

MEMORIA FINAL¹

Compromisos y Resultados

Proyectos de Innovación y Mejora Docente 2023/2024

Identificación del proyecto	
Código	sol-202300256982-tra
Título	Utilización de la Inteligencia Artificial como herramienta docente para mejorar y orientar al aprendizaje los procesos de evaluación en el ámbito universitario
Responsable	Miguel Ángel Gómez Ruiz

1. Describa los resultados obtenidos a la luz de los objetivos y compromisos que adquirió en la solicitud de su proyecto. Incluya tantas tablas como objetivos contempló.

Objetivo nº 1	<i>Analizar las propuestas de ChatGPT para la elección y redacción de resultados de aprendizaje, competencias y tareas de aprendizaje y evaluación en el contexto de asignaturas universitarias.</i>
Actividades que había previsto en la solicitud del proyecto:	<ol style="list-style-type: none"> 1. Extraer de ChatGPT una propuesta de resultados de aprendizaje y competencias enmarcadas en distintas asignaturas universitarias. 2. Consultar a la IA alternativas para tareas de aprendizaje y evaluación a desarrollar en el contexto de asignaturas universitarias. 3. Valorar mediante un cuestionario online la comprensión, corrección, coherencia y pertinencia de las propuestas realizadas. 4. Comparar las propuestas con las planteadas en la realidad de las asignaturas para reflexionar hasta qué punto pueden ser de utilidad las ideas generadas.
Actividades realizadas y resultados obtenidos:	<ol style="list-style-type: none"> 1. Nos situamos en ChatGPT (https://chatgpt.com/) 2. Contextualizamos la información que vamos a pedirle, especialmente sobre la asignatura con la que se hicieron las pruebas. 3. Le pedimos que nos dé distintas propuestas para la asignatura de: a) Resultados de aprendizaje; b) Competencias y c) Tareas de aprendizaje y evaluación. 4. Se compararon las propuestas realizadas por ChatGPT con la realidad de las asignaturas (ficha y desarrollo real). 5. Se interactuó con la herramienta y se concluyó cumplimentando el siguiente cuestionario: https://forms.gle/yM4DFmd8jh5BXyY99

¹ Esta memoria no debe superar las 6 páginas.

	<p>Las principales conclusiones fueron:</p> <ol style="list-style-type: none"> Los resultados de aprendizaje que propone ChatGPT son considerados demasiado genéricos y superficiales. Las competencias también son bastantes comprensibles, pero mejor formuladas y más útiles. Las tareas de aprendizaje y evaluación propuestas han sido valoradas de forma muy positiva. Son el componente al que más utilidad se le ha otorgado.
<p>Objetivo nº 2</p>	<p><i>Valorar las aportaciones de la IA en la fase de preparación del proceso de evaluación, concretamente en el establecimiento de criterios de evaluación, creación y diseño de instrumentos, así como la previsión de los agentes evaluadores.</i></p>
<p>Actividades que había previsto en la solicitud del proyecto:</p>	<ol style="list-style-type: none"> 1. Consultar con la IA la redacción de criterios de evaluación para tareas de aprendizaje concretas en el marco de asignaturas universitarias. 2. Realización con ChatGPT de distintos instrumentos de evaluación que reflejen los criterios establecidos para la evaluación de las tareas. 3. Plantear cómo podría desarrollarse la evaluación en cuestión atendiendo a las distintas modalidades participativas (autoevaluación, evaluación entre iguales, evaluación compartida y evaluación del profesorado). 4. Analizar mediante un cuestionario online la valía, utilidad, coherencia y corrección de las propuestas realizadas. Prestar especial atención a la posibilidad de transferencia de los instrumentos de evaluación al entorno EV/ALCOMIX-FLOASS instalado en el Campus Virtual de la Universidad de Cádiz.
<p>Actividades realizadas y resultados obtenidos:</p>	<ol style="list-style-type: none"> 1. Nos situamos en ChatGPT (https://chatgpt.com/) 2. Contextualizamos la información que vamos a pedirle, especialmente sobre la asignatura y la tarea concreta con la que se hicieron las pruebas. 3. Le pedimos que nos aportara criterios de evaluación para una o varias tareas de nuestras asignaturas. 4. Posteriormente, a partir de esos criterios, le pedimos a al sistema que nos creara instrumentos de evaluación. 5. Le consultamos cómo se podría desarrollar la evaluación según los agentes evaluadores (autoevaluación, evaluación entre iguales...). 6. Se interactuó con la herramienta y se concluyó cumplimentando el siguiente cuestionario: https://forms.gle/fvoxaFr3zBuReQPz7 <p>Las principales conclusiones fueron:</p> <ol style="list-style-type: none"> a. Los criterios de evaluación que redacta ChatGPT son valorados de manera positiva, destacando su corrección y validez. b. Los instrumentos de evaluación parecen tener menores posibilidades de utilizarse en la realidad. Se destaca el apoyo que puede aportar esta IA para realizar

	<p>instrumentos complejos.</p> <p>c. Sobre los agentes evaluadores se ha detectado que hace sugerencias que se basan en combinar varias o todas las modalidades existentes.</p> <p>d. La utilidad de las ideas generadas sobre criterios, instrumentos y modalidades es mayor que en la anterior etapa de diseño.</p>
<p>Objetivo nº 3 <i>Estudiar las funcionalidades que puede aportar ChatGPT en la retroalimentación personalizada y formativa a los estudiantes a partir de la realización de las tareas de aprendizaje y evaluación.</i></p>	
<p>Actividades que había previsto en la solicitud del proyecto:</p>	<p>1. Pedir la retroalimentación a ChatGPT de varias tareas textuales de aprendizaje, ya sean estas realmente hechas por estudiantes en cursos actuales y anteriores o bien, ejemplificadas por el docente.</p> <p>2. Analizar y reflexionar sobre la calidad y relevancia de la retroalimentación aportada por la IA. Considerar las posibilidades y limitaciones que podría tener esta herramienta para la realización de retroalimentación automática para los estudiantes universitarios. Esto se reflejará en un documento online que permita el desarrollo de ideas y comentarios por parte de los docentes.</p>
<p>Actividades realizadas y resultados obtenidos:</p>	<p>1. Nos situamos en ChatGPT (https://chatgpt.com/)</p> <p>2. Contextualizamos la información que vamos a pedirle, especialmente sobre la asignatura y la tarea concreta.</p> <p>3. Se informó sobre los criterios de evaluación a partir de los que tuvo que aportar la retroalimentación.</p> <p>4. Se utilizaron ejemplos de tareas reales sobre las asignaturas con las que se hicieron las pruebas.</p> <p>5. Analizamos y reflexionamos sobre la retroalimentación que nos aportó.</p> <p>6. Se interactuó con la herramienta y se concluyó cumplimentando el siguiente cuestionario: https://forms.gle/yXeNX4MTqC72unnda</p> <p>Las principales conclusiones fueron:</p> <p>a. Se aprecian las posibilidades formativas de la retroalimentación aportada por ChatGPT. Podría servir como base para los propios estudiantes para mejorar sus producciones, fomentando su aprendizaje autónomo.</p> <p>b. Como limitación, se destaca que la información que ofrece se puede considerar en ocasiones demasiado genérica, sin concretar en qué aspectos profundizar ni contextualizar la retroalimentación generada.</p>
<p>Objetivo nº 4 <i>Comprobar la pertinencia y coherencia de la IA en la cumplimentación de instrumentos de evaluación, la valoración y calificación de tareas de aprendizaje, así como en la realización de informes finales sobre la actividad y rendimiento del alumnado universitario.</i></p>	
<p>Actividades que había previsto en la</p>	<p>1. Probar la capacidad de ChatGPT para la cumplimentación de</p>

solicitud del proyecto:	<p><i>instrumentos de evaluación, la valoración del grado de consecución de criterios y la calificación de tareas de aprendizaje de forma automática en el ámbito universitario.</i></p> <p><i>2. Establecer la adecuación de las valoraciones y calificaciones a las tareas de la IA respecto a las que habrían realizado los docentes. Comparar los resultados para la obtención de posibles diferencias o similitudes estadísticamente significativas.</i></p> <p><i>3. Conocer las propuestas de informes finales personalizados a las tareas evaluadas que podría hacer ChatGPT.</i></p> <p><i>4. Valorar el posible interés de esos informes personalizados automáticos para la formación de los estudiantes mediante un cuestionario online.</i></p>
Actividades realizadas y resultados obtenidos:	<p>1. Nos situamos en ChatGPT (https://chatgpt.com/)</p> <p>2. Contextualizamos la información que vamos a pedirle, especialmente sobre la asignatura y la tarea concreta.</p> <p>3. Se informó sobre los criterios de evaluación a partir de los que tuvo que realizar un instrumento con el que valoró y calificó los ejemplos aportados.</p> <p>4. Se introdujo varios ejemplos de tareas reales y le pedimos que complimentara el instrumento y que aportara una calificación final.</p> <p>5. Analizamos la adecuación de las valoraciones y calificaciones. Comparamos las propuestas con las notas reales.</p> <p>6. Le pedimos a la IA que realizara un informe final personalizado sobre cada una de las tareas evaluadas y valoramos su interés formativo.</p> <p>7. Se interactuó con la herramienta y se concluyó complimentando el siguiente cuestionario: https://forms.gle/WvPwVb8rBDu9WHdF8</p> <p>Las principales conclusiones a partir del análisis de los resultados obtenidos, fueron:</p> <p>a. Se ha comprobado que de forma general ChatGPT propone calificaciones más elevadas y homogéneas que las otorgadas de manera real por los docentes. No existe una relación lineal entre ambas calificaciones.</p> <p>b. Como aspecto positivo, se ha destacado que argumenta sus valoraciones. Algunos de los aspectos negativos mencionados son que resulta excesivamente repetitivo y genérico en sus valoraciones y que no parece existir un cálculo homogéneo ni sólido de las calificaciones.</p> <p>c. Se aprecia que los informes finales generados por ChatGPT tienen interés formativo.</p>

Objetivo nº 4

Informar y difundir las conclusiones sobre la utilidad de ChatGPT como herramienta docente para la mejora y la orientación de los procesos de evaluación universitarios hacia el aprendizaje.

Actividades que había previsto en la

Para finalizar, realizaremos una técnica DAFO (Debilidades,

solicitud del proyecto:	<p><i>Amenazas, Fortalezas y Oportunidades) por parte del equipo de trabajo, donde se valorará de forma conjunta los beneficios y limitaciones detectadas de la IA en las distintas fases del proceso de evaluación. Con toda la información disponible, se difundirán las conclusiones mediante dos vías fundamentales:</i></p> <p><i>1. Informe técnico por escrito que contendrá todos los detalles de los hallazgos realizados en este proyecto. Será publicado en repositorio institucional RODIN (https://rodin.uca.es/)</i></p> <p><i>2. Realización de uno o varios vídeos divulgativos (atendiendo a las posibilidades y contenidos finales del proyecto), que serán publicados en el canal de YouTube del Grupo de Investigación EVALfor</i></p>
Actividades realizadas y resultados obtenidos:	<p>Se realizó la DAFO prevista. En la misma se han especificado 9 Debilidades, 9 Amenazas, 6 Fortalezas y 6 Oportunidades (consultar Informe Final).</p> <p>De igual manera, se ha realizado el Informe Final escrito donde se recogen todos los detalles y resultados de las pruebas realizadas durante el proyecto.</p> <p>También se ha difundido un vídeo divulgativo con las principales conclusiones del proyecto de innovación.</p> <p>Tanto del Informe Final, como del vídeo divulgativo, se ofrece más información en el apartado 4 de esta memoria.</p>

- Realice una breve valoración sobre la influencia del proyecto ejecutado en la evolución de las asignaturas implicadas.

<p style="text-align: center;"><i>Análisis del impacto de la innovación en las asignaturas relacionadas con el proyecto</i></p> <p>Como se tenía previsto en el proyecto: “se debe señalar que la finalidad de este proyecto no será la utilización directa en las asignaturas implicadas, sino establecer un periodo de formación, estudio en profundidad y reflexión sobre las implicaciones de una aplicación real en próximos cursos, pero teniendo de referencia los contenidos y prácticas de las asignaturas presentes”.</p> <p>Las repercusiones de la utilización directa de ChatGPT en las asignaturas eran impredecibles, por lo tanto, se descartó desde el comienzo esta posibilidad.</p> <p>No obstante, se ha apreciado el valor de esta IA para la futura mejora en el diseño e implementación de los procesos de evaluación de las asignaturas implicadas, siempre teniendo en consideración las limitaciones que actualmente encontramos en esta herramienta.</p>

- Incluya en la siguiente tabla el número de alumnos matriculados y el de respuestas recibidas en cada opción y realice una valoración crítica sobre la influencia que el proyecto ha ejercido en la opinión de los alumnos.

Opinión de los alumnos al inicio del proyecto				
Número de alumnos matriculados:				
<i>Valoración del grado de dificultad que cree que va a tener en la comprensión de los contenidos y/o en la adquisición de competencias asociadas a la asignatura en la que se enmarca el proyecto de innovación docente</i>				
Ninguna dificultad	Poca dificultad	Dificultad media	Bastante dificultad	Mucha dificultad

Opinión de los alumnos en la etapa final del proyecto				
<i>Valoración del grado de dificultad que ha tenido en la comprensión de los contenidos y/o en la adquisición de competencias asociadas a la asignatura en la que se enmarca el proyecto de innovación docente</i>				
Ninguna dificultad	Poca dificultad	Dificultad media	Bastante dificultad	Mucha dificultad
<i>Los elementos de innovación y mejora docente aplicados en esta asignatura han favorecido mi comprensión de los contenidos y/o la adquisición de competencias asociadas a la asignatura</i>				
Nada de acuerdo	Poco de acuerdo	Ni en acuerdo ni en desacuerdo	Muy de acuerdo	Completamente de acuerdo
En el caso de la participación de un profesor invitado				
<i>La participación del profesor invitado ha supuesto un gran beneficio en mi formación</i>				
Nada de acuerdo	Poco de acuerdo	Ni en acuerdo ni en desacuerdo	Muy de acuerdo	Completamente de acuerdo
Valoración crítica sobre la influencia que ha ejercido el proyecto en la opinión de los alumnos				
Como se ha comentado anteriormente, este proyecto no tenía previsto la aplicación real en las asignaturas implicadas, por lo que no se ha trabajado de forma directa con el alumnado.				

4. Describa las medidas de difusión a las que se comprometió en la solicitud y las que ha llevado a cabo².

Descripción de las medidas comprometidas en la solicitud
Como se ha comentado en la descripción del Objetivo nº5 y teniendo en consideración la actualidad y necesidad de formación docente en este ámbito de la Inteligencia Artificial y su relación con la construcción de los procesos de evaluación en el ámbito universitarios, se pretende difundir los resultados de forma divulgativa mediante uno o varios vídeos que serán publicados en el canal de YouTube del Grupo de Investigación EVALfor (Evaluación en Contextos Formativos – SEJ509) disponible en: https://www.youtube.com/@evalfor . Esto se realizará en el último tramo de este proyecto.
Descripción de las medidas que se han llevado a cabo
Se han llevado a cabo las medidas previstas: 1. Informe final por escrito (67 páginas) que contiene todos los detalles de los hallazgos realizados en este proyecto. Se adjunta al procedimiento junto a esta memoria. También se encuentra en revisión en el repositorio institucional RODIN (https://rodin.uca.es/) a la espera de su difusión pública definitiva. 2. Difusión de las conclusiones y resultados de forma divulgativa mediante un vídeo publicado en el canal de YouTube del Grupo de Investigación EVALfor (Evaluación en Contextos Formativos – SEJ509) disponible en: https://www.youtube.com/@evalfor . El enlace directo al vídeo es el siguiente: https://youtu.be/t65R2hCK6PA

² Si en la solicitud no indicó compromiso de difusión de resultados este criterio no se tendrá en cuenta en la evaluación

Proyecto de innovación

Utilización de la Inteligencia Artificial como herramienta docente para mejorar y orientar al aprendizaje los procesos de evaluación en el ámbito universitario

Curso 2023/2024



INFOME FINAL

INFORME FINAL

Proyecto “Utilización de la Inteligencia Artificial como herramienta docente para mejorar y orientar al aprendizaje los procesos de evaluación en el ámbito universitario”. Código: sol-202300256982-tra. Convocatoria de Proyectos de Innovación y Mejora Docente. Vicerrectorado de Profesorado. Unidad de Formación e Innovación Docente. Universidad de Cádiz. Curso 2023/2024.

RESPONSABLE DEL PROYECTO

Miguel Ángel Gómez Ruiz

Departamento de Didáctica de la Universidad de Cádiz

PARTICIPANTES EN EL PROYECTO

Nicolás Ponce González, Laura Sánchez Calleja, María del Carmen Molinero Roca, Andrea González Elorza, Mónica López Gil, Gregorio Rodríguez Gómez, María Soledad Ibarra Sáiz, Jaione Cubero Ibáñez, Francisco Javier Amores Fernández, Cristina Goenechea Permisán y Ana Gázquez Romera del Departamento de Didáctica de la Universidad de Cádiz. Juan Antonio Caballero Hernández del Departamento de Ingeniería Informática de la Universidad de Cádiz.

ÍNDICE

1. Introducción.....	1
2. Etapa de diseño de la evaluación.....	4
2.1. Resultados de aprendizaje propuestos.....	4
2.2. Competencias propuestas.....	7
2.3. Tareas de aprendizaje y evaluación propuestas.....	10
2.4. Utilidad de las propuestas y aspectos aprovechables.....	15
3. Etapa de preparación de la evaluación.....	18
3.1. Redacción de criterios de evaluación.....	18
3.2. Realización de instrumentos de evaluación.....	23
3.3. Modalidades de evaluación propuestas.....	26
3.4. Utilidad de las propuestas y aspectos aprovechables.....	31
4. Etapa de retroalimentación al alumnado.....	35
4.1. Calidad y relevancia de la retroalimentación.....	35
4.2. Posibilidades y limitaciones de la retroalimentación.....	40
5. Etapa de valoración de las tareas.....	44
5.1. Complimentación de instrumentos, valoraciones y calificaciones.....	44
5.2. Realización de informes personalizados.....	52
6. Debilidades, amenazas, fortalezas y oportunidades.....	57
6.1. Debilidades.....	57
6.2. Amenazas.....	58
6.3. Fortalezas.....	59
6.4. Oportunidades.....	59
7. Conclusiones.....	61

1. INTRODUCCIÓN

La irrupción social de la Inteligencia Artificial (IA), concretamente del desarrollo de OpenAI denominado ChatGPT (Generative Pre-trained Transformer) a finales de 2022, ha supuesto reacciones extremas ante esta herramienta, que van desde considerarla como el mejor asistente virtual o “chatbot” de la historia hasta plantearse efectos catastrofistas para la democracia, el empleo e incluso la humanidad (García-Peñalvo, 2023). El impacto real de este modelo de lenguaje informático solo se ha comenzado a atisbar. Su uso generalizado, su acceso libre y su facilidad para la interacción, junto con la potencia de sus utilidades, lo ha convertido en un tema de debate recurrente en los medios de comunicación o en diferentes asociaciones profesionales. No obstante, uno de los ámbitos en el que desde su lanzamiento está especialmente presente, es en la educación.

El uso por parte del estudiantado, de diferentes niveles, de esta herramienta, está provocando importantes polémicas en las instituciones académicas. Su repentina aparición y las consecuencias desconocidas a medio y largo plazo, ha desencadenado una variedad de reacciones en los responsables educativos. Existen contextos en los que se ha decidido limitar o directamente prohibir su uso. Sin embargo, otras corrientes defienden la necesidad de integrar la IA en los procesos de enseñanza-aprendizaje para potenciar la creatividad y el pensamiento crítico del alumnado (Zhai, 2023).

