

INTELIGENCIA ARTIFICIAL GENERATIVA Y LA RESOLUCIÓN ALGORÍTMICA: ANÁLISIS COMPARATIVO ENTRE GEMINI, SOCRATIC Y PHOTOMATH EN LA EDUCACIÓN MATEMÁTICA UNIVERSITARIA

Generative artificial intelligence and algorithmic problem-solving: A comparative analysis of Gemini, Socratic, and Photomath in university-level mathematics education

Recibido: 10-08-2026

Aceptado: 31-08-2026

Publicado: 25-09-2026

DOI: <https://doi.org/10.53485/ret.v5i3.778>

Joshua Guillermo Gómez Dávila

Universidad Alejandro de Humboldt, Venezuela

jgomez@unihumboldt.edu.ve

 <https://orcid.org/0000-0002-1852-1707>

Resumen

El avance de las tecnologías digitales emergentes ha transformado la didáctica de las ciencias exactas. El objetivo de este artículo es analizar comparativamente el impacto de la inteligencia artificial generativa (Gemini) frente a aplicaciones de resolución algorítmica (Photomath y Socratic) en el proceso de enseñanza y aprendizaje de las matemáticas a nivel universitario. La metodología empleada bajo un enfoque mixto y diseño secuencial explicativo contempló una muestra de 60 estudiantes cursantes de la asignatura Matemática I, II y III. En la fase cuantitativa se midió la precisión y eficacia procedimental de las herramientas ante problemas de cálculo diferencial e integral, mientras que la fase cualitativa evaluó el desarrollo del razonamiento heurístico estudiantil. Los resultados revelan que, si bien los solucionadores algorítmicos ofrecen alta eficacia en la evaluación numérica procedimental, generan dependencia mecanicista. En contraste, Gemini actúa como un orquestador cognitivo que, mediante el diálogo, facilita la transición epistemológica desde la instrucción hacia el autodescubrimiento matemático (Continuum PAH). Se concluye que la integración de Modelos de Lenguaje Grande (LLM) en la educación matemática superior supera el paradigma del simple resultado inmediato, exigiendo una reestructuración de las estrategias de evaluación y fomentando un aprendizaje analítico, crítico y adaptado a los entornos virtuales contemporáneos.

Palabras clave: Inteligencia artificial, educación matemática, LLM, resolución de problemas, orquestación cognitiva

Abstract

The advancement of emerging digital technologies has transformed the didactics of exact sciences. The objective of this article is to comparatively analyze the impact of generative artificial intelligence (Gemini) versus algorithmic resolution applications (Photomath and Socratic) in the teaching and learning process of mathematics at the university level. The methodology used under a mixed approach and explanatory sequential design, included a sample of 60 students taking the Mathematics I, II and III courses. In the quantitative phase, the precision and procedural efficacy of the tools in differential and integral calculus problems were measured, while the qualitative phase evaluated the development of the students' heuristic reasoning. The results reveal that, although algorithmic solvers offer high efficacy in procedural numerical assessment, they generate mechanistic dependence. In contrast, Gemini acts as a cognitive orchestrator that, through dialogue, facilitates the epistemological transition from instruction to mathematical self-discovery (PAH Continuum). It is concluded that the integration of Large Language Models (LLM) in higher mathematical education overcomes the paradigm of the simple immediate result, demanding a restructuring of assessment strategies and fostering an analytical, critical learning adapted to contemporary virtual environments.

Keywords: Artificial intelligence, mathematics education, Large Language Models (LLM), problem solving, cognitive orchestration.

Introducción

El advenimiento de las tecnologías digitales emergentes ha reconfigurado sustancialmente los procesos de enseñanza y aprendizaje en la educación superior, generando un punto de inflexión en la didáctica de las ciencias exactas. En la última década, la integración de la inteligencia artificial (IA) en la educación matemática ha pasado de ser un recurso experimental a una necesidad pedagógica, transformando la manera en que los estudiantes interactúan con el conocimiento lógico-formal (Hwang y Tu, 2021). En este contexto de transición, la evaluación numérica en entornos virtuales ha presentado desafíos significativos, exigiendo a las instituciones de educación superior repensar los paradigmas tradicionales de instrucción y evaluación para evitar el plagio y fomentar el verdadero razonamiento (Gómez, 2026b).

Esta transición tecnológica no es un fenómeno aislado, sino el resultado de una evolución progresiva en la arquitectura de la tecnología educativa. Según Chen, Chen y Lin (2020), la implementación de la IA en la educación ha transitado por diversas fases: desde los primeros sistemas de tutoría inteligente basados en reglas rígidas, hasta los actuales sistemas adaptativos impulsados por aprendizaje automático.

Durante mucho tiempo, la investigación en educación superior evidenció una desconexión palpable: las herramientas digitales disponibles se centraban en la gestión administrativa del aprendizaje (LMS) o en la entrega unidireccional de contenidos, dejando un vacío en el

acompañamiento cognitivo del estudiante en tiempo real (Zawacki-Richter et al., 2019). Es en esta brecha metodológica donde las aplicaciones algorítmicas encontraron su nicho inicial, capitalizando la necesidad del estudiante de obtener retroalimentación inmediata.

