

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial (2018, 2019, 2022 y 2024)?

LUIS ADOLFO APOLÍN MONTES¹

Universidad César Vallejo - Perú

JOHNY SATURNINO GARAY SANTISTEBAN²

Universidad Nacional Santiago Antúnez de Mayolo - Perú

Recibido el 05-02-25; primera evaluación el 01-09-25;
segunda evaluación el 01-10-25; aceptado el 03-11-25

RESUMEN

El artículo analiza el desempeño de ChatGPT-4 en las pruebas de ingreso a la Carrera Pública Magisterial en Perú (2018, 2019, 2022 y 2024) mediante un enfoque descriptivo y transversal. Se evaluaron respuestas en Matemáticas, Ciencia y tecnología (CTA) y Comunicación, comparándolas con estándares oficiales. ChatGPT-4 aprobó todas las pruebas, destacando en Ciencia y tecnología con 184 puntos en 2024 y primaria con 173.5 en 2018. Sin embargo, obtuvo los puntajes más bajos en Matemáticas (137 en 2022 y 150 en 2024) y Comunicación (155 en 2024), lo que refleja sus dificultades para el análisis crítico e interpretación contextual. Se recomienda entrenar el modelo en contextos más complejos y rediseñar evaluaciones. ChatGPT-4 tiene potencial como herramienta educativa, siempre que se use éticamente.

Palabras clave: procesamiento de lenguaje natural, evaluación de competencias, educación digital, tecnología educativa, aplicaciones de IA en evaluación.

¹ Docente en la especialidad de Lengua y Literatura por el Instituto Superior Pedagógico Público de Huaraz y licenciado en Periodismo por la Universidad Nacional Santiago Antúnez de Mayolo, Huaraz. Perteneciente al colectivo artístico internacional “Quipa Pacha”. Ha publicado la colección de cuentos “TeZtimonio. El grito de los días” (2015), “Hermano. El silencio de los días” (2017), y la colección de ambos “Epitafio” (2020) por Pandemónium Editorial. Correo electrónico: lapolinm@ucvvirtual.edu.pe ORCID: <https://orcid.org/0009-0003-5854-8116>

² Bachiller en Matemática y licenciado en Matemática por la Universidad Nacional Santiago Antúnez de Mayolo. Maestría en Ciencias de la Educación Superior por la Universidad San Pedro. Doctor en Educación por la Universidad César Vallejo y estudios concluidos de Doctorado en Matemática por la Universidad Nacional del Santa. Con más de 15 años de experiencia en la carrera docente a nivel superior desarrollando cursos del Área de Matemática e investigación. Correo electrónico: jgarays@unasam.edu.pe ORCID: <https://orcid.org/0000-0001-6329-4438>



Is ChatGPT-4 Capable of Passing the Entrance Exams for the Public Teaching Career (2018, 2019, 2022, and 2024)?

ABSTRACT

This article examines the performance of ChatGPT-4 in the entrance examinations for the public teaching career in Peru (2018, 2019, 2022, and 2024) using a descriptive, cross-sectional approach. Responses in Mathematics, Science and Technology (CTA), and Communication were evaluated and compared against official scoring standards. The results indicate that ChatGPT-4 passed all examinations, achieving its highest scores in Science and Technology (184 points in 2024) and in the Primary level (173.5 points in 2018). However, lower performance was observed in Mathematics (137 points in 2022 and 150 points in 2024) and Communication (155 points in 2024), suggesting limitations in critical analysis and contextual interpretation. The findings highlight the need to train language models in more complex and context-sensitive scenarios, as well as to reconsider the design of assessment instruments. Overall, ChatGPT-4 demonstrates potential as an educational support tool, provided its use is guided by ethical principles.

Keywords: natural language processing; competency assessment; digital education; educational technology; AI applications in assessment.

O ChatGPT-4 é capaz de ser aprovado nos exames de admissão à carreira pública docente (2018, 2019, 2022 e 2024)?

RESUMO

Este artigo analisa o desempenho do ChatGPT-4 nos exames de admissão à carreira pública docente no Peru (2018, 2019, 2022 e 2024), adotando uma abordagem descritiva e transversal. As respostas nas áreas de Matemática, Ciência e Tecnologia (CTA) e Comunicação foram avaliadas em comparação com os critérios oficiais de pontuação. Os resultados mostram que o ChatGPT-4 foi aprovado em todos os exames, com destaque para o desempenho em Ciência e Tecnologia (184 pontos em 2024) e no nível Primário (173,5 pontos em 2018). Contudo, observou-se desempenho inferior em Matemática (137 pontos em 2022 e 150 pontos em 2024) e em Comunicação (155 pontos em 2024), o que evidencia limitações relacionadas à análise crítica e à interpretação contextual. Recomenda-se o treinamento do modelo em contextos mais complexos, bem como a reformulação dos instrumentos de avaliação. De modo geral, o ChatGPT-4 apresenta potencial como ferramenta de apoio educacional, desde que utilizado de forma ética.

Palavras-chave: processamento de linguagem natural; avaliação de competências; educação digital; tecnologia educacional; aplicações de IA na avaliação.

1. INTRODUCCIÓN

La inteligencia artificial (IA) ha pasado de ser una tecnología emergente a una presencia cotidiana que impacta múltiples campos, incluida la educación. Su implementación requiere no solo un desarrollo ético, sino también una adecuada alfabetización digital. Autores como Moreno Padilla (2019) señalan que la IA debe ser un apoyo y no un reemplazo del docente, mientras que García-Peñalvo (2023) enfatiza la necesidad de formar a educadores y estudiantes en su uso responsable, es decir, alentando el desarrollo de competencias digitales y pensamiento crítico. Según Urquilla Castañeda (2022), la IA ha transformado el aprendizaje personalizado y la calidad educativa mediante evaluaciones y estrategias inteligentes, aunque no sustituirá al sistema tradicional. Por su parte, Arana (2021) subraya que la interacción humano-máquina es clave para el avance educativo en el siglo XXI, proponiendo un equilibrio entre tecnología, interacción social y nuevas tendencias.

En este contexto, ChatGPT, una herramienta avanzada de generación de texto ha mostrado utilidad en la creación de contenido y apoyo a docentes y estudiantes (Flores, 2023), aunque plantea preocupaciones éticas vinculadas con la integridad académica y presenta limitaciones en el abordaje de temas complejos.

En Perú, las pruebas de ingreso a la Carrera Pública Magisterial (CPM) garantizan la calidad docente al evaluar competencias pedagógicas y disciplinares. Este estudio analiza la capacidad de ChatGPT-4 para responder dichas pruebas en 2018, 2019, 2022 y 2024, y evalúa su desempeño en las áreas de Matemáticas, Ciencia y tecnología (CTA) y Comunicación. Asimismo, examina la precisión de sus respuestas, identifica fortalezas y limitaciones, y reflexiona sobre su potencial educativo y los desafíos éticos de su integración en los procesos de enseñanza y evaluación.

1.1. Revisión de la literatura

ChatGPT, desarrollado por OpenAI, es un ejemplo destacado de inteligencia artificial generativa basada en el modelo *generative pre-trained transformer* (GPT), capaz de producir texto coherente y relevante en conversaciones. Díaz y Ribera (2024) explican que su funcionamiento se apoya en un modelo probabilístico entrenado con grandes volúmenes de datos de internet, respaldado por la infraestructura de Microsoft y los avances en procesamiento del lenguaje natural. Aunque presenta limitaciones, su aplicación en educación ha demostrado beneficios al facilitar la creación de materiales didácticos y personalizar

el aprendizaje (González Díaz, 2018). En la misma línea, Obando Sandoval et al. (2025) concluyen que la inteligencia artificial transforma la educación al personalizar contenidos, optimizar tareas docentes y generar experiencias inmersivas, aunque advierten sobre los riesgos éticos como, por ejemplo, la privacidad de datos y la exclusión digital, subrayando que el programa debe complementar, no reemplazar, la interacción humana.

