

TRANSDISCIPLINARIEDAD

EVALUACIÓN DE *PROMPTS* EN EDUCACIÓN SUPERIOR: UN ENFOQUE TRANSDISCIPLINAR PARA EL DESARROLLO DE COMPETENCIAS PROFESIONALES

Santos Andrés GUTIÉRREZ FIGUEROA

Universidad Europea de Madrid

Resumen: Este estudio aborda el vacío metodológico en la evaluación de *prompts* en la educación superior desde un enfoque transdisciplinar. La investigación, realizada con 37 estudiantes de Administración y Dirección de Empresas de la Universidad Europea de Madrid, empleó la metodología CRAFT (Contexto, Función, Acción, Formato, Tono) para el diseño de prompts estructurados en entornos digitales simulados. Los resultados evidenciaron un «efecto techo» —el 89% de los alumnos obtuvo la máxima puntuación—, lo que reveló la insuficiencia del instrumento inicial para discriminar los niveles de dominio. Ante este hallazgo, se propone una rúbrica multidimensional que evalúa seis criterios: claridad, adecuación contextual, jerarquización, criterios de desempeño, capacidad de adaptación y fomento de competencias. El estudio concluye que la ingeniería de prompts requiere de instrumentos de evaluación con mayor granularidad, articulados con marcos éticos y de gobernanza institucional.

Palabras clave: ingeniería de *prompts*, evaluación de competencias, educación superior, enfoque transdisciplinar, rúbrica de evaluación.

I. INTRODUCCIÓN

La irrupción de la inteligencia artificial (IA) en la educación ha transformado radicalmente la manera en que se diseñan, implementan y evalúan las actividades docentes. En este novedoso paradigma, las instrucciones específicas que guían la interacción entre usuarios y sistemas de IA generativa, también denominados «*prompts*», se presentan como un elemento transformador en la mediación pedagógica y la evaluación de competencias profesionales. La calidad del prompt, definida por su estructura, claridad y capacidad de adaptación a contextos diversos, determina en gran medida la efectividad de las experiencias de aprendizaje y el desarrollo de habilidades transferibles al ámbito laboral. Gracias al auge de los modelos de atención, especialmente a partir del artículo seminal de Vaswani *et al.* (2017), se ha impulsado la creación de sistemas de IA capaces de interpretar y responder a instrucciones complejas, lo que ha abierto la puerta a nuevas formas de evaluación adaptativa y personalizada.

La rápida adopción de la IA generativa en la educación superior ha superado, sin embargo, el desarrollo de metodologías de evaluación robustas y estandarizadas. Mientras que la ingeniería de *prompts* se consolida como una competencia profesional crítica, los instrumentos de evaluación existentes a menudo carecen de la granularidad necesaria para diferenciar entre una competencia superficial y un dominio avanzado. Este desajuste da lugar a un vacío de evaluación, manifestado en fenómenos como el «efecto techo», donde una gran proporción de estudiantes alcanza las máximas calificaciones sin que ello refleje necesariamente una excelencia estratégica. Esta situación evidencia una crisis en las prácticas de evaluación tradicionales, que resultan insuficientes para medir las habilidades cognitivas de orden superior implicadas en la colaboración efectiva entre los seres humanos y la IA.

Este estudio aborda directamente dicho vacío. El avance tecnológico ha propiciado el desarrollo de enfoques transdisciplinarios que integran los conocimientos de distintas áreas —pedagogía, informática, psicología y ética, entre otras— y que se apoyan en el análisis de datos para identificar patrones de aprendizaje y diseñar rúbricas de evaluación más precisas y justas. En este contexto, el presente artículo se articula en torno a las siguientes preguntas de investigación:

1. ¿Cómo puede un marco de evaluación transdisciplinar y basado en datos medir eficazmente la calidad de los *prompts* diseñados por estudiantes?
2. ¿De qué manera el diseño de *prompts* estructurados impacta en el desarrollo de competencias profesionales específicas en un entorno de aprendizaje simulado?

3. ¿Qué elementos debe contener una rúbrica de evaluación para superar el «efecto techo» y discriminar con mayor granularidad los niveles de dominio en la ingeniería de *prompts*?

