



Recibido: 19 de diciembre de 2025  
Revisado: 20 de abril de 2026  
Aceptado: 15 de junio de 2026

Dirección de los autores:

<sup>1</sup> Departamento Economía de la Empresa. Facultad de Ciencias de la Economía y Empresa. Universidad Rey Juan Carlos. Paseo de los Artilleros, s/n, 28032 - Vicálvaro – Madrid (España)

<sup>2</sup> Facultad de Derecho, Empresa y Gobierno. Universidad Francisco de Vitoria. Ctra. Pozuelo-Majadahonda km. 1.800. 28223, Pozuelo de Alarcón - Madrid (España)

E-mail / ORCID

[emilio.velasco@urjc.es](mailto:emilio.velasco@urjc.es)



<https://orcid.org/0000-0001-8765-9753>

[sofia.ruiz@ufv.es](mailto:sofia.ruiz@ufv.es)



<https://orcid.org/0000-0002-7325-7777>

## ARTÍCULO / ARTICLE /

# Inteligencia artificial generativa, esfuerzo percibido y pensamiento crítico: evidencias de un estudio experimental con ChatGPT en Educación Superior

## Generative Artificial Intelligence, Perceived Effort, and Critical Thinking: Evidence from an Experimental Study with ChatGPT in Higher Education

Emilio Velasco-Bartolome<sup>1</sup> y Sofía Ruiz-Campo<sup>2</sup>

**Resumen:** La aparición de la Inteligencia Artificial (IA) ha tenido un profundo impacto en diversos sectores, incluida la educación. Aunque la integración de la tecnología en las aulas ha sido una constante durante décadas, tanto el profesorado como el alumnado se enfrentan a nuevos desafíos debido a la irrupción de los sistemas de procesamiento del lenguaje natural (PLN). El profesorado debe afrontar el análisis y la evaluación de la originalidad de los ensayos. Mediante el uso de la IA, el alumnado puede responder preguntas teóricas y resolver ejercicios complejos, pero se enfrenta al desafío de colaborar e interactuar con una herramienta que se basa en instrucciones expresadas en lenguaje natural, pero que también puede proporcionar respuestas completamente incorrectas. Nuestro estudio tiene como objetivo evaluar la eficacia de ChatGPT para resolver un estudio de caso con preguntas cuantitativas. Pretendemos comparar el rendimiento del alumnado que utiliza la herramienta con el de quienes no la emplean, desde la formulación de instrucciones hasta la validación o evaluación de los resultados. Los datos recopilados mediante cuestionarios administrados antes y después de la prueba muestran resultados sistemáticamente mejores entre el alumnado que no utilizó ChatGPT. Sin embargo, las expectativas sobre la exactitud de los resultados y la confianza en los resultados obtenidos son similares en ambos grupos. Proponemos interpretar este fenómeno como un efecto que denominamos «ceguera ante el fraude», que describe una situación en la que el alumnado parece no ser consciente de la probabilidad de obtener resultados inexactos al utilizar una herramienta de PLN basada en IA, lo que se manifiesta en una confianza excesiva en los resultados obtenidos. Este fenómeno puede afectar significativamente a los resultados de aprendizaje y debería contrarrestarse mediante el desarrollo de habilidades de pensamiento crítico en la educación.

**Palabras-clave:** Inteligencia Artificial, ChatGPT, Ilusión de capacidad explicativa, Pensamiento crítico, Educación Superior.

**Abstract:** The emergence of Artificial Intelligence (AI) has had a profound impact on diverse sectors, including education. While the integration of technology into classrooms has been a constant for decades, both educators and students face new challenges due to the advent of NLP systems. Educators deal with analyzing and evaluating the originality of an essay. With the application of AI, students can answer theoretical questions and solve complex exercises, but grapple with the challenge of collaborating and engaging with a tool that relies on natural language instructions but can also provide outright incorrect responses. Our study aims to assess the effectiveness of ChatGPT in solving a case study with quantitative questions. We seek to compare the performance of students who use the tool with those who do not, from formulating instructions to validating or assessing the results. The data collected through pre-, and post-test questionnaires reveal consistently better results in students who did not use ChatGPT. However, the expectation of result accuracy and confidence in the outcomes obtained are similar in both groups. We suggest this to be interpreted as an effect that we call "Blind to Fraud" describing the situation in which students appear unaware of the likelihood of inaccuracy when using an AI-based NLP tool, which shows overconfidence in the results obtained. This can significantly impact learning outcomes and should be counterbalanced through critical thinking skills in education.

**Keywords:** Artificial Intelligence, ChatGPT, Illusion of Explanatory Depth, Critical Thinking, Higher Education.

## 1. Introducción

El acceso generalizado a los modelos de Procesamiento del Lenguaje Natural (PLN) basados en Inteligencia Artificial (IA) ha tenido un impacto significativo en diversos sectores (Odden et al., 2021), que abarcan desde la economía, los negocios y la medicina, hasta la educación. Estos modelos de PLN están revolucionando la enseñanza y el aprendizaje, puesto que facilitan experiencias de aprendizaje personalizadas y asistencia adaptada a las necesidades del estudiante. Históricamente, las universidades han impulsado la innovación pedagógica al incorporar tecnologías destinadas a fortalecer las relaciones entre docentes y estudiantes, así como a dotar a estos últimos de competencias aplicables a la resolución de problemas reales. Esta tendencia incluye el aprendizaje basado en el empleo de tecnología, las redes sociales y la interactividad global (Harasim, 2000; Raja y Nagasubramani, 2018). De este modo, las universidades no sólo se han ocupado de incorporar técnicas de enseñanza innovadoras, sino también de su integración en el proceso de aprendizaje.

Parece claro que los modelos de PLN están aquí para quedarse. Los estudiantes ya están familiarizados con ellos, los utilizan y conocen sus ventajas. La evidencia procedente de un estudio global, en el que participaron más de 23.000 estudiantes de 109 países, confirma que, si bien los alumnos valoran ChatGPT por su capacidad para simplificar información compleja y resumir contenidos, también lo consideran menos fiable como apoyo al aprendizaje en el aula (Ravšelj et al., 2025). La literatura reciente subraya que los sistemas de IA basados en PLN, como ChatGPT, presentan a la vez oportunidades y riesgos pedagógicos significativos para las instituciones de educación superior (Fuchs, 2023). Una revisión sistemática del uso de ChatGPT en múltiples ámbitos identifica las preocupaciones en torno a su precisión y su fiabilidad, las restricciones al pensamiento crítico y a la resolución de problemas, y las cuestiones éticas y de privacidad, como los desafíos más destacados asociados a su uso (Sok y Heng, 2024). Estos hallazgos resultan especialmente relevantes en contextos educativos, donde la dificultad del alumno para detectar o verificar respuestas incorrectas, puede tener consecuencias directas sobre los resultados de su aprendizaje. Si bien las instituciones académicas y el profesorado pueden intentar impedir su uso, lo que exige esfuerzos considerables centrados en la detección del plagio, también pueden aceptar dicho uso y adaptar las metodologías de enseñanza, con el objetivo de mostrar los beneficios de la IA, sin que los estudiantes adopten el papel pasivo de copiar y pegar respuestas. Entre otros retos, el profesor debe preparar y adaptar sus clases y actividades para el uso de la IA, de modo que esta se convierta en un complemento y colaborador y no en un sustituto del trabajo de los estudiantes (y de los propios docentes). Este enfoque puede enriquecer el aprendizaje al implicar activamente al alumno, ya que éste necesita adquirir nuevas competencias, como la capacidad de formular preguntas, objetivos e instrucciones de manera eficaz en lenguaje natural, además de validar o evaluar las respuestas obtenidas.

Entre las alternativas disponibles, ChatGPT puede ser considerado el modelo de PLN más popular. Desarrollado por OpenAI, emplea un extenso conjunto de datos para generar respuestas textuales o gráficas a las consultas, comentarios e indicaciones de los usuarios (Gilson et al., 2023). Considerando esta tendencia en el ámbito universitario hacia el uso de las TIC (Tecnologías de la Información y la Comunicación), el presente estudio analiza empíricamente si el uso de ChatGPT puede, tal como

sugiere la literatura, complementar el proceso de aprendizaje y mejorar el conocimiento de los estudiantes.

El estudio se llevó a cabo mediante una metodología experimental con estudiantes españoles e internacionales. La investigación tiene como objetivo evaluar el rendimiento, el esfuerzo y la confianza de los estudiantes que utilizan ChatGPT para resolver un caso práctico de carácter cuantitativo. Más allá de su capacidad para el procesamiento de datos, se partió del supuesto de que ChatGPT podría ofrecer un valor estratégico para la resolución de la tarea. Entre las presunciones sobre su utilidad, se destacaron las siguientes tareas específicas: la identificación clara de los problemas empresariales; la identificación de los datos clave para resolver un problema; la identificación de las técnicas estadísticas necesarias para resolver un problema a partir de los datos proporcionados; el cálculo y los resultados numéricos obtenidos mediante la aplicación de las técnicas pertinentes; el análisis de la validez de los resultados obtenidos; la relevancia de los datos en relación con el problema; la interpretación contextualizada de los resultados; y la aportación de soluciones adecuadas mediante la aplicación del conocimiento.

