



ORIGINAL BREVE

## Performance of the ChatGPT tool in solving residency exams

### Desempeño de la herramienta ChatGPT en la resolución de exámenes de residencia

Javier Gonzalez-Argote<sup>1</sup>  , William Castillo-González<sup>1,2</sup>  

<sup>1</sup>Fundación Salud, Ciencia y Tecnología. Ciudad Autónoma de Buenos Aires, Argentina.

<sup>2</sup>Universidad de Ciencias Empresariales y Sociales. Ciudad Autónoma de Buenos Aires, Argentina.

**Citar como:** Gonzalez-Argote J, Castillo-González W. Performance of ChatGPT tool in the resolution of residency exams in Argentina. *Seminars in Medical Writing and Education*. 2024; 3:56. <https://doi.org/10.56294/mw202456>

Enviado: 08-09-2023

Revisado: 15-11-2023

Aceptado: 20-01-2024

Publicado: 22-01-2024

Editor: Dr. José Alejandro Rodríguez-Pérez 

#### ABSTRACT

**Introduction:** artificial intelligence is classified as a tool of interest at the present time. Through its application, organizational processes and decision-making are transformed; while promoting innovative development.

**Objective:** to describe the performance of the ChatGPT tool in solving residency exams.

**Method:** an observational, descriptive, retrospective study was carried out. As the universe of the study, all test-type questions (single selection) were analyzed. The variables analyzed were each of the questions belonging to each exam (correct or incorrect answers). Descriptive statistics were applied.

**Results:** Syllabus B stood out within group 1 with the highest number of correct answers (209 for 69,66 %). For its part, within group 2, syllabus D was predominant with 141 correct answers (70,5 %). The exams related to nursing stood out.

**Conclusions:** the use of artificial intelligence tools such as ChatGPT is variable in the field of medical sciences. Their performance in solving scientific questions is heterogeneous. It may vary with respect to the format of the question and the topic addressed.

**Keywords:** Science, Technology and Society; Medical Education; Professional education; Artificial intelligence; Technology.

#### RESUMEN

**Introducción:** la inteligencia artificial se cataloga como una herramienta de interés en los momentos actuales. Mediante su aplicación se transforman los procesos organizativos y toma de decisiones; a la par que fomenta el desarrollo innovador.

**Objetivo:** describir el desempeño de la herramienta ChatGPT en la resolución de exámenes de residencia.

**Método:** se realizó un estudio observacional, descriptivo, retrospectivo. Como universo del estudio se analizaron todas las preguntas de tipo test (selección única). Las variables analizadas fueron cada una de las preguntas pertenecientes a cada examen (respuestas correctas o incorrectas). Se aplicó la estadística descriptiva.

**Resultados:** Sobresalió el temario B dentro del grupo 1 con el mayor número de acierto (209 para un 69,66 %). Por su parte, dentro del grupo 2 resulto predominante el temario D con 141 aciertos (70,5 %). Destacaron los exámenes referentes a enfermería.

**Conclusiones:** la utilización de las herramientas de inteligencia artificial como ChatGPT es variable en el ámbito de las ciencias médicas. Su desempeño en la resolución de interrogantes científicas es heterogéneo. Puede variar con respecto al formato de la interrogante y la temática abordada.

**Palabras clave:** Ciencia, Tecnología y Sociedad; Educación Médica; Educación Profesional; Inteligencia Artificial; Tecnología.

## INTRODUCCIÓN

El avance científico tecnológico de la humanidad, ha encontrado formas y medios para su estandarización y empleo en cada sector de la sociedad. Desde los procesos automatizados en las grandes industrias, el empleo necesario (e indiscutible) de las tecnologías de la información y las comunicaciones hasta los más exquisitos procedimientos y medios diagnósticos y terapéuticos (terapias de biotecnologías, nanotecnologías entre otras) están presentes en el quehacer cotidiano del hombre.

