

## PAPER

# We Can Rely on ChatGPT as an Educational Tutor: A Cross-Sectional Study of its Performance, Accuracy, and Limitations in University Admission Tests

Saul Beltozar-Clemente<sup>1</sup>(✉), Enrique Díaz-Vega<sup>2</sup>, Raul Tejeda-Navarrete<sup>3</sup>, Joselyn Zapata-Paulini<sup>4</sup>

<sup>1</sup>Dirección de cursos básicos, Universidad Científica del Sur, Lima, Perú

<sup>2</sup>Departamento de ciencias, Universidad Privada del Norte, Lima, Perú

<sup>3</sup>Area de ciencias, Universidad Tecnológica del Perú, Lima, Perú

<sup>4</sup>Escuela de Posgrado, Universidad Continental, Lima, Perú

[sbeltozar@cientifica.edu.pe](mailto:sbeltozar@cientifica.edu.pe)

## ABSTRACT

The aim of this research was to evaluate the performance of ChatGPT in answering multiple-choice questions without images in the entrance exams to the National University of Engineering (UNI) and the Universidad Nacional Mayor de San Marcos (UNMSM) over the past five years. In this prospective exploratory study, a total of 1182 questions were gathered from the UNMSM exams and 559 questions from the UNI exams, encompassing a wide range of topics including academic aptitude, reading comprehension, humanities, and scientific knowledge. The results indicate a significant ( $p < 0.001$ ) and higher proportion of correct answers for UNMSM, with 72% (853/1182) of questions answered correctly. In contrast, there is no significant difference ( $p = 0.168$ ) in the proportion of correct and incorrect answers for UNI, with 52% (317/552) of questions answered correctly. Similarly, in the World History course ( $p = 0.037$ ), ChatGPT achieved its highest performance at a general level, with an accuracy of 91%. However, this was not the case in the language course ( $p = 0.172$ ), where it achieved the lowest score of 55%. In conclusion, to fully harness the potential of ChatGPT in the educational setting, continuous evaluation of its performance, ongoing feedback to enhance its accuracy and minimize biases, and tailored adaptations for its use in educational settings are essential.

## KEYWORDS

ChatGPT, performance, entrance exams, university

## 1 INTRODUCTION

Education plays a fundamental role in society as it has a profound and lasting influence on the development of individuals and communities, making it a key driver for social development [1]. Technology is playing an important role in this

Beltozar-Clemente, S., Díaz-Vega, E., Tejeda-Navarrete, R., Zapata-Paulini, J. (2024). We Can Rely on ChatGPT as an Educational Tutor: A Cross-Sectional Study of its Performance, Accuracy, and Limitations in University Admission Tests. *International Journal of Engineering Pedagogy (ijEP)*, 14(1), pp. 50–60. <https://doi.org/10.3991/ijep.v14i1.46787>

Article submitted 2023-09-16. Revision uploaded 2023-11-04. Final acceptance 2023-11-25.

© 2024 by the authors of this article. Published under CC-BY.

phenomenon, altering the way people learn and teach [2]. The impact of the GPT-3 (Generative Pre-trained Transformer 3) language model on educational paradigms has been very significant in recent years [3]. This technological advance [4] has brought about a profound change in the way educators approach the teaching [5] and learning process [6]. ChatGPT has facilitated the development of synchronous personalized teaching and self-learning systems [7], wherein the system can adjust its content based on the individual needs of the students [8]. This allows for more effective and efficient teaching by addressing cognitive difficulties in a more specific manner [9]. Therefore, we can assert that one of the most significant contributions of ChatGPT to education is the personalization of learning [10].

The free availability of ChatGPT version 3.5 to answer questions and provide information has revolutionized the way students access knowledge [11]. This breaks the traditional learning scheme, as students now focus on developing their information search, selection, and evaluation skills [12]. This has opened the door to a more efficient and effective teaching and learning process [13], where students can receive constructive feedback and targeted resources to address their cognitive gaps [14]. In today's world, it is crucial to cultivate digital literacy as a fundamental skill for educators and learners to effectively engage with AI systems such as ChatGPT [15]. This fosters the advancement of critical skills essential for professionals in this era, including critical thinking and problem-solving [16]. Students can leverage technology to ask intricate questions and collaborate on research projects [17], utilizing ChatGPT as a dependable source of information, thereby fostering independent learning and skill development for diverse learning styles [18].

GPT-3-based technology has paved the way for greater inclusion in the classroom [19]. Machine translation and adaptation of content for various learning styles have made it possible to meet the needs of students with diverse learning preferences, thus contributing to a more equitable educational environment [20]. Despite the aforementioned benefits, the integration of ChatGPT in education raises ethical and privacy challenges [21]. It is essential to carefully monitor data collection and automated decision-making to safeguard the rights and privacy of students [22] and to ensure that this technology benefits all students in a fair and equitable manner [23].