En diferentes contextos evaluativos, el desempeño de ChatGPT ha mostrado variabilidad. Gencer y Aydin (2023) reportaron éxito en exámenes médicos en turco, señalando que GPT-4 alcanzó el nivel de un estudiante promedio de medicina. Beltozar Clemente et al. (2024) encontraron resultados mixtos en las pruebas de admisión de la Universidad Nacional Mayor de San Marcos y la Universidad Nacional de Ingeniería, con un desempeño destacado en historia universal (91%) pero bajo en lenguaje (55%), lo que evidencia la necesidad de ajustes continuos y un uso ético. Ruiz Mendoza et al. (2024) evaluaron la eficacia de ChatGPT-4 en la creación de ítems de examen frente a diseñadores humanos, y encontraron una alta aceptación en claridad, neutralidad y alineación curricular, lo que resalta su potencial en el diseño educativo. Estos trabajos coinciden en señalar que, aunque la inteligencia artificial puede desenvolverse en evaluaciones de selección, su rendimiento es inconsistente y depende del área de conocimiento.

En el ámbito educativo, se han observado tanto beneficios como limitaciones. Sandoval Hernández et al. (2023) destacan que el uso de *prompts* puede enriquecer la enseñanza, aunque requiere de supervisión ética. Larico Hanco (2023) señala que ChatGPT favorece la personalización del aprendizaje universitario, siempre que se use con responsabilidad para evitar dependencias. Estrada-Araoz (2024) encontró que los estudiantes perciben el programa como un apoyo valioso para la comprensión de temas, pero subraya la necesidad de un uso crítico. En la educación médica, Borrome (2024) documentó buenos resultados en un examen de oftalmología en la Universidad de Buenos Aires, aunque con limitaciones en preguntas que demandaban juicio clínico. De manera complementaria, González Morcillo et al. (2024) mostraron que la técnica IF-AT, combinada con gamificación, aumentó la participación y comprensión universitaria, aunque recomendaron cautela al generalizar los resultados.

La literatura internacional refuerza estas conclusiones. Cerame et al. (2024) señalaron que GPT-4 superó a su predecesor en el examen Médico Interno Residente (MIR), estableciendo un nuevo estándar en educación médica, aunque sin sustituir el juicio clínico humano. Freire (2023) advirtió que, si bien ChatGPT puede asistir en la generación de ideas, es menos confiable en

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

preguntas complejas que requieren pensamiento crítico. Rodríguez-Echevarría (2024) confirmó esta limitación al mostrar que GPT-4 responde con precisión en exámenes teóricos de ingeniería informática, pero enfrenta dificultades en problemas que exigen razonamiento complejo. Por su parte, Sánchez Mendiola (2024) cuestionó la pertinencia de los exámenes tradicionales frente a la capacidad de la IA para generar respuestas automáticas y propuso evaluaciones centradas en competencias prácticas. Finalmente, Galli y Kanobel (2023) insistieron en que el uso de la IA en educación debe ser crítico y ético, pues los sesgos y errores del modelo requieren supervisión humana y alfabetización digital.

En conjunto, la revisión de literatura indica que ChatGPT-4 presenta fortalezas en contextos evaluativos estructurados, pero evidencia debilidades en tareas que exigen razonamiento crítico o interpretación contextual. Esta situación resalta la necesidad de un análisis más profundo de las bases conceptuales que sustentan su aplicación en evaluaciones de selección. Responder con precisión a ítems de opción múltiple no garantiza la posesión de las competencias que dichas pruebas buscan medir, lo cual plantea implicaciones metodológicas y éticas que deben considerarse en entornos formales de examen.

2. METODOLOGÍA

Este estudio empleó un enfoque descriptivo transversal y un diseño no experimental para evaluar la capacidad de ChatGPT-4 de responder correctamente a preguntas de las pruebas de ingreso a la CPM aplicadas por el Ministerio de Educación del Perú (Minedu) en 2018, 2019, 2022 y 2024. Estas pruebas oficiales, previamente validadas mediante revisión de expertos, alineación curricular y estandarización nacional, garantizan su confiabilidad, por lo que no se aplicó un protocolo adicional de validación de ítems.

La variable independiente utilizada fue la formulación de preguntas y, la dependiente, la capacidad de ChatGPT-4 para generar respuestas correctas, medida en términos de precisión, correlación y alineación con los estándares oficiales. La población corresponde a todas las pruebas aplicadas a nivel nacional, y la muestra incluyó ítems representativos de Matemáticas, Ciencia y tecnología (CTA), Comunicación, Educación básica regular (EBR) inicial y primaria.

Las pruebas de 2018 y 2019 se aplicaron en un solo día con variación en el orden de las preguntas, mientras que en 2022 y 2024 se administraron en dos fechas con versiones distintas por área y nivel, lo que garantiza la representatividad y diversidad de la muestra de conocimientos evaluados. A continuación, se detalla la distribución de las pruebas seleccionadas:

Pruebas para el Concurso público de ingreso a la CPM en instituciones educativas públicas de educación básica – 2018 y 2019 (90 ítems)

EBR Matemáticas:

- EBR secundaria Matemática - Forma 1

EBR Ciencia y tecnología (CT):

- EBR secundaria CT - Forma 1

EBR Comunicación:

- EBR secundaria Comunicación - Forma 1

EBR Inicial:

- EBR inicial - Forma 1

EBR Primaria:

- EBR primaria - Forma 1

Prueba para el Concurso público de ingreso a la CPM en instituciones educativas públicas de educación básica – 2022 (75 ítems)

En la prueba de ese año, se aplicaron tres cuadernillos diferenciados para la evaluación de habilidades generales en dos fechas distintas, generando seis formas en total. Aunque todos medían las mismas competencias cognitivas con estructura y formato consistentes, no se estableció una correspondencia directa con los cuadernillos de Conocimientos pedagógicos, curriculares y disciplinares de la especialidad. Ante ello, se optó por una asignación aleatoria para vincular ambos tipos de cuadernillos. Los de habilidades generales evaluaban el análisis de textos, identificación de ideas explícitas e implícitas, estructuración de premisas, reconocimiento de relaciones lógicas, resolución de problemas y razonamiento inductivo y deductivo (Minedu, 2024). La asignación aleatoria permitió garantizar una distribución equitativa de los instrumentos de evaluación y evitó la repetición de combinaciones entre cuadernillos.

