

Diseño y validación inicial de la escala CREEAI-6: un instrumento para medir creencias sobre capacidades de la IA en el contexto argentino

Design and initial validation of the CREEAI-6 scale: an instrument for measuring beliefs about AI capabilities in the Argentine context

Joaquín Ignacio Mezzadra
Universidad de Flores, Argentina
joaquin.mezzadra@uflouniversidad.edu.ar
 [0009-0000-7962-4307](https://orcid.org/0009-0000-7962-4307)

Bruno Gambino
Universidad de Buenos Aires, Argentina
bgambino@sociales.uba.ar
 [0009-0004-6754-1925](https://orcid.org/0009-0004-6754-1925)

Verónica Bustamante-Serón
Universidad de Valencia, España
bustamante.vero@gmail.com
 [0009-0009-7163-7731](https://orcid.org/0009-0009-7163-7731)

Ismael Ruiz Ortiz
Universidad de Valencia, España
ismael.p310@gmail.com
 [0000-0003-2350-5637](https://orcid.org/0000-0003-2350-5637)

Gastón Becerra
Universidad de Buenos Aires, Argentina | Conicet
gaston.becerra@sociales.uba.ar
 <https://orcid.org/0000-0001-9432-8848>

Resumen

Este artículo presenta el diseño y la evaluación psicométrica inicial de la CREEAI-6, una escala breve de seis ítems para evaluar creencias sobre capacidades específicas de la inteligencia artificial (IA) en el contexto

argentino. Los ítems fueron desarrollados a partir de investigaciones previas sobre representaciones sociales de la IA en Argentina (Becerra, Mezzadra & Movía, 2025), cubriendo seis áreas de contenido: decisiones justas, respuestas certeras, orientación psicológica, compañía emocional, investigación científica y enseñanza/aprendizaje. La escala se administró a una muestra de 426 adultos argentinos usuarios de IA. Los resultados mostraron una consistencia interna adecuada ($\alpha = .804$; $\omega = .811$). El análisis factorial exploratorio (AFE) mostró una solución compatible con una estructura predominantemente unifactorial, que explicó el 42% de la varianza total, con cargas factoriales que oscilaron entre .553 y .767. No obstante, los índices de ajuste global del modelo fueron subóptimos (RMSEA = .152; TLI = .807), por lo que la unidimensionalidad de la escala debe considerarse preliminar y requerirá confirmación mediante análisis factorial confirmatorio en muestras independientes. Como evidencia de validez basada en la relación con otras variables, la CREEAI-6 presentó una correlación positiva y moderada con la actitud general hacia la IA (AIAS-4; $\rho = .487$, $*p < .001$) y diferencias significativas en función de la frecuencia de uso de IA y del sexo, en línea con lo esperado teóricamente. Se discuten las implicancias de contar con un instrumento breve con indicadores psicométricos iniciales adecuados para medir creencias sobre capacidades de la IA en población argentina y se sugieren futuras líneas de investigación para ampliar la evidencia de validez del instrumento.

Palabras clave: creencias; inteligencia artificial; validación de escalas; análisis factorial exploratorio

Abstract

This article presents the design and initial psychometric evaluation of the CREEAI-6, a brief 6-item scale for assessing beliefs about specific capabilities of artificial intelligence (AI) in the Argentine context. The items were developed based on previous research on social representations of AI in Argentina (Becerra, Mezzadra & Movía, 2025), covering six content areas: fair decision-making, accurate responses, psychological guidance, emotional companionship, scientific research, and teaching and learning. The scale was administered to a sample of 426 Argentine adults who are AI users. Results showed adequate internal consistency ($\alpha = .804$; $\omega = .811$). Exploratory factor analysis (EFA) showed a solution compatible with a predominantly unifactorial structure, accounting for 42% of the total variance, with factor loadings ranging from .553 to .767. However, global model fit indices were suboptimal (RMSEA = .152; TLI = .807), so the unidimensionality of the scale should be considered preliminary and will require confirmation through confirmatory factor analysis in independent samples. As evidence of validity based on relationships with other variables, the CREEAI-6 showed a positive and moderate correlation with general attitude toward AI (AIAS-4; $\rho = .487$, $p <$

.001) and significant differences by frequency of AI use and sex, consistent with theoretical expectations. The implications of having a brief instrument with adequate initial psychometric indicators for measuring beliefs about AI capabilities in the Argentine population are discussed, and future research directions to further expand the validity evidence of the instrument are suggested.

Keywords: beliefs; artificial intelligence; scale validation; exploratory factor analysis

Recibido: 18/08/2026

Aceptado: 20/10/2026

DOI: [10.63790/cdvs.v19i1.517](https://doi.org/10.63790/cdvs.v19i1.517)

Calidad de Vida y Salud se encuentra bajo la licencia de Creative Commons [Atribución-NoComercial-CompartirIgual 4.0.](https://creativecommons.org/licenses/by-nc-sa/4.0/)



1. Introducción

El desarrollo acelerado de la inteligencia artificial (IA) ha generado un creciente interés por comprender cómo las personas evalúan, atribuyen significados y se relacionan con estas tecnologías. Lejos de constituir un objeto técnico neutro, la IA se ha instalado en el centro de debates públicos, políticos y académicos que la asocian tanto con promesas de eficiencia e innovación como con riesgos éticos, sociales y epistemológicos (Chen et al., 2025; Grassini, 2023; Schepman & Rodway, 2022). Esta heterogeneidad de sentidos ha llevado a la investigación psicosocial a preguntarse no solo por las actitudes generales hacia la IA, sino también por las creencias específicas que las personas sostienen acerca de lo que estos sistemas pueden o no hacer en distintos ámbitos de la vida cotidiana (Becerra et al., 2025a, 2025b).

Las creencias sobre las capacidades de la IA constituyen un componente central de la evaluación que los usuarios realizan sobre estas tecnologías. A diferencia de las actitudes, definidas como disposiciones evaluativas globales (Ajzen, 1991), las creencias refieren a juicios más específicos sobre el desempeño de la IA en tareas concretas, y se construyen a partir de la integración de información proveniente de experiencias directas, discursos públicos y representaciones sociales disponibles (Hoff & Bashir, 2015; Becerra et al., 2025b). Esta distinción resulta clave para comprender los procesos de aceptación y adopción tecnológica, ya que una persona puede mantener una actitud general favorable hacia la IA y, al mismo tiempo, ser escéptica respecto de su capacidad para tomar decisiones éticas o brindar apoyo emocional (Kornowicz & Thommes, 2025; Santiago-Torner et al., 2025).

La literatura especializada ha mostrado que las creencias sobre las capacidades de la IA no se organizan de manera homogénea, sino que varían sistemáticamente según el ámbito de tarea considerado. Así, mientras que en funciones instrumentales, como el procesamiento de información o la automatización de rutinas, la atribución de competencia tiende a ser más elevada, en áreas de aplicación que implican juicio ético, sensibilidad interpersonal o razonamiento contextual, como la toma de decisiones con implicaciones morales o el acompañamiento emocional, la confianza en la IA se reduce significativamente (Pagano et al., 2023; Muthukumar, 2025; Alfahl, 2025). Esta variabilidad sugiere que una orientación general hacia las capacidades de la IA puede expresarse de manera diferente según el ámbito de aplicación, el riesgo percibido y las expectativas de confiabilidad asociadas a cada tarea (Candrian & Scherer, 2022; Chen et al., 2025).

A pesar del creciente interés por las creencias sobre la IA, la mayoría de los instrumentos disponibles se han desarrollado en contextos anglosajones o europeos y se centran predominantemente en dimensiones actitudinales generales (Grassini, 2023; Schepman & Rodway, 2020, 2022; Sindermann et al., 2021). Son escasas las herramientas diseñadas para capturar la especificidad de las creencias sobre capacidades en diferentes espacios de tarea, y prácticamente inexistentes en el contexto latinoamericano (Morales-García et al., 2024; Luque-Reca et al., 2026). Esta carencia resulta

particularmente problemática si se considera que las representaciones sociales sobre la IA están mediadas por coordenadas culturales, históricas y discursivas específicas. En Argentina, investigaciones previas del equipo han documentado discursos sociales de pregnancia en el sentido común en torno a la IA, que incluyen tanto expectativas sobre su utilidad pedagógica y científica como preocupaciones vinculadas a la justicia algorítmica, la fiabilidad de sus respuestas y su posible rol en esferas afectivas o terapéuticas (Becerra et al., 2025a, 2025b). Estas dimensiones configuran un entramado complejo que exige ser abordado con instrumentos sensibles al contexto local.

Áreas de contenido para la evaluación de creencias sobre capacidades de la IA

Con el propósito de operacionalizar el constructo de creencias sobre capacidades de la IA, el presente estudio toma como punto de partida investigaciones previas orientadas a relevar e identificar las discusiones sociales más relevantes sobre el fenómeno en el contexto argentino (Becerra et al., 2025a; Becerra et al., 2025b). A partir de algunas de las dimensiones allí identificadas, se han seleccionado seis áreas de contenido que resultan centrales para comprender la evaluación social de la IA. Cada uno de estos ámbitos fue utilizado como base para la redacción de un ítem o reactivo de la escala CREEAI-6, con la expectativa de que, en conjunto, operen como indicadores de una creencia general sobre las capacidades de la IA, en lugar de constituir dimensiones factoriales independientes. Esta estrategia responde al objetivo de contar con un índice breve que permita capturar la orientación evaluativa global de las personas respecto de lo que la IA puede o no hacer en distintos ámbitos de aplicación.

La cuestión de la equidad algorítmica constituye uno de los ejes centrales del debate contemporáneo sobre la IA. Lejos de operar como sistemas axiológicamente neutros, los modelos de aprendizaje automático incorporan y amplifican desigualdades estructurales presentes en los datos de entrenamiento, fenómeno que ha sido documentado tanto en las métricas de evaluación de equidad como en los propios mecanismos diseñados para su mitigación (Nguyen-Duc et al., 2024). De manera particular, se ha evidenciado que los modelos de lenguaje natural tienden a perpetuar estereotipos de género y asimetrías socioeconómicas, lo que interpela de forma directa la presunción de objetividad sobre la que frecuentemente se legitima el uso de estas tecnologías (Pagano et al., 2023). A esta problemática se añade la dificultad de atribuir responsabilidad cuando las decisiones asistidas por IA generan perjuicios, dado que la cadena de imputabilidad se diluye entre desarrolladores, operadores e instituciones, dificultando la rendición de cuentas (Nguyen-Duc et al., 2024). Estudios con población universitaria confirman que la privacidad y el uso indebido de datos figuran entre las principales preocupaciones éticas asociadas a la IA (Alfahl, 2025).