La inteligencia artificial (IA) se cataloga como una herramienta de interés en los momentos actuales. Mediante su aplicación se transforman los procesos organizativos y toma de decisiones; a la par que fomenta el desarrollo innovador. A su vez, es un servicio al alcance de todos.<sup>(1,2)</sup> En este escenario, destaca la herramienta ChatGPT (*Generative Pre-training Transformer*) como un sistema de inteligencia artificial generativa creado sobre la base de más de 175 millones de parámetros que domina más de 8 millones de documentos y fuentes de información; lo que permite su capacidad para la generación de respuestas coherentes.<sup>(3,4)</sup>

Su aplicación en el contexto universitario es inmensurable. Estas herramientas pueden mejorar el entorno de estudio de los estudiantes, en especial su interacción con espacios virtuales. A su vez, agilizan la adquisición de información.<sup>(5,6)</sup> Sin embargo, es indiscutible la necesidad de evaluar la calidad, de manera continua, de los procesos en aras de evitar errores académicos u otros inconvenientes que pueden afectar el desempeño del proceso docente educativo.<sup>(7)</sup>

En el ámbito de las ciencias de la salud, se ha visto documentado el desempeño de la IA en este sector. Un enfoque ha dirigido a evaluar su accionar frente a exámenes teóricos en las ciencias médicas, donde se recogen resultados variables en cada estudio, aunque la media de respuesta se encuentra entre el 50 y 70 % de acierto.<sup>(7)</sup> A su vez, se han valorado aspectos relacionados con su aplicación en el perfeccionamiento de los métodos diagnósticos en especial por imágenes.<sup>(8)</sup> En materias de investigación científica en este sector se preconiza su aplicación para la optimización del tiempo en la búsqueda bibliográfica y procesamiento de análisis; a la par que debe vigilarse posible sesgos en el análisis y protección de la información que deviene en implicaciones morales.<sup>(9)</sup>

Sin lugar a duda, el avance tecnológico es palpable en cada sector de la sociedad, en especial en las ciencias de la salud.<sup>(10)</sup> La diversidad de usos de la IA hace necesario de controles sistemáticos en aras de analizar de manera continua sus beneficios y prevenir posibles complicaciones. Por tal motivo el objetivo del presente estudio es describir el desempeño de la herramienta ChatGPT en la resolución de exámenes de residencia.

## MÉTODOS

Se realizó un estudio observacional, descriptivo, retrospectivo sobre el desempeño de ChatGPT en la resolución de los exámenes de residencia del año 2022. Como universo del estudio se analizaron todas las preguntas de tipo test (selección única); no se aplicaron técnicas de muestreo por lo que se trabajó con la totalidad del universo. Las variables analizadas fueron cada una de las preguntas pertenecientes a cada examen (respuestas correctas o incorrectas).

Para la recolección de la información se utilizó la herramienta ChatGPT mediante la siguiente interrogante: ¿Puedes responder a las siguientes preguntas de selección múltiple sobre medicina solamente con los ítems correctos? Posterior, se aplicaron los exámenes correspondientes a medicina, enfermería, bioquímica y matemática (divididos en 4 temarios). Solo se analizaron las respuestas correctas con respecto al total. Se aplicó estadística descriptiva.

Se tuvieron las normas éticas para el desarrollo de investigaciones en ciencias de la salud y la II Declaración de Helsinki.

## RESULTADOS

Sobresalió el temario B dentro del grupo 1 con el mayor número de acierto (209 para un 69,66 %). A su vez, los exámenes de enfermería destacaron dentro del propio grupo con 147 respuestas correctas (24,5 %) en comparación con el resto.

Tabla 1. Distribución de las respuestas según los temarios y especialidades

Materia	Grupo 1				Grupo 2				T
	Temario A		Temario B		Temario C		Temario D		
	Aciertos	%	Aciertos	%	Aciertos	%	Aciertos	%	
Bioquímica	67/100	67	69/100	69	-	-	-	-	136
Enfermería	72/100	72	75/100	75	68/100	68	74/100	74	289
Medicina	68/100	68	65/100	65	68/100	68	67/100	67	268
Total	207/300	69	209/300	69,66	136/200	68	141/200	70,5	693

Dentro del grupo 2 resulto predominante el temario D con 141 aciertos (70,5 %); con mayor representación de enfermería (142 aciertos; 35,5 %).

## DISCUSION

La puesta en práctica de las nuevas tecnologías en las distintas ramas de las ciencias potencian los procesos productivos.<sup>(11)</sup> A su vez, abren un espacio para el debate creciente sobre sus posibles limitaciones o implicaciones en su uso.

