

Assessing the Psychological Impact of Generative AI on Data Science Education: An Exploratory Study

Henrique Ramos Pinto^{1*}, Vitor Meneghetti Ugulino de Araujo², Cleydson de Souza Ferreira Junior³, Lutero Lima Goulart³, Gabriel Silva Aguiar³, João Vitor Cardoso Beltrão³, Paloma Duarte de Lira³, Samuel José Fernandes Mendes³, Filipe de Lima Vaz Monteiro³, Erlon Lacerda Avelino³

¹Department of Center for Education, Federal University of Paraíba, Brazil; ²Department of Informatics, Federal University of Paraíba, Brazil; ³Department of Center for Informatics, Federal University of Paraíba, Brazil

ABSTRACT

The integration of AI in educational settings, particularly through Large Language Models (LLMs), is accelerating, reshaping pedagogical approaches due to students' interactions with new learning tools. This study assesses the impact of generative AI on the educational experiences of computer and data science students at the Center for Informatics, University of Paraiba (CI/UFPB), Brazil. Through Exploratory Factorial Analysis (EFA) of five psychometric scales, the research examines students' acceptance of LLMs, their associated burnout levels, technology anxiety, and the prevalence of metacognitive and dysfunctional learning strategies associated with LLMs. Results indicate significant adoption of AI technologies among students, accompanied by a low incidence of technology anxiety, manifesting as fears of losing jobs to AI. However, a significant correlation was observed between academic burnout and dysfunctional learning strategies in conjunction with LLMs reflects an advanced learning approach, yet challenges with functional learning strategies persist. This study contributes to the discourse on AI in education, highlighting the need for educational frameworks that support effective AI adoption while addressing the psychological demands on students.

Keywords: Artificial Intelligence (AI); AI for education; ChatGPT; Educational technologies; Large Language Models (LLMs); Exploratory Factorial Analysis (EFA); Data science education; Psychological impact of AI; Generative AI; Education; Natural Language Processing (NLP)

INTRODUCTION

In the era of Artificial Intelligence (AI), the academic landscape is undergoing a rapid transformation [1,2]. AI-powered technologies, such as ChatGPT, now provide nuanced responses to a broad range of topics in mere seconds, aiding undergraduate students in their coursework and learning processes [3-5]. Despite these significant benefits, professors express concerns about the responsible use of such technologies. They emphasize the risk of students developing an excessive reliance on these tools, potentially undermining their long-term creative and problem-solving skills [4,6,7]. Additionally, ethical issues such as AI system bias, plagiarism, and lack of transparency need to be considered [1,5]. As AI continues to influence education, universities worldwide are grappling with emerging challenges and opportunities. For instance, the Chinese University of Hong Kong was among the first Chinese institutions to ban the use of ChatGPT, aiming to uphold academic integrity [7]. Students caught using ChatGPT could face penalties ranging from grade reduction to course failure, or even dismissal, on grounds of academic plagiarism and misconduct [7]. In contrast, the Hong Kong University of Science and Technology has embraced the use of ChatGPT and other Large Language Models (LLMs), asserting their responsibility as educators to prepare students for an AI-driven world where tasks can be completed in a timely and cost-effective manner [7]. Popularized by ChatGPT and commonly referred to as generative AI, LLMs are powerful tools with remarkable capabilities for generating human-like text. Beyond major tech giants such as Google and OpenAI, smaller research groups are increasingly training their own LLMs [8]. Stemming from the machine learning branch of AI, these

Correspondence to: Pinto PHR, Department of Center for Education, Federal University of Paraíba, Brazil, E-mail: pedro.clave22@gmail.com Received: 18-Jan-2024, Manuscript No. IJOAT-24-29221; Editor assigned: 22-Jan-2024, PreQC No. IJOAT-24-29221 (PQ); Reviewed: 05-Feb-2024, QC No. IJOAT-24-29221; Revised: 12-Feb-2024, Manuscript No. IJOAT-24-29221 (R); Published: 19-Feb-2024, DOI: DOI: 10.35248/09764860.24.15.272

Citation: Pinto PHR, de Araujo VMU, Ferreira Junior CDS, Goulart LL, Aguiar GS, Beltaro JVC, et al. (2024) Assessing the Psychological Impact of Generative AI on Data Science Education: An Exploratory Study. Int J Adv Technol. 15:272

Copyright: © 2024 Pinto PHR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited..

models require vast amounts of training data, easily sourced from the internet, and are increasingly feasible due to the surge in computational power in recent years. As of 12 PM (GMT-5) on July 18, 2023, there were 15,821 LLMs registered with hugging face, a popular machine learning repository [8]. The expanding number of LLMs encompasses various architectures, settings, training methods, and families, reflecting not only the growing presence of these models but also the emergence of diverse types. Each type comes with unique capabilities and applications, marking a varied and rapidly evolving landscape in AI technology [8,9]. This burgeoning field of generative AI has significant implications for computer and data science education, as highlighted by Tu, et al. [4], necessitating a shift in both curriculum content and pedagogical approaches. The study by these authors underscored that LLMs can execute all stages of data analysis with just a few command prompts, presenting the potential for students to manipulate conventional exam questions, and thus emphasizing the need to adapt assessment practices. Aligning with the findings of Cooper [3]; Chiu, et al. [9]; Rahman, et al. [10], and Zhai, et al. [5], the authors remain optimistic about integrating AI into the educational landscape, despite the challenges. Among the foremost advantages of LLMs in education is their ability to provide personalized learning experiences. By analyzing students' responses and learning patterns, LLMs can customise educational content to individual needs, catering to diverse learning styles and abilities. This approach not only enhances student engagement but also improves learning outcomes.

