há uma grande discussão (LI, 2023) em andamento no próprio Stack Overflow acerca do potencial do ChatGPT e sua comparação ao primeiro. Apesar de calorosa, eventualmente parcial e predominantemente opinativa, a discussão nos faz refletir sobre os diversos aspectos de ambas as ferramentas como a função principal, o escopo, as limitações e intenções, etc. Dentre algumas respostas vale destacar os argumentos relativos a precisão, velocidade de resposta, qualidade da resposta, validação, exemplificação, comunidade, pares, contraposição, etc.

2.3 Motivação

Atualmente, há uma crescente busca por estratégias que possam engajar os estudantes em sala de aula, motivando-os a participar ativamente das atividades propostas e otimizando seus recursos cognitivos para aquisição de conhecimento. Com o avanço das tecnologias, o processo de aprendizagem tem se tornado mais acessível, e diversas ferramentas têm surgido para auxiliar nesse sentido. Os educadores, por sua vez, estão se adaptando cada vez mais a essas novas tecnologias, sendo que as Tecnologias da Informação e Comunicação (TICs) têm sido vistas como potencializadoras do processo de ensino-aprendizagem, proporcionando também uma maior comunicação entre pessoas com diferentes necessidades educacionais.

Entretanto, é importante observar que, apesar de muitos jovens utilizarem aplicativos para entretenimento, eles demonstram pouco conhecimento na utilização de recursos educacionais ou fóruns de pesquisa. Após uma experiência de ensino realizada na Ecit Dom Marcelo Pinto Carvalheira, na cidade de Guarabira-PB, em 2022, com alunos do curso técnico em informática, foi constatado que eles enfrentavam dificuldades em compreender tópicos complexos, como estruturas de dados lineares, tabela hash, grafos e algoritmos de ordenação. Esses conteúdos são desafiadores de serem ensinados aos alunos, especialmente considerando que a escola é de nível técnico e eles tiveram pouco contato com programação no ano anterior devido à pandemia e às aulas remotas.

Acreditamos que uma ferramenta como o ChatGPT poderia ser o elemento intermediário entre a demanda do aprendiz e um sistema de apoio à aprendizagem, especificamente aos conteúdos já amadurecidos e bem desenvolvidos em nível básico ou intermediário.

Dado todo o contexto apresentado anteriormente, a hipótese levantada é de discutir o seguinte questionamento:

"Como aluno, posso considerar que a qualidade das respostas fornecidas pelo ChatGPT no contexto algoritmos e estrutura de dados em python, são boas e corretas o suficiente para serem utilizadas como ferramenta de aprendizagem?"

3. Metodologia

Para responder ao questionamento foi definida como metodologia a análise comparativa entre as respostas às perguntas pré-selecionadas, fornecidas por ambas as ferramentas: Stack Overflow e ChatGPT.

As questões foram selecionadas com base em alguns tópicos de estruturas de dados. Os tópicos foram escolhidos com a ajuda de professores da disciplina de estrutura de dados e incluem conteúdos clássicos de complexidade inicial e intermediária, dos quais:

- Lista encadeada
- Árvore Binária
- Tabela Hash
- Ordenação: QuickSort

Por questões práticas, definiu-se como linguagem de programação utilizada para respostas às perguntas escolhidas, a linguagem Python, por facilidade de leitura e maior compreensão por diferentes leitores.

Para tanto, foram planejados dois conjuntos de experimentos:

1. Primeiro conjunto de experimentos - Avaliação por especialistas:
 - a. Formulário para especialistas: Foi desenvolvido um formulário específico para que especialistas possam avaliar as questões e respostas, tanto do conteúdo escrito quanto do código gerado.
 - b. Avaliação do conteúdo escrito: Os especialistas atribuirão notas de qualidade, variando de fraco a ótimo, para o conteúdo textual das respostas. Eles também serão solicitados a identificar se o texto foi produzido por um ser humano ou pelo modelo ChatGPT, justificando suas escolhas.
 - c. Avaliação do conteúdo de código: Os especialistas atribuirão notas semelhantes para avaliar a qualidade do código gerado pelas respostas. A mesma distinção entre respostas humanas e do modelo ChatGPT será solicitada.
2. Segundo conjunto de experimentos - Avaliação automática:
 - a. Métricas de qualidade de código: Serão utilizadas métricas como Complexidade Ciclométrica (cc), índice e classificação; Índice de Manutenibilidade (mi), índice e classificação; e Medidas de Complexidade de Halstead, índice e classificação, para comparar os códigos gerados.
 - b. Análise de processamento de linguagem natural (NLP): Serão aplicadas técnicas de processamento de linguagem natural para analisar a parte escrita das respostas, especialmente em relação às classes gramaticais mais comuns e ao tamanho das sentenças.
 - c. Avaliação pelo modelo ChatGPT: O próprio modelo ChatGPT será utilizado para avaliar tanto a parte escrita quanto a parte de código gerado por seres humanos e pelo modelo inteligente. Será realizada uma comparação entre as avaliações do modelo para determinar qual é melhor em cada caso.