II. OBJETIVOS

1. Evaluar la calidad del *prompt*

El objetivo principal del artículo consiste en analizar la calidad del *prompt* como elemento fundamental en la evaluación de actividades docentes en las que se ha estimulado la utilización de modelos de inteligencia artificial generativa con una orientación transdisciplinar.

2. Objetivos secundarios

Con el objeto de valorar el impacto del *prompt* en el desarrollo de competencias profesionales en entornos simulados, de manera específica, se plantean los siguientes objetivos:

- Examinar la estructura de la instrucción, el modelo aplicado y su jerarquización, para identificar los elementos que contribuyen a su calidad y efectividad en el ámbito pedagógico.

- Analizar el papel de la conceptualización y desarrollo de *prompts* en sistemas de IA generativa.

- Implementar un modelo estructurado de diseño de *prompts* en tres fases —contextualización, operativa y evaluación—, con la integración de principios transdisciplinarios y lógicas complementarias.

- Valorar, a partir de una experiencia educativa concreta, cómo los *prompts* bien diseñados potencian la adquisición de competencias transversales y profesionales críticas para la empleabilidad.

- Proponer una rúbrica básica para la evaluación de *prompts* en contextos educativos, fundamentada en los hallazgos de la investigación y coherente con los desafíos del mercado laboral contemporáneo.

III. METODOLOGÍA

La metodología adoptada en este estudio combina la experimentación en el aula con una exhaustiva investigación bibliográfica, lo que se ha logrado con la integración de enfoques cualitativos y cuantitativos que han ofrecido una visión holística del fenómeno analizado.

1. Diseño de la investigación y participantes

La experiencia educativa se llevó a cabo en una asignatura universitaria del último curso del grado de Administración y Dirección de Empresas, en la que participaron 37 estudiantes. La actividad consistió en la utilización de modelos de IA generativa para argumentar la validación de productos en entornos digitales simulados, a través de *prompts* diseñados específicamente para el desarrollo de competencias profesionales críticas. Para el diseño de los *prompts* se sugirió una estructura en tres fases: contextualización (definición del escenario y objetivos), operativa (formulación de instrucciones y criterios de desempeño) y evaluación (análisis de resultados y retroalimentación).

2. El fundamento académico de las metodologías de *prompting* estructurado

La ingeniería de *prompts* ha emergido como una disciplina práctica para optimizar la interacción con los modelos grandes de lenguaje (LLM) y mejorar la precisión y fiabilidad de sus respuestas. Aunque en la práctica han proliferado marcos mnemotécnicos de origen profesional, como CRAFT, sus principios fundamentales están sólidamente respaldados por la investigación académica sobre la interacción humano-IA. Este fenómeno, en el que la práctica de la comunidad informa y acelera la formalización académica, es característico de los campos tecnológicos emergentes.

El «*prompting* estructurado» se define como la práctica de descomponer una solicitud compleja en componentes claros y explícitos que guían el proceso de razonamiento del modelo. La literatura académica ha validado la eficacia de sus elementos clave. Por ejemplo, definir un:

- Rol o una personalidad para la IA (p. ej., «Actúa como un analista de marketing experto»), proporcionar un,
- Contexto detallado sobre la tarea y sus objetivos, especificar una,
- Acción concreta (p. ej., «resume», «analiza», «compara»), y determinar el,
- Formato de salida (p. ej., «en una tabla», «en una lista con viñetas») son técnicas demostradas para mejorar significativamente la calidad y relevancia de las respuestas generadas en dominios tan diversos como el desarrollo de software y la educación de pacientes.

En este estudio, se sugirió a los estudiantes el uso de la metodología CRAFT (Contexto, Rol/Función, Acción, Formato, Tono) como una instanciación práctica de estos principios de *prompting* estructurado, con la finalidad de proporcionar una base metodológica sólida y académicamente justificable para la actividad.

