

# Você sabe diferenciar um resumo escrito por humanos do gerado pelo ChatGPT?

Francisco das C. F. Almeida<sup>1</sup>, Yuska P. C. Aguiar<sup>2</sup>, Rafael M. Magalhaes<sup>1</sup>

<sup>1</sup>Universidade Federal da Paraíba (UFPB- Campus IV)  
Rio Tinto – PB – Brazil

<sup>2</sup>Universidade Federal da Paraíba (UFPB- Campus I)  
João Pessoa – PB – Brazil

{francisco.chagas,rafael}@dcx.ufpb.br, yuska@ci.ufpb.br

**Resumo.** *O ChatGPT gera textos artificialmente levando a discussões sobre produtividade, autoria e ética. Este estudo explora a percepção de acadêmicos sobre a origem, artificial ou humana, de resumos de Trabalhos de Conclusão de Curso (TCC) em Engenharia de Software. 23 TCCs foram analisados para extração de informações úteis na geração de resumos via ChatGPT. Os resumos foram disponibilizados, em agrupamentos aleatórios de 6, para classificação. A análise realizada por 25 avaliadores sugere que a maior facilidade em identificar corretamente resumos de origem humana, enquanto que os de origem artificial exigem um olhar técnico para classificação correta.*

## 1. Introdução

O *ChatGPT* (*Generative Pre-Trained Transformer*) foi construído pela OpenAI e disponibilizado no final de 2022 [1, 4] enquanto modelo de processamento de linguagem natural que utiliza técnicas de *Deep Learning* para processar grandes volumes de dados textuais [2]. A partir de uma plataforma web, o *ChatGPT* permite que humanos estabeleçam conversas a partir de um *prompt* de perguntas e demandas, cujas respostas são geradas artificialmente. Como seu treinamento acontece de forma não supervisionada e a partir de diversas fontes (livros, artigos, sites da web, etc.), reflexões sobre plágio, direitos autorais e produtividade têm surgido [2, 5].

Embora seu propósito não seja educacional, o *ChatGPT* tem sido utilizado como recurso auxiliar em atividades acadêmicas [2, 5].

Com base no exposto, este estudo se configura como uma análise de caso, no qual professores universitários da área de Tecnologia da Informação são convidados a classificar a origem de um conjunto de resumos de artigos, que podem ter sido produzidos por humanos ou gerados pelo ChatGPT. Além disso, o estudo tem como objetivo identificar as limitações da ferramenta, assim como a percepção das pessoas participantes sobre o potencial de uso do ChatGPT (aliado ou desafio?).

Logo, o escopo dessa pesquisa foram 23 Trabalhos de Conclusão de Curso (TCCs) da área de Engenharia de Software no formato de artigo defendidos entre os anos de 2015-2022, no curso de Sistemas de Informação da UFPB. Deste *corpus* foram extraídos os resumos de origem humana, assim como informações relevantes para geração automática de seus respectivos pelo *ChatGPT* a partir de *prompts* específicos. Os 46 resumos foram disponibilizados em um sistema *online*, a partir do qual pessoas

com perfil acadêmico e da área de Computação, foram convidadas a classificar a origem de um subconjunto de 6 resumos aleatórios. Ainda, foi disponibilizada a opção de “*não sei identificar*” e um campo aberto para comentar a classificação dada. A análise comparativa de “acertos” e “erros” das 75 classificações recebidas foi apoiada pela observação da estrutura gramatical dos resumos.

De forma complementar, foi realizada uma análise morfológica e gramatical dos resumos (frequência de ocorrência) com o objetivo de identificar a existência de padrões que pudessem dar indícios de uma origem automática com geração via *ChatGPT*. A estrutura de cada resumo também foi observada para identificação de padrão de escrita ou de construção de frases.