Pero si existe un aspecto educativo en los que se ponen todos los focos cuando se habla de ChatGPT, es su impacto en los procesos de evaluación. Normalmente se asume que es una herramienta que el alumnado puede utilizar de manera fraudulenta para superar los requerimientos de asignaturas o titulaciones completas. Pero quizá, lo realmente incómodo para los docentes es plantearse si ciertas tareas educativas se han vuelto obsoletas o la necesidad de modificar objetivos de aprendizaje, actividades, nuestro papel como docentes o las propias prácticas de evaluación (García-Peñalvo, 2023; Zhai, 2023).

Teniendo en consideración la actualidad, relevancia e impacto de la IA, concretamente de ChatGPT en la educación y especialmente en la evaluación, con este proyecto se ha pretendido analizar las posibilidades de esta herramienta para resignificarla desde “enemigo” a “aliado” de los docentes en los procesos de evaluación, algo que necesariamente viene por la mejora de las competencias del profesorado ante el actual y futuro avance de los “chatbots” (Tlili et al., 2023), así como en las propias tendencias e investigaciones sobre el campo de la evaluación.

En las últimas décadas, se han realizado diversos avances teóricos y prácticos en el ámbito evaluativo que siguen sin implantarse de forma generalizada en la educación formal, especialmente, en la universidad. A pesar de las distintas corrientes y conceptualizaciones, existe un propósito común de relacionar estrechamente la evaluación con el aprendizaje del alumnado para fomentar su capacidad de autorregulación. En este sentido, la importancia central de la retroalimentación, la participación de los estudiantes, la transparencia del proceso, la coherencia de su diseño e implantación, así como su relevancia académica, social y profesional, son algunos de los aspectos fundamentales dentro de este tópico de investigación (Gómez-Ruiz y Quesada-Serra, 2020).

Con este Proyecto de Innovación y Mejora Docente se ha pretendido analizar en profundidad las posibilidades y limitaciones de la IA, concretamente de ChatGPT, como herramienta docente en el apoyo, asesoramiento y realización de propuestas para mejorar los procesos de evaluación completos en las asignaturas universitarias. Este análisis requirió profundizar en las aportaciones de este sistema de procesamiento de lenguaje natural a las distintas fases de la evaluación, comenzando con la etapa de diseño (redacción de resultados de aprendizaje, competencias y establecimiento de tareas de aprendizaje/evaluación), la etapa de preparación (establecimiento de criterios de evaluación, diseño de instrumentos y delimitación de los agentes evaluadores), la etapa de retroalimentación y la etapa final de valoración (cumplimentación de instrumentos, calificaciones y realización de informes finales).

Aunque se planteó un estudio completo de todo el proceso de evaluación, se prestó especial atención a las posibilidades de ChatGPT para la aportación de retroalimentación personalizada y automática, ya que es una de las mayores dificultades para el profesorado en el contexto universitario, pero que a su vez es la más importante, ya que es imprescindible ofrecer información de calidad a los estudiantes sobre sus desempeños, adecuada además en tiempo y forma, para poder extraer el máximo potencial formativo de las actividades desarrolladas por el alumnado (Ajjawi y Boud, 2018).

En último lugar, se debe señalar que la finalidad de este proyecto no fue la utilización directa en las asignaturas implicadas, sino establecer un periodo de formación, estudio en profundidad y reflexión sobre las implicaciones de una aplicación real en próximos cursos, pero teniendo de referencia los contenidos y prácticas de las asignaturas presentes. Igualmente, y siendo conscientes de la novedad y relevancia de la Inteligencia Artificial, se ha difundido un vídeo divulgativo sobre las conclusiones obtenidas durante este estudio disponible en: <https://youtu.be/t65R2hCK6PA>

Referencias bibliográficas:

- Ajjawi, R. y Boud, D. (2018). Examining the nature and effects of feedback dialogue. *Assessment & Evaluation in Higher Education*, 43(7), 1106–1119. <https://doi.org/10.1080/02602938.2018.1434128>
- García-Peñalvo, F. J. (2023). La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico. *Education in the Knowledge Society (EKS)*, 24, e31279. <https://doi.org/10.14201/eks.31279>
- Gómez-Ruiz, M.A. y Quesada-Serra, V. (2020). Análisis de las calificaciones compartidas en la modalidad participativa de la evaluación colaborativa entre docente y estudiantes. *Revista Electrónica de Investigación y Evaluación Educativa*, 26 (1). <https://doi.org/10.7203/relieve.26.1.16567>
- Tlili, A., Shehata, B., Adarkwah, M.A. et al. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, art.15. <https://doi.org/10.1186/s40561-023-00237-x>
- Zhai, X. (2023). ChatGPT User Experience: Implications for Education. *SSRN*. <http://doi.org/10.2139/ssrn.4312418>

2. ETAPA DE DISEÑO DE LA EVALUACIÓN

El objetivo propuesto para esta etapa era analizar las propuestas de ChatGPT para la elección y redacción de resultados de aprendizaje, competencias y tareas de aprendizaje y evaluación en el contexto de asignaturas universitarias. Para llevarlo a cabo, se desarrollaron las siguientes acciones:

1. Extraer de ChatGPT una propuesta de resultados de aprendizaje y competencias enmarcadas en distintas asignaturas universitarias.
2. Consultar a la IA alternativas para tareas de aprendizaje y evaluación a desarrollar en el contexto de asignaturas universitarias.
3. Valorar mediante un cuestionario online la comprensión, corrección, coherencia y pertinencia de las propuestas realizadas.
4. Comparar las propuestas con las planteadas en la realidad de las asignaturas para reflexionar hasta qué punto pueden ser de utilidad las ideas generadas.

2.1. RESULTADOS DE APRENDIZAJE PROPUESTOS

Los resultados de aprendizaje que ha propuesto ChatGPT en las 8 pruebas realizadas por los docentes implicados en el proyecto de innovación, se han valorado utilizando una escala Likert compuesta por los valores nada (1), poco (2), algo (3), bastante (4) y totalmente (5), considerando si son comprensibles, correctos, coherentes y pertinentes.

A continuación, se ofrecen dos ejemplos de resultados de aprendizajes aportados por esta inteligencia artificial en el contexto de asignaturas universitarias:

Ejemplo 1

Resultados de aprendizaje propuestos por ChatGPT para la asignatura "Gestión de Proyectos"

Aquí tienes una propuesta de resultados de aprendizaje y competencias para la asignatura optativa "Gestión de Proyectos" en el Grado de Administración y Dirección de Empresas:

Resultados de Aprendizaje:

Comprender los Fundamentos de la Gestión de Proyectos: Los estudiantes deberán adquirir conocimientos sólidos sobre los principios y conceptos fundamentales en la gestión de proyectos, incluyendo la planificación, ejecución y control.

Aplicar Herramientas y Técnicas de Planificación: Desarrollar la capacidad de utilizar herramientas y técnicas modernas de planificación de proyectos, como diagramas de Gantt, redes PERT/CPM, y software de gestión de proyectos.

Gestión Efectiva de Recursos: Aprender a asignar y gestionar eficientemente los recursos disponibles en un proyecto, teniendo en cuenta restricciones de tiempo, presupuesto y personal.

Evaluación y Mitigación de Riesgos: Adquirir habilidades para identificar, evaluar y mitigar los riesgos asociados a los proyectos, utilizando técnicas apropiadas de gestión de riesgos.

Liderazgo y Trabajo en Equipo: Desarrollar habilidades de liderazgo y trabajo en equipo para gestionar con éxito proyectos que involucren a diversos miembros y partes interesadas.

Ejemplo 2

Resultados de aprendizaje propuestos por ChatGPT para la asignatura "Conflicto, Desarrollo Sostenible y Medio Ambiente"

Aquí tienes algunos resultados de aprendizaje que podrías trabajar en la asignatura "Conflicto, Desarrollo Sostenible y Medio Ambiente":

1. Conocimiento teórico y conceptual

- Los estudiantes serán capaces de **definir y explicar** los conceptos clave relacionados con el desarrollo sostenible, el conflicto ambiental y la cultura de paz.
- Los estudiantes podrán **identificar y analizar** las relaciones entre los conflictos sociales y ambientales, y cómo estos influyen en el desarrollo sostenible.

2. Habilidades analíticas y críticas

- Los estudiantes serán capaces de **analizar críticamente** casos de conflictos ambientales, comprendiendo las causas, actores involucrados, y los impactos sobre el entorno y las comunidades.
- Los estudiantes podrán **evaluar diferentes estrategias** de intervención para la resolución pacífica de conflictos ambientales y proponer soluciones sostenibles.
- Serán capaces de **interpretar políticas y normativas ambientales** internacionales y locales, reflexionando sobre su eficacia en la prevención y gestión de conflictos.

3. Habilidades de investigación

- Los estudiantes serán capaces de **realizar investigaciones** sobre un conflicto ambiental específico, identificando las dinámicas socio-políticas y ecológicas que lo caracterizan.
- Serán capaces de **proponer soluciones innovadoras** basadas en el desarrollo sostenible y la cultura de paz, aplicando teorías y herramientas de análisis aprendidas en la asignatura.

4. Competencias en mediación y resolución de conflictos

- Los estudiantes adquirirán competencias para **aplicar herramientas de mediación** y diálogo en conflictos ambientales, promoviendo la cultura de paz.
- Serán capaces de **diseñar procesos de negociación** y facilitación para resolver conflictos socioambientales, considerando las perspectivas de todos los actores involucrados.

5. Conciencia ética y compromiso social

- Los estudiantes desarrollarán una **conciencia crítica sobre la justicia ambiental**, entendiendo el impacto desigual que los conflictos ambientales tienen sobre las comunidades más vulnerables.

- Serán capaces de **articular la importancia de los derechos humanos** en el contexto de la sostenibilidad y la resolución de conflictos, comprometiéndose con soluciones éticas y equitativas.
- Fomentarán una **actitud proactiva hacia el activismo ambiental**, y la promoción del desarrollo sostenible como herramienta para la paz.

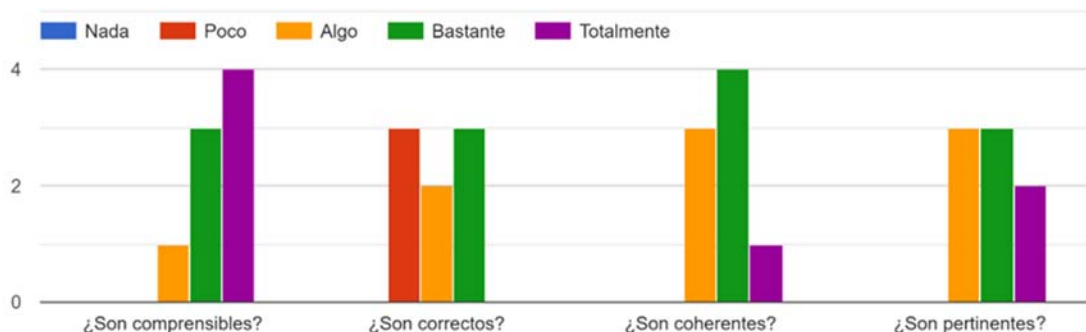
6. Trabajo en equipo y colaboración

- Los estudiantes podrán **trabajar en equipo** para elaborar proyectos colectivos de análisis y resolución de conflictos ambientales.
- Desarrollarán habilidades de **colaboración interdisciplinaria**, integrando conocimientos de diversas áreas (política, ecología, economía) para abordar problemas complejos.

Como se puede observar en la Figura 1, el criterio mejor valorado y el único cuya media supera el 4, es que estos resultados de aprendizaje se han considerado como bastante comprensibles ($\bar{X}=4,4$ y $\sigma=0,7$). En el extremo opuesto, y con una media que se sitúa en el 3 ($\sigma=0,9$), el criterio con menor valoración es el relacionado con su corrección. Respecto a que sean coherentes y pertinentes, se valoran de una forma bastante similar, considerando su coherencia con $\bar{X}=3,8$ ($\sigma=0,7$) y su pertinencia con $\bar{X}=3,9$ ($\sigma=0,8$).

Figura 1

Valoraciones sobre los resultados de aprendizaje propuestos por ChatGPT



Las valoraciones realizadas son justificadas con distintos comentarios y reflexiones por parte de los docentes participantes en el proyecto, a continuación, se reproducen algunas de estas apreciaciones:

“Presentan cierta coherencia ya que utiliza términos o conceptos que son acordes y/o afines dentro del marco de la asignatura y en el ámbito educativo actual, como son: evaluar críticamente, diseñar proyectos, reflexionar de forma crítica, resolución de problemas, trabajo en equipo, inclusión y diversidad, desarrollo emocional y social, autonomía y pensamiento crítico. Lo que me lleva a reflexionar que lo que configura y crea está basado más en el continente que realmente en el contenido. Me

explico, es como el alumnado cuando en sus trabajos emplean palabras que escuchan en las asignaturas porque suenan correctas, adecuadas y acordes a lo que se están trabajando, pero no han construido realmente su sentido y significado pues carece de argumentación y fundamentación desde teorías en las que enmarcar sus aportaciones. Y es esto lo que aprecio en la respuesta del Chat GPT, que reproduce y no construye con criterio formativo especializado en la materia y contrastado y desarrollado de diferentes fuentes." (P05)

“Los resultados de aprendizaje son coherentes e interesantes. No obstante, no llegan a ser correctos en su formulación. De hecho, parecen más competencias que resultados de aprendizajes, aunque esto es algo común si no se tiene una formación específica al respecto.... son genéricos, sin contexto o nivel de rendimiento. También parecen algo abstractos y superficiales, aunque realmente pueden considerarse coherentes y pertinentes.” (P07)

2.2. COMPETENCIAS PROPUESTAS

Considerando los mismos criterios y la misma escala que se utilizaron para valorar los resultados de aprendizaje, se han analizado las competencias que aporta ChatGPT para el diseño de varias asignaturas universitarias. A continuación, se ofrecen dos ejemplos de dichas competencias escritas por esta inteligencia artificial:

Ejemplo 3

Competencias propuestas por ChatGPT para la asignatura “Tutoría y Familia”

Aquí te propongo algunas competencias que podrían desarrollarse en la asignatura "Tutoría y Familia", con un enfoque en la Acción Tutorial dentro del contexto de la Educación Primaria:

1. Competencia en la gestión de la Acción Tutorial

Capacidad para planificar, implementar y coordinar la acción tutorial en el aula, atendiendo a las necesidades académicas, emocionales y sociales del alumnado.

2. Competencia en comunicación y colaboración con las familias

Habilidad para establecer canales de comunicación efectivos con las familias, promoviendo su participación activa en el proceso educativo y creando un entorno colaborativo para el desarrollo integral del alumno.

3. Competencia en orientación y apoyo emocional

Capacidad para identificar y atender las necesidades emocionales y sociales de los alumnos, proporcionando orientación personalizada y herramientas que favorezcan su desarrollo emocional.

4. Competencia en la mediación y resolución de conflictos

Habilidad para mediar en situaciones de conflicto, tanto dentro del aula como en la relación entre la escuela y las familias, aplicando técnicas de resolución pacífica de conflictos.

5. Competencia en atención a la diversidad

Capacidad para diseñar e implementar estrategias inclusivas que favorezcan la atención a la diversidad del alumnado, incluyendo estudiantes con necesidades educativas especiales, diversidad cultural o socioeconómica.

6. Competencia en evaluación del proceso tutorial

Habilidad para diseñar herramientas y criterios de evaluación que permitan medir la efectividad de la acción tutorial y su impacto en el desarrollo académico y personal de los estudiantes.

7. Competencia en el uso de tecnologías para la tutoría

Capacidad para integrar las Tecnologías de la Información y la Comunicación (TIC) en el proceso tutorial, facilitando el seguimiento, la orientación y la interacción con las familias.

8. Competencia en empatía y escucha activa

Habilidad para aplicar la escucha activa y empatía en las interacciones con los alumnos y las familias, favoreciendo un clima de confianza y apoyo mutuo.

9. Competencia en el diseño de programas de orientación

Capacidad para desarrollar y aplicar programas de orientación académica, personal y profesional que respondan a las características y necesidades del alumnado.

10. Competencia en la dinamización de grupos

Habilidad para gestionar y dinamizar grupos de estudiantes en el aula, fomentando el trabajo en equipo, la cohesión y la convivencia positiva.

11. Competencia en trabajo en equipo y colaboración interprofesional

Capacidad para colaborar con otros docentes, personal de apoyo y profesionales externos en el diseño e implementación de estrategias tutoriales.

12. Competencia ética y responsabilidad profesional

Desarrollo de una actitud ética y responsable en la función tutorial, asegurando la confidencialidad, el respeto por la diversidad y el bienestar integral del alumnado.

Estas competencias aseguran que los futuros docentes puedan gestionar de manera integral el rol tutorial, con un enfoque humanístico y práctico, que va más allá de lo académico para abordar las dimensiones emocionales y sociales de los estudiantes.

Ejemplo 4

Competencias propuestas por ChatGPT para la asignatura "Diseño y desarrollo del currículum II"

¡Claro! Desarrollar competencias específicas para tu asignatura es crucial para guiar el proceso de aprendizaje de los estudiantes. Aquí tienes algunas competencias que podrías considerar para "Diseño, desarrollo del currículum II" en el Grado en Educación Primaria:

1. Competencia en Conocimiento Normativo:

El estudiante será capaz de analizar y aplicar la normativa educativa vigente, identificando sus implicaciones en el diseño y desarrollo del currículo en Educación Primaria.

2. Competencia en Inclusión y Equidad:

El estudiante será capaz de diseñar un currículo inclusivo que atienda a la diversidad de los estudiantes, considerando diferentes estilos de aprendizaje, necesidades especiales y contextos culturales.

3. Competencia en Organización Escolar:

El estudiante será capaz de evaluar modelos de organización escolar que fomenten la inclusión y la equidad, proponiendo estrategias para una implementación efectiva del currículo en el contexto escolar.

4. Competencia en Colaboración y Participación:

El estudiante será capaz de promover la colaboración entre docentes, personal escolar, estudiantes y padres, reconociendo la importancia de la participación comunitaria en el diseño y desarrollo del currículo.

5. Competencia en Evaluación Inclusiva:

El estudiante será capaz de desarrollar métodos de evaluación inclusivos que permitan medir el progreso de todos los estudiantes, garantizando la equidad en la evaluación del aprendizaje.

6. Competencia en Tendencias Curriculares:

El estudiante será capaz de explorar y analizar tendencias actuales en el diseño curricular, incorporando enfoques emergentes y tecnológicos para mejorar la calidad educativa.

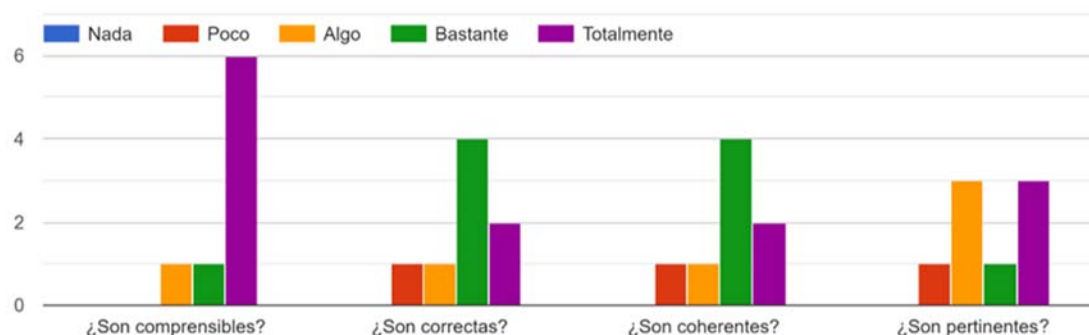
Recuerda adaptar estas competencias a los objetivos específicos de tu programa y a la normativa educativa local. Además, puedes subdividir estas competencias en objetivos más específicos si es necesario. ¿Necesitas más ayuda con algún aspecto específico?

