

INTELIGÊNCIA ARTIFICIAL NA ANÁLISE DE DADOS QUANTITATIVOS DE PESQUISA EDUCACIONAL: CONFIABILIDADE DO DATA ANALYST GPT (CHATGPT) COMPARADO AO SPSS E JAMOVI

INTELIGENCIA ARTIFICIAL EN EL ANÁLISIS DE DATOS CUANTITATIVOS DE INVESTIGACIÓN EDUCATIVA: CONFIABILIDAD DE DATA ANALYST GPT (CHATGPT) COMPARADO CON SPSS Y JAMOVI

ARTIFICIAL INTELLIGENCE IN THE ANALYSIS OF EDUCATIONAL RESEARCH QUANTITATIVE DATA: RELIABILITY OF DATA ANALYST GPT (CHATGPT) COMPARED TO SPSS AND JAMOVI



Cassio SANTOS¹
e-mail: cassiosantos@ie.ulisboa.pt

Como referenciar este artigo:

SANTOS, C. Inteligência Artificial na Análise de Dados Quantitativos de Pesquisa Educacional: Confiabilidade do Data Analyst GPT (ChatGPT) comparado ao SPSS e JAMOVI. **Nuances: Estudos sobre Educação**, Presidente Prudente, v. 35, n. 00, e024013, 2024. e-ISSN: 2236-0441. DOI: <https://doi.org/10.32930/nuances.v35i00.10682>



| **Submetido em:** 20/06/2024
| **Revisões requeridas em:** 15/07/2024
| **Aprovado em:** 12/08/2024
| **Publicado em:** 11/10/2024

Editores: Profa. Dra. Rosiane de Fátima Ponce
Prof. Dr. Paulo César de Almeida Raboni
Editor Adjunto Executivo: Prof. Dr. José Anderson Santos Cruz

¹ Unidade de Investigação e Desenvolvimento em Educação e Formação (UIDEF), Instituto de Educação (IE), Universidade de Lisboa (ULisboa), Lisboa – Portugal. Professor e Pesquisador.

RESUMO: A incorporação da Inteligência Artificial (IA) no cenário educacional e de pesquisa marca uma era transformadora, oferecendo oportunidades sem precedentes para aprimorar a forma como aprendemos e realizamos pesquisas. Este artigo explora o potencial do modelo de linguagem baseado em IA, *Data Analyst GPT*, desenvolvido pela OpenAI, como uma ferramenta confiável para realizar análises de dados quantitativos. A metodologia envolveu o uso do *Data Analyst GPT* e de dois softwares estatísticos padrão, SPSS e JAMOVI, para realizar uma análise estatística completa em um conjunto de dados educacionais típico, abrangendo vários testes estatísticos padrão, como testes de normalidade, análise de correlação (Pearson e Spearman), análise de variáveis categóricas e testes de comparação de médias (teste t, ANOVA, Tukey, Mann-Whitney U e Kruskal-Wallis), e seus resultados foram comparados.

PALAVRAS-CHAVE: Data Analyst GPT. ChatGPT. SPSS. JAMOVI. Inteligência Artificial (IA).

RESUMEN: La incorporación de la Inteligencia Artificial (IA) en el ámbito educativo y de investigación marca una era transformadora, ofreciendo oportunidades sin precedentes para mejorar la forma en que aprendemos y realizamos investigaciones. Este artículo explora el potencial del modelo de lenguaje basado en IA, *Data Analyst GPT*, desarrollado por OpenAI, como una herramienta confiable para llevar a cabo análisis de datos cuantitativos. La metodología involucró el uso de *Data Analyst GPT* y dos softwares estadísticos estándar, SPSS y JAMOVI, para realizar un análisis estadístico completo en un conjunto de datos educativos típico, abarcando varias pruebas estadísticas estándar, como pruebas de normalidad, análisis de correlación (Pearson y Spearman), análisis de variables categóricas y pruebas de comparación de medias (prueba t, ANOVA, Tukey, Mann-Whitney U y Kruskal-Wallis), y sus resultados fueron comparados. Los resultados demuestran una consistencia comparable a la de los software estadísticos estándar.

PALABRAS CLAVE: Data Analyst GPT. ChatGPT. SPSS. JAMOVI. Inteligencia Artificial (IA).

ABSTRACT: The integration of Artificial Intelligence (AI) into the educational and research landscape marks a transformative era, offering unparalleled opportunities for enhancing the way we learn and conduct research. This article explores the potential of the AI-based language model, *Data Analyst GPT*, developed by OpenAI, as a reliable tool for conducting quantitative data analysis. The methodology involved employing *Data Analyst GPT* and two standard statistical software packages, SPSS and JAMOVI, to conduct an end-to-end statistical analysis on a typical educational data set, covering several standard statistical tests such as normality, correlation analysis (Pearson's and Spearman's), Categorical Variables Analysis, and mean comparison tests (Test t, ANOVA, Tukey, Mann-Whitney U and Kruskal-Wallis), and their results were compared. The results demonstrate a consistency comparable to that of standard statistical software.

KEYWORDS: Data Analyst GPT. ChatGPT. SPSS. JAMOVI. Artificial Intelligence (AI).

Introdução

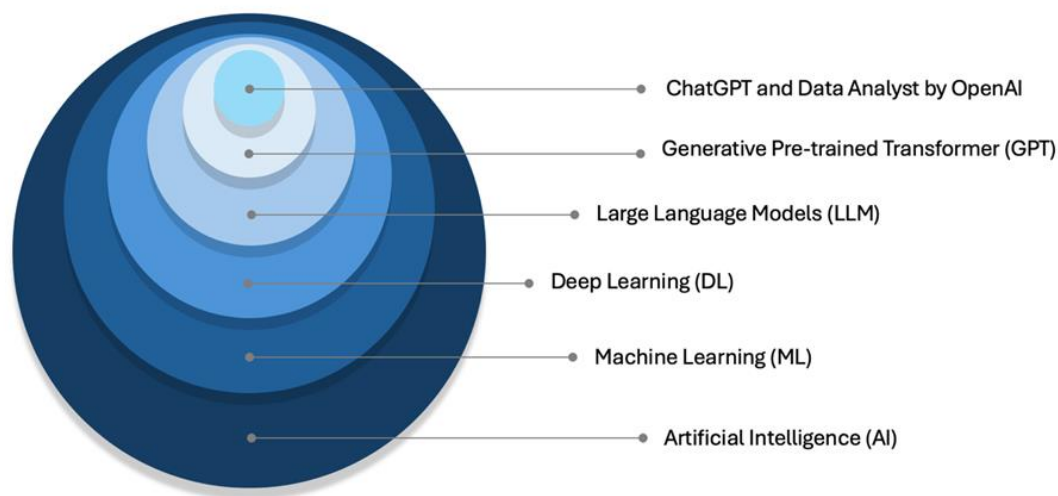
A Inteligência Artificial (IA) ocupa atualmente uma posição central em diversos setores da sociedade, com partes interessadas dedicadas à elaboração de diretrizes focadas na ética (Comissão Europeia, 2019; UNESCO, 2021), na responsabilidade em pesquisa (Comissão Europeia, 2024) e em estratégias educacionais voltadas para educadores (Comissão Europeia, 2022). Instituições de ensino superior, como a Universidade de Stanford (2021), têm contribuído para esse debate, desenvolvendo guias abrangentes. Grandes organizações têm publicado orientações para o uso ético da IA, sinalizando um esforço coordenado para a integração responsável da IA em escala global (European Commission, 2019; UNESCO, 2021).

A UNESCO manifestou preocupação com a utilização da IA no ensino superior, abordando desde o desenvolvimento de Guias de Início Rápido (UNESCO, 2023a) até questões mais específicas, como o uso do ChatGPT, além de tópicos mais amplos para auxiliar as partes interessadas no melhor uso da IA (UNESCO, 2023b). Ademais, destacam-se iniciativas voltadas para a criação de estruturas focadas em competências relacionadas à IA (Ehlers *et al.*, 2023). No contexto europeu, marcos importantes incluem a aprovação, pela Comissão Europeia, das primeiras diretrizes globais para o uso da IA (Comissão Europeia, 2021) e a formulação das Diretrizes Éticas para IA Confiável (Comissão Europeia, 2019).

