



**VNiVERSiDAD
D SALAMANCA**

MEMORIA DEL PROYECTO DE INNOVACION DOCENTE (ID2023/251)

**“INCORPORACIÓN DE CHATGPT PARA LA EVALUACIÓN PRÁCTICA DE LA
DOCENCIA DE BIOQUÍMICA DEL GRADO DE MEDICINA”**

Profesora coordinadora: Verónica González Núñez

Categoría: Profesora Titular de Universidad

Departamento: Bioquímica y Biología Molecular

e-mail: vgnunez@usal.es

ext. 5337

Salamanca, 29 de Mayo de 2024

Desde el Curso Académico 202223 se imparte la asignatura de “*Bioquímica*” en el primer semestre del primer Curso en el Grado en Medicina del Plan Aprobado en 2022, si bien esencialmente sigue siendo la misma asignatura que se viene impartiendo en el Grado en Medicina desde el curso 2011-11. Esta asignatura es de carácter básico, con una carga docente de 6 créditos ECTS y con un grado de presencialidad del 50%. Proporciona un conocimiento básico a nivel molecular sobre los mecanismos responsables del desarrollo de la vida y los procesos patológicos. Las prácticas de laboratorio y prácticas en aula de informática suponen 20 h por alumno, y los seminarios 6 h por alumno, con una presencialidad del 100%. Tal y como se puede observar, esta asignatura tiene una elevada carga práctica, tal y como suele ser habitual en grados de alto coeficiente de experimentalidad.

Por otra parte, la integración de la Inteligencia Artificial (IA) en la educación superior es un tema de gran relevancia en el ámbito de la innovación docente. En la actualidad, los profesores nos enfrentamos al desafío de adaptar nuestros métodos de enseñanza para satisfacer las necesidades cambiantes de los estudiantes digitales. La IA ofrece una variedad de herramientas innovadoras que pueden revolucionar la forma en que impartimos nuestra docencia e interactuamos con los estudiantes. Desde sistemas de tutoría virtual hasta plataformas de aprendizaje adaptativo, la IA proporciona nuevas oportunidades para personalizar la experiencia educativa y promover un aprendizaje más activo y participativo. Además, al analizar grandes conjuntos de datos, la IA puede ayudar a los educadores a identificar patrones de aprendizaje, mejorar la retroalimentación y optimizar los planes de estudio. En este contexto, la exploración de cómo la IA puede mejorar la práctica docente se convierte en un área de investigación crucial en la búsqueda de estrategias pedagógicas efectivas y centradas en el estudiante en el EEES.

Por todo ello, en este proyecto de innovación docente nos propusimos explorar el potencial de ChatGPT para innovar en la docencia de nuestra asignatura.

PROFESORADO QUE FORMA ESTE GRUPO DE INNOVACION DOCENTE

- Verónica González Núñez, Profesora Titular de Universidad, Departamento de Bioquímica y Biología Molecular.
- María Elena Díaz Rodríguez. Profesora Asociada, Departamento de Bioquímica y Biología Molecular.

OBJETIVOS DEL PROYECTO Y RESULTADOS OBTENIDOS

A continuación se van a referir objetivos propuestos inicialmente en la solicitud y se detallará cómo se han abordado.

1.- Preparar preguntas tipo test relativas a las prácticas impartidas en la asignatura de Bioquímica del Grado en Medicina.

Cada práctica de la asignatura de Bioquímica tiene unos objetivos concretos y los alumnos deben adquirir unas competencias concretas. Basada en la experiencia del equipo docente en los años previos, se han generado una serie de preguntas tipo test para analizar los conocimientos en cada una de las prácticas adquiridos por los alumnos. Dichas preguntas han sido similares a las que hemos planteado en años precedentes al alumnado. Para cada una de las prácticas se han preparado preguntas de tipo test, para las que se darán 5 respuestas posibles siendo sólo una correcta.

2.- Usando la aplicación ChatGPT, generar una segunda serie de preguntas relativas a las mismas prácticas.