Habilidades generales

- Primer día de aplicación: Habilidades generales - Forma 1
- Primer día de aplicación: Habilidades generales - Forma 2
- Primer día de aplicación: Habilidades generales - Forma 3
- Segundo día de aplicación: Habilidades generales - Forma 1
- Segundo día de aplicación: Habilidades generales - Forma 2
- Segundo día de aplicación: Habilidades generales - Forma 3

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

EBR Matemáticas:

- Segundo día de aplicación: EBR secundaria Matemática - Forma 1
- Segundo día de aplicación: EBR secundaria Matemática - Forma 2

EBR Ciencia y tecnología (CT):

- Segundo día de aplicación: EBR secundaria CT - Forma 1³

EBR Comunicación:

- Segundo día de aplicación: EBR secundaria Comunicación – Forma 1
- Segundo día de aplicación: EBR secundaria Comunicación – Forma 2

EBR Inicial:

- Segundo día de aplicación: EBR inicial - Forma 1⁴

EBR Primaria:

- Primer día de aplicación: EBR primaria - Forma 1⁵

Tabla 1. Cuadro de distribución aleatorio de cuadernillos por fecha

Día de aplicación	Habilidades generales	Área
Primer día	Habilidades generales - Forma 1	EBR primaria - Forma 1
Primer día	Habilidades generales - Forma 3	No asignada
Segundo día	Habilidades generales - Forma 3	EBR secundaria Matemática - Forma 1
Segundo día	Habilidades generales - Forma 3	EBR secundaria Matemática - Forma 2
Segundo día	Habilidades generales - Forma 2	EBR secundaria CT - Forma 1
Segundo día	Habilidades generales - Forma 3	EBR secundaria Comunicación - Forma 1
Segundo día	Habilidades generales - Forma 1	EBR secundaria comunicación - Forma 2
Segundo día	Habilidades generales - Forma 2	EBR inicial - Forma 1

³ La forma 1 aplicada el mismo día para EBR secundaria fue exactamente la misma que la forma 2.

⁴ La forma 1 aplicada el mismo día para EBR inicial fue exactamente la misma que la forma 2.

⁵ La forma 1 aplicada el mismo día para EBR primaria fue exactamente la misma que la forma 2.

Prueba para el concurso público de ingreso a la CPM en instituciones educativas públicas de educación básica – 2024 (75 ítems)

En la prueba de este año, se aplicaron seis cuadernillos diferenciados para la evaluación de habilidades generales en dos fechas distintas. Aunque todos los cuadernillos medían las mismas competencias cognitivas con una estructura y formato consistentes, no se estableció correspondencia directa con los cuadernillos de Conocimientos pedagógicos, curriculares y disciplinares de la especialidad, ni en sus fechas de aplicación. Ante ello, se optó por una asignación aleatoria para su vinculación. Los cuadernillos de habilidades generales evaluaban el análisis de textos, identificación de ideas explícitas e implícitas, estructuración de premisas, reconocimiento de relaciones lógicas, resolución de problemas y razonamiento inductivo y deductivo (Minedu, 2024). Este método garantizó una distribución equitativa de los instrumentos de evaluación y evitó la repetición de combinaciones entre cuadernillos.

EBR Matemáticas:

- Primer día de aplicación: EBR secundaria Matemática

EBR Ciencia y tecnología (CT):

- Primer día de aplicación: EBR secundaria CT

EBR Comunicación:

- Segundo día de aplicación: EBR secundaria Comunicación

EBR Inicial:

- Primer día de aplicación: EBR inicial

EBR primaria:

- Segundo día de aplicación: EBR primaria

Tabla 2. *Cuadro de distribución aleatorio de cuadernillos por fecha*

Día de aplicación	Habilidades generales	Área
Segundo día	Habilidades generales (HG01-11)	EBR primaria
Primer día	Habilidades generales (HG03-13)	EBR secundaria Matemática
Primer día	Habilidades generales (HG07-23)	EBR secundaria CT
Segundo día	Habilidades generales (HG05-21)	EBR secundaria Comunicación
Primer día	Habilidades generales (HG02-12)	EBR inicial

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

Para el estudio se incluyeron las pruebas de ingreso a la CPM de Matemáticas, Ciencia y tecnología (CTA), Comunicación, EBR inicial y EBR primaria, aplicadas a nivel nacional en 2018, 2019, 2022 y 2024. También se seleccionaron pruebas de ambos días de aplicación (forma 1 y forma 2) para asegurar la representatividad de la muestra. Se excluyeron aquellas pruebas de áreas no especificadas en los criterios de inclusión.

La recolección de datos se realizó mediante recopilación documental de las pruebas de ingreso a la CPM aplicadas a nivel nacional en 2018, 2019, 2022 y 2024, obtenidas de fuentes oficiales del Minedu y otras entidades responsables. Digitalizadas y almacenadas en una base de datos para su análisis, las preguntas se ingresaron manualmente en GPT-4 para asegurar la precisión de la información. Las respuestas generadas fueron documentadas junto con metadatos relevantes (como el tiempo de respuesta y nivel de confianza), y luego comparadas con las respuestas correctas mediante análisis cualitativo y cuantitativo. La recolección se estructuró en un formulario diseñado para capturar de manera sistemática cada respuesta, incluyendo campos para identificar la prueba, el área de conocimiento, la pregunta, la opción elegida por GPT-4, la respuesta oficial correcta, la evaluación de precisión y un apartado para observaciones adicionales, lo que permitió una evaluación precisa del desempeño de GPT-4 en relación con los estándares oficiales.

El trabajo de campo consistió en replicar las condiciones del examen en un entorno controlado. Las preguntas seleccionadas se ingresaron en ChatGPT-4, y las respuestas generadas por el modelo fueron comparadas con las claves oficiales proporcionadas por el Minedu. Este procedimiento posibilitó la observación y registro del comportamiento del modelo frente a los distintos tipos de preguntas y áreas evaluadas.

Así mismo, se llevó a cabo un análisis de datos que incluyó tanto un enfoque cuantitativo como cualitativo. En el análisis cuantitativo, se calculó el porcentaje de respuestas correctas en cada área y año para medir la precisión del modelo. Además, el análisis cualitativo identificó los patrones de desempeño, así como las fortalezas y limitaciones del modelo en diferentes contextos evaluativos. Estos resultados proporcionarán una base sólida para futuras investigaciones sobre la aplicabilidad de la inteligencia artificial en procesos educativos y evaluativos.

Además del análisis cuantitativo de aciertos y errores, se llevó a cabo un análisis cualitativo de las respuestas generadas por ChatGPT-4. Aunque las pruebas estaban estructuradas en formato de opción múltiple, el modelo no solo seleccionó alternativas, sino que en diversos casos produjo explicaciones o

razonamientos intermedios. Estos elementos permitieron examinar la calidad del proceso de respuesta más allá de la coincidencia con la clave oficial.

En primer lugar, la precisión se valoró verificando la correspondencia entre la alternativa elegida por el modelo y la respuesta correcta establecida en los exámenes, así como la exactitud de la explicación cuando esta era proporcionada. La relevancia se analizó considerando si la respuesta y su justificación se relacionaban directamente con la pregunta planteada, descartando aquellas en las que ChatGPT-4 ofrecía información tangencial o no vinculada con el contenido evaluado. La coherencia se evaluó en términos de consistencia interna, contrastando la opción seleccionada con la argumentación dada; se identificaron casos en los que el modelo podía elegir la alternativa correcta, pero fundamentarla con razonamientos contradictorios o insuficientes. Por último, la correlación y alineación se midieron en función de la correspondencia entre las respuestas del modelo y los criterios de evaluación establecidos por el Minedu, observando si mantenían un patrón de razonamiento compatible con el enfoque pedagógico esperado en cada área: resolución lógica en Matemática, interpretación contextual en Comunicación y aplicación conceptual en Ciencia y tecnología.

Este procedimiento cualitativo permitió detectar patrones de desempeño en el modelo, diferenciando entre errores derivados de confusiones conceptuales, fallas de razonamiento lógico, elecciones aparentemente aleatorias o dificultades de interpretación contextual. De este modo, se obtuvo una comprensión más profunda de las fortalezas y limitaciones de ChatGPT-4 en contextos evaluativos, complementando la información proporcionada por los porcentajes de aciertos.