## 2. O que é o *ChatGPT* ?

O *ChatGPT* adota uma arquitetura de modelo de linguagem natural não supervisionada [8] para treinamento de inteligência artificial a partir de com um conjunto massivo de dados não rotulados, sendo necessário, geralmente, milhões de dados para que ele possa ser ajustado e apto para realizar uma tarefa específica [1, 7, 8]. O *ChatGPT* é treinado para conversar com humanos a partir de um *prompt* e fornecer respostas adequadas de acordo com as informações obtidas por meio do seu treinamento. O *ChatGPT* é capaz de responder as perguntas, reconhecer possíveis erros e recusar solicitações inadequadas, mantendo um formato de diálogo [1]. A demanda enviada pelo *prompt* é utilizada como base para buscar informações pertinentes a fim de gerar respostas coerentes. A versão gratuita do *ChatGPT* utiliza a tecnologia GPT-3.5, que é inferior a versão da versão paga, GPT-4, *ChatGPT Plus*. Em termos gerais, divergem no poder de gerar informações mais precisas [7, 8], onde o GPT-4 apresenta uma redução de 82% na probabilidade de responder a solicitações de conteúdo proibido, além de uma melhoria de 40% na produção de respostas factuais em comparação ao GPT-3.5 [9]. Ainda, a versão do GPT-4 possui capacidades aprimoradas de raciocínio, fornece respostas mais criativas e processa instruções mais complexas [3].

## 3. Metodologia

Para condução da pesquisa foram executadas as seguintes etapas: (i) definição do escopo com a seleção dos Trabalhos de Conclusão de Curso (TCCs) a serem adotados como *corpus*; (ii) análise do *corpus* para extração das informações relevantes e definição do *prompt* específico de geração automática dos resumos para cada artigo; (iii) execução dos *prompts* no *ChatGPT Plus 4.0*, explorando 3 propostas de resumos de origem automática, sendo escolhido um de forma aleatória; (iv) modelagem e desenvolvimento de uma plataforma *online* para classificação dos resumos por parte das pessoas participantes no estudo, (v) difusão da pesquisa e coleta de dados; (vi) análise de dados das respostas obtidas, assim como análise da estrutura gramatical, tamanho das sentenças e observações de padrões nas estruturas dos resumos.

Nos cursos de Bacharelado em Sistemas de Informação (BSI) e Licenciatura em Ciência da Computação (LCC) da Universidade Federal da Paraíba (UFPB), desde 2010.2, foram defendidos e aprovados mais de 400 TCCs. Os TCCs podem assumir os formatos monografia, artigo científico ou artigo tecnológico e são disponibilizados em um repositório *online* da instituição. Para viabilizar a execução do estudo, seu escopo foi definido para artigos científicos, do curso de BSI e da área de Engenharia de Software, resultando em um *corpus* de 23 trabalhos. Estes foram analisados manualmente e informações de caracterização foram extraídas e documentadas (título, palavras-chave,

orientador, autor, problemática, metodologia, resultados relevantes e resumo do autor). A partir destas e seguindo a estrutura de resumos IMRaD (Introdução, Métodos, Resultados e Discussão) [6] foram criados *prompts* específicos para cada artigo de forma que o resumo correspondente fosse gerado artificialmente pelo *ChatGPT Plus 4.0*. Para garantir a equivalência entre os resumos (original e artificial) adotou-se uma margem de + ou - 20% para a quantidade de palavras com base no resumo original. Cada resumo artificial foi gerado 3 vezes e apenas um deles escolhido, aleatoriamente, para compor o *corpus* final.