La confianza en la veracidad de los sistemas de IA representa una dimensión crítica cuya evaluación no puede disociarse del ámbito de aplicación. La evidencia empírica disponible revela perfiles de rendimiento marcadamente heterogéneos: mientras que en contextos de diagnóstico de enfermedades infecciosas se registran niveles de precisión relativamente elevados, éstos coexisten con una variabilidad considerable en indicadores complementarios y con una degradación significativa del rendimiento ante pruebas de consistencia temporal (Alzarea et al., 2025). Esta asimetría se hace aún más patente en el ámbito de la salud mental, donde modelos con altas puntuaciones en entornos controlados experimentan contracciones sustanciales de su capacidad predictiva cuando son validados con datos clínicos reales (Bobkov et al., 2025; Trelles et al., 2025). Subyace a estos hallazgos una tensión epistemológica de mayor calado, en tanto la brecha observada entre la identificación estadística de síntomas y la formulación de recomendaciones terapéuticas pone en cuestión si el rendimiento algorítmico refleja una comprensión semántica genuina o constituye, más bien, el producto de correlaciones estadísticas sofisticadas desprovistas de razonamiento clínico sustantivo (Muthukumar, 2025).

El interés por las aplicaciones terapéuticas de la IA ha experimentado una notable expansión en el periodo pospandémico, en parte como respuesta a las brechas estructurales de accesibilidad a la atención psicológica convencional que el contexto sanitario global hizo visibles (Santiago-Torner et al., 2025). En este marco, la evidencia disponible ofrece un panorama ambivalente. Por un lado, intervenciones digitales fundamentadas en enfoques cognitivo-conductuales han mostrado efectos favorables sobre la resiliencia y la reducción del agotamiento profesional en contextos laborales (Leightley et al., 2026), lo que ha contribuido a consolidar la percepción de estos sistemas como una vía de acceso legítima al apoyo psicológico de primer nivel.

Por otro lado, los datos relativos a situaciones de mayor complejidad clínica revelan los límites actuales de dicha autonomía. La necesidad de intervención humana inmediata en la amplia mayoría de los episodios de crisis detectados algorítmicamente (Bobkov et al., 2025) indica que la fiabilidad del sistema decrece precisamente cuando las consecuencias de un error son más graves. A ello se suma la escasez de ensayos clínicos aleatorizados con seguimiento longitudinal que permitan valorar la sostenibilidad de los efectos, la seguridad a largo plazo y las implicaciones éticas derivadas de la expansión del autocuidado automatizado (Santiago-Torner et al., 2025).

La atribución de capacidades afectivas a los sistemas de IA representa uno de los terrenos conceptualmente más disputados dentro del debate académico sobre estas tecnologías. La trayectoria de los robots de asistencia social (SARs) ilustra con claridad esta reconfiguración, puesto que su concepción ha evolucionado desde la de meros sustitutos deficitarios del cuidado humano hacia la de agentes capaces de mitigar estados de ansiedad y soledad, particularmente en poblaciones vulnerables y entornos institucionales (Nichol et al., 2024). Esta revalorización funcional se ve

reforzada por hallazgos que apuntan en una dirección contraintuitiva. En situaciones de comunicación de malas noticias, las respuestas generadas por modelos de lenguaje han sido percibidas como más empáticas que las emitidas por profesionales médicos, un resultado que no implica superioridad moral del sistema artificial, sino que evidencia que ciertas propiedades discursivas, como la consistencia tonal, la validación explícita y la ausencia de elementos defensivos, pueden producir efectos relacionales significativos en condiciones de alta carga institucional (Muthukumar, 2025). No obstante, estos hallazgos obligan a problematizar la distinción entre empatía sintética y autenticidad relacional, así como a interrogar los efectos a largo plazo sobre los patrones de apego y el desarrollo socioemocional. El riesgo de una hiperconectividad desprovista de reciprocidad subjetiva genuina permanece como una de las principales incertidumbres que la investigación futura deberá abordar (Santiago-Torner et al., 2025).

Por otro lado, la capacidad de la IA para transformar la producción científica ocupa un lugar central en el imaginario tecnológico contemporáneo, aunque la literatura especializada ofrece una valoración marcadamente ambivalente al respecto. Si bien la incorporación de estas herramientas ha ampliado el acceso a capacidades avanzadas de análisis, síntesis bibliográfica y exploración de grandes corpus de datos, lo que supone una reconfiguración parcial de los procesos hipotético-deductivos tradicionales, dicha expansión conlleva también riesgos epistemológicos de considerable magnitud. Entre ellos se destacan la proliferación de malas praxis asociadas al uso opaco de sistemas generativos como estrategia de productividad artificial, la dilución de la autoría intelectual y la emergencia de una monocultura metodológica que homogeneiza los criterios de producción y validación del conocimiento (Feenstra et al., 2024). A estas preocupaciones se suma la evidencia de una divergencia significativa entre las evaluaciones cualitativas de calidad científica realizadas por expertos y los resultados de los sistemas automatizados de clasificación bibliométrica (Bryce et al., 2020), lo que pone en cuestión la fiabilidad de los mecanismos de evaluación mediados algorítmicamente.

El debate sobre el papel de la IA en los procesos de enseñanza y aprendizaje ha consolidado, con relativo consenso, la idea de que su función más plausible no es la sustitución de la mediación docente, sino el fortalecimiento de la autoeficacia profesional mediante la automatización de tareas operativas de baja complejidad, lo que permite redirigir la atención hacia las dimensiones relacionales y pedagógicas de mayor valor formativo (Shi et al., 2025). En este marco, la disponibilidad de retroalimentación inmediata y la adaptación algorítmica de contenidos a perfiles individualizados de desempeño emergen como los mecanismos centrales a través de los cuales la IA puede contribuir a la eficiencia académica (Peng & Li, 2025). No obstante, la distribución de actitudes entre el estudiantado revela una delimitación cognitiva clara de la legitimidad atribuida a estos sistemas: la confianza en su potencial transformador se reduce de manera considerable cuando se trata de competencias clínicas o profesionales de alta

incertidumbre (Alfahl, 2025), lo que sugiere que la aceptación algorítmica está modulada por el gradiente de responsabilidad percibida en cada área. A ello se añade el riesgo, documentado empíricamente, de que la dependencia sostenida de sistemas de guía automatizada inhiba el desarrollo de procesos metacognitivos de orden superior (Peng & Li, 2025).

Sobre el presente estudio

En función del marco teórico expuesto, este estudio se propone diseñar y validar la escala CREEAI-6 (creencias sobre capacidades específicas de la inteligencia artificial), un instrumento breve de seis ítems que integra creencias en seis áreas de contenido —justicia algorítmica, veracidad de respuestas, orientación psicológica, compañía emocional, investigación científica y enseñanza/aprendizaje— en un índice general de creencia sobre las capacidades de la IA. La escala se fundamenta en investigaciones previas del equipo sobre representaciones sociales de la IA en Argentina (Becerra et al., 2025a, 2025b) y se orienta a cubrir la necesidad de contar con un instrumento breve, de fácil administración y con propiedades psicométricas adecuadas para su uso en el contexto local. De este modo, se espera contribuir a la comprensión de los patrones de creencias sobre la IA en población argentina y proporcionar una herramienta útil para futuras investigaciones sobre aceptación, adopción y usos de estas tecnologías. Para alcanzar éste, se propusieron como objetivos específicos: a) evaluar la fiabilidad de la escala mediante los coeficientes alfa de Cronbach y omega de McDonald; b) explorar su estructura factorial mediante un análisis factorial exploratorio (AFE); y c) analizar la evidencia de validez basada en la relación con otras variables, incluyendo la actitud general hacia la IA (AIAS-4) y variables sociodemográficas.

A partir de los objetivos planteados y del marco teórico desarrollado, se formularon las siguientes hipótesis de investigación:

- H1#. La escala CREEAI-6 presentará una estructura unifactorial, con cargas factoriales superiores a .50 y una varianza explicada superior al 40%.
- H2#. La escala mostrará una consistencia interna adecuada, con valores de alfa de Cronbach y omega de McDonald $\geq .70$.
- H3#. La puntuación total de la CREEAI-6 se correlacionará de manera positiva y moderada con la actitud general hacia la IA (AIAS-4; ρ entre .30 y .60).
- H4#. Se observarán diferencias estadísticamente significativas en las puntuaciones de la CREEAI-6: a) los varones puntuarán más alto que las mujeres y b) las personas con mayor frecuencia de uso de IA presentarán puntuaciones más altas que aquellas con menor frecuencia de uso.

2. Método

Se realizó un estudio instrumental de diseño transversal y enfoque cuantitativo no experimental, orientado al diseño y validación de la escala CREEAI-6 (creencias sobre capacidades específicas de la inteligencia artificial) en el contexto argentino.

Participantes

La investigación se llevó a cabo con una muestra inicial de 525 adultos, reclutada mediante un procedimiento de muestreo no probabilístico, a través de convocatorias en redes sociales, listas de correo de instituciones educativas y grupos profesionales de los sectores de educación, tecnología y ciencias sociales. Con el fin de garantizar la pertinencia cultural y lingüística de la escala CREEAI-6, se restringió el análisis a las personas residentes en Argentina. La muestra final quedó conformada por 426 participantes argentinos mayores de 18 años (el resto provenía de España, Uruguay y otros países hispanohablantes), todos ellos con experiencia de uso de herramientas de inteligencia artificial en el último año, aunque con distinta frecuencia. No se excluyó a quienes declararon no haber utilizado IA.

La edad media de los participantes fue de 41.2 años ($DE = 13.6$), con un rango etario de 18 a 74 años. En cuanto al género, el 62.7% se identificó como mujer ($n = 267$), el 36.4% como varón ($n = 155$) y el 1% restante como no binario o prefirió no responder ($n = 4$). El nivel educativo de la muestra era elevado: el 84.5% poseía estudios universitarios o de posgrado, con un 57.5% de grado completo. Respecto de la frecuencia de uso de IA, el 22.1% reportó un uso esporádico, el 29.8% ocasional, el 45.3% frecuente, y solo el 2.8% ($n = 12$) indicó no haber utilizado IA. Dadas estas características —predominio de niveles educativos altos y elevada familiaridad con la IA—, los resultados deben considerarse preliminares y no extensivos a la población general argentina.