Los resultados se contrastan y evalúan teniendo en cuenta las siguientes conjeturas (presentadas en la Tabla 1), basadas en Lo (2023), en nuestra propia experiencia y en la literatura relativa a los resultados de competencias del estudiante en la educación superior (Nikolaos et al., 2019; Liesa-Orus et al., 2020).

**Tabla 1.** Las conjeturas de la presente investigación establecen que «El grupo que utiliza ChatGPT frente al grupo que no lo utiliza...»

| Id. | Conjeturas                                                                                                          |
|-----|---------------------------------------------------------------------------------------------------------------------|
| C.a | Obtiene una mejor calificación en el caso de estudio                                                                |
| C.b | Identifica mejor las necesidades y objetivos de la empresa objeto del caso de estudio.                              |
| C.c | Muestra una mejor precisión en el uso de las técnicas estadísticas apropiadas para las necesidades de conocimiento. |
| C.d | Muestra una menor precisión en el cálculo de los resultados.                                                        |
| C.e | Muestra mayor precisión en el análisis de la validez estadística de los resultados.                                 |
| C.f | Muestra mayor precisión al interpretar los resultados y sugerir soluciones basadas en el análisis.                  |
| C.g | Muestra una mayor tasa de respuesta y consecuentemente responde a más preguntas.                                    |
| C.h | Tiene mayores expectativas de calificación                                                                          |
| C.i | Percibe un menor esfuerzo al resolver el caso de estudio.                                                           |
| C.j | Muestra una menor confianza en los resultados debido a su dificultad de supervisar o verificar los resultados.      |

Partiendo del atractivo que el uso de herramientas digitales representa para los jóvenes (Irfan et al., 2023), el estudio explora, además las percepciones del alumno (C.h a C.j) ante el empleo de la herramienta y las consecuencias de su uso en el aprendizaje.

Los hallazgos del estudio pueden resultar de utilidad para las instituciones educativas y para el profesorado, ya que cuantifican los beneficios derivados del uso de

la IA basada en PLN frente a su no utilización. Asimismo, el análisis de las percepciones del alumnado tras completar una tarea empleando esta tecnología muestran que el uso adecuado de ChatGPT implica también el desarrollo de un conjunto de competencias que deben cultivarse durante la etapa universitaria, en particular dos de ellas: la capacidad de formular preguntas de manera ordenada y orientada a la obtención de un resultado (identificación de los objetivos de la tarea), y la capacidad de interpretar los resultados obtenidos, lo que incluye coherencia, capacidad de resolución de problemas y fiabilidad, entre otros aspectos.

En este contexto, introducimos el concepto de «ceguera ante el fraude» (blindness to fraud) para describir la situación en la que los estudiantes parecen no ser conscientes de la probabilidad de inexactitud en los resultados al emplear una herramienta de IA basada en PLN y, por consiguiente, carecen de conciencia sobre su grado real de comprensión de la materia. Esta investigación se puede reproducir en múltiples asignaturas, con distintas competencias implicadas, además de ofrecer un marco de reflexión sobre cómo diseñar tareas que puedan llevarse a cabo en «colaboración» con la IA.

Por último, los profesores y las instituciones pueden examinar qué aprendizaje complementario resulta necesario (comunicación con la IA, aprendizaje en la formulación de instrucciones o prompting, pensamiento computacional y competencias para verificar los resultados obtenidos) para lograr una integración eficaz de la IA en el aula, es decir, para acelerar y personalizar el aprendizaje del alumnado universitario, evitando en particular el efecto de «ceguera ante el fraude». El artículo continúa con la revisión de la literatura, con el análisis de la metodología, seguido de los resultados y las conclusiones del estudio.

### **1.1. Revisión de la literatura**

El empleo de herramientas digitales en el ámbito educativo se ha ampliado desde la Teoría de la Actividad de Vygotsky y Cole (1978). Según su planteamiento, los individuos existen dentro de un marco sociocultural y configuran constantemente su entorno, interactuando y desarrollando nuevos mecanismos culturales. De acuerdo con Leontiev (1978), la interacción es algo más que una mera reacción, provoca cambios internos en el individuo, que favorecen su propio desarrollo. Asimismo, Westin et al. (2015) también examinaron el papel de las TIC en sus interacciones con estudiantes y docentes, y cómo estas mejoran la interactividad y el aprendizaje.

No obstante, no todos los ejemplos de adopción de tecnologías en el aula ofrecen resultados favorables, ya que los resultados educativos dependen en gran medida de los enfoques pedagógicos del profesorado y de la forma en que la tecnología se integra en los procesos de aprendizaje (Moss et al., 2007). En este contexto, se puede deducir que la incorporación de nuevas tecnologías digitales no reduce el papel del profesor, sino que transforma sus responsabilidades hacia la orientación del alumno en el uso eficaz y crítico de las herramientas tecnológicas (Rawlins y Kehrwald, 2010). El profesor debería, por tanto, adaptar sus metodologías para garantizar que las herramientas basadas en IA maximicen no solo el acceso a los contenidos, sino también el desarrollo de procesos de aprendizaje significativos y de competencias analíticas.

Desde esta perspectiva, el empleo de la IA ofrece enormes posibilidades para la mejora de las prácticas de aprendizaje, enseñanza y apoyo en la educación superior (Smith, 2022). Por ejemplo, los chatbots basados en IA pueden ofrecer apoyo inmediato al alumno, mejorando su experiencia de aprendizaje, su motivación y su grado de implicación (O'Connor y McAndrew, 2021). Sin embargo, el profesor se enfrenta a un doble desafío: por un lado, debe adaptar su metodología a la IA (Firat, 2023), y, por otro, debe desarrollar métodos de evaluación que favorezcan la detección de trabajos cuya autoría pueda atribuirse sustancialmente a la IA. Simultáneamente, el alumno se enfrenta al reto de «colaborar» con la IA para mejorar sus competencias, lo que exige habilidades tanto para formular consultas en lenguaje natural como para evaluar las respuestas obtenidas, en cuanto a su corrección y en cuanto a su relevancia respecto a los objetivos de la indagación.

Desde el lanzamiento de ChatGPT, numerosos estudiosos han examinado sus implicaciones para la educación superior desde múltiples perspectivas, incluidas las oportunidades para la productividad y el aprendizaje, los riesgos para la evaluación y la integridad académica, las preocupaciones éticas y las aplicaciones disciplinares. Entre los numerosos ejemplos de publicaciones sobre esta materia se encuentran Arif et al. (2023); Farrokhnia et al. (2023); Fauzi et al. (2023); Gilson et al. (2023); Mhlanga (2023); Rospigliosi (2023); Rudolph et al. (2023); Thurzo et al. (2023); Tlili et al. (2023). Tajik y Tajik (2023) dan un paso más al categorizar los efectos de ChatGPT, o del despliegue de otros modelos de IA basados en PLN, en tres categorías: los alumnos, los profesores y la propia institución educativa.

Efectivamente, según estudios recientes, el empleo de la IA en la educación superior puede dar lugar a mejoras significativas en diversos ámbitos. Las herramientas educativas basadas en IA se han asociado con un mayor grado de implicación de los alumnos, experiencias de aprendizaje personalizadas y un apoyo académico mejorado (Yang et al., 2020; Tate et al., 2023). En particular, Nisar y Aslam (2023) constataron que ChatGPT puede funcionar como una herramienta eficaz de autoaprendizaje y consulta rápida para estudiantes de farmacología, lo que pone de relieve su potencial para apoyar los procesos de aprendizaje autónomo. De manera similar, Firat (2023) sostiene que ChatGPT puede transformar los entornos de educación abierta y autodirigida al fomentar que los alumnos asuman una mayor responsabilidad sobre sus propios resultados de aprendizaje.

El estudio subraya la necesidad de que el alumno conozca y domine la formulación de preguntas relevantes, la interacción constructiva con la herramienta de PLN y la evaluación de sus respuestas. Ello implica superar los desafíos relacionados con la formulación de las consultas, la síntesis de la información y la validación de los datos, con el fin de resolver con eficacia casos prácticos complejos.

Otra dificultad clave a la que puede enfrentarse el alumno es la posibilidad de depender excesivamente de la tecnología, lo cual puede menoscabar el desarrollo de competencias de pensamiento crítico, tales como la capacidad de evaluar la calidad y la fiabilidad de las fuentes, tomar decisiones fundamentadas y generar ideas nuevas. El alumno puede ser víctima de lo que Rozenblit y Keil (2002) denominan «la ilusión de profundidad explicativa», (Illusion of explanatory depth) por la cual se asume tener una comprensión exhaustiva de una materia a pesar de poseer únicamente un conocimiento superficial. Este efecto se ha constatado de manera más notoria en tareas que exigen conocimiento explicativo y patrones causales complejos, pero no en

otras actividades como seguir un procedimiento o aportar descripciones de hechos. Chromik (2021) presenta empíricamente la realidad de este efecto.