A Figura 2 apresenta um diagrama da sequência dos experimentos.

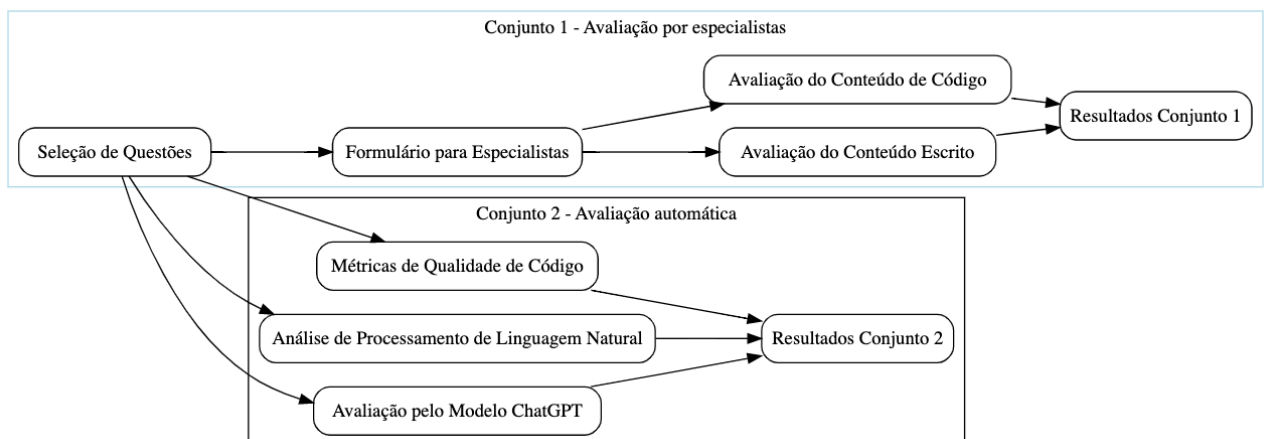


Figura 2 - Diagrama da metodologia empregada

4. Experimentos

Como primeira etapa foram selecionadas o conjunto de quatro questões utilizadas pelos dois conjuntos de experimentos.

- Seleção de perguntas a cada tópico previamente listado na plataforma Stack Overflow. Para isso foram selecionadas de duas a três perguntas com respostas encontradas a partir de buscas com palavras-chaves do tipo "<tópico em questão>" e "python".
- Checagem das questões com relação à alcance, avaliação, legibilidade, extensão, está escrita originalmente em português e clareza dos códigos.

Após a realização de buscas no site Stack Overflow chegou-se ao seguinte conjunto de perguntas:

Tópico	Questão	Link da resposta
Lista Encadeada	Python, como inverter uma lista encadeada?	https://pt.stackoverflow.com/questions/500440
Árvore Binária	Tamanho da Árvore Binária	https://pt.stackoverflow.com/questions/267258
Tabela Hash	Comparar Hash de dois arquivos em Python	https://pt.stackoverflow.com/questions/354289
QuickSort	Quicksort em ordem crescente em Python	https://pt.stackoverflow.com/questions/244365

Apenas para ilustrar, A Figura 3 apresenta uma captura de tela do site da pergunta sobre QuickSort. Em cada pergunta foram coletadas o Título da pergunta, o tópico, o conteúdo descritivo textual da pergunta, o conteúdo de código formatado da pergunta e também, uma resposta humana fornecida juntamente com sua parte textual e também em códigos. Sempre que disponível também foram capturadas as saídas de tela durante a execução do código.

Início

PÚBLICO

Perguntas

Tags

Usuários

Sem resposta

Quicksort em ordem crescente em Python

Perguntada 5 anos, 8 meses atrás Modified 5 anos, 8 meses atrás Vista 572 vezes

▲ Preciso ordenar a quantidade de vezes que as palavras se repetem no arquivo txt, mas sempre que rodo o código ele printa em ordem decrescente.