De manera general, las competencias que ha ido proponiendo ChatGPT tienen una mejor valoración que los resultados de aprendizaje. El criterio con la puntuación más alta vuelve a ser la comprensión que alcanza una media de 4,6 ($\sigma=0,7$). No obstante, en este caso, los criterios restantes se sitúan en valores similares entre sí, en todos los casos algo por debajo del 4, así, la corrección ($\bar{X}=3,9$ y $\sigma=0,9$), la coherencia ($\bar{X}=3,9$ y $\sigma=0,9$) y la pertinencia ($\bar{X}=3,8$ y $\sigma=1,1$) presentan puntuaciones considerables

que parecen indicar que, en el caso de las competencias, su formulación es más adecuada y útil que en los resultados de aprendizaje. En la Figura 2 se puede observar un resumen de las valoraciones aportadas por los investigadores respecto a las competencias que ha propuesto la inteligencia artificial.

Figura 2

Valoraciones sobre las competencias propuestas por ChatGPT



Se resaltan a continuación varios comentarios e impresiones de los docentes implicados en el proyecto sobre las competencias propuestas después de haber realizado el análisis sobre su formulación y utilidad.

“En el caso de las competencias no ha necesitado tantas especificaciones como para los resultados de aprendizaje, igualmente considero que se necesita cierto conocimiento en el diseño de competencias para añadir los prompts adecuados que permitan a la herramienta generar la información que necesitamos.” (P03)

"Las competencias parecen adecuadas. Algo genéricas y con mucha variedad en las propuestas, pero interesantes y con una formulación que mezcla habilidades y capacidades, pero que no desencajarían en la realidad." (P06)

2.3. TAREAS DE APRENDIZAJE Y EVALUACIÓN PROPUESTAS

Para concluir con las propuestas que realiza ChatGPT sobre el diseño de la evaluación, se consideró cómo de comprensibles, correctas, coherentes y pertinentes eran las descripciones sobre las tareas de aprendizaje y evaluación que el alumnado podría realizar en distintas asignaturas universitarias. Estas tareas deberían favorecer la adquisición de competencias anteriormente formuladas, así como ayudar a la

consecución de los resultados de aprendizaje también aportados. Veamos previamente dos ejemplos de estas propuestas de tareas.

Ejemplo 5

Tareas de aprendizaje y evaluación propuestas por ChatGPT para la asignatura "Procedimientos y recursos tecnológicos de evaluación"

Aquí tienes algunas ideas para tareas de aprendizaje y evaluación en la asignatura "Procedimientos y recursos tecnológicos de evaluación":

- 1. Investigación sobre herramientas:** Realizar un análisis comparativo de diferentes herramientas tecnológicas de evaluación, presentando sus características, ventajas y desventajas en un informe.
- 2. Diseño de un instrumento de evaluación:** Crear un instrumento de evaluación digital (cuestionario, rúbrica, etc.) para un contexto educativo específico, justificando su diseño y alineación con los objetivos de aprendizaje.
- 3. Implementación de una evaluación en línea:** Planificar y llevar a cabo una evaluación utilizando una plataforma LMS, reflexionando sobre la experiencia y los resultados obtenidos.
- 4. Análisis de datos:** Recoger datos de una evaluación real o simulada y utilizar herramientas de análisis para interpretar los resultados, identificando áreas de mejora.
- 5. Estudio de caso:** Desarrollar un estudio de caso sobre la implementación de tecnologías de evaluación en una institución educativa, incluyendo análisis de impacto y recomendaciones.
- 6. Reflexión crítica:** Escribir un ensayo reflexivo sobre los desafíos éticos y pedagógicos del uso de tecnologías en la evaluación, basándose en la literatura actual y experiencias personales.
- 7. Proyecto colaborativo:** Trabajar en grupos para diseñar un sistema integral de evaluación que incorpore diferentes herramientas y metodologías tecnológicas, presentando el proyecto a la clase.
- 8. Presentación de resultados:** Elaborar y presentar un informe final sobre el impacto de una herramienta tecnológica específica en un contexto de evaluación, incluyendo feedback de compañeros y docentes.

Ejemplo 6

Tareas de aprendizaje y evaluación propuestas por ChatGPT para la asignatura "Conflicto, Desarrollo Sostenible y Medio Ambiente"

Para desarrollar y evaluar la asignatura "Conflicto, Desarrollo Sostenible y Medio Ambiente", te propongo una serie de tareas de aprendizaje y evaluación que integran tanto el trabajo teórico como práctico, promoviendo la participación activa del alumnado.

1. Estudio de caso (individual o grupal)

Descripción:

Los estudiantes analizarán un conflicto ambiental real, como el acceso al agua, la minería, la deforestación o el cambio climático. Deberán investigar las causas, los actores involucrados, las consecuencias socioambientales y las posibles soluciones.

Objetivos:

- Desarrollar habilidades analíticas para comprender la complejidad de los conflictos ambientales.
- Evaluar el papel de los actores locales, internacionales, gubernamentales y no gubernamentales.

Evaluación:

- Informe escrito que incluya análisis del conflicto, el contexto político y socioeconómico, y una propuesta de intervención sostenible.
- Exposición oral para presentar los hallazgos al resto de la clase.

2. Proyecto de intervención (grupal)

Descripción:

Los estudiantes diseñarán un proyecto de intervención para un conflicto ambiental en un contexto específico. Este proyecto debe incluir un plan de resolución pacífica del conflicto basado en el desarrollo sostenible, detallando acciones concretas, actores implicados y un cronograma de implementación.

Objetivos:

- Desarrollar habilidades para planificar intervenciones sostenibles que promuevan la paz y la resolución de conflictos.
- Fomentar la colaboración y el trabajo en equipo.

Evaluación:

- Documento del proyecto con justificación, análisis del conflicto, propuesta de intervención y recursos necesarios.
- Presentación grupal del proyecto a través de una simulación o exposición.

3. Debates sobre temas controversiales (en clase o en línea)

Descripción:

Organizar debates sobre temas controversiales relacionados con el conflicto y el desarrollo sostenible, como la explotación de recursos naturales vs. la protección ambiental, o la justicia ambiental en países en desarrollo.

Objetivos:

- Desarrollar habilidades argumentativas y de pensamiento crítico.
- Fomentar la reflexión sobre las diferentes perspectivas y soluciones a los conflictos ambientales.

Evaluación:

- Participación activa en los debates, valorando la capacidad para presentar argumentos fundamentados.
- Reflexión escrita posterior sobre el debate, evaluando las posturas y propuestas presentadas.

4. Diario reflexivo (individual)

Descripción:

A lo largo del curso, los estudiantes llevarán un diario donde reflexionarán sobre los contenidos aprendidos, su impacto personal y profesional, y cómo aplicar los conocimientos adquiridos a casos reales.

Objetivos:

- Fomentar la autorreflexión y la capacidad de conectar la teoría con la práctica.
- Desarrollar una conciencia crítica sobre el papel del desarrollo sostenible en la resolución de conflictos.

Evaluación:

- Entrega del diario reflexivo al final del curso, evaluando la profundidad de las reflexiones y la capacidad para integrar los aprendizajes.

5. Ensayo crítico (individual)

Descripción:

Elaboración de un ensayo crítico sobre un tema relevante de la asignatura (por ejemplo, "El impacto del cambio climático en los conflictos por recursos naturales" o "El papel de las políticas internacionales en la resolución de conflictos ambientales").

Objetivos:

- Desarrollar habilidades de investigación, escritura y argumentación crítica.
- Analizar problemas complejos y proponer soluciones fundamentadas.

Evaluación:

- Se evaluará la coherencia argumentativa, el uso adecuado de fuentes académicas, la originalidad del enfoque y la claridad expositiva.

6. Simulación de negociación o mediación (grupal)

Descripción:

Organizar una simulación de negociación en la que los estudiantes representen diferentes actores involucrados en un conflicto ambiental (gobierno, empresas, ONGs, comunidades locales) y traten de alcanzar un acuerdo pacífico y sostenible.

Objetivos:

- Desarrollar competencias en mediación y negociación de conflictos.
- Experimentar el proceso de toma de decisiones en escenarios de conflictos reales.

Evaluación:

- Evaluación de la participación activa en la simulación, la capacidad de defender intereses, escuchar y proponer soluciones.
- Reflexión grupal sobre los desafíos enfrentados durante la negociación y cómo mejorar el proceso.

7. Lecturas críticas y fichas de análisis (individual)

Descripción:

Asignar lecturas de textos clave sobre conflictos ambientales y desarrollo sostenible. Los estudiantes deberán entregar fichas de análisis crítico donde resuman las ideas principales, identifiquen puntos controversiales y propongan preguntas para la discusión en clase.

Objetivos:

- Mejorar la comprensión de los textos académicos y su capacidad para extraer información relevante.
- Fomentar el pensamiento crítico y la participación en las discusiones en clase.

Evaluación:

- Fichas de análisis evaluadas según la capacidad de síntesis, profundidad crítica y la formulación de preguntas que fomenten la reflexión en clase.

8. Presentación de una propuesta educativa (grupal o individual)

Descripción:

Desarrollar una propuesta educativa para sensibilizar a la población sobre la importancia del desarrollo sostenible en la resolución de conflictos ambientales. Puede ser un taller, un curso online o una campaña educativa.

Objetivos:

- Aplicar los conocimientos adquiridos en el ámbito de la educación para la paz y el desarrollo sostenible.
- Fomentar la creatividad y el compromiso social en la resolución de problemas ambientales.

Evaluación:

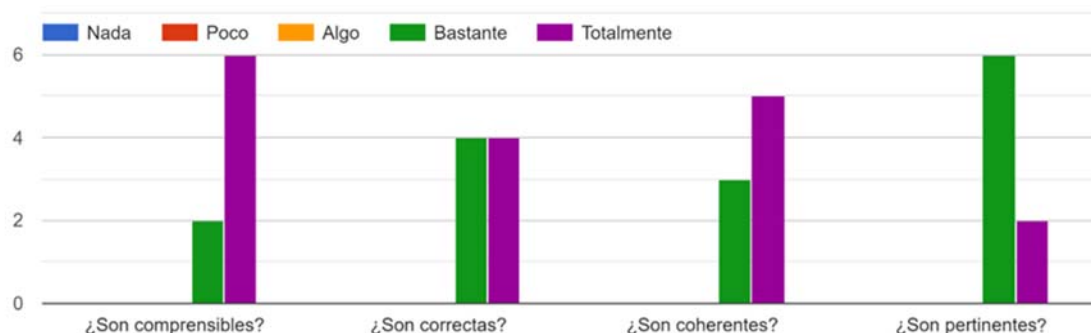
- Evaluación de la originalidad, viabilidad y el impacto potencial de la propuesta.
- Presentación de la propuesta ante la clase o en un foro público simulado.

Estas tareas combinan el aprendizaje teórico con la experiencia práctica, desarrollando habilidades de análisis, investigación, mediación y comunicación, además de promover una conciencia crítica y ética sobre los desafíos globales relacionados con el conflicto y el desarrollo sostenible. ¿Qué te parecen estas propuestas?

Las valoraciones realizadas a las propuestas de tareas de aprendizaje y evaluación por la inteligencia artificial han sido muy positivas. De hecho, todas las puntuaciones en los distintos criterios están entre el 4 (bastante) y el 5 (totalmente) como se puede apreciar en la Figura 3. De forma específica, se considera que son comprensibles con una media de 4,8 ($\sigma=0,7$), que son correctas con una media de 4,5 ($\sigma=0,5$), que son coherentes con una media de 4,6 ($\sigma=0,5$) y que son pertinentes con una media de 4,3 ($\sigma=0,4$). Como se puede comprobar, son estas tareas de aprendizaje y evaluación las que más utilidad se considera que pueden tener dentro del proceso de diseño de la evaluación de las asignaturas universitarias con las que se han realizado las pruebas.

Figura 3

Valoraciones sobre las tareas de aprendizaje y evaluación propuestas por ChatGPT



Estas puntuaciones positivas sobre las tareas generadas por ChatGPT también se pueden apreciar en las reflexiones y consideraciones de los participantes en el proyecto de innovación y que a continuación se ejemplifican:

“La propuesta de tareas de aprendizaje y evaluación son las que aprecio con mejor construcción. Se tratan de tareas tipo, que suelen plantearse en el desarrollo de la asignatura basadas en un papel activo del alumnado. Podrían extrapolarse a cualquier asignatura, orientándolas al contenido de la misma.” (P04)

“Ha aportado 10 tareas, las cuales varias son utilizadas de manera real en la asignatura. No solo hace una descripción de lo que se trataría, sino que también aporta de manera automática posibles criterios de evaluación. Además, propone estrategias de evaluación específicas. Es cierto que son ideas generales que habría que detallar y especificar para convertirlas en tareas aprovechables para los estudiantes, pero de lo que sería esperable, hace propuestas realmente interesantes y coherentes.” (P08)

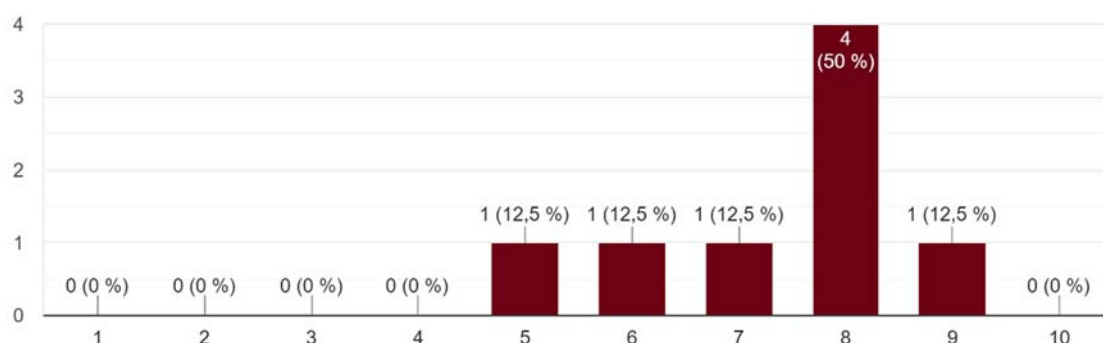
2.4. UTILIDAD DE LAS PROPUESTAS Y ASPECTOS APROVECHABLES

En último lugar, se consultaba a los investigadores hasta qué punto consideraban de utilidad las propuestas realizadas para el diseño de la evaluación por ChatGPT, teniendo de referencia la información real de las asignaturas, y qué aspectos concretos de las ideas generadas por la inteligencia artificial opinaban que podrían ser aprovechables para incorporarlas en un futuro a las materias con las que se han realizado las distintas pruebas.

Sobre la utilidad de las propuestas, utilizando una escala de 1 (nada útiles) a 10 (totalmente útiles), la mitad de las puntuaciones se sitúan en un 8, siendo el mínimo 5 y el máximo 9 como se puede apreciar en la Figura 4. De manera global, la media de estas valoraciones se ha situado en el 7,4 ($\sigma=1,2$), por lo que se puede considerar que se aprecia notablemente el apoyo de ChatGPT para el diseño de resultados de aprendizaje, competencias y tareas de aprendizaje y evaluación en las asignaturas universitarias implicadas en el estudio.

Figura 4

Valoraciones sobre la utilidad de las ideas generadas por ChatGPT para la etapa de diseño de la evaluación



Estas valoraciones se justifican mediante distintos argumentos, entre los que podemos destacar los siguientes:

“Creo que las propuestas sirven de manera general para ayudarnos a reflexionar sobre qué cuestiones queremos trabajar. Eso sí, debemos tener en cuenta que, como comentaba previamente, no se hace hincapié a cuestiones muy esenciales en mi opinión en las ciencias sociales, que tienen que ver en cómo se crea el conocimiento y los valores y cuestiones culturales intrínsecas.” (P02)

“La propuesta planteada por la IA es aplicable a la realidad, incluso presentando algunos resultados de aprendizaje y competencias similares a los ya planteados y otros que no se han tenido en cuenta y podrían incluirse. En relación a las tareas de aprendizaje y evaluación no son iguales a las planteadas, aunque la propuesta también podría ser aplicable, quizá en el aspecto relativo a la evaluación, la información no está suficientemente desarrollada, pero puede ser debido a una falta de solicitud de más información.” (P03)

“Las tareas de aprendizaje y evaluación creo que son bastante útiles y los resultados, aunque no son correctos en su formulación, tampoco lo son los que aparecen en la ficha 1B de la asignatura, me parecen que tienen cierta afinidad con la asignatura y que reconstruidos podrían ser útiles. Sin embargo, las competencias no las veo útiles. Por ello, valoro con un 5.” (P04)

“Son propuestas interesantes, pueden ser un buen punto de inicio para el diseño de la evaluación en las asignaturas. Es cierto que los resultados de aprendizaje sería lo que debería mejorar más en su formulación, pero por su contra, las tareas de aprendizaje propuestas son coherentes y podrían ser realmente útiles.” (P06)

Respecto a qué aspectos concretos podrían ser los más aprovechables de las ideas generadas por la inteligencia artificial, con la intención de su posible incorporación en un futuro a las asignaturas, se ha realizado un análisis textual de las respuestas aportadas por los participantes que se puede encontrar en la Tabla 1. Como se aprecia, todos (100%) consideran que son las ideas relacionadas con las tareas de aprendizaje y evaluación las más aprovechables. Sobre las competencias, son 2 las menciones realizadas (25%), el resto solo se citan una única vez.

Tabla 1

Aspectos aprovechables de las ideas generadas por ChatGPT para la etapa de diseño de evaluación

Aspecto	Frecuencia
Ideas relacionadas con las tareas de aprendizaje y evaluación	8
Ideas relacionadas con las competencias	2
Ideas relacionadas con los resultados de aprendizaje	1
Reducción de las competencias	1
Mayor relación entre competencias, resultados de aprendizajes y tareas	1

Debajo de estas líneas se ofrecen ejemplos sobre las contestaciones de los investigadores a este cuestionamiento sobre los aspectos aprovechables:

“Las ideas que propone sobre Tareas de Aprendizaje y Evaluación son bastante interesantes para llevarlas a cabo dentro del aula.” (P01)

“Algunas competencias, pero, sobre todo, ideas para la realización de tareas de aprendizaje y evaluación, como comentaba, es lo que más ayuda podría servir esta aportación.” (P06).

3. ETAPA DE PREPARACIÓN DE LA EVALUACIÓN

El objetivo propuesto para esta etapa era valorar las aportaciones de la IA en la fase de preparación del proceso de evaluación, concretamente en el establecimiento de criterios de evaluación, creación y diseño de instrumentos, así como la previsión de los agentes evaluadores. Para llevarlo a cabo, se ejecutaron las siguientes acciones:

1. Consultar con la IA la redacción de criterios de evaluación para tareas de aprendizaje concretas en el marco de asignaturas universitarias.
2. Realización con ChatGPT de distintos instrumentos de evaluación que reflejen los criterios establecidos para la evaluación de las tareas.
3. Plantear cómo podría desarrollarse la evaluación en cuestión atendiendo a las distintas modalidades participativas (autoevaluación, evaluación entre iguales, evaluación compartida y evaluación del profesorado).
4. Analizar mediante un cuestionario online la valía, utilidad, coherencia y corrección de las propuestas realizadas. Prestar especial atención a la posibilidad de transferencia de los instrumentos de evaluación al entorno EVALCOMIX-FLOASS instalado en el Campus Virtual de la Universidad de Cádiz.