A evolução da série *Generative Pre-Trained Transformer* (GPT), desenvolvida pela OpenAI, começou em 2018 com o lançamento inaugural, seguido pelo modelo GPT-3, em 2020, que representou um avanço significativo nas capacidades da IA. Aproveitando a base estabelecida pelo GPT-3, o ChatGPT foi lançado em 2022, com foco específico na IA conversacional e em aplicações de chatbot (Li *et al.*, 2024; Singh-Harjit, Singh-Avneet, 2023).

A relação entre IA, Aprendizado de Máquina (ML), Aprendizado Profundo (DL), Modelos de Linguagem de Grande Escala (LLM) e GPT é estruturada de maneira hierárquica e especializada. O ML (Lary *et al.*, 2016) fornece a estrutura geral para a aprendizagem a partir de dados; o DL (Alzubaidi *et al.*, 2021) oferece técnicas avançadas para aprendizado a partir de conjuntos de dados complexos e de alta dimensionalidade; os LLM (Chang, 2023; Fan *et al.*, 2023; Li *et al.*, 2024; Naveed *et al.*, 2023) processam e geram linguagem natural em grande escala, facilitando a interação natural entre humanos e computadores; e o GPT, que utiliza DL e LLM, oferece aos usuários respostas coerentes e contextualmente relevantes para suas perguntas (Gimpel *et al.*, 2023). O *GPT Data Analyst* é uma versão personalizada do ChatGPT, otimizada para análise de dados. A Figura 1 ilustra essa hierarquia de especialização.

Figura 1 - Hierarquia e Especialização da Inteligência Artificial



Fonte: Elaborado pelo autor (2024).

A implementação da Inteligência Artificial (IA) na educação teve um impacto significativo, refletido nas melhorias da eficiência do processo educacional, na promoção do aprendizado global, na personalização do ensino, na criação de conteúdo mais inteligente e na otimização da gestão educacional em termos de eficácia e eficiência (Montenegro-Rueda *et al.*, 2023). A relação entre IA e pesquisa no ensino superior é dual: “A primeira está relacionada à pesquisa sobre IA, enquanto a segunda trata da pesquisa utilizando ou apoiada por ferramentas de IA” (UNESCO, 2023b, p. 38, tradução nossa). A IA é capaz de processar grandes volumes de dados (Gimpel *et al.*, 2023), aprender automaticamente a identificar padrões complexos e tendências ocultas, além de apresentar flexibilidade para se adaptar a diferentes tipos de dados e contextos de pesquisa. Portanto, pode aprimorar a compreensão e a interpretação de dados quantitativos no campo da educação.

A literatura acadêmica contém uma vasta quantidade de pesquisas focadas no uso da IA na Educação (Al-Ghonmein, Al-Moghrabi, 2024; Crawford *et al.*, 2024; Ding *et al.*, 2023; Jia, Tu, 2024). No entanto, os estudos que exploram o potencial dessa tecnologia na análise de dados ainda são relativamente escassos (Huang *et al.*, 2024; Mohammadi, Nguyen, 2024; Sufi, 2024; Walter, 2024).

Seguindo as diretrizes estabelecidas no documento “Diretrizes Vivas sobre o Uso Responsável da IA Generativa em Pesquisa” (Comissão Europeia, 2024), é crucial ressaltar a importância da responsabilidade e integridade por parte dos pesquisadores em relação à produção científica apoiada por IA. Esse documento enfatiza que os pesquisadores devem ser, em última instância, responsáveis pelo conteúdo científico gerado ou apoiado por ferramentas

de IA, adotando uma postura crítica e cientes das limitações inerentes à IA generativa, como vieses e imprecisões.

A transparência no uso dessas ferramentas de IA é um ponto-chave destacado nas diretrizes de pesquisa (Comissão Europeia, 2024). Recomenda-se que os pesquisadores descrevam quais ferramentas de IA generativa foram utilizadas em seus processos de pesquisa, incluindo informações como nome, versão e data da ferramenta, além de como ela influenciou o processo de pesquisa. A documentação adequada de insumos (prompts) e resultados, sempre que relevante, é incentivada para promover a abertura e a replicabilidade da pesquisa. Por fim, as diretrizes encorajam os pesquisadores a se engajarem em aprendizado contínuo sobre o uso adequado das ferramentas de IA generativa. Dado o rápido desenvolvimento dessas tecnologias e o surgimento constante de novas aplicações, é fundamental que os pesquisadores se mantenham atualizados sobre as melhores práticas, participem de treinamentos e compartilhem conhecimento com colegas e outras partes interessadas, a fim de maximizar os benefícios dessas ferramentas avançadas para a pesquisa.

Diversas instituições expressaram preocupações sobre privacidade, confidencialidade e direitos de propriedade intelectual, seja pelo fato de que “modelos como o ChatGPT são opacos em relação ao conjunto de dados utilizado para treiná-los” (UNESCO, 2021, p. 7) ou ao compartilhar informações sensíveis ou protegidas com ferramentas de IA, em que “os pesquisadores devem estar atentos ao fato de que os dados ou entradas geradas ou carregadas (texto, dados, prompts, imagens, etc.) podem ser utilizadas para outros fins, como o treinamento de modelos de IA” (Comissão Europeia, 2024, p. 6).

A pesquisa demonstra que o ChatGPT pode auxiliar na análise e síntese de grandes conjuntos de dados, facilitando a identificação rápida e eficaz de padrões e insights que seriam difíceis de descobrir manualmente. Além disso, o modelo pode ser utilizado para gerar propostas de pesquisa, revisões de literatura e outros documentos relacionados à pesquisa (Atlas, 2023, p. 24).

Este artigo tem como objetivo investigar a confiabilidade do *Data Analyst GPT*, a versão personalizada e otimizada do ChatGPT-4 para análise de dados, proporcionando uma interface conversacional inteligente e versátil para a análise de dados de pesquisa quantitativa. O desempenho do *Data Analyst GPT* será comparado ao obtido com o uso de dois pacotes de software estatístico padrão, SPSS e JAMOVI.

Metodologia

Esta seção detalha os procedimentos do estudo, apresentando o *Data Analyst GPT* e o *software* de *benchmark*. No *Data Analyst GPT*, foram desenvolvidos *prompts* personalizados para executar os cálculos, enquanto os testes do SPSS e JAMOVI foram conduzidos de acordo com os procedimentos descritos em seus respectivos manuais de suporte

Data Analyst GPT

Os testes baseados em IA foram realizados utilizando o *Data Analyst GPT*, versão do ChatGPT otimizada para análise de dados, com o plano de assinatura Plus. No ChatGPT, o “*Data Analyst*” GPT é acessado por meio da seção “Explorar GPTs”, onde um arquivo Excel (*xlsx) contendo o conjunto de dados a ser analisado pode ser enviado. No momento da redação deste texto, o modelo GPT-4 era o mais avançado da série GPT.

Software estatístico padrão como *benchmark*

Dois pacotes de *software* estatístico padrão foram selecionados para serem utilizados como *benchmark*: o SPSS (versão 29.0.2.0 [20]) e o JAMOVI (versão 2.3.21.0), ambos operando em um sistema Mac OS. O SPSS é um *software* estatístico amplamente utilizado que permite diversos tipos de análises, transformações e formas de saída (Alili; Krstev, 2019). O JAMOVI² (R Core Team, 2021; The Jamovi Project, 2022) é um *software* estatístico gratuito e de código aberto, que foi adotado pela comunidade de pesquisa devido à sua facilidade de uso e à ampla gama de funções estatísticas, abrangendo desde análises básicas até técnicas multivariadas e univariadas avançadas (Alghami, Hussin, 2022; Marek *et al.*, 2023).

Conjunto de dados

Este artigo é baseado em pesquisas já publicadas. O conjunto de dados original foi adaptado, e novas variáveis foram adicionadas para abranger uma gama mais ampla de testes estatísticos necessários para verificar a confiabilidade do *Data Analyst GPT*. O conjunto de dados contém uma variedade de variáveis, permitindo o teste de diferentes hipóteses e cenários.

² Disponível em: <https://www.jamovi.org/about.html>. Acesso em: 15 set. 2024.

Testes Estatísticos

Este artigo aborda os testes estatísticos mais utilizados na pesquisa educacional, incluindo testes de normalidade, análise de correlação, análise de variáveis categóricas e testes de comparação de médias.

Normalidade

Dois abordagens complementares foram empregadas para avaliar a normalidade da distribuição dos dados: a análise estatística de Shapiro-Wilk e a inspeção visual da distribuição dos dados por meio de análises gráficas, como *boxplots* e histogramas. Essas abordagens proporcionam uma avaliação robusta da normalidade, permitindo a identificação visual de assimetrias, valores atípicos e a forma geral da distribuição. Ambas as análises foram realizadas nas variáveis “points_1” e “points_2”.

