Areté. Revista Digital del Doctorado en Educación
Vol. 10 (ee), 119 – 139, octubre, 2024. ISSN: 2443 – 4566
https://doi.org/10.55560/arete.2024.ee.10.9

EDUCACIÓN E INTELIGENCIA ARTIFICIAL: DESEMPEÑO DE CHATBOTS Y PROFESORES DE MATEMÁTICA EN LA RESOLUCIÓN DE PROBLEMAS GEOMÉTRICOS

EDUCATION AND ARTIFICIAL INTELLIGENCE: CHATBOTS AND MATHEMATICS TEACHERS PERFORMANCE AT SOLVING GEOMETRY PROBLEMS

ANA ROSA CORICA
UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES, FACULTAD CS. EXACTAS, NIEM, CONICET, TANDIL, ARGENTINA

PATRICIA SUREDA
UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES, FACULTAD CS. EXACTAS, NIEM, CONICET, TANDIL, ARGENTINA

VERÓNICA PARRA
UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES, FACULTAD CS. EXACTAS, NIEM, CONICET, TANDIL, ARGENTINA

SILVIA SCHIAFFINO
UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES, FACULTAD CS. EXACTAS, ISISTAN, CONICET, TANDIL, ARGENTINA

DANIELA GODOY
UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES, FACULTAD CS. EXACTAS, ISISTAN, CONICET, TANDIL, ARGENTINA

Fecha de recepción: 01 mayo 2024
Fecha de aceptación: 02 septiembre 2024

RESUMEN

La Inteligencia Artificial Generativa (IAGen) ha surgido como una tecnología disruptiva cuya amplia adopción impactó rápidamente en el contexto educativo. Este trabajo tiene por objetivo analizar el desempeño de tres chatbots en la resolución de cuatro problemas geométricos, dos de geometría clásica y dos de geometría fractal. Además, se comparan las soluciones a tres de los problemas con las aportadas por ocho profesores de matemática en servicio. El análisis y comparación de soluciones se realizan según el tipo de soluciones matemáticas proporcionadas, el rigor, procedimiento y justificación matemática empleados. El método general de análisis es exploratorio y para cada tipo de problemas se presentan categorías específicas de análisis. Los resultados obtenidos dan cuenta de las dificultades de los chatbots en resolver problemas geométricos, sobre todo, en las representaciones gráficas y ubicaciones espaciales, así como también de una clara diferencia a su favor en las

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

validaciones matemáticas. Se concluye en la potencialidad de la IAGen en la resolución de problemas siempre y cuando se analicen críticamente las respuestas que ofrecen.

PALABRAS CLAVE: Geometría clásica; Geometría fractal; IA Generativa; Educación; Profesores de matemática

ABSTRACT

Generative artificial intelligence (GenAI) has emerged as a disruptive technology with a wide adoption that has greatly impacted on education. This work aims at analyzing the performance of three chatbots at solving four Geometry problems, two of them belonging to classical geometry and the other two in the area of fractal Geometry. Also, we compare chatbots' solutions against those proposed by eight in-service Mathematics teachers. The solutions analysis and comparison are carried out taking into account the type of solutions suggested, the soundness, the procedures employed, and mathematical justifications given. The general method is exploratory, and for each problem type we present particular categories of analysis. The results obtained evidence chatbots' shortcomings when solving Geometry problems, mainly when dealing with graphical representations and spatial locations, as well as their advantage at mathematical validations. We conclude that GenIA has a great potential to contribute to Geometry problem solving as long as the answers are critically judged.

KEYWORDS: Classical geometry; Fractal geometry; Generative AI; Education; Mathematics teachers

1. INTRODUCCIÓN

Este trabajo presenta los resultados de las investigaciones que las autoras han estado desarrollando respecto al uso de la Inteligencia Artificial (IA) en la resolución de problemas matemáticos, particularmente, problemas geométricos. Encontrar soluciones para problemas de geometría puede resultar una tarea especialmente desafiante para la Inteligencia Artificial Generativa (IAGen) basada en LLMs (del inglés Large Language Models o Grandes Modelos de Lenguaje), ya que no sólo implica el conocimiento de conceptos fundamentales (como por ejemplo, ciertos teoremas) y su correcta aplicación, sino que involucra además el uso de habilidades de razonamiento espacial. De hecho, incluso los modelos multimodales más avanzados presentan dificultades para comprender con precisión figuras geométricas y las relaciones entre los elementos fundamentales, como puntos y líneas, e interpretar con precisión conceptos básicos como los grados de un ángulo (Gao et al., 2023).

La Inteligencia Artificial Generativa (IAGen) ha surgido como una tecnología disruptiva cuya amplia adopción impactó rápidamente en el contexto educativo. El empleo de recursos provenientes de IAG en estos ámbitos puede generar oportunidades de su uso por parte de docentes y estudiantes, que pueden favorecer los procesos de enseñanza-aprendizaje (Flores-Vivar y García-Peñalvo, 2023). Particularmente, la IA conversacional (IAConv), con capacidad de responder preguntas formuladas en lenguaje natural de manera amigable, fluida y coherente, representada por chatbots como ChatGPT[1], Bard/Gemini[2] o Copilot[3], se ha convertido en una tecnología con el potencial de llevar a cabo una transformación profunda

[1] https://chat.openai.com/

[2] https://gemini.google.com/

[3] https://copilot.microsoft.com/

en las prácticas educativas, tanto en la formación docente como dentro de las aulas. Los modelos fundacionales, que subyacen a estos productos de la IA conversacional, es decir los LLMs, son pre-entrenados con cantidades masivas de textos para capturar las relaciones intrínsecas entre las palabras y patrones del lenguaje, para generar nuevos textos a partir de ello. Las facilidades en cuanto a la generación de contenido, a partir de los cada vez más potentes LLMs, y la simplicidad de la interacción en lenguaje natural, ofrecen múltiples oportunidades para la asistencia activa a estudiantes y docentes en los procesos de enseñanza-aprendizaje. Sin embargo, estudios recientes respecto a las capacidades de los LLMs, para generar contenido en entornos educativos, han mostrado que presentan una serie de problemáticas que necesitan ser abordadas para capitalizar su potencial (Gao et al., 2023, Corica et al., 2024, Parra et al. 2024, Sureda et al. 2023).

Por su naturaleza probabilística, los LLMs producen contenido que, si bien responde en forma general a la instrucción recibida, no siempre lo hace con exactitud o de una forma que se ajuste a las necesidades de la tarea que se esté desarrollando. Es por ello, que en este trabajo, nos enfocamos en entender cómo estos LLMs se desempeñan en la resolución de problemas, por un lado de geometría clásica y por el otro, de geometría fractal. El objetivo de esta investigación se dirige a estudiar las capacidades de modelos de IAGen de propósito general accesibles a todo público, como ChatGPT y otros, para el estudio de problemas de geometría clásica y geometría fractal de potencial aplicación para el estudio en la escuela secundaria. Exploramos las respuestas que pueden aportar a estas tareas recursos provenientes de la IAGen en comparación con las que ofrecen profesores de matemática para las mismas. El artículo se organiza de la siguiente manera: en la sección 2 se presenta el método general. Aquí se detallan las decisiones metodológicas comunes a los cuatro problemas geométricos analizados. La sección 3 se centra en los dos problemas de la geometría clásica: se presentan el problema del decágono (3.1), el problema de la pizza (3.2) con sus respectivos enunciados, aspectos metodológicos propios de ambos y luego, el análisis y la discusión de los resultados obtenidos de estos primeros dos problemas (3.3). La sección 4 se ocupa, de forma análoga, de los dos problemas de la geometría fractal: se presentan el copo de nieve de Koch (4.1), el triángulo de Sierpinski (4.2) con sus respectivos enunciados, aspectos metodológicos de cada problema y luego, el análisis y la discusión de los resultados obtenidos de estos primeros dos problemas (4.3). En la sección 5 se discuten de forma integrada los resultados de las secciones anteriores y finalmente, en la sección 6 se presentan de forma concisa las conclusiones obtenidas.

