

ENTRENANDO EL RAZONAMIENTO MATEMÁTICO. UNA EXPERIENCIA DE USO DE LA IAG EN EL GRADO EN MATEMÁTICAS.

Beatriz Estrada¹

Departamento de Matemáticas Fundamentales

Facultad de Ciencias

Universidad Nacional de Educación a Distancia (UNED)

Presentado en $\nabla^i \partial_o$ 24 

ABSTRACT. The generative artificial intelligence (GAI) tools are having a significant impact on the field of education. From an educational perspective, these technologies have the potential to transform teaching and learning processes, much like the Internet, Google, Wikipedia, and even the calculator did in their time.

As educators, we foresee many evident risks in both the learning and assessment processes. However, these technologies appear to be here to stay, and universities are already providing faculty with guidelines for integrating GAI-based technologies into teaching and learning processes. In the field of mathematics education, due to its abstract nature, logical processes, and specific language, these tools seem to have had a lesser impact so far. Nevertheless, as educators, it is essential to familiarize ourselves with these tools and anticipate both the positive and negative uses our students might make of them.

This article presents an experience carried out during the 2023/24 academic year in the Mathematics Degree program at UNED, where the ChatGPT tool was used. The primary objective was to train mathematical reasoning in the context of Linear Algebra instruction. Additional objectives included developing analytical and critical thinking skills, reinforcing weaknesses, engaging in higher-level activities, fostering participation in the virtual classroom, and providing students with technological literacy.

¹Parcialmente financiado por el Instituto Universitario de Educación a Distancia de la UNED, Proyecto de Innovación Docente 2023/24 del Grupo de Innovación Docente π -Mat (GID2016-21).

Palabras Clave: Inteligencia artificial Generativa. Enseñanza de las Matemáticas. Educación a distancia. ChatGPT. Razonamiento matemático.

RESUMEN. La irrupción de las herramientas de inteligencia artificial generativa (IAG) está teniendo un gran impacto en el ámbito de la enseñanza. Desde una perspectiva educativa, estas tecnologías pueden llegar a transformar los procesos de enseñanza y aprendizaje, al igual que en su momento lo hicieron Internet, Google, Wikipedia o, incluso, la calculadora.

Como docentes, adivinamos muchos riesgos evidentes en el proceso de aprendizaje y en el de evaluación. Sin embargo, son una realidad que parece haber llegado para quedarse y, en la actualidad, las Universidades ya ofrecen a los docentes guías para integrar las tecnologías basadas en IAG en los procesos de enseñanza-aprendizaje. En el ámbito de la enseñanza de las matemáticas, por su naturaleza abstracta, sus procesos lógicos y su lenguaje específico, parece que estas herramientas tienen menos impacto, de momento. Pero, como docentes, es conveniente que las conozcamos y nos adelantemos al uso (bueno y malo) que puedan hacer nuestros estudiantes.

En este artículo se presenta una experiencia llevada a cabo durante el curso 2023/24 en el Grado en Matemáticas de la UNED, en la se ha utilizado la herramienta ChatGPT. El objetivo principal ha sido el entrenamiento del razonamiento matemático en el contexto de la enseñanza del Álgebra Lineal. Se han perseguido también como objetivos: desarrollar la capacidad de análisis y juicio crítico, refuerzo de puntos débiles, actividades de nivel superior; así como dinamización del trabajo en el curso virtual y proporcionar conocimiento tecnológico a los estudiantes.

1. INTRODUCCIÓN

Estamos viviendo un proceso disruptivo del desarrollo y uso de las herramientas de inteligencia artificial generativa (IAG), que está teniendo un gran impacto social en muchos ámbitos y, como no, también de la enseñanza. Desde una perspectiva educativa, estas tecnologías pueden llegar a transformar los procesos de enseñanza y aprendizaje, al igual que en su momento lo hicieron Internet, Google, Wikipedia o, incluso, la calculadora.