On the other hand, gaining admission to a Peruvian public university poses challenges for young people seeking to pursue higher education. These obstacles include economic barriers, educational inequality, limited academic programs, and inadequate preparation [24] [25]. The quality of education in Peru varies significantly based on geographic location and resource availability [26]. Young people from rural areas and disadvantaged communities often encounter substantial academic disadvantages [27]. The Peruvian public university is highly selective in its admissions process, requiring high scores in the admission exams [28]. This generates intense competition among applicants vying for a spot in their university classrooms [29]. Intense pressure and competition have a significant impact on students' mental health, as the stress and anxiety related to test preparation and test-taking can be overwhelming [30]. The high demand and competition were evident in the last admission process in 2023 for the National University Mayor de San Marcos (UNMSM), where 25,832 applicants applied for a total of 2,705 vacancies [31]. Similarly, for the National University of Engineering, 4,200 applicants applied for a total of 1,178 vacancies [32].

The objective of the research was to measure the performance of ChatGPT in answering non-multiple-choice questions in the admission exams of the two most

prestigious and demanding public universities in Latin America over the last five years: the National University of Engineering (UNI) and UNMSM. To assess its potential as an innovative technological tool in the field of education and be considered a virtual tutor in preparing thousands of university students for their academic training. The study not only emphasizes the models' performance but also provides information on their capabilities and limitations for their application in educational settings.

## 2 RELATED WORK

The following is a research study that examines the advantages and drawbacks of integrating ChatGPT in educational environments. The authors of [33] conducted a study on the performance of ChatGPT in the Chinese National Medical Licensing Examination (NMLE), evaluating its performance in exams over a period of three years. The study found that ChatGPT's performance was lower than that of medical students, and the proportion of correct answers was correlated with the year in which the test questions were published. This led to the conclusion that ChatGPT's knowledge and interpretation abilities for NMLE were still not comparable to those of medical students in China. Similarly, [34] analyzed ChatGPT's ability to pass the Royal College of Surgeons Orthopaedic Fellow (FRCS Orth) Part A exam. ChatGPT faced 240 simulated FRCS Orth Part A questions and obtained an overall score of 67.5%. The conclusion was that ChatGPT was unable to pass the exam. Several factors contributed to this outcome, including a lack of critical thinking, limited clinical experience, and an inability to meet the rigorous requirements of the exam.

They also [35] evaluated the performance of ChatGPT in their GPT-4 model on the European Board of Ophthalmology (EBO) exam and its potential role in medical education. They obtained a success rate of 91% on the EBO exam, demonstrating a high level of competence in the knowledge and application of ophthalmology. Likewise, in [36], a study was conducted on the performance of the last five medical specialty examinations (MSE), which included a total of 1177 questions. The lowest success rate was 54.3%, and the highest success rate was 70.9% for correct answers. There was no statistical difference between the correct and incorrect answers given to the clinical and basic science questions ( $p = 0.66$ ). However, there was a statistically significant difference between the short-questions group and the long-questions group ( $p = 0.03$ ). Similarly, [37] conducted a study to evaluate the performance of ChatGPT on non-imaging radiology board-style exam questions and explore its strengths and limitations. ChatGPT has been developed with 150 multiple-choice questions designed to match the style, content, and difficulty of the Canadian Royal College and American Board of Radiology exams. As a result, they achieved a 69% correct question rate (104 out of 150). The model demonstrated better performance on questions involving lower-order thinking (84%, 51 out of 61) than on those involving higher-order thinking (60%, 53 out of 89) ( $p = 0.02$ ).

## 3 METHODOLOGY

The admission exams for the UNMSM for the years 2023, 2022, 2020, 2019, and 2018 have been chosen. The following subjects were chosen: verbal skills,

language, literature, psychology, civics, Peruvian history, world history, geography, economics, philosophy, and biology. ChatGPT can perform adequately in these subjects, but it has limitations in processing graphics for mathematics courses. Similarly, the entrance exams for the UNI for the years 2023, 2022, 2020, 2019, and 2018 were chosen. The following areas were chosen for verbal reasoning and humanities, where ChatGPT can perform adequately due to its limitation in processing graphics for mathematics courses. This is to achieve optimal performance and a comprehensive evaluation of ChatGPT across various thematic areas.

To organize the questions in each subject for every university, we collected the official exams and answer keys provided by each university, which are publicly available. In both universities, the last five years were considered, except for 2021, for which we found no record of exams. The final dataset consisted of 1232 questions for UNMSM, of which 50 questions were excluded because they were graphically worded, making it difficult for ChatGPT to process them effectively. Likewise, for UNI, the final dataset comprised 560 questions, with only one question being excluded.