2. Método general

Dos de los problemas considerados en este trabajo corresponden a la geometría clásica y otros dos, a la geometría fractal. Estos problemas se detallan en las secciones (y sus correspondientes sub-secciones) siguientes y se los denomina, respectivamente: el problema del decágono (3.1), el problema de la pizza (3.2), el copo de nieve de Koch (4.1) y el triángulo de Sierpinski (4.2).

Los cuatro problemas fueron propuestos a tres chatbots diferentes, cada uno de ellos soportado por un modelo de lenguaje distinto. En estos casos se utilizó OpenAI ChatGPT, usando GPT 3.5 (Brown et al., 2020) como modelo de lenguaje, Microsoft Bing Chat o

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

Copilot, utilizando GPT 4 (Open AI, 2023), y Google Gemini, modelo sucesor de PaLM (Chowdhery et al., 2022). En todas las situaciones, se propuso como primer *prompt* o instrucción la resolución de los problemas tal como se enuncian en las siguientes secciones y subsecciones correspondientes. En todos ellos, excepto en el problema del decágono (3.1), se realizaron tres interacciones adicionales a continuación del prompt inicial. Esto permitió obtener respuestas más amplias a las obtenidas en un principio. El segundo *prompt* empleado para los tres problemas restantes fue: *¿Podrías demostrar matemáticamente los resultados obtenidos?* La pregunta *¿Podrías generar el código necesario para realizar la gráfica correspondiente?* se empleó como tercer *prompt* tanto en el problema de la pizza (3.2) como en el problema del copo de nieve de Koch (4.1) mientras que el tercer *prompt* para el problema del triángulo de Sierpinski se refirió al diseño de una propuesta para enseñar fractales en el nivel secundario de Argentina. Este aspecto, por cuestiones de espacio, no se analiza en este trabajo. De cada una de las respuestas obtenidas por los chatbots se generaron categorías de análisis de manera inductiva, es decir, emergentes de los datos, basadas en patrones o recurrencias presentes (Isaza, 2002). Por este motivo, las categorías y subcategorías usadas para el análisis de cada problema se detallan seguidas a cada uno.

El análisis realizado para ambos tipos de problemas de geometría, clásica y fractal, se orientó a responder las siguientes preguntas de investigación: (1) ¿Qué tipo de soluciones matemáticas proporcionan las IAGen a los problemas planteados?, (2) ¿Las soluciones de IAGen y los profesores de Matemática difieren en rigor, procedimiento y justificación matemática? y (3) ¿IAGen es capaz de proporcionar justificaciones matemáticas de las soluciones al mismo nivel que los profesores de matemáticas?

3. Problemas de la geometría clásica: resolución, análisis y discusión de resultados

La Geometría clásica (GC) se define como la ciencia de las figuras geométricas, lo cual incluye objetos como puntos, ángulos, segmentos, rectas, superficies, etc (Alexander y Koeberlein, 2011). Se destaca que la GC resulta ser útil para el estudio debido a que permite modelar y describir el espacio físico, cobrando relevancia para el estudio de problemas de diversas disciplinas (física, biología, química, astronomía, geología, etc.), como así también integrándose para el estudio de problemas de otras áreas de la matemática como el análisis, el álgebra, la aritmética, la estadística, etc. (Bressan et al., 2000). Esta codisciplinaridad vincula experiencias individuales y grupales que producen diferentes niveles de sofisticación del conocimiento, útiles para resolver problemas, producir obras de arte, interpretar hechos o dar explicaciones, entre otras cosas (Camargo y Acosta, 2012).

En la actualidad hay una fuerte presencia de la geometría en los diferentes diseños curriculares, pero pese a ello, diversos autores destacan su ausencia en las aulas (Barrantes y Balletbo, 2012; Fernández-Nieto, 2018). La enseñanza de la geometría en Argentina no es ajena a este fenómeno (Corica y Marin, 2014). Entre las razones de la pérdida de espacio y sentido del estudio de la geometría, tanto en los colegios como en la formación docente, se encuentran la dificultad de los docentes en hallar o plantear verdaderos problemas geométricos; la falta de claridad en el sentido de las nociones geométricas en los diseños curriculares (Itzcovich, 2005); la elección docente de desplazar el estudio de la geometría, priorizando otras áreas como el álgebra, la aritmética o el estudio de funciones (Abrate et al., 2006; Corica y Marin, 2014; Itzcovich, 2005; Perez y Guillén, 2007). Es interesante estudiar las respuestas de los chabots a problemas de Geometría clásica ya que estos se diferencian de los problemas numéricos analíticos, en la necesidad de generación y/o interpretación de representaciones gráficas en el plano. Por ejemplo, en la construcción de un decágono o en la partición de un círculo en tres regiones de igual superficie (problema de la pizza).

3.1. Problema del Decágono

El siguiente problema de Geometría, al que denominamos el problema del decágono, es un problema planteado en las Olimpiadas de mayo de 2018[4] (Fauring y Gutierrez, 2020), un concurso Iberoamericano de Matemáticas. Este concurso tiene dos niveles, el primer nivel es para estudiantes que en el año anterior al concurso tengan menos de 13 años al 31 de diciembre, que es el nivel donde se presentó este problema. El enunciado del mismo es el siguiente:

> *Sea ABCDEFGHIJ un polígono regular de 10 lados que tiene todos sus vértices en una circunferencia de centro O y radio 5. Las diagonales AD y BE se cortan en P y las diagonales AH y BI se cortan en Q. Calcular la medida del segmento PQ.*

La resolución propuesta desde la misma competencia se basa en la representación gráfica del decágono y la identificación del segmento que se propone calcular (PQ). Esto implica el trazado de segmentos que unen los vértices del decágono con el centro del mismo y sus diagonales. El análisis de los triángulos y trapecios que resultan de las construcciones permite concluir que los mismos son isósceles. A partir de este análisis se concluye que el segmento solicitado tiene la misma longitud que el radio de la circunferencia en el que se encuentra inscripto el decágono. Esta resolución permite hallar el valor exacto de la longitud del segmento PQ, que es 5.

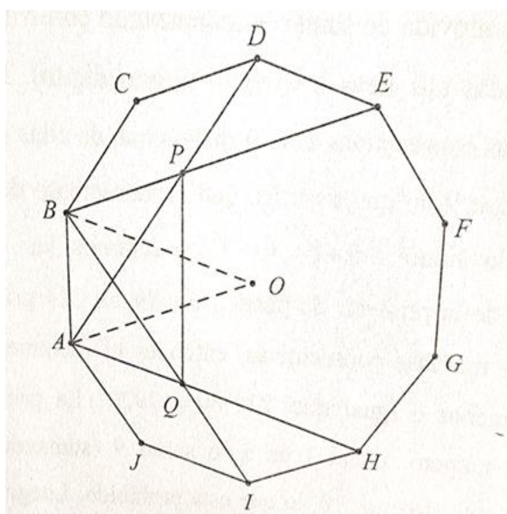

Figura 1: Representación del problema del decágono

Fuente: Fauring y Gutierrez (2020)

[4] https://www.oma.org.ar/enunciados/enunciados_Mayo 2018.pdf

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

El método de análisis para este caso consistió en identificar los errores cometidos por los chatbots en la resolución del problema y se los agrupó en categorías (Parra et al., 2024). Se consideraron para ello tres respuestas de cada chatbots, de manera de ponderar la aleatoriedad en las mismas, contabilizando un total de 9 soluciones diferentes al mismo problema. Las categorías de los errores identificadas son las siguientes:

Errores de construcción: provenientes de la representación en el plano de los elementos geométricos indicados en la respuesta, por ejemplo, asegurar que un ángulo central es de 72° cuando en realidad la amplitud es otra.