Según Lo (2023), ChatGPT obtiene un rendimiento notablemente bueno en economía, satisfactorio en programación y decepcionante en ámbitos como las matemáticas. Con independencia de las áreas en las que demuestre una mayor competencia, esta variación en el rendimiento resulta interesante para un caso práctico como el planteado en este experimento, en el que el rendimiento puede examinarse en una diversidad de ámbitos, incluidos la interpretación de enunciados, el cálculo y la redacción de propuestas, todos ellos basados en los hallazgos a partir de la resolución del caso práctico.

## 2. Método

Este estudio utiliza una metodología experimental con carácter exploratorio cuyos resultados se analizan mediante estadística inferencial.

### 2.1. El experimento

Un total de 30 estudiantes, pertenecientes a la misma clase, fueron divididos aleatoriamente en dos grupos con el objetivo de que resolvieran un caso de estudio: 1) grupo de control (50% de los estudiantes) que utilizaron el software estadístico (SPSS) estudiado durante la asignatura; y 2) otro grupo que utilizó exclusivamente ChatGPT. Todos los estudiantes habían completado el contenido explicitado en la guía docente de la materia (Market Research II) y tenían la misma experiencia trabajando con casos de estudio similares utilizando SPSS para su resolución.

La tarea del experimento requería que los estudiantes resolvieran un caso de estudio de 13 preguntas con el objetivo de evaluar tres áreas de competencias distintas que debían haber adquirido una vez concluido el curso. Para la rúbrica de competencias, las preguntas se agruparon por categorías, según el esquema que se presenta en la Tabla 2.

**Tabla 2.** El caso de estudio abarcaba tres áreas de competencias y utilizó 13 preguntas.

| Categorías de competencias evaluadas a través del caso de estudio                                                                                                       | Preguntas correspondientes a cada competencia | Grupo de competencias evaluadas mediante las preguntas   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|----------------------------------------------------------|
| Categoría 1: Identificación de problemas en un contexto de empresa, necesidades y objetivos que pueden ser abordados mediante una investigación cuantitativa de mercado | 1                                             | Lectura comprensiva y crítica de un contexto de empresa. |

| Categorías de competencias evaluadas a través del caso de estudio                                                                        |                                                 | Preguntas correspondientes a cada competencia | Grupo de competencias evaluadas mediante las preguntas                        |
|------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|-----------------------------------------------|-------------------------------------------------------------------------------|
| Categoría 2: Identificación y aplicación de técnicas de análisis estadístico para resolver los problemas. Se divide en tres componentes: | 2.1. Técnicas estadísticas elegidas             | 4, 7 y 10                                     | Técnicas estadísticas cálculo y análisis de validez.                          |
|                                                                                                                                          | 2.2. Resultados obtenidos en el análisis        | 2, 5, 8 y 11                                  |                                                                               |
|                                                                                                                                          | 2.3. Evaluación de la validez de los resultados | 6, 9 y 12                                     |                                                                               |
| Categoría 3: Interpretación de los resultados considerando los objetivos y necesidades de la empresa objeto de estudio                   |                                                 | 3 y 13                                        | Selección e interpretación de resultados y soluciones propuestas a la empresa |

## 2.2. Constructo e instrumentos de validación

Para esta investigación exploratoria se empleó un método de muestreo por conveniencia, debido a la accesibilidad de los sujetos dentro del entorno académico. Con el fin de fortalecer la validez interna, se utilizó un diseño de aula dividida (split-classroom) con el objetivo de realizar un ensayo controlado aleatorio (ECA) a pequeña escala.

Los estudiantes participantes fueron asignados aleatoriamente al grupo de ChatGPT o al de SPSS. Este enfoque es metodológicamente más robusto que el uso de grupos intactos, ya que elimina el sesgo de selección. Además, dado que todos pertenecen a la misma clase, comparten una formación previa idéntica, lo que minimiza el impacto de las variables del entorno y permite una evaluación comparativa más precisa de los resultados en un contexto de aula natural (validez ecológica). Este enfoque de muestreo se adaptó del marco de Verenikina (2010) al contexto de una universidad pública de Madrid.

Puesto que esta investigación constituye un experimento exploratorio, nuestro análisis inicial se basa principalmente en la estadística inferencial, la cual incluye pruebas T para grupos independientes y tamaños del efecto (D de Cohen, Delta de Glass) atribuibles al uso de ChatGPT. Ocasionalmente, cuando el tamaño del efecto es pequeño o para aislar variables específicas, se aplican correlaciones.

Junto con el estudio de caso de 13 preguntas, se solicitó a cada estudiante que completara dos cuestionarios (Tablas 3 y 4). El primer cuestionario se cumplimentó antes de acceder al estudio de caso y recoge datos sociodemográficos (preguntas 1-3) y académicos (preguntas 4-11). El segundo cuestionario evalúa las percepciones sobre la tarea (preguntas 12-17) y se distribuyó tras finalizar el estudio de caso, pero antes de conocer los resultados de los ejercicios.



El proceso de desarrollo del cuestionario se realizó bajo la supervisión de uno de los coautores de la investigación, pedagogo especialista en la metodología de estudio de caso, y fue revisado por tres expertos.

**Tabla 3.** Primer cuestionario (antes de la tarea): variables sociodemográficas, académicas y de percepción\*

| Item | Contenido de la pregunta                                                                                                                                                                                                                |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | Edad                                                                                                                                                                                                                                    |
| 2    | País de origen                                                                                                                                                                                                                          |
| 3    | Grado de referencia que estudia                                                                                                                                                                                                         |
| 4    | ¿Cuál es tu percepción sobre tus habilidades actuales utilizando IA?                                                                                                                                                                    |
| 5    | ¿Cuál es tu calificación promedio (media de tus calificaciones anteriores) en tu grado universitario?                                                                                                                                   |
| 6    | ¿Cuál es tu calificación promedio obtenida en las actividades previas de la asignatura Market Research II?                                                                                                                              |
| 7    | De acuerdo con tu estimación, ¿a cuántas sesiones de Market Research II, expresado como porcentaje, has asistido durante el semestre? (Considera 1 como menos del 10%, 2 como el 20%, etc., siendo el máximo 10 que representa el 100%) |
| 8    | ¿Cuánto confías en tu actual conocimiento de Market Research II?                                                                                                                                                                        |
| 9    | ¿Cuánto confías en tu actual conocimiento sobre cómo interpretar las necesidades y objetivos de negocio de una empresa?                                                                                                                 |
| 10   | ¿Cuánto confías en tu actual conocimiento sobre las diferentes técnicas estadísticas que se aplican en la investigación de mercados?                                                                                                    |
| 11   | ¿Cuánto confías en tu actual conocimiento sobre cómo interpretar los resultados de un análisis de datos para proporcionar sugerencias o soluciones de negocio?                                                                          |

\* Escala 1-10

El segundo cuestionario (Tabla 4), incluye la expectativa de la calificación y otros resultados obtenidos después de completar la actividad. Exclusivamente para estudiantes que utilizaron ChatGPT se incorporaron cinco preguntas más, preguntas de la 13 a la 17.

**Tabla 4.** Segundo cuestionario (después de la tarea): expectativa de los resultados obtenidos por los estudiantes \*

| Item | Contenido de la pregunta                                                                                                                                                                      |
|------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 12   | ¿Qué calificación consideras que has obtenido en el caso de estudio?                                                                                                                          |
| 13   | ¿Cómo de fácil ha sido encontrar el modo (lenguaje, estructura, procedimientos, etc.) para cooperar con ChatGPT?                                                                              |
| 14   | En lo que se refiere al esfuerzo requerido para solucionar el caso de estudio... ¿cómo de fácil ha sido?                                                                                      |
| 15   | ¿Cuánta confianza tienes en tus respuestas?                                                                                                                                                   |
| 16   | Imagina que este caso se hubiera hecho en un entorno profesional y que se refiriera a la situación real de un cliente, ¿cuánta confianza tendrías para presentar tus resultados a un cliente? |
| 17   | Si tuvieras que distribuir la calificación entre la IA y tú, ¿cuál sería tu reparto?                                                                                                          |

\* Escala 1-10



El hecho de comparar datos proporcionados por los propios estudiantes facilita el contraste entre los resultados obtenidos mediante los dos enfoques alternativos de recogida de información. Por un lado, se indaga a través de escalas de valoración numérica (con un rango de 1 a 10) correspondientes tanto a las preguntas sobre la percepción de su conocimiento global de la materia, como a su nivel de competencia en cada una de las habilidades críticas implicadas en la resolución del estudio de caso (preguntas 1 a 11). Para este grupo de preguntas, el alfa de Cronbach arroja un resultado de 0.885, lo que indica un nivel excelente de fiabilidad del constructo.