À princípio, cogitou-se a adoção do *Google Forms* para a coleta de dados. No entanto, os recursos disponíveis não são suficientes para atender às necessidades específicas do estudo. Ou seja, não é possível configurar um banco de resumos a partir do qual um número específico de resumos de origens distintas (humana e artificial) sejam selecionados, de forma aleatória, para ser apresentado aos participantes da pesquisa. Dessa forma, se fez necessário desenvolver uma aplicação *web* adequada ao contexto e especificidade da coleta de dados. Para tanto, adotou-se a tecnologia *React* para o *front-end*, *PHP* para o *back-end* e *MySQL* para o banco de dados. A aplicação foi organizada em três telas, das quais a primeira era apenas informativa contendo breve explicação sobre a pesquisa e dados das pessoas pesquisadoras, indicação sobre participação voluntária e garantia de anonimato. Uma segunda tela para coleta de dados sobre o participante, contemplando: instituição de ensino, curso e área de conhecimento nos quais atua, nível de formação e tempo de atuação na área. E, a tela para que a pessoa participante pudesse classificar a origem de 6 resumos, escolhidos aleatoriamente de um total de 56 resumos, sendo 3 de origem humana e 3 gerados pelo *ChatGPT*. As opções de classificação são: “Eu acho que o resumo foi gerado por uma pessoa”, “Eu acho que o resumo foi gerado pelo ChatGPT” e “Eu não sei avaliar a origem do resumo”. Ainda, foi disponibilizado um campo aberto e opcional para justificar a classificação dada (Figura 1).

1) Este estudo investigou os vieses cognitivos na tomada de decisão dos profissionais da indústria de software no contexto da arquitetura de software. A pesquisa, por meio de entrevistas semiestruturadas com engenheiros de software, identificou a presença de tais vieses na rotina diária desses profissionais. Os resultados demonstraram que os vieses cognitivos são universais e impactam diretamente o processo decisório relativo à arquitetura de software. Notou-se ainda que cada profissional, mesmo exposto aos mesmos vieses, apresentou percepções e exemplos distintos, reforçando a complexidade e a subjetividade do fenômeno.

- Eu acho que o resumo foi gerado por uma pessoa
- Eu acho que o resumo foi gerado pelo ChatGPT
- Eu não sei identificar a origem do resumo

Por que você chegou a essa conclusão?

**Figura 1: Tela de classificação da origem dos resumos**

A pesquisa foi divulgada em listas de email direcionadas para docentes da área de Computação. A participação foi voluntária e anonimizada. A análise das respostas foi direcionada para a análise de acertos, erros e dúvidas da classificação, considerando os comentários fornecidos.

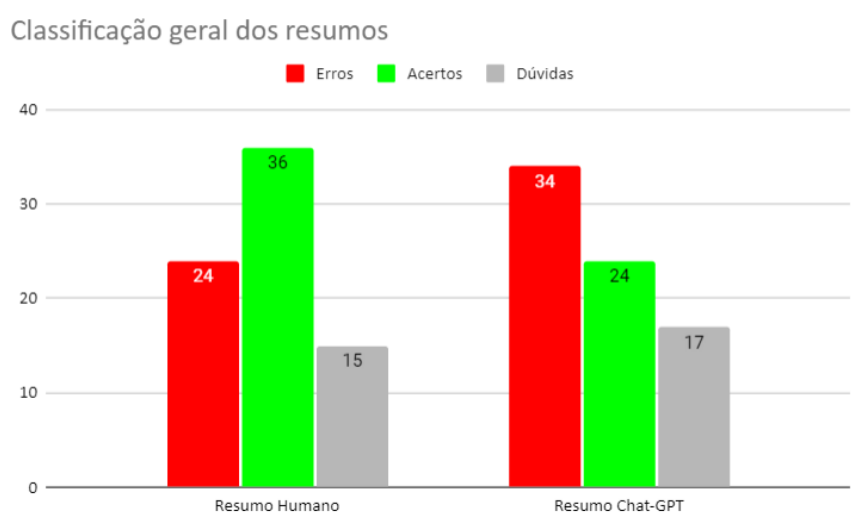
Ainda, fez-se uso do *Google Colab* para análise da estrutura gramatical, do tamanho das sentenças e dos padrões dos resumos, humanos e artificiais. Dentre os elementos gramaticais foram contabilizados artigo definido (DET); adjetivo (ADJ); advérbio (ADV); preposições (ADP); verbo (VERB); verbo auxiliar (AUX); pronomes (PRON); conjunção coordenativa (CCONJ); numeral (NUM); conjunção subordinativa (SCOMJ); substantivo próprio (PROPN); substantivos (NOUN) e outros (X).