Finalmente, cabe señalar algunas limitaciones metodológicas de este estudio. Al tratarse de un diseño no experimental y transversal, los resultados no permiten establecer relaciones causales, sino únicamente descriptivas. Además, la investigación se basó en la recopilación documental y en la digitalización manual de los ítems, lo que podría introducir sesgos en el registro. Aunque la muestra seleccionada fue representativa, no incluyó la totalidad de las especialidades evaluadas en las pruebas de la CPM, limitando la generalización de los hallazgos. Del mismo modo, las condiciones de aplicación fueron simuladas en un entorno controlado, sin considerar factores humanos como la presión del tiempo o el contexto sociocultural. Por último, los resultados corresponden únicamente a GPT-4 en la versión utilizada y no necesariamente a otras versiones de IA, por lo que se recomienda prudencia en la extrapolación de los hallazgos.

2.1. Procedimientos

El proceso de investigación comenzó con la selección de pruebas representativas de las pruebas de ingreso a la CPM de los años 2018, 2019, 2022 y 2024, de acuerdo con los criterios especificados en el apartado de población de estudio. Estos exámenes fueron elegidos para abarcar distintas áreas evaluativas y proporcionar un marco robusto para la prueba de ChatGPT-4.

Una vez seleccionadas las pruebas, se ingresaron en ChatGPT-4 para generar respuestas automáticas, y se permitió que la inteligencia artificial replicara el proceso de respuesta en un entorno controlado. Este procedimiento se enfocó en observar el comportamiento del modelo frente a diferentes tipos de preguntas incluidas en cada prueba.

Las respuestas de ChatGPT-4 se evaluaron en función de la precisión, al comparar cada opción seleccionada con las respuestas correctas oficiales de los exámenes. Dado que estos son de opción múltiple, el análisis se centró en verificar si la respuesta del modelo coincidía con la opción correcta proporcionada por el Minedu. Además, se verificó el cumplimiento de estándares, asegurando que las respuestas elegidas por ChatGPT-4 estuvieran alineadas con los criterios de evaluación establecidos para cada examen.

Específicamente, se consideraron algunas particularidades en las pruebas. Las preguntas de Comprensión lectora y Razonamiento lógico de los años 2018 y 2019 fueron idénticas, mientras que las preguntas de Conocimientos curriculares y pedagógicos de la especialidad variaron. En contraste, las pruebas de ingreso a la CPM de los años 2022 y 2024, aplicados en fechas distintas, incluyeron variaciones tanto en la subprueba de Habilidades Generales como en las preguntas de Conocimientos pedagógicos, curriculares y disciplinares de la especialidad. Estas especificidades permitieron realizar una evaluación más precisa y contextualizada del rendimiento de ChatGPT-4.

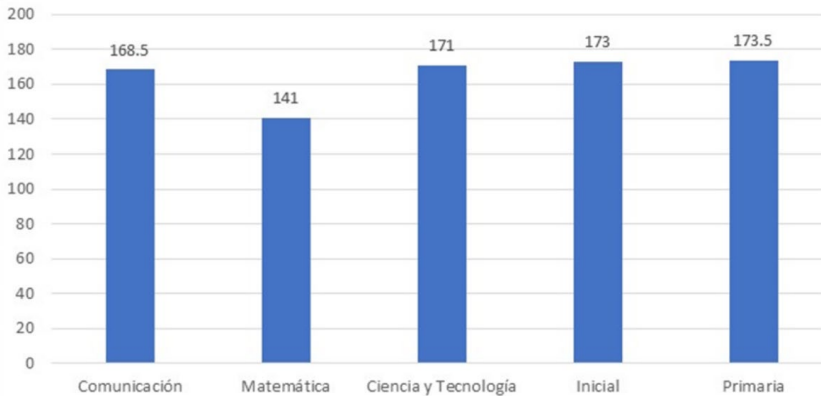
3. RESULTADOS

3.1. Evaluación del rendimiento de ChatGPT-4 en la prueba de ingreso a la CPM (2018)

Como se observa en la Figura 1, los resultados del desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2018 evidencian diferencias notables entre áreas. El modelo alcanzó el puntaje más alto en el área de primaria (173.5 puntos), seguido de inicial (173) y Ciencia y tecnología (171), lo que refleja un desempeño consistente en dominios de contenido general y pedagógico.

En contraste, el puntaje más bajo se registró en Matemática (141), lo que sugiere limitaciones del modelo en la resolución de problemas numéricos o lógicos. Estos resultados confirman que ChatGPT-4 muestra mayor precisión en áreas que demandan comprensión conceptual más que razonamiento formal.

Figura 1. Resultados del desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2018 por áreas y niveles

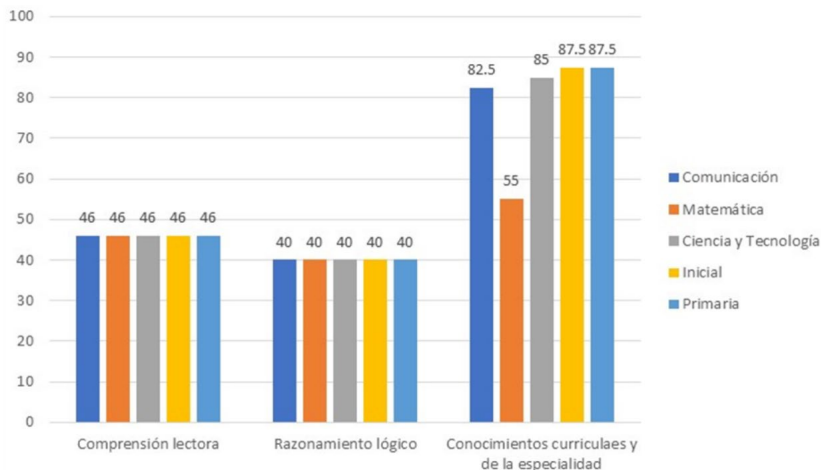


Nota. La Figura 1 muestra los puntajes promedio obtenidos por ChatGPT-4 en las áreas evaluadas en 2018. El puntaje mínimo aprobatorio fue de 120 puntos.

Como se aprecia en la Figura 2, los resultados comparativos revelan una marcada variación entre las subpruebas de Conocimientos curriculares y de la especialidad, donde el área de Matemática registra el rendimiento más bajo (55 puntos), en contraste con los puntajes superiores de Comunicación (82.5) y primaria (87.5). Este patrón evidencia una diferencia sustantiva en la capacidad de ChatGPT-4 para abordar tareas de naturaleza conceptual frente a aquellas que requieren razonamiento lógico formal. Asimismo, los puntajes obtenidos en Comprensión lectora y Razonamiento lógico se mantienen uniformes (46 y 40 puntos, respectivamente), lo que indica una consistencia del modelo ante preguntas compartidas entre las áreas.