3.1. REDACCIÓN DE CRITERIOS DE EVALUACIÓN

Los criterios de evaluación para tareas concretas redactados por ChatGPT en las 7 pruebas realizadas por los investigadores respecto a otras tantas asignaturas universitarias, se han valorado utilizando una escala Likert compuesta por los valores nada (1), poco (2), algo (3), bastante (4) y totalmente (5), teniendo en consideración los criterios referidos a su validez, corrección, coherencia y utilidad.

Para comenzar, reproducimos dos ejemplos de criterios de evaluación redactados por esta inteligencia artificial para tareas de aprendizaje y evaluación en el ámbito universitario:

Ejemplo 7

Criterios de evaluación redactados por ChatGPT para la tarea “Diseño de un proyecto de investigación aplicable en Educación Primaria”

Claro! Ajustaré la puntuación para que sea sobre 10 puntos totales. La escala se mantendrá proporcional para que refleje adecuadamente la importancia de cada criterio. Aquí está la versión ajustada:

1. Cumplimiento de la estructura del proyecto (1 punto):

- Resumen claro y conciso.
- Introducción que contextualice el tema y justifique la importancia de la investigación.
- Objetivos bien definidos y alineados con el tema.
- Marco conceptual que demuestre comprensión de la literatura existente.
- Cuestiones de investigación claramente formuladas.

2. Calidad del diseño metodológico (2 puntos):

- Método de investigación adecuado para abordar las cuestiones planteadas.
- Diseño de investigación bien fundamentado.
- Técnicas e instrumentos de recogida de datos bien seleccionados y justificados.
- Muestreo representativo y adecuado.
- Estrategias de análisis de datos coherentes con los objetivos.

3. Rigor académico, Normativa APA y Calidad de las Fuentes (2.5 puntos):

3.1. Normativa APA (1 punto):

- Correcta aplicación de la norma APA en citas y referencias.
- Ortografía y gramática correctas.
- Coherencia en la presentación de tablas y figuras según las pautas APA.

3.2. Calidad de las Fuentes (1.5 puntos):

- Selección de fuentes de información relevantes, confiables y actualizadas.
- Evaluación crítica de la pertinencia y credibilidad de las fuentes utilizadas.
- Integración adecuada de citas para respaldar afirmaciones y argumentos.

4. Claridad y coherencia en la presentación oral (1.5 puntos):

- Resumen oral claro y atractivo.
- Presentación de cada sección de manera organizada.
- Uso adecuado de recursos visuales (si aplica).
- Respuestas convincentes a preguntas durante la defensa.

5. Originalidad y relevancia del proyecto (1.5 puntos):

- Originalidad en la elección del tema y enfoque de investigación.
- Relevancia y aplicabilidad del proyecto en el contexto de la Educación Primaria.

6. Colaboración grupal (1.5 puntos):

- Participación equitativa de todos los miembros del grupo.
- Evidencia de colaboración efectiva en la elaboración del proyecto.
- Coherencia en el estilo de escritura y presentación, indicando una contribución equitativa de todos los miembros.

7. Reflexión crítica (1 punto):

- Capacidad para reflexionar críticamente sobre la propia investigación.
- Consideración de posibles limitaciones y sesgos.
- Discusión de las implicaciones prácticas y teóricas de los resultados previstos.

8. Precisión terminológica y coherencia conceptual (1 punto):

- Uso preciso y consistente de la terminología relacionada con la investigación.

- Coherencia conceptual a lo largo de todo el proyecto.
- Evitar ambigüedades y asegurarse de que los términos utilizados estén claramente definidos.

9. Adecuación y claridad en la redacción (1.5 puntos):

- Redacción clara, precisa y sin ambigüedades.
- Estructura de las oraciones y párrafos que faciliten la comprensión.
- Uso apropiado de conectores para establecer relaciones lógicas entre las ideas.
- Evitar jergas innecesarias y lenguaje técnico excesivamente complejo.

10. Relevancia y actualidad de las fuentes de información (1 punto):

- Selección de fuentes de información relevantes y actualizadas.
- Consideración de investigaciones recientes que respalden o enriquezcan el proyecto.
- Evaluación crítica de la pertinencia de las fuentes utilizadas en relación con los objetivos de investigación.

La suma total de los puntos sigue siendo 10, y la distribución proporciona una evaluación equitativa de los diferentes aspectos del proyecto. ¡Espero que esta versión sea de ayuda! Si tienes más preguntas o necesitas más ajustes, estoy aquí para ayudar.

Ejemplo 8

Criterios de evaluación redactados por ChatGPT para la tarea “Realización de un programa de tutoría para el desarrollo personal y social del alumnado”

Para la tarea en la asignatura Tutoría y Familia en la que el alumnado debe diseñar un programa de tutoría con 8 actividades enfocadas en el desarrollo personal y social, podrías considerar los siguientes criterios de evaluación:

1. Pertinencia y coherencia del programa (25%)

- Relevancia de los objetivos: Que las actividades estén claramente orientadas al desarrollo personal y social del alumnado.
- Coherencia interna: Que las actividades propuestas guarden una relación lógica y progresiva entre sí y con los objetivos del programa.

2. Variedad y creatividad de las actividades (20%)

- Diversificación de estrategias: Que las actividades incluyan diferentes metodologías o enfoques (trabajo en grupo, dinámicas participativas, reflexión personal, juegos, etc.).
- Innovación: Originalidad en las propuestas que motiven y capten el interés del alumnado.

3. Adaptación a las características del alumnado (15%)

- Adecuación a la etapa de Educación Primaria: Que las actividades sean apropiadas en cuanto a la edad y desarrollo cognitivo y emocional de los alumnos.
- Inclusividad: Consideración de la diversidad (necesidades especiales, diferencias culturales, etc.).

4. Desarrollo de competencias clave (15%)

- Competencia emocional: Que las actividades favorezcan la conciencia y gestión de las emociones.
- Competencia social y cívica: Fomento del respeto, la empatía, el trabajo en equipo, y la resolución pacífica de conflictos.

5. Descripción clara y detallada de las actividades (15%)

- Instrucciones precisas: Claridad en la explicación de cada actividad, incluyendo los recursos necesarios, tiempos y pasos a seguir.
- Criterios de evaluación de las actividades: Si incluyen formas de medir el éxito o la participación de los alumnos.

6. Reflexión y justificación pedagógica (10%)

- Fundamentación teórica: Explicación del porqué de las actividades desde una perspectiva pedagógica y cómo contribuyen al desarrollo integral del alumnado.
- Conexión con el marco legal: Relación con la normativa educativa vigente (LOMLOE, etc.).

7. Presentación y formato (5%)

- Claridad y estructura del documento: Que el programa esté bien organizado, con una estructura clara (introducción, actividades, conclusiones, etc.).
- Ortografía y redacción: Corrección gramatical y ortográfica.

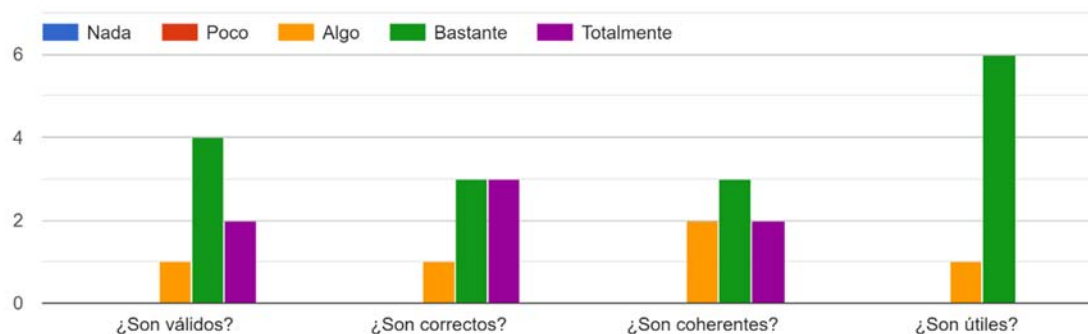
Estos criterios pueden ayudar a los estudiantes a estructurar su trabajo de manera más efectiva y a ti a evaluar el grado de cumplimiento de los objetivos formativos.

¿Te gustaría profundizar en algún aspecto o ajustar alguno de los criterios?

Pasando a describir el análisis de las valoraciones realizadas sobre los criterios de evaluación redactados por ChatGPT por parte del profesorado participante en el proyecto, encontramos, de forma global, puntuaciones altas en los distintos aspectos valorados. Destacaríamos positivamente la corrección de los criterios propuestos ($\bar{X}=4,3$ y $\sigma=0,7$), mientras que el extremo opuesto se colocaría su utilidad ($\bar{X}=3,9$ y $\sigma=0,3$), siendo este último el único aspecto que se valora de media por debajo del 4 (bastante). Justo en esa puntuación se encuentra su coherencia ($\bar{X}=4$ y $\sigma=0,8$), mientras que su validez se sitúa en una media de 4,2 ($\sigma=0,8$). La distribución concreta de las puntuaciones se puede comprobar en la Figura 5.

Figura 5

Valoraciones sobre los criterios de evaluación redactados por ChatGPT



Algunos de los argumentos que justifican las puntuaciones otorgadas por parte de los investigadores, se reproducen literalmente a continuación:

“Se presentan criterios de evaluación que pueden ser de utilidad como punto de partida, aunque hay que tener cuidado con la información aportada, por ejemplo, se plantea como criterio de evaluación "Competencias Generales: Evaluar las competencias generales del estudiante incluyendo la capacidad para recopilar, analizar y sintetizar información de manera efectiva." cuando se entiende que un criterio de evaluación permite verificar en el medio de evaluación en qué grado se alcanzan las competencias a desarrollar, por lo que no tendría sentido incluirlo como un criterio como tal. Por otro lado, no se centra en algunos aspectos esenciales a la hora de valorar el medio de evaluación en cuestión, como podrían ser aspectos formales del trabajo.” (P01)

“El principal problema con los criterios que se proponen en un primer momento es que no se agrupan adecuadamente en mi opinión. Con los diversos intentos he conseguido que se incluyeran aspectos que me parecían importantes y que en las primeras versiones no me proponía y que fusionase algunos que estaban íntimamente relacionados. En mi opinión, los ítems deben tener distintos aspectos y no ser tan similares. El criterio relacionado con la colaboración grupal es una propuesta que me parece difícilmente medible tal y como se propone y me ha generado muchas dudas el mantenimiento de este criterio. También, y no ignoro que, por supuesto hay que tener en cuenta el poco tiempo que llevo usando la herramienta, creo que se tarda mucho en obtener una propuesta definitiva. Eso sí, al final, si estos criterios los copias en un documento a parte, puedes ir reutilizando muchos de ellos en otras tareas.” (P03)

“Los criterios son bastante coherentes, a pesar de tener una descripción limitada de la tarea a realizar. De hecho, muchos de ellos se utilizan realmente. Los criterios se explican y además ofrece ponderaciones e incluso indicadores para su utilización. Podrían ser de bastante utilidad.” (P05)

3.2. REALIZACIÓN DE INSTRUMENTOS DE EVALUACIÓN

Tratando las posibilidades para la realización de instrumentos de evaluación a partir de los criterios propuestos por ChatGPT, se han analizado y valorado siguiendo el mismo procedimiento que en el caso anterior sobre la redacción de los criterios de evaluación. Veamos, en primer lugar, algunos ejemplos de qué tipo de instrumento ha propuesto esta inteligencia artificial. En este caso, para mantener exactamente el formato con el que se ha realizado la propuesta, estos ejemplos se tratan de capturas de pantalla.

Ejemplo 9

Captura de pantalla (parcial) de un instrumento realizado por ChatGPT para la evaluación de “Proyectos en Ciencias Económicas y Empresariales”

Criterio	Nivel 5 (Excelente)	Nivel 4 (Bueno)	Nivel 3 (Aceptable)	Nivel 2 (Insuficiente)	Nivel 1 (Deficiente)
Relevancia del Tema	Enfoque excepcional en un tema altamente relevante y significativo.	Elección sólida de un tema relevante para el campo de estudio.	El tema es adecuado, aunque podría ser más específico o detallado.	La elección del tema es poco clara o no del todo pertinente.	El tema elegido carece de relevancia y significado.
Claridad en los Objetivos	Objetivos claramente definidos y alineados con el propósito del proyecto.	Objetivos bien definidos, aunque podrían ser más específicos.	Objetivos presentes, pero la claridad y especificidad podrían mejorar.	Los objetivos son vagos o no están alineados con el propósito del proyecto.	Objetivos ausentes o inadecuados.
Metodología Aplicada	Selecciona y aplica una metodología de manera excepcional, demostrando un profundo entendimiento.	La metodología es adecuada y aplicada con precisión.	La metodología es adecuada, pero la aplicación podría mejorarse.	La elección de la metodología es cuestionable o mal aplicada.	La metodología seleccionada no es apropiada para el proyecto.
Contribución al Conocimiento	Hace una contribución significativa y original al avance del conocimiento en Ciencias Económicas y Empresariales.	Contribuye de manera sustancial al campo de estudio.	Contribuye de manera adecuada, aunque la originalidad podría mejorar.	La contribución es limitada o poco clara.	La contribución al conocimiento es insignificante o inexistente.
Análisis Crítico	Realiza un análisis crítico profundo y efectivo de la literatura y los conceptos.	Realiza un análisis crítico sólido y aplicado a la información relevante.	Realiza un análisis crítico aceptable, pero podría profundizar más.	El análisis crítico es limitado o no está completamente relacionado con el proyecto.	Falta de análisis crítico o es superficial.

Ejemplo 10

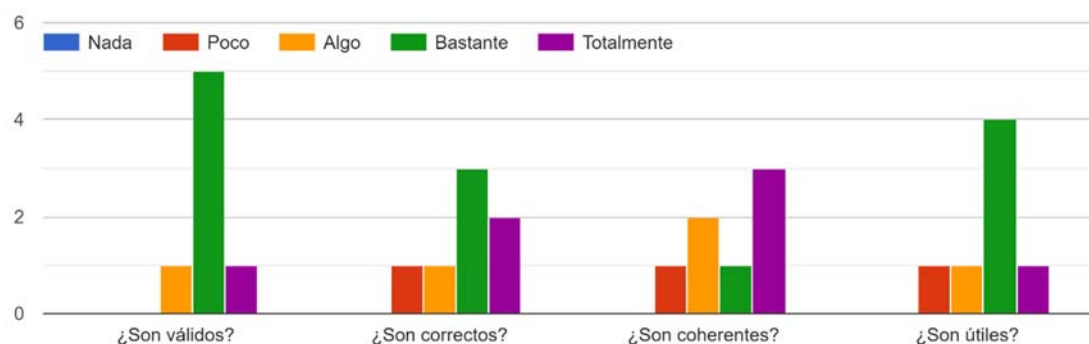
Captura de pantalla de un instrumento realizado por ChatGPT para la evaluación de una “Exposición Oral Individual”

Instrumento de Evaluación: Exposición Oral Individual		
Criterios	Descripción	Valoración
Claridad y organización del contenido (30%)		
Estructura lógica y coherente	La exposición tiene una introducción clara, un desarrollo ordenado y una conclusión bien definida.	1 2 3 4 5
Conexión con el temario	El tema elegido se relaciona directamente con los contenidos de conflicto, desarrollo sostenible y medio ambiente.	1 2 3 4 5
Argumentación fundamentada	Los argumentos presentados están respaldados por fuentes y datos relevantes.	1 2 3 4 5
Dominio del tema (25%)		
Profundidad del análisis	El estudiante demuestra un conocimiento profundo y un análisis crítico del tema.	1 2 3 4 5
Relación con el contexto global	Se establece una conexión entre el tema y el contexto global o los Objetivos de Desarrollo Sostenible.	1 2 3 4 5
Uso de terminología adecuada	Se emplea correctamente el vocabulario específico relacionado con el tema.	1 2 3 4 5
Habilidades de comunicación (20%)		
Claridad en la pronunciación	La pronunciación es clara, con un tono de voz adecuado.	1 2 3 4 5
Ritmo y fluidez	La exposición es fluida, sin pausas excesivas ni interrupciones innecesarias.	1 2 3 4 5
Lenguaje comprensible y persuasivo	El lenguaje empleado es adecuado, accesible y capaz de convencer al oyente.	1 2 3 4 5
Originalidad y creatividad (15%)		
Enfoque novedoso	Se presenta un enfoque original o novedoso del tema elegido.	1 2 3 4 5
Propuesta de soluciones innovadoras	El estudiante propone reflexiones o soluciones creativas relacionadas con el tema.	1 2 3 4 5
Técnica de grabación y formato (10%)		
Calidad de audio y vídeo	La calidad del sonido es clara, sin ruidos de fondo, y el vídeo (si aplica) tiene una buena resolución.	1 2 3 4 5
Cumplimiento de duración y formato	El vídeo/audio cumple con la duración establecida y se entrega en el formato solicitado.	1 2 3 4 5

Como se puede observar en la Figura 6, las valoraciones realizadas sobre las posibilidades de la realización de instrumentos de evaluación por ChatGPT son, en general, más negativas que respecto a la elaboración de los criterios de evaluación. El único aspecto que llega a la media de 4 es la validez ($\sigma=0,7$). La corrección y la coherencia tienen valoraciones medias similares ($\bar{X}=3,9$), aunque con algo más de desviación la coherencia ($\sigma=1,1$) que la corrección ($\sigma=1$). Por debajo se sitúa la utilidad, que solo alcanza una media de 3,7 ($\sigma=0,9$), algo que puede llamar la atención respecto a que se perciben ciertas carencias para que estos instrumentos se puedan aplicar directamente en la realidad educativa.