We used the GPT-3.5 version to conduct this research. To ensure accuracy in the performance of ChatGPT, we utilized the official multiple-choice questions and their respective answer keys provided by the admissions office of each university. All questions developed by ChatGPT were text-based and were directly copied and pasted in order to incorporate them. Graphical questions were excluded from our study to minimize potential bias in the responses provided by ChatGPT.

The evaluation for UNMSM covered the following subjects: verbal ability (285 questions), language (144 questions), literature (78 questions), psychology (103 questions), civics (83 questions), the history of Peru (61 questions), world history (43 pre-questions), geography (72 questions), economics (87 questions), philosophy (85 questions), and biology (141 questions). Similarly, the assessment for UNI included 330 verbal reasoning questions and 229 questions for all humanities courses. ChatGPT's performance in terms of providing correct answers was measured using the binomial distribution test. Performance based on estimated test scores between courses of the same exam was evaluated using a two-tailed paired sample t-test.

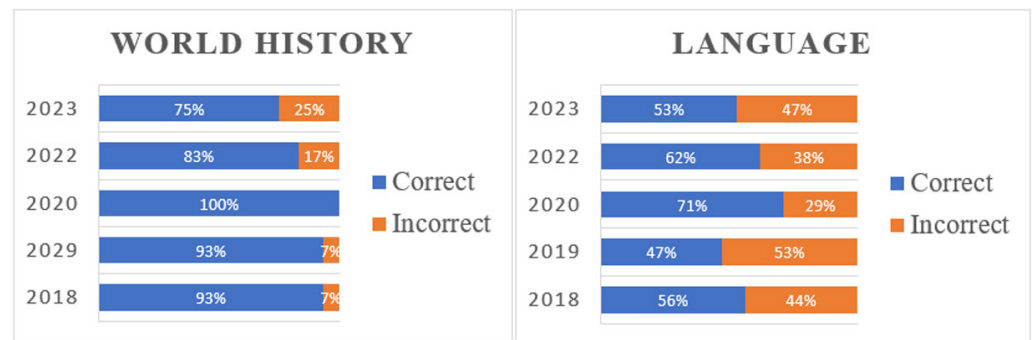
## 4 RESULTS

The performance of ChatGPT varied across the different admission tests evaluated, showing discernible patterns based on the course, university, and year of evaluation. When aggregating the questions from all the evaluations over the years, the overall proportion of correct answers in the admission exams for UNMSM was significantly higher than the proportion of incorrect answers ( $p < 0.001$ ). However, no significant differences were observed between the proportion of correct and incorrect answers in the tests corresponding to UNI ( $p = 0.168$ ), as indicated in Table 1.

**Table 1.** ChatGPT performance in UNMSM – UNI admission tests

Years	Courses	Answered Questions	Correct Questions	Result	Statistical Test
<b>National University of San Marcos (UNMSM)</b>					
2023-2022-2020-2019-2018	Verbal Ability	285	215	75%	0.039
	Language	144	79	55%	0.172
	Literature	78	56	72%	0.038
	Psychology	103	75	73%	0.035
	Civic Education	83	65	78%	0.005
	History of Peru	61	45	74%	<0.001
	World History	43	39	91%	0.037
	Geography	72	42	58%	0.354
	Economy	87	64	74%	<0.001
	Philosophy	85	62	73%	0.016
	Biology	141	111	79%	0.007
	<b>Total</b>		<b>1182</b>	<b>853</b>	<b>72%</b>
<b>National University of Engineering (UNI)</b>					
2023-2022-2020-2019-2018	Verbal reasoning	330	188	57%	0.181
	Humanities	229	129	56%	0.302
	<b>Total</b>	<b>559</b>	<b>317</b>	<b>57%</b>	<b>0.168</b>

In the overall UNMSM tests, ChatGPT achieved the highest performance in the world history course in terms of the percentage of correct answers, reaching a maximum of 91% (39/43) ( $p = 0.037$ ) at the global level. On the contrary, the model encountered significant challenges in the overall language courses, achieving a 55% correct response rate (79/144) ( $p = 0.172$ ) and a correct response rate in geography 58% (42/72) ( $p = 0.354$ ). The proportion of correct to incorrect responses in these courses was not statistically significant, as depicted in Figure 1. ChatGPT demonstrated outstanding performance in the UNMSM exams, scoring in geography (8/8), and history of Peru 100% in economics (8/8), and world history 100% (4/4) in 2020. ChatGPT performance was consistently maintained in the UNMSM exams, with average scores of 71% in 2018, 75% in 2019, 75% in 2020, 71% in 2022, and 72% in 2023. This performance demonstrated a significant and higher difference in correct answers compared to incorrect answers ( $p < 0.001$ ).