O teste de Shapiro-Wilk foi utilizado para comparar os dados de uma amostra com um conjunto de dados que segue uma distribuição normal, ou seja, com a mesma média e desvio padrão. Neste teste, resultados não significativos ($p > 0,05$) indicam que a distribuição dos dados da amostra não difere de uma distribuição normal, sugerindo que os dados seguem uma distribuição normal. Por outro lado, um resultado significativo ($p < 0,05$) indica que a distribuição dos dados é diferente de uma distribuição normal (Dancey; Reidy, 2020; Field, 2024).

A abordagem visual permite que pesquisadores e analistas realizem uma inspeção detalhada e intuitiva da distribuição dos dados, facilitando a identificação de características importantes, como assimetria e curtose. Por meio do uso de gráficos específicos, como *boxplots* e histogramas, é possível observar padrões, tendências e desvios que podem não ser evidentes por métodos puramente numéricos ou estatísticos (Field, 2024).

Um *boxplot* é uma representação gráfica eficaz que destaca as características essenciais de um conjunto de dados, sendo especialmente útil quando os dados se ajustam a uma distribuição normal. Em seu centro, encontra-se a “mediana”, elegantemente contida dentro de uma caixa. Os limites superior e inferior dessa caixa representam os quartis superior e inferior, respectivamente, demarcando a amplitude interquartil que abrange os 50% centrais dos pontos de dados. A partir da caixa, “bigodes” se estendem até os pontos de dados mais altos e mais baixos, delineando a dispersão geral dos dados. Da mesma forma, um histograma serve como uma ferramenta gráfica que retrata a distribuição de frequência de um conjunto de dados. Ele

facilita a visualização da distribuição, ilustrando a frequência de ocorrência de cada valor. Essa visualização é alcançada ao segmentar o conjunto de dados em intervalos definidos, ou “bins”, e contabilizar as observações dentro destes. Esses bins são projetados para serem sequenciais, distintos e uniformes em tamanho (Field, 2024).

Prompt no *Data Analyst GPT*: “É necessário realizar um teste de Shapiro-Wilk na coluna de dados 'point_1', com os resultados apresentados em uma tabela no formato APA. A tabela deve incluir o valor do teste estatístico, o valor de p (arredondado para três casas decimais) e os graus de liberdade. Além disso, solicito a geração de um *boxplot* e de um histograma para a coluna 'points_1', com o objetivo de avaliar visualmente sua distribuição”.

Prompt no *Data Analyst GPT*: “É necessário realizar um teste de Shapiro-Wilk na coluna de dados 'point_2', com os resultados apresentados em uma tabela no formato APA. A tabela deve incluir o valor do teste estatístico, o valor de p (arredondado para três casas decimais) e os graus de liberdade. Além disso, solicito a geração de um *boxplot* e de um histograma para a coluna 'points_2', com o objetivo de avaliar visualmente sua distribuição”.

Análise de Correlação

Foram empregados os testes de Spearman e Pearson para analisar as correlações entre as variáveis “points_1” e “number_of_devices”.

Os testes de correlação são usados para avaliar a força e a direção da associação entre duas variáveis quantitativas. O teste de correlação de Spearman, também conhecido como rho (ρ), é utilizado em situações onde os dados não atendem aos pressupostos de normalidade ou em casos de variáveis ordinais, fornecendo uma medida robusta de correlação sem assumir uma relação linear específica. Em contrapartida, a correlação de Pearson, representada pelo símbolo r , é indicada para dados com distribuição normal e uma relação linear, sendo uma medida da força e direção dessa linearidade. Ambos os testes variam entre -1 e 1, indicando uma forte relação linear quando os valores se aproximam de -1 ou 1, seja negativa ou positiva, respectivamente (Dancey; Reidy, 2020; Field, 2024).

Coefficiente de Spearman na variável “points_1”:

Prompt no *Data Analyst GPT*: “É necessário calcular o coeficiente de Spearman entre a coluna de dados 'points_1' e a coluna de dados 'number_of_devices', com os resultados

apresentados em uma tabela no formato APA. A tabela deve incluir o valor do teste estatístico e o valor de p (arredondado para três casas decimais)”.

Coefficiente de Pearson na variável “points_2”:

Prompt no Data Analyst GPT: “É necessário calcular o coeficiente de Pearson entre a coluna de dados 'points_2' e a coluna de dados 'number_of_devices', com os resultados apresentados em uma tabela no formato APA. A tabela deve incluir o valor do teste estatístico e o valor de p (arredondado para três casas decimais)”.

Análise de Variáveis Categóricas

Foi utilizado o teste Qui-quadrado nas variáveis categóricas '*proficiency_level*' e '*situation*' para verificar a independência entre elas. O teste Qui-quadrado é uma ferramenta estatística utilizada para comparar as frequências observadas com as frequências esperadas em diferentes categorias de uma variável categórica. Esse teste auxilia na determinação de diferenças significativas entre as categorias, verificando se as frequências observadas se desviam significativamente das esperadas ao acaso. O teste é amplamente utilizado em pesquisas para testar hipóteses sobre a associação ou independência entre categóricas (Dancey; Reidy, 2020; Field, 2024).

Prompt no Data Analyst GPT: “É necessário realizar um teste Qui-quadrado entre a coluna de dados 'proficiency_level' e a coluna de dados 'situation', com os resultados apresentados em uma tabela no formato APA. A tabela deve incluir o valor do teste estatístico, o valor de p (arredondado para três casas decimais) e os graus de liberdade”.

Testes de Comparação de Médias

Fatores com dois grupos

Para analisar o conjunto de dados com fatores que possuem dois grupos, foram aplicados o teste U de Mann-Whitney para variáveis não paramétricas e o teste t para variáveis paramétricas. O **teste t para amostras independentes** baseia-se na suposição de que as populações das quais as amostras são extraídas apresentam distribuições normais com variâncias iguais. Esse teste é amplamente reconhecido pela sua rigorosidade na comparação de médias (Dancey; Reidy, 2020; Field, 2024). Em contrapartida, o **teste U de Mann-Whitney**, que não pressupõe normalidade ou igualdade de variâncias entre os grupos, é utilizado para dados que não satisfazem os pressupostos paramétricos (Dancey; Reidy, 2020; Field, 2024).

Teste Mann-Whitney U na variável “point_1”

O seguinte prompt foi solicitado ao *Data Analyst GPT*: “Realize o teste U de Mann-Whitney na coluna de dados 'points_1', comparando a variável 'gender', onde 0 corresponde a Masculino e 2 a Feminino. Os resultados devem ser apresentados em uma tabela de dados no formato APA, incluindo o valor do teste estatístico e o valor de p, arredondado para três casas decimais”.

Teste t na variável “point_2”

O seguinte prompt foi solicitado ao *Data Analyst GPT*: “Realize o teste t na coluna de dados 'points_2', comparando a variável 'gender', onde 0 corresponde a Masculino e 2 a Feminino. Os resultados devem ser apresentados em uma tabela de dados no formato APA, incluindo o valor do teste estatístico, o valor de p (arredondado para três casas decimais) e os graus de liberdade”.

Mais de dois grupos

Quando há mais de dois grupos, é utilizado o **teste de Kruskal-Wallis** para variáveis não paramétricas e a **Análise de Variância (ANOVA)** para variáveis paramétricas. O **teste de Kruskal-Wallis** é uma alternativa não paramétrica ao ANOVA de uma via e é empregado quando os pressupostos de normalidade não são atendidos. Esse teste avalia se as medianas dos rankings entre dois ou mais grupos diferem significativamente entre si. A ANOVA, por sua

vez, é usada para comparar as médias entre três ou mais grupos, assumindo normalidade e homogeneidade de variâncias (Dancey; Reidy, 2020; Field, 2024).

Quando a ANOVA detecta diferenças significativas entre as médias dos grupos, são necessários **testes post-hoc** para identificar quais grupos específicos diferem entre si. O teste de **Tukey** foi utilizado para realizar essas comparações pareadas, devido à sua eficácia em controlar a taxa de erro do Tipo I (Dancey; Reidy, 2020; Field, 2024).

Teste de Kruskal-Wallis na variável “point_1”

O seguinte prompt foi solicitado ao *Data Analyst GPT*: “Realize o teste de Kruskal-Wallis na coluna de dados 'points_1', comparando a variável 'level_of_education', onde 2 corresponde a Bacharelado, 3 a Mestrado e 4 a Doutorado. Os resultados devem ser apresentados em uma tabela de dados no formato APA, incluindo o valor do teste estatístico, o valor de *p* (arredondado para três casas decimais) e os graus de liberdade”.