Figura 6

Valoraciones sobre los instrumentos de evaluación realizados por ChatGPT



Vamos a poner varios comentarios y reflexiones de los docentes implicados que vendrían a justificar sus puntuaciones respecto a los instrumentos de evaluación realizados por la inteligencia artificial:

“Se le ha propuesto la realización de tres instrumentos: rúbrica (bastante clara y coherente, bien estructurada y muy útil), escala de valoración (nada útil, en un primer lugar lo que propone son los valores de la escala sin ningún tipo de ítem a valorar, en un segundo lugar se le propone que la realice de forma más completa y detallada, aportando dimensiones y unidades de logro, y es cierto que mejora, pero lo que propone se parece más a una rúbrica que a una escala) y una lista de control (se ajusta a los distintos aspectos del informe por lo que se considera que es útil para evaluar la tarea propuesta).” (P02)

“Lo más útil, para mí, es la elaboración de rúbricas. Las rúbricas son instrumentos complejos que requieren mucho cuidado en su elaboración para que los descriptores de los niveles estén redactados de forma que recojan todos los cambios que deben darse entre un nivel y otro. Por ello, me parece que esta herramienta nos ayuda muchísimo a elaborar rúbricas que pueda luego directamente copiar a otro documento. En cuanto al resto de instrumentos, creo que, al menos de momento, no es muy útil y que teniendo la propuesta de criterios se tarda menos en diseñar con otras herramientas instrumentos alternativos a la rúbrica. Las cuestiones de los pesos para la conversión de lo cualitativo en cuantitativo y poder hacer el sumatorio de la calificación final, tampoco he conseguido que me plantease propuestas que me convencieran mucho. Eso al final creo que un docente sabe qué criterio opina que puede ser más importante y planteará unos pesos directamente sin necesidad de utilizar la herramienta.” (P03)

Para completar el análisis sobre los instrumentos de evaluación propuestos por ChatGPT, se consultó a los participantes sobre si consideran factible transferir estos instrumentos al entorno EvalCOMIX-FLOASS. Según los argumentos utilizados, en las 7 pruebas realizadas (100%) se considera que sí, que esa transferencia sería posible, pero siempre teniendo en cuenta que se debería realizar un proceso previo de adaptación y revisión, es decir, podría servir de base para la creación de instrumentos en el entorno, aunque no para trasladarlos de manera directa. Algunas de las contestaciones sobre este aspecto se pueden leer a continuación:

“Creo que no hay ningún problema en la transferencia. Al final en el entorno EvalCOMIX-FLOASS los instrumentos se van elaborando y se puede copiar y pegar el listado de los criterios o descriptores de una rúbrica en vez de tener que elaborarlo por tu cuenta directamente en el entorno. Lo que sería muy interesante es poder explorar alguna manera de descargar los instrumentos o listados que genera la IA y poder tenerlos ya en un excel o similar que importar al entorno.” (P03)

“Sí, sería posible, pero en el caso de la rúbrica que ha propuesto, habría que dividir dimensiones, subdimensiones y poner una escala dentro de cada nivel de logro de los criterios, pero teniendo en cuenta la estructura y ponderaciones que ya propone, la mayoría del trabajo ya estaría enfocado.” (P07)

3.3. MODALIDADES DE EVALUACIÓN PROPUESTAS

En última instancia, para considerar los elementos más importantes de la etapa de preparación de la evaluación, se consultó con ChatGPT sobre las modalidades de evaluación o agentes evaluadores que consideraba oportuno para ejecutar las evaluaciones sobre el contexto de las tareas explicadas en cada prueba y con los criterios e instrumentos que había realizado. Se reproducen, a continuación, un par de ejemplos de las contestaciones escritas de esta inteligencia artificial.

Ejemplo 11

Modalidades de evaluación propuestas por ChatGPT para la tarea “Diseño de un programa de actividades de tutoría para el desarrollo personal y social del alumnado”

Dado el tipo de tarea (diseño de un programa de actividades de tutoría), puedes aplicar diferentes modalidades de evaluación dependiendo del enfoque pedagógico que desees implementar. Aquí te explico algunas opciones y recomendaciones para este caso:

1. Evaluación del profesorado

- Descripción: El profesor/a evalúa el trabajo de los estudiantes utilizando el instrumento (rúbrica) que hemos diseñado.
- Recomendación: Esta modalidad es adecuada si deseas centrar la evaluación en el juicio experto del profesor/a, garantizando que los estudiantes reciban una retroalimentación precisa y profesional. Además, es útil si quieres asegurarte de que se cumplan ciertos estándares académicos.
- Ventajas:
 - Retroalimentación basada en criterios pedagógicos.
 - Garantiza objetividad y alineación con los objetivos del curso.
- Desventajas: Puede limitar la implicación del alumnado en su proceso de evaluación.

2. Evaluación entre iguales (Co-evaluación)

- Descripción: Los estudiantes evalúan el trabajo de sus compañeros utilizando la misma rúbrica.
- Recomendación: Esta modalidad es excelente para fomentar el aprendizaje colaborativo, el análisis crítico y la auto-reflexión. Los estudiantes se benefician de observar cómo otros abordan el mismo reto, lo que puede enriquecer su propio aprendizaje.
- Ventajas:
 - Desarrolla habilidades críticas y reflexivas.
 - Fomenta el aprendizaje entre iguales y la colaboración.
- Desventajas:
 - Los estudiantes podrían no sentirse tan seguros evaluando a otros, o puede haber sesgos de amistad.
 - Requiere supervisión por parte del profesor para garantizar equidad.

3. Autoevaluación

- Descripción: Cada estudiante evalúa su propio trabajo basándose en los mismos criterios de la rúbrica.
- Recomendación: Esta modalidad es ideal si deseas que los estudiantes tomen más responsabilidad sobre su proceso de aprendizaje. Los obliga a reflexionar críticamente sobre la calidad de su propio trabajo, lo cual puede llevar a un aprendizaje más profundo.
- Ventajas:
 - Promueve la reflexión crítica y la auto-regulación.
 - Fomenta la autonomía del alumnado.
- Desventajas:
 - Algunos estudiantes podrían tener dificultades para ser objetivos.
 - Puede no ser tan efectiva si los estudiantes no tienen una buena comprensión de los criterios.

4. Evaluación compartida (Profesor/a y estudiante)

- Descripción: El profesor/a y el estudiante evalúan el trabajo de forma conjunta. El estudiante realiza una autoevaluación y luego se compara con la evaluación del profesor/a, generando un diálogo en torno al proceso.
- Recomendación: Esta modalidad es la más enriquecedora si quieres combinar la reflexión crítica del estudiante con el juicio experto del docente. Facilita una retroalimentación más dialogada y personalizada.
- Ventajas:

- Aumenta la implicación del alumnado en su evaluación.
 - Permite al profesor/a dar retroalimentación más precisa y enfocada.
 - Desarrolla habilidades de autoevaluación en los estudiantes, mientras asegura la calidad mediante la evaluación del profesorado.
- Desventajas:
- Requiere más tiempo y coordinación para discutir los resultados.
 - Los estudiantes pueden sentirse incómodos al enfrentarse a posibles discrepancias en las evaluaciones.

Modalidad recomendada:

Dado que la tarea es compleja y fomenta el desarrollo de competencias críticas y creativas, una **evaluación compartida** entre el docente y los estudiantes puede ser la opción más rica y formativa. Esta modalidad equilibra la retroalimentación experta del profesor/a con la reflexión personal del estudiante, permitiendo a este último ser partícipe activo de su propio proceso de aprendizaje.

Otra opción complementaria es la **evaluación entre iguales**, que puede fomentar el aprendizaje colaborativo y dar una perspectiva externa al propio trabajo.

¿Te gustaría implementar alguna de estas modalidades en particular o combinar varias?

Ejemplo 12

Modalidades de evaluación propuestas por ChatGPT para la tarea “Informe de investigación mediante Observación Sistemática”

Por supuesto, aquí te doy una sugerencia sobre cómo podrías llevar a cabo la evaluación según los diferentes agentes evaluadores: autoevaluación, evaluación entre iguales y evaluación del profesorado.

Autoevaluación:

Instrucciones para los Estudiantes:

- Antes de la entrega del informe, proporciona a cada estudiante una copia de la rúbrica detallada.
- Pídeles que revisen la rúbrica cuidadosamente y que se autoevalúen en cada criterio utilizando la escala de valoración proporcionada.
- Anima a los estudiantes a reflexionar sobre sus fortalezas y áreas de mejora en relación con cada dimensión del informe.
- Puedes solicitar que proporcionen comentarios adicionales sobre su contribución individual al trabajo grupal.

Beneficios de la Autoevaluación:

- Fomenta la autorreflexión y el desarrollo de habilidades de evaluación.
- Permite que los estudiantes se responsabilicen de su propio aprendizaje y desempeño.
- Facilita la identificación de áreas de mejora personal y del grupo.

Evaluación entre Iguales:

Instrucciones para los Estudiantes:

- Organiza una sesión donde los miembros del grupo se proporcionan mutuamente retroalimentación basada en la rúbrica.
- Pide a cada estudiante que evalúe el trabajo de sus compañeros utilizando la rúbrica y la escala de valoración.
- Anima a que proporcionen comentarios específicos y constructivos sobre las fortalezas y las áreas de mejora de cada miembro del grupo.
- Promueve la discusión para garantizar una comprensión común de las evaluaciones entre iguales.

Beneficios de la Evaluación entre Iguales:

- Fomenta la comunicación efectiva y la colaboración dentro del grupo.
- Proporciona diferentes perspectivas sobre el desempeño de cada miembro.
- Ayuda a los estudiantes a desarrollar habilidades de retroalimentación constructiva.

Evaluación del Profesorado:

Instrucciones para el Profesorado:

- Utiliza la rúbrica y la escala de valoración como base para evaluar cada informe de manera individual y grupal.
- Realiza una revisión cuidadosa de cada dimensión, proporcionando comentarios específicos y constructivos.
- Considera la coherencia entre las evaluaciones individuales y grupales al realizar la evaluación final.
- Puedes utilizar una lista de verificación adicional para asegurar que todos los elementos y requisitos se cumplan.

Beneficios de la Evaluación del Profesorado:

- Aporta una evaluación objetiva y experta.
- Proporciona orientación específica para el desarrollo académico de los estudiantes.
- Garantiza la coherencia en la aplicación de los criterios de evaluación.

Consideraciones Finales:

- Combina las evaluaciones de autoevaluación, evaluación entre iguales y del profesorado para obtener una evaluación integral.
- Anima a los estudiantes a utilizar las retroalimentaciones recibidas para su crecimiento académico.
- Fomenta la transparencia en el proceso de evaluación, explicando claramente los criterios y cómo se aplicarán.

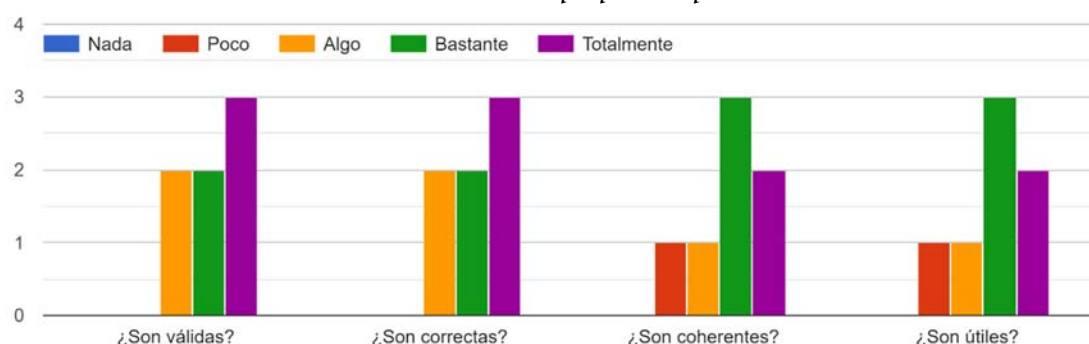
Esta combinación de autoevaluación, evaluación entre iguales y del profesorado permite una evaluación completa y equitativa del informe de investigación. Cada agente evaluador aporta perspectivas valiosas que contribuyen al desarrollo académico y la mejora continua de los estudiantes.

En el caso de las modalidades de evaluación que propone, llama la atención que en prácticamente todos los casos realiza sugerencias que se basa en combinar varias modalidades, tal y como se puede apreciar en los ejemplos anteriores. Esto, que

podría ser interesante si consideramos que ChatGPT realiza estas propuestas para implicar a los estudiantes en los procesos evaluativos, con los beneficios formativos que trae consigo, al ser repetido de forma sistemática en todo tipo de tarea y circunstancia, también puede hacer entender que no termina de discernir en cada caso cuál sería la mejor opción para llevarlo a cabo en la realidad. De hecho, tal y como se puede apreciar en la Figura 6, los investigadores involucrados en este proyecto de innovación, consideran que las modalidades que propone ChatGPT son más válidas y correctas ($\bar{X}=4,1$ y $\sigma=0,8$ en ambos casos) que coherentes y útiles ($\bar{X}=3,8$ y $\sigma=1$ en ambos casos).

Figura 6

Valoraciones sobre las modalidades de evaluación propuestas por ChatGPT



Esta interpretación, se hace evidente en varios comentarios y argumentos que vienen a justificar las puntuaciones otorgadas por parte del profesorado que realizó las distintas pruebas. A continuación, se pueden leer algunas de estas opiniones:

“Lo que proporciona son sugerencias sobre cómo se deberían llevar a cabo las distintas modalidades de evaluación (Autoevaluación, Evaluación entre Iguales, Evaluación Compartida y Evaluación del Profesorado) además de los beneficios que aporta cada una de ellas.” (P02)

“En primer lugar, he de decir que está claro que no es algo que la IA tenga muy trabajado ni que yo haya fomentado. Con esto quiero contextualizar que, por ejemplo, la propuesta que hace de evaluación compartida creo que no es del todo correcta. Además, hace propuestas muy genéricas (proporcionar instrumentos, fomentar la retroalimentación constructiva, etc.). Cuando le he pedido que me propusiera instrumentos concretos para solucionar esta cuestión la propuesta me ha parecido muy floja y sin coherencia metodológica o argumentación sobre la propuesta.” (P03)

“Aquí hace un compendio de posibilidades, pero básicamente que el estudiante haga una autoevaluación, que luego sea evaluado por el profesor y también por los compañeros, es decir, propone que se haga prácticamente todo lo posible.” (P06)

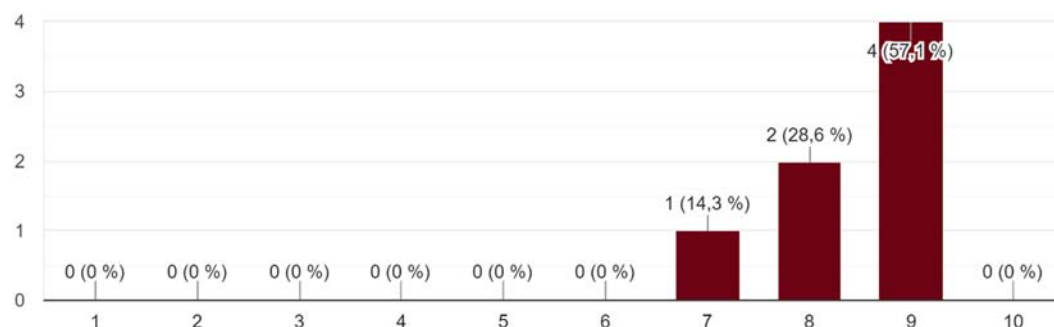
3.4. UTILIDAD DE LAS PROPUESTAS Y ASPECTOS APROVECHABLES

Al igual como se consultó sobre la fase de diseño, en esta fase de preparación se volvió a interrogar a los participantes sobre su percepción respecto a la utilidad de las ideas generadas por ChatGPT para los criterios, instrumentos y modalidades de evaluación, teniendo siempre de referencia las fichas e información real de las asignaturas con las que se hicieron las pruebas. De igual forma, se les consultó sobre los aspectos concretos que podrían aprovecharse para incorporarlos en un futuro en las asignaturas de las propuestas realizadas por la inteligencia artificial.

En primer lugar, sobre la utilidad de las propuestas realizadas por ChatGPT en la etapa de preparación y teniendo en cuenta la escala de 1 (nada útiles) a 10 (totalmente útiles), la media de las valoraciones de las 7 pruebas realizadas en esta fase, se sitúa en un 8,4 ($\sigma=0,7$), es decir, una puntuación bastante elevada y que supera el 7,4 de la valoración sobre la utilidad en la anterior etapa de diseño. En la Figura 7 se puede observar la distribución de las puntuaciones otorgadas por los participantes.

Figura 7

Valoraciones sobre la utilidad de las ideas generadas por ChatGPT para la etapa de preparación de evaluación



De manera textual, a continuación, ofrecemos algunas de las justificaciones con las que los investigadores argumentaban sus puntuaciones sobre la utilidad de ChatGPT en esta etapa de preparación de la evaluación:

“Considero que la propuesta de la IA es bastante acertada. Con respecto a los Criterios de Evaluación, se presentan más claros y específicos que en la realidad, siendo estos último más generales. Las modalidades de evaluación comunes en ambas propuestas son Autoevaluación, Evaluación entre Iguales y Evaluación del Profesorado, en la ficha de la asignatura se nombran estas modalidades y en la IA se explica el proceso y los beneficios de cada una. En la ficha de la asignatura no se especifica qué tipo de instrumentos se van a utilizar en cada una de las modalidades de evaluación por lo que no se puede hacer una comparativa sobre esto.” (P02)

“Anteriormente, ya destacaba que ChatGPT puede ayudar mucho a que "el síndrome del folio en blanco" a la hora de abordar los diseños de los procesos evaluativos se disminuya o, incluso, desaparezca. Pero esto no quiere decir que todas las propuestas que nos dé sean positivas. Es necesario tener un criterio claro sobre lo que se quiere, darle primero muchas vueltas sobre la asignatura y luego reflexionar sobre lo que esta herramienta propone. Creo que la síntesis entre cuestiones propuestas por la IA y lo que se genere por parte de los docentes es el mejor camino. No hay que olvidar que uno de los aspectos más interesantes a la hora de utilizar estas herramientas es ahorrar tiempo y poder ayudar a nuestro cerebro a hacer una "lluvia de ideas" con la máquina. En este sentido, creo que mientras no acabemos perdiendo más tiempo intentando "corregir" a la máquina, que si lo hiciéramos nosotros creo que es muy positivo. Aun así, no sustituye el diálogo y la creación colectiva de los procesos evaluativos.” (P03)

“Son muy interesantes y realistas la mayoría de propuestas, a pesar de no tener un contexto demasiado detallado, podrían servir para mejorar la realización y evaluación de la tarea planteada como ejemplo. Las propuestas pueden ser de utilidad para complementar lo ya existente.” (P07)

Abordando los aspectos específicos que se pueden aprovechar de las ideas generadas por la inteligencia artificial, se puede observar en la Tabla 2, que en el 100% de las pruebas realizadas se aprecia que podría servir de ayuda para la construcción de instrumentos de evaluación. También se destacan las ideas relacionadas con la generación de criterios de evaluación (6 pruebas, 85,7%). Ya en menor medida, también se comentan los beneficios para reflexionar sobre las modalidades de evaluación, para dotar de transparencia el proceso de evaluación o como apoyo para

involucrar a los estudiantes en los procesos evaluativos e incluso, de manera concreta, realizar una evaluación compartida entre docente y estudiantes.

Tabla 2

Aspectos aprovechables de las ideas generadas por ChatGPT para la etapa de preparación de la evaluación

Aspecto	Frecuencia
Ayuda para la construcción de instrumentos de evaluación	7
Ideas relacionadas con los criterios de evaluación	6
Ideas relacionadas con las modalidades de evaluación	2
Utilidad para involucrar a los estudiantes en la evaluación	1
Apoyo para dotar de transparencia el proceso de evaluación	1
Facilita la evaluación compartida entre alumnado y docente	1

Algunos ejemplos sobre las contestaciones y argumentaciones de los docentes implicados en las pruebas, en el contexto de este proyecto de innovación, se ofrecen a continuación:

“Como aspectos concretos la ayuda a incorporar algunos criterios no contemplados de un modo directo en el diseño de la asignatura, como, por ejemplo, criterios relativos a las conclusiones o aspectos éticos. También puede ayudar a la construcción de instrumentos, en este caso en concreto se está trabajando con una escala de valoración, pero podría ayudar a transformar este instrumento en una rúbrica que podría ayudar al alumnado a realizar la autoevaluación y la evaluación entre iguales, además de facilitar los procesos de coevaluación entre el alumnado y el docente” (P01)

“Varios de los ítems creo que son muy útiles. Al final, al haber construido los criterios e instrumentos de cero, me ha ayudado a reflexionar sobre los instrumentos que tenemos actualmente. La organización que se ha realizado de los mismos al final también me gusta bastante. La rúbrica de cinco niveles me parece muy útil, ya que es posible que para el alumnado pueda ser más útil que una escala de valoración como se usa en esta tarea.” (P03)

“Mejorar criterios e instrumento de evaluación. También podría servir para reflexionar sobre la pertinencia sobre las combinaciones de las distintas modalidades evaluativas disponibles.” (P07)

4. ETAPA DE RETROALIMENTACIÓN AL ALUMNADO

El objetivo propuesto para esta etapa era estudiar las funcionalidades que puede aportar ChatGPT en la retroalimentación personalizada y formativa a los estudiantes a partir de la realización de las tareas de aprendizaje y evaluación. Para llevarlo a cabo, se realizaron las siguientes acciones:

1. Pedir la retroalimentación a ChatGPT de varias tareas textuales de aprendizaje, ya sean estas realmente hechas por estudiantes en cursos actuales y anteriores o bien, ejemplificadas por el docente.
2. Analizar y reflexionar sobre la calidad y relevancia de la retroalimentación aportada por la IA. Considerar las posibilidades y limitaciones que podría tener esta herramienta para la realización de retroalimentación automática para los estudiantes universitarios. Esto se reflejará en un documento online que permita el desarrollo de ideas y comentarios por parte de los docentes.