In fact, LLMs are now employed in a range of applications, including virtual assistants, customer service, content creation, and researchers are also benefiting from their support in academic pursuits [1,10,11]. Several prestigious publishers, including Taylor and Francis, Nature, and Elsevier, have revised their authorship policies to accommodate this new research paradigm. As standard practice, these publishers disallow listing LLMs like ChatGPT as authors, emphasizing a "human-centric" approach by detailing the use of AI technologies in the methods section without granting them co-authorship [1]. Furthermore, researchers remain responsible for the integrity of their academic publications.

State-of-the-art

As a leading-edge educational tool, AI chatbots offer the unique advantage of scalable, individualized tutoring, providing support that autonomously adapts to the learning pace and style of each student [12]. Expanding on this concept, Yin et al. [13] investigated a micro-learning chatbot environment compared to a traditional classroom setting with 91 students over a brief 40-minute intervention. Although the average post-test scores were similar, chatbot users surpassed their peers by half a standard deviation. In a related study, Essel, et al. [14] provided 68 undergraduates with access to an AI assistant, KNUSTbot, and these students also outperformed the control group by half a standard deviation. Corroborating these findings, a systematic review by Okonkwo, et al. [2] emphasized that AI chatbots, akin to KNUSTbot, have become a mainstay in online education and serve as an effective technological tool to boost students' learning engagement. However, it's important to distinguish between pre-transformer

chatbots and LLMs. The advent of transformer architecture has profoundly reshaped the Natural Language Processing (NLP) field in a relatively short period [15], and currently, ChatGPT is hailed as the state-of-the-art in conversational technologies [5,9]. These AI models have evolved to the degree that they can proficiently tackle standard assessments in law schools [6] and even devise intricate solutions for programming paradigms [4]. Moreover, Cooper [3] underscores the benefits of LLMs for educators and institutions, emphasizing its prowess in personalizing coursework, tailoring assessments, and meticulously crafting quizzes and science units. Regrettably, while there is an abundance of reviews and opinion pieces on ChatGPT, there is a dearth of experimental studies evaluating student performance using LLMs as a one-to-one tutoring method. This gap in empirical research may be attributed to the novelty of the subject, suggesting that it might be premature for comprehensive, controlled experiments to have been conducted and published. However, in an encouraging development, Urban et al. [16] recently released a preprint detailing their research, which involved experimental and control groups focusing on creative problem-solving performance among university students. Their findings suggest that students utilizing ChatGPT demonstrated a capacity to formulate solutions that were more innovative, detailed, and closely aligned with task objectives compared to those not using such advanced tools. This study represents a preliminary but significant step in understanding the potential benefits of ChatGPT and similar LLMs in enhancing student learning outcomes.

Delving deeper, two surveys have provided valuable insights into the current state of research on the psychological impact of ChatGPT in educational settings. In their study, Siregar; Hasmayni et al. [17] highlighted the significant positive effect of ChatGPT on students' learning motivation. Leveraging validated psychometric scales for their analysis, they found that approximately 57.3% of the variance in student motivation could be attributed to the use of ChatGPT. Separately, Sallam et al. [18] introduced a TAMbased survey instrument, the TAME-ChatGPT (Technology Acceptance Model Edited for ChatGPT Adoption), specifically designed to assess the successful integration and application of this technology in healthcare education. Their primary goal was to develop and validate an appropriate psychometric scale for measuring the usage and integration of ChatGPT, facilitating subsequent studies on this construct and related variables such as student anxiety, perceived risk, and behavioral/cognitive factors.

When inferring psychological variables, it is important to utilize validated psychometric scales, as such variables cannot be directly observed [19]. Therefore, the credibility of a psychological instrument is intrinsically linked to the rigor of its validation process, and a critical aspect of this process is factorial analysis [20]. This statistical method evaluates the common variance among items on a psychometric scale to ensure they measure the same underlying construct [21]. Considering the nascent stage of research on ChatGPT, especially in its application within educational contexts, there is a noticeable gap in the psychometric assessment of its impact. In response to this knowledge deficit, an exploratory study was embarked upon.

OPEN OACCESS Freely available online

MATERIALS AND METHODS

Objectives

The primary objective of this research is to explore five psychological constructs potentially related to the impact of generative AI on computer and data science education. To address this, psychometric scales specifically designed to measure those constructs were developed, encompassing AI acceptance, motivation to learn, technology anxiety related to AI, academic burnout, and metacognitive and dysfunctional learning strategies when studying with Large Language Models (LLMs). Consequently, a fundamental part of the research objectives includes validating these customised-made psychological instruments. This validation process is critical to ensure that the scales accurately capture the intended constructs and are capable of providing meaningful information for understanding and improving the educational process in the context of emerging AI technologies.

Hypothesis being tested

- H1: Al-driven technologies are fundamentally reshaping educational practices and attitudes, especially in technologically advanced fields.
- H2: High levels of academic burnout are related to increased technology anxiety and the adoption of dysfunctional learning strategies.
- H3: Effective use of LLMs, combined with metacognitive strategies, significantly improves students' learning and motivation.

Procedures and participants

The study was conceptualized as a probabilistic samplingbased survey, leveraging opinion-driven questionnaires. Data collection was conducted exclusively at the Center for Informatics, University of Paraíba (CI/UFPB), in Brazil, using Google Forms. This platform automatically notified participants of any missing values, ensuring the completeness and accuracy of each submission. To maximize participation and reach, a diverse dissemination strategy was employed. This strategy included using WhatsApp for communication within CI/UFPB academic groups; distributing informative leaflets with QR codes for straightforward access to the online form; and facilitating educational discussions in classroom environments led by faculty members and students. This effort yielded a substantial dataset, with 178 respondents: 143 males, accounting for 80.3% of the total, and 35 females, comprising 19.7%. Designed as a non-interventional study, the procedures adhered strictly to the highest ethical standards in research involving human subjects. Sociocultural values, participant autonomy, and anonymity were respected throughout the data collection process. All potential risks were carefully measured and mitigated. The deployment of the Informed Consent Statement was integral to this process, ensuring participants' complete awareness of the research's nature and their rights. To maintain data integrity and reduce potential response bias, the psychometric scales were presented to participants in a randomized sequence. This approach is effective in minimizing the influence of question order on responses, thereby providing a

more balanced and objective view of the results.