Por último, el tamaño de la muestra final ($N=426$) supera las recomendaciones metodológicas para la realización de análisis factoriales, las cuales sugieren un mínimo de 400 participantes para obtener resultados estables (Goretzko et al., 2021). Asimismo, la muestra mantiene una proporción de más de 70 participantes por cada uno de los seis ítems de la escala, superando ampliamente el criterio tradicional de 10 a 1 recomendado en la literatura psicométrica (Costello & Osborne, 2005; Hair et al., 2019). Esta conjunción de un tamaño muestral elevado y cargas factoriales robustas ($>.70$) asegura la estabilidad de las estimaciones y la validez de las conclusiones (Ondé & Alvarado, 2020).

Instrumentos

Se administró una encuesta autoadministrada en línea que incluyó los siguientes instrumentos.

La escala de creencias sobre capacidades específicas de la IA (CREEAI-6)

Se construyó una escala *ad hoc* de seis ítems (uno por cada área de contenido) para evaluar el grado de acuerdo con seis capacidades atribuidas a la IA: a) tomar decisiones justas y objetivas (C5); b) dar respuestas certeras y sin prejuicios (C2); c) ofrecer orientación psicológica (C1); d) brindar compañía y apoyo emocional (C6); e) hacer investigación científica (C3); y f) enseñar y guiar el aprendizaje (C4).

El proceso de construcción de los ítems se desarrolló en dos fases. En primer lugar, la selección de las seis áreas de contenido se basó en un consenso intersubjetivo entre los investigadores del equipo, a partir de los resultados de investigaciones previas sobre representaciones sociales de la IA en Argentina (Becerra, Mezzadra & Movía, 2025). Estas investigaciones identificaron, mediante análisis de discursos sociales, cuáles eran las áreas de mayor pregnancia en el sentido común respecto de las capacidades atribuidas a la IA. El consenso del equipo permitió delimitar las seis áreas que, por su relevancia y recurrencia en el corpus analizado, resultaban más representativas del fenómeno. En segundo lugar, una vez redactados los ítems, se realizó una prueba piloto con jueces expertos en psicología social y en temáticas vinculadas a las nuevas tecnologías. Estos jueces evaluaron la claridad, la pertinencia y la adecuación de cada reactivo a su área de contenido correspondiente, y sus observaciones fueron incorporadas en la versión final del instrumento. Cabe señalar que este proceso constituye una aproximación cualitativa a la validez de contenido. No se realizó un estudio formal de validación de contenido con índices cuantitativos de acuerdo interjueces, lo que se reconoce como una limitación del estudio y una línea prioritaria para futuras investigaciones.

La decisión de utilizar un único ítem por área de contenido responde a dos criterios principales. En primer lugar, el objetivo de la escala es capturar una creencia general sobre las capacidades de la IA, más que evaluar de forma exhaustiva cada área por separado. En segundo lugar, la elección de una escala breve (seis ítems) facilita su administración en estudios con baterías extensas, reduciendo la carga de respuesta y el tiempo de cumplimentación (Clark & Watson, 2019). No obstante, esta estrategia implica que no es posible calcular la consistencia interna de cada área/ámbito de manera independiente, lo que se reconoce como una limitación del estudio. Las respuestas se registran en una escala tipo Likert de 5 puntos (1 = muy en desacuerdo; 5 = muy de acuerdo).

Tabla 1

Ítems de la escala CREEAI-6

Ítem	Contenido
C1	Creo que la IA puede ofrecer orientación y atención psicológica.
C2	Creo que la IA puede dar respuestas certeras y sin prejuicios a mis preguntas.
C3	Creo que la IA puede hacer investigación científica.
C4	Creo que la IA puede enseñar y guiar el aprendizaje.
C5	Creo que la IA puede tomar decisiones justas y objetivas.
C6	Creo que la IA puede brindar compañía y apoyo emocional.

Variables sociodemográficas y de uso

El instrumento recopiló datos relativos a la edad, sexo, nivel educativo, área profesional, país de residencia y frecuencia de uso de herramientas de IA (nunca, esporádicamente, ocasionalmente, frecuentemente).¹

Escala de actitud hacia la IA (AIAS-4)

Para el análisis de validez basada en la relación con otras variables (validez nomológica), se administró la versión adaptada al español argentino de la AI Attitude Scale (AIAS-4) de Grassini (2023), previamente validada en el contexto local (Mezzadra et al., 2026). Esta escala breve de cuatro ítems evalúa la actitud general hacia la IA como constructo unidimensional, abarcando la percepción de utilidad personal, el impacto en la vida y el trabajo, las intenciones futuras de uso y la valoración de su efecto global. Las respuestas se registran en una escala tipo Likert de cinco puntos (1 = totalmente en desacuerdo; 5 = totalmente de acuerdo). Su inclusión responde al objetivo de examinar la relación entre las creencias sobre capacidades específicas de la IA (CREEAI-6) y la actitud general hacia la IA (AIAS-4), esperando una correlación positiva y moderada entre ambos constructos.

La encuesta se administró en línea mediante un formulario web diseñado específicamente para este estudio.² La difusión del cuestionario se realizó a

¹ La frecuencia de uso de herramientas de IA se midió con una pregunta de opción única con cuatro categorías ancladas temporalmente: 1) nunca, 2) esporádicamente (menos de una vez al mes), 3) ocasionalmente (al menos una vez a la semana) y 4) frecuentemente (a diario o casi a diario).

² Accesible en <https://studio--ethical-compass-ycikd.us-central1.hosted.app/>

través de un muestreo no probabilístico de tipo bola de nieve. El estudio se desarrolló en el marco de las actividades de investigación de la Licenciatura en Sociología de la Universidad de Flores (Argentina). De acuerdo con la normativa argentina vigente (Resolución 1480/2011 del Ministerio de Salud de la Nación), las investigaciones que utilizan encuestas anónimas, sin recolección de datos sensibles ni información identificatoria, no requieren evaluación por un comité de ética en investigación. El diseño y los procedimientos del estudio fueron evaluados y aprobados por las instancias académicas correspondientes de la institución. Todos los participantes otorgaron su consentimiento informado electrónico antes de acceder al cuestionario. El anonimato de las respuestas fue garantizado al no solicitarse ningún dato de identificación personal que pudiera vincular las respuestas con los participantes. La recolección de datos se llevó a cabo durante los meses de octubre y noviembre de 2025, período en el que se obtuvo la muestra final de participantes argentinos.

Análisis de datos

Todos los análisis se realizaron utilizando diferentes funciones del software Jamovi (versión 2.7.30) sobre la submuestra de participantes argentinos ($N = 426$). Adicionalmente, los análisis complementarios, como el análisis paralelo y los intervalos de confianza para las cargas factoriales, se llevaron a cabo mediante el entorno RStudio a través de Posit Cloud, lo que permitió una mayor flexibilidad en los procedimientos de *bootstrapping*. Para los análisis, se verificaron los supuestos de normalidad mediante la prueba de Shapiro-Wilk, observándose desviaciones significativas en todos los ítems ($p < .001$), por lo que se utilizaron métodos robustos cuando correspondió.

Se calculó la fiabilidad de la escala mediante el coeficiente alfa de Cronbach (α) y el omega de McDonald (ω), considerando valores $\geq .70$ como aceptables. Siguiendo las recomendaciones de Clark y Watson (2019) y Hair et al. (2019), se esperaron correlaciones interítem moderadas (entre .30 y .80) como evidencia de que los ítems miden el mismo constructo sin resultar redundantes. Asimismo, se examinaron las correlaciones ítem-total corregidas (esperándose $\geq .30$) y el α si se elimina el elemento para evaluar la contribución de cada ítem a la consistencia interna de la escala.

Dado que los seis ítems de la escala CREEAI-6 se responden mediante una escala Likert de 5 puntos, los datos obtenidos son de naturaleza ordinal. El uso de técnicas estadísticas que asumen normalidad y continuidad en las variables (como el análisis factorial con estimación de máxima verosimilitud) puede resultar inapropiado para este tipo de datos (Li, 2016; Muthén & Muthén, 2017). Por ello, para el análisis factorial exploratorio (AFE) se optó por el método de factorización por eje principal (Principal Axis Factoring), que se basa en la matriz de correlaciones y no requiere el supuesto de normalidad multivariante (Hair et al., 2019). Previamente al AFE, se verificó la adecuación muestral mediante el índice de Kaiser-Meyer-Olkin (KMO), y se esperó un valor

$\geq .70$ para considerar la muestra adecuada (Kaiser, 1974) y la prueba de esfericidad de Bartlett: se esperó que resultara significativa ($p < .05$) para confirmar que la matriz de correlaciones no era una matriz identidad (Pérez & Medrano, 2010). Asimismo, el número de factores a retener se determinó mediante el criterio de autovalores mayores que 1 (Kaiser-Guttman), la inspección del gráfico de sedimentación (codo de Cattell) y la interpretabilidad teórica de la solución. Como se esperaba que los eventuales factores subyacentes pudieran estar correlacionados, se aplicó una rotación oblicua (Promax). Se consideraron aceptables las cargas factoriales $\geq .40$, y se examinó la presencia de cargas cruzadas (ítems con cargas $\geq .30$ en más de un factor). Se reportaron las comunialidades (o unicidades) de cada ítem.

Análisis complementario con matriz de correlaciones policóricas

Dado que la matriz de Pearson puede atenuar las correlaciones cuando los datos son ordinales, se realizó un AFE complementario sobre la matriz de correlaciones policóricas, más adecuada para este tipo de datos (Li, 2016; Muthén & Muthén, 2017). Esta matriz se calculó mediante el paquete *polycor* (Fox, 2019) en el entorno R (R Core Team, 2024) utilizando el método de estimación de dos pasos. Este método produce estimaciones ligeramente más conservadoras que los procedimientos de máxima verosimilitud implementados en otros programas, lo que se refleja en coeficientes de correlación algo inferiores. Seguidamente, se utilizó el mismo método de extracción (factorización por eje principal), el mismo criterio de retención de factores y la misma rotación que en el análisis principal. Adicionalmente, se realizó un análisis paralelo basado en la matriz policórica con 1000 simulaciones, siguiendo las recomendaciones de Horn (1965) y Finch (2020). Los resultados de este análisis complementario se reportan al final del apartado de resultados.