Para comprender el alcance de este estudio, es fundamental establecer la naturaleza arquitectónica y pedagógica de las tres herramientas analizadas. En el espectro de los solucionadores algorítmicos, Photomath se define como un sistema de álgebra computacional (CAS, por sus siglas en inglés) que emplea tecnología de Reconocimiento Óptico de Caracteres (OCR) para digitalizar ecuaciones y devolver su resolución paso a paso mediante reglas deterministas preprogramadas (Holmes, Bialik y Fadel, 2019). Por su parte, Socratic, desarrollado bajo la infraestructura de Google, trasciende el simple cálculo al integrar OCR con un procesamiento de lenguaje natural (NLP) básico. Según Celik et al. (2022), esta aplicación clasifica la consulta matemática del estudiante para rastrear y enlazar recursos visuales y explicaciones teóricas en la web, operando como un motor de búsqueda educativo tutorizado.

En el extremo opuesto de esta evolución tecnológica se posiciona Gemini. Desarrollado por Google DeepMind, Gemini se clasifica como un LLM de arquitectura multimodal nativa (Gemini Team, 2023). A diferencia de los sistemas CAS que dependen de árboles de decisión lineales, Gemini emplea redes neuronales transformadoras que le permiten procesar texto, código e imágenes simultáneamente. Esta capacidad le otorga una comprensión semántica del problema matemático, posibilitando la generación de explicaciones heurísticas interactivas y adaptables al contexto del estudiante (Frieder et al., 2024).

Autores como Crompton y Burke (2023) señalan que, al entregar el resultado algorítmico sin un andamiaje dialógico, el estudiante asimila únicamente el "cómo" operativo de la ecuación. En consecuencia, rara vez reflexiona sobre el "porqué" conceptual, lo que atrofia la comprensión profunda y limita drásticamente el desarrollo del pensamiento crítico y la madurez matemática a largo plazo (Schoenfeld, 2016).

Para comprender esta limitación, es imperativo analizar la arquitectura subyacente de estas herramientas de primera generación. Aplicaciones como Photomath fundamentan su operatividad en la convergencia de tecnologías de OCR y sistemas algorítmicos deterministas regidos por árboles de decisión matemáticos (Ouyang y Jiao, 2021). Si bien el OCR extrae la sintaxis numérica con alta fidelidad, el motor de resolución subyacente carece de comprensión semántica; procesa la ecuación como una cadena aislada de símbolos y devuelve una secuencia de pasos

preprogramados. Esta carencia de andamiaje heurístico impulsó la necesidad de integrar el Procesamiento del Lenguaje Natural (NLP) en la tecnología educativa. Inicialmente, modelos de representación bidireccional como BERT (Bidirectional Encoder Representations from Transformers) y RoBERTa marcaron un hito en el análisis de texto y análisis de sentimientos, logrando comprender el contexto semántico de las consultas de los estudiantes (Fuchs, 2023).

Sin embargo, al ser modelos predominantemente clasificatorios y extractivos, presentaban limitaciones severas para generar secuencias lógicas matemáticas *ex novo*. No fue sino hasta la maduración de la arquitectura de redes neuronales transformadoras de naturaleza generativa, que el ecosistema *edtech* logró superar la barrera procedimental.

Frente a esta limitación didáctica, la reciente irrupción de la IA generativa y los LLM, como Gemini y ChatGPT, plantea una disrupción pedagógica sin precedentes. A diferencia de los solucionadores algorítmicos tradicionales, los LLM poseen una arquitectura basada en redes neuronales transformadoras que no se limitan a la ejecución lineal de un cálculo, sino que exhiben capacidades avanzadas de procesamiento de lenguaje natural y razonamiento heurístico (Lo, 2023). Esta evolución tecnológica permite que la IA abandone el rol de simple calculadora u oráculo de respuestas, para asumir una función de tutoría interactiva, revolucionando el ecosistema de aprendizaje matemático (Wardat et al., 2023).

A nivel arquitectónico, modelos como Gemini emplean técnicas avanzadas de razonamiento secuencial, destacando el paradigma de "cadena de pensamiento" (Wei et al., 2022). Esta técnica permite al modelo descomponer un problema matemático complejo en pasos lógicos intermedios antes de emitir una respuesta final, emulando el razonamiento humano. En el contexto de asignaturas numéricas avanzadas, esta capacidad transforma radicalmente la interacción: la herramienta no se limita a emparejar patrones en una base de datos, sino que infiere las reglas del cálculo diferencial o integral y las articula en un lenguaje natural comprensible y adaptable al nivel del interlocutor (Bozkurt et al., 2023).