**Fig. 1.** Historical ChatGPT answer accuracy for the highest and lowest performing courses respectively in the UNMSM exams

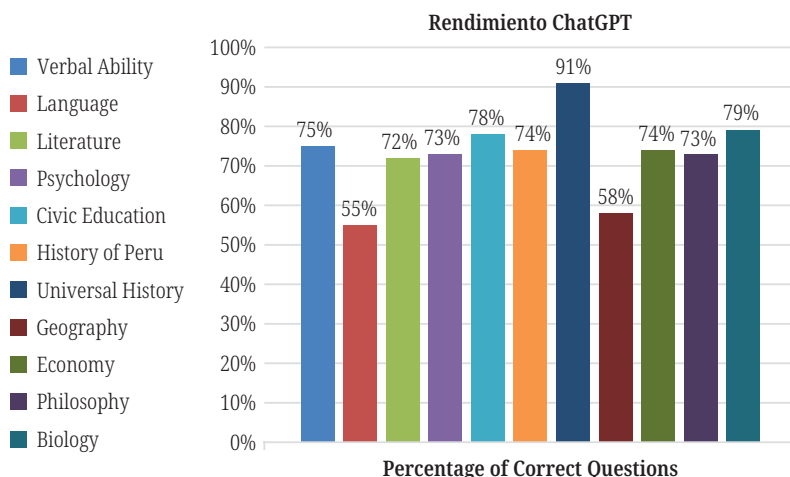


Fig. 2. Overall ChatGPT performance for each course in the UNMSM exams

Likewise, he achieved moderate success in verbal ability courses with a 75% (215/285) completion rate, economics with a 74% (64/87) completion rate, psychology with a 73% (75/103) completion rate, philosophy with a 73% (62/85) completion rate, and literature with a 72% (56/78) completion rate, as shown in Figure 2. The ChatGPT achieved an average performance of 56% (129/229) for the humanities courses and 57% (188/330) for the verbal reasoning course in the total UNI tests. This global performance indicates that there is no significant difference between correct and incorrect answers ( $p = 0.168$ ). Likewise, ChatGPT’s highest performance in the UNI exam was in the verbal reasoning course, achieving 71% (50/70) in 2023. Paradoxically, it faced difficulties in answering correctly in the same verbal reasoning course, obtaining 48% (38/80) in 2018, as indicated in Figure 3. In addition, the performance of ChatGPT on the exams for UNI was not significantly different ( $p = 0.302$ ) in the years 2018 (51%), 2019 (57%), 2020 (53%), 2022 (59%), and 2023 (62%).

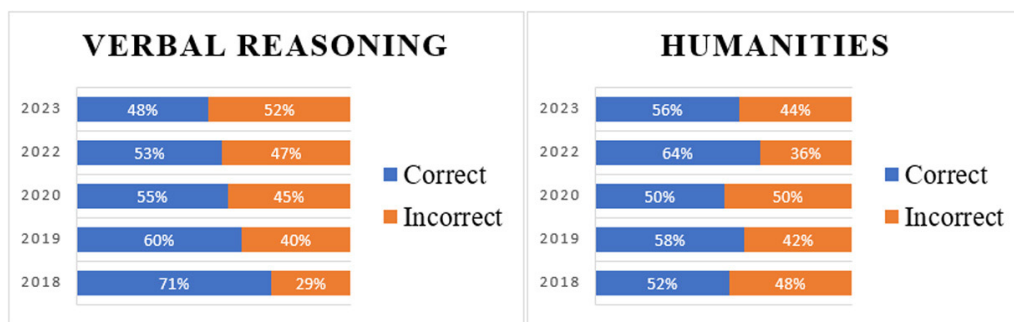


Fig. 3. Historical ChatGPT response accuracy for the highest and lowest performing courses respectively in the UNI exams

ChatGPT performance was below expectations in the literature course, with 43% (3/7) in 2018 and 38% (3/8) in 2020, as well as in the geography course, with 48% (10/21) in 2023. In addition, the performance of the GPT-3.5 model varied significantly over the years in the literature course 43% (2108), 67% (2019), 38% (2020), 86% (2022), and 78% (2023), unlike in other course, where its performance remained relatively constant on average. Similarly, consistent performance in the exams of both UNMSM and UNI has been observed over the years, as depicted in Figure 4.