Por otro lado, se utilizaron escalas de tipo Likert para evaluar el grado de acuerdo con una proposición. Este método adecuado para establecer comparaciones cuando las alternativas son difíciles de evaluar porque dependen de las percepciones manifestadas por los individuos. Este tipo de escala se emplea para valorar la confianza de los estudiantes en los resultados obtenidos, en una escala de 1 a 10, donde 1 indica «Ninguna confianza» y 10 «Absoluta confianza» (preguntas 12 a 17). Al aplicar nuevamente el alfa de Cronbach, el resultado es sobresaliente (0,943), demostrando una alta fiabilidad del constructo.

Finalmente, cuando se utiliza el alfa de Cronbach para comprobar el grado de fiabilidad global del constructo, esta vez contemplando la totalidad de las preguntas y escalas empleadas, se obtiene un resultado de 0,800, lo que indica que tanto las escalas como el propio constructo poseen validez estadística.

### 3. Resultados

Una vez comprobada la pertinencia del constructo de medida, nuestro objetivo prioritario es evaluar la comparabilidad de los grupos. No hay una diferencia relevante en la información recogida para las preguntas 1, 2 y 3, puesto que se trata de estudiantes de edades comprendidas entre 20 y 21 años, principalmente europeos y con estudios previos similares.

Con respecto a las preguntas 4 a 11 (Tabla 3), los resultados obtenidos en cada grupo (Tabla 5) también eran muy parecidos. El resultado mostraba un p-valor bajo (<0,05), y, por lo tanto, se considera estadísticamente significativo.

**Tabla 5.** Información recopilada de los dos grupos de estudiantes mediante el cuestionario inicial. Preguntas 4-11 \*

| Preguntas #                                                                      | Usuarios de ChatGPT | NO usuarios de ChatGPT |
|----------------------------------------------------------------------------------|---------------------|------------------------|
| 4: Habilidades utilizando IA                                                     | 7.33                | 7.14                   |
| 5: Promedio de calificaciones                                                    | 7.27                | 7.15                   |
| 6: Calificaciones previas en Market Research II                                  | 6.87                | 7.14                   |
| 7: Asistencia durante el semestre                                                | 7.13                | 7.07                   |
| 8: Confianza en su conocimiento de Market Research II                            | 5.73                | 5.00                   |
| 9: Confianza en su habilidad para identificar objetivos y necesidades de negocio | 6.00                | 5.64                   |
| 10: Confianza en la elección de la técnica estadística apropiada                 | 5.20                | 4.79                   |

| Preguntas #                                                   | Usuarios de ChatGPT | NO usuarios de ChatGPT |
|---------------------------------------------------------------|---------------------|------------------------|
| 11: Confianza en la interpretación de resultados y soluciones | 5.80                | 5.14                   |

\* Escala 1-10

Dado que no hay diferencias apreciables entre los dos grupos de estudiantes, proseguimos con la investigación.

### **3.1. Resultados de las competencias en la materia y resultados en la tarea [Conjeturas C.a) a C.f)]**

Nos adentramos, pues, en el análisis de las conjeturas presentadas en la Tabla 2. El resumen puede encontrarse en la Tabla 6, a la que se ha añadido una columna para mostrar la calificación obtenida en el caso de estudio.

**Tabla 6.** Resultado y coeficiente de variación en el caso de estudio dividido en categorías \* para ambos grupos de estudiantes.

| Categoría de competencias                               | Usuarios de ChatGPT | Coeficiente de variación | NO usuarios de ChatGPT | Coeficiente de variación |
|---------------------------------------------------------|---------------------|--------------------------|------------------------|--------------------------|
| 1. Identificación de necesidades y objetivos de negocio | 1.72                | 0.11                     | 3.85                   | 0.08                     |
| 2.1. Técnicas estadísticas aplicadas                    | 3.75                | 0.05                     | 6.63                   | 0.05                     |
| 2.2. Resultados obtenidos en el análisis                | 0.66                | 0.17                     | 5.90                   | 0.06                     |
| 2.3. Validez de los resultados                          | 5.00                | 0.06                     | 5.64                   | 0.06                     |
| 3. Interpretación de los resultados                     | 2.17                | 0.05                     | 5.95                   | 0.04                     |
| Calificación obtenida en el caso de estudio             | 1.72                | 0.69                     | 4.91                   | 0.61                     |

\* Escala 1-10

### **3.2. Conjetura C.a): Calificación en el caso de estudio**

La conjetura C.a) asume que «el grupo que usa ChatGPT obtiene una mejor calificación en el caso de estudio». Los datos muestran claramente que los usuarios de ChatGPT rinden como promedio peor que los que no lo utilizan, aunque muestran una percepción de habilidad en el uso de herramientas de IA relativamente mayor. No obstante, el promedio en el caso de estudio puede estar construido sobre una gran dispersión de los resultados, por lo que se procedió a estudiar el coeficiente de variación.

El coeficiente de variación es bajo y muy similar en ambos grupos de estudiantes (excepto para la autoevaluación de la habilidad en el cálculo de resultados para los usuarios de ChatGPT: Competencia 2.2). Por lo tanto, el promedio de resultados puede ser considerado sólido para el análisis. Esto nos permite rechazar la conjetura C.a), aunque un análisis más profundo será llevado a cabo posteriormente para determinar la posible causa.

### 3.3. Conjeturas C.b) a C.f): Rendimiento y precisión

#### *El efecto de la calificación previa en la asignatura*

Como la diferencia de rendimiento en el caso de estudio es notable, queríamos dilucidar si había alguna relación entre la calificación previa en la asignatura o la calificación global en el grado (Pregunta 5) y el resultado obtenido en el caso de estudio. La correlación resulta estadísticamente válida ( $p < 0,05$ ) sólo para la relación entre la Calificación previa en la asignatura y la Competencia #1. "Identificación de las necesidades y objetivos de negocio", que muestra una correlación positiva de intensidad moderada (+0,435) y la Competencia #4. "Validez de los resultados" (+0,431). En consecuencia, ambas competencias pueden ser explicadas por la calificación actual en la asignatura para ambos grupos. El resto, muestran p-valores superiores a 0,05 y son, en consecuencia, rechazadas.

Para completar la indagación de estas competencias, se realiza un T-Test de medias comparadas para muestras independientes porque la explicación más plausible, descartada la competencia en la asignatura, es que la variable explicativa sea el uso o no de ChatGPT en la tarea. El resultado obtenido es estadísticamente significativo (sig.  $< 0,05$ ), pero muestra un tamaño de efecto bajo (Tabla 7). Se eligió la D de Cohen para explorarlo, porque las desviaciones estándar eran similares.

**Tabla 7.** D de Cohen para medir el tamaño de efecto en las competencias 2.1, 2.2 y 3.

| Competencias                          | (2.1) Técnicas estadísticas aplicadas | (2.2) Resultados obtenidos en el análisis | (3) Interpretation of results obtained |
|---------------------------------------|---------------------------------------|-------------------------------------------|----------------------------------------|
| Estimación de puntos en la D de Cohen | 0.28                                  | 0.34                                      | 0.17                                   |

Por lo tanto, el uso de ChatGPT no puede ser estadísticamente atribuido a una mayor o menor demostración de rendimiento en las competencias analizadas. De este modo, podemos poner en duda la conclusión de Lo (2023), que sintetiza cualitativamente los resultados de diferentes estudios acerca del distinto rendimiento de ChatGPT en diferentes áreas de conocimiento. Según nuestros datos, la influencia directa del uso de ChatGPT en el rendimiento, para las competencias analizadas separadamente, debe ser rechazada, porque la calificación previa en la materia parece tener una influencia más significativa. Esto nos permite rechazar, bien por la ausencia de correlación, bien por el pequeño tamaño de efecto, nuestras conjeturas C.b) a C.f), excepto C.c), puesto que los usuarios de ChatGPT no obtienen un mejor resultado en el caso de estudio (C.b); no rinden mejor al identificar las necesidades de negocio o sus objetivos (C.d), ni tienen una mayor precisión al analizar la validez estadística de los resultados (C.e), ni, finalmente, son más precisos interpretando y sugiriendo soluciones basadas en los datos obtenidos durante el caso de estudio (C.f). En lo que se refiere a C.c) podemos confirmar la conjetura porque los usuarios de ChatGPT muestran un menor nivel de precisión calculando los resultados.