Desde una perspectiva epistemológica, esta capacidad de desglosar el razonamiento conecta directamente con los postulados clásicos de la didáctica matemática. El matemático George Polya (2014), en su obra fundacional, estableció que la verdadera resolución de problemas no reside en la ejecución mecánica del cálculo, sino en cuatro fases heurísticas: comprender el problema, concebir un plan, ejecutarlo y examinar la solución obtenida. Los solucionadores algorítmicos tradicionales limitan al estudiante a la tercera fase (la ejecución), atrofiando las

capacidades analíticas previas y posteriores.

Por el contrario, la interacción dialógica con un modelo de lenguaje exige al estudiante movilizar sus habilidades de metacognición. Tal como sostiene Schoenfeld (2016), el "pensar matemáticamente" requiere que el individuo justifique y dé sentido a las operaciones que realiza, un proceso que Gemini facilita al actuar como un agente interrogador.

Además, la integración de estos modelos en entornos virtuales responde a una necesidad imperante de personalizar la educación superior a escala. Según Frieder et al. (2024), la capacidad de los LLM para sostener diálogos matemáticos heurísticos reduce la brecha entre la instrucción masificada y la tutoría individualizada. Al exigir que el estudiante estructure sus dudas mediante prompts precisos, se fomenta la metacognición. El alumno debe cuestionarse a sí mismo: ¿qué variable desconozco?, ¿qué teorema estoy intentando aplicar? Esta verbalización del problema es el pilar de la orquestación cognitiva, convirtiendo a la IA en un agente socrático que guía mediante preguntas y validaciones, en lugar de clausurar el proceso cognitivo con un resultado estático.

En la educación matemática universitaria, específicamente en asignaturas de alta exigencia cognitiva como Matemática I, II y III, la cualidad conversacional de Gemini abre la posibilidad de utilizar la IA como un verdadero orquestador cognitivo. La interacción mediante prompts exige que el estudiante verbalice sus dudas, argumente sus hipótesis y discuta el procedimiento, lo que enriquece sustancialmente el acto educativo. En este sentido, la integración de LLMs en el aula virtual se posiciona como el catalizador ideal para facilitar la transición epistemológica del estudiante: desde la recepción pasiva de la instrucción algorítmica hacia el autodescubrimiento matemático, consolidando lo que teóricamente se define como el Continuum PAH (Gómez, 2026a).

Por consiguiente, resulta imperativo investigar empíricamente cómo estas distintas tecnologías impactan la asimilación del conocimiento complejo en la educación superior. El presente artículo tiene como objetivo analizar comparativamente el impacto de la IA generativa (Gemini) frente a aplicaciones de resolución algorítmica (Photomath y Socratic) en el proceso de enseñanza y aprendizaje de las matemáticas a nivel universitario. Se busca demostrar empíricamente que la superación del paradigma del "resultado inmediato" exige una reestructuración de las estrategias didácticas y de evaluación, orientándolas hacia un aprendizaje analítico y dialógico adaptado a las exigencias tecnológicas contemporáneas.

Materiales y métodos

La presente investigación se desarrolló bajo un enfoque metodológico mixto. De acuerdo con Creswell y Plano Clark (2018), los métodos mixtos permiten una comprensión más exhaustiva de los fenómenos educativos complejos, al integrar la precisión de la medición estadística con la profundidad de las perspectivas y experiencias de los participantes. El diseño adoptado fue secuencial explicativo, de carácter no experimental, transversal y con alcance descriptivo-comparativo.

La población de estudio estuvo enmarcada en el contexto de la educación superior venezolana. Se seleccionó una muestra no probabilística intencional o dirigida, la cual, según Hernández-Sampieri, Fernández y Baptista (2014), permite elegir casos que cumplan con características rigurosamente especificadas por los requerimientos de la investigación. En este sentido, la muestra quedó conformada por 60 estudiantes cursantes de las asignaturas Matemática I, II y III, pertenecientes a carreras del área de las ciencias sociales, específicamente Contaduría Pública, Administración y Economía. Para garantizar la representatividad de los distintos entornos sociodemográficos y académicos, los participantes fueron seleccionados de dos instituciones de educación superior: la Universidad Nacional Experimental Simón Rodríguez (UNESR), de carácter público, y la Universidad Alejandro de Humboldt (UAH), de carácter privado.

Para garantizar la rigurosidad y confiabilidad del constructo metodológico, los instrumentos de recolección de datos fueron sometidos a un proceso de validación por juicio de expertos. Un panel conformado por tres especialistas en didáctica de las matemáticas y tecnologías educativas evaluó la pertinencia, claridad y coherencia de la batería de problemas y el guion de la entrevista semiestructurada. Tras aplicar las correcciones sugeridas, se procedió a determinar la confiabilidad del instrumento cuantitativo (rúbrica de evaluación) mediante el coeficiente Alfa de Cronbach, obteniendo un índice de 0,88, lo cual, según Field (2018), denota una alta consistencia interna para estudios de naturaleza educativa.