3. Análisis del «efecto techo»: una implicación metodológica crítica

Un hallazgo central de la fase inicial de la investigación fue la detección de un pronunciado «efecto techo» en la evaluación. El hecho de que el 89% de los alumnos alcanzara la máxima puntuación en el diseño de *prompts* no se interpretó como un indicador de éxito universal, sino como una limitación metodológica crítica del instrumento de evaluación inicial.

Este resultado puso de manifiesto que la rúbrica empleada carecía de la granularidad suficiente para discriminar entre *prompts* que eran meramente funcionales y aquellos que demostraban un dominio avanzado, caracterizado por el pensamiento crítico, la creatividad estratégica y una comprensión profunda de la interacción con la IA. Este «efecto techo» se convirtió en el catalizador principal de la investigación, ya que evidenció la necesidad de diseñar un instrumento de valoración más sofisticado, capaz de distinguir y fomentar la excelencia en esta habilidad, lo que justifica el objetivo central de este artículo: la propuesta de una nueva rúbrica.

4. Triangulación de datos y revisión bibliográfica

La revisión bibliográfica se centró en fuentes académicas relevantes sobre inteligencia artificial en educación, diseño y evaluación de *prompts*, enfoques transdisciplinares y análisis de datos en contextos educativos. Se prestó especial atención al artículo de Vaswani *et al.* (2017) y a estudios recientes sobre la integración de la IA en la enseñanza y el desarrollo de competencias profesionales (UNESCO, 2024; Moll & Weseley, 2023; Bhardwaja, Ghose & AlMaadeed, 2024). Asimismo, se analizaron marcos teóricos sobre transdisciplinariedad (Compaso, s.f.), la ética y responsabilidad en la aplicación de la IA (Khelif, Ahmed & Souissi, 2024; Stahl, 2023), así como lo referente al diseño de rúbricas de evaluación adaptativas (Rubric Best Practices, 2019).

Se reconoce como una limitación del presente estudio la dependencia de los datos de rendimiento estudiantil y el análisis de los *prompts* generados. Para reforzar la validez de los hallazgos, futuras investigaciones deberían incorporar una triangulación de datos más robusta, incluyendo observaciones del docente, autoevaluaciones de los estudiantes y análisis de los entregables finales generados con la IA, tal y como sugieren las buenas prácticas metodológicas.

IV. RESULTADOS

Los resultados de la experiencia educativa y el análisis bibliográfico han permitido identificar una serie de hallazgos clave sobre la calidad del prompt y su impacto en el desarrollo de competencias profesionales.

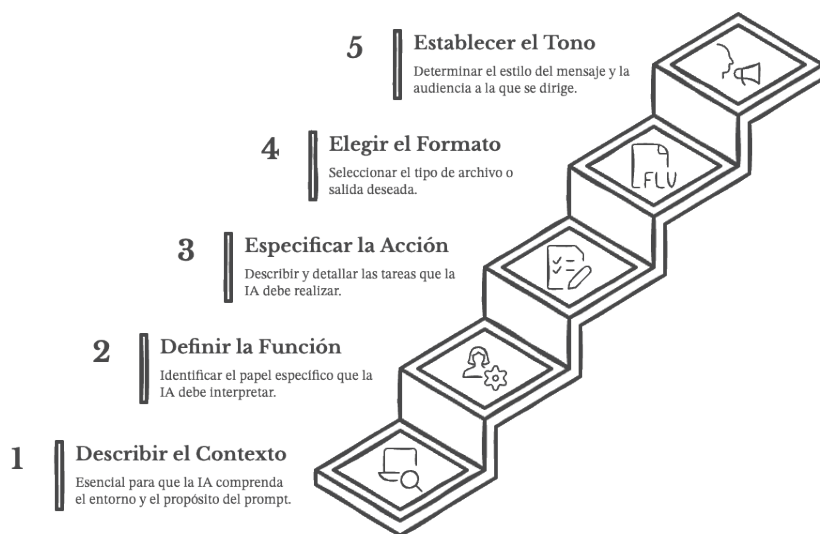
1. Estructura y calidad del *prompt*

La estructura del prompt emerge como un factor determinante para la efectividad de las actividades docentes en las que sea aplicada la IA. Los *prompts* diseñados en tres fases —contextualización, operativa y evaluación— facilitan una comprensión clara de los objetivos, las tareas a realizar y los criterios de éxito, lo que se traduce en una mayor motivación y desempeño por parte de los estudiantes. La jerarquización de instrucciones y la claridad en la formulación de los criterios de desempeño contribuyen a reducir la ambigüedad y a orientar de manera precisa la interacción con la IA, lo que favorece la adquisición de competencias transversales como el pensamiento crítico, la resolución de problemas y la comunicación efectiva (UNESCO, 2024; Moll & Weseley, 2023).