Instruments

Five psychometric scales were made, each featuring items rated on a 7-point Likert scale to capture the nuances of participants' responses. The scales ranged from "strongly disagree" to "strongly agree" and from "never" to "always". Each scale contributes to a comprehensive understanding of the multifaceted relationship between students' experiences and generative AI technologies. The data derived from these instruments are poised to provide a rich foundation for analysis, aiming to elucidate the psychological and educational dynamics at play. Below is a description of each scale employed in the study, outlining their respective domains and the constructs they are intended to measure:

Academic Burnout Model, 4 items (ABM-4): Initially designed to assess work-related burnout [22], this scale measures the extent of specific aspects students might feel as a result of prolonged engagement with intense academic activities. These aspects include study-demand exhaustion, the emotional impact of academic pressures, and the depletion experienced from academic endeavors.

AI Technology Anxiety Scale, 3 items (AITA-3): This scale measures the anxiety students may feel when interacting with generative AI technologies, including fears of job displacement or subject matter obsolescence. The scale was adapted from Wilson, et al. which defined technology anxiety as "the tension from the anticipation of a negative outcome related to the use of technology deriving from experiential, behavioral, and physiological elements".

Intrinsic Motivation Scale, 3 items (IMOV-3): Designed to assess the level of students' intrinsic motivation towards learning, this scale, adapted from Siregar et al. [17], evaluates the inherent satisfaction and interest in the learning process itself. This scale is particularly relevant given the literature suggesting a significant relationship between the use of AI technologies and students' motivation to learn.

Learning Strategies Scale with Large Language Models, 6 items (LS/LLMs-6): This scale encompasses two distinct dimensions involving LLMs, one focusing on Dysfunctional Learning Strategies (DLS/LLMs-3) and the other on Metacognitive Learning Strategies (MLS/LLMs-3), each consisting of three items. The DLS/LLMs-3 sub-scale investigates potential counterproductive learning strategies that students might adopt when using LLMs, which could impede effective learning, and includes both unique and conventional items of dysfunctional strategies [23]. In contrast, the MLS/LLMs-3 sub-scale assesses the self-regulatory practices that students employ while learning with LLMs, aimed at enhancing learning outcomes. This subscale is specifically customised to include items directly related to LLMs. Based on Oliveira et al. [24]; Pereira et al. [25], these scales provide a comprehensive assessment of learning strategies in the context of LLMs, covering both the metacognitive techniques that enhance learning and the dysfunctional methods that may potentially hinder it.

LLMs Acceptance Model Scale, 5 items (TAME/LLMs-5): This scale evaluates students' readiness to integrate Large Language

Models (LLMs) into their learning process. It has been adapted from Sallam, et al. [18] to provide a more streamlined approach for assessing AI technology acceptance and intention to use.

Adaptation review and validity

All scales used in the study were specifically adapted to align with the unique characteristics of the participants and the research context, which focused on the impact of generative AI on computer and data science education at CI/UFPB. This institution is renowned for its technologically proficient student body, many of whom are familiar with or actively engaged in technology and programming. Adapting these tools was crucial to accurately capture the nuances of an environment where students are both learning about and working with AI technology, ensuring relevance and sensitivity to the depth of their responses. Moreover, to enhance participant engagement and ensure response accuracy, the study employed a concise questionnaire design. This brevity was vital to prevent participant fatigue and maintain their interest, enabling the collection of accurate and meaningful data reflective of their experiences and perceptions in a rapidly evolving educational context. The study focused on using established tools with a smaller set of items, ideal for foundational research. By distilling these scales to their core elements, we ensured their psychometric soundness while making them more user-friendly and less burdensome for the respondents.

Data Analysis

The collected data underwent a meticulous statistical examination, with Exploratory Factorial Analysis (EFA) serving as a pivotal technique for dimensionality reduction. This approach was important in simplifying the complex data structure and unveiling the fundamental dimensions within the observed variables, thereby validating the psychological properties of the scales. For descriptive statistics, both box plots and violin plots were utilized to provide visual representations of data distribution and variance. Additionally, Spearman correlation analysis, particularly suited for the data's non-parametric nature, was employed to comprehend the relationships between variables. These combined methods offered a comprehensive understanding of the data's characteristics and interrelations. Python and R were utilized for the analysis in this study, with the study's code and dataset available on GitHub. See Appendix A for a summary of the key libraries used and their applications.

EFA decision processes

Factorial analysis is not a singular technique but rather a group of associated methods that should be considered and applied in concert [19]. The objectives of EFA are multifaceted and include the reduction of variables to a smaller number of factors, assessment of multicollinearity, development of theoretical constructs, and testing of proposed theories [21]. The sequential and linear approach to EFA demands careful consideration of various methodological steps to ensure the validity and reliability of the results. The decisions made throughout this process are detailed below:

Sample adequacy: Prior to factor extraction, it's imperative to evaluate whether the data set is suitable for factor analysis. To

this end, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure were employed [19,21]. Bartlett's test is used to test the hypothesis that the correlation matrix is not an identity matrix, essentially assessing whether the variables are interrelated and suitable for structure detection. A significant result from Bartlett's test allows for the rejection of the null hypothesis, indicating the factorability of our data [21]. On the other hand, the KMO measure evaluates the proportion of variance among variables that could be attributed to common variance. With the KMO index ranging from 0 to 1, values above 0.5 are considered suitable for factor analysis [19,21].