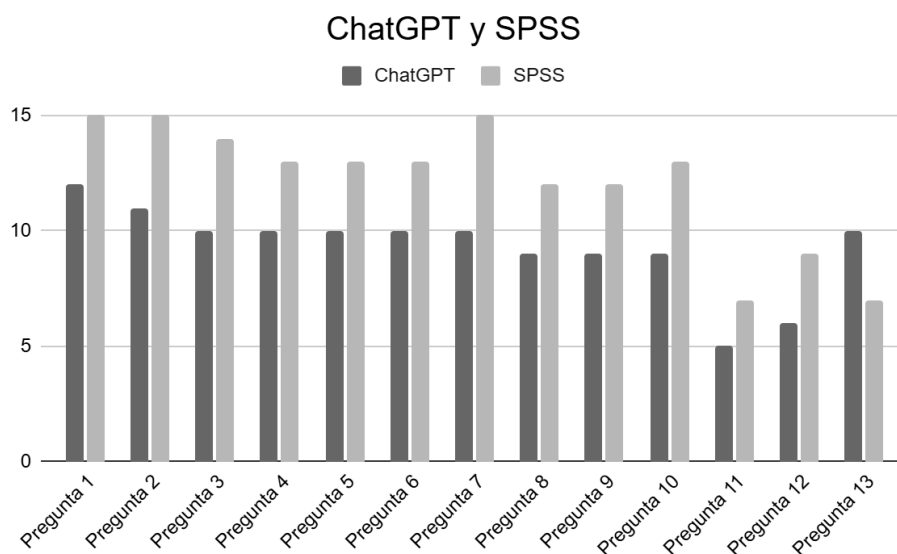
Se puede añadir, en relación con C.a) que al considerar la calificación final obtenida en el caso de estudio para ambos grupos de estudiantes (y no únicamente el

resultado en cada competencia), la diferencia puede ser correctamente explicada por el tamaño de efecto, puesto que, en este caso, la D de Cohen es de 2,24. Dicho de otro modo, la calificación del caso de estudio depende notoriamente del uso o no de ChatGPT para su resolución. En este caso, los estudiantes que utilizaron ChatGPT mostraron un nivel de rendimiento significativamente menor que los que no lo usaron. Así que, el uso de ChatGPT afectó negativamente al proceso de resolución del caso de estudio en su conjunto. Aunque la calificación final depende de las calificaciones parciales en cada competencia, y la significación estadística mostró que los resultados en cada competencia, aunque eran claramente peores, no eran atribuibles al uso o no de ChatGPT, es posible afirmar que el foco no debe ponerse en cada competencia, como concluía Lo (2023) en su estudio de revisión, sino en la tarea principal en la que la contribución de ChatGPT parece decisiva y reduce notablemente el rendimiento de los estudiantes.

En resumen, sin menospreciar las competencias implicadas en la tarea, es mejor orientarse a esclarecer el impacto que tiene el uso de IA en el desarrollo de soluciones para tareas complejas.

### 3.4. Conjetura C.g): Tasa de Respuesta

Otro aspecto considerable para el rendimiento es la tasa de respuesta de ambos grupos de estudiantes. La Figura 1 ofrece un resumen del número de respuestas por pregunta en el caso de estudio.



**Figura 1.** Número de estudiantes que responden cada pregunta del caso de estudio, utilizando ChatGPT o SPSS (no usuarios de ChatGPT)

En todas las competencias (ver Tabla 2), excepto para la Categoría 3: Interpretación de los resultados (pregunta 13), los usuarios de ChatGPT respondieron menos preguntas, lo que contradice nuestra Conjetura C.g): «El grupo que utiliza

ChatGPT tiene una mayor tasa de respuesta y, en consecuencia, es capaz de responder más preguntas».

No obstante, es llamativo que 10 de los 15 estudiantes que utilizaron ChatGPT respondieron a la última pregunta, pero sus resultados fueron significativamente más bajos, (2,17/10), que los de aquellos que no utilizaron ChatGPT, que obtuvieron, en promedio 5,95/10. La pregunta pedía que se desarrollara un buyer persona utilizando la información de los resultados obtenidos durante el caso de estudio. Los estudiantes que utilizaron ChatGPT respondieron, aunque en la mayoría de los casos la respuesta no estaba basada en los resultados del caso de estudio ni en la situación descrita de la empresa. A pesar de todo, los estudiantes utilizaron la respuesta dada por ChatGPT, confiando en una respuesta que les parecía bien estructurada, narrativa, y plausible.

En consecuencia, el uso de ChatGPT no produjo un mayor número de respuesta a las preguntas, y, por lo tanto, no supuso una mayor facilidad para los estudiantes al completar el caso de estudio. Por eso, la conjetura C.g) debe ser rechazada.

### **3.5. Conjetura C.h): Expectativas sobre la calificación**

Los estudiantes no solo fueron calificados tras resolver el caso de estudio, sino que también se les preguntó acerca de la calificación que esperaban haber conseguido en la tarea. Aunque los que utilizaron ChatGPT esperaban una calificación mayor que los que utilizaron SPSS, sus calificaciones reales fueron muy inferiores a las que esperaban. Por su parte, los estudiantes que no utilizaron ChatGPT subestimaron sus calificaciones. La Tabla 8 muestra la disparidad entre la calificación estimada y la real para ambos grupos.

**Tabla 8.** Calificación real y estimada para ambos grupos de estudiantes tras la resolución del caso de estudio

|                        | <b>Real</b> | <b>Estimada</b> |
|------------------------|-------------|-----------------|
| Usuarios de ChatGPT    | 1.72        | 5.00            |
| NO usuarios de ChatGPT | 4.91        | 4.31            |

Al aplicar una prueba T de muestras independientes a la expectativa de calificación obtuvimos un valor de 2.52 en estimación de puntos para la D de Cohen, lo que evidencia que el uso o no de ChatGPT influye claramente en la expectativa de calificación. De este modo, aunque el uso de ChatGPT influye sobre el rendimiento en la tarea, reduciéndolo, afecta de manera notable a la expectativa de calificación, derivando hacia una gran sobrestimación del rendimiento. Esto nos permite confirmar nuestra Conjetura C.h) que suponía que el grupo de usuarios de ChatGPT tenía expectativa de una mayor calificación.

De acuerdo con nuestro análisis de datos, podemos concluir que, si bien el uso de ChatGPT no produjo un mejor rendimiento en la calidad de la resolución de la tarea, ni una mejor calificación, ni una mayor tasa de respuestas, sorprendentemente, incrementó la expectativa de calificación. Las secciones posteriores de este texto exploran los motivos y las implicaciones de esta sobrestimación.

### 3.6. Conjetura C.i): Esfuerzo necesario

Además de la calificación, estábamos interesados en medir el grado de esfuerzo percibido, puesto que suponíamos que el uso de ChatGPT debía reducirlo. Al comparar el esfuerzo percibido en completar la tarea para los usuarios de ChatGPT frente a aquellos que no lo utilizaron, resulta evidente que los usuarios de ChatGPT percibieron un menor esfuerzo (Tabla 9).

**Tabla 9.** Esfuerzo percibido para resolver el caso de estudio en ambos grupos de estudiantes.

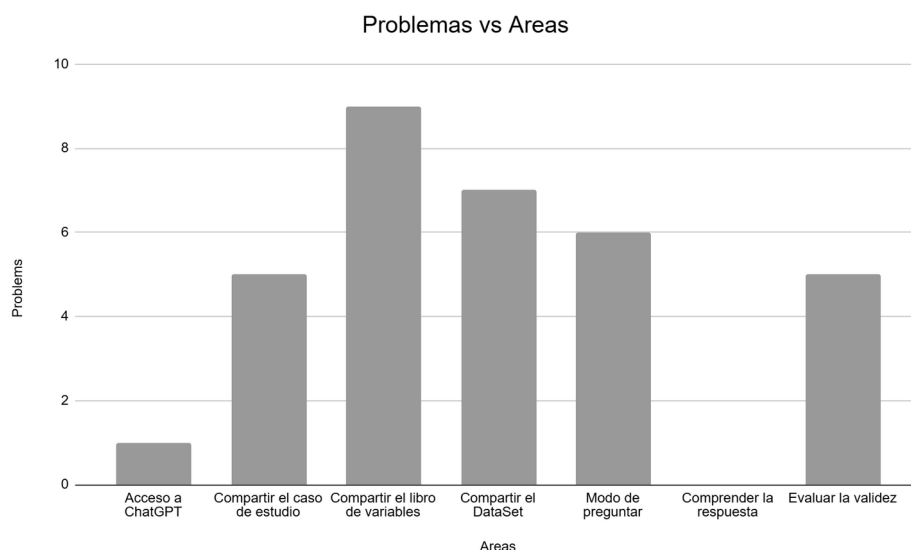
|                           | Usuarios de ChatGPT | NO usuarios de ChatGPT |
|---------------------------|---------------------|------------------------|
| Esfuerzo *                | 5.5                 | 7.15                   |
| Desviación estándar       | 3.23                | 1.67                   |
| Coefficiente de variación | 58.69%              | 23.4%                  |

\* Escala 1-10

Dado el alto coeficiente de variación entre los estudiantes que utilizaron ChatGPT, se hizo un análisis para determinar si el conocimiento previo de la materia o su calificación en el grado pudiera explicar esa variación. El resultado del análisis de correlación entre la calificación en la asignatura, por un lado, y la calificación en el Grado, por otro, con el esfuerzo percibido, no arrojan una validez estadística suficiente. Teniendo en cuenta que la diferencia es grande y que no puede ser explicada por una variable de confusión, decidimos realizar un T-Test de grupos independientes.