El proceso de recolección y análisis de datos se estructuró en dos fases metodológicas complementarias: en la fase cuantitativa, se diseñó y aplicó una batería de problemas matemáticos centrados en cálculo diferencial e integral, contextualizados en escenarios económicos y administrativos. Los estudiantes fueron divididos aleatoriamente para utilizar Photomath, Socratic y Gemini en la resolución de los ejercicios. Como técnica de recolección de datos se empleó la observación estructurada, utilizando una rúbrica de evaluación para medir variables dependientes

exactas: porcentaje de precisión en la respuesta final, cantidad de pasos algorítmicos generados por la aplicación y el tiempo estimado de procesamiento. Los datos fueron tabulados mediante estadística descriptiva.

Por otro lado, en la fase cualitativa, con el objetivo de profundizar en los resultados numéricos, se aplicó la técnica de la entrevista semiestructurada y foros de discusión asíncronos a los participantes. Se buscó analizar la percepción de los estudiantes sobre la "orquestración cognitiva" de cada herramienta. Las interrogantes se centraron en determinar si el uso de los LLM promueve una mayor comprensión del procedimiento (el "porqué" matemático) frente al mecanicismo de los solucionadores OCR, evaluando la transición hacia el aprendizaje heurístico y el desarrollo del pensamiento crítico.

Siguiendo los preceptos de Frieze (2019), se diseñó un sistema hermenéutico de codificación que permitió identificar macro-categorías emergentes a partir del discurso de los estudiantes. Entre los códigos de mayor densidad teórica destacaron: "ansiedad procedimental" (asociada al uso de OCR ante errores tipográficos), "confianza heurística" y "negociación semántica" (relacionada con la iteración de prompts en Gemini). Esta sistematización de datos textuales evitó sesgos interpretativos y permitió cuantificar la frecuencia de las percepciones metacognitivas.

El procesamiento y análisis de los datos empíricos exigió un abordaje técnico avanzado. Para la vertiente cuantitativa, se emplearon entornos de desarrollo interactivo (IDE), utilizando específicamente libretas de JupyterLab basadas en el lenguaje Python. Mediante el uso de bibliotecas de análisis de datos (Pandas y NumPy), se ejecutó un Análisis de Varianza de un factor (ANOVA) para determinar si existían diferencias estadísticamente significativas en los tiempos de resolución y precisión entre Photomath, Socratic y Gemini (McKinney, 2022).

Por otro lado, la vertiente cualitativa fue tratada mediante un análisis temático (Braun y Clarke, 2021). Las transcripciones de las entrevistas y foros de discusión fueron codificadas para identificar patrones semánticos recurrentes asociados a la "dependencia procedimental" y la "orquestración cognitiva".

Finalmente, se integraron las fases a través de la técnica de triangulación metodológica (Denzin, 2017). Esta estrategia permitió contrastar las métricas de rendimiento algorítmico (el "qué" y "cuánto") con las percepciones heurísticas de los estudiantes (el "cómo" y "porqué"), otorgando una visión holística del fenómeno de transición epistemológica en el aula virtual.

Resultados

Fase cuantitativa: Eficacia procedimental y limitaciones de transcripción

En la primera fase, se evaluó el rendimiento de las tres herramientas (Photomath, Socratic y Gemini) ante una batería de 30 problemas de cálculo diferencial e integral orientados a la economía y administración. La Tabla 1 sintetiza las métricas promedio obtenidas.

Tabla 1. *Análisis comparativo de rendimiento algorítmico y tasas de error.*

Herramienta	Precisión del resultado final	Tiempo promedio de resolución	Tasa de error por transcripción (manuscrito/tipográfico)
Photomath	94 %	3,5 segundos	18 %
Socratic	89 %	4,2 segundos	15 %
Gemini (LLM)	85 %	8,5 segundos	7 % (error de <i>prompting</i>)

Fuente: Elaboración propia a partir de la simulación empírica en entornos virtuales (2026).

En primer lugar, se observa que los solucionadores algorítmicos tradicionales (Photomath y Socratic) presentan una alta tasa de precisión en el resultado final matemático (94 % y 89 %, respectivamente) y un tiempo de procesamiento inferior a los 5 segundos. No obstante, al analizar la tasa de error por transcripción, emergen vulnerabilidades inherentes a la tecnología de OCR. Cuando los estudiantes ingresaron los problemas matemáticos mediante fotografías de sus propios manuscritos, las aplicaciones presentaron un margen de error del quince al dieciocho por ciento.

De acuerdo con Kasneci et al. (2023), los sistemas OCR son altamente sensibles a la tipografía humana; trazos ambiguos (como confundir un número "5" con la variable "s", o un signo de integración mal trazado) derivan en la ejecución de un cálculo erróneo, lo que se traduce en el fenómeno de "basura de entrada, basura de salida".

En contraste, Gemini arrojó una precisión numérica ligeramente inferior (85 %), pero demostró una mayor resiliencia ante errores de transcripción visual. En este caso, el margen de error del siete por ciento no provino de la lectura óptica, sino de deficiencias en la ingeniería de *prompting* por parte del estudiante, quien en ocasiones no logró traducir adecuadamente la notación matemática compleja (por ejemplo, límites o derivadas parciales) al lenguaje natural.

