

¿Se puede evaluar con IAG?

Mónica Patricia Gruden
grudenmonica@gmail.com
Universidad Nacional de Moreno, Argentina

Resumen

En este trabajo comento mi experiencia docente en una materia de primer año en la carrera de Ciencias Económicas y Jurídicas de la Universidad Nacional de Moreno. Me cuestiono si es posible evaluar con Inteligencia Artificial Generativa (en adelante IAG) y para analizar esta posibilidad planteo tres experiencias. Una evaluación donde le pido a Chat GPT que ponga en conversación a autores de distintas corrientes de administración para que los alumnos los identifiquen, otra experiencia en donde tengo que iterar prompts para lograr una evaluación de múltiples opciones con Perplexity, que no solo tenga el formato deseado, sino que, además, identifique los temas importantes con los significados esperados y, por último, la dinámica de evaluar la lectura a través de una IA que personifica a personajes de la historia. En este último caso no es IAG, pero la experiencia de interacción de los alumnos fue enriquecedora y amerita el planteo de analizar si es producto de la humanización de la IA.

Palabras claves: Inteligencia artificial generativa, Educación superior, Evaluación con IAG, Estrategias didácticas en administración con IAG.

Líneas temáticas: Experiencia con IA en aula en la educación superior.

Introducción

Existe una tensión entre las oportunidades y desafíos que tiene la IAG en donde nos tenemos que ocupar de detectar el plagio, superar los usos banales, desafío de uso profundo y reflexivo, co-creación de contenido y co-evaluación entre docentes para diseñar las propias asignaturas. (Andreoli, Aubert, Cherbavaz, & Perillo, 2024)

En este artículo, me interesa profundizar la co-evaluación entre docentes para diseñar estrategias didácticas con IAG en una asignatura de primer año llamada IOPA (Instituciones, Organizaciones y Principios de Administración) de cursada obligatoria para las carreras de Ciencias Económicas y Jurídicas de la Universidad Nacional de Moreno. La co-evaluación significa en mi caso compartir y consensuar las herramientas que seleccionamos para la evaluación con nuestro par pedagógico.

Los alumnos deben asistir a clase con la bibliografía leída para poder analizar los temas y realizar actividades con los contenidos. Paso a relatar las experiencias en función de los problemas con los que me encontré. En el primer caso, mi objetivo es la evaluación del aprendizaje y para ello, le ofrezco a los alumnos una actividad donde deben reconocer autores de las corrientes de pensamiento en Administración que fue elaborada con chat GPT y que la mayoría no pudo responder. Segundo caso, la experiencia de encontrar el mejor prompt que pueda hacer que Perplexity realice un multiple choice de un PDF preguntando lo relevante y que esté vinculado a la importancia que se le dio al tema en clase. En la tercera experiencia, comento una evaluación del aprendizaje con IA que no es generativa, pero me pareció pertinente compartir porque los alumnos

luego de la lectura de un autor tenían que hacerle preguntas y se generó una dinámica interesante a causa de la humanización del autor.

Diseño pedagógico

Los alumnos deben asistir a clase con la bibliografía leída para poder analizar los temas y realizar actividades con los contenidos por esto he realizado distintos tipos de evaluación.

Es importante considerar que la evaluación puede ser: diagnóstica inicial, evaluación formativa y evaluación sumativa. (Anjovich & Cappelletti, 2017) Pero también hay que considerar las cuatro dimensiones de la evaluación (Barbera, 2006), la evaluación del aprendizaje que es al final (donde se ven los resultados), la evaluación para el aprendizaje donde se analiza el proceso de enseñanza-aprendizaje, la evaluación como aprendizaje para que el alumno pueda conocer sus fortalezas y debilidades y la evaluación desde el aprendizaje donde el docente busca mejorar. En las experiencias que realicé me enfoqué más en la evaluación para el aprendizaje.

Cuando seleccionamos el medio para evaluar, consideramos que existen riesgos porque dudamos de los ensayos, trabajos de investigación, presentaciones de proyectos, disertaciones, ya que puede haber plagio o sesgo según la IA elegida. Algunos autores sugieren confiar en portafolios, reportes de pares, observación directa, registros de desempeño o exámenes orales (Sánchez Mendiola, 2023). Pero cuando seleccionamos por ejemplo un multiple choice o un verdadero o falso y el docente es el que consulta a una IAG también tenemos que tomar recaudos. A continuación, describo mi experiencia en tres situaciones de evaluación distintas:

1. El lenguaje en que se expresa la consigna o la forma de preguntar, confunde al alumno (evaluación desarrollada en chat GPT).