En paralelo, usando la aplicación *ChatGPT* se ha generado otro grupo de preguntas sobre las mismas prácticas. Para ello nos hemos abierto una cuenta en la aplicación on-line "*Open AI*" y dentro del producto "*ChatGPT*". Desde la cuenta abierta, hemos guiado a esta aplicación a través del chat para generar preguntas tipo test, indicándole de manera concreta la asignatura y la práctica para la que necesitamos que nos formule las preguntas, además de las características de las preguntas que debe plantear (preguntas tipo test con 5 opciones y únicamente una respuesta correcta).

3.- En las pruebas de evaluación continua y en el examen final, plantear a los alumnos preguntas generadas con ambas metodologías.

Una vez generados de forma independiente estos dos grupos de preguntas, han sido planteadas a los alumnos en la evaluación continua y en el examen final. Así, se han realizado una prueba de evaluación continua a mitad del cuatrimestre y un examen final. Estas pruebas han constado de una serie de preguntas de tipo test sobre las prácticas impartidas; se han incluido dos preguntas por cada práctica realizada: una ha sido generada por el equipo docente y otra por ChatGPT sin que los alumnos conozcan el origen de las mismas. De la misma manera se ha obrado en el examen final.

4.- Analizar si las preguntas elaboradas con ayuda de AI son una herramienta fiable para la evaluación objetiva del aprovechamiento y de adquisición de competencias por los estudiantes.

Una vez que los alumnos han respondido tanto a las preguntas generadas por el equipo docente como por ChatGPT, se han comparado los resultados obtenidos para cada grupo de preguntas. Para ello se ha analizado el grado de dificultad para cada una de ellas, mediante el número y porcentaje de respuestas que los alumnos hayan contestado correctamente. Posteriormente, se han analizado los resultados por grupos: preguntas generadas por el equipo docente frente a ChatGPT. El número total de alumnos matriculados en la asignatura es de unos 240 por lo que se han podido realizar estudios estadísticos para ver qué grupo de preguntas ha sido contestado correctamente en mayor proporción. De esta manera, hemos podido ponderar la capacidad de integrar nuevas tecnologías de AI en general, en concreto ChatGPT, en la práctica docente y en su evaluación.

RESULTADOS OBTENIDOS Y VALORACION

La incorporación a la vida en general y la docencia en particular de las TICs y de la inteligencia artificial (IA) hace que tengamos que replantearnos su uso en dichos ámbitos. Por ello, en este proyecto de innovación docente, nos planteamos utilizarla como herramienta en la docencia y más concretamente evaluar la docencia práctica y de seminarios de la asignatura Bioquímica mediante preguntas generadas con la aplicación de IA ChatGPT y por el equipo docente, con el fin de comparar la capacidad de ambas aproximaciones para determinar el grado de aprovechamiento de los estudiantes y de adquisición de competencias.

Con este fin y siguiendo el plan de trabajo inicialmente propuesto, el grupo de profesores del equipo preparamos preguntas tipo test relativas a las prácticas y seminarios impartidos en la asignatura "Bioquímica" del grado en Medicina y basadas en la experiencia del equipo docente en los años previos. Para cada una de las prácticas/seminarios se planteó una pregunta tipo test para las que se daban 5 respuestas posibles siendo sólo una de las 5 opciones correcta.

En paralelo, y utilizando la aplicación ChatGPT se generó un segundo grupo de preguntas relativas a las mismas prácticas/seminarios. Para ello abrimos una cuenta en la aplicación online "Open-AI" y utilizamos la herramienta "ChatGPT" a la que pedimos que plantearse

preguntas con características similares a las planteadas por el equipo docente para cada una de las practicas/seminarios.