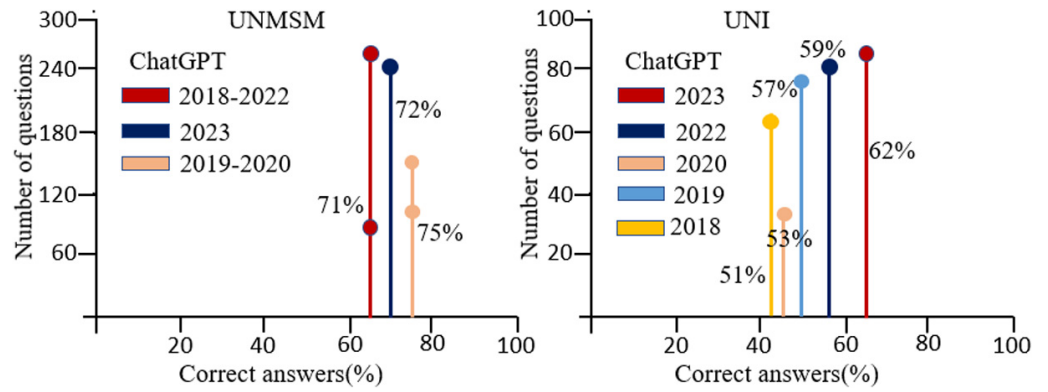


Fig. 4. Percentage and number of correct ChatGPT answers in admission exams over the years

In addition, it is observed that in the accumulated questions there is a density of correct answers in the Verbal Ability course compared to the total number of questions answered, as well as a lower density of correct answers in the Geography course, as indicated in Figure 5.

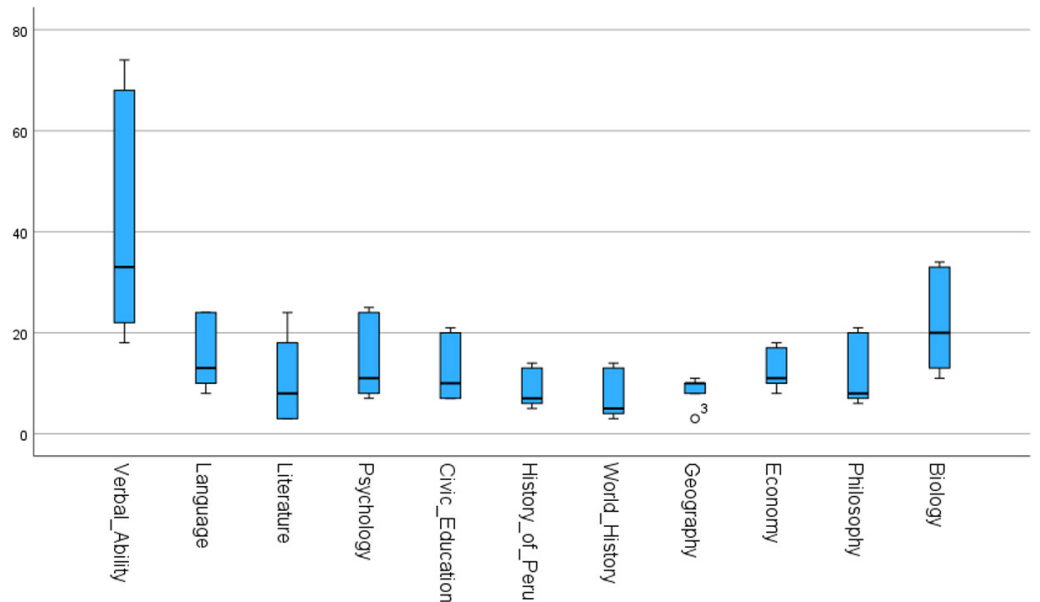


Fig. 5. Density of questions answered correctly by ChatGPT with respect to the total number of questions, each course in the UNMSM exams

## 5 DISCUSSION

### A) Principal findings

The research aimed to assess ChatGPT’s performance in answering multiple-choice questions from the most recent admission exams of the two leading public universities in Latin America, UNMSM and UNI, to evaluate its potential as an innovative technological tool for the educational system and its suitability as a virtual tutor for thousands of students preparing for university education.

The analysis of the results obtained in this research reveals interesting patterns in the performance of ChatGPT in a variety of courses taken in the admission exams at both UNMSM and UNI over several years. First, it is noteworthy that ChatGPT

achieved its best performance in the World History course, with an exceptionally high correct answer rate of 91% overall. This indicates that the model has a profound understanding of this field and can offer highly precise answers on historical topics. In addition, during the UNMSM-specific exams, ChatGPT demonstrated outstanding performance in courses such as geography and Peruvian history, achieving 100% correct answers. These results suggest that the model is capable of providing accurate answers in specific contexts. It is important to note that these cases are limited to certain years and courses.

On the other hand, the model encountered significant challenges in the language course, achieving a correct response rate of only 55%. This result indicates that the model struggled significantly to understand and effectively answer questions related to this area. It is important to consider that these courses often involve a high level of linguistic complexity, which could account for the lower success rates.