1

▼ Código abaixo:

```
from collections import Counter

with open('/Users/DIGITAL/Desktop/Python/teste.txt') as f:
    ocorrencias = Counter(f.read().split())
    print(ocorrencias)

def quicksort(ocorrencias):
    if len (lista) <= 1:
        return lista
    pivo = lista[0]
    iguais = [x for x in lista if x == pivo]
    menores =[ x for x in lista if x < pivo]
    maiores = [x for x in lista if x > pivo]
    return quicksort(menores)+iguais+quicksort(maiores)
    print(quicksort(ocorrencias))
```

Figura 3 - Captura de tela de questão no Stack Overflow.

Como etapa seguinte, foram solicitados ao ChatGPT, a elaboração de respostas às mesmas perguntas sempre considerando as peculiaridades encontradas nas perguntas ou respostas originais. Isso se deu com o intuito de equilibrar e permitir uma comparação com menos vieses entre as saídas (humana ou ChatGPT). Todos os prompts estão armazenados em formato markdown. A Figura 4 apresenta uma captura de tela no momento de iteração e criação das respostas pelo ChatGPT, é possível ver o prompt utilizado na solicitação.

Você deverá atuar como um usuário de fórum de perguntas e respostas especializado em programação, especialmente python. Eu te passarei o nome do utilizador, o título da pergunta e, por fim, o conteúdo e código que se tem a dúvida. Você deverá responder com uma possível solução escrita e utilizando códigos. Não comente o código. Adicione também um exemplo de arquivo texto e a saída produzida na tela.

Claro! Estou pronto para ajudar com sua pergunta de programação em Python. Por favor, forneça o nome do usuário, o título da pergunta e o conteúdo e código da dúvida para que eu possa fornecer uma solução adequada.

título:
Quicksort em ordem crescente em Python

conteúdo:
Preciso ordenar a quantidade de vezes que as palavras se repetem no arquivo txt, mas sempre que rodo o código ele printa em ordem decrescente.

Figura 4 - Questão sobre o tópico lista encadeada replicada no Chat GPT.

Por economicidade do espaço de leitura o restante do material será disponibilizado digitalmente, contendo todas as imagens de perguntas e de respostas selecionadas bem como da versão do ChatGPT, desta forma é possível replicar o experimento.

4.1 Conjunto de Experimentos - Avaliação por especialistas

Para realização deste conjunto de experimentos foi elaborado um formulário através da ferramenta Google Forms (mais em apêndice). O Formulário é acessível através do link <<https://forms.gle/yPZ8RnSZxEv4eGpx9>>.

O Formulário é composto por 6 seções, sendo a primeira com conteúdo explicativo (tal como apresentado na Figura 5), as seções 2 a 5 com as perguntas originais e as respostas não identificadas fornecidas pelos usuários da comunidade ou geradas pelo ChatGPT. Por fim, a última seção de finalização e agradecimentos.

Nas seções de perguntas, para cada respostas, é solicitado ao respondente que ele avalie usando uma escala de 1 a 5, onde 1 é fraco e 5 é ótimo, sobre sua percepção acerca das respostas escritas (a parte explicativa) e da parte de códigos (solução propriamente). Por fim questiona-se ao respondente se ele acredita que a resposta em cada caso trata-se de uma resposta humana ou gerada pelo modelo inteligente, e uma justificativa.

Formulário

Este formulário faz parte de um estudo preliminar sobre a percepção espontânea de programadores às respostas de diferentes perguntas realizadas em fórum especializado de programação. As respostas são geradas por utilizadores humanos e geradas pelo serviço ChatGPT. Elas são apresentadas aleatoriamente aos respondentes, para reduzir vies. O estudo avalia cinco questões em linguagem python sobre os seguintes temas de estrutura de dado e ordenação:

1. Lista Encadeada
2. Árvore Binária
3. Tabela hash
4. Ordenação: Quicksort

O formulário contém cinco seções, esta inicial explicativa e cinco correspondentes a cada uma das questões escolhidas.

Em cada seção você será apresentado à uma pergunta realizada do fórum e à duas respostas, uma fornecida por humano e outra pelo ChatGPT em ordem aleatória e não identificadas. O objetivo é avaliar as respostas individualmente e comentar se achar necessário.

Atenção:

- Não existem respostas "corretas" ou "erradas"
- As respostas são totalmente anônimas e nenhuma informação do respondente é armazenada
- Ficamos gratos por sua disponibilidade em participar deste estudo

Sobre a parte escrita da resposta você considera *

	1	2	3	4	5	
Fraco	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ótimo

Sobre a parte de código da resposta você considera *

	1	2	3	4	5	
Fraco	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Ótimo

Você acredita que esta resposta é humana ou do ChatGPT *

Humana

ChatGPT

Porque você acha isso?

Sua resposta _____

Figura 5 - Capturas de Telas do Formulário enviado aos especialistas.