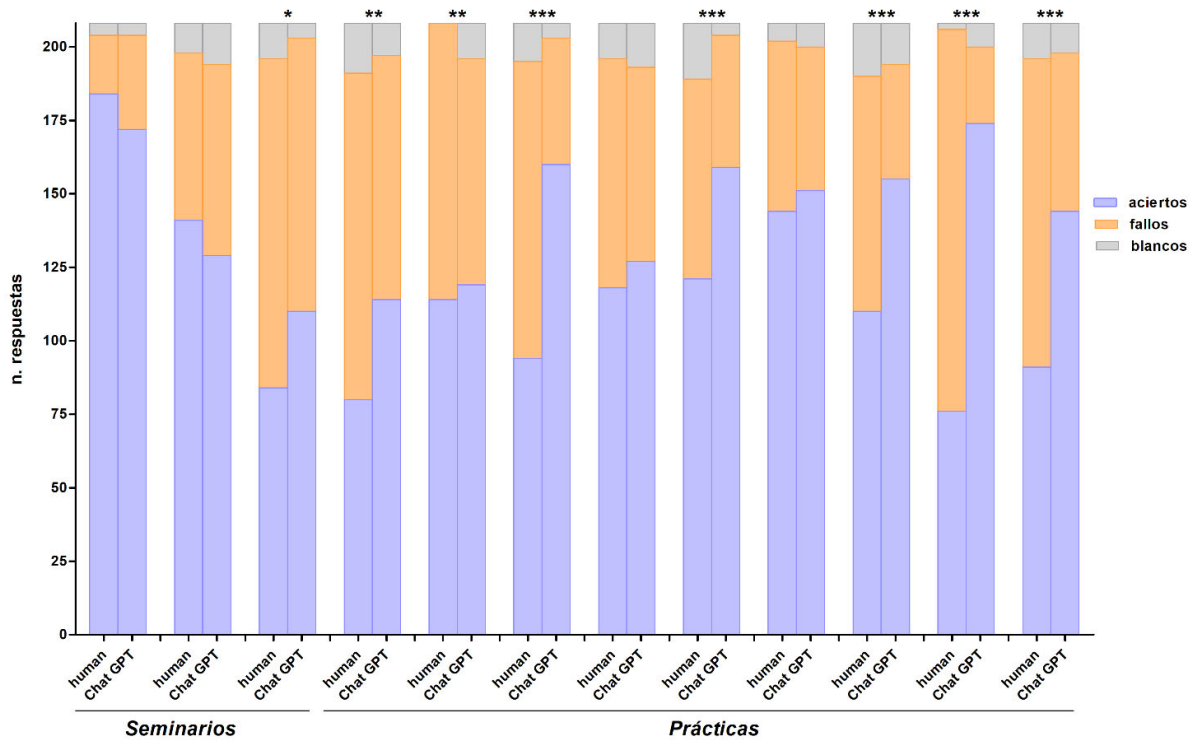
Inicialmente propusimos seguir esta estrategia para preparar dos preguntas de cada práctica o semanario (una hecha por el equipo docente y otra por *AI*) y, sin que los alumnos lo supiesen, plantearlas en la prueba de evaluación continua que se realiza a mitad del trimestre y en el examen final de la asignatura para posteriormente analizar el grado de respuesta de los alumnos a ambos cuestionarios y su corrección para evaluar la posibilidad de usar esta herramienta con fines docentes.

Cuando generamos las distintas preguntas a través de la inteligencia artificial nos encontramos con diversos problemas. En primer lugar, había que dirigir de manera muy precisa a este programa para que plantease la preguntas, y sabiendo muy bien cómo orientarlo. Aun así, en la mayor parte de las preguntas que planteó, encontramos problemas con las respuestas generada o cuales consideraba que eran correctas. De hecho, a la hora de trasladar las preguntas a los alumnos, en más de la mitad de los casos tuvimos que corregirlas ya que, no estaban bien formuladas o las respuestas que proponía la aplicación eran claramente erróneas. Por otro lado, nos dimos cuenta de que después de que formulase varias de las preguntas, empezó a pedir nuestra opinión respecto a las respuestas que nos daba o a preguntarnos nuestra opinión sobre cuál de las respuestas considerábamos correcta. Es decir, estaba aprendiendo de nosotros.