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

Figura 2. Resultados comparativos de evaluación por subpruebas, áreas y niveles

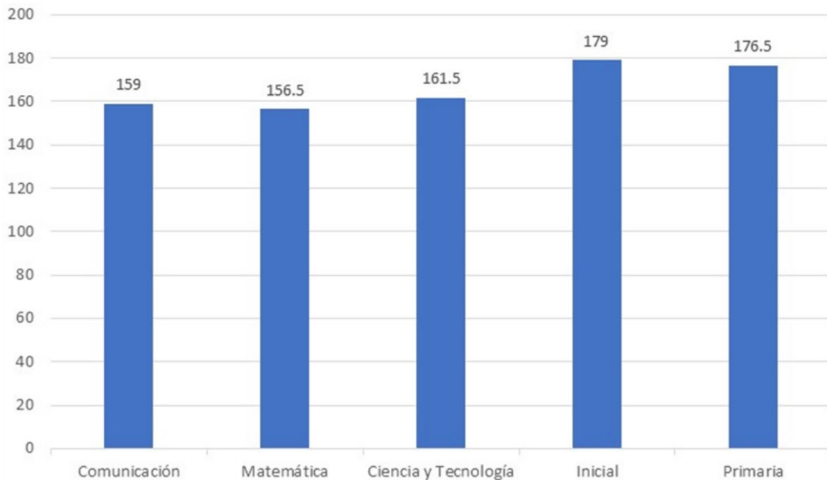


Nota. La Figura 2 presenta la comparación de puntajes por subpruebas, áreas y niveles. Se destaca el menor desempeño en Matemática y la uniformidad de resultados en Comprensión Lectora y Razonamiento Lógico.

3.2. Evaluación del rendimiento de ChatGPT-4 en la prueba de ingreso a la CPM (2019)

Como se muestra en la Figura 3, los resultados del desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2019 evidencian una tendencia similar a la observada en el año anterior, aunque con ligeras variaciones en los puntajes promedio por área. El modelo obtuvo su mejor rendimiento en inicial (179 puntos) y primaria (176.5 puntos), lo que confirma su mayor precisión en contenidos generales y pedagógicos. En contraste, el área de Matemática presentó nuevamente el puntaje más bajo (156.5), seguida de Comunicación (159), lo que sugiere persistentes dificultades en tareas que requieren razonamiento formal y comprensión lingüística avanzada. Estos resultados consolidan el patrón de desempeño diferenciado según el tipo de competencia evaluada.

Figura 3. Resultados del desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2019 por áreas y niveles

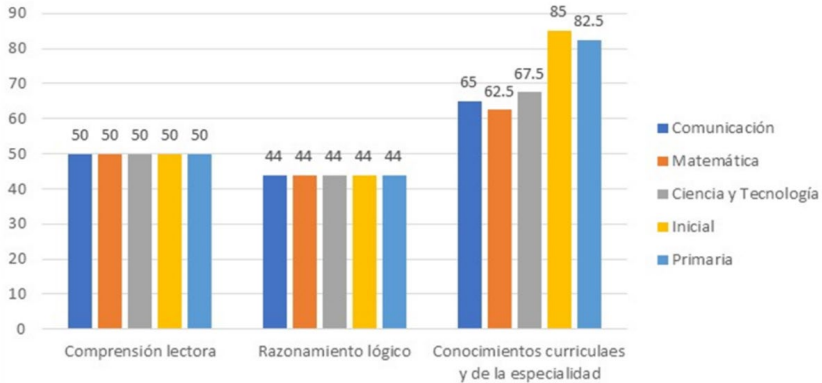


Nota. La Figura 3 presenta los puntajes promedio obtenidos por ChatGPT-4 en las áreas evaluadas en 2019. El puntaje mínimo aprobatorio fue de 120 puntos.

Como se observa en la Figura 4, los resultados comparativos por subpruebas en la evaluación de 2019 muestran diferencias notables en el tipo de conocimiento evaluado. Las puntuaciones en Comprensión Lectora y Razonamiento Lógico se mantienen constantes (50 y 44 puntos, respectivamente) en todas las áreas, lo que evidencia una homogeneidad en los ítems compartidos. En cambio, en la subprueba de Conocimientos curriculares y de la especialidad, se aprecian variaciones significativas: primaria (85 puntos) e inicial (82.5 puntos) presentan los valores más altos, mientras que Matemática (62.5 puntos) y Ciencia y tecnología (67.5 puntos) muestran un rendimiento menor. Este comportamiento sugiere que ChatGPT-4 mantiene un mejor desempeño en contenidos pedagógicos y generales frente a los de razonamiento abstracto o formal.

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

Figura 4. Resultados comparativos de evaluación por subpruebas, áreas y niveles

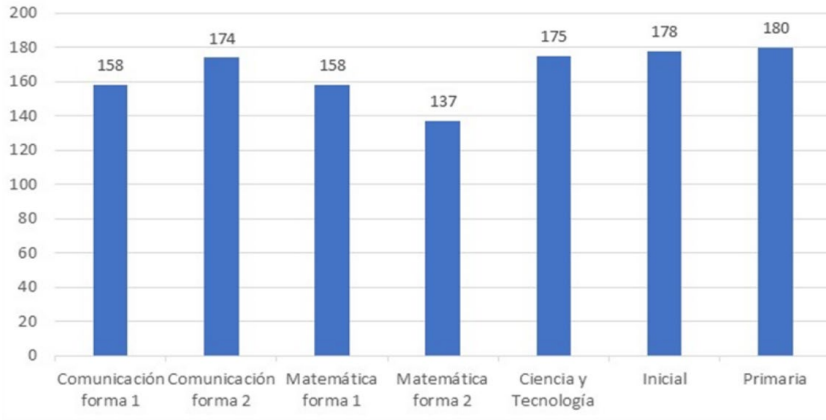


Nota. La Figura 4 muestra los resultados comparativos por subpruebas, áreas y niveles, destacando la uniformidad en Comprensión lectora y Razonamiento lógico, y las variaciones en conocimientos curriculares y de la especialidad.

3.4. Evaluación del rendimiento de ChatGPT-4 en la prueba de ingreso a la CPM (2022)

Como se aprecia en la Figura 5, los resultados del desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2022 muestran una tendencia general de aprobación en todas las áreas evaluadas, aunque con diferencias significativas entre ellas. El modelo alcanzó los puntajes más altos en primaria (180 puntos) e inicial (178 puntos), seguidos de Ciencia y tecnología (175 puntos), lo que evidencia una sólida comprensión en dominios de conocimiento general y pedagógico. En contraste, los resultados más bajos se registraron en Matemática forma 2 (137 puntos) y Comunicación forma 1 (158 puntos), lo que revela las limitaciones del modelo en la resolución de problemas cuantitativos y en tareas de interpretación lingüística. Estos resultados confirman la variabilidad del rendimiento de ChatGPT-4 según el tipo y complejidad de los ítems evaluados.

Figura 5. Resultados del desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2022 por áreas y niveles

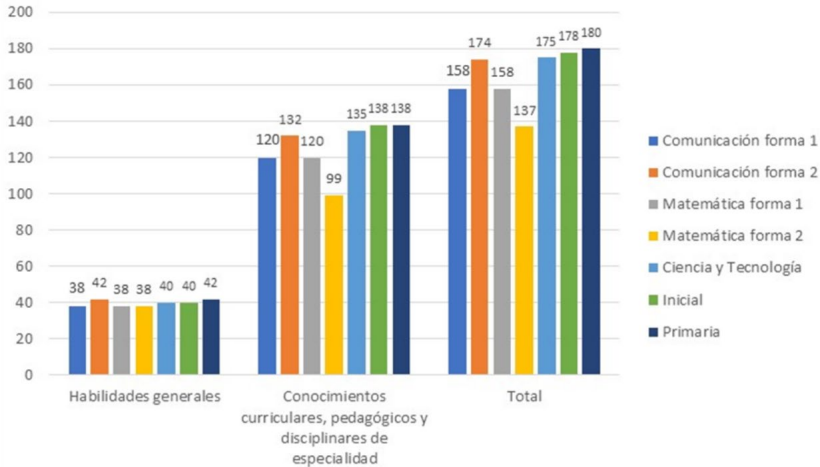


Nota. La Figura 5 presenta los puntajes promedio obtenidos por ChatGPT-4 en las distintas áreas y formas evaluadas en 2022. El puntaje mínimo de aprobación total fue de 110 puntos, y el de la subprueba de Conocimientos pedagógicos, curriculares y disciplinares fue de 84 puntos.