Para examinar la evidencia de validez basada en la relación con otras variables (AERA, APA & NCME, 2014), se realizaron los siguientes análisis:

Correlación con la actitud general hacia la IA (AIAS-4)

Se calculó el coeficiente de correlación Rho de Spearman entre la puntuación total de la CREEAI-6 (promedio de los seis ítems) y la puntuación total de la AIAS-4 (actitud general hacia la IA). Se esperaba una correlación positiva y moderada (ρ entre $.30$ y $.60$), lo que indicaría que ambos instrumentos miden constructos relacionados pero diferenciados (Campbell & Fiske, 1959). Adicionalmente, se calcularon las correlaciones de Spearman entre la puntuación total de actitud (AIAS-4) y cada uno de los seis ítems de la CREEAI-6 por separado. Este análisis permitió identificar qué áreas

específicas de capacidad se asocian más fuertemente con la actitud general hacia la IA, aportando evidencia adicional de validez nomológica.

Validez de grupos conocidos

Para examinar la validez de grupos conocidos de la escala CREEAI-6, se compararon las puntuaciones totales de la escala en función de variables que la literatura ha mostrado asociadas sistemáticamente con las evaluaciones sobre IA (Grassini, 2023; Møgelvang & Grassini, 2025; Schepman & Rodway, 2022). Dado que la prueba de Shapiro-Wilk indicó desviaciones significativas de la normalidad en las puntuaciones ($p < .001$), se utilizaron procedimientos no paramétricos para todas las comparaciones entre grupos:

- **Sexo.** Se realizó la prueba U de Mann-Whitney para comparar las puntuaciones totales de la CREEAI-6 entre varones y mujeres. Se hipotetizó que los varones presentarían puntuaciones significativamente más altas que las mujeres, en línea con la evidencia previa sobre actitudes y creencias hacia la IA (Møgelvang & Grassini, 2025). El tamaño del efecto se estimó mediante la correlación biseriada de rangos.
- **Frecuencia de uso de IA:** se realizó la prueba H de Kruskal-Wallis para comparar las puntuaciones totales de la CREEAI-6 entre los grupos de frecuencia de uso (nunca, esporádicamente, ocasionalmente, frecuentemente). Se esperaba una relación positiva y significativa entre la frecuencia de uso y las creencias en capacidades de la IA (a mayor uso, mayor acuerdo con las capacidades atribuidas). Ante la detección de diferencias significativas, se realizaron comparaciones *post hoc* mediante la prueba de Dunn con ajuste de Bonferroni para identificar diferencias específicas entre pares de grupos. El tamaño del efecto se estimó mediante el coeficiente ϵ^2 (epsilon cuadrado).

En todos los análisis de validez de grupos conocidos se reportaron los estadísticos de contraste (U para Mann-Whitney; H para Kruskal-Wallis; z para Dunn), grados de libertad (cuando correspondía), niveles de significación (p) y tamaños del efecto (correlación biseriada de rangos para Mann-Whitney; ϵ^2 para Kruskal-Wallis). Se adoptó un nivel de significación de $\alpha = .05$.

2. Resultados

Análisis descriptivos de los ítems

La Tabla 2 presenta los estadísticos descriptivos de los seis ítems de la escala CREEAI-6. Las medias oscilaron entre 1.87 (C3) y 3.37 (C6). Las

desviaciones típicas fueron similares en todos los ítems (DE entre 0.921 y 1.167), reflejando una variabilidad moderada en las respuestas. Los valores de asimetría y curtosis se encontraron dentro de rangos aceptables (Kline, 2016).

Tabla 2

Estadísticos descriptivos de los ítems de la escala CREEAI (N = 426)

Ítem	Contenido	M	DE	Asimetría	Curtosis
C1	Orientación psicológica	2.37	0.987	0.439	-0.319
C2	Respuestas certeras	2.63	1.048	0.221	-0.793
C3	Investigación científica	1.87	0.921	0.917	0.273
C4	Enseñanza y aprendizaje	2.23	1.073	0.536	-0.607
C5	Decisiones justas	2.91	1.167	0.014	-1.067
C6	Compañía emocional	3.37	1.055	-0.652	-0.431

Nota. M = media; DE = desviación típica.

Consistencia interna

La escala CREEAI-6 mostró una adecuada consistencia interna, con un alfa de Cronbach de .804 (IC 95% [.769, .833]) y un omega de McDonald de .811 (IC 95% [.772, .836]) (Tabla 3). La coincidencia entre ambos coeficientes y la precisión de sus intervalos de confianza, que en ningún caso incluyen valores inferiores a .70, respaldan la fiabilidad de la escala (Nunnally & Bernstein, 1994; Clark & Watson, 2019). A su vez, las correlaciones ítem-total corregidas (Tabla 4) oscilaron entre .495 y .668, todas superando el criterio mínimo de .30 (Field, 2018). El ítem con mayor contribución a la coherencia interna fue C3 (“Creo que la IA puede hacer investigación científica”; ítem-total = .668), mientras que C5 (“Creo que la IA puede tomar decisiones justas y objetivas”) presentó la correlación más baja, aunque aceptable (ítem-total = .495).

El análisis de α si se elimina el elemento indicó que la eliminación de cualquier ítem no mejoraba sustancialmente la consistencia interna (valores entre .752 y .791), lo que sugiere que todos los ítems contribuyen positivamente a la fiabilidad de la escala.

Tabla 3

Fiabilidad de la escala CREEAI-6

	Alfa de Cronbach [IC 95%]	ω de McDonald [IC 95%]	N elementos
CREEAI-6	.804 [.769, .833]	0.811 [.772, .836]	6

Tabla 4

Fiabilidad de cada elemento de la escala CREEAI-6

Si se descarta el elemento				
Ítem	Contenido	Correlación del elemento con otros	Alfa de Cronbach	ω de McDonald
C1	Orientación psicológica	.635	.757	.769
C2	Respuestas certeras	.529	.780	.789
C3	Investigación científica	.668	.752	.759
C4	Enseñanza y aprendizaje	.549	.776	.783
C5	Decisiones justas	.495	.791	.797
C6	Compañía emocional	.519	.783	.794

Análisis factorial exploratorio (AFE)

Para explorar la estructura factorial subyacente de la escala CREEAI-6, se realizó un análisis factorial exploratorio (AFE) utilizando el módulo Factor de Jamovi. Dado que se esperaba que los eventuales factores pudieran estar correlacionados, se utilizó el método de extracción de factorización por eje principal con rotación oblicua (Promax). Cabe señalar que, al retenerse un único factor, la rotación no modificó la solución factorial.

Previamente al AFE, se verificó la adecuación de los datos. El índice de Kaiser-Meyer-Olkin (KMO) fue de 0.798, valor considerado bueno y que indica que las correlaciones entre los ítems son lo suficientemente fuertes para el análisis factorial (Kaiser, 1974). Los valores de KMO individuales oscilaron entre 0.758 (C4) y 0.858 (C6), superando el umbral mínimo de 0.50. La prueba

de esfericidad de Bartlett resultó significativa: $\chi^2(15) = 785$, $p < .001$, lo que confirma que la matriz de correlaciones no es una matriz identidad.

El análisis reveló que un solo factor presenta un autovalor superior a 1 (autovalor = 2.52), explicando el 42.0 % de la varianza total. El segundo factor presentó un autovalor de 0.25, muy por debajo del criterio de Kaiser, y el gráfico de sedimentación mostró un punto de inflexión claro después del primer factor. Las cargas factoriales estandarizadas de la solución unifactorial (Tabla 5) fueron todas significativas y oscilaron entre .553 y .767, superando el umbral mínimo de .50 recomendado (Hair et al., 2019). Los intervalos de confianza al 95 % indican que las estimaciones son precisas y estables, ya que en todos los casos los límites inferiores superan el umbral de .50, con la excepción de C5 (límite inferior = .470) y C6 (límite inferior = .506), cuyos IC son ligeramente más amplios, pero no comprometen la interpretación de la escala. El ítem con mayor carga fue C3 (Investigación científica; carga = .767; IC 95% [.709, .823]), seguido de C1 (Orientación psicológica; carga = .718; IC 95 % [.650, .783]).

Tabla 5

Cargas factoriales y unicidades del AFE (solución unifactorial) con IC al 95%

Ítem	Descripción del reactivo	Carga factorial [IC 95%]	Unicidad
C3	Investigación científica	.767 [.709, .823]	.412
C1	Orientación psicológica	.718 [.650, .783]	.484
C4	Enseñanza y aprendizaje	.643 [.560, .723]	.587
C2	Respuestas certeras	.607 [.522, .689]	.631
C6	Compañía emocional	.575 [.506, .644]	.669
C5	Decisiones justas	.553 [.470, .636]	.694

Nota. La extracción de factores se realizó mediante el método de eje principal con rotación Promax (oblicua); la unicidad se definió, por consiguiente, como la diferencia entre la unidad y la comunalidad. Los intervalos de confianza (IC 95%) se obtuvieron mediante bootstrapping con 1000 réplicas.

El modelo unifactorial mostró los siguientes índices de ajuste, propios del análisis exploratorio: RMSEA = 0.152 (IC 90% [0.126, 0.180]), TLI = 0.807, BIC = 43.4, $\chi^2(9) = 97.9$, $p < .001$. Si bien estos valores se sitúan por debajo de los criterios habituales para modelos confirmatorios ($CFI \geq .90$; $RMSEA \leq .08$), en el contexto de un análisis factorial exploratorio la interpretación de estos índices debe ser cautelosa. El AFE tiene como objetivo principal

identificar la estructura subyacente de los datos —en este caso, una solución compatible con una estructura predominantemente unifactorial, respaldada por el autovalor > 1 , la varianza explicada, el gráfico de sedimentación y las cargas factoriales—, más que obtener un ajuste exacto del modelo (Fabrigar et al., 1999; Hair et al., 2019). La decisión de retener un factor se fundamenta en estos criterios empíricos y en la interpretabilidad teórica de la solución. En conjunto, los resultados indican que la escala CREEAI presenta una consistencia interna adecuada ($\alpha = .804$, $\omega = .811$), una solución compatible con una estructura predominantemente unifactorial que explica el 42 % de la varianza total, con cargas factoriales que oscilan entre .553 y .767. Vale aclarar que la unidimensionalidad de la escala debe considerarse preliminar, dado que el ajuste global del modelo fue subóptimo y requerirá confirmación mediante análisis factorial confirmatorio en muestras independientes.