Inicialmente planteamos seguir esta estrategia para generar preguntas en la prueba de evaluación continua y en el examen final. Como no queríamos que penalizase en la nota global de los alumnos aquellas preguntas generadas por *AI*, en la prueba continua sugerimos a los alumnos que las contestasen para entender y ver como serían las preguntas de prácticas en el examen final, aunque no contarían en la evaluación. No obstante, la mayor parte de los alumnos no las contestaron por lo que no hemos podido interpretar correctamente esos resultados y llegar a conclusiones al respecto.

En cuanto al examen final, en éste incluimos para cada una de las prácticas/seminarios dos preguntas, como habíamos planteado inicialmente. En este examen las preguntas de prácticas no tienen penalización, por lo que las respuestas incorrectas no restan. A pesar de ello, hay un porcentaje de alumnos que sigue sin contestar alguna pregunta. Un total de 208 alumnos se presentaron al examen y a todos ellos les planteamos 24 preguntas, 12 generadas por nosotros y 12 por *AI*, correspondientes a 9 prácticas y 3 seminarios. Recogimos para cada una de las preguntas el número de aciertos, fallos y blancos (no contestados), como se recoge

en la gráfica y comparamos las dos preguntas formuladas para cada práctica o seminario (Gráfica 1). Al contrario que en la evaluación continua, en este caso los alumnos contestaron la mayor parte de las preguntas formuladas. Globalmente respondieron de manera correcta a más preguntas planteadas por IA que por el equipo docente (68,7% frente a 54,4 % respectivamente).



Gráfica 1. Representación gráfica de las respuestas a las distintas preguntas planteadas por el equipo docente (*human*) o IA (ChatGPT) para cada uno de los seminarios o prácticas. Los asteriscos indican el grado de significación estadística obtenido para la prueba de Chi Cuadrado que compara la tasa de éxito entre la pregunta formulada por humanos y por ChatGPT.

Hemos analizado de manera individualizada las respuestas obtenidas para cada par de preguntas de la misma práctica/seminario, como se muestra en la gráfica. Además, determinamos si había diferencias estadísticamente significativas en su contestación a las dos preguntas pareadas de cada práctica o seminario utilizando el test estadístico de la prueba de Chi-cuadrado, cuyos resultados se muestran en la tabla 1. Las diferencias estadísticamente significativas en este test se muestran con asteriscos sobre las columnas con las respuestas.

Prueba de Chi-cuadrado			
Preguntas	Valor Chi-square	p-valor	Significación
Seminarios			
1 y 2	3,174	0,2046	ns
3 y 4	1,725	0,4222	ns
5 y 6	8,128	0,0172	*
Prácticas			
7 y 8	11,29	0,0035	**
9 y 10	13,8	0,001	**
11 y 12	44,07	< 0,0001	***
13 y 14	1,664	0,4352	ns
15 y 16	19,62	< 0,0001	***
17 y 18	1,209	0,5464	ns
19 y 20	22,27	< 0,0001	***
21 y 22	111,3	< 0,0001	***
23 y 24	28,49	< 0,0001	***

Tabla 1. Análisis estadístico de las respuestas a las preguntas pareadas planteadas para cada práctica o seminario, utilizando la prueba estadística de Chi-cuadrado. Leyenda: ns, no significativo; *: p-valor < 0.05; **: p-valor < 0.01; ***: p-valor < 0.001.

En conclusión, podemos afirmar que la herramienta on-line ChapGPT podría utilizarse para la elaboración de cuestionarios destinados a evaluar el aprendizaje de los alumnos. No obstante, para poder utilizarse es necesaria una supervisión muy estrecha y dirección para que la aplicación genere correctamente las preguntas que quieren prepararse. Por otro lado, y analizando las contestaciones de los alumnos a las preguntas planteadas por los dos métodos, Chat GPT no es capaz de elaborar de manera tan precisa preguntas sobre habilidades y destrezas, que es lo que debe evaluarse en unas prácticas. Las preguntas son más teóricas, y por ende, más fáciles de responder para los alumnos. Este hecho debe ser tomado en consideración si se pretende implantar herramientas de AI en la docencia universitaria, especialmente en titulaciones con alto grado de experimentalidad y donde es necesario evaluar el grado de adquisición de destrezas y habilidades por parte del alumnado.