Factor retention: Determining the optimal number of factors to retain is an important aspect of EFA, as it defines the dimensionality of the constructs. In this study, three distinct methods were adopted to identify the optimal number of factors: The Kaiser-Guttman criterion [21], parallel analysis [26], and the factor forest approach, which involves a pre-trained machine learning model [27]. Factor extraction plays a significant role in simplifying data complexity and revealing the dataset's underlying structure, thereby ensuring the dimensions of the constructs measured are captured accurately, and the analysis truly reflects the data's nature.

Internal reliability assessment: A critical component of EFA is the evaluation of the internal reliability of the scales. Reliability assessment refers to the process of examining how consistently a scale measures a construct. Ensuring consistent measurement is pivotal, as it confirms that any observed variations in data accurately reflect differences in the underlying construct, rather than resulting from measurement error or inconsistencies [28]. For this purpose, we employed two key metrics: McDonald's omega (ω) and Cronbach's alpha (α). The omega metric provides an estimate of the scales' internal consistency, presenting a robust alternative to the traditionally utilized Cronbach's alpha [28].

Factor extraction: A fundamental next step involves selecting an appropriate method for factor extraction, which dictates how factors are derived from the data. In this research, Principal Axis Factoring (PAF) was employed. Unlike methods that require multivariate normality, PAF is adept at handling data that may not fully meet these criteria, making it a suitable choice in exploratory contexts, especially with smaller sample sizes [20,21]. PAF's capability to uncover latent constructs within the data without imposing stringent distributional assumptions aligns well with the exploratory nature of the survey [20].

Rotation method: The final step in factorial analysis often involves choosing a rotation method to achieve a theoretically coherent and interpretable factor solution. In this study, promax rotation, a widely used method for oblique rotation, was selected [29]. The rationale for using an oblique rotation like promax lies in its suitability for scenarios where factors are presumed to be correlated. Unlike orthogonal rotations, which assume factors are independent, oblique rotations acknowledge and accommodate the possibility of inter-factor correlations [19,29].

Use of AI tools

The ChatGPT-4 model played a substantial role in this project, being utilized not only for correcting grammar but also for refining paragraphs, and assisting with coding in data analysis.

Additionally, a specialized model, SciChat, was developed, specifically designed to assist in enhancing the writing process for this research. This model was custom-designed to respond to queries related to scientific writing, providing more focused and effective support. The diverse application of these AI tools ensured the maintenance of high scientific accuracy and integrity throughout the project, with every output rigorously checked for precision and reliability.

RESULTS

LLMs usage preferences among students

In the assessment of LLMs' preference at the CI/UFPB, the data revealed a predominant usage of ChatGPT 3.5, with a staggering 92.7% of the respondents utilizing this free version. ChatGPT 4, despite being a paid version, is used by 5.6% of the participants, showcasing a willingness to invest in more advanced AI tools. Bing Chat, another LLM, is used by 23% of the students, indicating a diversity in the AI platforms engaged by the students for their educational pursuits. A smaller, yet significant, fraction of students have adopted Bard, comprising 18% of the users, and only a minority, 4.5%, reported not using any LLMs at all, which underscores the widespread penetration of these technologies in the academic environment.

Validity assessment through factorial analysis

The results of the EFA revealed diverse psychometric properties across the scales. Primarily, both Bartlett's test and the KMO measure affirmed the suitability of the data for factor analysis. As indicated in Table 1, the KMO values for all scales exceeded the benchmark of 0.5, suggesting the sample adequacy for each scale [21]. Moreover, the significance of Bartlett's tests across the board (p-value < 0.05) rejected the null hypothesis that our correlation matrix is an identity matrix, underscoring the data's aptness for structure detection [19] (Table 1).

Table 1: Sample adequacy

Scales	KMO Test	Bartlett's test of sphericity	
ABM-4	0.742	0	
AITA-3	0.562	0	
IMOV-3	0.581	0	
LS/LLMs-6	0.682	0	
TAME/LLMs-5	0.689	0	

As detailed in Table 2, the factorial retention procedure further elucidates the dimensionality of the scales, reinforcing the diverse psychometric properties observed during EFA. Remarkably, the ABM-4, AITA-3, and IMOV-3 scales each indicated a singular factorial structure, as evidenced by uniform retention values across the chosen methods: Kaiser criterion, parallel analysis, and factor forest algorithm. Conversely, the LS/LLMs-6 and TAME/LLMs-5 scales exhibited a more complex structure. The LS/LLMs-6 scale steadily retained two factors across all three methods, indicating a clear factorial structure, and the TAME/LLMs-5 scale presented a variation between one and two factors. Hence, the decision to retain only one factor for this scale was a strategic response to the

inherent limitations of parallel analysis in processing categorical and rank-ordered data. Recognizing these technical limitations, the study shifted its focus to the factor forest method [27] and traditional Kaiser's criterion [21] (Table 2).

Table 2: Factorial retention

Scales	Kaiser criterion	Parallel analysis	Factor forest	
ABM-4	1	1	1	
AITA-3	1	1	1	
IMOV-3	1	1	1	
LS/LLMs-6	2	2	2	
TAME/LLMs-5	1	2	1	

After evaluating sample adequacy and determining the number of factors to retain, the study progressed to assess the internal reliability of the scales. This evaluation, detailed in Table 3, involved an analysis of Cronbach's alpha (α), McDonald's omega (ω), and the cumulative variance explained for each scale [19,28]. Notably, the ABM-4 and LS/LLMs-6 scales displayed strong reliability, with both ω values surpassing the 0.7 benchmark, suggesting high consistency. The TAME/LLMs-5 scale showed a well-balanced reliability profile, evidenced by closely matched α and ω values. In contrast, the AITA-3 and IMOV-3 scales, while still reliable, recorded slightly lower scores, indicating moderate consistency. All the scales exhibited acceptable levels of cumulative variance explained. This measure indicates the proportion of total variance in the observed variables that is accounted for by the factors. Higher values suggest that the factors extracted during the EFA process are effectively capturing the underlying structure of the dataset [20]. In the context of this study, the cumulative variance explained by each scale, although varying, was within an acceptable range. This suggests that the scales are adequately capturing the constructs they are intended to measure, thereby supporting their validity. For instance, even though the TAME/LLMs-5 scale had the lowest cumulative variance explained (0.426), it still provided a significant portion of the variance, contributing to a meaningful understanding of the empirical data (Table 3).