Como se muestra en la Figura 6, el desempeño de ChatGPT-4 en las distintas subpruebas evidencia diferencias notorias entre las áreas evaluadas. En Habilidades generales, los puntajes se mantienen homogéneos, con un rango estrecho entre 38 y 42 puntos, lo que sugiere un desempeño equilibrado en tareas de razonamiento general. Sin embargo, en la subprueba de Conocimientos curriculares, pedagógicos y disciplinares de la especialidad, se aprecia una mayor dispersión: el área de Matemática forma 2 registra el puntaje más bajo (99), mientras que primaria (132) y Ciencia y tecnología (138) obtienen valores superiores. En el puntaje total, se destacan nuevamente las áreas de Primaria (180) e Inicial (178), confirmando el patrón de fortaleza del modelo en dominios conceptuales y pedagógicos frente a los lógico-numéricos.

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

Figura 6. Resultados comparativos de evaluación por subpruebas, áreas y niveles

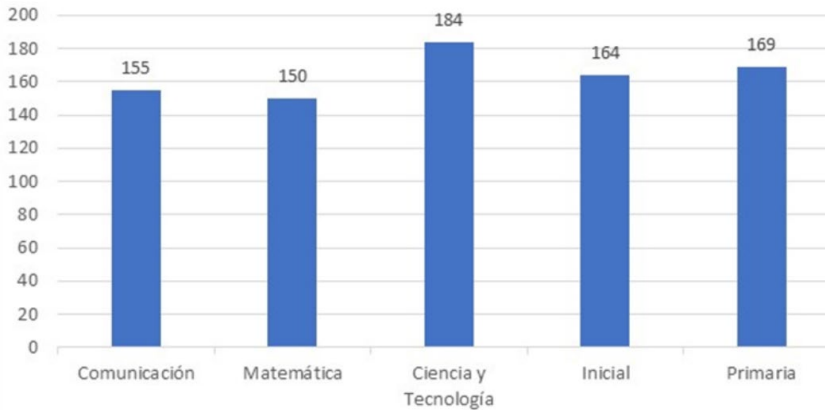


Nota. La Figura 6 presenta los resultados comparativos por subpruebas, áreas y niveles. Se observa homogeneidad en Habilidades generales y mayores diferencias en Conocimientos curriculares, pedagógicos y disciplinares, con el menor desempeño en Matemática forma 2.

3.5. Evaluación del rendimiento de ChatGPT-4 en la prueba de ingreso a la CPM (2024)

Como se observa en la Figura 7, el desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2024 evidencia un rendimiento particularmente alto en el área de Ciencia y tecnología (184 puntos), lo que sugiere una sólida competencia en la comprensión y resolución de problemas con base científica. En segundo lugar, destacan primaria (169 puntos) e inicial (164 puntos), que reflejan una consistencia en dominios conceptuales de carácter pedagógico y general. En contraste, las áreas de Comunicación (155 puntos) y Matemática (150 puntos) presentan los valores más bajos, lo que indica que el modelo mantiene sus limitaciones en tareas que requieren razonamiento lógico formal o interpretación lingüística compleja.

Figura 7. Resultados del desempeño de ChatGPT-4 en la prueba de ingreso a la CPM 2024 por áreas y niveles

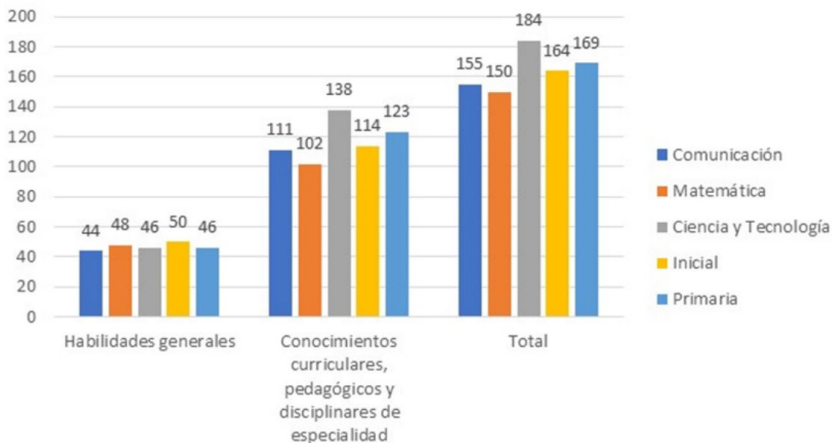


Nota. La Figura 7 muestra los puntajes promedio obtenidos por ChatGPT-4 en las áreas evaluadas en 2024. Se destaca el mejor desempeño en Ciencia y Tecnología y los menores puntajes en Comunicación y Matemática.

Como se aprecia en la Figura 8, los resultados comparativos por subpruebas en la evaluación de 2024 muestran un comportamiento diferenciado según el tipo de conocimiento medido. En Habilidades generales, los puntajes se mantienen equilibrados, oscilando entre 44 y 50 puntos, lo que evidencia un rendimiento estable en razonamiento general. En cambio, en la subprueba de Conocimientos curriculares, pedagógicos y disciplinares de la especialidad, se observan variaciones más amplias: Ciencia y tecnología (138 puntos) y Comunicación (123 puntos) presentan los valores más altos, mientras que Matemática (102 puntos) alcanza el puntaje más bajo. En el total general, los resultados más elevados corresponden nuevamente a Ciencia y tecnología (184 puntos) y primaria (169 puntos), confirmando el patrón de superioridad del modelo en áreas conceptuales y de aplicación pedagógica frente a los campos numéricos.

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

Figura 8. Resultados comparativos de evaluación por subpruebas, áreas y niveles



Nota. La Figura 8 presenta los puntajes comparativos por subpruebas, áreas y niveles. Se observan diferencias notorias entre las categorías evaluadas, con mejores resultados en Comunicación y Ciencia y tecnología.

3.6. Análisis cualitativo de las respuestas generadas por ChatGPT-4

El análisis cualitativo permitió identificar patrones recurrentes en la forma en que el modelo razona, justifica y selecciona sus respuestas. A partir de la revisión de más de 400 ítems se construyeron tres categorías principales, cada una con subniveles representativos.

Tabla 3. *Categorías emergentes del análisis cualitativo de las respuestas de ChatGPT-4*

Categoría principal	Descripción	Ejemplo textual o patrón observado
1. Razonamiento lógico-coherente	El modelo selecciona la alternativa correcta y sustenta su decisión mediante inferencias adecuadas y consistentes con el contenido del ítem. Este tipo de respuesta fue más frecuente en Ciencia y tecnología y en primaria.	“La opción correcta es la B, porque la energía potencial depende de la altura y la gravedad; al aumentar la altura, la energía también aumenta”.
2. Respuestas correctas con justificación parcial	ChatGPT-4 acierta la alternativa, pero su explicación resulta ambigua, redundante o incompleta. Se detectó principalmente en Comunicación e inicial, donde el modelo tiende a generalizar sin argumentar a profundidad.	“La respuesta es la C porque el texto habla de sentimientos, y esa es la idea principal”.
3. Errores conceptuales o de interpretación contextual	El modelo elige una alternativa incorrecta debido a una lectura superficial del enunciado, confusión terminológica o falta de razonamiento formal. Fue común en Matemática y en algunos ítems de análisis crítico en Comunicación.	“La respuesta es la A ya que se trata de una ecuación cuadrática” (cuando el problema correspondía a una relación lineal).