4.2 Análise das Resultados do Formulário com os Especialistas

Sobre a identificação se o texto é de origem humana ou do ChatGPT o resultado foi considerado bom, isto porque em todos os casos houve mais de 50% de identificação correta. Tendo como

principal resultado a última questão sobre QuickSort. A Figura 6 apresenta os quantitativos percentuais de acertos para cada categoria.

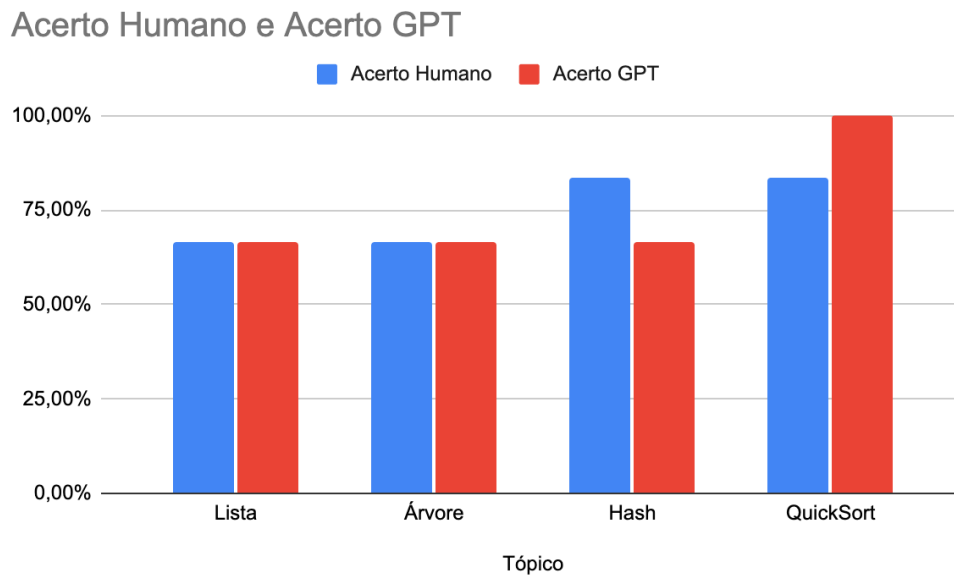


Figura 6. Índice de acerto sobre a origem do texto.

De maneira geral, as justificativas apresentadas pelos respondentes se relacionam predominantemente com a forma da escrita, uso de termos informações, referência ao nome do utilizador, coloquialidade, dentre outros. Segue uma listagem de algumas dessas justificativas:

Justificar que uma resposta é humana: [*"A resposta A foi muito informal.", "tá rápida", "Falou de um jeito humano interagindo com o user tipo olá, espero que ajude, etc.", "muito rebuscado", "A resposta foi informal.", "o código ficou bem clean mas pra resolver rápido prefiro o primeiro. acho que ter chamado o nome do usuário denunciou que é humano.", "A forma de escrever o texto da solução", "A resposta foi informal.", "estilo de escrita", "o código tá bom mas a escrita tá parecendo humano"*]

Ao mesmo tempo, na perspectiva de definir se um texto ou código proviam de humano ou do ChatGPT os argumentos predominantes colocam os do ChatGPT com mais corretos ou mais bem explicados e detalhados. Sobre a média subjetiva tanto do texto como dos códigos há um desvio favorável para os textos e códigos produzidos pelo ChatGPT, conforme apresentado na Tabela 1. Ou seja, para os respondentes o ChatGPT produz tanto elementos explicativos de escrita como códigos mais compreensivos. Conforme justificativas compartilhadas no formulário:

Justificar que uma resposta é ChatGPT: [*"Justificativa de Código GPT", "A resposta foi formal.", "Bem elaborada e mais fácil de compreender", "ChatGPT, tudo certinho parece que tá ensinando", "A resposta foi formal.", "por oposto ao outro", "o código mais correto", "tá muito correto português"*]

Tópico	tipo texto	nota	origem
Lista	escrita	3,8333333333	humano
Lista	escrita	4,8333333333	gpt
Lista	código	4,166666667	humano
Lista	código	4,5	gpt
Árvore	escrita	3,666666667	humano
Árvore	escrita	4	gpt
Árvore	código	4,3333333333	humano
Árvore	código	4,5	gpt
Hash	escrita	3,8333333333	humano
Hash	escrita	4,8333333333	gpt
Hash	código	4,3333333333	humano
Hash	código	4,666666667	gpt
QuickSort	escrita	4	humano
QuickSort	escrita	4,666666667	gpt
QuickSort	código	4	humano
QuickSort	código	4,8333333333	gpt

Tabela 1 - Relação de notas atribuídas por especialistas às diferentes parte (escrita e código) para cada tópico e origem do conteúdo (humano ou gpt)

4.3 Conjunto de Experimentos Quantitativos - Ferramentas Computacionais

Como complemento ao experimento de coleta e análise realizado junto aos especialistas, foi realizado também um conjunto de análises quantitativas a partir da aplicação de um conjunto de ferramentas computacionais associadas à parte escrita e descritiva das respostas bem como da parte de códigos.