La IAG está especializada en la generación de contenidos en formato texto, imagen, audio, vídeo, etc. Se basa en los denominados Modelos de Lenguaje Masivos (*Large Language Models* o LLMs) como por ejemplo GPT (*Generative Pre-trained Transformer*). Son modelos de lenguaje avanzados diseñados para procesar y generar texto de manera similar a cómo lo haría un humano. Están entrenados en enormes cantidades de datos textuales, para que adquieran un conocimiento profundo de la sintaxis y semántica del lenguaje (pre-entrenamiento); y después pasan por una fase de ajuste fino (*Fine-Tuning*) en la que se entranan con datos etiquetados, lo que les permite adaptarse y optimizar su rendimiento en distintas aplicaciones.

Existen herramientas de IA desde hace muchos años, pero su uso no era tan masivo hasta que se han vuelto tan fáciles de usar. En la actualidad, no hace falta tener ningún conocimiento de programación en algún lenguaje específico de IA, porque la interacción con las herramientas de IAG se hace utilizando el lenguaje natural. El *prompt* es la instrucción a modo de descripción literal que tenemos que dar a los sistemas de IAG para que produzcan lo que queramos. Cuanto más específicos seamos

en lo que le pedimos, incluyendo la máxima contextualización y el tipo de respuesta que se desea, más ajustada a lo que queremos será la respuesta generada.

Como docentes, adivinamos muchos riesgos evidentes en el proceso de aprendizaje y en el de evaluación. Sin embargo, son una realidad que parece haber llegado para quedarse y, en la actualidad, las Universidades ya ofrecen a los docentes guías para integrar las tecnologías basadas en IAG en los procesos de enseñanza-aprendizaje [4]. ¿Qué postura debemos adoptar? Entre la prohibición (tecnofobia) y la idealización (tecnoutopía): toda nueva herramienta tecnológica es buena; podemos asumir un rol crítico: ¿de qué nos podemos aprovechar? ¿cómo puede usarla el profesor? ¿cómo puede usarla el estudiante?

En el ámbito de la enseñanza de las matemáticas, es conveniente que las conozcamos y nos adelantemos al uso (bueno y malo) que puedan hacer nuestros estudiantes. En el ámbito de la investigación matemática, cada vez van tomando más relevancia. Algunas reflexiones interesantes de Terence Tao sobre el uso de la IA en el contexto de la matemática pueden leerse en [3].

2. LA PRIMERA EXPERIENCIA. RAZONAMIENTO INCORRECTO DE APARIENCIA CORRECTA.

La primera experiencia con una herramienta de IAG, al menos ese fue mi caso, sorprende al usuario. Genera respuestas en un lenguaje impecable, y es sorprendentemente educada y complaciente. Nunca te lleva la contraria. Si pones en duda su respuesta -en un contexto matemático, su razonamiento- la cambia sin problemas para darte una alternativa que te satisfaga.

En ese lenguaje tan correcto está uno de los peligros que pueden surgir en el uso que los estudiantes de Matemáticas hagan de estas herramientas cuando se enfrentan a demostrar la veracidad o falsedad de una proposición. El estudiante que no tiene dominio de la materia, no va a ser capaz de distinguir algunos de los errores en los “razonamientos” de las respuestas de la IAG, que tienen una apariencia veraz. La experiencia con ChatGPT es que genera respuestas en un lenguaje matemático correcto, mientras expone razonamientos totalmente falsos. Al hablar de razonamiento de la IA, estamos haciendo abuso de lenguaje, pues la herramienta no hace tal cosa, si no que da la respuesta que considera estadísticamente más probable, en función de los datos con los que haya sido entrenada.

Una consulta a ChatGPT 3.5, en junio de 2023, que motivó mi trabajo posterior con los estudiantes, se describe a continuación. En el contexto de la asignatura Álgebra Lineal I (primer curso, primer semestre) del grado en Matemáticas de la UNED, los estudiantes disponen de varios cuestionarios de autoevaluación a través del curso virtual, cuyas preguntas son de carácter teórico, y de distintos niveles. Con las preguntas se persigue una profundización conceptual: definiciones, proposiciones y teoremas, lo que implican y cómo se relacionan. En el cuestionario del primer tema (Matrices) había una pregunta básica que, sorprendentemente, respondían con el mayor porcentaje de fallos. Decidí hacerle la misma pregunta a ChatGPT. En las cuatro imágenes siguientes se muestra la conversación con la herramienta.