Análisis complementario con matriz de correlaciones policóricas

Con el fin de evaluar la robustez de la solución factorial frente al tipo de matriz de correlaciones utilizada, se replicó el AFE sobre la matriz de correlaciones policóricas. La matriz policórica mostró correlaciones positivas y estadísticamente significativas entre todos los ítems, con coeficientes que oscilaron entre .296 (C4 con C5) y .640 (C3 con C4).

El análisis factorial exploratorio sobre esta matriz reveló una solución unifactorial prácticamente idéntica a la obtenida con la matriz de Pearson. El factor único presentó un autovalor de 3.082 y explicó el 42.0 % de la varianza total. Las cargas factoriales estandarizadas oscilaron entre .553 (C5) y .767 (C3), con comunales que variaron entre .31 (C5) y .59 (C3). La Tabla 6 presenta las cargas factoriales y las unicidades de esta solución complementaria.

Tabla 6

Cargas factoriales y unicidades del AFE complementario (matriz policórica, solución unifactorial) con IC al 95%

Ítem	Descripción del reactivo	Carga factorial [IC 95%]	Unicidad
C3	Investigación científica	.767 [.711, .815]	.412
C1	Orientación psicológica	.718 [.642, .784]	.484
C4	Enseñanza y aprendizaje	.643 [.564, .714]	.587
C2	Respuestas certeras	.607 [.513, .684]	.631
C6	Compañía emocional	.575 [.497, .645]	.669
C5	Decisiones justas	.553 [.461, .631]	.694

Nota. Extracción mediante factorización por eje principal. La matriz de correlaciones policóricas se calculó con el paquete polycor. Los intervalos de confianza (IC 95%) se obtuvieron mediante bootstrapping con 1000 réplicas. Las cargas factoriales son idénticas a las obtenidas con la matriz de Pearson (ver Tabla 5), lo que confirma la robustez de la solución.

Los intervalos de confianza al 95% indican que las estimaciones son razonablemente precisas y estables. La mayoría de los límites inferiores superan el umbral de .50 (C3, C1, C4, C2), mientras que C6 (.497) y C5 (.461) presentan límites inferiores ligeramente por debajo de ese umbral, lo que sugiere que estos dos reactivos podrían requerir una revisión en futuras versiones de la escala. No obstante, sus cargas factoriales puntuales (.575 y .553, respectivamente) superan el criterio mínimo de .50, por lo que su contribución al factor general sigue siendo adecuada.

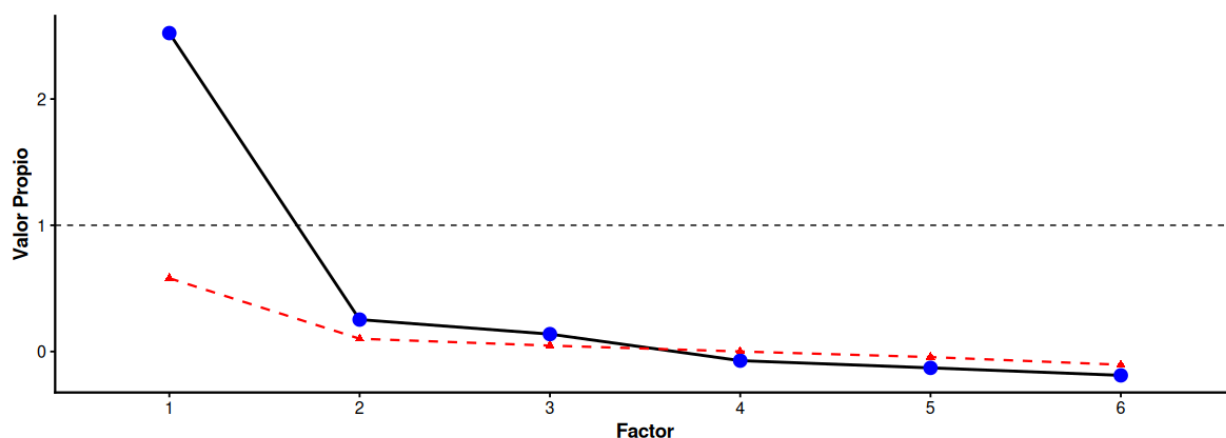
Los índices de ajuste del modelo unifactorial sobre la matriz policórica fueron los siguientes: RMSEA = .152 (IC 90% [.126, .180]), TLI = .807, BIC = 43.44, RMSR = .048, $\chi^2(9) = 97.93$, $p < .001$. Estos valores son idénticos a los obtenidos con la matriz de Pearson, lo que indica que el ajuste subóptimo del modelo no se debe al tipo de matriz de correlaciones utilizada, sino a la estructura misma de los datos (posible covarianza residual entre algunos ítems, como C3 y C4, que correlacionan .64). El valor del RMSR (.048), por debajo del umbral de .05, indica que los residuos entre las correlaciones observadas y las reproducidas por el modelo son pequeños en promedio, lo que respalda la adecuación global de la solución unifactorial.

Finalmente, se realizó un análisis paralelo basado en la matriz policórica (1000 simulaciones), siguiendo las recomendaciones de Horn (1965) y Finch (2020). Este análisis mostró que solo el primer factor presentó un autovalor real superior al valor propio simulado correspondiente. Los autovalores reales para los factores 2 y 3 (0.799 y 0.783, respectivamente) fueron inferiores a sus valores simulados (≈ 1.05 y ≈ 1.02), lo que indica que no superaron el criterio

de significación práctica. El gráfico del análisis paralelo (Figura 1) muestra visualmente que la línea de autovalores reales cae por debajo de la línea simulada después del primer factor. Si bien el *output* automático del análisis paralelo sugirió la posibilidad de retener tres factores, la inspección del gráfico y la comparación de los valores absolutos de los autovalores indicaron que los factores 2 y 3 carecían de significación práctica, por lo que no se exploró ninguna solución alternativa. En consecuencia, el análisis paralelo respalda de manera convergente una estructura unifactorial.

Figura 1

Análisis paralelo basado en la matriz de correlaciones policóricas



Nota. La línea negra con puntos azules representa los autovalores reales de la matriz policórica; la línea roja punteada representa los autovalores simulados (percentil 95) de 1000 matrices aleatorias. El punto de corte se produce después del primer factor.

Validez basada en la relación con otras variables

Para evaluar la validez de la escala basada en su relación con otras variables, se analizó la correlación entre la puntuación total de la CREEAI-6 y la puntuación total de la AIAS-4 (actitud general hacia la IA), previamente adaptada y validada en el contexto argentino (Mezzadra et al., 2026). Dado que las pruebas de normalidad indicaron desviaciones significativas (Shapiro-Wilk $p < .001$), se utilizó el coeficiente de correlación Rho de Spearman. El análisis reveló una correlación positiva, moderada y estadísticamente significativa entre ambas escalas: $\rho = .487$ ($N = 426$, $p < .001$) (Tabla 6). Este valor indica que las personas con actitudes más favorables hacia la IA tienden a atribuirle mayores capacidades en los ámbitos evaluados. La magnitud del coeficiente ($\rho = .487$; $\rho^2 = .237$) resulta relevante por dos razones: a) es suficientemente alta para confirmar que ambos constructos están relacionados (el 24 % de varianza compartida), y b) no es excesivamente alta

(por debajo del umbral de .70), lo que indica que los instrumentos no son redundantes y miden constructos diferenciados (Hair et al., 2019).

Tabla 7

Correlación entre la escala CREEAI-6 (creencias) y la AIAS-4 (actitud)

Variable	1	2
CREEAI-6 (creencia total)	-	
AIAS-4 (actitud total)	0.487***	-

Nota. N = 426. *** $p < .001$. Coeficiente Rho de Spearman.

Para explorar con mayor detalle esta relación, se calcularon las correlaciones entre la actitud general (AIAS-4) y cada uno de los seis ítems de la escala CREEAI-6 (Tabla 7). Los coeficientes oscilaron entre $\rho = .256$ (C2) y $\rho = .453$ (C6), todos estadísticamente significativos ($p < .001$). La magnitud de estas asociaciones fue moderada para la mayoría de los ítems (ρ entre .325 y .453), excepto para C2 (respuestas certeras), que mostró una relación pequeña ($\rho = .256$) (Cohen, 1988). El ítem con mayor varianza compartida fue C6 (compañía emocional) ($\rho^2 = .205$), mientras que C2 presentó la menor ($\rho^2 = .066$). Este patrón sugiere que la actitud general hacia la IA se asocia más fuertemente con la atribución de capacidades socioemocionales (compañía, orientación psicológica) y éticas (decisiones justas) que con capacidades puramente instrumentales (respuestas certeras).

Tabla 8

Correlaciones (Rho de Spearman) entre la actitud general (AIAS-4) y cada ítem de la escala CREEAI-6

Ítem	Contenido	ρ con AIAS-4	p
C6	Compañía emocional	.453	< .001
C1	Orientación psicológica	.379	< .001
C5	Decisiones justas	.357	< .001
C4	Enseñanza y aprendizaje	.327	< .001
C3	Investigación científica	.325	< .001
C2	Respuestas certeras	.256	< .001

Nota. N = 426. Todas las correlaciones son significativas al nivel $p < .001$.

La Tabla 8 presenta la matriz de correlaciones de los seis ítems de la escala CREEAI-6; todos los coeficientes resultaron positivos y estadísticamente significativos ($p < .001$), oscilando entre $\rho = .292$ (C4 con C5) y $\rho = .658$ (C3 con C4). Este patrón de correlaciones moderadas es consistente con la homogeneidad de los ítems. La correlación más alta se observó entre ítems conceptualmente afines (investigación y enseñanza; $\rho = .658$), mientras que la ausencia de correlaciones extremadamente altas ($> .80$) sugiere que los ítems no son redundantes (Clark & Watson, 2019). No obstante, la coherencia interna y la validez estructural de la escala se evalúan de manera más rigurosa mediante los coeficientes de fiabilidad (α y ω) y el AFE, presentadas en las secciones anteriores (Kline, 2016; Hair et al., 2019).

Tabla 9

Matriz de correlaciones entre los ítems de la escala CREEAI-6 (N = 426)

Ítem	C1	C2	C3	C4	C5	C6
C1	1.000	-				
C2	.561	1.000	-			
C3	.491	.406	1.000	-		
C4	.389	.350	.658	1.000	-	
C5	.391	.306	.411	.292	1.000	-
C6	.431	.297	.393	.361	.397	1.000

Nota. Coeficiente Rho de Spearman. Todas las correlaciones son significativas con $p < .001$.