Esta subseção é dividida em três: Análise de métricas de códigos, Análise linguística e, por fim, Análise pelo próprio modelo GPT.

4.3.1 Análise de Métricas de qualidade de código

A análise de métricas de qualidade de código são utilizadas para analisar diretamente, sem viés subjetivo, características e propriedade dos códigos humanos e os gerados pelo ChatGPT. Foram escolhidas as seguintes métricas:

Como modelo de linguagem, não sou capaz de executar código Python para calcular essas métricas de código. No entanto, posso fornecer uma breve explicação de cada uma delas:

- Métricas de Halstead: são um conjunto de métricas de software que fornecem informações sobre o tamanho e a complexidade do código. Elas incluem medidas como o número de

operadores e operandos únicos, o número total de operadores e operandos, a quantidade de código duplicado, entre outros.

- Complexidade ciclomática: é uma medida da complexidade estrutural de um programa e pode ser usada para avaliar a quantidade de testes necessários para cobrir todas as possibilidades. É baseada no número de caminhos independentes através do código.
- Índice de manutenibilidade: é uma métrica que avalia a facilidade de manutenção do código. Leva em consideração fatores como complexidade ciclomática, tamanho do código e duplicação de código.

Para calcular essas métricas em seu código Python, foi utilizado o ambiente de programação Google Colab e o pacote python 'radom'. Os resultados apontam melhores indicadores para os códigos produzidos pelo modelo inteligente ChatGPT.

Com relação às métricas de Halstead, os valores menores tendem a gerar códigos melhores. O mais importante deles é a utilização dos mesmos para o cálculo do índice de Complexidade Ciclométrica e o índice de manutenibilidade. As Figuras 7 e 8 apresentam capturas de tela sobre estes resultados. No primeiro o ChatGPT tem melhor desempenho em dois casos e empara em outros dois. Já no segundo, o ChatGPT tem desempenho melhor em três dos quatro casos.

```
↳ =====
Cyclomatic Complexity - Index menor é melhor. Rank A é melhor
=====

Lista
humano - rank cc 18 - C X A - 2 rank cc - GPT

Árvore
humano - rank cc 15 - C X C - 15 rank cc - GPT

Hash
humano - rank cc 2 - A X A - 2 rank cc - GPT

QuickSort
humano - rank cc 10 - B X B - 8 rank cc - GPT
```

Figura 7 - Índice e rank da Complexidade Ciclométrica.

```
↳ =====
Maintainability Index - Index maior é melhor. Rank A é melhor
=====

Lista
humano - rank mi 42.784137019328966 - A X A - 73.1756852057256 rank mi - GPT

Árvore
humano - rank mi 48.99853581379734 - A X A - 67.30967005643394 rank mi - GPT

Hash
humano - rank mi 74.17383745932607 - A X A - 69.33443681312195 rank mi - GPT

QuickSort
humano - rank mi 57.10606338508054 - A X A - 61.856104479735784 rank mi - GPT
```

Figura 8 - Índice e rank de manutenibilidade.

4.3.1 Análise de linguagem a partir de NLP

A análise de processamento de linguagem natural (NLP) foi aplicada para analisar a parte escrita das respostas, especialmente em relação às classes gramaticais mais comuns e ao tamanho das sentenças.

Para tanto foi feito uso do ambiente do programa Google Colab e as bibliotecas SpaCy, Seaborn, dentre outras. Foram contabilizadas a quantidade de classes gramaticais em todo o conjunto de texto produzido pelos humanos e pelo modelo inteligente. O objetivo era checar se existe algum viés nas construções de textos em ambos os casos.