En relación con las ciencias de la salud y sus distintas aristas (asistencial, docente y científico-investigativo) el desempeño de la inteligencia artificial, en el caso particular de la herramienta ChatGPT es variable y amplio. Autores como Castillo-González<sup>(12)</sup> reconoce su utilidad en las distintas etapas del proceso editorial, en especial mediante la corrección de estilos de redacción de artículos para mejorar la coherencia científica. Sin embargo, destaca la importancia de la creatividad humana como fiscalizador del proceso editorial e investigativo dentro de las ciencias médicas. Criterio que es sustentado por Vega-Jiménez et al.<sup>(13)</sup> al mostrar la importancia de que los autores firmantes de las investigaciones declaren el uso de cualquier herramienta de IA para la confección de investigaciones, en aras de evitar posibles conflictos futuros relacionados con la autoría y el contenido de los trabajos.

En cuanto a la formación médica de pregrado y posgrado, las herramientas generadoras de contenido digital como es el caso del ChatGPT ayudan a la didáctica del aprendizaje mediante la optimización del tiempo y el acceso rápido efectivo a la información necesaria.<sup>(14)</sup> Referente a los resultados mostrados, Carrasco et al.<sup>(15)</sup> muestra resultados similares en relación con los porcentajes de aciertos por exámenes, a su vez estos autores destacan en su estudio que las preguntas que analizan varios elementos para su respuesta, acumulan mayor porcentaje de errores. A su vez, Alfertshofer et al.<sup>(16)</sup> concuerda con los resultados del presente estudio al analizar el desempeño de la herramienta en exámenes similares en distintos países con un promedio de aciertos entre un 22 % para los exámenes aplicadas en Francia y un 73 % para los aplicados en Italia.

Estos resultados pueden servir de base para la puesta en práctica de protocolos en las distintas instituciones de formación académicas en aras de fomentar el uso de las herramientas de inteligencia artificial solo para el mejoramiento de los procesos educativos; en aras de mantener la integridad del proceso docente educativo. Criterios que guardan relación con los expuestos por Vega-Jiménez et al.<sup>(17)</sup>

A su vez, es válido destacar que el proceso médico asistencial implica múltiples aspectos para arribar a un diagnóstico médico. El mismo conjuga el conocimiento teórico (demostrado en los distintos exámenes teóricos-prácticos durante los años de formación) y habilidades prácticas como escuchar, sentir y la capacidad de comunicación con el paciente. Elementos claves del adecuado interrogatorio y examen físico que debe realizar el personal médico; que se complementa con la experiencia del personal asistencial. De igual manera, se requieren la realización de determinados procedimientos diagnósticos (según el padecimiento sospechado) para arribar al correcto diagnóstico clínico.

Criterios que sustentan el refrán: “no existen enfermedades; sino enfermos” y el análisis del paciente como un ser biopsicosocial sobre el cual interactúan múltiples procesos, no solo la afección aguda sino factores agravantes, desencadenantes entre otros; se sustenta el concepto de salud como el estado de bienestar físico, mental y social y no solo la ausencia de enfermedad.<sup>(18)</sup> Aspectos que deben ser tomados en consideración para el diagnóstico de las distintas patologías y su posterior conducta terapéutica. Por tal motivo, el uso y empleo de estas herramientas deben ser consideradas como medios auxiliares para el diagnóstico y no como elemento principal del mismo. Gutiérrez-Cirlos et al.<sup>(19)</sup> expone criterios similares.

## CONCLUSIONES

La utilización de las herramientas de inteligencia artificial como ChatGPT es variable en el ámbito de las ciencias médicas. Su desempeño en la resolución de interrogantes científicas es heterogéneo. Puede variar con respecto al formato de la interrogante y la temática abordada.

## REFERENCIAS BIBLIOGRAFICAS

1. Castillo-González W. The importance of human supervision in the use of ChatGPT as a support tool in scientific writing. *Metaverse Basic and Applied Research* 2023;2:29-29. <https://doi.org/10.56294/mr202329>.

2. Castillo-Gonzalez W. ChatGPT and the future of scientific communication. *Metaverse Basic and Applied Research* 2022;1:8-8. <https://doi.org/10.56294/mr20228>.