ANOVA na variável “point_2”

O seguinte prompt foi solicitado ao *Data Analyst GPT*: “Realize o teste ANOVA na coluna de dados 'points_2', comparando a variável 'level_of_education', onde 2 corresponde a Bacharelado, 3 a Mestrado e 4 a Doutorado. Além disso, realize o teste de Levene para verificar a homogeneidade das variâncias. Os resultados devem ser apresentados em tabelas de dados no formato APA, incluindo o valor do teste estatístico, o valor de *p* (arredondado para três casas decimais) e os graus de liberdade”.

Prompt solicitado ao *Data Analyst GPT*: “Realize o teste ANOVA na coluna de dados 'points_2', comparando a variável 'situation', onde 1 corresponde a Empregado, 2 a Aposentado, 3 a Desempregado e 4 a Estudante. Além disso, realize o teste de Levene para verificar a homogeneidade das variâncias. Os resultados devem ser apresentados em tabelas de dados no formato APA, incluindo o valor do teste estatístico, o valor de *p* (arredondado para três casas decimais) e os graus de liberdade. Caso sejam identificadas diferenças estatisticamente significativas, aplique o teste post-hoc de Tukey”.

Resultados

Para fins de comparação, os mesmos testes estatísticos foram realizados utilizando dois *softwares* estatísticos padrão, o SPSS e o JAMOVI. Isso permitiu a comparação direta dos resultados obtidos com os fornecidos pelo *Data Analyst GPT*.

Testes de Normalidade

Duas abordagens distintas foram adotadas para a realização dos testes de normalidade: a análise estatística e a análise gráfica.

Análise Estatística

O teste de Shapiro-Wilk foi aplicado para verificar a normalidade das distribuições, conforme apresentado na Tabela 1.

Tabela 1 - Resultados do Teste de Shapiro-Wilk

Software	"point_1"			"point_2"		
	estatística	df	valor p	estatística	df	valor p
Data Analyst GPT	0.994	845	<.001	0.998	845	0.555
SPSS	0.994	846	<.001	0.998	846	0.555
JAMOVI	0.994	-	<.001	0.998	-	0.555

Fonte: Elaborado pelo autor (2024).

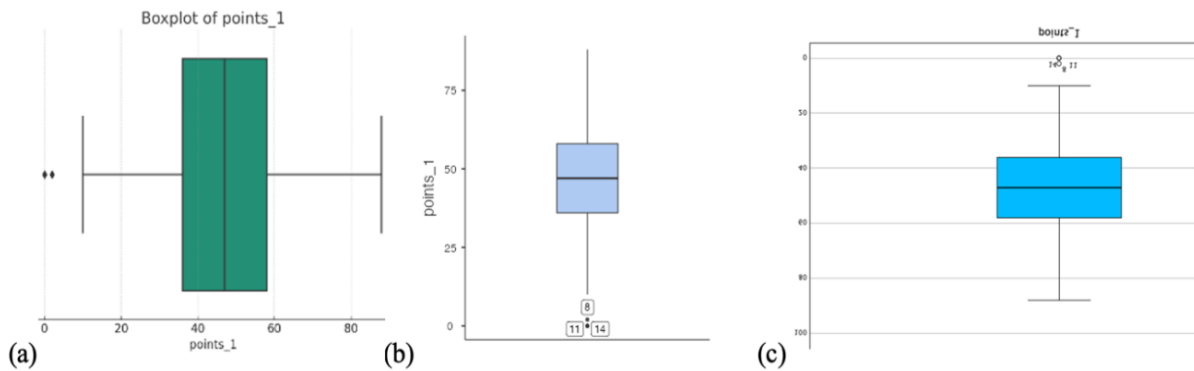
Os resultados indicam consistência nos valores estatísticos e nos p-valores obtidos pelo *Data Analyst GPT* em comparação com o SPSS e o JAMOVI no teste de Shapiro-Wilk. No entanto, uma diferença significativa foi observada nos graus de liberdade (df), com o *Data Analyst GPT* documentando 845, o SPSS apresentando 846 e o JAMOVI omitindo essa informação nas variáveis "point_1" e "point_2".

Ao selecionar os testes estatísticos adequados, é essencial determinar se a amostra segue uma distribuição normal. Os resultados da análise estatística indicam que a variável "point_1" não segue uma distribuição normal ($p < 0,05$), sugerindo uma possível assimetria ou excesso de curtose. Em contrapartida, a variável "point_2" apresentou características de normalidade ($p > 0,05$), o que sugere que sua distribuição está de acordo com a distribuição normal. Essa distinção é fundamental para a escolha apropriada dos testes estatísticos, garantindo a validade e a confiabilidade das análises.

Análise Gráfica

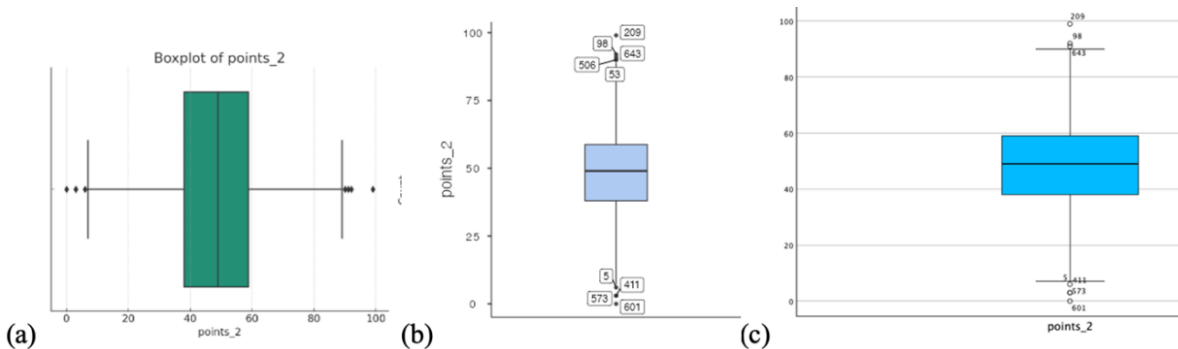
Boxplots e histogramas elaborados por ([a] *Data Analyst GPT*, [b] SPSS e [c] JAMOVI) foram utilizados para verificar a comparabilidade ao testar a normalidade das distribuições das variáveis “point_1” e “point_2”. Os *boxplots* para a variável “point_1” estão apresentados no Gráfico 2, enquanto os *boxplots* para “point_2” estão no Gráfico 3. Os histogramas referentes à variável “point_1” são mostrados no Gráfico 4, e os dados de “point_2” estão representados no Gráfico 5.

Gráfico 2 – Boxplots obtidos para a variável “point_1”



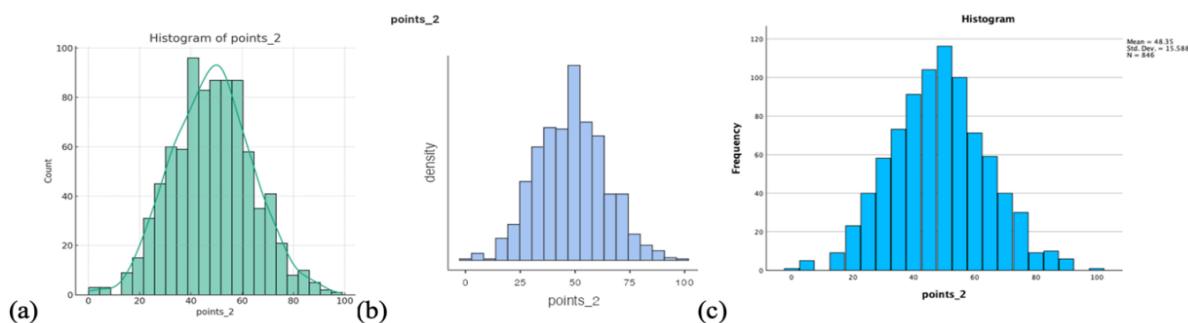
Fonte: Elaborado pelo autor (2024).