Validez de grupos conocidos

Para examinar la validez de grupos conocidos de la escala CREEAI-6, se compararon las puntuaciones totales de la escala entre varones y mujeres. Dado que la prueba de Shapiro-Wilk indicó desviaciones significativas de la normalidad ($W = 0.985$, $p < .001$), se optó por la prueba U de Mann-Whitney como procedimiento estadístico no paramétrico. Los resultados mostraron diferencias estadísticamente significativas ($U = 15.279$, $p < .001$), con un tamaño del efecto moderado (correlación biseriada de rangos = 0.251) (Tabla 9). Las mujeres presentaron una media de 2.45 (DE = 0.707), mientras que los varones obtuvieron una media de 2.78 (DE = 0.756) (Tabla 10). La Figura 2 ilustra esta diferencia de forma gráfica, lo que muestra que los varones tienden a atribuir mayores capacidades a la IA en los ámbitos.

Tabla 10

Diferencias por sexo en la puntuación total de la CREEAI-6 (prueba U de Mann-Whitney)

Escala	U	p	Correlación biseriada de rangos
CREEAI-6	15.279	< .001	0.251

Tabla 11

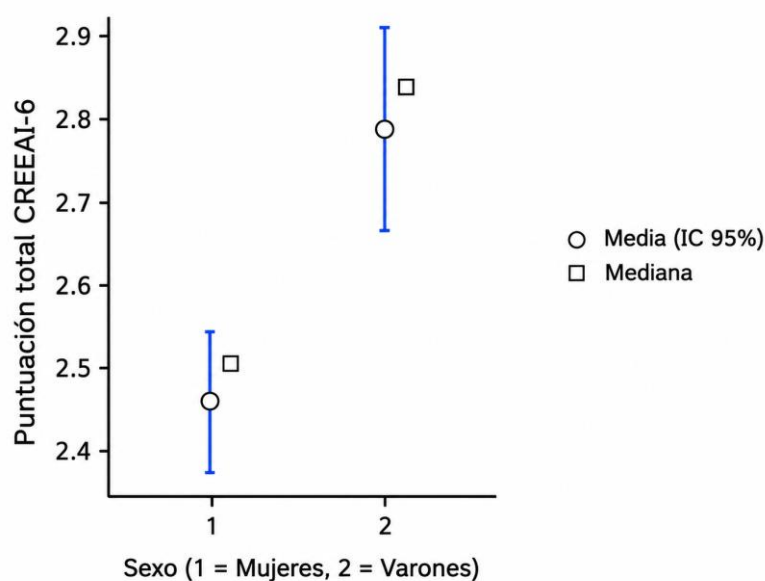
Estadísticos descriptivos de la puntuación total de la CREEAI-6 según sexo

Sexo	n	Media	DE	Mediana
Mujeres	270	2.45	0.707	2.50
Varones	151	2.78	0.756	2.83

Nota. La muestra presenta una distribución de sexos desbalanceada (64.1% mujeres, 35.9% varones), lo que se reconoce como una limitación para la generalización de los resultados.

Figura 2

Puntuaciones medias de la CREEAI-6 por sexo



Nota. Las barras representan las medias; los segmentos verticales indican el intervalo de confianza del 95%. La línea horizontal dentro de cada barra representa la mediana. Sexo: 1 = Mujeres (n = 270), 2 = Varones (n = 151).

Para examinar la validez de grupos conocidos en función de la experiencia con herramientas de IA, se compararon las puntuaciones totales de la escala entre los distintos niveles de frecuencia de uso (1 = Nunca, 2 = Esporádicamente, 3 = Ocasionalmente, 4 = Frecuentemente). Dado que la prueba de Shapiro-Wilk indicó desviaciones significativas de la normalidad ($p < .001$), se optó por la prueba H de Kruskal-Wallis como procedimiento estadístico no paramétrico. Los resultados mostraron diferencias estadísticamente significativas entre los grupos ($H = 42.4$, $gl = 3$, $p < .001$),

con un tamaño del efecto moderado ($\varepsilon^2 = 0.0998$) (Tabla 11). Las comparaciones *post hoc* con la prueba de Dunn (ajuste de Bonferroni) indicaron que el grupo de uso *frecuente* presentó rangos promedio significativamente más altos que los grupos de uso *ocasional* ($p = .008$), *esporádico* ($p = .024$) y *nunca* ($p < .001$). Asimismo, el grupo de uso *ocasional* puntuó significativamente más alto que los grupos de *nunca* ($p = .011$) y *esporádico* ($p = .024$). La diferencia entre los grupos de uso *esporádico* y *nunca* no alcanzó significación estadística ($p = .675$) (Tabla 12). Las medias y medianas (Tabla 13) mostraron un patrón ascendente: los participantes que reportaron un mayor uso de IA presentaron puntuaciones más altas en la escala de creencias, desde el grupo de *nunca* ($M = 1.89$) hasta el de uso *frecuente* ($M = 2.81$) (Figura 3).

Tabla 12

Prueba H de Kruskal-Wallis para las diferencias por frecuencia de uso

Escala	H	gl	p	ε^2
CREEAI-6	42.4	3	< .001	0.0998

Tabla 13

Comparaciones post hoc (Dunn) con ajuste de Bonferroni entre grupos de frecuencia de uso

Comparación	z	p (Bonferroni)
Nunca vs. Esporádico	-1.59	.675
Nunca vs. Ocasional	-3.12	.011
Nunca vs. Frecuente	-4.42	< .001
Esporádico vs. Ocasional	-2.88	.024
Esporádico vs. Frecuente	-5.41	< .001
Ocasional vs. Frecuente	-3.20	.008

Nota. Las comparaciones se basan en rangos promedio.

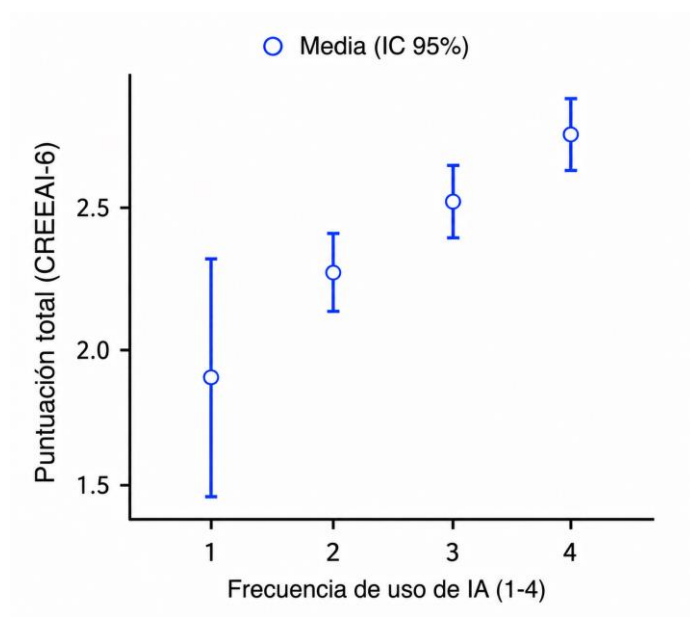
Tabla 14

Estadísticos descriptivos de la puntuación total de la CREEAI-6 según frecuencia de uso de IA

Frecuencia de uso	n	Media	DE	Mediana
Nunca	12	1.89	0.629	1.83
Esporádicamente	94	2.25	0.697	2.17
Ocasionalmente	127	2.53	0.734	2.50
Frecuentemente	193	2.81	0.687	2.83

Figura 3

Puntuaciones medias de la CREEAI-6 según frecuencia de uso de IA



Nota. Las barras representan las medias; los segmentos verticales indican el intervalo de confianza del 95%. La codificación de la variable frecuencia de uso se estableció como: 1 (nunca), 2 (esporádicamente), 3 (ocasionalmente) y 4 (frecuentemente).

4. Discusión y conclusiones

El presente estudio tuvo como objetivo diseñar y presentar evidencia psicométrica inicial de la CREEAI-6, una escala breve de seis ítems para evaluar creencias sobre capacidades específicas de la inteligencia artificial (IA)

en el contexto argentino. Los resultados obtenidos sugieren que la escala presenta indicadores de fiabilidad adecuados y evidencia exploratoria compatible con una estructura predominantemente unifactorial, así como relaciones teóricamente esperadas con la actitud general hacia la IA y con variables sociodemográficas. No obstante, como se discutirá, la evidencia de estructura interna obtenida en este estudio debe considerarse preliminar, dado que el ajuste global del modelo unifactorial resultó subóptimo. En conjunto, estos resultados posicionan a la CREEAI-6 como un instrumento prometedor para la evaluación de creencias sobre capacidades de la IA en población argentina, cuya confirmación requiere de estudios adicionales.

La escala CREEAI-6 mostró una consistencia interna adecuada, con valores de $\alpha = .804$ y $\omega = .811$, superando el umbral mínimo recomendado de .70 (Nunnally & Bernstein, 1994). Estos valores se sitúan en el rango esperado para escalas breves de nueva creación (Clark & Watson, 2019). Las correlaciones ítem-total corregidas (entre .495 y .668) indican que todos los ítems contribuyen de manera adecuada a la coherencia global de la escala, sin que ninguno de ellos resulte redundante o problemático (Field, 2018).

El análisis factorial exploratorio (AFE) mostró una solución compatible con una estructura predominantemente unifactorial, respaldada por el autovalor > 1 (2.52), la varianza explicada (42 %), el gráfico de sedimentación, el análisis paralelo y las cargas factoriales, que oscilaron entre .553 y .767. No obstante, los índices de ajuste global del modelo (RMSEA = .152; TLI = .807) se sitúan por debajo de los criterios habitualmente aceptados ($CFI \geq .90$; $RMSEA \leq .08$), lo que indica un ajuste subóptimo. Este resultado debe interpretarse con cautela y no permite afirmar que la unidimensionalidad de la escala haya quedado confirmada. La decisión de concebir los seis reactivos como indicadores de una única creencia general responde a un modelo reflectivo. Su justificación se sostiene en dos argumentos. Primero, investigaciones previas del equipo (Becerra, Mezzadra & Movía, 2025) documentaron un núcleo figurativo compartido en torno a las capacidades de la IA: quienes atribuyen mayores capacidades en un dominio tienden a atribuirles también en otros. Segundo, la evidencia empírica de este estudio respalda esa interpretación, ya que las correlaciones entre los seis reactivos fueron todas positivas y moderadas (.296 a .640) y un único factor explicó el 42 % de la varianza total. Si los ítems midieran constructos independientes, no se observaría este patrón. Aun así, se reconoce que la heterogeneidad conceptual de los seis contenidos introduce un límite a la interpretación reflectiva del modelo. Es posible que parte de la varianza de cada reactivo sea específica de su área de contenido y no compartida con el factor general, lo que podría explicar el ajuste subóptimo del modelo unifactorial. Futuros estudios deberán evaluar empíricamente si el modelo reflectivo unifactorial se sostiene frente a modelos alternativos (un modelo bifactor, un modelo formativo o una estructura multidimensional) mediante análisis factorial confirmatorio.