Errores conceptuales: provenientes de definiciones incorrectas, aplicación de propiedades sin garantizar las condiciones necesarias, confundir unidades de longitud con las de amplitud, etc. Por ejemplo, aplicar el Teorema de Pitágoras en triángulos que no son rectángulos.

Contradicciones: provenientes de inconsistencias entre deducciones y representaciones en el plano. Por ejemplo, deducir que un ángulo es agudo mientras que en la representación construida a partir de esa deducción es un ángulo llano.

3.2. Problema de la Pizza

El problema de la pizza se deriva de un teorema matemático denominado, análogamente, *El Teorema de la pizza*. Este teorema, propuesto por el matemático Upton, en 1967, fue en principio publicado como un problema a resolver y luego derivó en el teorema donde se propone particionar el círculo (pizza) de forma análoga a la propuesta original, pero desplazando del centro la intersección de las líneas de corte. El enunciado del problema es el siguiente:

La pizza es un alimento que forma parte de nuestra dieta y que todos sabemos partir. El corte tradicional en forma de sector circular es el más extendido, pero ¿es la única forma de dividir una pizza en tres partes iguales? (Camacho et al., 2015).

Este problema no es de solución única ni inmediata, siendo su demostración más conocida la propuesta por Carter y Wagon en el libro *Proofs without words II* (Nelsen, 1993). Para el corte que se proponga, es necesario detallar la forma de hacer el corte, la construcción geométrica del mismo y validar que las tres superficies generadas tengan igual área. Esto se realiza con conocimientos y procedimientos geométricos de diferentes niveles de complejidad dependiendo del tipo de corte propuesto. Por ejemplo, algunos posibles cortes son: división de la pizza en trozos tradicionales (sectores circulares), en trozos concéntricos (coronas circulares), en forma de lágrima, usando rectas paralelas, en forma de “T”, en forma de “V” y la división de la pizza a partir del Mosaico de disco monoédrico. La Figura 2 ilustra algunas de estas alternativas.

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

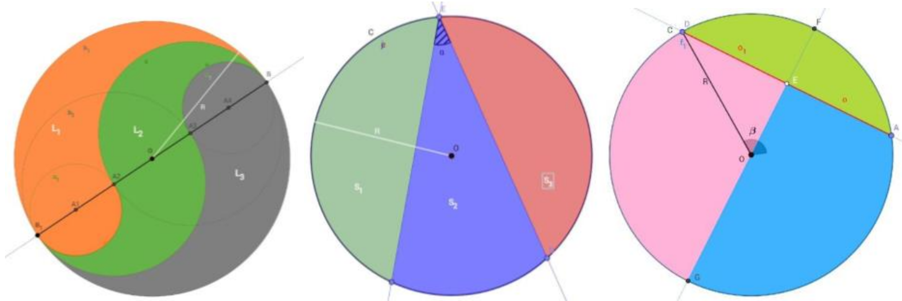

Figura 2: (producción propia). Corte en forma de lágrima (izq.), corte en forma de "V" (central) y corte en forma de "T" (der.)

El método de análisis empleado para este problema consistió observar los siguientes aspectos:

Tipo de corte propuesto: dadas las alternativas de resolución anteriores, se evaluó la capacidad de los chatbots de brindar una o más soluciones al problema planteado más allá del corte típico.

Marco preponderante utilizado: el marco general utilizado para resolver el problema, entre los que se cuentan los enfoques: algebráico, geométrico, aritmético, coloquial, gráfico y empírico. Estos marcos se definen en relación a las técnicas que se proponen para poder desarrollar el corte seleccionado y la justificación de la respuesta.

Nivel de validación matemática: en este punto se analizó el procedimiento propuesto para validar matemáticamente que las tres superficies de la pizza generadas por el corte seleccionado tienen igual área. Las mismas fueron desde la categoría nula (no se propuso ninguna justificación para la solución propuesta), luego validaciones bajas o medias, hasta una validación máxima o alta (donde la justificación matemática se centró en calcular algebraicamente el área de cada superficie de pizza, así como la demostración de igualdad de las áreas de cada superficie).

Tipo de resolución: este punto corresponde a los razonamientos explicitados en el desarrollo de la resolución y al tipo de resultados que éstos permiten obtener. Los mismos se clasificaron en procedimientos que permiten llegar o no (aunque sea parcialmente) a resultados correctos, incorrectos o si se advirtieron inconsistencias, independientemente del tipo de marco empleado y del nivel de validación ofrecida.

3.3. Análisis y discusión de resultados de los problemas de geometría clásica

Respecto a la pregunta (1): ¿Qué tipo de soluciones matemáticas proporcionan las IAGen a los problemas dados? El desempeño general de los chatbots en la generación de un texto para responder el problema del decágono fue algo decepcionante, ya que no alcanzaron en su mayoría la respuesta correcta y cometieron un número considerable de errores, de diferente tipo en la descripción de la solución. Según la categorización de errores propuesta, que incluye errores de construcción, conceptuales y contradicciones, se observó que la mayoría se identificó en las dos primeras categorías, mientras la aparición de contradicciones fue más infrecuente. Los errores de construcción están relacionados específicamente con la construcción en el plano, ya que tiene que ver con la traducción de una especificación

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

geométrica, dada en el texto, a una representación gráfica del mismo. Mientras que los errores conceptuales devienen de la incorrecta aplicación de nociones geométricas. Para este problema, no se realizó una comparación con las respuestas de los profesores, por lo que no permite responder a las preguntas (2) y (3).

Para el problema de la pizza, teniendo en cuenta la pregunta de investigación (1), se compararon las soluciones de los chatbots con las de ocho profesores de matemática en servicio (en el nivel secundario de Argentina). Las respuestas se compararon según los tipos de cortes propuestos, el marco de resolución predominante y en términos del tipo de resolución, clasificándolas en respuestas correctas, incorrectas e inconsistentes. Se encontró que las respuestas de los profesores fueron mayormente correctas y sin inconsistencias. En cambio, hubo un alto número de respuestas incorrectas entre las generadas por los tres chatbots, al igual que inconsistencias. Solo ChatGPT proporcionó alguna respuesta correcta en este problema. Entre las inconsistencias se cuentan, por ejemplo, proponer un tipo de corte (espiral) y luego resolverlo con otro (sectores circulares).

Respecto a la segunda pregunta de investigación (2) ¿Las soluciones de IAGen y los profesores de Matemática difieren en rigor, procedimiento y justificación?, un aspecto analizado en este caso fue la variedad en las soluciones al problema de la pizza propuestas por unos y otros. En total se mencionaron doce tipos de cortes diferentes, siendo el de mayor frecuencia el denominado sector circular (y sus variantes). Los profesores fueron quienes ofrecieron mayor diversidad en los tipos de cortes propuestos (7 tipos de cortes diferentes), en comparación con los propuestos por los chatbots, quienes realizaron propuestas de a lo sumo 3 tipos de cortes diferentes. Aun cuando los chatbots ofrecieron menos variedad de respuestas, una de ellas, realizada por ChatGPT, incluyó un corte en rectángulos, para una pizza que el chatbot supone rectangular y que valida correctamente en el marco algebraico. La particularidad de esta respuesta es que, dado que el enunciado no especifica explícitamente una pizza circular, es interesante notar que los chatbots respondieron en su mayoría de acuerdo con la forma más probable de la pizza, mientras que los profesores lo hicieron debido a que se sostiene sobre una idea adquirida socialmente respecto del formato circular estándar de una pizza.