El tamaño de efecto, utilizando en este caso Glass Delta, reveló un resultado estadísticamente significativo, con un tamaño de efecto grande (3,228), lo que indica que el uso de ChatGPT tiene un impacto significativo en las diferencias en el esfuerzo percibido. A continuación, se analizó cómo el esfuerzo percibido influía en la expectativa de la calificación obtenida para ambos grupos. En ambos casos los resultados son estadísticamente significativos ( $\text{sig.} < 0,05$ ) y mostraron una correlación de Pearson moderada y negativa (-0,4), es decir, que cuanto menor es el esfuerzo percibido, mayor es la expectativa de calificación. Este hecho debe contribuir al mayor optimismo observado entre los que usaron ChatGPT, puesto que su esfuerzo percibido es notablemente menor.

Un factor que potencialmente podría estar influyendo en los estudiantes que utilizaron ChatGPT, podría ser la dificultad en el uso mismo de la herramienta, más que el esfuerzo percibido en la resolución del caso. Analizamos este aspecto y le pedimos a los estudiantes que evaluaran (con un escala de 0 a 10, siendo 0 extremadamente fácil, y 10 extremadamente difícil) cómo de fácil o difícil era interactuar con ChatGPT. El resultado fue 3,71, lo que indica un nivel de dificultad bajo a moderado. En la figura 2, se puede observar dónde residen las mayores dificultades en una tabla de frecuencias.



**Figura 2.** Frecuencia de los problemas al interactuar con ChatGPT en diferentes áreas.

En lo que se refiere al esfuerzo necesario para completar el caso de estudio, los usuarios de ChatGPT utilizaron una media de 46,9 interacciones con el modelo de LNP, casi una interacción por minuto, con un alto Coeficiente de Variación (0,62).

Un análisis de correlación entre el esfuerzo necesario para resolver el caso de estudio y la facilidad de interacción y cooperación con la herramienta mostró una alta correlación negativa (-0,57), lo que indica que, cuanto más difícil se percibía la interacción más esfuerzo se reportaba. No parece arriesgado asumir, en consecuencia, que gran parte del esfuerzo percibido fue dedicado a interactuar con el modelo de lenguaje. Esto nos permite considerar que el esfuerzo dedicado al caso fue percibido como menor o que el foco del esfuerzo fue más orientado a utilizar la herramienta que al caso de estudio y que los estudiantes aceptaron las respuestas de manera acrítica una vez obtenidas. El hecho de que ningún estudiante reportara dificultades en entender las respuestas confirma con claridad este hecho.

Además, como mostramos previamente, hubo una diferencia importante en el número de interacciones necesarias dentro del grupo de estudiantes que utilizaron ChatGPT. Para analizar esta alta desviación estándar intragrupo, se realizó un análisis de correlación mediante Pearson, tanto con la media del grado universitario, como con la media de las calificaciones en la asignatura. El resultado no muestra una relación significativa estadísticamente con la calificación en el grado, pero sí con la media de calificaciones en la materia, arrojando un resultado de -0,522 (significancia bilateral 0,06). Dicho de otro modo, los estudiantes que rindieron mejor en la materia (pruebas de evaluación previas), necesitaron significativamente menos interacciones con ChatGPT.

#### *Distribución de la carga de trabajo y de la calificación entre los estudiantes y la IA*

Proseguimos con el análisis de la distribución de la aportación a la calificación. Cuando les pedimos que dividieran la nota obtenida entre ellos y ChatGPT según una escala de



0 a 10, donde 0 significaba “La IA lo hizo todo” y 10 significaba que el estudiante hizo todo funcionando la IA como una simple calculadora, la respuesta promedio fue de 3,82, lo que indica que los estudiantes consideraban que contribuyeron ligeramente menos que la IA a la resolución del caso (menos del 50%).

A continuación, se analizan los motivos de esa distribución. Consideramos la facilidad de cooperación con la IA y el resultado fue una intensa correlación negativa de -0,675 (con significancia estadística). Así que, cuanto más difícil era cooperar con la IA, menos tendían los estudiantes a reconocer la contribución de ésta. Esto refuerza la idea de que el esfuerzo percibido por los estudiantes surge primeramente de la dificultad de utilizar la herramienta de LNP, más que del caso de estudio en sí mismo.

Adicionalmente, aparecía una correlación negativa, -0,593, válida estadísticamente, entre su expectativa de calificación y la distribución de la contribución. En otras palabras, cuanto más alta era su expectativa de calificación, menos contribución asumían estos de la IA.

El estudio mostró que:

- a) El esfuerzo percibido depende más de la usabilidad de la herramienta que de la dificultad de la tarea.
- b) El reparto de la contribución está fuertemente influido por la facilidad de cooperar con la IA.
- c) Un valor más alto de rendimiento esperado reduce la atribución de contribución de la IA.

### **3.7. Conjetura C.j): Evidencia de la confianza en los resultados**

Los estudiantes estaban informados de que podían utilizar ChatGPT para cualquier tarea si lo consideraban necesario para resolver el caso de estudio. Sin embargo, incluso para el cálculo de algo tan sencillo como un promedio, el 60% de los usuarios de ChatGPT dieron una respuesta incorrecta, y el otro 40% no respondió a esta pregunta, evitando el uso de algo tan sencillo como una calculadora. Por el contrario, el 93,3% de los estudiantes que no tenían acceso a ChatGPT respondieron correctamente. Los usuarios de ChatGPT confiaron en sus respuestas sin verificarlas, se atribuyeron el mérito de las soluciones, e informaron de que no tenían ninguna dificultad en entender las respuestas.

El estudio indagó, además, el nivel de confianza que ambos grupos de estudiantes tenían en las respuestas que dieron (Tabla 10), tanto en sus respuestas con fines de evaluación académica, como también con fines de uso en un entorno profesional. No se encontraron diferencias entre grupos, aunque como podía suponerse de antemano, ambos mostraron más confianza en la validez académica, donde el grado de responsabilidad es menor, que en la profesional.

**Tabla 10.** Confianza expresada en las respuestas dadas al caso de estudio.

|                        | <b>Confianza en el ámbito académico</b> | <b>Confianza en un entorno profesional</b> |
|------------------------|-----------------------------------------|--------------------------------------------|
| Usuarios de ChatGPT    | 4.21                                    | 3.54                                       |
| NO usuarios de ChatGPT | 4.31                                    | 3.62                                       |

El análisis no permite confirmar la conjetura de que la calificación promedio del grado o de la materia correlaciona con la confianza. Tampoco un análisis de Medias para grupos independientes mediante técnicas T arroja ninguna luz sobre el asunto. En consecuencia, nuestra conjetura inicial C.j): «Los estudiantes que utilizan ChatGPT muestran una menor confianza en los resultados debido a su dificultad para revisarlos o verificarlos», no puede ser confirmada o rechazada.

Como se esperaba, una mayor expectativa de calificación implica una mayor confianza, tanto en los resultados en el ámbito académico (0,893) como en el profesional (0,809), siendo ambas correlaciones significativas desde el punto de vista estadístico.

Estos resultados subrayan dos percepciones:

- a) Precisión en las respuestas: los usuarios de Chat GPT muestran una mayor autoconfianza, a menudo sobrestimando la precisión de sus resultados debido a una ‘ceguera al fraude’, a pesar de la discrepancia con el rendimiento real.
- b) Uso distinto de las respuestas. Esta excesiva confianza tiene muy poco impacto en la utilidad percibida, ya sea en el entorno académico como en el entorno profesional.

Para concluir el análisis de los resultados obtenidos en este estudio, se presenta la Tabla 11, que resume la confirmación o no de las conjeturas.

**Tabla 11.** Resumen de la confirmación o rechazo de las conjeturas y técnicas estadísticas empleadas.

|      | <b>Statistical technique</b>        | <b>Status</b> | <b>Reason</b>                                    |
|------|-------------------------------------|---------------|--------------------------------------------------|
| C.a) | Prueba T para grupos independientes | Rechazada     | D de Cohen >2                                    |
| C.b) | Correlación de Pearson              | Rechazada     | Sig = 0,435. ChatGPT es una variable confusora.  |
| C.c) | Prueba T para grupos independientes | Rechazada     | No atribuible a ChatGPT. D de Cohen =0,28        |
| C.d) | Prueba T para grupos independientes | Rechazada     | No atribuible a Chat GPT. D de Cohen =0,34       |
| C.e) | Correlación de Pearson              | Rechazada     | Sig = 0,431. Chat GPT es una variable confusora. |
| C.f) | Prueba T para grupos independientes | Rechazada     | No atribuible a ChatGPT. D de Cohen =0,17        |

|      | Statistical technique               | Status                     | Reason                                                                       |
|------|-------------------------------------|----------------------------|------------------------------------------------------------------------------|
| C.g) | Chi-cuadrado                        | Rechazada                  | La cuenta de usuarios de ChatGPT es menor que la cuenta de usuarios de SPSS. |
| C.h) | Prueba T para grupos independientes | Confirmada                 | D de Cohen = 2,52                                                            |
| C.i) | Prueba T para grupos independientes | Confirmada                 | Delta de Glass = 3,228                                                       |
| C.j) | Prueba T para grupos independientes | Ni confirmada ni rechazada | Significancia > 0,05 en todos los test.                                      |

#### 4. Conclusiones

La IA, y en particular los modelos de PLN, está transformando la educación superior, lo que suscita preocupaciones en torno a la seguridad de los datos, el rendimiento y la legitimidad en los entornos educativos. Las instituciones educativas y sus miembros tienen la responsabilidad de garantizar que la IA beneficie tanto a los alumnos como a los profesores, por ello, es necesario contar con una investigación objetiva que permita la toma de decisiones, en lugar de basarse únicamente en argumentos subjetivos o persuasivos de expertos sobre la IA.