En la configuración de la arquitectura del *prompt*, la investigación se ha apoyado en la metodología CRAFT, orientada hacia un enfoque estructurado de la ingeniería de instrucciones. En virtud de ello, se pueden crear *prompts* eficaces a través de un proceso dividido en cinco elementos clave, a saber: Contexto, Función, Acción, Formato/Enfoque y Tono/Giro. Esta metodología garantiza una orientación precisa a la IA para que derive hacia resultados más específicos y deseados. Para ello, cada elemento proporciona al modelo un conjunto de introducciones o «*inputs*» dirigidas a afinar la respuesta del modelo de IA (Thomson, 2024):

- Contexto: información contextual indispensable para que la Inteligencia Artificial comprenda la situación y los objetivos a lograr.
- Función: indicación precisa del rol o la personalidad que la IA debe asumir al proporcionar su respuesta.
- Acción: definición exacta de la acción deseada de la IA (e.g., crear, resumir, redactar, traducir).
- Formato/Enfoque: determina la estructura del resultado esperado (por ejemplo, mediante el uso de viñetas, párrafos, o una longitud específica) e indica el enfoque principal que se debe adoptar.
- Tono/Target: especifica el tono o estilo deseado para la respuesta, así como la audiencia a la que se dirige el contenido.

FIGURA 1. *Representación gráfica de la Metodología CRAFT, en la que se pueden observar esquemáticamente los pasos recomendados para la construcción de un prompt efectivo*



Fuente: Elaboración propia con apoyo en la herramienta Napkin.ai

2. Rúbrica propuesta para la evaluación avanzada de *prompts*

Para abordar el «efecto techo» y el vacío de evaluación identificado, se desarrolló una rúbrica de evaluación más detallada. A diferencia del instrumento inicial, esta propuesta busca operacionalizar con mayor precisión los distintos niveles de dominio. Cada criterio se describe a continuación para clarificar las expectativas en cada nivel de desempeño.

— **Claridad y precisión:** Evalúa si la instrucción es inequívoca. Un nivel insuficiente presenta un prompt confuso, mientras que un nivel excelente no deja lugar a ambigüedades y guía a la IA de forma determinista.

— **Adecuación al contexto:** Mide la alineación del prompt con los objetivos de aprendizaje y el escenario propuesto. Un prompt excelente está perfectamente alineado y demuestra que el estudiante comprende el propósito estratégico de la tarea.

- Jerarquización y secuenciación: Valora la organización lógica de las tareas dentro del prompt. Un prompt excelente descompone una tarea compleja en pasos lógicos y secuenciales e imita un flujo de trabajo profesional.
- Inclusión de criterios de desempeño: Evalúa si el prompt especifica cómo debe ser una respuesta de calidad. Un prompt excelente incluye criterios claros, explícitos y medibles que la IA debe seguir, lo que obliga al estudiante a definir el éxito antes de actuar.
- Capacidad de adaptación: Mide la flexibilidad del prompt para ser ajustado a diferentes perfiles o escenarios. Un prompt excelente puede ser fácilmente modificado para distintos públicos o estilos y demuestra un pensamiento de diseño avanzado.
- Fomento de competencias: Valora si el prompt va más allá de la simple recuperación de información para fomentar habilidades superiores. Un prompt excelente está diseñado para provocar análisis, síntesis o creación, al tiempo que promueve activamente competencias profesionales clave.