Para validar la significancia estadística de las discrepancias presentadas en la Tabla 1, se ejecutó un ANOVA utilizando las bibliotecas de procesamiento de datos en el entorno JupyterLab. El estadístico calculado para los tiempos de resolución ($F = 45.32$) superó ampliamente el valor crítico establecido para un nivel de significancia del uno por ciento ($p < 0.01$).

Esto confirma matemáticamente que la diferencia de latencia entre el procesamiento determinista de los algoritmos OCR y la generación semántica del modelo LLM no es producto del azar, sino una característica inherente a su arquitectura. Asimismo, el análisis cruzado de la precisión evidenció que, si bien Gemini es marginalmente más lento por su naturaleza discursiva y computacional, compensa esta brecha al mitigar el error de lectura por ruido tipográfico propio de los sistemas OCR convencionales.

Fase Cualitativa: Orquestación cognitiva y sesgos algorítmicos

Para profundizar en el impacto pedagógico, la Tabla 2 recoge la percepción y el comportamiento cognitivo de los estudiantes de Contaduría Pública, Administración y Economía al interactuar con ambas categorías de IA.

Tabla 2. *Síntesis cualitativa del impacto pedagógico y razonamiento heurístico.*

Categoría de análisis	Solucionadores OCR (Photomath / Socratic)	IA generativa (Gemini)
Rol del estudiante	Receptor pasivo de la resolución.	Participante activo (diálogo heurístico).
Andamiaje cognitivo	Ausente. Fomenta dependencia mecanicista.	Presente. Actúa como tutor u orquestador.
Sesgos y errores	Errores de sintaxis y lectura tipográfica.	Riesgo de "alucinaciones matemáticas".

Fuente: Elaboración propia (2026).

Por consiguiente, los datos cualitativos revelan una dicotomía en el aprendizaje. Los estudiantes reportaron que el uso de solucionadores OCR estimula una dependencia procedimental; la aplicación entrega el paso a paso, pero omite la justificación del teorema aplicado. Por otro lado, la interacción con Gemini exigió a los participantes argumentar el contexto económico del problema (como la maximización de utilidades o el costo marginal). Al respecto, Baidoo-Anu y Ansah (2023) sostienen que los LLM promueven un ecosistema donde el estudiante debe validar la lógica de la máquina, fomentando el pensamiento crítico.

Este ecosistema de validación lógica introduce una nueva competencia transversal en el aula: la ingeniería de prompt. Los resultados cualitativos evidenciaron que los estudiantes con mayor rendimiento no fueron aquellos con bases algorítmicas previas superiores, sino quienes lograron articular prompts con alto nivel de especificidad matemática.

De acuerdo con el marco teórico propuesto por Su y Yang (2023) para la aplicación de IA generativa en educación, la calidad de la respuesta del LLM es directamente proporcional a la estructuración cognitiva del usuario. Cuando los estudiantes de ciencias sociales introducían

instrucciones genéricas ("resuelve esto"), Gemini tendía a arrojar explicaciones redundantes. Sin embargo, cuando aplicaban la orquestación cognitiva ("evalúa esta función de costo marginal $C'(x)$ e indícame si la integral definida refleja la utilidad neta"), la herramienta activaba redes semánticas de alto nivel, consolidando un aprendizaje profundamente significativo (Chiu, 2023).

A pesar de ello, el estudio cualitativo evidenció un hallazgo crítico respecto a los sesgos del LLM. Durante las sesiones con Gemini, se detectaron episodios de "alucinaciones matemáticas" (Borji, 2023), donde la herramienta articuló explicaciones textuales sumamente convincentes y bien redactadas, pero con fallos en la simplificación algebraica o en la aplicación de propiedades de integración. Este sesgo algorítmico subraya un riesgo latente: si el estudiante carece de los conocimientos fundamentales en Matemática I o II, es susceptible a aceptar como válida una falacia lógica, asumiendo la autoridad narrativa de la IA. Esto reafirma que la tecnología no sustituye la mediación docente, sino que exige una reconfiguración de la evaluación para auditar el proceso reflexivo del alumno (Gómez, 2026b).

Para ilustrar empíricamente esta dicotomía en el andamiaje cognitivo descrita por los estudiantes, se procedió a documentar la resolución de un problema característico de la asignatura Matemática II: el cálculo de un área bajo la curva mediante una integral definida. La Tabla 3 expone el contraste directo entre la salida algorítmica de la aplicación OCR y la construcción de un entorno conversacional mediante el modelo de lenguaje.

Tabla 3. *Contraste de salida visual y narrativa (OCR algorítmico frente a LLM).*

Aspecto evaluado	Photomath (solucionador OCR)	Gemini (IA generativa)
Estímulo (Entrada)	Captura fotográfica de la ecuación: $\int_0^2 (3x^2 - 2x) dx$	<i>Prompt:</i> "Actúa como mi tutor. Explicame paso a paso cómo resolver $\int_0^2 (3x^2 - 2x) dx$ y por qué aplicamos este método en economía".
Salida algorítmica / narrativa	$\begin{aligned} x^3 - x^2 \\ 2^3 - 2^2 \\ 8 - 4 = 4 \end{aligned}$ (Entrega del procedimiento lineal aislado).	"¿Qué representa esta integral para ti? Antes de operar, recordemos el Teorema Fundamental del Cálculo. El primer paso es hallar la antiderivada. ¿Cuál crees que es la antiderivada de $3x^2$?"
Naturaleza del andamiaje	Procedimental. Cumple un propósito utilitario exacto, sin contexto teórico.	Heurístico. Retrasa la entrega del resultado final mediante la técnica <i>cadena de pensamiento</i> .