Continuando con la pregunta (2), se identificaron también los marcos de resolución utilizados en cada caso. Las respuestas de los profesores mostraron mayor amplitud en cuanto a los marcos posibles de resolución que los chatbots. Estos últimos proveyeron casi exclusivamente respuestas en lo que podríamos llamar un marco de resolución coloquial. Es decir, una explicación en palabras en lenguaje natural, que no incluye construcciones matemáticas ni expresiones algebraicas que contribuyan al entendimiento del problema. Por otro lado, debido a que los profesores cuentan con una mayor cantidad de herramientas matemáticas, dada su formación específica, recurrieron a marcos geométricos, numéricos y gráficos. Es importante notar en este punto que, a pesar de que el marco algebraico es matemáticamente el más adecuado para la resolución de este problema, no fue el preponderantemente elegido por los profesores, mientras que por parte de los chatbots fue usado solamente en una respuesta.

El problema de la pizza permitió explorar también las soluciones apuntando a dar respuestas a la pregunta (3), referida a los niveles de las justificaciones matemáticas y la validación de las soluciones. En este caso, sobresalió la ausencia de validaciones en las soluciones de los chatbots, siendo ChatGPT el único que consiguió responder en una de las oportunidades con una validación considerada de alta calidad, mientras que Gemini alcanzó alguna validación baja. El tipo de validación más típico entre los profesores fue también del tipo baja, solo en algún caso nula, y también consiguieron brindar algunas soluciones con validación media y alta. Es importante notar que a los chatbot se les solicitó una validación en el segundo *prompt*, mientras que los profesores en todos los casos ofrecieron al menos una validación aún sin haberse pedido explícitamente. Puede atribuirse esta diferencia a que dentro de la formación de un profesor de Matemática, está instalado el supuesto que toda afirmación debe demostrarse.

4. Problemas de la geometría fractal: resolución, análisis y discusión de resultados

La Geometría Fractal (GF) comprende el estudio de conjuntos irregulares, que surgieron a finales del siglo XIX y se caracterizan por tener propiedades geométricas y analíticas particulares centradas en el concepto de autosemejanza (Chavil et al., 2020). Los fractales se pueden encontrar en diversas formas que surgen en la naturaleza, lo que conduce a varias aplicaciones prácticas. Estas características, que resultan interesantes de estudiar, no sólo desde la matemática en sí misma, sino también por la gran cantidad de aplicaciones reales que tienen los fractales en la vida cotidiana, parecen, sin embargo, no ser suficientes para incluirlos en el ámbito escolar. En los lineamientos oficiales (establecidos por el Ministerio de Educación) de la escuela secundaria de Argentina se reconoce la importancia del estudio de fractales. Incluso algunos de los 24 diseños curriculares del país - por ejemplo el de la Provincia de Buenos Aires - proponen estudiarlo, pero ese estudio rara vez se concreta dentro de las aulas. Por un lado, hay quienes aluden que una de las posibles razones de esta ausencia, es la formación de los profesores de matemática en el área de la geometría fractal (Chavil et al., 2020; Martin et al., 2019). Por otro lado, hay quienes refieren a una falta de recursos específicos para enseñar esta geometría en la escuela. Es interesante estudiar las respuestas de los chabots a problemas de Geometría fractal ya que estos se diferencian de los problemas de Geometría Clásica. Por ejemplo, los fractales son curvas con perímetro infinito, pero encierran áreas finitas (copo de nieve de Koch) e incluso áreas nulas (Triángulo de Sierpinski) (Figura 3)

Figura 3: Copo de Koch (izq.). Triángulo de Sierpinski para tres iteraciones (der.)

4.1. El copo de nieve de Koch

El problema del Copo de nieve de Koch se puede enunciar como sigue:

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

El copo de nieve de Koch (también conocido como estrella de Koch o isla de Koch) es una curva fractal y uno de los primeros fractales que se han descrito. Se publicó en un artículo de 1904 titulado Sobre una curva continua sin tangentes, construible a partir de la geometría elemental por el matemático sueco Helge von Koch. El copo de nieve de Koch se puede construir comenzando con un triángulo equilátero, sustituyendo cada lado por la curva de Koch. Calcular el perímetro y el área del copo de Koch.

El método de análisis de este caso se concentró en el método de construcción del fractal, como así también los cálculos de perímetro y área:

Construcción del fractal: en lo que concierne a la construcción del copo de nieve de Koch, se analiza si ésta fue hecha a partir del triángulo equilátero (es decir, asumiendo en el procedimiento que los tres lados son iguales), si se definió y validó la partición de cada lado del triángulo en tres segmentos de igual longitud y si fue capaz de obtener una representación gráfica precisa del copo de nieve.

Cálculo del perímetro del fractal: en este cálculo se analiza la construcción de la sucesión requerida, en la que cada término es el perímetro en una iteración y cada iteración proviene de tomar el segmento central de la trisección de cada lado y construir allí un nuevo triángulo equilátero, su límite cuando la sucesión tiende a infinito y si se alcanza el valor correcto del perímetro dado por el cálculo del límite de la sucesión, que es infinito.

Cálculo del área del fractal: de forma análoga al punto anterior, este cálculo involucra tres aspectos a considerar: el cálculo de la altura y base del triángulo, las sumas sucesivas de las áreas de los triángulos que se agregan a cada iteración (lo que conduce, inevitablemente, a una serie geométrica infinita), la determinación de la razón de la serie y finalmente el cálculo del área como resultado de dicha serie, que da como resultado: $\frac{2\sqrt{3}}{5}L^2$.

4.2. El Triángulo de Sierpinski

En el caso de Triángulo de Sierpinski, se utilizó el siguiente enunciado como primer actividad (prompt), para luego pedir una demostración matemática como en los problemas anteriores:

El triángulo de Sierpinski es un fractal, con Waclaw Sierpinski en 1915 como su mentor. Para su construcción se parte de un triángulo equilátero. El primer paso consiste en dividirlo en cuatro triángulos equiláteros iguales (lo que se consigue uniendo los puntos medios de los lados) y eliminar el triángulo central, es decir quedándose con los tres triángulos equiláteros de los vértices. El segundo paso de la construcción consiste en hacer lo mismo que se ha hecho en el primer paso sobre cada uno de los tres triángulos obtenidos en el paso anterior. Y se repite el proceso infinitas veces, obteniendo como resultado final el triángulo de Sierpinski. ¿Puedes calcular el perímetro y el área del triángulo de Sierpinski?

De manera similar al estudio realizado con el copo de nieve de Koch, se consideraron las siguientes tres categorías de análisis:

Construcción del fractal: en este caso, la construcción del fractal tiene como aspecto más general la cantidad de iteraciones detalladas antes de presentar el término general, sea del perímetro o del área.

Cálculo del perímetro del fractal: esto involucra en este problema determinar si se construye la sucesión correspondiente, si se presenta su término general y, finalmente, si se da el valor correcto del perímetro que debe ser infinito.

Cálculo del área del fractal: análogamente para el área se observa si se construye la sucesión correspondiente, si se presenta su término general y si se alcanza el valor correcto del área, que es cero.

4.3. Análisis y discusión de resultados de los problemas de geometría fractal

En el caso de estos problemas de Geometría fractal se enfocó el análisis de las respuestas de chatbots y profesores en su capacidad de describir el método de construcción, cálculo del área y del perímetro de cada fractal. Con relación al fractal copo de nieve de Koch y considerando la primera pregunta de investigación, (1) ¿Qué tipo de soluciones matemáticas proporcionan las IAGen a los problemas planteados?, es posible afirmar que ChatGPT no da el valor correcto del perímetro ni del área del copo de Koch. Proporciona las expresiones que permitirán calcular ambos valores, pero no indica el resultado. De los cuatro chatbots restantes, todos dan el valor correcto del perímetro, mientras que sólo dos de ellos, Copilot creativo y Copilot preciso, obtienen el valor correcto del área. Por su parte, el profesor da ambos valores correctos, tanto del perímetro como del área.