However, there was moderate success overall in courses that involved verbal skills and humanities subjects such as economics, psychology, philosophy, and literature, with correct response rates ranging from 72% to 75%. These results suggest that ChatGPT performs well in subjects that require a deep understanding of language and concepts related to the humanities.

In terms of performance in the various UNI courses, the ChatGPT achieved correct response rates of 56% for the humanities courses and 57% for the verbal reasoning course overall, indicating average performance in these academic domains. It is important to note that there were significant differences in ChatGPT performance within the same course across different years. For example, in the literature course, correct response rates varied significantly from year to year. This suggests that the model struggled to maintain consistent performance in this course over the years.

In summary, these results offer valuable information on the strengths and weaknesses of ChatGPT in different university courses. While the model exhibits profound expertise in specific areas and can offer flawless responses in particular contexts, it struggles with courses involving linguistic complexity. The variation in its performance in the same course over the years also emphasizes the necessity for ongoing assessment and tailored adjustments to maximize its effectiveness in the educational setting.

## **B) Study limitations**

We have identified some limitations, including the fact that we evaluated the performance of ChatGPT on a small number of admission exams from the aforementioned universities. This may not be representative of all tests used at other universities in Latin America or on other continents. Another significant limitation is the potential for bias in the responses produced by the model. Since ChatGPT learns from real data, it was trained on a corpus of data until 2021. This limits its understanding of information beyond that time, which could lead to biased responses or inadvertent biases on sensitive topics. This raises ethical concerns and requires careful oversight when using the model in educational settings. In summary, although ChatGPT is a promising tool, it is crucial to acknowledge and tackle these limitations to guarantee its responsible and efficient use in educational activities.

## **6 CONCLUSIONS**

In conclusion, the data presented in our research depicts a mixed picture of ChatGPT performance on the UNMSM and UNI admission exams. To fully harness



the potential of ChatGPT in the educational setup, continuous evaluation of its performance, ongoing feedback to enhance accuracy and minimize biases, and tailored adaptations for its use in educational contexts are essential. ChatGPT serves as a support tool in the academic training of students. This must be accompanied by sufficient development and raising awareness among educators and students. The objective is to understand the limitations of the model and its ethical and responsible use, which will enable its effective integration into academic training and educational research, maximizing its usefulness and minimizing potential challenges.

## 7 REFERENCES

- [1] S. T. H. Pham, "Exploring the lived experience of educators and business executives in the phenomenon of artificial intelligence in education," in *Phenomenological Studies in Education*, 2023, pp. 182–206. <https://doi.org/10.4018/978-1-6684-8276-6.ch010>
- [2] M. Hasnain, "Commentary: Future thoughts on ChatGPT and its applications in education," *Journal of Research in Innovative Teaching and Learning*, vol. 16, no. 2, pp. 284–285, 2023. <https://doi.org/10.1108/JRIT-09-2023-100>
- [3] S. S. Gill *et al.*, "Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 19–23, 2024. <https://doi.org/10.1016/j.iotcps.2023.06.002>
- [4] A. L. V. Ubaldo, V. Y. G. Barreto, F. Sierra-Liña, and M. Cabanillas-Carbonel, "Application of learning software in basic education students with intellectual disabilities: A systematic review of the literature," *International Journal of Engineering Pedagogy*, vol. 13, no. 1, pp. 45–64, 2023. <https://doi.org/10.3991/ijep.v13i1.35677>
- [5] Y. Dosymov *et al.*, "Effectiveness of computer modeling in the study of electrical circuits: Application and evaluation," *International Journal of Engineering Pedagogy*, vol. 13, no. 4, pp. 93–112, 2023. <https://doi.org/10.3991/ijep.v13i4.34921>
- [6] H. Ibrahim *et al.*, "Perception, performance, and detectability of conversational artificial intelligence across 32 university courses," *Scientific Reports*, vol. 13, p. 12187, 2023. <https://doi.org/10.1038/s41598-023-38964-3>
- [7] C. K. Y. Chan and W. Hu, "Students' voices on generative AI: Perceptions, benefits, and challenges in higher education," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 43, 2023. <https://doi.org/10.1186/s41239-023-00411-8>
- [8] C. K. Y. Chan, "A comprehensive AI policy education framework for university teaching and learning," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, 2023. <https://doi.org/10.1186/s41239-023-00408-3>
- [9] A. M. Elkhatat, "Evaluating the authenticity of ChatGPT responses: A study on text-matching capabilities," *International Journal for Educational Integrity*, vol. 19, no. 1, 2023. <https://doi.org/10.1007/s40979-023-00137-0>
- [10] S. S. Gill *et al.*, "Transformative effects of ChatGPT on modern education: Emerging Era of AI Chatbots," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 19–23, 2024. <https://doi.org/10.1016/j.iotcps.2023.06.002>
- [11] S. G. Kim, "Using ChatGPT for language editing in scientific articles," *Maxillofac Plast Reconstr Surg*, vol. 45, no. 1, 2023. <https://doi.org/10.1186/s40902-023-00381-x>
- [12] M. Salvagno, F. S. Taccone, and A. G. Gerli, "Can artificial intelligence help for scientific writing?" *Critical Care*, vol. 27, no. 1, 2023. <https://doi.org/10.1186/s13054-023-04380-2>
- [13] S. Beltozar-Clemente, O. Iparraquirre-Villanueva, J. Zapata-Paulini, and M. Cabanillas-Carbonell, "Changing mathematical paradigms at the university level: Feedback from a flipped classroom at a peruvian university," *International Journal of Engineering Pedagogy*, vol. 13, no. 6, pp. 76–89, 2023. <https://doi.org/10.3991/ijep.v13i6.40763>