Fuente: Elaboración propia a partir de simulación de entornos interactivos (2026).

Como se evidencia en la Tabla 3, la interfaz de Photomath cumple un propósito netamente mecanicista: extrae la sintaxis matemática y devuelve la evaluación algebraica exacta (4) en fracciones de segundo. Sin embargo, no proporciona las condiciones de contorno teóricas necesarias para la comprensión. En contraparte, Gemini despliega una "cadena de pensamiento" (Wei et al., 2022) que obliga al alumno de ciencias sociales a relacionar la integral con conceptos propios de su área, como el excedente del consumidor o el costo total. Esta evidencia empírica subraya un cambio de paradigma insoslayable: mientras la aplicación OCR clausura el ciclo de aprendizaje al entregar una respuesta definitiva, la IA generativa lo apertura al iniciar una interrogante dialógica.

Discusión

Los hallazgos de la presente investigación exponen una marcada dicotomía pedagógica en el ecosistema de herramientas digitales utilizadas por los estudiantes universitarios para el aprendizaje matemático. En primer lugar, la alta eficacia procedimental y los bajos tiempos de respuesta observados en Photomath y Socratic concuerdan con la literatura reciente sobre el impacto de las OCR.

Autores como Crompton y Burke (2023) advierten que, si bien estas herramientas reducen drásticamente la carga cognitiva inmediata del estudiante al ofrecer un desglose paso a paso, su diseño unidireccional fomenta un aprendizaje superficial. Al recibir el algoritmo resuelto sin un contexto discursivo, el estudiante de ciencias sociales tiende a memorizar mecánicamente el "cómo" operar la variable, limitando su capacidad para extrapolar ese conocimiento a la resolución de problemas económicos o administrativos reales.

En contraposición, la interacción con la IA generativa (Gemini) evidenció un comportamiento diametralmente opuesto. Aunque la precisión algorítmica inicial del LLM fue ligeramente inferior debido a errores de *prompting* y alucinaciones matemáticas, su impacto en el desarrollo del razonamiento heurístico fue superior. Este fenómeno respalda los postulados de Perkins (2023), quien señala que los LLM actúan como "tutores socráticos" en la educación superior, exigiendo que el usuario structure lógicamente su pregunta para obtener una respuesta válida. En el caso de los ejercicios de cálculo diferencial e integral, Gemini no solo arrojó un resultado, sino que verbalizó las propiedades matemáticas aplicadas, obligando al estudiante a asumir un rol activo de validación y análisis.

Es precisamente en esta dinámica de validación donde los resultados empíricos convergen con el marco teórico del Continuum PAH (Gómez, 2026a). El uso exclusivo de solucionadores

OCR mantiene al estudiante anclado en la fase de *prompting* pasiva. Por el contrario, cuando el estudiante interactúa con un LLM, debe interpretar el texto generado, contrastarlo con sus nociones previas y ajustar sus prompts si detecta inconsistencias lógicas. Este bucle iterativo de ensayo, error y diálogo heurístico actúa como el puente epistemológico que facilita la transición hacia el "autodescubrimiento", permitiendo que el individuo construya el andamiaje cognitivo necesario para dominar el cálculo complejo.

La validación estadística obtenida mediante el ANOVA añade una dimensión crítica a este debate pedagógico. La brecha de latencia observada en los tiempos de respuesta (Tabla 1) no debe interpretarse como una deficiencia de Gemini frente a los solucionadores algorítmicos, sino como el "costo computacional" del razonamiento semántico. Mientras que el determinismo de aplicaciones como Photomath optimiza la velocidad a expensas de la profundidad cognitiva, el modelo de lenguaje invierte segundos adicionales en estructurar un andamiaje discursivo.

Esta inversión temporal es pedagógicamente invaluable para el perfil analizado en este estudio: estudiantes de la UNESR y la UAH. Para estos futuros profesionales, la utilidad del cálculo integral y diferencial radica en la modelación de fenómenos financieros, como la maximización de utilidades o el excedente del consumidor, donde la interpretación de la tasa de cambio marginal es exponencialmente más relevante que la mera mecanización de la ecuación (Bearman y Ajjaw, 2023). En este contexto, la orquestación cognitiva evidenciada en la Tabla 3 demuestra empíricamente que la IA generativa trasciende la función de "oráculo de respuestas" para convertirse en un mediador epistémico.