Table 3: Internal reliability assessment

Scales	Kaiser criterion	Parallel analysis	Factor forest	
ABM-4	0.702	0.702	0.528	
AITA-3	0.619	0.664	0.579	
IMOV-3	0.578	0.625	0.545	
LS/LLMs-6	0.64	0.739	0.618	
TAME/LLMs-5	0.656	0.663	0.426	

At last, the cornerstone of EFA is the factorial extraction procedure. The factorial loadings in EFA are critical as they represent the strength and direction of the relationship between observed variables (questionnaire's items) and underlying latent factors [19,21]. Essentially, these loadings measure how much variance in an item is explained by the factor, providing insights

into how well each variable aligns with a particular factor. Analyzing the EFA loadings presented in Table 4, it's evident that most items demonstrate strong correlations with their respective factors, indicating a coherent pattern of associations across the scales, and the effectiveness of the factorial analysis in simplifying the data's complexity into a small number of factors. It's important to note that while most scales appeared to be unidimensional, indicating a single underlying construct, the LS/LLMs-6 scale was an exception. This scale exhibited a bidimensional structure, with Factor 1 representing metacognitive learning strategies (MLS/LLMs-3 sub-scale) and Factor 2 representing dysfunctional learning strategies (DLS/LLMs-3 sub-scale). This bifurcation in the LS/LLMs-6 scale implies the presence of two distinct constructs within the scale: One related to effective learning strategies and another related to counterproductive strategies. This differentiation is essential for understanding the subsequent discussion and results' portrayal (Table 4).

Table 4:	Factorial	loadings
----------	-----------	----------

Scales	Kaiser criterion	Parallel analysis	Factor forest
	Item 1	0.7	-
	Item 2	0.74	-
ABM-4	Item 3	0.72	-
	Item 4	0.74	-
	Item 1	0.53	-
AITA-3	Item 2	0.85	-
	Item 3	0.86	-
	Item 1	0.83	0
	Item 2	0.85	0.05
	Item 3	0.86	0.04
LS/ LLMIS-0	Item 4	0.11	0.68
	Item 5	0.02	0.71
	Item 6	0.09	0.76
	Item 1	0.72	-
IMOV-3	Item 2	0.82	-
	Item 3	0.67	-
	Item 1	0.69	-
	Item 2	0.69	-
TAME/LLMs-5	Item 3	0.63	-
	Item 4	0.74	-
	Item 5	0.47	-

Psychometric scales' descriptive statistics

The exploration into the scales' characteristics will be visualized through the strategic use of box and violin plots, which will illustrate the core tendencies and variations within the students' data. Box plots are designed to highlight central measures—the mean (indicated by a white dot) and median (depicted as a red line) along with the spread of responses. This spread is represented by quartiles, which divide the data into four equal parts. The interquartile range, marking the distance between the first and third quartiles, encompasses the central 50% of the data, shown as a blue box. In parallel, violin plots will provide insights into the data's density and distribution, offering a more nuanced interpretation of variability and frequency across different values. This dual approach allows for a comprehensive understanding of the dataset, from central tendencies to the diversity of responses.

Beginning with an assessment of students' mental health, the ABM-4 scale uncovers significant patterns in students' exhaustion, stress and depletion (Figure 1). The mean and median, positioned above the midpoint scale of 16, suggest a high level of academic burnout among the students. The box plot displays a wide interquartile range, indicating a diverse dispersion in the severity of students' experiences. Additionally, the violin plot's density is prominently stretched, reinforcing the spectrum of responses, from the median to more extreme reports. In contrast, the AITA-3 scale, which focuses on technology anxiety related to AI, presents a different data distribution (Figure 2). The box plot demonstrates that the interquartile range is below the mid-scale value of 12, indicating a trend toward lower anxiety levels within the group. This observation is further supported by the violin plot, where the data concentration is heaviest at the lower end of the scale. Results indicates that 75% of students' anxiety levels are comfortably below the midpoint, and none have reached the maximum level of technology anxiety.



Figure 1: Academic Burnout Model, 4 items (ABM-4)



Regarding learning strategies, the analysis of the MLS/LLMs-3 sub-scale through its descriptive statistics unveils a median that is marginally above the midpoint, indicating a propensity for higher engagement with metacognitive strategies (Figure 3). Nevertheless, the wide interquartile range indicates a variety in students' responses. The corresponding violin plot supports this conclusion, displaying a response density that decreases slowly as one moves away from the median. Furthermore, the DLS/LLMs-3 visualizations shed light on another dimension of student learning strategies (Figure 4). When comparing these sub-scales, it appears that students are generally more consistent in their use of metacognitive strategies than in avoiding dysfunctional ones, which seem to be more scattered. This is reinforced by the violin plot's bimodal peaks, which imply two main clusters of responses among the participants.

OPEN OACCESS Freely available online



Figure 4: Dysfunctional Learning Strategies (DLS/LLMs-3)

Delving deeper, IMOV-3 provides an intriguing overview of the distribution of students' inherent enthusiasm for learning (Figure 5). The median value of 15, with the entire interquartile range situated above the midpoint of 12, reflects a collective tendency toward a more motivated approach to learning. An outlier, represented as a solitary dot, hints at an exceptional case where a student's motivation significantly diverges from the norm. This data leads us to a plausible conclusion that students demonstrate strong motivation for learning, evidenced by 75% of them scoring beyond the midpoint threshold. Lastly, TAME/LLMs-6 provides insights into students' perceptions and acceptance of LLMs (Figure 6). The box plot reveals a small interquartile range positioned above the scale's midpoint, indicating a cohesive attitude toward LLMs among the respondents. This compact range suggests a consensus in acceptance levels, with only a few outliers indicating some reservations. Mirroring this, the violin plot's expanded middle section reflects a majority consensus, which narrows at both ends to represent fewer students with extreme viewpoints, be they highly skeptical or exceptionally receptive to LLMs (Table 5 and 6).