TABLA 1. La siguiente tabla muestra el desarrollo de la propuesta de una rúbrica para la evaluación del prompt aplicado al ámbito de la práctica docente

<i>Criterios</i>	<i>1-3 Insuficiente</i>	<i>4-6 Regular</i>	<i>7-8 Bueno</i>	<i>9-10 Excelente</i>
Claridad y precisión	La instrucción es confusa o poco clara	La instrucción es comprensible, pero carece de precisión	La instrucción es clara y precisa en su mayoría	La instrucción es completamente clara y precisa, sin ambigüedades.
Adecuación al contexto	No se relaciona con el contexto, ni con los objetivos	Relación débil con el contexto los objetivos	Relación adecuada con el contexto y los objetivos	Totalmente alineado con el contexto y los objetivos de aprendizaje
Jerarquización y secuenciación	No hay jerarquización, ni secuenciación	Jerarquización y secuenciación poco clara	Tareas jerarquizadas y secuenciadas de forma lógica	Tareas perfectamente jerarquizadas y secuenciadas

<i>Criterios</i>	<i>1-3 Insuficiente</i>	<i>4-6 Regular</i>	<i>7-8 Bueno</i>	<i>9-10 Excelente</i>
Inclusión de criterios	No se incluyen criterios de desempeño	Criterios vagos y poco medibles	Criterios explícitos y medibles en su mayoría	Criterios de desempeño claros, explícitos y completamente medibles
Capacidad de adaptación	No se considera la diversidad de perfiles	Adaptación mínima a diferentes estilos de aprendizaje	Adaptación adecuada a varios perfiles y estilos	Totalmente adaptable a diferentes perfiles estilos
Fomento de competencias	No se fomenta ninguna competencia	Fomento limitado de competencias	Fomento adecuado de competencias transversales	Fomento excepcional de competencias transversales y profesionales

Fuente: Elaboración propia

V. DISCUSIÓN

La evaluación del prompt en contextos educativos representa un campo emergente de investigación con implicaciones profundas para el desarrollo de competencias y la transformación de las prácticas docentes. Los resultados confirman la importancia de la calidad de las instrucciones como elemento central en la mediación pedagógica y ponen de manifiesto la necesidad de adoptar enfoques transdisciplinares y basados en datos.

1. El *prompt* como puente transdisciplinar

La conceptualización del prompt como un elemento de conexión entre diversas disciplinas, tradicionalmente independientes, considera que la integración de conocimientos y metodologías de diferentes áreas posibilita la creación de actividades docentes más enriquecidas, complejas y congruentes con las exigencias del ámbito laboral. Los principios transdisciplinares, que reconocen la existencia de múltiples niveles de realidad y lógicas complementarias, facilitan la creación de experiencias de aprendizaje más inclusivas, equitativas y relevantes para los estudiantes (Compaso, s.f.; Moll & Weseley, 2023).

El enfoque transdisciplinar también contribuye a superar las limitaciones de los enfoques disciplinares tradicionales, que tienden a

fragmentar el conocimiento y a limitar la transferencia de competencias entre contextos. En este sentido, los *prompts* elaborados de manera óptima ejercen la función de agentes catalizadores en la integración del conocimiento, lo que facilita la construcción de competencias transferibles y la preparación de los estudiantes para enfrentar los retos del mercado laboral actual (OECD, 2025).

Tras la evaluación explícita del *prompting* respecto de su composición dentro de la actividad docente se ha detectado un impacto directo en el desarrollo de competencias fundamentales de la titulación en la que se ha aplicado, como:

— El análisis crítico: La construcción de un prompt sofisticado no pide un resumen, sino que exige a la IA «evaluar, comparar, criticar», lo que empuja al alumno a elaborar un pensamiento crítico en primer lugar.

— La integración de conocimientos: Los alumnos de la disciplina en la que se desarrollaron los estudios han debido integrar su propio conocimiento de las ideas y áreas de negocio en las que están desarrollando la actividad para formular un prompt que fuese estratégicamente relevante.

— El desarrollo de metodologías innovadoras: Se logró posicionar al *prompting* como una habilidad esencial dentro del arsenal de metodologías ágiles y de innovación.