El estudio se planteó originalmente con el objetivo de evaluar el esfuerzo y los resultados del alumnado al emplear la IA, pero condujo a un hallazgo inesperado relativo a la confianza manifestada en el uso de la IA basada en PLN.

El estudio presenta dos áreas principales de investigación. En primer lugar, se examinan de manera desagregada las competencias requeridas para el empleo de IA en educación universitaria. Los resultados indican que el uso de ChatGPT, por sí solo, no mejora los resultados de competencias de manera individual, por el contrario, el conocimiento previo sobre la materia parece constituir un factor más explicativo del rendimiento académico cuando se considera cada competencia de forma separada. De hecho, al medir el rendimiento global en la prueba sin desagregar por competencias, el uso de ChatGPT afecta negativamente a los resultados. Estas calificaciones pueden deberse tanto a cuestiones relacionadas con el funcionamiento de la propia herramienta, como a la forma en que el alumno interactúa con ella, lo que pone de manifiesto la necesidad de formarles previamente en el uso eficaz y crítico de las herramientas basadas en IA. En este sentido, estudios recientes han subrayado que el valor educativo de ChatGPT depende en gran medida de cómo el alumnado y el profesorado aprovechan su potencial dentro de procesos de aprendizaje estructurados (Zhu et al., 2023). Si bien han aparecido estudios previos que han subrayado el potencial de ChatGPT como herramienta de autoaprendizaje y apoyo académico en la educación superior (Nisar y Aslam, 2023), nuestros hallazgos sugieren que su eficacia depende en gran medida de la capacidad del alumno para evaluar y validar de forma crítica los resultados generados. En su estudio, Aiyappa et al. (2023) sostienen que, a pesar de su aparente utilidad, la evaluación de ChatGPT y de otros grandes modelos de lenguaje similares revela limitaciones asociadas a la contaminación de los datos, dado que dichos modelos se entrenan con la totalidad de la web. Sin embargo, en nuestra investigación, el problema principal parece surgir en una etapa posterior, más allá del

propio proceso de verificación de los hechos (como a Irfan y Murray, 2023). Según nuestros hallazgos, existe una cuestión más de fondo, que puede abordarse de manera más directa por parte del profesorado, las instituciones y el propio alumnado. Los estudiantes tendieron a depositar un grado excesivo de confianza en las respuestas percibidas como personalizadas y autorizadas, casi como si estuvieran consultando a un experto personal. Este comportamiento resulta coherente con el fenómeno de la descarga cognitiva (*cognitive offloading*) (Gerlich, 2025), según el cual la delegación de tareas cognitivas en herramientas externas reduce la implicación de los individuos en el razonamiento independiente y se asocia con un declive importante de las competencias de pensamiento crítico. Aunque la calificación media obtenida en el grado universitario puede explicar algunos resultados relacionados con las competencias, ChatGPT parece ejercer una influencia limitada sobre el resto de las habilidades. En consecuencia, la presencia de un sistema de IA conversacional y aparentemente personalizado puede eclipsar las competencias analíticas y de pensamiento crítico previamente adquiridas por el alumno.

El segundo punto estudiado se centra en el esfuerzo percibido por los alumnos. Los estudiantes que utilizaron ChatGPT declararon niveles de esfuerzo más bajos, en particular aquellos con un mejor conocimiento previo de la materia. Resulta interesante señalar que el número de preguntas formuladas a ChatGPT no afectó a la tasa total de respuesta a las preguntas del caso de estudio. Los estudiantes que declararon un menor esfuerzo y que emplearon ChatGPT presentaron mayores expectativas de calificación y mostraron una mayor confianza en sus respuestas.

Así pues, se sospechó que existía un efecto oculto en el alumno por el uso de la herramienta. Las herramientas de IA basadas en modelos de PLN están diseñadas para mantener conversaciones con los usuarios, transmitiendo una sensación de fluidez y naturalidad en el intercambio que puede reforzar la percepción de validez de los resultados obtenidos. El alumnado tiende a considerar que no obtiene una respuesta recopilada de internet (como ocurriría con un navegador), sino una respuesta creada específicamente por una Inteligencia, solo para ellos y para la ocasión.

En este punto, comenzamos a estudiar el concepto de Ilusión de Profundidad Explicativa (*Explanatory Illusion Depth*) (Rozenblit y Klein, 2002) con el fin de intentar explicar nuestros hallazgos. Según estos autores, los individuos pueden llegar a creer que poseen conocimiento sobre el mundo simplemente porque un determinado estado de opinión social (procedente de un grupo alineado con sus creencias) refuerza el conocimiento explicativo. Aun cuando la persona pudiera verificar los hechos, la Ilusión de Profundidad Explicativa le lleva a pensar algo que, en conjunto, no resulta cierto. Sin embargo, esta heurística no se correspondía plenamente con nuestros hallazgos, dado que, en nuestro caso, no implicaba necesariamente un estado de opinión social ni una narrativa explicativa.

El efecto observado y medido (un diálogo cercano, una baja fricción en las respuestas y resultados individualizados), se alinean estrechamente con lo que Thaler y Sunstein denominaron un «empujón» (*nudge*) (Thaler y Sunstein, 2008). Este concepto hace referencia a una arquitectura de elección diseñada específicamente para orientar nuestras acciones y percepciones por debajo de nuestro radar racional.

Posteriormente, tanto Yeung (2017) como Lanzing (2019) se aventuraron a proponer la categoría del «hiper empujón» (hypernudging). En este modelo, la arquitectura de elección no se limita a apelar a mecanismos básicos compartidos por todas las personas, sino que se individualiza mediante algoritmos y macrodatos (Big Data), creando un espacio de toma de decisiones íntimo, personal y único.

El alumnado prioriza su confianza en un sistema que le «susurra al oído» como un experto, por encima de su propio conocimiento (incluso cuando lo posee). No se les ocurre consultar otras fuentes ni verificar la información por sí mismos, aun cuando resultaría muy sencillo hacerlo. Existe un cierto «resplandor» (no sería exagerado calificarlo de confianza) en la interacción con la IA, que los deja «ciegos ante el fraude».

El fenómeno observado, al que hemos denominado «ceguera ante el fraude», parece ser propio del uso de herramientas de IA en el contexto del PLN. Se manifiesta como una percepción inicial de esfuerzo reducido, atribuible principalmente al uso de la herramienta y no a la tarea concreta en sí. Este esfuerzo reducido conduce, a su vez, a unas expectativas de calificación infladas, generando una sensación injustificada de autoatribución del logro. Los estudiantes que esperaban calificaciones más altas tendieron a atribuirse a sí mismos una parte significativa del trabajo a la hora de repartirlo entre ellos y la IA.

A diferencia de la «ilusión de profundidad explicativa», el fenómeno de la «ceguera ante el fraude» permite que los individuos se atribuyan por completo las respuestas o explicaciones, lo que da lugar a una reflexión de carácter moral que dejaremos al margen aquí por razones de espacio. La naturaleza privada de las interacciones con la IA desempeña un papel importante en esta autoatribución, ya que facilita borrar cualquier rastro de aportación externa. Comprender este fenómeno va más allá de enseñar a los alumnos a formular preguntas o a verificar resultados. Parece tratarse de un sesgo que tiene su origen en la propia interacción con los sistemas de IA basados en PLN.

Por todo ello, consideramos que emerge una nueva subtarea, tanto para el alumnado y el profesorado como para las instituciones. Como sugiere Joyner (2023), junto con la formación en pensamiento crítico y en la capacidad de cuestionar las respuestas, debería hacerse hincapié en la evaluación crítica de los procesos que generan dichas respuestas. Utilizar IA basada en modelos de PLN de manera eficaz implica desarrollar un conjunto específico de competencias, en particular dos habilidades clave que deberían cultivarse durante la etapa universitaria: la capacidad de formular consultas de manera estructurada y orientada a la consecución de objetivos, y la capacidad de interpretar los resultados con coherencia, competencia en la resolución de problemas y fiabilidad, entre otros atributos.