En respuesta a la pregunta de investigación (2), ¿Las soluciones de IAGen y los profesores de Matemática difieren en rigor, procedimiento y justificación matemática? Se analizaron los procedimientos de construcción de ambos fractales que ofrecieron los chatbots en relación con la que presentaron los profesores. En el caso del copo de Koch, los resultados de los chatbots fueron contrastados con las respuestas de un profesor de matemática. Este profesor, Gemini y las tres versiones de Copilot (equilibrado, creativo y preciso) proponen un triángulo equilátero genérico de lado L. En la versión equilibrada además de explicitar que el triángulo es equilátero, indica que cada lado del mismo se reemplaza por la curva de Koch. Para continuar, tanto el profesor como los chatbots se refieren a dividir cada lado del triángulo en tres partes iguales, pero ninguno menciona cómo hacerlo y validarlo desde la matemática. En cuanto al cálculo del perímetro, el profesor y Copilot en dos de sus modalidades, creativo y equilibrado, construyen la sucesión, obteniendo de forma correcta el término general de la misma, mientras que el resto de los chatbots no lo describen. El profesor afirma que la sucesión tiende a infinito y que entonces es divergente sin validar este resultado. De los chatbots, salvo Copilot creativo, el resto no emplea la noción de límite de manera explícita para calcular el perímetro. Para calcular el área, ni el profesor ni los chatbots presentan la construcción de la serie geométrica necesaria para determinar el valor del área. Aún sin proveer validación, Copilot creativo y preciso, determinan el valor correcto de la razón geométrica. Respecto al resultado del área, tanto el profesor como los chatbots, exceptuando a Gemini, obtienen el valor correcto de la misma. Es importante aclarar que la determinación de la sucesión geométrica y de su razón no debe considerarse como el sinónimo del cálculo del área. Se trata de procedimientos y objetivos matemáticos diferentes. Si bien, Gemini indica que el área es finita, concluye en un valor erróneo. Copilot más creativo obtiene el valor esperado para el área, pero presenta una contradicción en la notación.

Por último, para el copo de Koch se analizaron las representaciones gráficas del fractal. El profesor propone las imágenes correctas para las primeras tres iteraciones, en cambio los chatbots (a quienes se les pidió el código para generar un gráfico) evidenciaron dificultades para alcanzar una gráfica correcta. En primer lugar, en casi todos los casos, los

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

chatbots requirieron más de una interacción para alcanzar un gráfico correcto. En algunos casos comenzaron graficando la curva de Koch, en lugar del copo de nieve. Luego, al indicar que el copo es una figura cerrada a diferencia de la curva, mostró la falta de interpretación de lo que significa una figura cerrada, tratando de solucionar la gráfica primero coloreando la superficie sin cerrar la figura y luego, cerrando la figura, pero perdiendo las propiedades del copo de nieve de Koch. Por otro lado, una de las respuestas (Copilot creativo) arrojó un resultado interesante, ya que generó un código que solicita al usuario el número de recursiones que desea, y genera el gráfico de acuerdo a ello. Es decir, que logró extraer el concepto de recursión del fractal y lo incorporó a la interacción con el usuario en la generación de gráficos.

En resumen, como respuesta a la pregunta de investigación 2 para el caso del fractal de Koch, obtuvimos diferencias significativas entre las soluciones de IAGen y los profesores de Matemática, en cuanto al rigor, procedimientos y justificaciones. Respecto al procedimiento, ChatGPT es el más débil desde el punto de vista matemático. En un punto intermedio, identificamos al profesor, ya que sus procedimientos son matemáticamente adecuados. Copilot creativo presenta un procedimiento matemático más sólido, convirtiéndose en el chatbot más adecuado para resolver este problema. Respecto a la validación, Copilot creativo es el que más frecuentemente presenta justificaciones matemáticas adecuadas, mientras que el profesor y ChatGPT son los que menos lo hacen. Respecto a las gráficas, en la mayoría de los casos, los chatbots confundieron la curva de Koch con el copo de nieve.

Siguiendo con el copo de Koch, y en busca de dar respuesta a la pregunta de investigación (3), ¿IAGen es capaz de proporcionar justificaciones matemáticas de las soluciones al mismo nivel que los profesores de matemáticas? Según se desprende de lo anterior, efectivamente es posible afirmar que las respuestas de algunos de los chatbots estuvieron al nivel de una resolución matemática humana. En relación ahora al fractal Triángulo de Sierpinski y considerando la primera pregunta de investigación, (1) ¿Qué tipo de soluciones matemáticas proporcionan las IAGen a los problemas planteados?, es posible afirmar que todos los chatbots dan el valor correcto del perímetro a excepción de Copilot creativo, quien afirma que el perímetro es cero, lo cual es un error pues el perímetro es infinito. Respecto al área, todos los chatbots dan el valor correcto (que es cero) excepto Copilot creativo quien afirma que el área del triángulo de Sierpinski tiene al área inicial, lo cual es un error. Los tres profesores concluyeron en el valor correcto tanto del perímetro como del área.

Con relación a la pregunta (2), ¿Las soluciones de IAGen y los profesores de Matemática difieren en rigor, procedimiento y justificación matemática? y respecto a la construcción del triángulo de Sierpinski, los procedimientos desarrollados por los profesores de matemáticas y los chatbots difieren en el número de iteraciones realizadas. Los tres profesores realizan 4 iteraciones antes de formular la expresión general tanto del perímetro como del área; mientras que cuatro de los chatbots (ChatGPT y Copilot en sus tres versiones) presentan directamente el término general sin ninguna iteración previa. Sólo Gemini propone cuatro iteraciones antes de introducir el término general. En la construcción de la secuencia del perímetro, dos profesores proponen una solución adecuada. El tercero construye una serie pero esto no es correcto ya que el perímetro no representa una suma infinita, sino una sucesión. Por otro lado, ninguno de los chatbots alude explícitamente a la noción de sucesión, siendo éste el concepto clave para determinar el perímetro. Considerando que, como ya se indicó en la respuesta a la primera pregunta, algunos de los chatbots alcanzan valores correctos habiendo presentado expresiones algebraicas incorrectas (o no presentarlas) en el procedimiento, es posible suponer que esto se debe a la forma de construcción del texto, donde la solución (el perímetro infinito) tiene mayor presencia en los textos usados para entrenar los modelos que los pasos intermedios. Es por esto que aciertan en esta parte de la solución. Intuitivamente, los profesores pueden inferir que al eliminar áreas en cada iteración, el valor tenderá a cero. Sin embargo, los chatbots llegaron a esta solución por ser la más probable en los datos observados, a pesar de la construcción hecha en los pasos intermedios del razonamiento. Esto conduce a inferir en diferencias de procedimiento entre las IAGen y los profesores pues los primeros no logran presentar un resultado que se desprenda del procedimiento realizado.

En cuanto a las validaciones, ni los profesores ni dos de los chatbots (Copilot equilibrado y Copilot preciso) ofrecen justificaciones matemáticas al problema del triángulo de Sierpinski. En cambio, tres de los chatbots ofrecen al menos una validación en el orden de lo coloquial. ChatGPT propone esta forma coloquial, pero lo hace incorrectamente, aludiendo a que el área se reduce a la mitad en cada iteración porque estamos eliminando un triángulo central. La respuesta de Copilot creativo sugiere consultar demostraciones en sitios web externos y ofrece algunos enlaces. Finalmente, Gemini justifica a partir de series infinitas, lo cual también es incorrecto. Nuevamente, si bien los chatbots no alcanzaron justificaciones avanzadas, intentaron hacerlo desde lo coloquial. Es posible entonces identificar aquí diferencias en el nivel de validaciones a favor de los chatbots. Esta diferencia podría deberse a que los chatbots aprenden textos provenientes de Internet, que contienen demostraciones matemáticas. En cambio, en la práctica docente, se encuentra naturalizado el énfasis en la ejercitación práctica omitiendo o dejando de lado, las demostraciones de los resultados obtenidos (Bosch y Gascón, 2009). Esto podría ser un motivo por el cual estos profesores, que se desempeñan en el nivel secundario de Argentina no ofrecieron validaciones de sus soluciones.