El ajuste limitado puede deberse a varias razones complementarias. En primer lugar, el AFE tiene como propósito explorar la estructura subyacente

de los datos más que probar un modelo teórico, por lo que los índices de ajuste global son menos informativos que los criterios de retención de factores en este contexto (Fabrigar et al., 1999). En segundo lugar, el modelo presenta solo nueve grados de libertad, lo que hace que el RMSEA sea especialmente sensible a pequeñas desviaciones y tienda a sobreestimar el desajuste (Kenny et al., 2015). En tercer lugar, la presencia de covarianza residual entre algunos reactivos —particularmente entre C3 y C4 (investigación científica y enseñanza/aprendizaje), que correlacionan .64— podría estar contribuyendo al desajuste del modelo unifactorial. Estas covariaciones sugieren que, si bien los seis reactivos comparten una orientación general hacia las capacidades de la IA, algunos de ellos podrían estar capturando contenidos adicionales que no están siendo modelados explícitamente. En particular, la covarianza entre C3 y C4 podría reflejar la afinidad temática entre ambos contenidos, que comparten un carácter académico y epistémico, en contraste con otros reactivos de naturaleza más emocional o ética. La heterogeneidad conceptual de los seis contenidos evaluados, si bien no invalida la existencia de un factor general, podría estar introduciendo varianza específica no capturada por el modelo unifactorial. Si bien teóricamente la escala fue diseñada para cubrir seis áreas de contenido (justicia algorítmica, veracidad, orientación psicológica, compañía emocional, investigación científica y enseñanza/aprendizaje), los datos ofrecen únicamente un respaldo preliminar a la idea de que estos ámbitos podrían operar como indicadores de una creencia general sobre las capacidades de la IA. Este hallazgo, que debe ser confirmado en futuros estudios con muestras independientes mediante un análisis factorial confirmatorio (Kline, 2016), es consistente con investigaciones previas sobre representaciones sociales de la IA en Argentina (Becerra et al., 2025a, 2025b), que documentaron la existencia de un núcleo figurativo compartido en torno a las capacidades de la IA, más que una diferenciación nítida entre ámbitos.

Respecto de la relación con la actitud general hacia la IA, la CREEAI-6 mostró una correlación positiva y moderada con la actitud general hacia la IA (AIAS-4; $\rho = .487$, $p < .001$), lo que indica que ambos constructos están relacionados en la dirección esperada sin resultar redundantes (Hair et al., 2019). Este hallazgo es consistente con el marco teórico que sustenta el estudio dado que las personas con actitudes más favorables hacia la IA tienden a atribuirle mayores capacidades en las áreas evaluadas. La magnitud del coeficiente ($\rho = .487$) es similar a la reportada en otros estudios de validación que examinan la relación entre actitudes generales y creencias específicas sobre tecnología (Schepman & Rodway, 2022; Grassini, 2023), y sugiere que ambos constructos comparten aproximadamente un 24 % de varianza, lo que respalda su distinción conceptual.

Las correlaciones diferenciales entre la actitud general y cada ítem de la escala aportan evidencia adicional de validez nomológica. El ítem con mayor correlación con la actitud general fue C6 (“compañía emocional”; $\rho = .453$), seguido de C1 (“orientación psicológica”; $\rho = .379$) y C5 (“decisiones justas”; $\rho = .357$). Este patrón sugiere que la actitud general hacia la IA se asocia más

fuertemente con la atribución de capacidades en áreas socioemocionales y éticas que con áreas puramente epistémicas. Una posible interpretación de este hallazgo, que deberá ser contrastada en investigaciones futuras, es que las personas con una valoración global más positiva de la IA tiendan a extender dicha confianza a sus capacidades en áreas sensibles desde el punto de vista ético y relacional (Schepman & Rodway, 2020). En contraste, el ítem con menor correlación fue C2 (“respuestas certeras”; $\rho = .256$), lo que sugiere que la precisión factual de la IA podría ser percibida como un aspecto más específico, menos central en la formación de la actitud general, o bien que responde a otros factores no evaluados en este estudio. Si bien el diseño correlacional no permite establecer relaciones causales, esta asociación diferencial es consistente con la literatura que señala que las valoraciones de utilidad personal y los juicios éticos sobre la IA suelen ejercer un mayor peso en la actitud general que la evaluación técnica de sus capacidades (Schepman & Rodway, 2020).

Asimismo, la CREEAI-6 mostró diferencias significativas en función de variables teóricamente relevantes, lo que aporta evidencia preliminar de validez de grupos conocidos. En primer lugar, se observaron diferencias significativas por sexo ($U = 15.279$, $p < .001$), con los varones presentando puntuaciones más altas ($M = 2.78$; $DE = 0.756$) que las mujeres ($M = 2.45$), consistentes con la literatura previa que reporta actitudes más favorables hacia la tecnología y la IA en varones (Møgelvang & Grassini, 2025; Grassini, 2023; Schepman & Rodway, 2022). El tamaño del efecto (correlación biseriada de rangos = 0.251) indica una diferencia moderada, que se alinea con la magnitud reportada en otros estudios de validación de escalas sobre IA (Morales-García et al., 2024). A pesar de ello, dado que la distribución de la muestra estuvo desbalanceada por sexo (64.1 % mujeres, 35.9 % varones), estos resultados deben interpretarse con cautela y requerirán confirmación en muestras más equilibradas.

En segundo lugar, se encontraron diferencias significativas en función de la frecuencia de uso de IA ($H = 42.4$, $gl = 3$, $*p < .001$; $\epsilon^2 = .0998$), con un patrón ascendente claro, en tanto los participantes con mayor frecuencia de uso presentaron puntuaciones más altas en la escala de creencias. Las comparaciones *post hoc* con la prueba de Dunn (ajuste de Bonferroni) indicaron que el grupo de uso frecuente presentó rangos promedio significativamente más altos que el grupo de uso ocasional ($p = .008$), el de uso esporádico ($p < .001$) y el grupo que nunca la usa ($p < .001$). Asimismo, el grupo de uso ocasional puntuó significativamente más alto que los grupos de uso esporádico ($p = .024$) y nunca ($p = .011$). La única comparación que no alcanzó significación estadística fue la que enfrentó a los grupos de uso esporádico y nunca ($p = .675$). Pero este último resultado debe interpretarse con especial cautela, dado que el grupo que declaró no haber utilizado IA estuvo compuesto por solo 12 participantes (2.8 % de la muestra), lo que limita severamente la potencia estadística de las comparaciones que lo involucran y restringe la generalización de los hallazgos a la población de personas que no utilizan IA. En conjunto, estos resultados muestran una asociación positiva

entre la frecuencia de uso de IA y las creencias más favorables sobre sus capacidades, un patrón que se alinea con la literatura sobre aceptación tecnológica, en la que la experiencia previa con la tecnología se relaciona con evaluaciones más positivas (Chen et al., 2025; Hoff & Bashir, 2015).

Sin embargo, el estudio presenta limitaciones que deben ser consideradas. En primer lugar, y de manera destacada, el ajuste global del modelo unifactorial fue subóptimo (RMSEA = .152; TLI = .807), lo que indica que la estructura unidimensional de la CREEAI-6 no ha quedado confirmada en este estudio y debe considerarse preliminar. Si bien el RMSR (.048) sugiere un ajuste absoluto adecuado y las cargas factoriales fueron robustas, la presencia de covarianza residual entre algunos reactivos (particularmente C3 y C4) y el reducido número de grados de libertad del modelo aconsejan interpretar estos resultados con cautela. En suma, dado que los seis reactivos representan áreas de contenido conceptualmente heterogéneas, no puede descartarse que el modelo reflectivo unifactorial no sea el más adecuado para representar la estructura latente del constructo. Futuros estudios deberán comparar este modelo con otros alternativos y confirmar la estructura mediante análisis factorial confirmatorio (AFC) en muestras independientes, modelando explícitamente las covariaciones residuales si fuera necesario (Kline, 2016). En segundo lugar, la muestra se caracteriza por un nivel educativo elevado (84.5 % con estudios universitarios o de posgrado) y una alta familiaridad con la IA, lo que limita la generalización de los resultados a poblaciones con menor nivel educativo o menor exposición a estas tecnologías. Asimismo, la distribución de la muestra estuvo desbalanceada por sexo (64.1 % mujeres, 35.9 % varones), lo que podría afectar la generalización de los resultados a poblaciones con una distribución más equilibrada. Futuros estudios deberían incluir muestras más diversas y balanceadas en términos sociodemográficos para ampliar la evidencia de validez de la escala (Clark & Watson, 2019). En tercer lugar, el grupo que declaró no haber utilizado IA estuvo compuesto por solo 12 participantes (2.8 % de la muestra), lo que limita severamente la potencia estadística de las comparaciones *post hoc* que lo involucran y restringe la generalización de los hallazgos a la población de personas que no utilizan IA. Futuros estudios deberían reforzar este segmento poblacional para obtener conclusiones más robustas sobre las diferencias en función de la frecuencia de uso. En cuarto lugar, el uso de un único ítem por cada área de contenido, si bien se justifica por la brevedad del instrumento, impide calcular la consistencia interna de cada área de manera independiente y limita la profundidad del constructo evaluado (Clark & Watson, 2019). Futuros estudios podrían ampliar la escala con ítems adicionales por área para evaluar la fiabilidad de posibles dimensiones por separado y realizar análisis factoriales confirmatorios de orden superior. En quinto lugar, el diseño transversal impide establecer relaciones causales entre la frecuencia de uso de IA y las creencias sobre sus capacidades. Investigaciones longitudinales podrían examinar si la experiencia de uso modifica las creencias a lo largo del tiempo (Chen et al., 2025). Finalmente, dado el carácter exploratorio de la escala CREEAI-6, no se ha evaluado la invarianza factorial

de la escala en función de variables como el sexo, la edad o la frecuencia de uso, lo que limita la capacidad para afirmar que la estructura de la escala se mantiene estable en diferentes subgrupos poblacionales (Kline, 2016). Futuras investigaciones deberán realizar análisis de invarianza factorial y confirmar la estructura unifactorial mediante un AFC en muestras independientes para corroborar la estabilidad de la estructura de la escala.