	ABM-4	AITA-3	IMOV-3	MLS/ LLMs-3	DLS/ LLMs-3	TAME/ LLMs-5
ABM-4	1	0.27	-0.14	0.05	0.41	0.16
AITA-3	0.27	1	0	0.09	0.34	0.16
IMOV-3	-0.14	0	1	-0.03	-0.31	-0.04
MLS/ LLMs-3	0.05	0.09	-0.03	1	0.11	0.6
DLS/ LLMs-3	0.41	0.34	-0.31	0.11	1	0.13
TAME/ LLMs-5	0.16	0.16	-0.04	0.6	0.13	1

Table 6: Significance matrix (p-values)

Table 5: Spearman's Correlation Matrix (p)

	ABM-4	AITA-3	IMOV-3	MLS/ LLMs-3	DLS/ LLMs-3	TAME/ LLMs-5
ABM-4	-	0	0.06	0.46	0	0.04
AITA-3	0	-	0.99	0.23	0	0.03
IMOV-3	0.06	0.99	-	0.64	0	0.06
MLS/LLMs-3	0.46	0.23	0.64	-	0.13	0
DLS/LLMs-3	0	0	0	0.13	-	0.07
TAME/ LLMs-5	0.04	0.03	0.6	0	0.07	-

DISCUSSION

An overwhelming acceptance of LLMs

The findings unveil significant psychological and behavioral patterns among data science students, especially their overwhelming acceptance of LLMs like ChatGPT and Bard. This trend reflects a forward-thinking approach to incorporating AI technologies into their academic toolkit. With only eight students reporting no use of LLMs, and considering the descriptive results from the TAME/LLMs-6 scale, the data underscores a pervasive, technology-oriented ethos at CI/UFPB. As documented in recent studies [1,3-6], the influence of LLMs is reshaping educational practices across institutions worldwide. These AI technologies are not merely transient tools but are becoming integral to the future of teaching and learning, with their impact evolving more rapidly in fields like data science, which are inherently connected to technological advancements.

The role of AI in computer and data science education

This technology-friendly environment likely contributes to the notably low levels of AI-related anxiety, as seen in the descriptive results of the AITA-3 scale. The students' regular interactions with advanced technological tools seem to buffer them from the typical apprehensions concerning new technological integrations. Rather than viewing AI as a threat to their skills or future job prospects, they appear to recognize its potential to enhance their capabilities and autonomy (H1 plausible).

However, a minor positive correlation was observed between the acceptance of LLMs and technology anxiety (ρ =0.16, p-value=0.03). This finding somewhat diverges from Wilson, et al. [23], which suggested that anxiety regarding technology typically has a negative correlation with the acceptance, usage, and integration of such tools. Nevertheless, it is essential to distinguish that the ATAS scale from Wilson, et al. [23] is concerned with general technology anxiety and not crafted for the evolving AI context, whereas the AITA-3 is dedicated to exploring the societal issues triggered by those technologies. This distinction implies that for students regularly using LLMs, the perceived effectiveness of AI might paradoxically induce more anxiety about its societal integration, highlighting a nuanced relationship between familiarity with AI and perceptions of its broader implications. The concern about the displacement of programming jobs by AI mirrors a general expectation of profound changes across various sectors. This anticipated shift accent the critical need for strategic upskilling in education, encouraging programmers to expand their expertise beyond the conventional pipeline. Now, "students need to learn to view themselves as product managers rather than software engineers", which not only prepares them for the evolving demands of the job market but also positions them to navigate the future of work with agility and foresight. It is known that the emergence of digital technologies, such as calculators, smartphones, and GPS systems, has profoundly impacted human cognition [1]. Similarly, AI is poised to bring about significant psychological changes, but the specifics of these changes remain largely unknown. These evolving cognitive landscapes, influenced by AI's unique interactions and capabilities, underscore the need for new forms of literacy and adaptability in the 21st century, which goes beyond traditional digital navigation skills. A crucial aspect of effective AI interaction is the skill to craft precise prompts, a capability that varies among individuals; Some find it easier to formulate than others [1,30]. As AI becomes increasingly integral in various aspects of life, the skill of prompt formulation should be recognized and developed with the same emphasis as the overall digital literacy.

Metacognition in modern learning environments

The integration of LLMs with metacognitive strategies, which refer to the conscious control over cognitive processes involved in learning such as organizing, prioritizing, and actively monitoring one's comprehension and progress, indicates a sophisticated approach to learning [24,25]. According to the MLS/LLMs-6 descriptive results, students are not merely relying on LLMs; rather, they are thoughtfully incorporating them into their study habits, utilizing their capabilities to enhance understanding and refine problem-solving skills. This strategic application likely contributes to the favorable perception of LLMs, as demonstrated by the strong positive correlation between the TAME/LLMs-6 and the MLS/LLMs-3 (ρ =0.60, p<0.01). This indicates that increased acceptance of AI corresponds with heightened metacognitive engagement, suggesting that students

are employing these technologies in a purposeful and efficient manner in their learning habits, a conclusion supported by their high motivation levels as reflected in IMOV-3 descriptive results (H3 plausible).