- [14] V. Ratten and P. Jones, "Generative artificial intelligence (ChatGPT): Implications for management educators," *International Journal of Management Education*, vol. 21, no. 3, p. 100857, 2023. <https://doi.org/10.1016/j.ijme.2023.100857>
- [15] M. C. Keiper, "ChatGPT in practice: Increasing event planning efficiency through artificial intelligence," *Journal of Hospitality, Leisure, Sport & Tourism Education*, vol. 33, p. 100454, 2023. <https://doi.org/10.1016/j.jhlste.2023.100454>
- [16] A. Taşgin and C. Dilek, "The mediating role of critical thinking dispositions between secondary school student's self-efficacy and problem-solving skills," *Thinking Skills and Creativity*, vol. 50, p. 101400, 2023. <https://doi.org/10.1016/j.tsc.2023.101400>
- [17] P. P. Ray, "Refining the application of artificial intelligence in the water domain: Exploring the potential of ChatGPT," *Science of the Total Environment*, vol. 892, p. 164638, 2023. <https://doi.org/10.1016/j.scitotenv.2023.164638>
- [18] G. Karaali, "Artificial intelligence, basic skills, and quantitative literacy," *Numeracy*, vol. 16, no. 1, 2023. <https://doi.org/10.5038/1936-4660.16.1.1438>
- [19] K. Masters, "Ethical use of artificial intelligence in health professions education: AMEE Guide No. 158," *Medical Teacher*, vol. 45, no. 6, pp. 574–584, 2023. <https://doi.org/10.1080/0142159X.2023.2186203>
- [20] M. Farrokhnia, S. K. Banihashem, O. Noroozi, and A. Wals, "A SWOT analysis of ChatGPT: Implications for educational practice and research," *Innovations in Education and Teaching International*, pp. 1–15, 2023. <https://doi.org/10.1080/14703297.2023.2195846>
- [21] P. P. Ray, "ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 121–154, 2023. <https://doi.org/10.1016/j.iotcps.2023.04.003>
- [22] A. Strzelecki, "To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology," *Interactive Learning Environments*, pp. 1–14, 2023. <https://doi.org/10.1080/10494820.2023.2209881>
- [23] M. Ajevski, K. Barker, A. Gilbert, L. Hardie, and F. Ryan, "ChatGPT and the future of legal education and practice," *Law Teacher*, vol. 57, no. 3, pp. 352–364, 2023. <https://doi.org/10.1080/03069400.2023.2207426>
- [24] F. Nancy, S. M. Héctor, C. Ulises, V. Rosa, and S. Milagros, "Investigative competencies from the approach of the national project to 2036: Perspective of peruvian education," *Revista de Gestão Social e Ambiental*, vol. 17, no. 4, p. e03466, 2023. <https://doi.org/10.24857/rgsa.v17n4-013>
- [25] J. L. Y. Vargas, "Derroteros de la educación peruana en el XXI: Interculturalizar, decolonizar y subvertir," *Ensaio*, vol. 25, no. 97, pp. 918–942, 2017. <https://doi.org/10.1590/s0104-40362017002500988>
- [26] J. Saavedra and M. Gutierrez, "Peru: A wholesale reform fueled by an obsession with learning and equity," in *Audacious Education Purposes: How Governments Transform the Goals of Education Systems*, 2020, pp. 153–180. [https://doi.org/10.1007/978-3-030-41882-3\\_6](https://doi.org/10.1007/978-3-030-41882-3_6)
- [27] C. Brock and H. Lawlor, *Education in Latin America*, pp. 1–196, 2018. <https://doi.org/10.4324/9781351005180>
- [28] A. L. Gallegos, J. O. León, L. R. Zavala, A. R. Díaz, and K. B. Dulanto, "Reflections and background on the foundations of the Peruvian university and its autonomy," *Universidad y Sociedad*, vol. 14, no. 1, pp. 155–164, 2022.
- [29] J. F. Luesia, I. Benítez, R. Company-Córdoba, I. Gómez-Gómez, and M. Sánchez-Martín, "Assessing the relevance of academic competencies in college admission tests from a higher-order thinking perspective: A systematic review," *Thinking Skills and Creativity*, vol. 48, p. 101251, 2023. <https://doi.org/10.1016/j.tsc.2023.101251>
- [30] O. J. Mamani-Benito, R. F. Carranza-Esteban, T. Caycho-Rodríguez, I. E. Corrales-Reyes, and C. R. Mejia, "Research lines that include mental health as a study priority in peruvian universities," *Revista Cubana de Medicina Militar*, vol. 50, no. 4, 2021.