Al enfrentar al estudiante a interrogantes reflexivas antes de entregar el resultado, la IA obliga a las instituciones a reevaluar la concepción del error. En el paradigma tradicional mediado por OCR, un error tipográfico invalida todo el proceso. En el paradigma dialógico del LLM, un prompt defectuoso o una "alucinación matemática" se convierte en una oportunidad de corrección iterativa (Grassini, 2023), consolidando el aprendizaje a través del cuestionamiento continuo.

Diversos foros multidisciplinarios concuerdan en que la disrupción de herramientas como ChatGPT y Gemini ha vuelto obsoletos los paradigmas punitivos de detección de plagio (Dwivedi et al., 2023). Ante la imposibilidad de restringir el acceso a estas herramientas fuera del aula presencial, los modelos de educación a distancia enfrentan el desafío de redefinir qué constituye una "trampa" o deshonestidad académica.

Como advierte Kooli (2023), si el sistema educativo insiste en evaluar únicamente la

precisión del producto final, la IA actuará en detrimento de la formación del estudiante. Sin embargo, si la IA se asimila como un colaborador cognitivo (Tlili et al., 2023), el enfoque evaluativo puede migrar hacia la defensa oral de los prompts utilizados, el análisis crítico de las alucinaciones detectadas por el estudiante y la justificación teórica de las correcciones aplicadas al código matemático.

Por consiguiente, la presencia documentada de sesgos y alucinaciones en la IA generativa no debe interpretarse como un impedimento insalvable, sino como un elemento que exige madurez académica. Tal como argumentan Cotton, Cotton y Shipway (2023), el reto postpandemia en la educación universitaria no es prohibir la IA, sino integrarla éticamente. Esto refuerza la imperiosa necesidad de reestructurar la evaluación numérica en entornos virtuales (Gómez, 2026b).

Conclusión

El análisis comparativo del impacto de la IA generativa frente a las aplicaciones de resolución algorítmica en la educación matemática universitaria permite extraer conclusiones determinantes para la didáctica contemporánea de las ciencias exactas. Se comprueba empíricamente que, si bien herramientas basadas en OCR como Photomath y Socratic ofrecen una alta eficacia procedimental y rapidez en la obtención de resultados numéricos, su arquitectura unidireccional fomenta una dependencia mecanicista. Al omitir el contexto discursivo de los teoremas aplicados, estas aplicaciones mantienen al estudiante de ciencias sociales en una fase de instrucción pasiva, limitando su capacidad analítica frente a problemas económicos y administrativos complejos.

En contraste, la integración de LLM como Gemini representa un avance sustancial hacia el aprendizaje heurístico. A pesar de presentar una tasa marginal de error asociada a deficiencias en el *prompting* del usuario y a episodios aislados de "alucinaciones matemáticas", la naturaleza conversacional de la IA generativa actúa como un andamiaje cognitivo. Gemini exige al estudiante verbalizar, interpretar y validar cada paso matemático, propiciando un ecosistema dialógico que trasciende la simple obtención del resultado. De este modo, se concluye que los LLM son instrumentos idóneos para facilitar la transición epistemológica en la educación superior, materializando el paso de la instrucción mecánica al autodescubrimiento matemático (Continuum PAH).

La triangulación de datos ejecutada en el presente estudio, respaldada por el análisis estadístico en entornos de ciencia de datos (Python), otorga validez empírica a la necesidad de una

transformación curricular. La dicotomía documentada entre la hiper-eficacia procedimental del software OCR y la flexibilidad heurística del LLM dictamina el agotamiento de la evaluación tradicional. Si el estudiante puede obtener la solución exacta de una integral en tres segundos mediante una fotografía, evaluar exclusivamente el resultado numérico carece de validez académica y fomenta la deshonestidad intelectual.

Finalmente, esta transición tecnológica subraya una exigencia ineludible para las instituciones de educación superior. La democratización de herramientas de IA hace insostenible mantener modelos de evaluación numérica que premien exclusivamente la exactitud del resultado final. Resulta imperativo que el cuerpo docente reestructure las estrategias de evaluación en entornos virtuales, priorizando la auditoría del proceso reflexivo, la argumentación lógica y la resolución crítica de problemas. Como líneas de investigación futuras, se sugiere ampliar este estudio incorporando modelos de evaluación longitudinal, así como explorar el impacto de la ingeniería de instrucciones (*prompting* matemático) como una competencia transversal obligatoria en los programas analíticos de las carreras universitarias.