4.1. CALIDAD Y RELEVANCIA DE LA RETROALIMENTACIÓN

El proceso seguido para que ChatGPT aporte retroalimentación, ha sido, en primer lugar, contextualizarle las características de la tarea en cuestión, informar al sistema sobre los criterios de evaluación a partir de los cuales tendría que realizar esta retroalimentación y, en último lugar, escribirle o adjuntarle las tareas concretas realizadas por los estudiantes. Las pruebas se han realizado con 7 tareas de otras tantas asignaturas universitarias. Para analizar la retroalimentación, se ha utilizado una escala Likert compuesta por los valores nada (1), poco (2), algo (3), bastante (4) y totalmente (5), considerando si son de calidad, relevantes y si son formativas.

A continuación, se ofrecen dos ejemplos de la retroalimentación aportada por esta inteligencia artificial en el contexto de las tareas universitarias propuestas:

Ejemplo 13

Retroalimentación aportada por ChatGPT para una tarea sobre un informe comparativo de dos documentos académicos/profesionales en el campo de la economía y empresa.

El trabajo realizado por el grupo (*anonimizado*) presenta varios aspectos positivos y áreas de mejora en línea con los criterios de evaluación, la estructura del trabajo y los requisitos formales establecidos. Aquí tienes algunas observaciones:

Aspectos positivos:

1. Estructura del trabajo: El informe sigue una estructura clara y ordenada, comenzando con una introducción, seguida de un desarrollo comparativo y conclusiones personales, y finalmente las referencias bibliográficas. Esto facilita la comprensión del contenido para el lector.

2. Cumplimiento de los requisitos formales: El trabajo cumple con los requisitos formales establecidos, incluyendo la portada con la información requerida, el uso del tipo de letra Times New Roman (12 pt), el interlineado sencillo, los márgenes de 2,5 cm y la justificación completa del texto. Además, las páginas están numeradas y se incluyen las referencias bibliográficas según la normativa APA 7ª Edición.

3. Análisis comparativo detallado: El desarrollo comparativo del informe es exhaustivo y detallado, abordando cada punto de comparación de manera clara y precisa. Se destacan las diferencias y similitudes entre los documentos analizados, lo que enriquece la comprensión del lector sobre los mismos.

Áreas de mejora:

1. Claridad en la introducción: La introducción podría ser más clara en cuanto a la presentación del propósito del informe comparativo y la relevancia de los documentos seleccionados. Sería beneficioso incluir una declaración de la importancia de comparar estos documentos específicos y qué se espera lograr con el análisis.

2. Coherencia en la redacción: Algunas partes del informe podrían mejorar en términos de coherencia y fluidez en la redacción. Se sugiere revisar la estructura de las oraciones para garantizar una mejor comprensión y conexión entre las ideas presentadas.

3. Detalles adicionales en las conclusiones personales: Las conclusiones personales podrían ampliarse para incluir reflexiones más profundas sobre las implicaciones de las diferencias y similitudes identificadas en los documentos. Además, podría ser útil agregar recomendaciones o sugerencias basadas en el análisis realizado.

En general, el trabajo realizado por el grupo (*anonimizado*) es sólido y demuestra un buen entendimiento de los documentos analizados. Con algunas mejoras en la introducción, la coherencia en la redacción y la profundidad de las conclusiones personales, el informe podría alcanzar un nivel aún más alto de calidad y claridad en la presentación de sus hallazgos.

Ejemplo 14

Retroalimentación aportada por ChatGPT para una tarea sobre una presentación oral de un conflicto medioambiental en la actualidad.

Retroalimentación de la presentación:

1. Pertinencia e importancia del tópico seleccionado (9/10)

El tema seleccionado, los juicios climáticos y su relación con la justicia climática, es extremadamente relevante en el contexto actual de lucha contra el cambio climático. La estudiante hace una conexión clara entre el cambio climático y los derechos humanos, lo cual es

fundamental en los debates actuales de justicia ambiental y social. Además, la inclusión de casos recientes de tribunales internacionales, como el TEDH, aporta un enfoque innovador y actualizado.

2. Originalidad y fluidez de la presentación oral (7/10)

La originalidad del enfoque está bien marcada al centrar el análisis en los juicios climáticos, un tema que no es ampliamente cubierto en los medios tradicionales. Sin embargo, el guion puede beneficiarse de una mayor fluidez y cohesión entre las secciones. El uso de citas y referencias externas es apropiado, pero en algunos momentos corta el ritmo narrativo. Mejorar las transiciones entre las ideas puede hacer que la presentación sea más dinámica.

3. Interés suscitado con el tema presentado (8/10)

El tema de la justicia climática, vinculado a los derechos humanos y a los juicios en tribunales internacionales, es por sí solo un tema atractivo y de gran interés para la audiencia. La estudiante consigue captar la atención, especialmente al incluir casos recientes y específicos que añaden urgencia al tema. No obstante, para mantener un mayor interés, podría haber introducido algunos ejemplos más visuales o comparativos que ayudaran a hacer el contenido aún más accesible.

4. Claridad y calidad del contenido de la presentación (7/10)

La presentación contiene un contenido bien fundamentado y basado en fuentes sólidas. Sin embargo, la estructura general puede mejorarse. Algunos apartados son un poco densos y podrían desglosarse en ideas más claras y concisas. A veces, la cantidad de información legal y técnica puede abrumar al oyente, por lo que sería útil simplificar ciertos conceptos o explicar de manera más didáctica términos clave, como la relación entre cambio climático y violaciones de derechos humanos.

5. Relación entre las ideas trabajadas (6/10)

Aunque la estudiante aborda un tema complejo con profundidad, la relación entre las diferentes ideas no siempre está claramente delineada. Algunos segmentos parecen introducirse abruptamente, como el apartado sobre el IPCC o los detalles sobre los casos legales, que podrían beneficiarse de mejores introducciones o transiciones. También sería útil una mayor conexión entre las ideas expuestas y la conclusión, destacando cómo cada ejemplo fortalece su argumento central.

6. Profundidad de análisis en los argumentos aportados (8/10)

La estudiante demuestra una comprensión profunda del tema y de los desafíos legales asociados con el cambio climático. Se aprecia un análisis crítico en relación con los casos judiciales y su potencial impacto en las políticas climáticas globales. La mención del Acuerdo de París y su impacto en las leyes nacionales es un ejemplo de la capacidad de la estudiante para conectar distintos niveles de gobernanza y derecho internacional. Sin embargo, podría haberse profundizado más en el impacto de estos juicios climáticos en otros estados o en el papel de las ONGs en estos litigios.

Resumen de retroalimentación:

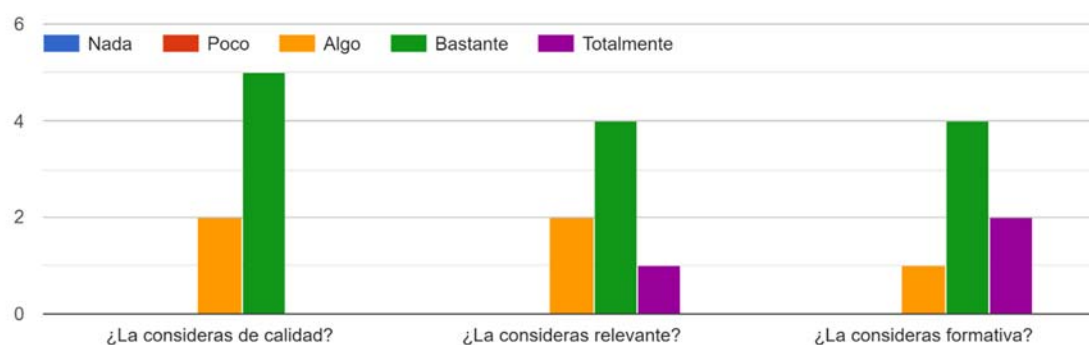
La estudiante ha logrado elegir un tema muy pertinente y actual, mostrando una buena capacidad para integrar casos recientes con la teoría de la justicia climática. Aunque el contenido es sólido y bien fundamentado, sería beneficioso que trabajara en mejorar la claridad y fluidez de la

presentación, especialmente en la manera de conectar ideas y de presentar conceptos complejos de una manera más accesible para el público.

En el caso de la retroalimentación aportada por ChatGPT, el criterio mejor valorado y el único cuya media supera el 4, es que se considera como formativa ($\bar{X}=4,1$ y $\sigma=0,6$). En menor medida, tras el análisis realizado por los investigadores implicados en el proyecto de innovación, se percibe como relevante con una media en las puntuaciones de 3,9 ($\sigma=0,6$) y en último lugar que es una información que posee calidad ($\bar{X}=3,7$ y $\sigma=0,5$). Un resumen gráfico de las puntuaciones otorgadas en las 7 pruebas realizadas, se puede observar en la Figura 8.

Figura 8

Valoraciones sobre la retroalimentación aportada por ChatGPT



Las puntuaciones asignadas a cada criterio se han justificado con una serie de argumentos y comentarios por parte de los docentes participantes, reproduciéndose en las siguientes líneas varias de estas apreciaciones:

“Detecto en sus respuestas:

-Que las citas de fuentes en el texto no las valora en el criterio de normas APA. Por lo que no ofrece retroalimentación.

-Qué existen contradicciones, por ejemplo, en un criterio dice "Esto demuestra una comprensión sólida de cómo los diferentes aspectos están interconectados en el impulso hacia la innovación educativa" y en otro "Sería útil explorar con más detalle cómo estos aspectos interactúan entre sí y cómo contribuyen al cambio cultural en la escuela hacia una cultura de innovación y mejora continua."

-Repite la misma argumentación para varios trabajos: "Sería beneficioso profundizar en cómo estos aspectos interactúan entre sí y cómo

contribuyen al cambio cultural en la escuela hacia una cultura de innovación y mejora continua."

-Incluye aspectos que no aparecen en el fragmento que se le aporta para que ofrezca feedback. En uno de los fragmentos no se aporta nada de autoría sobre la innovación educativa y sin embargo, el chat dice: "Se cita a autores relevantes y se establecen conexiones claras entre los conceptos discutidos"”. (P02)

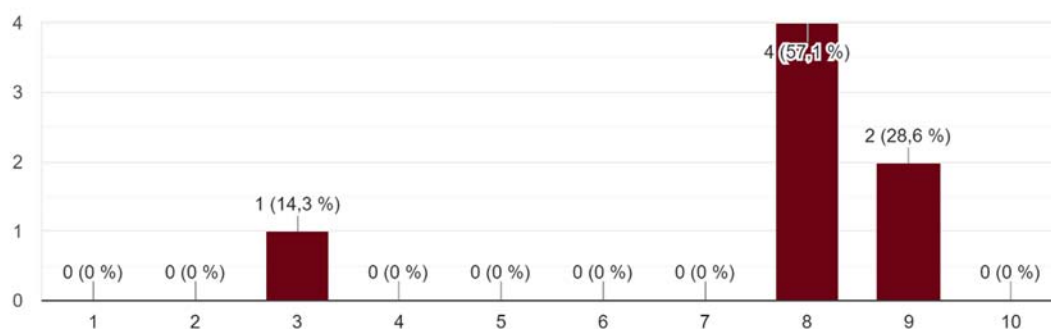
“Creo que mejoraría que aportase fuentes y fundamentación teórica en las afirmaciones que hace. He de decir que me ha sorprendido lo amplia y rica que ha sido la retroalimentación y que utilice ejemplos claros del propio texto. Sobre todo, creo que es muy formativo para iniciarse.” (P04)

“La retroalimentación ha sido relevante respecto a los criterios aportados y se puede considerar de calidad. Es cierto que podría haber especificado más algunas cuestiones, pero seguro que puede concretar si se le insiste en que se centrara en algo concreto. Puede ser de interés para mejorar la propuesta que se le ha puesto como ejemplo. En este caso, ha puesto cada criterio, ha asignado una calificación sobre 10 y ha justificado dicha calificación con lo bueno y malo en cada caso.” (P06)

De forma complementaria, se consultaba de manera específica a los investigadores que realizaron el análisis, en qué medida creían que ChatGPT podía servir para aportar retroalimentación personalizada y formativa de forma automática directamente a los estudiantes respecto a los trabajos que estos realizan. Utilizando una escala de 1 (la retroalimentación no sirve de nada) hasta 10 (sirve totalmente), como se puede observar en la Figura 9, existió una persona que únicamente lo valoró con un 3, mientras que la mayoría (57,1%) lo hizo mediante el valor 8, completándose las puntuaciones con 2 docentes (28,6%) que lo valoraron con un 9. De forma media, estas valoraciones alcanzaron el 7,6, con una desviación de 1,9, por lo que, aunque con limitaciones evidenciadas en los anteriores comentarios expuestos, se aprecia el potencial de la herramienta para ofrecer información que sirva de base para la mejora de las producciones expuestas por los estudiantes de cara a su desarrollo final.

Figura 9

Valoraciones sobre la utilidad para los estudiantes de la retroalimentación automática aportada por ChatGPT



4.2. POSIBILIDADES Y LIMITACIONES DE LA RETROALIMENTACIÓN

Para completar el análisis de la retroalimentación que ofrece ChatGPT a tareas realizadas por estudiantes universitarios, se les planteó a los investigadores que reflexionaran sobre las posibilidades y limitaciones de esta inteligencia artificial en ese cometido.

Comenzando por sus posibilidades, 6 docentes (85,7%) destacan que podría permitir la reflexión y mejora del alumnado a la hora de realizar sus trabajos, fomentando así también su aprendizaje autónomo. En menor medida, también se ha mencionado por parte de 3 participantes (42,9%) que podría servir para que tanto estudiantes como docentes comprendan y se planteen de forma explícita la importancia de los criterios de evaluación a la hora de que se le pueda ofrecer retroalimentación útil y de calidad. Completan estos aspectos positivos, como se puede observar en la Tabla 3, que podría descargar de trabajo a los docentes en esta función de aportar retroalimentación a los estudiantes o que ya se pueden introducir en ChatGPT directamente distintos tipos de archivos (audios, vídeos, escritos) a partir de los cuales ofrecer información para su mejora, aunque en la versión gratuita esta función se encuentra muy limitada.

Tabla 3

Posibilidades de la retroalimentación aportada por ChatGPT

Aspecto	Frecuencia
Permite la reflexión y mejora del alumnado en la realización de sus trabajos, ayudando a su aprendizaje autónomo	6
Que el alumnado y el profesorado comprendan la importancia de los criterios de evaluación y los requisitos de las tareas	3
Descargar de trabajo a los docentes en la aportación de retroalimentación	1
Posibilidad de subir archivos directamente para que dé información sobre su contenido (audios, vídeos, escritos)	1

Algunas de las respuestas escritas que han realizado los docentes participantes que han analizado las posibilidades de la retroalimentación automática de ChatGPT a tareas de estudiantes, se pueden leer en los párrafos que siguen:

“Creo que, si el estudiantado se acostumbra a utilizar la herramienta de manera adecuada, puede ayudar no solo a recibir información de calidad, sino, que además el alumnado conociendo las limitaciones de la herramienta deberá hacer el ejercicio de contrastar las propuestas planteadas. Todo esto permitirá un ejercicio de reflexión y mejora del trabajo antes de ser entregado al profesorado para recibir una nueva retroalimentación. Además, la necesidad de incluir los criterios de evaluación y requisitos también ayudará a que el alumnado dé importancia a dichos criterios y requisitos permitiéndole generar productos de mejor calidad y, por lo tanto, ayudando al aprendizaje de estos.” (P01)

“Creo que esta herramienta puede ayudar enormemente a la hora de generar más comentarios al alumnado sobre sus productos y descargar de muchas tareas al profesorado. El problema que creo que tiene es que, en mi opinión, debe ser supervisado y, cuando sea necesario, editado por el profesorado por lo que el automatismo se ve reducido. Dicho esto, sobre todo para niveles inferiores creo que puede ser una buena herramienta que trabajar también con el propio alumnado para que pueda avanzar en un aprendizaje más autónomo. Siguiendo el trabajo que están haciendo investigadores como Nicols o McCallum sobre "internal/active feedback", creo que podría ser muy interesante seguir explorando ya no solo el que chat GPT haga comentarios, sino la ejemplificación para que

el alumnado pueda hacer comparaciones entre sus productos y otros facilitados por ChatGPT.” (P04)

“Atención porque no he tenido que escribir (copiar y pegar). Ya acepta y analiza archivos de hasta 100 Mb. Por lo que da la retroalimentación directamente desde el PDF. Procesamiento de archivos de audio y vídeo está disponible en GPT-4 (de forma gratuita con uso limitado). Quizá habría que quitar el nombre de los estudiantes para conservar el anonimato. Muy potente esta nueva funcionalidad. Sobre la retroalimentación, puede servir de base para mejorar el trabajo”. (P07)

Respecto a las limitaciones de esta herramienta a la hora de aportar retroalimentación, como se puede observar en la Tabla 4, se destaca que la información que redacta ChatGPT para la mejora de las tareas es en ocasiones demasiado genérica, sin tener en consideración el contexto concreto de su realización o en qué aspectos se debe hacer mayor hincapié (4 referencias; 57,1% de los participantes). Igualmente, el 42,9% de los docentes (3), mencionan que esta inteligencia no realiza comprobaciones sobre la veracidad de las fuentes bibliográficas que se utilizan en los trabajos ni tampoco analiza la posible existencia del plagio académico en ellos. Igualmente, salen otros aspectos mejorables, como las limitaciones existentes en su versión gratuita para realizar comentarios sobre archivos que reflejen tareas no textuales como audios o vídeos (2 referencias; 28,6% de los participantes) o los problemas para ofrecer mejoras respecto al formato de los trabajos presentados y sus elementos gráficos (1 participante).