Para finalizar con el triángulo de Sierpinski, y en busca de dar respuesta a la pregunta de investigación (3), ¿IAGen es capaz de proporcionar justificaciones matemáticas de las soluciones al mismo nivel que los profesores de matemáticas? Según se desprende de lo anterior, efectivamente es posible afirmar que las respuestas de algunos de los chatbots brindaron una resolución matemática mejor que la de los profesores de matemática.

5. Discusión general

Del análisis general de los cuatro problemas abordados y el análisis particular realizado en cada uno de ellos, enriquecido por las diferencias entre las Geometría clásica y la fractal, se pueden esbozar las siguientes respuestas sintéticas a las preguntas de investigación planteadas inicialmente. En términos de los tipos de respuesta de la IAGen y respondiente a la pregunta de investigación (1) se puede concluir que las IA conversacionales representadas por los chatbots cometen una serie de errores en el razonamiento geométrico

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

atribuible a distintas razones. La primera, es la presencia de errores conceptuales debidos a la aplicación incorrecta de ciertas nociones o conceptos geométricos (por ejemplo, la noción de figura cerrada en la gráfica de un fractal). La segunda, la presencia de inconsistencias por la falta de interpretación del contexto geométrico sobre el que se está trabajando (por ejemplo, deducir que un ángulo es agudo cuando es llano). Y, la tercera, la presencia de inconsistencia entre las distintas partes que componen una solución (por ejemplo, llegar a una solución correcta con pasos intermedios incorrectos). La mayoría de las fallas observadas en las respuestas al problema propuesto están relacionadas con dos críticas comunes que se hacen a los LLMs (Pavlick, 2023), la falta de estructura simbólica y la falta de *grounding*. Ambos aspectos cuestionan su capacidad para proporcionar representación y comprensión del lenguaje humano a pesar de contar con habilidades lingüísticas similares a las humanas.

La falta de estructura simbólica impide que el modelo realice un razonamiento formal y verifique los pasos del razonamiento. De ello surgen las inconsistencias y las respuestas como las obtenidas para el perímetro y área de los fractales cuyos valores son correctos a pesar de no proveer las estructuras algebraicas apropiadas para alcanzarlos. En este caso el LLM arma un texto donde la respuesta es la correcta, porque la misma es altamente probable en base a los textos observados durante el entrenamiento. En otras palabras, en muchos documentos dice que el perímetro del triángulo de Sierpinski es infinito, y es por ello que los chatbots incorporan este texto en la solución. Sin embargo, el texto intermedio de la respuesta no solo se construye con mayor incertidumbre, sino que además no sigue un proceso de deducción sino estocástico, los LLMs construyen el texto en base a la probabilidad de una secuencia de palabras.

La falta de grounding, es decir acceso o conciencia de los objetos en el mundo real a los cuales se refiere el lenguaje, conduce además a una mala interpretación de las nociones geométricas, incluso las más básicas. Esto es particularmente observable en los problemas geométricos que requieren razonamiento espacial y comprensión de representaciones visuales. Los profesores en este sentido, si bien pueden llegar a una solución incorrecta o incluso incompleta, es menos probable que cometan errores de interpretación en el espacio geométrico, de hecho, ellos no caen en inconsistencias como en las que recaen los chatbots.

En respuesta a la pregunta de investigación (2), en base a lo observado con el segundo problema de geometría clásica y los dos de fractales, también es posible realizar algunas observaciones respecto de los procedimientos provistos por unos y otros. En primer lugar, fue posible observar que los profesores ofrecieron mayor diversidad de soluciones (particularmente en el denominado problema de la pizza, que se prestaba para ello) y contaron con más cantidad de herramientas para aportar un marco teórico a las mismas. Los chatbots, por otro lado, recurrieron mayormente a explicaciones de los desarrollos en forma coloquial, lo cual es de esperar por ser una herramienta tecnológica ideada esencialmente para la construcción de lenguaje. Esto, sin embargo, resultó de todos modos ser algo efectivo en el caso de los problemas de Geometría fractal, en los cuales los profesores encontraron mayores dificultades también para desarrollar los procedimientos de cálculo requeridos.

En segundo lugar, nuevamente relacionado a la carencia de razonamiento simbólico que de soporte al desarrollo matemático y la forma en la que un LLM construye el texto, se puede mencionar que los chatbots en muchos casos se limitan a presentar la solución final (la cual adquieren por ser la de mayor presencia), pero fallan en varios puntos del desarrollo matemático. Por ejemplo, los profesores muestran varias iteraciones antes de formular la expresión general del triángulo de Sierpinski, mientras que los chatbots se limitan a presentarla sin ninguna iteración previa. Es importante destacar que el objetivo de una IA conversacional es crear un texto que se "parezca" a un razonamiento en el contexto del problema planteado, no el de dar una solución con exactitud. En este proceso también existe otro limitante en los datos de entrenamiento de los LLMs (Bubeck et al., 2023), ya que usualmente la descripción de la resolución de un problema, tal como se podría encontrar en la Web, no expresa el proceso de pensamiento que conduce a la solución (conjeturas, errores encontrados, vuelta atrás, correcciones, etc.), sino que son una exposición lineal de la solución encontrada. Un modelo entrenado con estos datos no podría entonces tampoco capturar este proceso y reproducirlo adecuadamente.

Finalmente, la pregunta de investigación (3) refería a las capacidades de los profesores y chatbots para brindar justificaciones y validaciones rigurosas desde un punto de vista matemático. En este sentido el resultado fue dispar para el problema de Geometría clásico y los dos de fractales. Para el primero, los profesores ofrecieron mejores formas de validación en términos generales, mientras que los chatbots prácticamente no pudieron hacerlos. En cambio, en los problemas de fractales ambos, profesores y chatbots, encontraron dificultades para alcanzar validaciones apropiadas. Los chatbots, dado su objetivo específico, que no es más que mantener una conversación, proveyeron validaciones en forma coloquial. En cambio, los profesores, posiblemente conscientes de la rigurosidad que requiere una demostración matemática y de la falta de adecuación de una respuesta verbal o coloquial, no ofrecieron validación alguna.

A continuación, la Tabla 1 presenta una síntesis con los resultados de ambas discusiones:

Tabla 1. Síntesis de resultados

Problema	**Geometría clásica**		**Geometría Fractal**	
	Del Decágono	De la pizza	Fractal de nieve de Koch	Triángulo de Sierpinski
Tipo de análisis	Se analizaron los tipos de errores.	Se analizó: -Tipos de cortes propuestos. -Marco de resolución predominante. -justificaciones matemáticas y la validación de las soluciones.	El análisis se concentró en el método de construcción del fractal, los cálculos de perímetro y los del área.	
Chatbots	Se observan tres tipos de errores. **-Errores de construcción.** **-Errores**	**Alto número de respuestas incorrectas, y con inconsistencias (*).**	**Construcción del fractal**. Proponen dividir cada lado del triángulo en tres partes iguales, pero	**Construcción del fractal.** Los chatbots (ChatGPT y Copilot en sus tres versiones) presentan