En el año 2023 decidí realizar una evaluación a los alumnos luego de terminar con la unidad referida a varios autores de Administración como Taylor, Fayol, Mayo, Weber, etc. Era necesario saber si antes de la evaluación sumativa habían logrado ver las diferencias de criterio, los distintos análisis y entornos que rodeaban a estos autores. Esta evaluación no era calificada, sino que orientaba a los alumnos para que vean si estaban preparados para la evaluación definitiva. Le pedí al chat GPT que realicé una discusión entre Taylor, Fayol y Weber. Utilicé las afirmaciones que me parecieron apropiadas, sin poner el nombre del autor reemplazándolos con una letra, para que los alumnos lo descubran. Cuando fui a clase y realicé esta actividad pensé que era fácil y que podían identificar a los autores y no fue así. Me sirvió para ver el grado de comprensión de lo visto e identificar que los alumnos pueden repetir la teoría, identificar lo más importante de cada autor, pero hacer esta evaluación implicaba relacionarlos y analizar desde un lenguaje que no era similar a la lectura que habían hecho. Fue útil para desarrollar más su pensamiento crítico y para darme cuenta que no servía usar la IAG directamente en una evaluación si previamente no habían experimentado con preguntas, opciones o verdadero o falsos, hechos con IAG.

2. Perfeccionar el prompt para que el resultado sea realmente lo que el docente necesita evaluar.

Al comenzar a elaborar un multiple choice se intenta que el prompt tenga todos los elementos necesarios: rol, tarea, contexto, formato, ejemplos o información complementaria. Se transfiere a la IAG lo que es importante del tema y pasa a ser este desafío una extensión de lo que ya se discute en las clases. ¿Quedó claro para el alumno lo más importante de cada tema? En una pareja pedagógica ¿hay coincidencia en los criterios sobre lo más importante?

Escribir e intentar prompts (Viñas, 2023) para obtener por ejemplo un multiple choice con Perplexity.

Primer prompt: Elabora un cuestionario como si fueras docente de IOPA en la Universidad Nacional de Moreno en primer año. Deber ser referido a los temas del PDF que adjunto. Diez preguntas basadas en ejemplos de la realidad enfocándose en los temas de las organizaciones, instituciones y complejidad. Tienen que con contar con tres opciones de respuesta con dos distractores y una sola respuesta correcta. También elabora una rúbrica para calificar con las respuestas correctas.

Se planteó el rol, pero no fue claro el producto, se pudo contextualizar, pero Perplexity no podía leer el PDF para hacer las preguntas. Hizo una rúbrica, pero no estaban cuáles eran las respuestas correctas. Primer intento fallido.

Segundo prompt: Elabora un cuestionario con el PDF de Góngora y Cikatelli con 10 reactivos con 3 opciones de respuesta y 2 distractores. Una sola respuesta correcta. Con una rúbrica para calificar al alumno.

En este caso no se contextualizó, ni se le dio el rol, pero la tarea estaba bien especificada y se le brindó a la IAG que seguía siendo Perplexity un PDF que pudo leer y se agregó la posibilidad de rúbrica y retroalimentación. No estaba la opción correcta. Las preguntas no eran sobre los temas prioritarios que se habían profundizado en clase. Otro intento fallido.

Tercer prompt: se agrega que la respuesta correcta deberá aparecer después de cada reactivo.

En este intento se logra tener la respuesta correcta para poder corregir más rápido, pero sigue existiendo un problema y es que las preguntas no apuntan a los conceptos más importantes señalados en clase por los profesores. Las preguntas tienen un criterio de importancia en función de los títulos propuestos en el libro.

Cuarto prompt: Como docente, elabora un cuestionario con el PDF de Góngora y Cikatelli que se adjunta. Los temas más importantes tienen que ver con las organizaciones, las instituciones y la complejidad. Debe tener 15 reactivos con 3 opciones de respuesta y 2 distractores. Una sola respuesta correcta. La respuesta correcta debe aparecer después de cada reactivo. También debes realizar como docente, una rúbrica para poder calificar a los alumnos.

En este caso se cumple con el rol, la tarea, el formato y se le pide que las preguntas sean sobre los temas importantes señalados. Pero tiene el error de usar el concepto de institución mal en varias preguntas del cuestionario.

Quinto prompt: se agrega que el concepto de instituciones responda a la nueva corriente neo-institucional pero, se nota en una de las preguntas, que entendió el significado pero en otra comete un error de interpretación.

Sexto prompt: Como un docente, elabora un cuestionario con el PDF de Góngora y Cikatelli que se adjunta. Los temas más importantes tienen que ver con las organizaciones, la complejidad y las instituciones consideradas como valores según Svarstein. Debe tener 15 reactivos con 3 opciones de respuesta y 2 distractores. Una sola respuesta correcta. La respuesta correcta debe aparecer después de cada reactivo. También debes realizar como docente una rúbrica para poder calificar a los alumnos.

Por suerte en el sexto intento logré que hiciera preguntas de opción considerando los temas importantes con la interpretación de los mismos de acuerdo al programa.