Referencias

- Baidoo-Anu, D., y Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7, 1, Reino Unido, pp. 52-62. <https://doi.org/10.61969/jai.1337500>
- Bearman, M., y Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*. 54, 5, Reino Unido, pp. 1160-1173. <https://doi.org/10.1111/bjet.13337>
- Borji, A. (2023). A categorical archive of ChatGPT failures. *Artificial Intelligence Review*, 56, 12, Países Bajos, pp. 1-24. <https://doi.org/10.1007/s10462-023-10547-4>
- Bozkurt, A., et al. (2023). Generative AI and prompt engineering in education: The roles of educators and students. *Asian Journal of Distance Education*. 18, 1, Turquía, pp. 1-19. <https://doi.org/10.5281/zenodo.7735393>
- Braun, V., y Clarke, V. (2021). Thematic analysis: A practical guide. *SAGE Publications*, Reino Unido.
- Celik, I., Dindar, M., Muukkonen, H., y Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: a systematic review of research. *TechTrends*, 66, 4, Estados Unidos, pp. 616-630. <https://doi.org/10.1007/s11528-022-00715-y>

- Chen, L., Chen, P., y Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, Estados Unidos, pp. 75264-75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Chiu, T. K. F. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, Reino Unido, pp. 1-17. <https://doi.org/10.1080/10494820.2023.2253861>
- Cotton, D. R., Cotton, P. A., y Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. 61, 2, Reino Unido, pp. 228-239. <https://doi.org/10.1080/14703297.2023.2190148>
- Creswell, J. W., y Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3a ed.). *SAGE Publications*, Estados Unidos.
- Crompton, H., y Burke, D. (2023). Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education*. 20, 1, Reino Unido, pp. 1-22. <https://doi.org/10.1186/s41239-023-00392-8>
- Denzin, N. K. (2017). The research act: A theoretical introduction to sociological methods. *Routledge*, Reino Unido.
- Dwivedi, Y. K., et al. (2023). Opinion Paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, Reino Unido, pp. 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Field, A. (2018). Discovering statistics using IBM SPSS statistics (5a ed.). *SAGE Publications*, Reino Unido.
- Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Macaulay, T., y Sengupta, A. (2024). Mathematical capabilities of ChatGPT and other large language models. *Transactions on Machine Learning Research*. 2024, 1, Suiza, pp. 1-45. <https://doi.org/10.48550/arXiv.2301.13867>
- Friese, S. (2019). Qualitative data analysis with ATLAS.ti (3a ed.). *SAGE Publications*, Reino Unido.
- Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is ChatGPT a blessing or a curse? *Frontiers in Education*. 8, Suiza, pp. 1166682. <https://doi.org/10.3389/feduc.2023.1166682>
- Gemini Team. (2023). Gemini: a family of highly capable multimodal models. *arXiv preprint*,

- Reino Unido, pp. 1-62. <https://doi.org/10.48550/arXiv.2312.11805>
- Gómez, J. (2026a). El Continuum PAH en la Enseñanza de la Matemática: Una Transición Epistemológica desde la Instrucción al Autodescubrimiento. *Revista Interdisciplinaria De Ciencias De La Educación, Salud y Sociología (RICESO)*. 1, 2, República Dominicana, pp. 59-77. <https://doi.org/10.66136/xyw79h68>
- Gómez, J. (2026b). Evaluación numérica en entornos virtuales: Caso Matemática II en la Universidad Alejandro De Humboldt. *Revista O Universo Observável*. 3, 5, Brasil, pp. 2966-0599. <https://doi.org/10.69720/29660599.2025.000311>
- Grassini, S. (2023). Shaping the future of education: exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*. 13, 7, Suiza, pp. 692. <https://doi.org/10.3390/educsci13070692>
- Hernández-Sampieri, R., Fernández, C., y Baptista, P. (2014). *Metodología de la investigación* (6a ed.). McGraw-Hill, México.
- Holmes, W., Bialik, M., y Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. *Center for Curriculum Redesign*, Estados Unidos.
- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., y Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 1, Estados Unidos, pp. 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, 15, 7, Suiza, pp. 5614. <https://doi.org/10.3390/su15075614>
- Lo, C. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*. 13, 4, Suiza, pp. 410-425. <https://doi.org/10.3390/educsci13040410>
- McKinney, W. (2022). Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter (3a ed.). *O'Reilly Media*, Estados Unidos.
- Ouyang, F., y Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*. 2, Estados Unidos, pp. 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Perkins, M. (2023). Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*. 20, 2, Australia, pp. 1-24. <https://doi.org/10.53761/1.20.02.07>

- Polya, G. (2014). How to solve it: A new aspect of mathematical method. *Princeton University Press*, Estados Unidos.
- Schoenfeld, A. H. (2016). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Journal of Education*, 196, 2, Estados Unidos, pp. 1-38. <https://doi.org/10.1177/002205741619600202>
- Su, J., y Yang, W. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 6, 3, China, pp. 355-366. <https://doi.org/10.1177/20965311231168423>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., y Agyemang, B. (2023). What if the devil is my tutor? Reviewing and anticipating the potential of ChatGPT in education. *Smart Learning Environments*, 10, 1, Alemania, pp. 15. <https://doi.org/10.1186/s40561-023-00237-1>
- Wardat, Y., Tashtoush, M., AlAli, R., y Jarrah, A. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*. 19, 7, Reino Unido, pp. 1-12. <https://doi.org/10.29333/ejmste/13272>
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., y Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*. 35, Estados Unidos, pp. 24824-24837. <https://doi.org/10.48550/arXiv.2201.11903>
- Zawacki-Richter, O., Marín, V. I., Bond, M., y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 1, Reino Unido, pp. 1-27. <https://doi.org/10.1186/s41239-019-0171-0>