— La aplicación de la competencia digital: a través de este tipo de ejercicios se produce un aumento de la competencia digital desde el simple uso de herramientas hacia la capacidad de interactuar con sistemas de IA generativa de un modo eficaz.

2. El modelo de atención y la personalización del aprendizaje

Con el auge de los modelos de atención y *transformers*, los sistemas de IA generativa actuales son capaces de adaptar sus respuestas a las necesidades específicas de cada estudiante y de cada contexto educativo (Vaswani *et al.*, 2017; Zhang *et al.*, 2023). Esta personalización es fundamental para el desarrollo de competencias profesionales en entornos simulados, ya que permite replicar escenarios complejos y realistas del mundo laboral.

De esta forma se facilita el aprendizaje experiencial y la transferencia de conocimientos. Además, la habilidad de los modelos de atención para centrarse en los aspectos más significativos de la instrucción contribuye a mejorar la eficacia de los *prompts* y el fortalecimiento de la adquisición de habilidades transversales (Moll & Weseley, 2023).

3. Evaluación adaptativa y análisis de datos

Como hemos observado, el análisis de datos posibilita la identificación de patrones de aprendizaje individualizados, lo que resulta fundamental para el diseño de rúbricas de evaluación adaptativas y personalizadas (Bhardwaja, Ghouse & AlMaadeed, 2024; OECD, 2025). En tal sentido, la evaluación basada en *prompts* de alta calidad se perfila así como una innovación metodológica que trasciende los límites disciplinares y responde de manera eficaz a los desafíos del mercado laboral contemporáneo.

De acuerdo a los resultados arrojados por la experiencia educativa se muestra que los estudiantes que interactúan con modelos de IA generativa a partir de instrucciones claras y jerarquizadas desarrollan un mayor nivel de argumentación, capacidad de análisis y resolución de problemas en entornos simulados, en comparación con aquellos que utilizan *prompts* menos estructurados.

4. Implicaciones éticas y la necesidad de una gobernanza institucional

No cabe duda que la integración de la IA en la educación plantea importantes desafíos éticos y de responsabilidad que deben ser abordados de manera proactiva para garantizar la equidad, la inclusión y la protección de los derechos de los estudiantes (Khlif, Ahmed & Souissi, 2024; Bhardwaja, Ghouse & AlMaadeed, 2024; Stahl, 2023). El diseño de los *prompts*, así como el de los sistemas de evaluación adaptativa debe incorporar principios éticos y de responsabilidad que aseguren la transparencia, la explicabilidad y la rendición de cuentas en el uso de la IA en contextos educativos (Mohammad & Ahmad, 2023).

La UNESCO (2024) ha enfatizado la necesidad de un enfoque centrado en el ser humano, que promueva la inclusión y la equidad en el acceso a las tecnologías de la IA y en la distribución de sus beneficios. En este sentido, la evaluación del *prompt* debe considerar no solo la calidad técnica de la instrucción, sino también su capacidad para promover valores éticos y democráticos, y para contribuir a la formación de ciudadanos críticos y responsables.

La conversación académica ha madurado, transitando desde una preocupación inicial por la integridad académica individual hacia un enfoque en la gobernanza institucional sistémica. El desafío ya no recae únicamente en el estudiante o el docente, sino en la responsabilidad de las instituciones académicas para crear un ecosistema de IA que sea seguro, equitativo y eficaz.

Una revisión sistemática conducida recientemente por García-López & Trujillo-Liñán (2025) subraya la urgencia de establecer marcos éticos y regulatorios para afrontar retos críticos como el sesgo algorítmico, que puede perpetuar desigualdades sociales; la privacidad de los datos de los estudiantes; y la propagación de desinformación generada por los modelos. Estos no son problemas aislados, sino riesgos sistémicos que demandan una respuesta coordinada.