Tabla 4

Limitaciones de la retroalimentación aportada por ChatGPT

Aspecto	Frecuencia
La retroalimentación es demasiado genérica, sin contextualización	4
No comprueba la veracidad de las fuentes utilizadas ni la posible existencia de plagio académico	3
Limitaciones en la versión gratuita para tareas no textuales	2
Problemas para ofrecer mejoras respecto al formato de los trabajos o a sus elementos gráficos	1

Para concluir con este análisis sobre las limitaciones de la retroalimentación aportada por ChatGPT, reproducimos varios comentarios y argumentaciones al respecto de los docentes implicados:

“Creo que queda mucho por mejorar en aspectos relativos a las fuentes bibliográficas, aunque la herramienta trabaja los aspectos formales de la citación la misma no verifica la veracidad de las fuentes y por lo tanto tampoco puede profundizar en la calidad de las fuentes bibliográficas utilizadas. Otro aspecto limitante es todo lo que tiene que ver con el formato, dado que la manera de incluir la información en la herramienta es de a través de un texto plano, sin formato, la herramienta no puede facilitar retroalimentación relacionada con este tipo de aspectos y cuando se le pide información al respecto se inventa la respuesta. Lo mismo sucede con imágenes creadas o tablas elaboradas.” (P01)

“No detecta plagio, ni las similitudes que existen (solo cambiar dos o tres palabras) del documento que se le ofrece para el análisis del centro, que relata algo de sus prácticas educativas y sus principios teóricos. No conoce lo que has trabajado en tu asignatura y, por tanto, no sabe el enfoque de la misma y los matices que consideran deben aparecer en los trabajos como elementos básicos o las lecturas obligatorias.” (P02)

“Aunque para el ejemplo se ha utilizado un guion escrito de una exposición, hay que tener en cuenta que el enfoque real de la tarea es una exposición oral, por lo que escaparía a poder dar retroalimentación real en actividades no textuales de manera continua, es decir, para dar retroalimentación a un estudiante valdría, pero si son muchos, no habría más remedio que tener Premium (Plus) que tiene un coste adicional. Sobre la retroalimentación, podría ser más concreto.” (P06)

5. ETAPA DE VALORACIÓN DE LAS TAREAS

El objetivo propuesto para esta etapa era comprobar la pertinencia y coherencia de la IA en la cumplimentación de instrumentos de evaluación, la valoración y calificación de tareas de aprendizaje, así como en la realización de informes finales sobre la actividad y rendimiento del alumnado universitario. Para llevarlo a cabo, se realizaron las siguientes acciones:

1. Probar la capacidad de ChatGPT para la cumplimentación de instrumentos de evaluación, la valoración del grado de consecución de criterios y la calificación de tareas de aprendizaje de forma automática en el ámbito universitario.
2. Establecer la adecuación de las valoraciones y calificaciones a las tareas de la IA respecto a las que habrían realizado los docentes. Comparar los resultados para la obtención de posibles diferencias o similitudes estadísticamente significativas.
3. Conocer las propuestas de informes finales personalizados a las tareas evaluadas que podría hacer ChatGPT.
4. Valorar el posible interés de esos informes personalizados automáticos para la formación de los estudiantes mediante un cuestionario online.

5.1. CUMPLIMENTACIÓN DE INSTRUMENTOS, VALORACIONES Y CALIFICACIONES

El proceso seguido para analizar este aspecto, comenzaba con la necesaria contextualización a ChatGPT sobre la asignatura y la tarea concreta a la que posteriormente le pediríamos su valoración. Posteriormente, se informaba a la inteligencia artificial sobre los criterios de evaluación a partir de los cuáles tendría que diseñar un instrumento con el que valorar y calificar varios ejemplos (de 3 a 5) de tareas reales, introduciendo finalmente dichas tareas en el sistema para que ejecutara su evaluación y propusiera una calificación final. Ofrecemos a continuación dos ejemplos sobre dichas valoraciones y calificaciones realizadas por ChatGPT a tareas reales del alumnado universitario.

Ejemplo 15

Valoración y calificación realizadas por ChatGPT para una tarea sobre el enfoque de la innovación educativa en Educación Infantil.

Evaluación del Trabajo utilizando la Rúbrica

A continuación, se evalúa el trabajo presentado por el alumnado en función de los criterios de la rúbrica.

Parte 1: Dibujo, póster o imagen del docente innovador

1. Corrección ortográfica, sintáctica y gramatical. Claridad expositiva de su discurso.

- Nivel: Bueno (5-6)
- Observaciones: Se presentan algunos errores gramaticales y de sintaxis que afectan la claridad del discurso en varias partes del trabajo.

2. Originalidad y creatividad del diseño

- Nivel: Muy Bueno (7-8)
- Observaciones: El diseño del docente innovador es creativo y presenta elementos originales, aunque podría haber sido más innovador.

Parte 2: Justificación y establecimiento de relaciones entre los aspectos

3. Corrección ortográfica, sintáctica y gramatical. Claridad expositiva de su discurso.

- Nivel: Bueno (5-6)
- Observaciones: Nuevamente, algunos errores ortográficos y gramaticales están presentes, pero el discurso es claro en su mayoría.

4. Elaboración de un discurso propio con precisión terminológica

- Nivel: Muy Bueno (7-8)
- Observaciones: El discurso es original y la terminología se utiliza correctamente en su mayoría, aunque con algunas imprecisiones.

5. Fundamentación de ideas y argumentación sustentada en hechos y teorías

- Nivel: Muy Bueno (7-8)
- Observaciones: Las ideas están bien fundamentadas y la argumentación es sólida, sustentada en hechos y teorías relevantes.

6. Establecimiento de relaciones entre los aspectos

- Nivel: Muy Bueno (7-8)
- Observaciones: Se establecen relaciones adecuadas entre los aspectos, con algunas conexiones menos claras.

Parte 3: Reflexión sobre la innovación educativa

7. Corrección ortográfica, sintáctica y gramatical. Claridad expositiva de su discurso.

- Nivel: Bueno (5-6)
- Observaciones: Aunque hay varios errores ortográficos y gramaticales, el discurso es claro y coherente en su mayoría.

8. Claridad y profundidad de la reflexión

- Nivel: Muy Bueno (7-8)
- Observaciones: La reflexión es clara y bien desarrollada, aunque podría profundizar más en algunos aspectos.

9. Conexión con ideas y valores educativos

- Nivel: Muy Bueno (7-8)
- Observaciones: La reflexión conecta adecuadamente con ideas y valores educativos, aunque podría ser más coherente en algunos puntos.

10. Comprensión de la complejidad de la innovación educativa

- Nivel: Muy Bueno (7-8)
- Observaciones: Demuestra una buena comprensión de la complejidad y controversia de la innovación educativa, aunque podría profundizar más.

Comentarios Generales

El trabajo presentado es sólido en términos de contenido y muestra un buen entendimiento de los conceptos de innovación educativa. Sin embargo, la claridad del discurso se ve afectada por errores ortográficos y gramaticales, que deberían corregirse para mejorar la presentación global. La justificación y las relaciones establecidas entre los aspectos son adecuadas, y la reflexión es profunda y bien argumentada, aunque se beneficiaría de una mayor profundidad en algunos puntos.

Calificación Final

Dado que cada criterio tiene un peso similar, se promedia la calificación de cada sección:

Parte 1:

- Corrección ortográfica: 5-6
- Originalidad: 7-8

Parte 2:

- Corrección ortográfica: 5-6
- Discurso propio: 7-8
- Fundamentación: 7-8
- Relaciones: 7-8

Parte 3:

- Corrección ortográfica: 5-6
- Claridad y profundidad: 7-8
- Conexión con valores: 7-8
- Comprensión de la complejidad: 7-8

Promedio: $(5+7+5+7+7+7+5+7+7+7) / 10 = 6.9$

Resultado

Calificación final: 7 (Muy Bueno)

El trabajo es muy bueno en términos de contenido y comprensión de los conceptos clave, aunque necesita mejoras en la claridad del discurso y la corrección ortográfica.

Ejemplo 16

Valoración y calificación realizadas por ChatGPT para una tarea sobre un informe analítico de aplicaciones móviles aplicados a la evaluación.

Aquí tienes la evaluación del informe analítico de (*anonimizado*) sobre las aplicaciones Edpuzzle y Classroom, utilizando la rúbrica que hemos creado.

Evaluación del Informe Analítico

Criterios	Descripción	Puntuación (0-10)
Pertinencia y utilidad	Ambas aplicaciones son relevantes para el ámbito evaluativo, se mencionan características útiles.	9
Corrección ortográfica y gramatical	Hay varios errores menores en la redacción (como "esta a alcance" en lugar de "está al alcance").	7
Adecuación a la estructura del informe	El informe sigue la estructura solicitada, aunque falta un poco de claridad en la separación de secciones.	8
Claridad expositiva y relevancia de los argumentos	Los argumentos son claros, aunque en algunos puntos pueden ser más concisos.	8
Cumplimiento de requisitos formales	Cumple con la mayoría de los requisitos formales establecidos.	8
Profundidad de argumentos en la entrevista	La profundidad de los argumentos es buena, pero podría beneficiarse de ejemplos concretos.	7
Congruencia y consistencia de valoraciones	Las valoraciones son coherentes, aunque se repiten algunas ideas sin necesidad.	8
Precisión terminológica y capacidad de síntesis	Uso adecuado de terminología, pero la síntesis podría mejorar en algunos párrafos.	7

Puntuación Total: 54 / 80

Calificación Final (escala 0-10): 7.75

Comentarios Adicionales:

- Aspectos Positivos: El informe muestra un buen entendimiento de las aplicaciones analizadas y su relevancia en el ámbito educativo. La descripción personal de las aplicaciones es adecuada y se enfatizan aspectos positivos y negativos.
- Mejoras: Se sugiere revisar la redacción para evitar errores gramaticales y mejorar la claridad. Además, proporcionar ejemplos concretos en el análisis podría enriquecer la evaluación.

Si necesitas más ajustes o deseas evaluar otro informe, ¡házmelo saber!

Como se ha podido comprobar en estos ejemplos aportados, ChatGPT plantea diversas estrategias para valorar a partir de instrumentos los trabajos concretos. Puede o no cumplimentar explícitamente dichos instrumentos, hace comentarios por criterios o de manera general, así como presenta distintas formas en el cálculo de la calificación final propuesta. Evidentemente, esta forma de actuar se verá influenciada por las órdenes específicas que se le escriba durante el propio proceso.

Para continuar con el análisis, se consideró de interés comparar las calificaciones que iba generando ChatGPT respecto a las calificaciones reales que fueron otorgadas por los docentes en las tareas objeto de las pruebas. En total, se utilizaron 26 tareas

realizadas por estudiantes universitarios en las distintas asignaturas que participaron en este proyecto de innovación. En la Tabla 5 se ofrece un resumen de las calificaciones establecidas por la inteligencia artificial y las establecidas por los docentes participantes en cada una de las tareas que se utilizaron de pruebas.

Tabla 5

Calificaciones aportadas por ChatGPT y calificaciones aportadas por los docentes en 26 tareas universitarias

Tarea	Calificación de ChatGPT	Calificación docente
1	7,00	6,00
2	9,50	8,80
3	6,00	3,50
4	7,00	6,00
5	8,00	7,00
6	7,75	7,10
7	9,00	8,90
8	7,00	4,50
9	9,50	8,60
10	4,00	8,40
11	7,90	9,20
12	8,50	7,00
13	8,90	8,90
14	8,80	8,80
15	7,00	8,25
16	8,50	9,20
17	9,00	8,40
18	6,70	3,40
19	10,00	4,00
20	8,30	9,30
21	8,10	8,50
22	9,30	7,20
23	7,70	8,10
24	8,00	7,00
25	8,50	8,50
26	7,70	4,30
\bar{x}	7,99	7,26
σ	1,27	1,90

Como se puede apreciar, la media de las puntuaciones calculadas por ChatGPT (7,99) son mayores que las otorgadas por los docentes (7,26), de igual manera, existe una

menor desviación de estas calificaciones (1,27) respecto a las puntuaciones reales (1,90), por lo que se evidencia, en primer lugar, que ChatGPT propone calificaciones, de manera general, más elevadas y homogéneas que las realizadas de por los docentes en las tareas reales utilizadas en estas pruebas.

De forma complementaria, también se realizó con los datos obtenidos una prueba de correlación de Pearson (Tabla 6) y una prueba de rango con signo de Wilcoxon (Tabla 7). Respecto a la prueba de correlación, se observa que la significación bilateral es de 0,136, mayor a 0,05, por lo que se aceptaría la hipótesis nula que llevaría a considerar que no existe una asociación lineal entre las calificaciones aportadas por ChatGPT y las calificaciones reales otorgadas por los docentes. Por su contra, el grado de significación en la prueba de Wilcoxon es de 0,049 que es menor a 0,05, por lo que se rechazaría la hipótesis nula que vendría a corroborar la existencia de diferencias estadísticamente significativas entre ambos tipos de calificaciones en las mismas tareas.

Tabla 6

Resultado de la prueba de correlación de Pearson entre las calificaciones aportadas por ChatGPT y las calificaciones aportadas por los docentes

		Calificación_ChateGPT	Calificación_Real
Calificación_ChateGPT	Correlación de Pearson	1	,300
	Sig. (bilateral)		,136
	N	26	26
Calificación_Real	Correlación de Pearson	,300	1
	Sig. (bilateral)	,136	
	N	26	26

Tabla 7

Resumen de la prueba de rangos con signo de Wilcoxon entre las calificaciones aportadas por ChatGPT y las calificaciones aportadas por los docentes

N total	26
Estadístico de prueba	73,500
Error estándar	32,835
Estadístico de prueba estandarizado	-1,964
Sig. asintótica (prueba bilateral)	,049

Para concluir con este análisis, se les pidió a los investigadores participantes en el proyecto de innovación que realizaran comentarios y reflexiones sobre la cumplimentación de instrumentos, la valoración de la consecución de criterios de evaluación y las calificaciones de tareas realizadas por ChatGPT. Tras el análisis de estos comentarios, solo se puede extraer un único aspecto positivo (Tabla 8) de forma clara, y es que, para 4 docentes (57,1%) es interesante que argumente sus valoraciones y la cumplimentación de los instrumentos, ya sea mostrándolos completos o explicando con qué puntuaciones los completaría. No obstante, sí que se han detectado distintos aspectos negativos a destacar (Tabla 9), como que resulta excesivamente repetitivo y genérico en sus valoraciones y en las calificaciones aportadas, que no parece existir un cálculo homogéneo ni sólido de las calificaciones que propone y las limitaciones existentes en la versión gratuita para que aporte sus valoraciones a partir de la subida directa de archivos, todas ellas con 3 referencias (42,9%). También se ha argumentado negativamente respecto a que no detecta el propio uso de ChatGPT en la redacción de algunas de las tareas con las que se han realizado las pruebas (2 referencias), que no posee un dominio conceptual específico del ámbito de la evaluación (1 referencia) o que da la sensación de no ser demasiado fiable en algunas valoraciones que las realiza sin el fundamento de ninguna evidencia (1 referencia).

Tabla 8

Aspectos positivos sobre la cumplimentación de instrumentos, la valoración de la consecución de criterios de evaluación y las calificaciones de tareas realizadas por ChatGPT

Aspecto	Frecuencia
Argumenta sus valoraciones y la cumplimentación de instrumentos de manera explícita o implícita	4

Tabla 9

Aspectos negativos sobre la cumplimentación de instrumentos, la valoración de la consecución de criterios de evaluación y las calificaciones de tareas realizadas por ChatGPT

Aspecto	Frecuencia
Resulta repetitivo y genérico, tanto en sus valoraciones como en las calificaciones	3
No existe un cálculo homogéneo ni sólido de sus calificaciones	3
Limitación de subir archivos en su versión gratuita	3
No detecta el propio uso de ChatGPT en las tareas escritas	2
No posee un dominio conceptual suficiente en el ámbito evaluativo	1
No parece fiable en algunas de sus valoraciones ya que las realiza sin evidencias	1

Algunos ejemplos de los comentarios y reflexiones realizados por los docentes involucrados en estas pruebas, han sido los siguientes:

“En general, es pasable pues es bastante descriptivo y aporta información sobre la cumplimentación del instrumento desde la valoración de los criterios, por lo que la calificación de la tarea es bastante visible y argumentada con cierta coherencia. Resulta significativo, lo repetitivo que es en sus propios argumentos y calificaciones (todas prácticamente iguales, decimal arriba o abajo) a pesar de que los trabajos presentan diferencias significativas, lo que me hace llegar a la conclusión de que está totalmente descontextualizado y que funciona como lo que es, una máquina.

Resaltar también que, en la primera corrección, realizó promedio de las calificaciones por cada uno de los criterios y en la segunda y tercera corrección no lo ha realizado (he tenido que continuar en días diferentes por problemas de funcionamiento de la propia herramienta) y al realizarlo de forma manual, la calificación real sería de 7,6 y en el informe final le pone un 7. Lo que iría en perjuicio del alumnado.” (P01)

“En primer lugar, he tenido que pedirle un par de veces que modifique el instrumento, es cierto que los criterios de evaluación que le aporté son bastante generales por lo que el instrumento creado tiene unos ítems muy generales. Por este motivo, las calificaciones no se asemejan a las que yo valoré cuando corregía. Por otra parte, una de las tareas que le he pedido

que califique está suspenso, entre otros motivos, porque está hecha con ChatGPT por lo que la IA piensa que está bien realizada.” (P03)

“Establece una puntuación por criterios, pero no muestra la rúbrica cumplimentada como tal. No obstante, aporta puntos fuertes y áreas de mejora, por lo que puede ser interesante. Ha funcionado introduciendo los trabajos en PDF directamente, algo muy interesante, no obstante, está muy limitado (solo 3 trabajos) en la versión gratuita, aunque también funciona pegando el texto. En el caso de subir los trabajos, se bloquea el chat cuando llega al máximo y hay que iniciar otro. Respecto a las calificaciones, llama la atención que siempre está por encima de la calificación real, llamando sobre todo la atención de un trabajo suspenso con 4 que propone que está perfecto... parece que tiende al notable en general.” (P05)

5.2. REALIZACIÓN DE INFORMES FINALES PERSONALIZADOS

La última petición que se le realizó a ChatGPT sobre el proceso de evaluación de las tareas universitarias con las que se hicieron las pruebas fue, tras su calificación, que realizara un informe final personalizado de estas tareas para valorar su posible interés formativo. Debajo de estas líneas se reproducen dos de estos informes finales generados por la inteligencia artificial.

Ejemplo 17

Informe final de una tarea sobre la elaboración de un informe comparativo de distintos tipos de documentos profesionales en Economía y Empresa

<p>Informe Final de Evaluación del Informe Comparativo Estudiante(s): Grupo (<i>anonimizado</i>) Asignatura: Gestión de Proyectos Fecha: [Fecha] Profesor/a: [Nombre del Profesor/a]</p> <p>Introducción En la asignatura de Gestión de Proyectos, se ha solicitado la elaboración de un informe comparativo que analice distintos tipos de documentos profesionales. El informe presentado por el Grupo (<i>anonimizado</i>) ha sido evaluado conforme a la rúbrica de evaluación establecida, abarcando diversos criterios que aseguran una valoración integral y objetiva del trabajo.</p> <p>Evaluación Detallada</p> <p>1. Adecuación a Requisitos y Estructura - Puntuación: 14/15</p>
--

- Comentarios: El informe cumple casi completamente con todas las normas y requisitos establecidos. La estructura general es adecuada, siguiendo el tipo de informe comparativo y el patrón elegido.

2. Portada e Índice

- Puntuación: 8/10

- Comentarios: La portada y el índice están completos y correctos. Sin embargo, hay pequeños detalles de presentación que podrían mejorarse para alcanzar la excelencia.

3. Claridad y Rigor de la Introducción

- Puntuación: 8/10

- Comentarios: La introducción es clara y muestra rigor académico. Explica y justifica adecuadamente el patrón de informe comparativo elegido.

4. Desarrollo Comparativo

- Puntuación: 17/20

- Comentarios: El desarrollo del informe es claro y muestra un buen nivel de rigor. El análisis de las similitudes y diferencias entre los documentos es detallado y bien fundamentado.

5. Adecuación de Tablas y Figuras

- Puntuación: 8/10

- Comentarios: Las tablas y figuras se utilizan adecuadamente y son claras. No obstante, su calidad y presentación podrían ser mejoradas ligeramente.