Gráfico 3 – Boxplots obtidos para a variável “point_2”



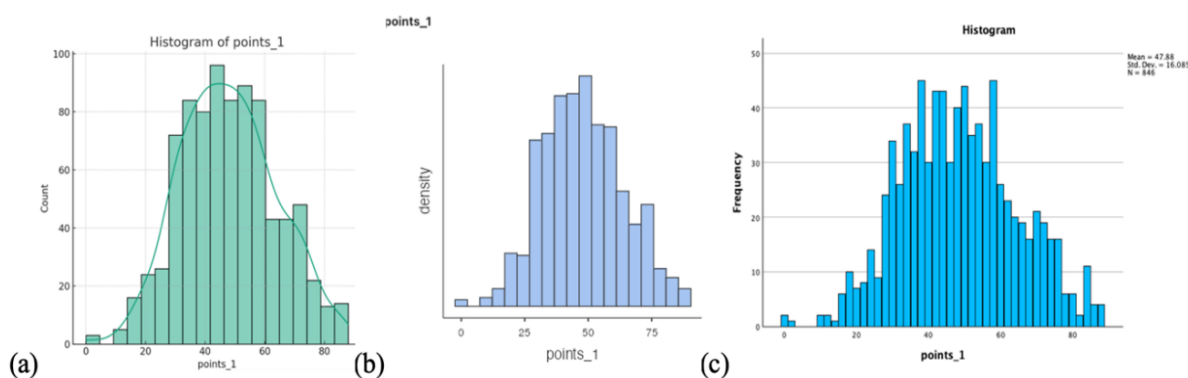
Fonte: Elaborado pelo autor (2024).

Gráfico 4 – Histogramas obtidos para a variável "point_1"



Fonte: Elaborado pelo autor (2024).

Gráfico 5 – Histogramas obtidos para a variável "point_2"



Fonte: Elaborado pelo autor (2024).

A análise gráfica realizada por meio de *boxplots* demonstra a confiabilidade do *Data Analyst GPT* em comparação com o SPSS e o JAMOVI. É importante ressaltar que, enquanto o *Data Analyst GPT* foi capaz de identificar apenas a presença de valores discrepantes, tanto o SPSS quanto o JAMOVI também conseguiram indicar quais desses valores eram considerados outliers.

Análise de Correlação

Com base nos resultados obtidos pelos testes de normalidade, o coeficiente de Spearman foi utilizado para a variável não paramétrica "point_1" ($p < 0,05$) e o coeficiente de Pearson para a variável paramétrica "point_2" ($p > 0,05$) em relação ao "number_of_devices", conforme pode ser observado na **Tabela 2**.

Tabela 2 – Coeficiente de Spearman (“point_1”) e coeficiente de Pearson (“point_2”)

Software	Coeficiente de Spearman		Coeficiente de Pearson	
	ρ (rho)	valor-p	ρ (rho)	valor-p
Data Analyst GPT	0.228	<.001	-0.025	0.468
SPSS	0.228	<.001	-0.025	0.468
JAMOVI	0.228	<.001	-0.025	0.468

Fonte: Elaborado pelo autor (2024).

Os resultados indicam a confiabilidade do *Data Analyst GPT* em comparação com o SPSS e o JAMOVI, tanto para o coeficiente de Spearman, aplicado a dados com distribuição não normal, quanto para o coeficiente de Pearson, utilizado para os dados com distribuição normal.

Análise de Variáveis Categóricas

Testes de Qui-quadrado foram realizados nas variáveis categóricas “points_2” e “situation”, conforme mostrado na **Tabela 3**.

Tabela 3 – Qui-quadrado

Software	estatística	df	valor-p
Data Analyst GPT	50.767	15	<.001
SPSS	50.767	15	<.001
JAMOVI	50.8	15	<.001

Fonte: Elaborado pelo autor (2024).

Os resultados indicam que o *Data Analyst GPT* foi capaz de fornecer resultados consistentes quando comparado ao SPSS e ao JAMOVI na análise das variáveis categóricas utilizando o teste de Qui-quadrado. É importante observar que, enquanto o JAMOVI reportou as estatísticas com uma casa decimal, os outros softwares as exibiram com três casas decimais. No entanto, essa diferença de formatação não impacta a interpretação dos resultados.

Testes de Comparação de Médias

Fatores com Dois Grupos

Com base no conhecimento obtido sobre a normalidade dos dados, o teste de Mann-Whitney U foi utilizado para a variável não paramétrica “point_1”, enquanto o teste t foi empregado para a variável paramétrica “point_2”, conforme pode ser observado na **Tabela 4**.

Tabela 4 – Mann-Whitney U com “gênero”

Software	Mann-Whitney U		Test t		
	estatística	valor-p	estatística	df	valor-p
Data Analyst GPT	92,001.5	0.390	0.810	844	0.418
SPSS	85,903.5	0.389	0.810	844	0.418
JAMOVI	85,904.0	0.390	0.810	844	0.418

Fonte: Elaborado pelo autor (2024).

Em relação ao valor p do teste de Mann-Whitney U (Tabela 4), tanto o *Data Analyst GPT* quanto o JAMOVI apresentaram resultados semelhantes. O SPSS, por sua vez, mostrou ligeiras variações no terceiro decimal, o que pode ser atribuído as diferenças de arredondamento. Além disso, uma discrepância notável foi observada no valor da estatística obtida pelo *Data Analyst GPT*. Enquanto o SPSS e o JAMOVI apresentaram resultados similares com pouca variação, essa diferença não afeta a interpretação dos resultados.

Mais de Dois Grupos

Os testes de Kruskal-Wallis foram empregados para variáveis não paramétricas, como “point_1” e “nivel_de_educacao”, enquanto os testes ANOVA foram utilizados para variáveis paramétricas, como “point_2”, juntamente com os testes de Levene, conforme pode ser visto na **Tabela 5**.

Tabela 5 – Testes de Kruskal-Wallis, ANOVA e Levene

Software	Teste de Kruskal-Wallis			Teste ANOVA			Teste de Levene		
	estatística	df	valor-p	estatística	df1	df2	valor-p	F	valor-p
Data Analyst GPT	9.741	2	0.008	0.882	2	843	0.414	0.281	0.755
SPSS	9.741	2	0.008	0.882	2	843	0.414	0.281 ^a	0.755 ^a
JAMOVI	9.74	2	0.008	0.882	2	843	0.414	0.252 ^b	0.777 ^b

^aCom base na mediana; ^bCom base na média.

Fonte: Elaborado pelo autor (2024).

Os resultados indicam que o *Data Analyst GPT* foi capaz de fornecer resultados comparáveis aos obtidos pelo SPSS e pelo JAMOVI, tanto para o teste de Kruskal-Wallis, aplicado a dados com distribuição não normal, quanto para o teste ANOVA, que considera dados normalmente distribuídos.

Em relação ao teste de Levene, o software SPSS ofereceu duas variantes: uma calculada com base na mediana e a outra com base na média. Em contraste, tanto o *Data Analyst GPT* quanto o JAMOVI forneceram apenas uma versão dos resultados. Os dados sugerem que o *Data*

Analyst GPT considerou a mediana para os cálculos, enquanto o JAMOVI utilizou a média. Essa correspondência demonstra uma forte consistência entre ambos os *softwares*.

Adicionalmente, foi realizado outro teste ANOVA para analisar a confiabilidade do teste post-hoc, considerando agora as variáveis “points_2” e “situation”. O objetivo foi avaliar a confiabilidade do *software* quando há uma diferença estatística significativa entre as médias. Os resultados dos testes ANOVA e de Levene são apresentados na **Tabela 6**, enquanto os resultados dos testes post-hoc de Tukey podem ser encontrados na **Tabela 7**.

Tabela 6 – Testes ANOVA e de Levene

Software	Teste ANOVA				Teste de Levene	
	estatística	df1	df2	valor-p	F	valor-p
Data Analyst GPT	3.372	3	842	0.018	0.108	0.956 ^a
SPSS	3.372	3	842	0.018	0.108 ^a	0.956 ^a
JAMOVI	3.37	3	842	0.018	0.029 ^b	0.993 ^b

^aCom base na mediana; ^bCom base na média.

Fonte: Elaborado pelo autor (2024).

Tabela 7 - Tukey post-hoc

Comparação	Analista de Dados GPT	SPSS	JAMOVI
		estatística	
Empregado	Aposentado	0.030	0.030
	Desempregado	0.900	0.946
	Estudante	0.484	0.483
Aposentado	Desempregado	0.177	0.177
	Estudante	0.030	0.030
Desempregado	Estudante	0.900	0.926

Fonte: Elaborado pelo autor (2024).