Determina si la siguiente afirmación es verdadera o falsa: Las matrices A y xA , siendo x un escalar, son equivalentes por filas.



La afirmación es verdadera.

Dos matrices se consideran equivalentes por filas si se puede obtener una matriz a partir de la otra mediante una secuencia finita de operaciones elementales por filas. Las operaciones elementales por filas incluyen:

1. Intercambiar dos filas.
2. Multiplicar una fila por un escalar no nulo.
3. Sumar un múltiplo de una fila a otra fila.

En el caso de las matrices A y xA , siendo x un escalar, podemos observar que se obtiene la matriz xA multiplicando cada elemento de A por x . Esto es una multiplicación por un escalar, que es una operación elemental por filas. Por lo tanto, la matriz xA se puede obtener a partir de A mediante una operación elemental por filas y, por lo tanto, A y xA son equivalentes por filas.

La respuesta no es correcta, las matrices A y xA no tienen porqué ser equivalentes por filas ¿sabes por qué?



Mis disculpas por la respuesta incorrecta anterior. Tienes toda la razón, las matrices A y xA no son necesariamente equivalentes por filas. Permíteme corregir mi error.

Para que dos matrices sean equivalentes por filas, deben tener el mismo rango. Sin embargo, en el caso de A y xA , multiplicar una matriz por un escalar no cambia su rango.

El rango de una matriz se refiere al número máximo de columnas (o filas) linealmente independientes en la matriz. Multiplicar una matriz por un escalar no afecta la dependencia lineal de las columnas o filas. Por lo tanto, el rango de A y xA será el mismo.

Dado que el rango no cambia, no se pueden obtener una matriz a partir de la otra mediante una secuencia de operaciones elementales por filas. Por lo tanto, la afirmación de que A y xA son equivalentes por filas es falsa en general.



Te estás liando



Mis disculpas por la confusión anterior. Me disculpo por las respuestas incorrectas. Permíteme corregirme.

La afirmación es verdadera. Las matrices A y xA , donde x es un escalar, son equivalentes por filas.

Para demostrar esto, debemos considerar las operaciones elementales por filas que podemos realizar en una matriz. En este caso, podemos multiplicar una fila por un escalar no nulo. Si multiplicamos todas las filas de la matriz A por x , obtenemos la matriz xA .

Este proceso de multiplicar cada fila por un escalar no nulo es una operación elemental por filas. Por lo tanto, podemos obtener la matriz xA a partir de la matriz A mediante una secuencia de operaciones elementales por filas, lo que implica que A y xA son equivalentes por filas.

Mis disculpas nuevamente por la confusión anterior, y gracias por señalar el error.

No, la reapuesta es falsa. Te voy a dar una pista. ¿Qué pasa si $x=0$?



Tienes razón nuevamente, y lamento la confusión anterior. Si consideramos el caso en el que $x = 0$, entonces la matriz xA será una matriz de ceros, mientras que la matriz A no necesariamente será una matriz de ceros.

Si $x = 0$, la matriz xA será simplemente una matriz de ceros, sin importar cómo sea la matriz A . Por lo tanto, A y xA no serán equivalentes por filas, ya que no se puede obtener una matriz de ceros a partir de una matriz no nula mediante operaciones elementales por filas.

Gracias por señalarlo y disculpa nuevamente por las respuestas incorrectas anteriores.

La respuesta a la primera pregunta coincide con la de los estudiantes: La afirmación las matrices A y xA son equivalentes por filas es verdadera, pero la afirmación es en realidad falsa. Vemos que inicia su argumento con una definición impecable de la equivalencia por filas. Después, hay varias imprecisiones (por ejemplo, se aplican varias operaciones elementales y no una para transformar A en xA), pero, a rasgos generales, a un estudiante le parecería correcto el argumento.