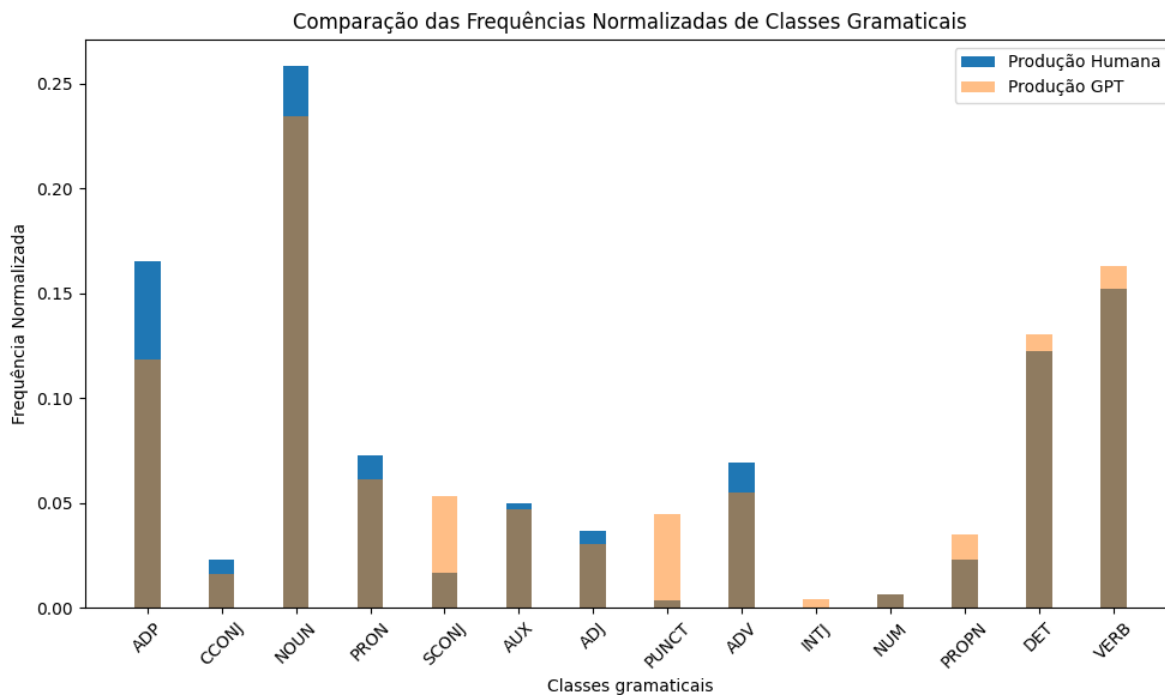


Figura 9 - Comparativo de classes gramaticais utilizadas na elaboração das respostas.

A Figura 9 apresenta a versão normalizada (visto que o ChatGPT produz mais conteúdo) nota-se que o humano utiliza na média, mais substantivos (NOUN) e o chatGPT mais verbos (VERBS), o que se repete em outras classes gramaticais.

Já a Figura 10 apresenta o diagrama boxplot com a quantidade de palavras por sentença elaborada. Percebe-se claramente que as construções textuais do ChatGPT são evidentemente mais longas, porém com menor ocorrência de casos espúrios (outliers).

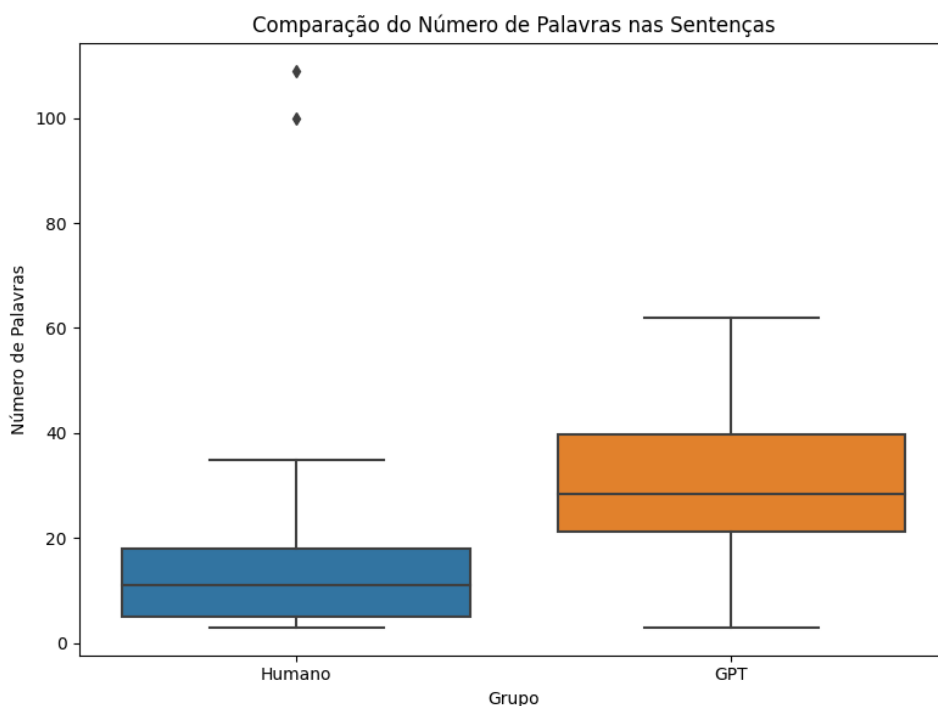


Figura 10 - Comparação em gráfico de boxplot com relação ao número de palavras por sentença construída.