Os resultados demonstram a confiabilidade do *Data Analyst GPT* em comparação com o SPSS e o JAMOVI nos testes ANOVA, especialmente quando há uma diferença estatisticamente significativa entre as médias, assim como nos testes post-hoc de Tukey. É relevante observar que a discrepância observada no teste de Levene origina-se da mesma variabilidade encontrada na ANOVA, que ocorre quando não há diferenças estatisticamente significativas entre as médias.

Discussão

Este artigo teve como objetivo avaliar a confiabilidade do *Data Analyst GPT* (ChatGPT) na análise de dados quantitativos, realizando uma comparação direta com os resultados obtidos de dois pacotes de software estatístico clássicos, o SPSS e o JAMOVI. Os testes selecionados para essa comparação direta incluíram: a) testes de normalidade; b) análise de correlação utilizando o coeficiente de Pearson para amostras paramétricas e o coeficiente de Spearman para amostras não paramétricas; c) o teste de qui-quadrado para a análise de variáveis categóricas; e d) testes de comparação de médias, incluindo o teste t e ANOVA para amostras paramétricas, e os testes Mann-Whitney U e Kruskal-Wallis para amostras não paramétricas.

Duas abordagens foram empregadas para avaliar a confiabilidade do *Data Analyst GPT* na análise de normalidade: o teste estatístico de Shapiro-Wilk para uma avaliação quantitativa e a análise gráfica com *boxplots* e histogramas para uma avaliação visual. Os testes realizados com o *Data Analyst GPT* e os softwares estatísticos de referência, nomeadamente SPSS e JAMOVI, requerem conhecimento intermediário de estatística e análise de dados (Huang *et al.*, 2024), especialmente na seleção dos testes estatísticos a serem aplicados. No caso do *Data Analyst GPT*, a execução é facilitada por uma interface de chat.

O teste de Shapiro-Wilk foi aplicado às variáveis “points_1” e “points_2” para verificar a normalidade dos dados, e os resultados foram equivalentes em termos de estatística do teste e valor p. No entanto, observa-se uma discrepância aparente nos Graus de Liberdade (df), com o *Data Analyst GPT* registrando 845, o SPSS 846 e o JAMOVI omitindo essa métrica. É importante esclarecer que, no contexto do teste de Shapiro-Wilk, o conceito de graus de liberdade não é normalmente utilizado, uma vez que esse teste se concentra em avaliar se uma amostra provém de uma distribuição normal, sem depender diretamente dos graus de liberdade que geralmente se aplicam a testes envolvendo variações ou desvios padrão. Portanto, a menção aos graus de liberdade nesse contexto pode não ser essencial, o que pode justificar, ao menos em parte, a ausência desses valores no teste de Shapiro-Wilk realizado pelo JAMOVI.

Os *boxplots* gerados para avaliar a distribuição dos dados foram informativos, pois ilustraram seus quartis e destacaram os outliers. Nos *boxplots* produzidos pelo SPSS (Gráfico 2b) e pelo JAMOVI (Gráfico 2c), os outliers são indicados na parte inferior; os *boxplots* elaborados pelo *Data Analyst GPT* (Gráfico 2a) não mostram tais outliers, mas apenas indicam sua existência. Portanto, observa-se uma limitação no *Data Analyst GPT* em relação à representação visual dos outliers, o que pode restringir uma análise mais aprofundada das

variações extremas nos dados. Contudo, em casos onde múltiplos outliers estão presentes, os visualizar torna-se desafiador, independentemente do *software* utilizado.

Os histogramas produzidos para avaliar a distribuição de frequência dos dados forneceram uma visualização clara e informativa. Eles permitem uma comparação equivalente, mesmo quando suas escalas foram ajustadas automaticamente e diferentes intervalos de dados foram definidos pelo *software*. Isso pode ocorrer tanto no eixo X (abscissa), que representa a frequência de cada intervalo, com a barra mais alta indicando a maior frequência de valores, quanto no eixo Y (ordenada), que reflete a contagem numérica das ocorrências correspondentes no eixo X. Os histogramas do *Data Analyst GPT* (Gráfico 4a) foram vantajosos. A inclusão de uma curva de densidade ou Estimativa de Densidade de Kernel (KDE) fornece uma perspectiva adicional sobre a distribuição geral dos dados, sugerindo a forma da distribuição subjacente de maneira mais contínua e integrada.

Os resultados dos testes de correlação de Spearman e Pearson, bem como do teste de Qui-quadrado, demonstram que o *Data Analyst GPT* apresenta desempenho comparável ao do SPSS e do JAMOVI. Isso ressalta a confiabilidade e a precisão da ferramenta na análise de correlação, afirmando sua capacidade de fornecer resultados analíticos robustos. Quanto ao teste de Mann-Whitney U para comparação de médias, os resultados indicam a confiabilidade do *Data Analyst GPT* em comparação com o SPSS e o JAMOVI em termos de valor-p, mesmo considerando uma leve diferença no terceiro decimal atribuível ao arredondamento. No entanto, o *Data Analyst GPT* apresentou uma diferença significativa nos valores da estatística U quando comparado aos outros *softwares*.

O teste Mann-Whitney U foi inicialmente concebido por Frank Wilcoxon (Wilcoxon, 1945) para analisar medidas de tendência central em amostras do mesmo tamanho. Posteriormente, Henry B. Mann e Donald R. Whitney (Mann; Whitney, 1947) estenderam sua aplicação para amostras de tamanhos diferentes. Dessa forma, os valores estatísticos do teste Mann-Whitney U podem ser derivados por meio de duas abordagens distintas: a formulação baseada em postos (Wilcoxon, 1945) e o método de comparação direta (Mann; Whitney, 1947). A formulação baseada em postos envolve a ordenação combinada de todos os valores de ambos os grupos, atribuindo postos a cada valor e utilizando estes para calcular a estatística U, ajustando efetivamente quaisquer empates (Wilcoxon, 1945). Em contraste, o método de comparação direta quantifica o número de vezes que um valor de um grupo excede o de outro, oferecendo uma abordagem intuitiva que, apesar de sua simplicidade, torna-se impraticável para análises de grandes amostras devido às demandas computacionais (Mann; Whitney, 1947).

A documentação do SPSS (IBM Corporation, 2022) menciona o uso da formulação baseada em postos, mas não foram encontrados documentos equivalentes para o *Data Analyst GPT* e o JAMOVI. A similaridade dos resultados das estatísticas U entre o SPSS e o JAMOVI (85.903,5 e 85.904,0, respectivamente) pode indicar que o JAMOVI também emprega o método baseado em postos, enquanto o *Data Analyst GPT* adota o método de comparação direta. Considerando o valor-p do teste de Mann-Whitney, pode-se concluir que o *Data Analyst GPT* forneceu resultados confiáveis em comparação com o SPSS e o JAMOVI.

Os testes de Kruskal-Wallis e ANOVA mostram que a confiabilidade do *Data Analyst GPT* é comparável à do SPSS e do JAMOVI. Essa equivalência também se estende aos resultados dos testes post-hoc de Tukey, aplicados quando a ANOVA indicou a presença de diferenças estatisticamente significativas entre os grupos, e ao teste de Levene para verificar a homogeneidade.

Diante da limitação de processamento no *Data Analyst GPT*, mesmo ao utilizar a versão paga, ChatGPT Plus (GPT-4o), o principal impacto percebido foi na confiança em sua disponibilidade. A OpenAI exibe uma mensagem³ indicando que o limite de uso será ajustado dinamicamente para priorizar o acesso ao GPT-4o pelo maior número de pessoas, de acordo com a demanda e o desempenho do sistema. Também é mencionado um limite de 40 mensagens a cada 3 horas, levantando dúvidas sobre a disponibilidade do *Data Analyst GPT*. Essa interrupção inesperada e a necessidade de pausar por aproximadamente duas horas antes de retomar as análises destacam uma preocupação significativa: a falta de clareza e transparência em relação às limitações atuais do ChatGPT Plus (GPT-4o) no momento da assinatura, especialmente para aqueles que dependem da ferramenta *Data Analyst GPT* para realizar análises de dados contínuas.

Outro fator limitante a ser considerado ao adotar o *Data Analyst GPT* é a ausência de especificação da versão da ferramenta. Enquanto neste artigo foram utilizadas e conhecidas as versões do JAMOVI 2.3.21.0 e do SPSS 29.0.2.0, a versão exata do *Data Analyst GPT* era desconhecida, sendo a única versão o modelo GPT-4o, uma vez que, como um modelo de inteligência artificial, possui capacidade de aprendizado. Portanto, é importante que a pesquisa realizada com o *Data Analyst GPT* seja acompanhada pelo prompt de execução (European Commission, 2024).