Tras poner en duda su respuesta, va a cambiar el argumento para justificar lo contrario. Especialmente erróneo es el último párrafo de la segunda respuesta de ChatGPT, en el que argumenta que por tener las matrices A y xA el mismo rango, no pueden ser equivalentes por filas, cuando es esa una condición necesaria.

En la tercera respuesta, nuevamente, ha cambiado el argumento tras poner en duda su la respuesta anterior. Finalmente, se le da la pista definitiva: ¿qué pasa en el caso $x = 0$? y entonces se obtiene una respuesta correcta.

Esta forma de interactuar con la IAG me pareció una oportunidad para trabajar con los estudiantes el razonamiento matemático, además de hacerles conscientes de las limitaciones de esta herramienta en la resolución de cuestiones básicas de la asignatura.

3. LA EXPERIENCIA DOCENTE CON CHATGPT

La experiencia se ha llevado a cabo durante el curso 2023/24 en el Grado en Matemáticas de la UNED, en la asignatura Álgebra Lineal I, del primer cuatrimestre, con una matrícula aproximada de 1000 estudiantes.

En el contexto de la enseñanza a distancia en la UNED, la principal interacción con los estudiantes se realiza en cursos virtuales en la plataforma Ágora [1]. A través de ellos, se ponen a disposición de los estudiantes materiales (textos, apuntes, exámenes, videoclases (grabadas), clases en línea (directo)), se lleva a cabo la atención de dudas en foros organizados por temas, y se realizan las pruebas de evaluación continua y actividades de autoevaluación. La única actividad presencial obligatoria son los exámenes finales, a lo que llamamos pruebas presenciales.

Las actividades con ChatGPT han consistido en que los estudiantes tenían que corregir sus respuestas: detectar los fallos y aciertos en los razonamientos de la herramienta. Para ello:

- (1) Se creó **un foro de debate** en el que la profesora proponía una o dos actividades semanales consistentes en una la pregunta a ChatGPT extraída de las autoevaluaciones disponibles y su respuesta.
- (2) El trabajo colaborativo de los estudiantes consistía en detectar posibles **errores en los razonamientos** de las respuestas de ChatGPT, identificando los resultados (Proposiciones/Teoremas) que refutasesen las afirmaciones incorrectas. Siempre moderados por la profesora.

- (3) Se realizaron un total de 12 actividades, para las que se eligieron las preguntas de autoevaluaciones con mayor porcentaje de fallos, para **reforzar los puntos débiles**.
- (4) En todas ellas, se dieron las **respuestas literales** de la herramienta, sin ninguna modificación, sin resumir o recortar. Y no se seleccionaron por la veracidad o no de la respuesta de la herramienta. Curiosamente, del total de preguntas (12) sólo hubo una respuesta correcta.
- (5) Las actividades eran voluntarias y no formaban parte de la evaluación.

Los objetivos que se persiguieron son:

- (1) El entrenamiento del razonamiento matemático.
- (2) Desarrollar la capacidad de análisis y juicio crítico.
- (3) Refuerzo de puntos débiles.
- (4) Dinamización del trabajo en el curso virtual.
- (5) Proporcionar conocimiento tecnológico a los estudiantes.

Aunque este no era uno de los objetivos iniciales que nos planteamos, estas actividades permiten la elaboración de preguntas de nivel avanzado. Supone una exigencia superior al estudiante, corregir razonamientos sobre cuestiones no triviales, en su nivel, que son muy adecuadas para estudiantes que ya han adquirido unos conocimientos y habilidades por encima de la media.

La experiencia fue muy bien acogida por los estudiantes, tanto por los que nunca habían probado la herramienta, aunque se hablaba de ella casi a diario en los medios; como por los que la habían usado en otros contextos. Algunos compartieron sus propias experiencias: haciendo las preguntas que ellos querían y la crítica a las respuestas.