Del análisis se desprenden algunos patrones transversales:

1. ChatGPT-4 muestra consistencia formal y redacción precisa, incluso en respuestas erradas, lo que puede generar una impresión de acierto.
2. Tiende a sustituir el razonamiento lógico por descripciones literales cuando los ítems requieren abstracción o contextualización.
3. En ítems de comprensión lectora, se observó que el modelo identifica correctamente las ideas explícitas, pero omite inferencias implícitas o relaciones pragmáticas, lo que explica su menor desempeño en el área de Comunicación.
4. En Matemática, los errores provienen de la interpretación de los datos o de la elección del procedimiento, no de la sintaxis del lenguaje matemático.
5. En cambio, en Ciencia y tecnología, el modelo mantiene una alta coherencia causal, aplicando definiciones y principios de forma correcta, lo que explica sus puntajes más altos.

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

En conjunto, las evidencias cualitativas confirman que ChatGPT-4 razona con coherencia y precisión en contextos estructurados, pero presenta limitaciones cuando el ítem demanda análisis crítico, contextual o inferencial. Estas observaciones complementan los resultados numéricos y aportan una visión más completa de la naturaleza de su desempeño evaluativo.

4. DISCUSIÓN

Los hallazgos muestran un desempeño variable de ChatGPT-4 según el área evaluada. En Ciencia y tecnología alcanzó los puntajes más altos, lo que sugiere eficacia en problemas bien definidos y de carácter técnico. En Matemáticas, aunque superó el mínimo de aprobación, sus resultados fueron los más bajos de manera constante, lo que refleja unas limitaciones para el razonamiento formal y en la resolución de problemas complejos. En Comunicación se observó un descenso mayor en el rendimiento, lo que apunta a dificultades para interpretar matices lingüísticos y contextuales.

Estos patrones dialogan con estudios previos. Gencer y Aydin (2023) reportan un desempeño exitoso de GPT-4 en un examen de cirugía torácica en Turquía, lo que resalta su eficacia en tareas técnicas y objetivas. Por su parte, Gonzáles Díaz (2018) señala que la inteligencia artificial puede optimizar materiales didácticos y personalizar el aprendizaje en disciplinas estructuradas, siempre que el contenido sea claro y con menor nivel de ambigüedad. Los resultados de este estudio muestran que, incluso en áreas estructuradas como Matemáticas, el modelo enfrenta dificultades cuando los ítems exigen abstracción formal o encadenamientos lógicos más complejos.

En Comunicación las limitaciones se relacionan con el razonamiento crítico y la interpretación contextual. Esto se relaciona con lo mencionado por Rodríguez-Echevarría (2024), quien indicaba que GPT-4 responde con precisión en preguntas teóricas de ingeniería informática, pero falla en aquellas que requieren razonamiento complejo. A su vez, estas limitaciones también se relacionan con estudios como el de Freire (2023), quien también advierte que, aunque ChatGPT es útil en la generación de ideas, resulta menos confiable ante preguntas que demandan un juicio crítico. La variabilidad de resultados entre los diferentes años de aplicación sugiere que el modelo es sensible a cambios en la formulación de los ítems, lo cual limita su adaptabilidad. Este aspecto coincide con lo señalado por Sandoval Hernández et al. (2023), quienes recomiendan un uso ético y supervisado de la inteligencia artificial, y con Estrada-Araoz (2024), quien enfatiza la necesidad de emplear ChatGPT de manera crítica para evitar dependencias y asegurar la precisión de la información.

Autores como Cerame et al. (2024) observaron que GPT-4 superó a su predecesor en el examen MIR, con altos niveles de precisión en preguntas técnicas, aunque sin sustituir el juicio clínico humano, lo que guarda relación con la menor eficacia observada en el área de Comunicación de este estudio; mientras que Ruiz Mendoza et al. (2024) destacan la capacidad de ChatGPT para crear ítems de examen de alta calidad, siempre que exista revisión humana que asegure pertinencia y alineación con el contexto evaluativo. Es necesario aclarar que esta habilidad no debe confundirse con la capacidad de responder correctamente a ítems ya elaborados. La presente investigación se concentra en la resolución de las pruebas de la CPM, mientras que el diseño de exámenes exige ciertos criterios pedagógicos, la validez del contenido y una adecuación curricular. Por esta razón, aunque ambas funciones pueden verse como complementarias, los resultados no permiten concluir que un buen desempeño en la resolución de preguntas se traduzca en la capacidad para generarlas.

En conjunto, los resultados de este estudio ofrecen implicaciones prácticas y éticas relevantes. Para la comunidad docente, ChatGPT-4 puede ser un recurso de apoyo en la preparación de materiales y en la retroalimentación personalizada, siempre que se emplee con una formación crítica y ética que evite dependencias indebidas. Para los diseñadores de evaluaciones, los hallazgos resaltan la necesidad de elaborar ítems que midan análisis, razonamiento y creatividad, menos susceptibles de ser resueltos mecánicamente por modelos de inteligencia artificial. En ambos casos se requiere supervisión humana y capacitación para garantizar la validez de las evaluaciones y la integridad de los procesos educativos.

5. CONCLUSIONES

Los resultados del estudio muestran que ChatGPT-4 es capaz de aprobar las pruebas de ingreso a la CPM, cumpliendo con los requisitos mínimos. Su desempeño es más efectivo en áreas estructuradas y factuales como Ciencia y tecnología y EBR primaria, mientras que en Matemática y Comunicación los puntajes bajos reflejan las limitaciones de la IA en problemas complejos y en interpretación lingüística avanzada.

El rendimiento varió significativamente entre años y áreas, lo que indica que su efectividad depende de la formulación y el estilo de las preguntas. Para mejorar su desempeño, se recomienda optimizar el modelo en áreas críticas mediante entrenamiento específico con contextos complejos y realizar evaluaciones continuas con supervisión humana para verificar estándares académicos.

Paralelamente, los diseñadores de exámenes pueden rediseñar las evaluaciones para incluir preguntas que fomenten el pensamiento crítico, más allá

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

de respuestas automáticas. La Figura 5, que desglosa los resultados por áreas y niveles, muestra una variabilidad en los puntajes que plantea interrogantes sobre las posibles causas, como la percepción de dificultad o inconsistencias en la redacción, lo que resalta la necesidad de un análisis exhaustivo para garantizar equivalencia y coherencia en las evaluaciones.

Por último, se sugiere integrar la IA como una herramienta complementaria, establecer pautas para su uso ético y fomentar la alfabetización digital entre estudiantes y docentes para facilitar un uso crítico y responsable, que reduzca su dependencia a modelos de IA y promueva una correcta aplicación que complemente (pero no reemplace) el desarrollo de competencias clave.