The meta-analysis by Theobald [31] highlights the intricate relationship between various factors in academic settings, emphasizing the positive impacts of cooperative learning on cognitive and metacognitive strategies. The analysis suggests that programs centered around feedback more effectively enhance metacognitive skills, resource management, and motivation. Notably, programs grounded in a metacognitive theoretical framework achieve greater success in academic achievement compared to those that focus solely on cognitive aspects-copying, memorizing, reading, summarizing etc. This insight is particularly relevant in the context of AI educational technologies, such as chatbot-based learning environments, which excel in offering personalized, immediate feedback. Studies by Chiu, et al. [9]; Urban, et al. [16] and Yin, et al. [13] have demonstrated the effectiveness of AI technologies in these areas. They argue that such technologies, by providing personalized feedback, can substantially aid students' development.

In contrast, the analysis of the DLS/LLMs-3 sub-scale reveals that despite general confidence in using LLMs, students recognize certain challenges. Difficulties in identifying inaccuracies in outputs from LLMs emphasize the complexities of relying on AI for learning. This complexity is further elucidated by the moderate negative correlation between the IMOV-3 and the DLS/LLMs-6 (p=-0.31, p<0.01), indicating that students employing more dysfunctional learning strategies tend to be less intrinsically motivated. The DLS/LLMs-3 also reflects conventional dysfunctional strategies, which include lack of self-regulation and subject understanding. This reinforces the necessity of teachers' support in optimizing students' use of AI for educational purposes, corroborating Chiu, et al. [23] findings which indicate that both student expertise and teacher assistance are crucial for effectively fostering learning competence with AIbased chatbots.

Metacognitive strategies encompass the processes of planning, monitoring, and evaluating one's understanding and learning. These higher-order cognitive processes involve self-regulation and control over learning activities [24,25]. The dichotomy between LLMs' usage in metacognitive versus dysfunctional strategies highlights the necessity for a balanced integration of LLMs in educational settings. It underscores the importance of equipping students with the skills to critically evaluate and effectively employ LLMs outputs, enabling them to discern and rectify errors, thus optimizing their learning journey and outcomes. As suggested by recent studies [3,4,5,16] strategic LLMs use, supported by a thorough understanding of their limitations, can empower students to navigate potential pitfalls and maximize the benefits of these advanced tools.

LLMs and mental health among students

The high scores on academic burnout highlight a critical aspect of student life, emphasizing how the pressures for academic achievement and the competitive nature of academia significantly contribute to elevated stress levels and impact students' overall educational experiences. An important aspect to consider is the interaction between this widespread stress and students' use of LLMs. Research findings reveal a positive and moderate correlation between ABM-4 and DLS/LLMs-3 (ρ =0.41, p-value=0.00). This correlation suggests that higher levels of academic burnout are associated with an increased reliance on ineffective learning strategies involving LLMs. Additionally, a small but significant correlation exists between ABM-4 and AITA-3 (ρ =0.27, p-value=0.00), indicating a relationship between academic stressors and students' apprehensions regarding AI's role in society (H2 plausible).

In the context of the scholarly consensus highlighted by Mofatteh [32], which elucidates the impact of psychological and academic variables on stress and anxiety levels in university students, the correlations identified in the referenced study acquire considerable significance. These correlations, when placed within a broader context, contribute to a nuanced understanding of student experiences. Recognized factors such as low self-esteem, personality traits like high neuroticism and low extraversion, and feelings of loneliness are known to increase susceptibility to stress, anxiety, and depression during university years. The findings from the study in question add an additional layer to this complex dynamic, which is particularly relevant in the current educational landscape where technology and digital tools have become increasingly integral to the learning process.

Limitations and prospects for future research

This study's insights must be contextualized within the scope of its methodological constraints. The relatively small sample size may limit the generalizability of our findings to broader populations. Additionally, the brevity of the psychometric scales, while necessary to cover a range of constructs without overburdening respondents, may yield less granularity compared to more extensive conventional scales. Such constraints could potentially affect the accuracy and depth of our insights into the multifaceted impacts of generative AI on education. Moreover the scales were not reviewed by professionals specializing in the constructs being measured, which may have provided further validation of the instruments used. Future research endeavors could aim to mitigate these limitations by employing larger and more diverse samples to enhance the representativeness of the results. Further refinement and validation of the scales by subject matter experts would bolster the reliability of the measurements. An exploration of correlations between scale results and sociodemographic variables could yield rich insights; for instance, comparing the metacognitive and dysfunctional learning strategies of newer students with those more advanced in their studies could reveal how adaptability and coping mechanisms evolve through the university experience. Such investigations could offer valuable information on the developmental trajectory of learning strategies in relation to the integration of AI in academic settings.

CONCLUSION

In summary, the study explains on the complex dynamics between students' engagement with LLMs and their psychological and

academic well-being. The findings reveal a dichotomy: On one side, there is a discernible trend of acceptance and strategic utilization of LLMs among data science students, indicating a positive shift towards the integration of AI in educational paradigms. On the other side, the study also unveils a nuanced interplay between the use of these advanced tools and academic stressors. The identified correlations between academic burnout, dysfunctional learning strategies, and AI-related anxiety highlight the necessity for educational institutions to cultivate not just students' outcomes but also a supportive environment for student development.

AUTHOR CONTRIBUTIONS

Conceptualization: Lira PD, Araújo VMU, Beltrão JVC, Aguiar G S, Ferreira Junior C S, Avelino EL and Mendes SJF; methodology: Lira PD, Araújo VMU, Ramos PHR, Ferreira Junior CS, and Mendes SJF; validation; Ramos PHR; formal analysis; Ramos PHR; Investigation: Lira PD; Mendes SJF, Monteiro FLV; Aguiar GS, Ramos PHR and Beltrão JVC. Resources: Ramos PHR, Monteiro FLV and Goulart LL; Data curation: Ramos PHR, Goulart LL; Writing original draft preparation: Ramos PHR; writing review and editing: Araújo VMU and Ferreira Junior CS; visualization: Ramos PHR; supervision: Araújo VMU; project administration: Araújo VMU; All authors have read and agreed to the published version of the manuscript.