## 5. References

- Aiyappa, R., An, J., Kwak, H., & Ahn, Y. Y. (2023). Can we trust the evaluation on ChatGPT? *arXiv preprint arXiv:2303.12767*. <https://doi.org/10.48550/arXiv.2303.12767>
- Arif, T. B., Munaf, U., & Ul-Haque, I. (2023). The future of medical education and research: Is ChatGPT a blessing or blight in disguise? *Medical Education Online*, 28(1): 2181052. <https://doi.org/10.1080/10872981.2023.2181052>
- Chromik, M., Eiband, M., Buchner, F., Krüger, A., & Butz, A. (2021). *I Think I Get Your Point, AI! The Illusion of Explanatory Depth in Explainable AI*. In 26th International Conference on Intelligent User Interfaces (IUI '21). Association for Computing Machinery, NY, USA, 307–317. <https://doi.org/10.1145/3397481.3450644>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*. <https://doi.org/10.1080/14703297.2023.2195846>
- Fauzi, F., Tuhuteru, L., Sampe, F., Ausat, A. M. A., & Hatta, H. R. (2023). Analysing the role of ChatGPT in improving student productivity in higher education. *Journal on Education*, 5(4), 14886-14891. <https://doi.org/10.31004/joe.v5i4.2563>
- Firat, M. (2023). How ChatGPT Can Transform Autodidactic Experiences and Open Education? <https://doi.org/10.31219/osf.io/9ge8m>
- Fuchs, K. (2023). Exploring the opportunities and challenges of NLP models in higher education: is ChatGPT a blessing or a curse? In *Frontiers in Education*, 8, 1166682. Frontiers. <https://www.frontiersin.org/articles/10.3389/feduc.2023.1166682/full>
- Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312. <https://doi.org/10.3389/feduc.2023.1166682>
- Harasim, L. (2000). Shift happens: Online education as a new paradigm in learning. *The Internet and Higher Education*, 3(1-2), 41-61. [https://doi.org/10.1016/S1096-7516\(00\)00032-4](https://doi.org/10.1016/S1096-7516(00)00032-4)
- Irfan, M., Murray, L., & Ali, S. (2023). Insights into Student Perceptions: Investigating Artificial Intelligence (AI) Tool Usability in Irish Higher Education at the University of Limerick. *Global Digital & Print Media Review*, VI(II), 48-63. [http://dx.doi.org/10.31703/gdpmr.2023\(VI-II\).05](http://dx.doi.org/10.31703/gdpmr.2023(VI-II).05)
- Irfan, M., & Murray, L. (2023). Empowering Critical skills of Teaching and Learning with Artificial Intelligence (AI) tools. University of Limerick. <https://doi.org/10.34961/researchrepository-ul.22788506.v1>
- Joyner, D., (2023). *ChatGPT in Education: Partner or Pariah?* XRDS 29, 3 (Spring 2023), 48–51. <https://doi.org/10.1145/3589651>
- Lanzing, M. (2019). "Strongly recommended" revisiting decisional privacy to judge hypernudging in self-tracking technologies. *Philosophy & Technology*, 32(3), 549-568. <https://doi.org/10.1007/s13347-018-0316-4>
- Leontiev, N., (1978). *Activity, consciousness and personality*. Prentice-Hall.
- Liesa-Orus, M., Latorre-Cosculluela, C., Vázquez-Toledo, S., & Sierra-Sánchez, V. (2020). The technological challenge facing higher education professors: Perceptions of ICT tools for developing 21st century

- skills. *Sustainability*, 12(13), 5339. <http://dx.doi.org/10.3390/su12135339>
- Lo, C. K. (2023). What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*, 13(4), 410. *MDPI* AG. <http://dx.doi.org/10.3390/educsci13040410>
- Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. <https://doi.org/10.2139/SSRN.4354422>
- Moss, G, Jewitt, C, Levacic, R, Armstrong, V, Cardini, A. & Castle, F. (2007). The interactive whiteboards, pedagogy, and pupil performance evaluation: an evaluation of the Schools Whiteboard Expansion (SWE). Project London Challenge. [https://www.researchgate.net/publication/267239102\\_The\\_Interactive\\_Whiteboards\\_Pedagogy\\_and\\_Pupil\\_Performance\\_Evaluation\\_An\\_Evaluation\\_of\\_the\\_Schools\\_Whiteboard\\_Expansion\\_SWE\\_Project\\_London\\_Challenge](https://www.researchgate.net/publication/267239102_The_Interactive_Whiteboards_Pedagogy_and_Pupil_Performance_Evaluation_An_Evaluation_of_the_Schools_Whiteboard_Expansion_SWE_Project_London_Challenge)
- Nikolaos, K., Tzafes, O., & Thanos, T. (2019). University students' skills in using the internet for educational purposes and the digital divide. *European Journal of Open Education and E-Learning Studies*, 4(1), 60-72. <http://www.oapub.org/edu>
- Nisar, S., & Aslam, M. S. (2023). Is ChatGPT a Good Tool for T&CM Students in Studying Pharmacology? <https://dx.doi.org/10.2139/ssrn.4324310>
- O'Connor, P. & McAndrew, S. (2021). Enhancing Learning Experiences with AI-Powered Chatbots in Higher Education. *Journal of Educational Technology*, 48(3), 167-185.
- Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? *An analysis using natural language processing. Sci. Educ.* 105, 653-680. doi: 10.1002/sce.21623
- Raja, R., & Nagasubramani, P. C. (2018). Impact of modern technology in education. *Journal of Applied and Advanced Research*, 3(1), 33-35. <https://doi.org/10.21839/jaar.2018.v3i1.165>
- Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., ... & Aristovnik, A. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLoS One*, 20(2), e0315011
- Rawlins, P., & Kehrwald, B. (2010). Education technology in teacher education: Overcoming challenges, realizing opportunities. In *Cases on digital technologies in higher education: Issues and challenges* (50-63). IGI Global. <https://doi.org/10.4018/978-1-61520-869-2.ch004>
- Rospigliosi, P. (2023) Artificial intelligence in teaching and learning: What questions should we ask of ChatGPT? *Interactive Learning Environments*, 31(1), 1-3. <https://doi.org/10.1080/10494820.2023.2180191>
- Rozenblit, L. & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive science*, 26(5), 521-562. [https://doi.org/10.1207/s15516709cog2605\\_1](https://doi.org/10.1207/s15516709cog2605_1)
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1). <https://doi.org/10.37074/jalt.2023.6.1.9>
- Smith, J. (2022). Personalized Learning through AI in Higher Education. *International Journal of Educational Technology*, 39(2), 79-94.
- Sok, S., & Heng, K. (2024). Opportunities, challenges, and strategies for using ChatGPT in higher education: A literature review. *Journal of Digital Educational Technology*, 4(1), ep2401. <https://doi.org/10.30935/jdet/13152>
- Tajik, E. & Tajik, F. (2023, April). A Comprehensive Examination of the Potential Application of ChatGPT In Higher Education Institutions. *TechRxiv*. <https://doi.org/10.36227/techrxiv.22589497>
- Tate, T., Doroudi, S., Ritchie, D. & Xu, Y. (2023). Educational Research and AI-Generated Writing: Confronting the Coming Tsunami. <https://doi.org/10.35542/osf.io/4mec3>
- Thurzo, A., Strunga, M., Urban, R., Surovkova, J., & Afrashtehfar, K. I. (2023). Impact of



- artificial intelligence on dental education: a review and guide for curriculum update. *Education Sciences*, 13(2), 150. <https://doi.org/10.3390/educsci13020150>
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 15. <https://doi.org/10.1186/s40561-023-00237-x>
- Verenikina, I. (2010). Vygotsky in twenty-first-century research. In EdMedia+ innovate learning (pp. 16-25). Association for the Advancement of Computing in Education (AACE). <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=2337&context=edupapers>
- Vygotsky, L. S., & Cole, M. (1978). Mind in society: Development of higher psychological processes. Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Westin, T., Wiklund, M., Mozellius, P., & Norberg, L. (2015). Regression From Game-Oriented to Traditional School. *Journal of Educational Technology Systems*, 43(4), 349–370. <https://doi.org/10.1177/0047239515558>
- Yang, D., Oh, E. S., & Wang, Y. (2020). Hybrid physical education teaching and curriculum design based on a voice interactive artificial intelligence educational robot. *Sustainability*, 12(19), 8000. <https://doi.org/10.3390/su12198000>
- Yeung, K., (2017) ‘Hypernudge’: Big Data as a mode of regulation by design, Information, Communication & Society, 20:1, 118-136, DOI: 10.1080/1369118X.2016.1186713
- Zhu, C., Sun, M., Luo, J., Li, T., & Wang, M. (2023). How to harness the potential of ChatGPT in education? Knowledge Management & eLearning, 15(2), 133–152. <https://doi.org/10.34105/j.kmel.2023.15.008>