Restaurar e manter a confiança dos usuários requer comunicação clara e abrangente por parte dos desenvolvedores sobre todos os aspectos operacionais, incluindo possíveis limites de

³ Disponível em: <https://help.openai.com/en/articles/7102672-how-can-i-access-gpt-4>. Acesso em: 15 set. 2024.

uso. Tal transparência no momento da assinatura é essencial para garantir que os usuários possam planejar adequadamente seu uso do *Data Analyst GPT*, evitando surpresas desagradáveis e assegurando que as expectativas de disponibilidade estejam alinhadas com a realidade operacional da ferramenta.

Outros aspectos relevantes, destacados por várias partes interessadas, incluem a privacidade, a confidencialidade e os direitos de propriedade intelectual (European Commission, 2019, 2024; UNESCO, 2021). Ao contrário do SPSS e do JAMOVI, nos quais o conjunto de dados é armazenado no software instalado nos computadores dos usuários, o *Data Analyst GPT* opera em um ambiente on-line, e não está claro como esse conjunto de dados é armazenado e como será utilizado, se para treinamento de inteligência ou para inclusão em uma base de conhecimento. Essa falta de clareza compromete o uso do *Data Analyst GPT* em pesquisas com dados confidenciais ou sensíveis.

É crucial destacar variações específicas das tecnologias de inteligência artificial projetadas para garantir a privacidade dos dados. Um exemplo inclui o ChatGPT Teams (Privacidade Empresarial) e os chats temporários (OpenAI, 2024), que são configurados para assegurar que os conjuntos de dados inseridos pelos usuários não sejam utilizados para treinamento de modelos ou inclusão em uma base de conhecimento. Da mesma forma, o Copilot da Microsoft (Universidad de Granada, 2024), uma ferramenta de assistência à codificação impulsionada por IA, segue diretrizes rigorosas para proteger as informações dos usuários, evitando o uso desses conjuntos de dados para aprimorar algoritmos de aprendizado de máquina.

No contexto europeu, recomenda-se que, ao utilizar IA com informações sensíveis ou protegidas, seja fundamental prestar atenção à privacidade, confidencialidade e direitos de propriedade intelectual (European Commission, 2024). Os pesquisadores devem proteger trabalhos não publicados ou sensíveis, evitando o envio desses dados para sistemas de IA on-line sem a garantia de que as informações não serão reutilizadas, como ao treinar futuros modelos de IA ou ao usar indevidamente os dados. Além disso, é importante não fornecer a terceiros dados pessoais sem o consentimento explícito do indivíduo.

Apesar dessas preocupações, há uma vantagem significativa na análise de dados fornecidos pelo *Data Analyst GPT*. Por meio dos *prompts*, pesquisadores e estudantes têm acesso a uma ferramenta de fácil utilização e altamente acessível, facilitando a realização de análises de dados com comandos simples e linguagem direta, além de processar solicitações em linguagem natural. Assim, a ferramenta tem o potencial de se tornar um importante aliado na

promoção da pesquisa, estimulando a inovação e apoiando a descoberta de novos *insights* de maneira intuitiva e acessível. Ademais, ela pode processar grandes volumes de dados sem exigir recursos computacionais avançados por parte do pesquisador, uma vez que os cálculos são realizados nos servidores de inteligência artificial.

Considerações finais

Este artigo demonstrou o potencial do *Data Analyst GPT* em ampliar o horizonte da pesquisa educacional, evidenciando sua confiabilidade na análise de dados quantitativos. Por meio de uma análise comparativa com softwares estatísticos padrão, como SPSS e JAMOVI, foi evidenciado que o *Data Analyst GPT* pode ser empregado de forma confiável como uma ferramenta estatística por pesquisadores e estudantes da educação. Sua interface amigável, que responde a comandos simples e linguagem direta, reduz a necessidade de dominar linguagens de programação complexas ou ter conhecimento técnico aprofundado. Isso representa um avanço significativo para a realização de estudos quantitativos, tornando a análise de dados mais acessível e menos intimidadora para os pesquisadores da educação.

Uma limitação significativa do *Data Analyst GPT* é a falta de especificação clara em relação à versão da ferramenta em uso, uma vez que opera em um ambiente dinâmico onde a versão exata pode não ser explicitamente conhecida, restando apenas o modelo subjacente. Outra limitação importante é a capacidade de processamento do *Data Analyst GPT*, mesmo na versão paga (ChatGPT Plus). A interrupção inesperada e a necessidade de uma pausa de aproximadamente duas horas antes de retomar as análises destacam uma preocupação relevante: a falta de clareza e transparência sobre as limitações de uso, especialmente para aqueles que dependem da ferramenta para análises de dados contínuas. Manter a confiança dos usuários requer uma comunicação clara e abrangente por parte dos desenvolvedores sobre todos os aspectos operacionais, incluindo possíveis limites de uso. Essa transparência é crucial para garantir que os usuários possam planejar adequadamente a utilização do *Data Analyst GPT*, evitando surpresas desagradáveis e alinhando suas expectativas de disponibilidade à realidade operacional da ferramenta. Além disso, o limite de upload de arquivos de 50 MB, embora suficiente para muitos conjuntos de dados quantitativos, pode ser uma restrição em estudos que envolvem conjuntos de dados maiores.

A dependência de uma ferramenta que opera em um ambiente on-line também levanta preocupações sobre a privacidade e a confidencialidade dos dados, uma vez que os detalhes

sobre como as informações são armazenadas e utilizadas não são totalmente transparentes. Os pesquisadores devem ter cautela ao utilizar dados confidenciais ou sensíveis com o *Data Analyst GPT*, particularmente em contextos em que a segurança é crítica.

As limitações específicas deste estudo incluem o número de testes estatísticos realizados e a abordagem de solicitação direta para a análise de dados, especificando os testes desejados. Além disso, a apresentação dos resultados gerados pelo *Data Analyst GPT* pode variar em versões futuras da ferramenta, à medida que melhorias na interface e nos métodos de visualização forem implementadas. No entanto, os resultados estatísticos em si, considerando que o *Data Analyst GPT* utiliza bibliotecas bem estabelecidas, não devem sofrer mudanças significativas, assegurando a replicabilidade e a confiabilidade dos resultados. É importante observar que essa característica, que envolve o uso de bibliotecas, não é exclusiva do *Data Analyst GPT*; softwares como o JAMOVI, que utiliza bibliotecas R, também compartilham essa consistência, embora sem a mesma interface gráfica. Em termos de experiência do usuário e apresentação de dados, outros softwares estatísticos tradicionais, como SPSS e JAMOVI, também estão sujeitos a atualizações que podem impactar esses aspectos.

REFERÊNCIAS

- AL-GHONMEIN, A. M.; AL-MOGHRABI, K. G. The potential of ChatGPT technology in education : advantages , obstacles and future growth. **IAES International Journal of Artificial Intelligence (IJ-AI)**, Jacarta, v. 13, n. 2, p. 1206–1213, 2024. DOI: 10.11591/ijai.v13.i2.pp1206-1213.
- ALGTHAMI, N. M. J.; HUSSIN, N. Meta-Analytic Evidence for Board Characteristics as Correlates of Firm Performance Among Saudi Arabian Businesses. **International Journal of Academic Research in Business and Social Sciences**, Islamabade, v. 12, n. 6, 4 jun. 2022. DOI: 10.6007/IJARBSS/v12-i6/13886.
- ALILI, A.; KRSTEV, D. Using SPSS for research and Data Analysis. **Knowledge International Journal**, Escópia, v. 32, n. 3, p. 363–368, 26 jul. 2019. DOI: 10.35120/kij3203363a.
- ALZUBAIDI, L.; ZHANG, J.; HUMAIDI, A. J.; AL-DUJAILI, A.; DUAN, Y.; AL-SHAMMA, O.; SANTAMARÍA, J.; FADHEL, M. A.; AL-AMIDIE, M.; FARHAN, L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. **Springer International Publishing**, Berlin, v. 8, 2021. DOI: 10.1186/s40537-021-00444-8.
- ATLAS, S. **ChatGPT for Higher Education and Professional Development: A Guide to Conversational AI**. Kingston: University of Rhode Island, 2023. v. 1.

CHANG, D. T. Concept-Oriented Deep Learning with Large Language Models. **ArXiv**, [S. l.], 2023. DOI: 10.48550/arXiv.2306.17089.