INSTITUTIONAL REVIEW

Due to the nature of the experiment being a low-risk survey utilizing opinion questionnaires, authorization from the Ethics Committee was not required, in accordance with Resolution 510/2016 of the Brazilian National Health Council.

INFORMED CONSENT

Informed consent was obtained from all participants.

DATA AVAILABILITY

Empirical data used for analysis is available on GitHub. \href{https://github.com/pintophr/Assessing-the-Psychological-Impact-of-Generative-AI-on-Data-Science-Education.git}

CONFLICTS OF INTEREST

The authors declare no conflict of interests.

REFERENCES

- 1. Malik T, Dwivedi Y, Kshetri N, Hughes L, Slade EL, Jeyaraj A, et al. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inform Manage. 2023;71:102642.
- 2. Okonkwo CW, Ade-Ibijola A. Chatbots applications in education: A systematic review. Comp Edu Artif Intell. 2021;2:100033.
- 3. Cooper G. Examining science education in chatgpt: An exploratory study of generative artificial intelligence. J Sci Educ Technol. 2023;32(3):444452.

- 4. Tu X, Zou J, Su WJ, Zhang L. What should data science education do with large language models. arXiv preprint arXiv:2307.02792. 2023.
- 5. Zhai X. ChatGPT user experience: Implications for education. Available at SSRN 4312418. 2022.
- 6. Choi JH, Hickman KE, Monahan AB, Schwarcz D. Chatgpt goes to law school. J Legal Educ. 2021;71:387.
- Hung J, Chen J. The benefits, risks and regulation of using ChatGPT in Chinese academia: A content analysis. Soc Sci. 2023;12(7):380.
- Gao S, Gao AK. on the origin of llms: An evolutionary tree and graph for 15,821 large language models. arXiv preprint arXiv:2307.09793. 2023.
- 9. Chiu TK, Moorhouse BL, Chai CS, Ismailov M. Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. Interact Learn Environ. 2023:1-7.
- Rahman MM, Terano HJ, Rahman MN, Salamzadeh A, Rahaman MS. ChatGPT and academic research: A review and recommendations based on practical examples. J Educ Manag Dev Stud. 2023;3(1):1-2.
- Ramos ASM. Generative artificial intelligence based on large language models: Tools for use in academic research. Scielo preprints 2023.
- Lee YF, Hwang GJ, Chen PY. Impacts of an AI-based cha bot on college students' after-class review, academic performance, selfefficacy, learning attitude, and motivation. Educ Technol Res Dev. 2022;70(5):1843-1865.
- Yin J, Goh TT, Yang B, Xiaobin Y. Conversation technology with micro-learning: The impact of chatbot-based learning on students' learning motivation and performance. J Edu Comp Res. 2021;59(1):154-177.
- 14. Essel HB, Vlachopoulos D, Tachie-Menson A, Johnson EE, Baah PK. The impact of a virtual teaching assistant (chatbot) on students' learning in Ghanaian higher education. Int J Edu Technol High Edu. 2022;19(1):1-9.
- Pahune S, Chandrasekharan M. Several categories of large language models (llms): A short survey. arXiv preprint arXiv:2307.10188. 2023.
- 16. Urban M, Děchtěrenko F, Lukavský J, Hrabalová V, Svacha F, Brom C, et al. Can ChatGPT improve creative problem-solving performance in university students?
- Siregar FH, Hasmayni B, Lubis AH. the analysis of chatgpt usage impact on learning motivation among Scout students. Int J Res Rev. 2023;10(7):632-638.
- Sallam M, Salim NA, Barakat M, Al-Mahzoum K, Ala'a B, Malaeb D, et al. Assessing health students' attitudes and usage of ChatGPT in Jordan: Validation study. JMIR Med Educ. 2023;9(1):e48254.

- 19. Damásio BF. Uses of exploratory factorial analysis in psychology. Avaliação Psicológica. 2012;11(2):213-228.
- Goretzko D, Pham TT, Bühner M. Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. Curr Psychol. 2021;40:3510-3521.
- Taherdoost HA, Sahibuddin SH, Jalaliyoon NE. Exploratory factor analysis; concepts and theory. Adv Pure Appl Math. 2022;27:375-382.
- Kristensen TS, Borritz M, Villadsen E, Christensen KB. The copenhagen burnout inventory: A new tool for the assessment of burnout. Work & stress. 2005;19(3):192-207.
- Wilson ML, Huggins-Manley AC, Ritzhaupt AD, Ruggles K. Development of the Abbreviated Technology Anxiety Scale (ATAS). Behav Res Methods. 2023;55(1):185-199.
- Oliveira AF, Caliatto SG. Exploratory factorial analysis of a scale of learning strategies. Educação: Teoria e Prática. 2018;28(59):548-565.
- Pereira CPDS, dos Santos AA, Ferraz AS. Learning strategies assessment scale for vocational education: Adaptation and psychometric studies. Revista Portuguesa de Educação. 2020;33(1);75–93.
- Crawford AV, Green SB, Levy R, Lo WJ, Scott L, Svetina D, et al. Evaluation of parallel analysis methods for determining the number of factors. Educ Psychol Meas. 2010;70(6):885-901.
- Goretzko D, Bühner M. One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis. Psychol Methods. 2020;25(6):776.
- Dunn TJ, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. Br J Psychol. 2014;105(3):399-412.
- 29. Sass DA, Schmitt TA. A comparative investigation of rotation criteria within exploratory factor analysis. Multivariate Behav Res. 2010;45(1):73-103.
- Shanahan M. Talking about large language models. arXiv preprint arXiv:2212.03551. 2022.
- Theobald M. Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. Contemp Educ Psychol. 2021;66:101976.
- 32. Mofatteh M. Risk factors associated with stress, anxiety, and depression among university undergraduate students. AIMS Public Health. 2021;8(1):36.