#### 4. Análise de dados

A coleta de dados contemplou a participação de 25 respondentes, acadêmicos de diferentes áreas da computação e instituições de ensino, com níveis de doutorado e mestrado, tendo experiência profissional de no mínimo 5 anos. Ao todo foram classificados 150 resumos, sendo metade de origem artificial e a outra metade humana.

A classificação dos resumos gerados pelo *ChatGPT Plus 4.0* (Gráfico 1) sugere uma tendência entre as pessoas participantes ao classificarem os resumos de origem artificial como sendo de origem humana (“erro”). Ou seja, os resumos originados pelo *ChatGPT* foram receberem 34 (45,3%) classificações como sendo de origem humana, 24 classificações corretas (32%) e 17 resumos não foram classificados (22,6%) pois as pessoas se mantiveram em dúvida.

As justificativas dadas pelos participantes quando classificaram os resumos como sendo de origem humana (“erros”) destacaram (i) a presença de palavras incomuns ou menos frequentes no texto, que poderia ser um indício forte que o resumo seria de um humano; (ii) a fluidez do texto, (iii) a presença de citações e (iv) a presença “*opiniões baseadas nos resultados*”. Para as classificações corretas, as características mencionadas pelas pessoas respondentes nas justificativas, incluem: (v) o uso de palavras comuns no texto, vocabulário mais genérico e menos específico; e a identificação de padrões de escrita.



**Gráfico 1 - Erros, Acertos e Dúvidas na classificação dos resumos de origem Humana e Artificial (ChatGPT)**

Para a classificação dos resumos humanos (Gráfico 1) percebe-se a tendência entre os participantes em categorizar os resumos corretamente. 36 (48%) das classificações foram assertivas, enquanto que 24 (32%) foram equivocadas (como de origem artificial), havendo 15 sem classificação por dúvida na origem (20%).

Dentre as justificativas apresentadas para a classificação correta, de origem humana, tem-se a estrutura do resumo, erros de ortografia, acentuação e o uso de crase, incomuns para textos gerados pelo *ChatGPT*. Também foi mencionado o estilo de escrita como sendo um possível diferencial, dado que o *ChatGPT* tende à escrita mais padronizada. Para as classificações de origem artificial quando humana (erradas), percebe-se que a dificuldade em realizar a classificação correta decorre muito de algo que está ausente em resumos gerados pelo *ChatGPT* que é uma peculiaridade humana em cometer erros. Dado um resumo sem erros ortográficos e gramaticais, as pessoas participantes buscaram fazer uma análise mais criteriosa em relação a padrões de escrita que indiquem geração artificial, como evidenciado a seguir: "Alguns padrões se repetem nos textos gerados pelo *ChatGPT*: 'Este estudo tem como objetivo...'; 'A metodologia empregada envolveu...'; 'A análise resultou na identificação...". Esses resultados indicam que o *ChatGPT* obteve sucesso ao simular um resumo humano, exigindo uma análise mais crítica para uma classificação precisa.