En este sentido, la literatura aboga por un enfoque institucional proactivo para equilibrar la innovación con la integridad. Hubbard y Gawthorpe (2024) proponen el marco «ETHICAL» (*Excite, Tailor, Identify, Collect, Apply, Learn, Iterate*) como un modelo práctico para que las instituciones desarrollen la alfabetización en IA del personal y establezcan políticas transparentes y responsables. Una gobernanza efectiva implica no solo crear normativas, sino también rediseñar las evaluaciones para fomentar el pensamiento crítico y desarrollar programas de formación en IA para toda la comunidad educativa. La rúbrica propuesta en este estudio puede considerarse, por tanto, no solo como una herramienta de evaluación para el aula, sino como un componente de una estrategia institucional más amplia para la gobernanza responsable de la IA.

VI. CONCLUSIONES

Este estudio ha explorado la evaluación del prompt en la educación superior, proponiendo un marco para medir y desarrollar esta competencia emergente. A continuación, se presentan los hallazgos principales, las limitaciones del estudio y las recomendaciones para futuras investigaciones, estructurando las conclusiones de manera clara y diferenciada.

1. Hallazgos principales

— La calidad de los *prompts* estructurados es un factor determinante en el desarrollo de competencias profesionales en entornos de aprendizaje intervenidos por IA, yendo más allá del mero uso instrumental de la tecnología.

— El fenómeno del «efecto techo» en las evaluaciones de los estudiantes no es un indicador de éxito, sino un síntoma de un vacío de evaluación que revela la insuficiencia de las rúbricas tradicionales para medir la competencia en ingeniería de *prompts*.

— Una rúbrica multidimensional, como la propuesta en este trabajo, que evalúa criterios como la claridad, la adecuación contextual, la jerarquización lógica y el fomento de competencias transversales, se

presenta como una solución viable para superar dicho vacío de evaluación y proporcionar una retroalimentación más granular.

— La integración responsable de la IA en la educación exige un doble enfoque: por un lado, la innovación pedagógica en el aula y, por otro, el desarrollo de marcos de gobernanza institucional robustos que aborden los desafíos éticos de manera sistémica.

2. Limitaciones del estudio

Este trabajo presenta varias limitaciones que deben ser reconocidas para contextualizar sus hallazgos y guiar futuras investigaciones:

— **Tamaño de la Muestra y Generalizabilidad:** El estudio se realizó con una cohorte específica de 37 estudiantes de una única titulación, lo que limita la posibilidad de generalizar los resultados a otras disciplinas o contextos educativos.

— **Falta de Triangulación de Datos:** La investigación se basó principalmente en el análisis de los *prompts* y las calificaciones de los estudiantes. Como se señaló en la revisión metodológica, se carece de una triangulación con datos cualitativos como entrevistas, grupos focales o autoevaluaciones, que podrían ofrecer una comprensión más profunda del proceso de aprendizaje.

— **Validación del Instrumento:** La rúbrica propuesta es un constructo teórico derivado de los hallazgos de este estudio. Aunque está fundamentada en la experiencia y la literatura, requiere una validación empírica rigurosa en poblaciones estudiantiles más amplias y diversas para confirmar su fiabilidad y validez.

3. Recomendaciones y futuras líneas de investigación

A partir de los hallazgos y limitaciones, se proponen las siguientes recomendaciones y líneas de investigación:

— **Para Educadores:** Se recomienda la adopción de metodologías de prompting estructurado en el diseño de actividades con IA y el uso de rúbricas granulares para evaluar y desarrollar de forma explícita las habilidades de interacción de los estudiantes con estas herramientas.

— **Para Instituciones:** Es imperativo que las instituciones de educación superior desarrollen políticas proactivas y claras sobre el uso ético de la IA, inspirándose en marcos como «ETHICAL». Esto debe ir acompañado de una inversión en la formación del profesorado para construir una cultura de alfabetización en IA generalizada.

— **Para Futuras Investigaciones:** Se propone una agenda de investigación centrada en: (1) realizar estudios de validación de la rúbrica

propuesta en diferentes disciplinas y niveles educativos; (2) emplear diseños de investigación de métodos mixtos para triangular los datos sobre el proceso de aprendizaje del estudiante; y (3) investigar el impacto a largo plazo de la formación en ingeniería de *prompts* sobre la empleabilidad y el desempeño profesional de los egresados.