6. Adecuación de Citas y Referencias Bibliográficas

- Puntuación: 9/10

- Comentarios: Las citas y referencias son correctas y están bien formateadas según la norma APA 7ª Edición. Hay una mínima falta de formato que se puede corregir fácilmente.

7. Claridad y Rigor de las Conclusiones

- Puntuación: 9/10

- Comentarios: Las conclusiones son claras, bien fundamentadas y muestran un alto rigor analítico. Sin embargo, podrían profundizarse un poco más.

8. Coherencia e Integración Argumental

- Puntuación: 14/15

- Comentarios: El informe es coherente y los argumentos están bien integrados. La narrativa fluye de manera lógica y estructurada.

9. Adecuación y Pertinencia de las Citas Bibliográficas y Fuentes Utilizadas

- Puntuación: 8/10

- Comentarios: Las fuentes utilizadas son pertinentes, relevantes y bien integradas en el texto. La selección de fuentes es adecuada y actualizada.

Calificación Final: 9.5/10

Aspectos a Destacar

- Desarrollo comparativo claro y profundo: El análisis detallado de las similitudes y diferencias entre los documentos muestra un buen nivel de comprensión y rigor académico.
- Claridad y coherencia: El informe es coherente y bien estructurado, facilitando la lectura y comprensión de los argumentos presentados.

Aspectos a Mejorar

- Portada e índice: Aunque la portada y el índice están correctos, podrían mejorarse algunos detalles de presentación y formato para alcanzar la excelencia.
- Conclusiones personales: Las conclusiones podrían ser más profundas y detalladas, proporcionando una reflexión más exhaustiva sobre el proceso comparativo y los aprendizajes obtenidos.

Recomendaciones

- Mejorar la presentación visual: Prestar mayor atención a los detalles de presentación en la portada, el índice y las tablas y figuras. Un diseño más pulido puede mejorar la percepción general del informe.
- Profundizar en las conclusiones: Ampliar las conclusiones personales con una reflexión más profunda sobre los aprendizajes y desafíos enfrentados durante el desarrollo del informe comparativo. Esto no solo enriquecerá el informe, sino que también contribuirá al crecimiento académico del grupo.

Este informe final busca contribuir a la formación del estudiante, proporcionando una retroalimentación detallada y constructiva que facilite el desarrollo de habilidades críticas y analíticas en la elaboración de documentos profesionales. Felicitamos al Grupo (*anonimizado*) por su esfuerzo y dedicación, y los animamos a seguir mejorando en sus futuros proyectos académicos.

[Nombre del Profesor/a] [Firma]

Ejemplo 18

Informe final de una tarea sobre la realización de una exposición oral entregada en formato vídeo sobre un problema medioambiental de en la actualidad

Informe formativo: "Video sobre medio ambiente"

Puntuación final: 8.1/10 (Notable)

Aspectos positivos:

- Tema pertinente: El enfoque ambiental es central para la asignatura, y tu presentación abarcó los aspectos importantes con claridad.
- Estructura organizada: El contenido fue presentado de manera lógica y con una buena secuencia de ideas, lo que permite una comprensión fluida.
- Claridad en las explicaciones: Lograste comunicar las ideas de forma comprensible, y se nota un esfuerzo por ser claro y directo.

Áreas de mejora:

- Profundización del análisis: Si bien los puntos clave fueron abordados, el análisis podría haber sido más profundo. Algunos conceptos podrían haberse desarrollado más, sobre todo en términos de ejemplos o datos que respalden los argumentos.
- Originalidad y enfoque: Aunque el tema es relevante, podría haberse explorado de manera más original o desde una perspectiva nueva. Agregar tu propio enfoque crítico puede hacer que el contenido se destaque más.
- Fluidez en la exposición: A veces, el discurso se sintió un poco pausado o ensayado. Más práctica puede ayudarte a lograr una exposición más fluida y natural.

Recomendaciones:

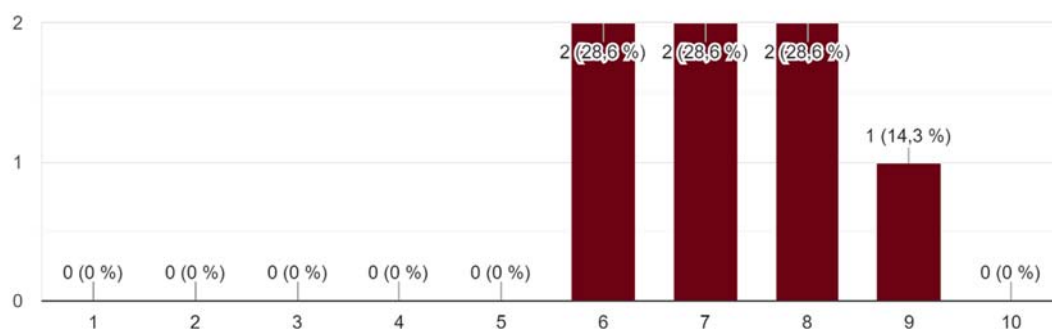
- Profundiza más en los argumentos presentados, ofreciendo análisis más críticos o ejemplos concretos que sustenten mejor los puntos de vista.
- Piensa en cómo darle un toque más personal o creativo a la exposición, lo que te ayudará a distinguir tu trabajo.
- Trabaja en mejorar la fluidez de la presentación para que el discurso sea más natural y continuo.

Con este informe formativo, el estudiante tendrá una idea clara de lo que ha hecho bien y de cómo puede mejorar en futuras presentaciones. Si tienes más ajustes que te gustaría realizar, no dudes en comentarlo. ¡Espero que le sea útil!

Estos informes finales fueron valorados en una escala de 1 (no tienen ningún interés) a 10 (tienen un gran interés) en las pruebas realizadas con distintas tareas de aprendizaje y evaluación contextualizadas en 7 asignaturas universitarias respecto a si se consideraban de interés para la formación de los estudiantes. La media de las puntuaciones otorgadas al interés formativo de los informes finales generados por ChatGPT se situó en 7,3 ($\sigma=1$), por lo que, de manera similar a como ocurría con la aportación de retroalimentación, se aprecia que estos informes puedan ofrecer información justificativa de las valoraciones de las tareas realizadas, a la vez que sirvan de base para mejorar otras tareas académicas similares. La distribución de las puntuaciones aportadas por los docentes se puede observar en la Figura 10.

Figura 10

Valoraciones sobre el interés formativo para los estudiantes de los informes finales a las tareas generados por ChatGPT



Estas valoraciones sobre los informes finales fueron justificadas por los investigadores del proyecto de innovación con distintos comentarios y reflexiones, entre los que ofrecemos algunos ejemplos en las siguientes líneas:

“La información que puede facilitar la herramienta puede ser de interés, aunque hay que tener mucho cuidado y deben contrastarse las valoraciones que esta emite con otras valoraciones que sean emitidas por el/la docente y otros estudiantes.” (P02)

“Creo que pueden aportar mucho, ya que cualquier información que se dé al alumnado que ayude a identificar sus errores y a mejorar su proceso de aprendizaje es positivo. El problema es que el instrumento debe estar bien hecho y que incluya unas dimensiones coherentes. En este caso comparar tareas ya evaluadas con otro instrumento (similar, pero no igual) se ha notado mucho.” (P04)

“Hace un análisis descriptivo de la tarea, realiza comentarios por los criterios de evaluación propuestos y también aporta recomendaciones. En general, se puede considerar de interés, pero realmente no termina de profundizar demasiado en sus valoraciones y sería recomendable que sea más específico. Más o menos, en la línea de lo comentado sobre la retroalimentación.” (P05)

6. DEBILIDADES, AMENAZAS, FORTALEZAS Y OPORTUNIDADES

Para finalizar, expondremos los principales resultados tras realizar una técnica DAFO (Debilidades, Amenazas, Fortalezas y Oportunidades) por parte del equipo de trabajo del proyecto de innovación, donde se ha valorado de forma conjunta los beneficios y limitaciones detectados de ChatGPT como herramienta docente en las distintas fases del proceso de evaluación.

6.1. DEBILIDADES

1. Debe utilizarse con cautela, dado que cuando se trata de contextos complejos y muy específicos, la retroalimentación o las valoraciones que realiza pueden ser incompletas o ser inexactas.
2. A la hora de valorar ciertos trabajos, aunque presenta argumentaciones que parecen ser correctas a nivel gramatical, a veces arroja información incorrecta, por ejemplo, un trabajo facilitado en el que no se le da la portada del mismo y la herramienta valora dicho apartado del trabajo.
3. El uso intensivo de la herramienta puede generar dependencia tecnológica, lo que puede producir un descenso en las capacidades, del alumnado y profesorado, que se ponen en juego a la hora de abordar los procesos evaluativos.
4. La herramienta tiene dificultades a la hora de valorar tareas que implican sustancialmente la creatividad del alumnado.
5. Requiere dedicar mucho tiempo en la aportación del contexto y de información suficiente para el entrenamiento de la herramienta para poder obtener resultados satisfactorios. De otra forma, será imposible que aprecie distintos matices o enfoques.
6. Existen limitaciones severas en la versión gratuita de esta inteligencia artificial para subir archivos de texto, de audio o de vídeo para su análisis. Que todo sea contenido escrito tiene muchas limitaciones para la evaluación educativa.
7. A la hora de realizar correcciones presenta limitaciones importantes como no detectar el plagio académico o la propia utilización de ChatGPT sin haberse citado correctamente.
8. Muchas propuestas son globales o generales siguiendo un prototipo, que en realidad pueden extrapolarse a cualquier asignatura, orientándolas al contenido de la misma.

9. No especifica las fuentes de las que se nutre, para hacer su propuesta, por lo que carece una fundamentación teórica sólida necesaria en la mayoría de trabajos académicos universitarios.

6.2. AMENAZAS

1. Cuestiones éticas cuando utilizamos creaciones y productos que no son propios (derechos de autoría, protección de datos, etc.).
2. Al igual que con otros inicios de tecnología que han impactado en los procesos tradicionales de enseñanza y aprendizaje, el alumnado y, especialmente, el profesorado puede manifestar resistencia al cambio.
3. Una mala utilización de la herramienta en el diseño y aplicación de la evaluación puede dejar en manos de los algoritmos, que pueden ser más o menos precisos, ciertas decisiones cruciales para los procesos de enseñanza y aprendizaje, además de las valoraciones en la consecución de las competencias desarrolladas en el alumnado.
4. No todo el alumnado puede tener el mismo acceso a la herramienta o, al menos, a sus versiones más avanzadas de pago, algo que puede incrementar la brecha digital.
5. Puede generar que el profesorado que no tiene formación específica en evaluación, no considere necesario formarse en este ámbito o en fomentar la reflexión sobre el proceso y acabe siendo algo mecanicista.
6. La deshumanización de un proceso vivo e interactivo como debe ser el evaluativo por una automatización con criterios que no conocemos detrás de sus respuestas. Por ejemplo, las retroalimentaciones son muy impersonales, el profesorado, normalmente, conoce a su alumnado y el *feedback* lo hace en relación a las características de este.
7. Si no realiza de una manera coherente y válida los instrumentos de evaluación, lo que aporta puede no tener suficiente interés formativo para los estudiantes.
8. Caer en el aislamiento y no hacer trabajo en equipos porque puedes contrastar tu visión con la que ofrece la herramienta sin necesidad de interactuar con otras personas.
9. Realizar un proceso de evaluación muy estandarizado y descontextualizado de la realidad del propio alumnado e incluso de la materia. No centrado en los procesos si no en los productos.

6.3. FORTALEZAS

1. La automatización de ciertas tareas evaluativas puede permitir a los docentes una descarga de trabajo para disponer de más tiempo para la enseñanza e interacción directa con el estudiantado.
2. El uso adecuado de la herramienta por parte del alumnado, tras un corto periodo de formación para ello, puede permitir una retroalimentación inmediata y accesible, que se encuentra disponible en cualquier momento y en cualquier lugar.
3. Los procesos de apoyo en el diseño y aplicación de la evaluación pueden adaptarse a cualquier ámbito de conocimiento y disciplina, siendo muchas de las propuestas que facilita, de interés y aprovechables.
4. Ayuda al profesorado a reflexionar sobre los componentes de la evaluación o a considerar distintas opciones que quizá puedan mejorar sus propuestas evaluativas e incluso su enseñanza.
5. Permite hacer lluvia de ideas sin tener que realizar reuniones con compañeros, lo cual puede ayudar a que el profesorado se sienta más acompañado y seguro en sus decisiones.
6. Intenta justificar todas las respuestas que aporta y suele indicar posibles modificaciones para la mejora de los distintos trabajos, sirviendo, al menos, como punto inicial para el aprendizaje autónomo del alumnado.

6.4. OPORTUNIDADES

1. Puede ser una herramienta de gran ayuda para el alumnado como apoyo para la realización de autoevaluaciones y en los procesos de autorregulación del aprendizaje, aunque es importante que los mismos estén formados en su uso, para utilizar esta inteligencia artificial una vez realizada una reflexión previa y cuando se utilice, se haga con cautela conociendo las fortalezas y limitaciones de la misma.
2. Al incorporar la herramienta en la realidad educativa, tanto el profesorado como el alumnado pueden desarrollar sus habilidades digitales.
3. Su correcta utilización puede permitir el fomento de la evaluación formativa, facilitando el acceso a una fuente de retroalimentación automática, aunque esta debe ser tratada con cautela por parte del alumnado.
4. Puede fomentar otros modelos más participativos de evaluación, como la evaluación entre iguales o la autoevaluación, ya que son prácticas que suele proponer.

5. Normalización en el uso de instrumentos de evaluación, tratando de acabar con uno de los mayores argumentos para no realizarlos: el poco tiempo disponible y el esfuerzo que conllevan.
6. Ahorro de tiempo para algunas tareas, sobre todo las que tienen que ver con dar ideas y ejemplos que puedas utilizar de manera real en las asignaturas universitarias. De esta forma, se dispone de un punto de partida, sin que sea necesario empezar directamente desde 0. También puede ofrecer ayuda ante una duda o situación concreta a resolver con respecto al proceso de evaluación.

7. CONCLUSIONES

Para finalizar con este informe final, realizaremos un repaso de las principales conclusiones que podemos extraer tras los análisis realizados en este proyecto de innovación durante las distintas etapas de la evaluación que hemos considerado sobre la utilización de la inteligencia artificial, concretamente de ChatGPT, como herramienta docente para mejorar y orientar al aprendizaje los procesos de evaluación en el ámbito universitario.

Etapas de diseño de la evaluación:

- Los resultados de aprendizaje que propone ChatGPT son considerados bastante comprensibles, pero solo algo correctos. Se destaca que estos se pueden considerar demasiado genéricos y superficiales. También, que podrían estar más relacionados a cómo se redactan las competencias que realmente a resultados de aprendizaje.
- Las competencias que propone tienen una mejor valoración que los resultados de aprendizaje. Se vuelve a considerar que estas competencias son bastantes comprensibles, pero mejor formuladas y más útiles que los anteriormente mencionados resultados de aprendizaje.
- Las tareas de aprendizaje y evaluación propuestas por la inteligencia artificial han sido valoradas de forma muy positiva. Son el componente al que más utilidad se le ha otorgado dentro del proceso de diseño de la evaluación.
- En general, se considera de notable utilidad las ideas generadas en esta etapa de diseño de la evaluación. Se destaca, con mucha diferencia, las aportaciones respecto a las tareas de aprendizaje y evaluación.

Etapas de preparación de la evaluación:

- Los criterios de evaluación que redacta ChatGPT son valorados de manera positiva, destacando su corrección y validez.
- Los instrumentos de evaluación, aunque no se aprecia que sean del todo mal contruidos, parecen tener menores posibilidades de utilizarse en la realidad, ya que su utilidad percibida es uno de los aspectos peores valorados. Se destaca, no obstante, la ayuda que puede aportar esta herramienta para realizar instrumentos complejos, como las rúbricas.
- Siguiendo con los instrumentos, se ha considerado factible transferirlos al entorno EvalCOMIX-FLOASS, siempre considerando realizar un necesario proceso previo de adaptación y revisión, ya que de manera directa sería prácticamente imposible.

- Sobre las modalidades, en la mayoría de caso se ha detectado que hace sugerencias que se basan en combinar varias o todas las modalidades existentes. Puede ser de interés o simplemente no termina de discernir qué sería mejor para cada caso. Pero sin duda, puede servir para que los docentes reflexionen al respecto de introducir estas modalidades.
- La utilidad de las ideas generadas sobre criterios, instrumentos y modalidades es mayor que en la anterior etapa de diseño. Se destaca la posible ayuda para la construcción de instrumentos de evaluación y las ideas relacionadas con la creación de criterios de evaluación.

Etapa de retroalimentación al alumnado:

- Se aprecian las posibilidades formativas de la retroalimentación aportada por ChatGPT. Concretamente, podría servir para tener una base para los propios estudiantes de cara a mejorar sus producciones finales, fomentando su aprendizaje autónomo. También se destaca que puede servir para plantearse, tanto estudiantes como docentes, la importancia de fijar criterios claros para que la retroalimentación se más útil.
- Como limitaciones, se destaca que la información que ofrece se puede considerar en ocasiones demasiado genérica, sin concretar en qué aspectos profundizar ni contextualizar la retroalimentación generada. También, que no comprueba la veracidad de las fuentes utilizadas ni la posible existencia de plagio académico, así como las limitaciones en su versión gratuita para aportar retroalimentación de tareas no textuales a partir de archivos de vídeo o audio.

Etapa de valoración de las tareas:

- Se ha comprobado, en las 26 pruebas realizadas, que de forma general ChatGPT propone calificaciones más elevadas y homogéneas que las otorgadas de manera real por los docentes. Las pruebas estadísticas realizadas confirman que no existe una relación lineal entre ambas calificaciones (correlación) y además la existencia de diferencias estadísticamente significativas entre las calificaciones aportadas por esta herramienta y las calificaciones reales otorgadas por los docentes.
- Como aspecto positivo de las valoraciones y calificaciones realizadas por ChatGPT, se ha destacado únicamente que argumenta sus valoraciones y la cumplimentación de los instrumentos. Los aspectos negativos han sido más variados, mencionándose especialmente que resulta excesivamente repetitivo y genérico en sus valoraciones y en las calificaciones aportadas, que no

parece existir un cálculo homogéneo ni sólido de las calificaciones que propone, las limitaciones existentes en la versión gratuita para que aporte sus valoraciones a partir de la subida directa de archivos y que no detecta el propio uso de ChatGPT en las tareas escritas.

- Se aprecia que los informes finales generados por ChatGPT tienen interés formativo para justificar las tareas realizadas, pero también para servir de base para mejorar otras tareas académicas similares, aunque sería necesario su revisión, concreción y profundización por parte del profesorado.

the \mathbb{R}^n space. The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

The \mathbb{R}^n space is a vector space over the real numbers, and the \mathbb{R}^n space is a vector space over the real numbers.

Asunto: RODIN: Aporte aprobado y archivado.

De: repositorio.digital@uca.es

Fecha: 30/09/2024, 15:46

Para: miguel.gomez@uca.es

Usted aportó: Informe final del Proyecto “Utilización de la Inteligencia Artificial como herramienta docente para mejorar y orientar al aprendizaje los procesos de evaluación en el ámbito universitario”

A la colección: Proyectos de Innovación y Mejora Docente

Su aporte ha sido aceptado y archivado en RODIN, y le ha sido asignado el siguiente identificador:

<http://hdl.handle.net/10498/33468>

Por favor use este identificador cuando cite su aporte.

¡¡Muchas gracias!!

Repositorio de Objetos de Docencia e Investigación de la Universidad de Cádiz