CRAWFORD, J.; ALLEN, K.-A.; PANI, B.; COWLING, M. When artificial intelligence substitutes humans in higher education: the cost of loneliness, student success, and retention. **Studies in Higher Education**, London, v. 49, n. 5, p. 1–15, 2024. DOI: 10.1080/03075079.2024.2326956.

DANCEY, C. P.; REIDY, J. **Statistics without maths for psychology**. 8. ed. London: Prentice Hall, 2020.

DING, L.; LI, T.; JIANG, S.; GAPUD, A. Students' perceptions of using ChatGPT in a physics class as a virtual tutor. **International Journal of Educational Technology in Higher Education**, Barcelona, v. 20, n. 1, p. 1–18, 2023. DOI: 10.1186/s41239-023-00434-1.

EHLERS, U.-D.; LINDNER, M.; SOMMER, S.; RAUCH, E. AICOMP - Future Skills in a World Increasingly Shaped By AI. **Ubiquity Proceedings**, London, 2023. DOI: 10.5334/uproc.91.

EUROPEAN COMMISSION. Ethics guidelines for trustworthy AI. **European Commission**, Bruxelas, p. 1–39, 2019.

EUROPEAN COMMISSION. **Proposal for a Regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts**. Brussels: European Commission, 2021.

EUROPEAN COMMISSION. **Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for Educators**. Brussels: European Commission, 2022.

EUROPEAN COMMISSION. **Living guidelines on the responsible use of generative AI in research**. Brussels: European Commission, 2024.

FAN, L.; LI, L.; MA, Z.; LEE, S.; YU, H.; HEMPHILL, L. A Bibliometric Review of Large Language Models Research from 2017 to 2023. **ArXiv**, [S. l.], p. 1–36, 2023. DOI: 10.48550/arXiv.2304.02020.

FIELD, A. **Discovering Statistics Using IBM SPSS Statistics**. 6. ed. London: SAGE Publications, 2024.

GIMPEL, H.; HALL, K.; DECKER, S.; LÄMMERMANN, L.; MÄDCHE, A.; RÖGLINGER, M.; RUINER, C.; SCHOCH, M.; SCHOOP, M.; URBACH, N.; VANDIRK, S. Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education. **Digital Annual Report**, Stuttgart, p. 1–54, 2023.

HUANG, Y.; WU, R.; HE, J.; XIANG, Y. Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: A comparative analysis with SAS, SPSS, and R. **Journal of global health**, New York, v. 14, n. 1088, p. 04070, 2024. DOI: 10.7189/jogh.14.04070.

IBM CORPORATION. **IBM SPSS Statistics Algorithms**. [S. l.: s. n.], 2022.

JIA, X.-H.; TU, J.-C. Towards a New Conceptual Model of AI-Enhanced Learning for College Students: The Roles of Artificial Intelligence Capabilities, General Self-Efficacy, Learning Motivation, and Critical Thinking Awareness. **Systems**, [S. l.], v. 12, n. 3, p. 74, 2024.

LARY, D. J.; ALAVI, A. H.; GANDOMI, A. H.; WALKER, A. L. Machine learning in geosciences and remote sensing. **Geoscience Frontiers**, Beijing, v. 7, n. 1, p. 3–10, 2016. DOI: 10.1016/j.gsf.2015.07.003.

LI, J.; DADA, A.; PULADI, B.; KLEESIEK, J.; EGGER, J. ChatGPT in healthcare: A taxonomy and systematic review. **Computer Methods and Programs in Biomedicine**, Amsterdam, v. 245, p. 108013, 2024. DOI: 10.1016/j.cmpb.2024.108013.

MANN, H. B.; WHITNEY, D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. **The Annals of Mathematical Statistics**, Washington, v. 18, n. 1, p. 50–60, mar. 1947. DOI: 10.1214/aoms/1177730491.

MAREK, J.; MAJ, E.; PRZYBYLA, O. K.; SKRZYNSKI, W.; PASICZ, K.; FABISZEWSKA, E.; PRUSZYNSKI, A.; ROWINSKI, O. The impact of studying on the hippocampal volume in medical students and its correlation with the results of the Final Medical Examination: a single-centre, prospective observational cohort study. **Polish Journal of Radiology**, Warsaw, v. 88, p. 22–30, 16 jan. 2023. DOI: 10.5114/pjr.2023.124433.

MOHAMMADI, S. S.; NGUYEN, Q. D. A User-Friendly Approach for the Diagnosis of Diabetic Retinopathy Using ChatGPT and Automated Machine Learning. **Ophthalmology Science**, New York, v. 4, n. 4, p. 100495, 2024. DOI: 10.1016/j.xops.2024.100495.

MONTENEGRO-RUEDA, M.; LÓPEZ-MENESES, E.; FERNÁNDEZ-CERERO, J.; FERNÁNDEZ-BATANERO, J. M. Impact of the Implementation of ChatGPT in Education: A. **Computers**, Bern, v. 12, n. 153, p. 1–13, 2023. DOI: 10.3390/computers12080153.

NAVEED, H.; KHAN, A. U.; QIU, S.; SAQIB, M.; ANWAR, S.; USMAN, M.; AKHTAR, N.; BARNES, N.; MIAN, A. A Comprehensive Overview of Large Language Models. **ArXiv**, p. 1–43, 12 jul. 2023. DOI: 10.48550/arXiv.2307.06435.

OPENAI. **Enterprise privacy at OpenAI**. Disponível em: <https://openai.com/enterprise-privacy>. Acesso em: 25 mar. 2024.

R CORE TEAM. **A Language and environment for statistical computing**. (Version 4.1) [Computer software], 2021.

SINGH-HARJIT; SINGH-AVNEET. ChatGPT: Systematic Review, Applications, and Agenda for Multidisciplinary Research. **Journal of Chinese Economic and Business Studies**, Washington, v. 21, n. 2, p. 193–212, 2023. DOI: 10.1080/14765284.2023.2210482.

STANFORD UNIVERSITY. **Artificial Intelligence Index Report 2021**. Stanford: Stanford University, 2021.

SUFI, F. Generative Pre-Trained Transformer (GPT) in Research: A Systematic Review on Data Augmentation. **Information**, Bern, v. 15, n. 2, p. 99, 2024. DOI: 10.3390/info15020099.

THE JAMOVI PROJECT. **Jamovi**. (Version 2.3) [Computer Software], 2024.

UNESCO. **Recommendation on the Ethics of Artificial Intelligence**. Paris: UNESCO, 2021.

UNESCO. **ChatGPT and Artificial Intelligence in Higher Education: Quick start guide**. Paris: UNESCO, 2023a.

UNESCO. **Harnessing the Era of Artificial Intelligence in Higher Education: A Primer for Higher Education Stakeholders**. Paris: UNESCO, 2023b.

UNIVERSIDAD DE GRANADA. **Inteligencia Artificial en la universidad: Centro de Producción de Recursos para la Universidad Digital (CEPRUD)**. 2024. Disponível em: <https://ceprud.ugr.es/formacion-tic/inteligencia-artificial>. Acesso em: 25 mar. 2024.

WALTER, Y. Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. **International Journal of Educational Technology in Higher Education**, Dublin, v. 21, n. 1, 2024. DOI: 10.1186/s41239-024-00448-3.

WILCOXON, F. Individual Comparisons by Ranking Methods. **Biometrics Bulletin**, Washington, v. 1, n. 6, p. 80, dez. 1945. DOI: 10.2307/3001968.

CRediT Author Statement

Reconhecimentos: Agradeço ao Professor Pedro Reis por suas valiosas contribuições à conceituação deste artigo.

Financiamento: Este trabalho foi apoiado por Fundos Nacionais através da FCT - Fundação Portuguesa para a Ciência e Tecnologia, IP, no âmbito da Unidade de Investigação e Desenvolvimento em Educação e Formação (UIDEF), UIDB/04107/2020. Mais detalhes podem ser encontrados em <https://doi.org/10.54499/UIDB/04107/2020>.

Conflitos de interesse: Não há interesses concorrentes.

Aprovação ética: Este artigo não requer aprovações éticas.

Disponibilidade de dados e material: Os conjuntos de dados anônimos utilizados e/ou analisados durante o estudo, juntamente com os resultados do Data Analyst GPT, SPSS e JAMOVI, estão disponíveis nas Informações Suplementares.

Contribuições dos autores: Autoria única.

Processamento e editoração: Editora Ibero-Americana de Educação.
Revisão, formatação, normalização e tradução.