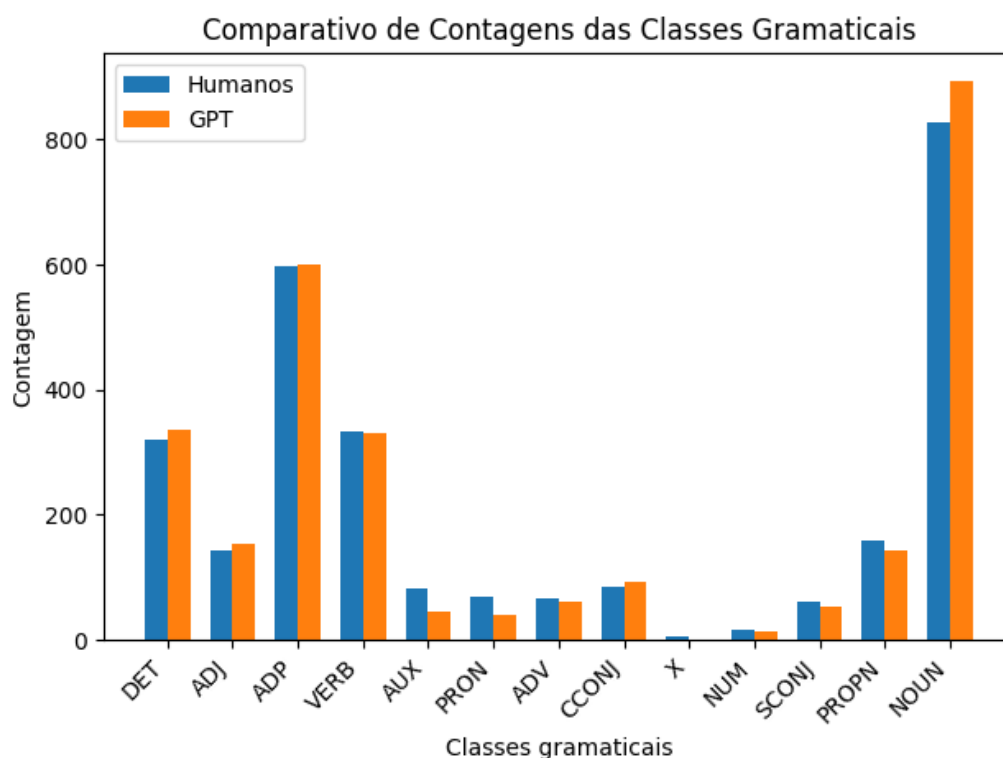
A análise da estrutura gramatical dos resumos foi motivada por justificativas apresentadas pelos participantes para classificações “de origem artificial” para os textos sem problemas de escrita (erros gramaticais ou ortográficos), mesmo quando estes eram de origem humana. Com o intuito de abordar essa questão, foi feito um levantamento no *Google Colab*, para quantificar o número de ocorrência de todas as classes gramaticais de cada resumo.

De acordo com o Gráfico 2, observa-se similaridades na construção de sentenças e na quantidade de palavras por tipo gramatical para os resumos de forma indistinta à sua origem. São mais frequentes, em uma proporção H:A (Humano e Artificial), Substantivos (NOUN 827: 893), preposições (ADP 597:601), Verbos (VERB 334: 332) e Artigos Definidos (DET 319: 336). Embora a distribuição seja proporcional, a quantidade de substantivos, adjetivos e artigos definidos é maior para os resumos gerados pelo *ChatGPT Plus 4.0*. Embora presentes em quantidade menor nos textos verbos auxiliares (AUX 83: 46), pronomes (PRON: 68:39), advérbios (ADV 67:62), numerais (NUM 15:13), conjunção subordinativa (SCONJ 60:53) e substantivo próprio (PROPN: 158: 143), são mais numerosos nos textos de origem humana. Aspectos sutis que podem interferir na percepção do leitor e classificação da origem do resumos.

Ainda que não se tenha observado diferenças significativas nos tamanhos das sentenças geradas por humanos e artificialmente, destaca-se a necessidade de examinar mais profundamente outros aspectos da estrutura dos resumos ao análise do tamanho médio dos resumos gerados por humanos e pelo *ChatGPT* revela uma interessante semelhança. Embora não tenham sido observadas diferenças significativas nos tamanhos das sentenças entre as duas origens, essa semelhança aparenta que a capacidade do *ChatGPT* em produzir resumos que se assemelha a resumos gerados por humanos.

Esse resultado é um possível indicador de que o *ChatGPT* é capaz de ajustar sua produção de resumos para ser semelhante aos humanos. No entanto, uma análise mais profunda pode ser necessária para entender se essa convergência no tamanho médio se estende a outros aspectos estruturais ou se há nuances específicas que diferenciam as composições geradas por humanos das geradas pelo *ChatGPT*.

Na análise da estrutura dos resumos, identificou-se um padrão de escrita nos textos gerados pelo *ChatGPT Plus 4.0*. “*Este estudo...*” foi uma construção presente como frase inicial para 14 dos 23 resumos gerados automaticamente. Outros casos menos frequentes foram: “*Este artigo*” para 4, “*Este trabalho*” para 3 e “*O objetivo*” e “*Esta pesquisa*”, 1 ocorrência para cada. Essa repetição sugere uma tendência na geração de texto, indicando uma preferência algorítmica por uma formulação específica. Este padrão de escrita não destaca-se de forma massiva para os resumos escritos por humanos, havendo apenas 2 casos de estruturas iniciais idênticas, para os 24.