REFERENCIAS BIBLIOGRÁFICAS

- Arana, C. (2021). Inteligencia Artificial Aplicada a la Educación: Logros, Tendencias y Perspectivas. *INNOVA UNTREF Revista Argentina de Ciencia y Tecnología*, 1(7). <https://www.revistas.untref.edu.ar/index.php/innova/article/view/1107>
- 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), 50-60. <https://doi.org/10.3991/ijep.v14i1.46787>
- Borrome, R. (2024). ChatGPT-4: su desempeño en un examen final de la carrera de médico especialista en oftalmología de la Universidad de Buenos Aires. *Revista OCE*, 10(3), 50-64. <https://www.revistaoce.com/index.php/revista/article/view/286/483>
- Cerame, Á., Juaneda, J., Estrella-Porter, P., de la Puente, L., Navarro, J., García, E., Sánchez, D. A., & Carrasco, J. P. (2023). ¿Es capaz GPT-4 de aprobar el MIR 2023? Comparativa entre GPT-4 y ChatGPT-3 en los exámenes MIR 2022 y 2023. *Revista Española de Educación Médica*, (2), 1-10. <https://revistas.um.es/edumed/article/view/604091/358071>
- Díaz, O., & Ribera, M. (2024). Introducción. En M. Ribera & O. Díaz Montesdeoca (Coords.), *ChatGPT y educación universitaria: posibilidades y límites de ChatGPT como herramienta docente* (pp. 7-23). Octaedro; Institut de Desenvolupament Professional, Universitat de Barcelona. <https://diposit.ub.edu/dspace/bitstream/2445/206141/1/9788410054011.pdf>
- Estrada-Araoz, E. G., Paredes-Valverde, Y., Quispe-Herrera, R., Gallegos-Ramos, N. A., Rivera-Mamani, F. A., & Romaní-Claros, A. (2024). Investigating the attitude of university students towards the use of ChatGPT as a learning resource. *Data and Metadata*, 3, 268. <https://doi.org/10.56294/dm2024268>

- Flores, V. (2023, 11 de noviembre). ¿Puede ChatGPT resolver correctamente los exámenes online? *Proctorizer*. Recuperado de <https://proctorizer.com/puede-chatgpt-resolver-correctamente-los-examenes-online/>
- Freire, D. (2023, 14 de mayo). La gente cree que puede usar ChatGPT para aprobar un examen universitario. Realmente no es tan sencillo. *La Vanguardia*. <https://www.lavanguardia.com/andro4all/tecnologia/la-gente-cree-que-puede-usar-chatgpt-para-aprobar-un-examen-universitario-realmente-no-es-tan-sencillo>
- Galli, M.G., & Kanobel, M.C. (2023). ChatGPT en Educación Superior: explorando sus potencialidades y sus limitaciones. *Revista Educación Superior y Sociedad*, 35(2), 174-195. <https://doi.org/10.54674/ess.v34i2.815>.
- García-Peñalvo, F. J. (2023). La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: Disrupción o Pánico. *Education in the Knowledge Society*, 24(1). <https://doi.org/10.14201/eks.31279>
- Gencer, A., & Aydin, S. (2023). Can ChatGPT pass the thoracic surgery exam? *The American Journal of the Medical Sciences*, 366(4), 291-295. <https://doi.org/10.1016/j.amjms.2023.08.001>
- González Díaz, E. (2018). *Aplicación para la creación de contenido educativo utilizando ChatGPT* [Trabajo de fin de grado, Universidad Rey Juan Carlos]. BURJC Digital. <https://burjcdigital.urjc.es/bitstream/handle/10115/36769/2023-24-ETSII-A-2243-2243059-e.gonzalezd.2018-MEMORIA.pdf?sequence=1&isAllowed=y>
- González Morcillo, C., Martínez Cid, S., Albusac Jiménez, J. A., Vallejo Fernández, D., Sánchez Sobrino, S., & Castro Sánchez, J. J. (2024). Adaptación de la técnica de evaluación inmediata IF-AT y gamificación grupal: diseño de preguntas efectivas en la era de ChatGPT. *Actas de las Jenui*, 9, 59-66. https://rua.ua.es/dspace/bitstream/10045/144622/1/JENUI_2024_008.pdf
- Larico Hanco, R. (2023). Impact of ChatGPT generative artificial intelligence on university teaching. *Chakiñan: Revista de Ciencias Sociales y Humanidades*, 14(1). <https://doi.org/10.37135/chakinan.001.14.07>
- Ministerio de Educación del Perú. (2024). Prueba nacional de nombramiento docente 2024. Evaluación Docente *PerúEduca*. <https://evaluaciondocente.perueduca.pe/nombramiento24/prueba-nacional/>
- Moreno Padilla, R. D. (2019). La llegada de la inteligencia artificial a la educación. *RITI Journal*, 7(14), 260-270. <https://doi.org/10.36825/RITI.07.14.022>
- Obando Sandoval, C. A., Añazco Naula, J. E., Toala Timaran, F. C., Livingston Pardo, P. M., Moran Valencia, S. F., & Diaz Mendoza, A. C. (2025). El Impacto de la Inteligencia Artificial Dentro del Proceso de Enseñanza Aprendizaje de la Educación. *Ciencia Latina Revista Científica Multidisciplinar*, 9(1), 7407-7424. https://doi.org/10.37811/cl_rcm.v9i1.16416

¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial?

- Rodriguez-Echeverría, R., Gutiérrez, J. D., Conejero, J. M., & Prieto, Á. E. (2024). Analysis of ChatGPT performance in computer engineering exams. *IEEE-RITA: Revista Iberoamericana de Tecnologías del Aprendizaje*. <https://doi.org/10.1109/RITA.2024.3381842>.
- Ruiz Mendoza, K. K., Pedroza Zúñiga, L. H., & López García, A. Y. (2023). Creación y jueceo de ítems: ChatGPT como diseñador y juez. *Trabalhos em Lingüística Aplicada*, 62(3). <https://doi.org/10.1590/010318138163473142>
- Sánchez Mendiola, M. (2024). La inteligencia artificial generativa y la evaluación: ¿Qué pasará con los exámenes? *Revista Investigación en Educación Médica*, 13(51), 1-10. <https://riem.facmed.unam.mx/index.php/riem/article/view/1228/1462>
- Sandoval Hernández, M. A., Morales Alarcón, G. J., Vázquez Leal, H., Huerta Chua, J., & Filobello Niño, U. A. (2023). El uso del *prompt* de ChatGPT como asistente en la educación. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 13(26). <https://doi.org/10.23913/ride.v13i26.1872>
- Urquilla Castaneda, A. (2022). Un viaje hacia la inteligencia artificial en la educación. *Realidad y Reflexión*, 22(56), 121-136. <https://doi.org/10.5377/ryr.v1i56.15776>
- Zúñiga Sánchez, O. (2024). El impacto de ChatGPT en la formación y producción académica: que no cunda el pánico. *Revista Investigación en Educación Médica*, 14(28). <https://doi.org/10.23913/ride.v14i28.1867>

Roles de autor: Apolín, L.: Conceptualización, Metodología, Curación de datos, Escritura – Borrador original. Garay, J.: Conceptualización, Visualización, Escritura – Revisión y edición.

Cómo citar este artículo: Apolín, L., & Garay, J. (2026). ¿Es capaz ChatGPT-4 de aprobar las pruebas de ingreso a la Carrera Pública Magisterial (2018, 2019, 2022 y 2024)? *Educación*, XXXV(68), 199-223. <https://doi.org/10.18800/educacion.202601.A009>

Primera publicación: 02 de febrero de 2026.

Este es un artículo de acceso abierto distribuido bajo los términos de Licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0), que permite el uso, la distribución y la reproducción sin restricciones en cualquier medio, siempre que se cite correctamente la obra original.