A pesar de estas limitaciones, la CREEAI-6 presenta indicadores de fiabilidad adecuados ($\alpha = .804$; $\omega = .811$) y evidencia exploratoria compatible con una estructura predominantemente unifactorial, así como relaciones teóricamente esperadas con la actitud general hacia la IA y con variables sociodemográficas. Estos resultados posicionan la escala como un instrumento con potencialidad para la evaluación de creencias sobre capacidades de la IA en población adulta argentina, particularmente en investigaciones que requieran una medida breve y de rápida administración. No obstante, la estructura factorial de la escala no ha quedado confirmada en este estudio, por lo que se recomienda que futuros estudios confirmen estas propiedades en muestras independientes mediante análisis factorial confirmatorio, evalúen la invarianza factorial en diferentes subgrupos poblacionales, amplíen la evidencia de validez mediante diseños longitudinales y muestras más diversas y exploren posibles fuentes de covarianza residual entre los reactivos.

5. Referencias

- AERA, A. NCME, American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement In Education (NCME) (2014). *Standards for educational and psychological testing*.⁷
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Alfahl, S. (2025). Knowledge attitudes and ethical concerns about artificial intelligence among medical students at Taibah University: A cross-sectional study. *Advances in Medical Education and Practice*, 16, 1609-1620. <https://doi.org/10.2147/AMEP.S528281>
- Alzarea, A. I., Alatawi, A. D., Alharbi, K. S., Alanazi, A. S., Alrashidi, A. H., Alshammari, M. S., Alshammari, T. K., & Alshammari, F. (2025). Evaluating AI performance in infectious disease education: A comparative analysis of ChatGPT, Google Bard, Perplexity AI, Microsoft Copilot, and Meta AI. *Frontiers in Medicine*, 12, Article 1679153. <https://doi.org/10.3389/fmed.2025.1679153>

- Becerra, G., Mezzadra, J. I., & Movia, G. (Eds.) (2025a). *Mitos y representaciones de la Inteligencia Artificial*. Editorial de la Universidad de Flores. <https://doi.org/10.63790/978-987-710-129-4>
- Becerra, G., Mezzadra, J., Ruiz Ortiz, I., & Gambino, B. (2025b). Representaciones sociales de la Inteligencia Artificial: una comparación entre 2021 y 2024. *Cultura y Representaciones Sociales*, 20(39), Article e0002048. <https://www.culturayrs.unam.mx/index.php/CRS/article/view/12860>
- Bobkov, A., Krasnov, G., Kolganov, A., Panov, M., & Burkin, M. (2025). Telepsychiatry and artificial intelligence: A structured review of emerging approaches to accessible psychiatric care. *Healthcare*, 13(11), Article 1348. <https://doi.org/10.3390/healthcare13111348>
- Bryce, C., Dowling, M., & Lucey, B. (2020). The journal quality perception gap. *Research Policy*, 49(7), Article 103957. <https://doi.org/10.1016/j.respol.2020.103957>
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105. <https://doi.org/10.1037/h0046016>
- Candrian, C., & Scherer, A. G. (2022). Decision-making with AI: The role of human oversight and responsibility attribution. *Journal of Business Ethics*, 180(2), 1-18. <https://doi.org/10.1007/s10551-021-04984-2>
- Chen, J., Xie, W., Xie, Q., Hu, A., Qiao, Y., Wan, R., & Liu, Y. (2025). A systematic review of user attitudes toward GenAI: Influencing factors and industry perspectives. *Journal of Intelligence*, 13(7), 78. <https://doi.org/10.3390/jintelligence13070078>
- Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, 31(12), 1412. <https://doi.org/10.1037/pas0000626>
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1), 1-9. <https://doi.org/10.7275/jyj1-4868>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272-299. <https://doi.org/10.1037/1082-989X.4.3.272>

- Feenstra, R. A., Carretero García, C., & Gómez Nicolau, E. (2024). Perception of research misconduct in a Spanish university. *Journal of Academic Ethics*. Advance online publication. <https://doi.org/10.1007/s10805-024-09526-7>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Finch, W. H. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis. *Educational and psychological measurement*, 80(2), 217-241. <https://doi.org/10.1177/001316441986576>
- Fox, J. (2019). *polycor: Polychoric and Polyserial Correlations*. R package version (0.7-10). <https://cran.r-project.org/package=polycor>
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 40(7), 3510-3521. <https://doi.org/10.1007/s12144-019-00300-2>
- Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. *Frontiers in Psychology*, 14, 1191628. <https://doi.org/10.3389/fpsyg.2023.1191628>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, 57(3), 407-434. <https://doi.org/10.1177/0018720814547570>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. <https://doi.org/10.1007/BF02289447>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36. <https://doi.org/10.1007/BF02291575>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Kornowicz, J., & Thommes, K. (2025). Algorithm, expert, or both? Evaluating the role of feature selection methods on user preferences and reliance.

PLoS ONE, 20(3), e0318874.
<https://doi.org/10.1371/journal.pone.0318874>

Leightley, D., Lavelle, G., Sun, S., & Fear, N. T. (2026). Effectiveness of AI-based interventions in workplace mental health: A systematic review and narrative synthesis. *British Medical Bulletin*, 157(1), 1-14.
<https://doi.org/10.1093/bmb/ldag007>

Luque-Reca, O., Peñacoba, C., Catalá, P., Gutiérrez Hermoso, L., & Grassini, S. (2026). Artificial intelligence attitudes in higher education: Spanish validation of the AIAS-4 and its associations with well-being and academic AI usage. *Current Psychology*, 45(4), 370.
<https://doi.org/10.1007/s12144-025-08566-5>

Mezzadra, J. I., Gambino, B., Bustamante-Serón, V., & Ruiz Ortiz, I. (2026). Adaptación y validación de la AI Attitude Scale (AIAS-4) en el contexto argentino. *Desarrollos en Ciencias Sociales Computacionales*, 1(1), e6.
<https://revistadesarrollos.uflo.edu.ar/index.php/RDCSC/article/view/8>

Møgelvang, A., & Grassini, S. (2025). Validating the AI attitude scale (AIAS-4) and exploring attitudinal differences in a large sample of Norwegian university students. *Discovery Education*, 4(1), Article 212.
<https://doi.org/10.1007/s44217-025-00657-6>

Morales-García, W. C., Sairitupa-Sanchez, L. Z., Morales-García, S. B., & Morales-García, M. (2024). Adaptation and psychometric properties of an attitude toward artificial intelligence scale (AIAS-4) among Peruvian nurses. *Behavioral Sciences*, 14(6), Article 437.
<https://doi.org/10.3390/bs14060437>

Muthukumar, K. (2025). Empathy AI in healthcare. *Frontiers in Psychology*, 16, Article 1680552. <https://doi.org/10.3389/fpsyg.2025.1680552>

Nguyen-Duc, A., Pham, N., & Hung, P. N. (2024). The landscape of ML fairness research in higher education: A systematic mapping study. *SSRN*.
<https://doi.org/10.2139/ssrn.4713827>

Nichol, B., Howarth, M., Nolan, C., & Victor, C. (2024). Exploring the impact of socially assistive robots on health and wellbeing across the lifespan. *International Journal of Nursing Studies*, 153, Article 104730.
<https://doi.org/10.1016/j.ijnurstu.2024.104730>

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

- Ondé, D., & Alvarado, J. M. (2020). Reconsidering the Conditions for Conducting Confirmatory Factor Analysis. *The Spanish Journal of Psychology*, 23, e55. <https://doi.org/10.1017/SJP.2020.56>
- Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., & Winkler, I. (2023). Bias and unfairness in machine learning models: A systematic review on datasets, tools, fairness metrics, and identification and mitigation methods. *Big Data and Cognitive Computing*, 7(1), Article 15. <https://doi.org/10.3390/bdcc7010015>
- Peng, J., & Li, Y. (2025). Frontiers of artificial intelligence for personalized learning in higher education: A systematic review of leading articles. *Applied Sciences*, 15(18), Article 10096. <https://doi.org/10.3390/app151810096>
- Pérez, E. R., & Medrano, L. A. (2010). Análisis factorial exploratorio: bases conceptuales y metodológicas. *Revista Argentina de Ciencias del Comportamiento (RACC)*, 2(1), 58-66. <https://doi.org/10.32348/1852.4206.v2.n1.15924>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Santiago-Torner, C., Corral-Marfil, J. A., & Tarrats-Pons, E. (2025). Artificial intelligence and the reconfiguration of emotional well-being (2020–2025): A critical reflection. *International Journal of Environmental Research and Public Health*, 22(1), Article 1. <https://doi.org/10.3390/ijerph22010001>
- Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in Human Behavior Reports*, 1, 100014. <https://doi.org/10.1016/j.chbr.2020.100014>
- Schepman, A., & Rodway, P. (2022). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human-Computer Interaction*, 39(13), 2724-2741. <https://doi.org/10.1080/10447318.2022.2085400>
- Shi, Q., Zhang, J., & Zhao, W. (2025). Progress or plateau? A 20-year systematic review of self-efficacy among TCFL teachers. *Frontiers in Psychology*, 16, Article 1678437. <https://doi.org/10.3389/fpsyg.2025.1678437>
- Sindermann, C., Sha, P., Zhou, M., Wernicke, J., Schmitt, H. S., Li, M., ... & Montag, C. (2021). Assessing the attitude towards artificial intelligence:

Introduction of a short measure in German, Chinese, and English language. *KI - Künstliche Intelligenz*, 35, 109-118.
<https://doi.org/10.1007/s13218-020-00689-0>

Trelles, A., Ruiz, T. F., & Rojo, A. P. (2025). Systematic review and meta-analysis of explainable machine learning models for clinical depression detection. *Behavioral Sciences*, 15(11), Article 1476.
<https://doi.org/10.3390/bs15111476>