**Gráfico 2 – Quantitativo das classes gramaticais presentes nos textos de origem humana e artificial (ChatGPT)**

## 5. Considerações finais

Neste estudo exploratório sobre textos acadêmicos de origem humana e artificial foi possível aplicar uma estratégia para análise comparativa sobre a percepção e a classificação da origem de resumos de TCCs por professores da áreas de Tecnologia, assim como sobre a estruturação gramatical destes. Embora os dados coletados não possam ser generalizados, diante do tamanho da amostra, alguns *insights* foram percebidos: (i) maior taxa de erro na classificação de resumos gerados artificialmente, (ii) equivalência na declaração de “não sei classificar” para ambos os tipos de resumo e (iii) similaridade na estrutura gramatical de ambos os tipos de resumos. Considerando este três aspectos pode-se dizer que as pessoas participantes foram levadas a pensar, mesmo sem saber, que o *ChatGPT Plus 4.0* conseguiu reproduzir o estilo de escrita humano. No entanto, o uso recorrente de inicialização dos textos a partir de um padrão

de escrita foi observado como um viés dos resumos gerados pelo *ChatGPT Plus 4.0*, sendo alvo de possível identificação de sua origem.

Mesmo considerando que a quantidade de pessoas participantes na classificação dos resumos foi reduzida - que invalida uma generalização dos resultados - e que o escopo considerou apenas resumos de TCC de Engenharia de Software do curso de BSI da UFPB - sendo estas as limitações atuais da pesquisa - as análises e resultados iniciais são motivadores para continuidade e ampliação da pesquisa. Espera-se que, como trabalho futuro, seja possível identificar estratégias de identificação da origem dos textos, assim como meios de utilizar os recursos de IA como insumo no processo de redação acadêmica, um aliado no processo de ensino aprendizagem, permeado por discussão sobre ética e autoria.

## Referências

- [1] ChatGPT. (2022). "Home." Disponível em: <https://openai.com/blog/chatgpt/>. Acesso em 25 de março de 2023.
- [2] FIA BUSINESS SCHOOL. (2023). "Chat GPT: o que é e como funciona." Disponível em: <https://fia.com.br/blog/chat-gpt/>. Acesso em 27 de março de 2023.
- [3] OpenAI. (2023). "ChatGPT — Notas de lançamento." Disponível em: <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>. Acesso em 03 de abril de 2023.
- [4] OpenAI. (s.d.). Documentação da API GPT. Beta. Disponível em: <https://beta.openai.com/docs/>. Acesso em 5 de abril de 2023.
- [5] Hardware.com.br. "ChatGPT já está começando a protagonizar problemas com plágios e direitos autorais". Disponível em: <https://www.hardware.com.br/noticias/2023-01/chatgpt-ja-esta-comecando-a-protagonizar-problemas-com-plagios-e-direitos-autorais.html>. Acesso em 05 de abril de 2023.
- [6] Pilatti, L. A., Cantorani, J. R. H., & Cechin, M. R. (2023). Como desenvolver a estrutura IMRaD em artigo original. *Retos*, 49, 914-925.
- [7] Ramos, A. S. M. (s.d.). "Inteligência Artificial Generativa baseada em grandes modelos de linguagem - ferramentas de uso na pesquisa acadêmica." Disponível em: <https://doi.org/10.1590/SciELOPreprints.6105>.
- [8] OpenAI. "Model Index for Researchers." OpenAI, 2021, Disponível em: <https://platform.openai.com/docs/model-index-for-researchers>. Acesso em 05 abril de 2023.
- [9] OpenAI. "GPT-4." OpenAI, Disponível em: <https://openai.com/product/gpt-4>. Acesso em 06 abril de 2023.