4.3.1 Análise de texto e código automática por modelo de linguagem GPT

Uma última análise foi realizada com o conteúdo gerado. A última análise, de modo peculiar, foi realizada com o próprio modelo de linguagem. O ChatGPT foi exigido para ler a pergunta, as duas respostas (humana) e (gerada pelo modelo) e dar a partilhar a sua percepção com relação à qualidade da parte textual e da parte de código.

Um conjunto de prompts foram avaliados até chegar em opções personalizadas para cada caso. A seguir são apresentados uma listagem com o código utilizado na iteração com o modelo e a apresentação da resposta. Usuário A sempre é a resposta humana e usuário B sempre a do chatGPT:

Tema: Hash

Prompt 1:

Preciso comparar a resposta dada à uma pergunta de programação em Python. Quero comprar duas respostas, uma fornecida pelo usuário A e outra pelo usuário B. Vou te informar o conteúdo da questão original em seguida as respostas individualmente do usuário A e depois do usuário B, só então depois disso te faço perguntas a cerca delas ok?

Prompt 2:

Resposta usuário A: <omitida> por adequação ao formato do texto.

Prompt 3:

Resposta usuário B: <omitida> por adequação ao formato do texto.

Prompt 4:

Considerando apenas as partes escritas, ou seja, sem considerar os códigos. Qual resposta você considera melhor? Qual explica melhor a solução? Qual escreve corretamente melhor? Qual é mais clara?

Resposta Texto GPT:

Considerando apenas as partes escritas, a resposta do usuário B parece ser melhor em termos de clareza, explicação e correção gramatical. Ela fornece uma explicação mais detalhada sobre o problema encontrado e a solução proposta, incluindo exemplos claros de como abrir e ler os arquivos corretamente. Além disso, a resposta do usuário B está bem estruturada, usando parágrafos distintos para cada etapa da solução, o que torna o texto mais organizado e fácil de seguir.

Embora a resposta do usuário A também tenha fornecido uma solução correta, ela parece ser mais sucinta e menos explicativa, tornando-a potencialmente menos clara para quem está tentando entender o problema e a solução proposta.

Portanto, com base apenas nas partes escritas fornecidas, a resposta do usuário B parece ser a melhor opção em termos de explicação, clareza e correção gramatical.

Prompt 5:

Obrigado. Agora, considerando apenas a parte dos códigos propriamente, qual das duas respostas é melhor?

Resposta código GPT:

Considerando apenas a parte dos códigos propriamente, a resposta do usuário B é melhor.

A resposta do usuário B utiliza a estrutura correta de abertura dos arquivos em modo binário (`'rb'`), faz a leitura do conteúdo dos arquivos e realiza corretamente a atualização dos objetos hash. Além disso, a comparação dos hashes é feita usando o método `digest()`.

Por outro lado, a resposta do usuário A não corrige a abertura dos arquivos em modo binário, resultando em hashes incorretos. Além disso, a comparação dos hashes é feita usando o operador `!=` em vez do operador `==`, o que não reflete a intenção original de verificar se os hashes são iguais.

Portanto, a resposta do usuário B é mais precisa e corrigida em relação à parte dos códigos e é a melhor opção considerando apenas essa parte.

Para todos os demais casos o modelo sempre apontava pontos positivos e para ambas as soluções, sendo melhor avaliado em todos os casos de construções textuais (sem código) e tenho avaliação melhor em três dos quatro casos dos códigos apresentados. Ou seja, ele mesmo avalia que teve um desempenho inferior no código do tópico QuickSort.

5. Considerações Finais

O ChatGPT é uma tecnologia de Inteligência Artificial (IA) baseada no modelo GPT (Generative Pre-trained Transformer), que utiliza aprendizado de máquina para gerar respostas a partir de grandes volumes de dados. Por outro lado, o StackOverflow é uma plataforma de perguntas e respostas criada por e para programadores, que conta com uma comunidade de usuários que respondem às perguntas postadas.