- [31] National University of San Marcos, “Oficina general de planificación.” Accessed: Oct. 06, 2023. [Online]. Available: <https://ogpl.unmsm.edu.pe/Cifras>
- [32] National Engineering University, “Universidad nacional de ingeniería – UNI.” Accessed: Oct. 06, 2023. [Online]. Available: <https://uni.edu.pe/>
- [33] X. Wang *et al.*, “ChatGPT performs on the Chinese national medical licensing examination,” *Journal of Medical Systems*, vol. 47, no. 1, 2023. <https://doi.org/10.1007/s10916-023-01961-0>
- [34] A. Saad, K. P. Iyengar, V. Kurisunkal, and R. Botchu, “Assessing ChatGPT’s ability to pass the FRCS orthopaedic part A exam: A critical analysis,” *Surgeon*, vol. 21, no. 5, pp. 263–266, 2023. <https://doi.org/10.1016/j.surge.2023.07.001>
- [35] C. Panthier and D. Gatinel, “Success of ChatGPT, an AI language model, in taking the French language version of the European board of ophthalmology examination: A novel approach to medical knowledge assessment,” *Journal Français D’Ophtalmologie*, vol. 46, no. 7, pp. 706–711, 2023. <https://doi.org/10.1016/j.jfo.2023.05.006>
- [36] A. D. Oztermeli and A. Oztermeli, “ChatGPT performance in the medical specialty exam: An observational study,” *Medicine*, vol. 102, no. 32, p. E34673, 2023. <https://doi.org/10.1097/MD.00000000000034673>
- [37] R. Bhayana, S. Krishna, and R. R. Bleakney, “Performance of ChatGPT on a radiology board-style examination: Insights into current strengths and limitations,” *Radiology*, vol. 307, no. 5, p. e230582, 2023. <https://doi.org/10.1148/radiol.230582>

## 8 AUTHORS

**Saul Beltozar-Clemente**, Degree in Mathematics and Physics from Universidad Nacional del Centro del Perú. Master in Education Sciences from the National University of Education – Peru. Diploma in Scientific Research from the Universidad de Ciencias y Humanidades – Peru. Leadership Certification in Agile Project Management by Tecnológico de Monterrey – Mexico. Professor at Universidad Privada del Norte, Universidad Científica del Sur, Universidad Tecnológica del Perú. International lecturer in Argentina, United Kingdom. Specialization in software development, AI, Machine Learning. Consultant and trainer in Information Technology in public and private institutions focused on education (E-mail: [sbeltozar@cientifica.edu.pe](mailto:sbeltozar@cientifica.edu.pe)).

**Enrique Díaz-Vega**, I have a degree in mathematics, a master’s degree in education and a doctorate in education sciences, in addition to several certificates and qualifications of good professional performance. Currently I am a professor in the Science Department at the Universidad Privada del Norte (E-mail: [enrique.diaz@upn.edu.pe](mailto:enrique.diaz@upn.edu.pe)).

**Raul Tejada-Navarrete**, Doctor in Education and Master in the specialty of Evaluation and Accreditation of Educational Quality, Bachelor in Education with a major in mathematics and physics, graduated as a professional technician in accounting. He is currently a professor at Universidad Cesar Vallejo and Universidad Tecnológica del Peru in Mathematics I and II, Financial Mathematics and Physics I and II, and a teacher appointed by MINEDU. His areas of research are basic and university education in curriculum and pedagogical management (E-mail: [C22618@utp.edu.pe](mailto:C22618@utp.edu.pe)).

**Joselyn Zapata-Paulini**, is a master’s student in environmental management and sustainable development at Universidad Continental and writes scientific articles. She is specialized in augmented reality and software development (E-mail: [70994337@continental.edu.pe](mailto:70994337@continental.edu.pe)).