VII. REFERENCIAS

- BHARDWAJ, T., GHOUSE, G. y ALMAADEED, N. (2024). «Trustworthy and responsible AI: A review of the landscape and a future research agenda». *Big Data & Society*, 11(1), 1-19. <https://doi.org/10.1177/17470161231180449>.
- Compaso: *Journal of Comparative Research in Anthropology and Sociology* (s.f.). <http://compaso.eu>.
- DWIVEDI, R., PANDEY, S. K. y JAISWAL, A. K. (2024). «Artificial intelligence (AI) and machine learning (ML) in drug discovery and development: Applications, challenges and future prospects». *The Journal of Science and Technology*, 4(2), 195-212. <https://doi.org/10.55525/tjst.1272369>.
- ELNEKIEDY, A. M., GHONEIM, O. A. y EL-BAKRY, H. M. (2021). «Deep learning models for detecting depression from speech: A review». *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 3154-3159). IEEE. <https://doi.org/10.1109/BIBM52615.2021.9669861>.
- GARCÍA-LÓPEZ, I. M. & TRUJILLO-LIÑÁN, L. (2025). «Ethical and regulatory challenges of Generative AI in education: a systematic review». *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1565938>.
- KHLIF, H., AHMED, K. y SOUISSI, M. (2024). «Navigating the Ethical Labyrinth of Artificial Intelligence Research: Unveiling Undisclosed Realities and Charting a Path Forward». *Journal of Management Inquiry*. Publicación anticipada en línea. <https://doi.org/10.1177/10762175241289886>.
- MOHAMMAD, S. y AHMAD, J. (2023). «Explainable artificial intelligence (XAI): A systematic meta-survey of the literature». *Discover Artificial Intelligence*, 3(1), 20. <https://doi.org/10.1007/s43681-023-00294-5>.
- MOLL, R. F. y WESELEY, J. (2023). «Role of AI in physics education: A review and future directions». *Physical Review Physics Education Research*, 19(2), 020150. <https://doi.org/10.1103/PhysRevPhysEducRes.19.020150>.
- OECD. (2025). *Perspectivas del empleo 2024: Nota de país España*. OECD Publishing.
- OKOLIE, C., NWOSU, A. C., MBACHU, C. N. y OKOLIE, C. C. (2024). «Ethical considerations in the application of artificial intelligence in healthcare: A narrative review». *Researches in Medical and Artificial Lungs*, 2(1), 100068. <https://doi.org/10.1016/j.rmal.2023.100068>.
- POURAN, A., VEYSEH, B., DERNONCOURT, F., MIN, B. y NGUYEN, T. H. (2022). «Generating complement data for aspect term extraction with GPT-2». En *Proceedings of the Third Workshop on Deep Learning for Low-Resource Natural Language Processing* (pp. 1-10). Association for Computational Linguistics. <https://aclanthology.org/2022.deeplo-1.1>.
- SANTIAGO, T. (2024). *La IA generativa como soporte a los roles de gestión estratégica*. UPM.

- STAHL, B. C. (2023). *Artificial intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies*. Zenodo. <https://doi.org/10.5281/zenodo.8115695>.
- THOMPSON, T. (2023, 22 de octubre). *Using the CRAFT method to create an AI editor*. Medium. <http://bit.ly/45fP7T3>.
- UNESCO. (2024). *El uso de la IA en la educación: decidir el futuro que queremos*. UNESCO.
- VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A. N., KAISER, Ł. y POLOSUKHIN, I. (2017). *Attention is all you need*. arXiv. <https://doi.org/10.48550/arXiv.1706.03762>.
- VYGOTSKY, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- ZHANG, Y., FANG, Y., WANG, J. y WANG, X. (2023). «Understanding the impact of generative AI on academic writing: A bibliometric analysis». *Applied Network Science*, 8(1), 72. <https://doi.org/10.1186/s13040-023-00339-9>.
- ZHOU, W., CHEN, Z., WANG, Z., WU, J., LI, H. y LIN, W. (2024). *PromptIQA: Boosting Performance in No-Reference Image Quality Assessment*. arXiv. <https://doi.org/10.48550/arXiv.2402.05888>.