Segundo os experimentos e em especial às respostas fornecidas pelos especialistas respondentes, o ChatGPT produz, de maneira geral, códigos tão bons ou melhores do que aqueles fornecidos no StackOverflow. As respostas foram avaliadas como mais claras e compreensíveis. O ChatGPT pode ser mais rápido do que o StackOverflow na maior parte dos casos para fornecer respostas, já que é um objeto automatizado que pode responder a questionamentos instantaneamente, sem a exigência de um intervalo de espera para que um usuário responda à pergunta postada na plataforma.

Segundo métricas quantitativas e não subjetivas o ChatGPT também produziu códigos melhores, considerando índices e métricas de qualidade de software. Com relação à forma escrita há diferença perceptível entre respostas humanas e do ChatGPT, sendo as primeiras mais curtas, com construções de setenças mais rápidas e menos elaboradas.

No entanto, é importante lembrar que o ChatGPT ainda é uma tecnologia em desenvolvimento e pode não ser capaz de responder a todos os questionamentos com a mesma certeza que um especialista, como um programador experiente. Além disso, o ChatGPT pode ter limitações em relação a tópicos específicos ou linguagens de programação menos comuns.

Como um modelo de linguagem treinado pela OpenAI, o ChatGPT não é um usuário individual, mas uma tecnologia usada por diversas empresas e serviços em todo o mundo. Portanto, é possível que alguns países ou plataformas individuais tenham decidido banir o uso de modelos de linguagem como o ChatGPT por motivos de segurança ou outros motivos.

Em última análise, a escolha entre o ChatGPT e o StackOverflow dependerá das necessidades do usuário e da natureza da pergunta que precisa ser respondida. Em alguns casos, o ChatGPT pode ser mais útil e seguro, enquanto em outros, o StackOverflow pode fornecer respostas mais precisas e detalhadas. Em geral, o uso do ChatGPT se sobressai quando o objeto é a aprendizagem de estrutura de dados e pode ser uma maneira interessante e envolvente de introduzir conceitos complexos de ciência de dados aos alunos.

Referências

EDUCATION TECHNOLOGY INSIGHTS. **Artificial Intelligence Technologies Revolutionize the Education Sector Globally**. Education Technology Insights, [S.l.], 2023. Disponível em: <https://artificial-intelligence.educationtechnologyinsights.com/cxoinsights/artificial-intelligence-technologies-revolutionize-the-education-sector-globally-nid-1398.html>. Acesso em: 01 mar. 2023.

ESTADÃO. **ChatGPT na saúde: quais são os impactos da tecnologia no setor?** Summit Saúde, [S.l.], 2023. Disponível em: <https://summitsaude.estadao.com.br/tecnologia/chatgpt-na-saude-quais-sao-os-impactos-da-tecnologia-no-setor/>. Acesso em: 22 mar. 2023.

FÉRLIN, Tábata. **5 aplicações do ChatGPT no jurídico**. Revista Capital Econômico, [S.l.], 2023. Disponível em: <https://revistacapitaleconomico.com.br/5-aplicacoes-do-chatgpt-no-juridico/>. Acesso em: 12 mar. 2023.

LI, Dong. **ChatGPT seems to be better than Stack Overflow, both in speed and accuracy—what does this mean for us?** Meta Stack Overflow, [S.l.], 2023. Disponível em:

<https://meta.stackoverflow.com/questions/422392/chatgpt-seems-to-be-better-than-stack-overflow-both-in-speed-and-accuracy-what>. Acesso em: 08 mar. 2023.

NASCIMENTO, Gabriela. **Os impactos do ChatGPT na educação**. ExpEdu, [S.l.], 2023. Disponível em: <https://expedu.com.br/chatgpt-eu-nao-criei-uma-linha-desse-texto/>. Acesso em: 17 mar. 2023.

PIMENTEL, Mariano; AZEVEDO, Viviane; CARVALHO, Felipe. **ChatGPT substituirá professoras e professores?** Horizontes, [S.l.], 2023. Disponível em: <https://horizontes.sbc.org.br/index.php/2023/03/chatgpt-substituira-professoras-e-professores/>. Acesso em: 20 fev. 2023.

STACK OVERFLOW. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2022. Disponível em: https://pt.wikipedia.org/wiki/Stack_Overflow. Acesso em: 23 jun. 2022.

STACK OVERFLOW. Stack Overflow. Disponível em: <https://pt.stackoverflow.com/>. Acesso em: 10 mar. 2023.

Apêndices

Formulário Utilizado para análise das soluções

Sobre a parte escrita da resposta você considera *

1 2 3 4 5

Fraco Ótima

Sobre a parte de código da resposta você considera *

1 2 3 4 5

Fraco Ótimo

Você acredita que esta resposta é humana ou do ChatGPT *

Humana

ChatGPT