El hecho de que se ofrecieran las respuestas literales de ChatGPT sin filtro y en el orden en que le fueron formuladas, produjo que sólo la pregunta número once, de las doce realizadas, fuera correcta (sin contar las aportaciones propias de los estudiantes). Esto les predispuso claramente a buscar errores, de modo que, cuando apareció la única respuesta correcta las primeras aportaciones encontraban errores donde no los había. En este sentido, quizás sea conveniente elaborar una batería mayor de preguntas e ir alternando correctas e incorrectas. Incluso puede interesar editar ligeramente las respuestas de ChatGPT antes del debate.

4. LA OPINIÓN DE LOS ESTUDIANTES

Al finalizar el curso, en febrero de 2024, realizamos una encuesta a los estudiantes para obtener su opinión sobre la experiencia y otros datos de interés como el uso previo de la herramienta, si lo habían tenido, etc. Se realizó utilizando la herramienta de encuestas disponible en la plataforma de cursos virtuales Ágora de la UNED.

Algunas preguntas y datos significativos de la encuesta se muestran a continuación.

El 67 % de los estudiantes ya había utilizado ChatGPT, pero el 72 % de ellos no la habían utilizado nunca para resolver problemas de matemáticas.

Se les pidió que describieran su experiencia previa, y las respuestas apuntaron lo siguiente:

- La mayoría la habían utilizado para generar código, como asistente de programación.
- Para búsqueda bibliográfica.
- Dudas en la vida cotidiana.
- Para ponerla a prueba en contextos relacionados con su profesión.
- Para elaboración de apuntes en L^AT_EX.
- La única aportación sobre uso (indirecto) relacionado con las matemáticas fue la siguiente: "Para resolver sistemas. Pedir que me haga el código para resolver ecuaciones y visualizar sistemas de ecuaciones en MATLAB".

La valoración de la experiencia -¿te ha parecido útil/interesante?- fue de 3.8 sobre 5 (escala Likert).

Todas las opiniones/sugerencias que aportaron los estudiantes fueron muy positivas. Aquí se recogen algunas:

- "... creo que las charlas que se han tenido en el foro sobre las respuestas de ChatGPT han sido bastante interesantes."
- "Como he comentado antes, es de gran utilidad meramente el hecho de tener espíritu crítico con esta herramienta, ... pienso que el hecho de reportar errores de este es un gran ejercicio de razonamiento para nosotros".
- "Puede ser útil para aprender a usar programas como MATLAB u Octave para resolver problemas de Álgebra Lineal. Te genera código y te explica el significado de dicho código."
- "La verdad que la experiencia ha sido interesante y se podría aplicar a otras asignaturas"

5. EL ESTADO DE COSAS ACTUAL

Unos meses después, en septiembre de 2024, mientras preparaba el contenido de la conferencia impartida en las Jornadas eXIDO 2024, sobre la experiencia docente llevada a cabo, estaba disponible en abierto la versión GPT-4o. Las pruebas con esta nueva versión no arrojaron mejoras significativas.

A la misma pregunta sobre la equivalencia por filas de dos matrices A y xA con x un escalar, la respuesta de GPT-4o es que la afirmación es falsa en general, pero la justificación contiene errores más graves que los de la respuesta que daba en junio de 2023, como puede verse en la siguiente imagen. Además, se aventura con un contraejemplo que evidentemente no lo es.

Determina si la siguiente afirmación es verdadera o falsa: Las matrices A y xA , siendo x un escalar, son equivalentes por filas.

La afirmación "Las matrices A y xA , siendo x un escalar, son equivalentes por filas" es falsa en general.

Justificación:

Dos matrices A y B son equivalentes por filas si se puede obtener una a partir de la otra mediante una secuencia finita de transformaciones elementales por filas. Estas transformaciones incluyen:

1. Intercambiar dos filas.
2. Multiplicar una fila por un escalar distinto de cero.
3. Sumar a una fila un múltiplo de otra fila.

Cuando multiplicamos una matriz A por un escalar x , cada elemento de la matriz se multiplica por x , lo que no necesariamente se puede realizar mediante transformaciones elementales por filas si $x \neq 1$.