Corica, A. R., Sureda, P., Parra V., Shiaffino S. y Godoy, D. (2024). Educación e inteligencia artificial: desempeño de chatbots y profesores de matemática en la resolución de problemas geométricos. *Areté. Revista Digital del Doctorado en Educación,* 10 (ee), 119 – 139. https://doi.org/10.55560/arete.2024.ee.10.9

conceptuales. **-Contradicciones**	-Proponen tres **tipos de cortes** diferentes. **Marco de resolución** coloquial. No se utilizó el algebraico que es el óptimo. Ausencia de **validaciones**. ChatGPT es el único que propone una validación considerada de alta calidad, mientras que Gemini alcanzó alguna validación baja. (*) ChatGPT proporcionó alguna respuesta correcta	ninguno menciona cómo hacerlo y validarlo desde la matemática. **Cálculos del perímetro.** Copilot en dos de sus modalidades, creativo y equilibrado, construyen la sucesión, obteniendo de forma correcta el término general de la misma, mientras que el resto de los chatbots no lo describen. **Cálculos del área.** Todos los chatbots, exceptuando a Gemini, ofrecen el valor correcto. Pero ningún chatbots presenta la construcción de la serie geométrica necesaria para determinar el valor del área. Copilot creativo y preciso, determinar el valor correcto de la razón geométrica, sin ofrecer validación matemática. **Validaciones:** Copilot creativo es el que más frecuentemente presenta justificaciones matemáticas adecuadas, mientras que ChatGPT es el que menos lo hace.	directamente el término general sin ninguna iteración previa. Sólo Gemini propone cuatro iteraciones antes de introducir el término general. **Cálculos del perímetro.** Todos los chatbots dan el valor correcto del perímetro a excepción de Copilot creativo, quien afirma que el perímetro es cero, lo cual es un error pues el perímetro es infinito. En la construcción de la sucesión ninguno de los chatbots alude explícitamente a la noción de sucesión, siendo éste el concepto clave para determinar el perímetro. **Cálculos del área.** Respecto al área, todos los chatbots dan el valor correcto (que es cero) excepto Copilot creativo quien afirma que el área del triángulo de Sierpinski tiene al área inicial, lo cual es un error. **Validaciones:** dos de los chatbots (Copilot equilibrado y Copilot preciso) no ofrecen justificaciones matemáticas al problema del triángulo de Sierpinski. Pero, tres de los chatbots ofrecen al menos una

				validación en el orden de lo coloquial.
Profesores	Ninguno	8 (ocho profesores)	1 (un profesor)	3 (tres profesores)
	No Corresponde	**Respuestas correctas y sin inconsistencias.** -Ofrecieron mayor diversidad de **cortes** (7 siete). **Marco de resolución** geométrico, numérico y gráfico. No se utilizó el algebraico que es el óptimo. La **validació**n más típica fue del tipo baja, solo en algún caso nula, y también consiguieron brindar algunas soluciones con validación media y alta.	**Construcción del fractal.** Propone dividir cada lado del triángulo en tres partes iguales, pero no menciona cómo hacerlo y validarlo desde la matemática. **Cálculos del perímetro.** El profesor construye la sucesión, obteniendo de forma correcta el término general. Luego afirma que tiende a infinito y que entonces es divergente sin validar este resultado. **Cálculos del área.** Obtienen el valor correcto del área. No presenta la construcción de la serie geométrica necesaria para determinar su valor. **Validaciones:** el profesor es quien menos la presenta.	**Construcción del fractal.** Los tres profesores realizan 4 iteraciones antes de formular la expresión general tanto del perímetro como del área. **Cálculos del perímetro.** Los tres profesores concluyeron en el valor correcto del perímetro. En la construcción de la sucesión dos profesores proponen una solución adecuada. El tercero construye una serie, pero esto no es correcto. **Cálculos del área.** Los tres profesores concluyeron en el valor correcto del área. **Validaciones:** ninguno de los profesores ofrece justificaciones matemáticas al problema del triángulo de Sierpinski.

6. Conclusión

El desempeño general de los chatbots en la resolución de problemas geométricos no fue matemáticamente correcto, ya que cometieron un número considerable de errores e incluso presentaron algunas inconsistencias en sus respuestas. En cambio, los profesores aportaron mayormente respuestas correctas y sin inconsistencias. Además, las respuestas de los profesores mostraron mayor amplitud de técnicas de resolución que los chatbots, en aquellos casos donde era posible resolver los problemas de más de una manera. Los chatbots proveyeron casi exclusivamente respuestas en lo que podríamos llamar un marco de resolución coloquial. Esto se debe a que los problemas encontrados en los últimos dos aspectos obedecen, en parte, a la brecha entre el problema para el que se desarrolló el modelo (dialogar) y la tarea que se le está asignando (razonar un problema matemático). McCoy et al. (2023) destacan la importancia de no ver a los LLMs como resolvedores de problemas matemáticos, sino más bien como un sistema estadístico de predicción de palabras que se utiliza para resolver problemas matemáticos. Entonces, los fracasos pueden entenderse directamente en términos de un conflicto entre la tarea de predicción de la siguiente palabra y la de resolución de problemas. Respecto a las validaciones, en líneas generales, ni los profesores ni los chatbots aportaron justificaciones con el rigor deseable desde la Matemática. Los profesores en general no brindaron validaciones mientras que los chatbots ofrecieron mayores instancias de justificaciones, aunque quedando en gran parte, en lo coloquial.

Los resultados obtenidos en este estudio podrían tener varias implicaciones prácticas: pueden ser una fuente para que todos los profesores evalúen cómo utilizar correctamente las herramientas de IAGen, y para los formadores de profesores para introducirlos en los cursos de formación con valoraciones adecuadas de sus capacidades y limitaciones. También, puede ser útil para el diseño de nuevas herramientas de asistencia automática que aprovechen las técnicas IAGen para guiar a formadores, profesores y estudiantes en sus actividades de formación, de enseñanza y aprendizaje. En conclusión, es posible afirmar que las metodologías que incluyan IAGen como parte activa en la resolución de problemas de la geometría o como base para la generación de recursos para la enseñanza de este tema, pueden valerse de las observaciones del presente trabajo para su efectiva incorporación. Alentamos el uso de recursos provenientes de la IAGen siempre y cuando se analicen críticamente las respuestas que ofrecen.

Agradecimientos

Este trabajo fue parcialmente financiado por los proyectos PIP CONICET (2023/2025), PIO 2024/2025 PEIDyT 2024/2025, SECAT-UNCPBA.

Referencias

Abrate, R. S., Delgado, G. I., & Pochulu, M. D. (2006). Caracterización de las actividades de Geometría que proponen los textos de Matemática. *Revista Iberoamericana De Educación*, 39(1), 1–9. https://doi.org/10.35362/rie3912598

Alexander, D. C. & Koeberlein, G. M. (2011). *Geometría Quinta Edición*. México D.F.: Cengage Learning.

Barrantes, M. & Balletbo, I. (2012). Tendencias actuales de la enseñanza-aprendizaje de la geometría en educación secundaria. *Revista Internacional de Investigación en Ciencias Sociales*, 8(1), 25-42.

Bosch, M., Gascón, J. (2009). Aportaciones de la Teoría Antropológica de lo Didáctico a la formación del profesorado de matemáticas de secundaria. En M.J. González, M.T.

González & J. Murillo (Eds.), *Investigación en Educación Matemática XIII* (pp. 89-113). Santander: SEIEM.

Bressan, A., Bogisic, B. & Crego, K. (2000). *Razones para enseñar Geometría en la Educación Básica. Mirar, construir, decir y pensar…* Buenos Aires: Ediciones Novedades Educativas.

Brown, T., Mann, B., Ryder, N., … Amodei, D. (2020). Language models are few-shot learners. In: H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan and H. Lin (Edit.) *Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20)* (pp. 1877-1901). Curran Associates Inc.: NY.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., and Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *CoRR* abs/2303.12712.

Camargo, L. & Acosta, M. (2012). La geometría, su enseñanza. *Tecné, Episteme y Didaxis (TED)*, 32, 4-8.