3. No hay internet en clase. ¿cuál es la dinámica para usar IA en una evaluación para el aprendizaje?:

Este año, 2024, hice que los alumnos, en grupos y luego de la lectura de Weber, vayan a una página española: hablaconlahistoria.es¹ y les pedí que le hagan 5 preguntas a Weber. Organicé grupos para que intercambiaran el conocimiento sobre la lectura realizada y que fueran a trabajar a la biblioteca para que no usen sus datos. Volvieron al aula y cada grupo debía exponer de a una las preguntas y sus respectivas respuestas. Por las preguntas que hacían me daba cuenta si habían leído o no y también sus inquietudes. Fue una clase donde en algún momento el aula quedó vacía, pero todos volvieron para contar lo que les había pasado y estaban ansiosos por compartirlo.

Fue muy interesante porque me encontré que, a pesar de estar interactuando con una inteligencia artificial, la clase la había humanizado. Entre muchas de las preguntas, le consultaron sobre su parecer acerca del trabajo de la mujer como líder, qué opinaba de Hitler y a pesar de que no lo conoció, la respuesta fue fabulosa ya que unió el liderazgo carismático con los conceptos de burocracia. Un alumno le preguntó como habían sido sus padres respecto de su formación y nos encontramos con una respuesta que tenía que ver con su biografía, pero como lo relataba la inteligencia artificial resultaba interesante a los efectos de la pregunta hecha. En conclusión, para esta actividad donde se evaluaba el grado de lectura de los alumnos y su interpretación, la interacción con la IA provocó un ambiente ameno de compartir la experiencia, en la interacción y las respuestas de la IA que los motivó mucho.

La clave para revivir a un personaje histórico en el mundo digital está en la recopilación de discursos, escritos, cartas, libros, etc. Estos datos forman un “ADN lingüístico” del personaje para que el modelo aprenda y emule su estilo y personalidad. Deben plantearse las implicancias éticas, ya que puede haber malentendidos o divulgaciones incorrectas de su legado. También se podrían plantear interrogantes sobre la privacidad y el consentimiento. (Vivas, 2023) Considero muy importante la mediación docente para poder resolver los dilemas éticos que se pueden suscitar y recordar constantemente que es una IA.

Conclusiones

¿Se puede evaluar con inteligencia artificial generativa? Por supuesto que se puede, pero como advertimos, se debe tener en cuenta la experimentación con la herramienta, probando prompts e iterar con paciencia y compartiendo con los pares pedagógicos. Si creemos que es algo rápido y nos facilita el trabajo, creo que es una percepción equivocada porque una evaluación pensada en el desarrollo del espíritu crítico del alumno, lleva tiempo. Las preguntas surgen en la IAG, si se hace un multiple choice, de la importancia que el autor puso en los apartados de su escrito, o sea, la inteligencia artificial va a tomar como importante lo que el autor considera como importante al separar en títulos o capítulos. Si nosotros tomamos algunos puntos del autor para comparar con el resto de la bibliografía de la materia y queremos evaluar esta comparación tenemos que aportar a la inteligencia artificial esta referencia, autor o concepto importante en el prompt. Si en la clase le damos importancia a un tema debemos incluir en el prompt esta consideración. Por lo que nos demanda tiempo y análisis. También se debe tener en cuenta, la forma de expresar las preguntas y que debemos en clase, hacer mini evaluaciones del mismo tipo que va a tener la evaluación sumativa, para que haya una coordinación entre la clase y la evaluación. Y por supuesto, siempre mediar en la actividad con IAG que se desarrolla con los alumnos, para evitar caer en dilemas éticos.

Bibliografía

Andreoli, S., Aubert, E., Cherbavaz, M., & Perillo, L. (2024). Entre humanos y algoritmos, percepciones docentes sobre la exploración con IAG en la Enseñanza del Nivel Superior. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 37, 63-77.
doi:10.24215/18509959.37.e6

Anjovich, R., & Cappelletti, G. (2017). La evaluación como oportunidad. Buenos Aires: Paidós, Voces de la educación.

Barbera, E. (2006). Aportaciones de la Tecnología a la E-Evaluación. *Revista de educación a distancia RED*. Obtenido de <https://revistas.um.es/red/article/view/24301>

Sánchez Mendiola, M. (2023). La inteligencia artificial generativa y la evaluación: ¿Qué pasará con los exámenes? *Investigación en Educación Médica*, 5-8.
doi:10.22201/fm.20075057e.2023.48.23550

Viñas, M. (2023). Chat GPT: 16 Instrucciones potentes "prompts" que aceleran tu forma de aprender o enseñar. Obtenido de Academy Totem Guard:
<https://www.totemguard.com/aulatotem/wp-content/uploads/2023/04/Chat-gpt-prompts.pdf>

Vivas, F. (2023). Invisible. La inteligencia artificial en nuestra vida. Ciudad Autónoma de Buenos Aires, Argentina: Sudamericana.