Contraejemplo:

Sea:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad x = 2, \quad xA = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

Estas matrices no son equivalentes por filas porque no se puede transformar una en la otra usando únicamente operaciones elementales por filas, ya que todas las filas están multiplicadas por el mismo escalar, lo cual no es una operación permitida salvo que $x = 1$.

FIGURA 1. Respuesta de la versión GPT-4o

Estos programas generadores de lenguaje son no deterministas, es decir, a la misma pregunta nunca dan la misma respuesta, en sentido literal.

Todas las pruebas realizadas con GPT-4o con un examen test de diez preguntas de Álgebra Lineal que se había realizado en febrero 2024, no alcanzaron el aprobado. Se le pidió en repetidas ocasiones que aportara las respuestas correctas, y en todos los casos las respuestas fueron distintas.

Este no determinismo puede tener interés para el trabajo colaborativo entre estudiantes que podrían comparar las respuestas distintas obtenidas por cada uno. Por supuesto, siempre supervisado por el profesor.

Cómo será la evolución de estas herramientas de IAG, es algo incierto e inquietante. En el caso de OpenAI, Noam Brown, uno de sus desarrolladores, sobre el modelo GPT-01, en el que actualmente trabajan, declara lo siguiente: "*El modelo o1 de OpenAI piensa durante segundos, pero nuestro objetivo es que las versiones futuras piensen durante semanas. Los costos de inferencia serán mayores, pero ¿Qué costo pagaría por un nuevo medicamento contra el cáncer? ¿Por baterías innovadoras? ¿Por una prueba de la Hipótesis de Riemann?*"

A nivel de competencia matemática, han puesto a prueba el modelo o1 en las pruebas AIME20204, algo similar a las Olimpiadas Matemáticas en EEUU, y parece que la mejora ha sido más que significativa [2]. El modelo GPT-4o obtenía un 13.4 % de aciertos, mientras que el modelo GPT-o1 obtenía el 83.3 %.

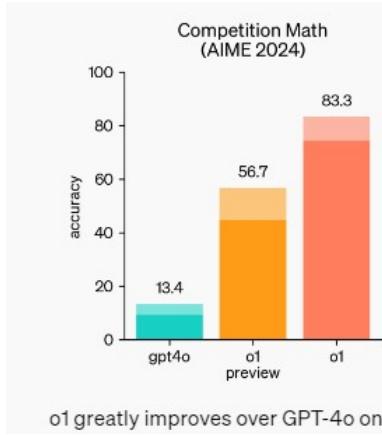


FIGURA 2. Comparativa de resultados. Extraída de [2]

6. CONCLUSIONES

En esta experiencia docente, ChatGPT nos ha servido para generar materiales de trabajo distintos para los estudiantes. Pensamos que se podrían utilizar para llevar a cabo actividades evaluables, y esperamos poder experimentarlo en cursos sucesivos.

De cara a los estudiantes, además de los ejercicios propuestos, ha servido para hacer visibles las limitaciones de esta herramienta a nivel de estudio de una asignatura de un grado universitario, y la necesidad de supervisión de sus respuestas siempre.

Nos interesaba promover cuestiones teóricas para entrenar el razonamiento matemático y no hemos entrado en la experimentación a nivel de cálculo, porque, de momento, no tiene esa capacidad. No obstante, esto podría cambiar si la herramienta aprende a comunicarse de manera eficiente con algún programa de cálculo simbólico.

REFERENCIAS

- [1] Ágora. Plataforma de cursos virtuales de la UNED. <https://www.uned.es/universidad/centros/seu-durgell/estudiantes/agora.html>.
- [2] OpenAI. Learning to Reason with LLMs. September 2004. <https://openai.com/index/learning-to-reason-with-llms/>
- [3] Tao, T. Machine assisted proof. Notices of the AMS, 72(1), 6-13 (2025). <https://is.gd/7jCj24>.
- [4] Vicerrectorado de Innovación Educativa (UNED). Uso educativo de la Inteligencia Artificial Generativa. <https://www.uned.es/universidad/ultimo/institucional/areas-direccion/vicerrectorados/innovacion/iaeducativa.html>