Camacho Machín, M., De La Fuente Martínez, C., Gámez Ruiz, J. L., González López, Ma. J., Jara Martínez, P., Marín del Moral, A., Ortega del Rincón, T., Recio Muñiz, T. J., Rico Romero, L. & Ruiz Hidalgo, J. F. (2015). *Construcción de modelos matemáticos y resolución de problemas*. Secretaría General Técnica: España.

Chavil, D. & Romero, I. & Rodríguez, J. (2020). Introducción al concepto de fractal en enseñanza secundaria usando realidad virtual inmersiva. *Desde el Sur*, 12(2), 615-629.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A.…, and Fiedel, N. (2022). PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240), 1-113.

Corica, A. & Marin, E. (2014). Actividad de estudio e investigación para la enseñanza de nociones de geometría. Sociedad *Canaria "Isaac Newton" de Profesores de Matemática; Números*, 85(3), 91-114.

Corica, A., Parra, V., Sureda, P., Schiaffino, S., Godoy, D (2024). Fractal de Koch: análisis de respuestas de IA generativa y un profesor de matemática. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología (TE&ET)*. 89-99 http://doi:10.24215/18509959.37.e8.

Gao, J., Pi, R., Zhang, J., Ye, J., Zhong, W., Wang, Y., Hong, L., Han, J., Xu, H., Li, Z., Kong, L. (2023). G-LLaVA: Solving geometric problem with multi-modal large language mode. arXiv:2312.11370

Fauring, P., Gutierrez, F. (2020). O*limpiadas de Mayo - XVII a XXIV*. Buenos Aires, Argentina: Red Olímpica.

Fernández-Nieto, E. L. (2018). La geometría para la vida y su enseñanza. *AiBi Revista De Investigación, Administración E Ingeniería*, 6(1), 33-61. https://doi.org/10.15649/2346030X.475

Flores-Vivar, J., García-Peñalvo, F. (2023) Reflections on the ethics, potential, and challenges of artificial intelligence in the framework of quality education (SDG4), *Comunicar*, 74, 37-47.

Isaza G. (2002) *Análisis, Interpretación y Construcción Teórica en la Investigación Cualitativa*. Centro de educación a distancia. Universidad de Manizales.

Itzcovich, H. (2005). *Iniciación al estudio didáctico de la geometría: De las construcciones a las demostraciones*. Buenos Aires: libros del Zorzal.

Martin, N., Parra, V., Fanaro, M. (2019). Enseñanza de fractales a partir de preguntas: descripción de una experiencia en un curso de matemática del último año de la escuela secundaria. *Épsilon - Revista de Educación Matemática*, 102, 25-34.

McCoy, R. T., Yao, S., Friedman, D., Hardy, M. y Griffiths, T. L. (2023). Embers of autoregression: Understanding large language models through the problem they are trained to solve. *CoRR* abs/2309.13638.

OpenAI: GPT-4 technical report. (2023). ArXiv, abs/2303.08774

Parra, V., Sureda, P., Corica, A., Schiaffino, S., Godoy, D. (2024). Can generative AI solve Geometry problems? Strengths and weaknesses of LLMs for geometric reasoning in Spanish (2024). *International Journal of Interactive Multimedia and Artificial Intelligence*, 8(5), 65-74.

Pavlick, E. (2023). Symbols and grounding in large language models. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381(2251), p. 20220041, https://doi.org/10.1098/rsta.2022.0041

Pérez, S.; Guillén, G. (2007). Estudio exploratorio sobre creencias y concepciones de profesores de secundaria en relación con la geometría y su enseñanza. En P. Bolea; M. Camacho y P. Flores (Eds), *Actas del XI Simposio de la Sociedad Española de Investigación en Educación Matemática (SEIEM)* (pp. 295-305). Universidad de La Laguna. Tenerife.

Sureda, P., Corica, A. R., & Parra, V. (2023). Inteligencia Artificial Generativa en la formación de Profesores de Matemática en servicio. *Unión - Revista Iberoamericana de Educación Matemática*, 19(69). 1-15.

Ana Rosa Corica. Investigadora Independiente del CONICET (Argentina) y Profesora Adjunta de la Facultad de Ciencias Exactas de la UNCPBA (Tandil, Argentina). Es Doctora en Ciencias de la Educación (2010), Licenciada en Educación Matemática (2005) y Profesora en Matemática y Física (2002). Directora y fundadora del Núcleo de Investigación en Educación Matemática (NIEM) de la UNCPBA. Tiene más de 20 años de experiencia en el área de Educación Matemática y Física y en la formación universitaria inicial y continua de profesores de matemática. Sus publicaciones más recientes son: “Can generative AI solve Geometry problems? Strengths and weaknesses of LLMs for geometric reasoning in Spanish”; e “Inteligencia Artificial Generativa en la formación de Profesores de Matemática en servicio”.

Patricia Sureda. Investigadora Asistente del CONICET (Argentina) y Profesora Adjunta de la Facultad de Ciencias Exactas UNCPBA (Tandil, Argentina). Doctora en Enseñanza de las Ciencias. Mención Matemática (2012), Licenciada en Educación Matemática (2006) y

Profesora en Matemática (2005). Integrante y fundadora del Núcleo de Investigación en Educación Matemática (NIEM) de la UNCPBA. Tiene 20 años de experiencia en el área de Educación Matemática y en la formación universitaria inicial y continua de profesores de matemática. Publicaciones recientes: "Can generative AI solve Geometry problems? Strengths and weaknesses of LLMs for geometric reasoning in Spanish"; e "Inteligencia Artificial Generativa en la formación de Profesores de Matemática en servicio".

Verónica Parra. Investigadora Adjunta del CONICET (Argentina) y Profesora Adjunta de la Facultad de Ciencias Exactas UNCPBA (Argentina). Doctora en Enseñanza de las Ciencias. Mención Matemática (2013), Licenciada en Educación Matemática (2008) y Profesora en Matemática (2005). Integrante y fundadora del Núcleo de Investigación en Educación Matemática (NIEM) de la UNCPBA. Tiene 20 años de experiencia en el área de Educación Matemática y en la formación universitaria inicial y continua de profesores de matemática. Publicaciones recientes: "Can generative AI solve Geometry problems? Strengths and weaknesses of LLMs for geometric reasoning in Spanish"; e "Inteligencia Artificial Generativa en la formación de Profesores de Matemática en servicio".

Silvia Schiaffino. Investigadora Principal de CONICET (Argentina) y Profesora Asociada con Dedicación Exclusiva en la Facultad de Ciencias Exactas de la UNCPBA (Tandil, Argentina). Es Magíster en Ingeniería de Sistemas (2001) y Doctora en Ciencias de la Computación (2004). Posee más de 20 años de experiencia en el área de Inteligencia Artificial. Sus publicaciones más recientes son: "Can Generative AI Solve Geometry Problems? Strengths and Weaknesses of LLMs for Geometric Reasoning in Spanish", "An approach for explaining group recommendations based on negotiation information" y "A Lightweight Approach for Building User Mobility Profiles".

Daniela Godoy. Investigadora Principal de CONICET (Argentina) y Profesora Asociada con Dedicación Exclusiva en la Facultad de Ciencias Exactas de la UNCPBA (Tandil, Argentina). Es Magíster en Ingeniería de Sistemas (2001) y Doctora en Ciencias de la Computación (2005). Posee más de 20 años de experiencia en el área de Inteligencia Artificial. Sus publicaciones más recientes son: "Can Generative AI Solve Geometry Problems? Strengths and Weaknesses of LLMs for Geometric Reasoning in Spanish", "A Study on Influential Features for Predicting Best Answers in Community Question-Answering Forums", y "I Want to Break Free! Recommending Friends from Outside the Echo Chamber".