3. González LN. EL IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN LOS NEGOCIOS. *Difusiones* 2023;25:153-61.

4. López KMG. Inteligencia artificial generativa: Irrupción y desafíos. *Enfoques* 2023;4:57-82.

5. Espinosa RDC, Caicedo-Eraza JC, Londoño MA, Pitre IJ. Inclusive Innovation through Arduino Embedded Systems and ChatGPT. *Metaverse Basic and Applied Research* 2023;2:52-52. <https://doi.org/10.56294/mr202352>.
6. Ferrer-Benítez M. Online dispute resolution: can we leave the initial decision to Large Language Models (LLM)? *Metaverse Basic and Applied Research* 2022;1:23-23. <https://doi.org/10.56294/mr202223>.
7. Toro-Espinoza MF, Montalván-Espinoza JA, Masabanda-Vaca MA. Aplicación de la inteligencia artificial en el aprendizaje universitario. *Reicomunicar* 2023;6:153-72. <https://doi.org/10.46296/rc.v6i12edespoc.0168>.
8. Ruibal-Tavares E, Calleja-López JR, Rivera-Rosas CN, Aguilera-Duarte LJ. Inteligencia artificial en medicina: panorama actual. *REMUS* 2023. <https://doi.org/10.59420/remus.10.2023.178>.
9. T S, Arumugam T, Pandurangan H, Panjaiyan K. Adopción de la Inteligencia Artificial en la Atención Sanitaria: Una perspectiva enfermera. *Salud, Ciencia y Tecnología* 2023;3:510. <https://doi.org/10.56294/saludcyt2023510>.
10. Cano CAG, Castillo VS, Gallego TAC. Unveiling the Thematic Landscape of Generative Pre-trained Transformer (GPT) Through Bibliometric Analysis. *Metaverse Basic and Applied Research* 2023;2:33-33. <https://doi.org/10.56294/mr202333>.
11. Luna GJJ. Study on the impact of artificial intelligence tools in the development of university classes at the school of communication of the Universidad Nacional José Faustino Sánchez Carrión. *Metaverse Basic and Applied Research* 2023;2:51-51. <https://doi.org/10.56294/mr202351>.
12. Castillo-González W, Lepez CO, Bonardi MC. Chat GPT: a promising tool for academic editing. *Data and Metadata* 2022;1:23. <https://doi.org/10.56294/dm202223>.
13. Jiménez JV, Leyva LLL, Leon AM. ChatGPT e inteligencia artificial, señal de alerta para el proceso editorial de revistas médicas. *Revista Cubana de Información en Ciencias de la Salud* 2023;34.
14. Ledo MJV, Olite FMD, Vera IA, Suárez I del RM, Domínguez AMA, Pedro JYP. Chat en la educación médica. *Educación Médica Superior* 2023;37.
15. Carrasco JP, García E, Sánchez DA, Porter E, De La Puente L, Navarro J, et al. ¿Es capaz “ChatGPT” de aprobar el examen MIR de 2022? Implicaciones de la inteligencia artificial en la educación médica en España. *Rev Esp Edu Med* 2023;4. <https://doi.org/10.6018/edumed.556511>.
16. Alfertshofer M, Hoch CC, Funk PF, Hollmann K, Wollenberg B, Knoedler S, et al. Sailing the Seven Seas: A Multinational Comparison of ChatGPT’s Performance on Medical Licensing Examinations. *Ann Biomed Eng* 2023. <https://doi.org/10.1007/s10439-023-03338-3>.
17. Jiménez JV, Gomez EEB, Álvarez PJR. ChatGPT e inteligencia artificial: ¿obstáculo o ventaja para la educación médica superior? *Educación Médica Superior* 2023;37.
18. Colectivo de autores. *Medicina general integral*. Tomo I. Salud y medicina. vol. Vol 1. 4ta ed. La Habana: Editorial Ciencias Medicas; 2022.
19. Gutiérrez-Cirlos C, Carrillo-Pérez DL, Bermúdez-González JL, Hidrogo-Montemayor I, Carrillo-Esper R, Sánchez-Mendiola M. ChatGPT: oportunidades y riesgos en la asistencia, docencia e investigación médica. *GMM* 2023;159:11757. <https://doi.org/10.24875/GMM.230001671>.

#### **FINANCIACIÓN**

Sin financiación.

#### **CONFLICTO DE INTERESES**

No existen conflictos de intereses.

#### **CONTRIBUCIÓN DE LOS AUTORES**

*Conceptualización:* Javier González Argote.

*Investigación:* Javier González Argote, William Castillo-González.

*Metodología:* Javier González Argote.

*Redacción-borrador original:* Javier González Argote, William Castillo-González.

*Revisión y edición de la redacción:* Javier González Argote, William Castillo-González.