

Assessment of Artificial Intelligence Literacy in Universities: A Critical Review

Mauricio Rojas-Contreras¹; Laura Villamizar Carrillo¹; Leydi Carolina Quintana Suárez¹; José Liviston Mendoza Bejarano¹; Oscar Albeiro Villamizar Peña¹

¹Universidad de Pamplona, Colombia, {mrojas, lauravillamizar, leydi.quintana, jliviston, oscar.villamizar3}@unipamplona.edu.co

Abstract– Artificial intelligence (AI) has become a transformative technology that redefines professional profiles and employability conditions, making it imperative for higher education institutions (HEIs) to prepare graduates to face these changes. This critical review examines the literature on the assessment of AI literacy (AIL) in the university context, supported by a multi-method approach that integrates critical review and hermeneutic triangulation. Existing conceptual frameworks, such as those from UNESCO and Ng et al., are analyzed, which converge on key dimensions: technical knowledge, application, critical evaluation, and ethics. A critical gap is evident between the high dependence on self-report instruments, which may be biased by individual overconfidence, and the emerging need for performance-based assessments that measure actual competence. Validated tools such as the AILQ (Artificial Intelligence Literacy Questionnaire) and the GLAT (Generative AI Literacy Assessment Test) performance test are highlighted, analyzing their structure and applicability. Institutional responses are explored, from "AI Across the Curriculum" curricular integration to the development of ethical policies. The findings indicate that while there is consensus on the dimensions of AIL, challenges persist in assessment standardization, academic integrity, and algorithmic bias. The conclusion underscores the urgency of adopting a multimodal approach that combines self-reports with performance tests to obtain a reliable measure of AIL, ensuring that graduates are truly prepared for an AI-driven labor market.

Keywords-- Artificial intelligence literacy assessment, Artificial intelligence, AI literacy in universities, Artificial intelligence in universities.

Evaluación de la alfabetización en inteligencia artificial en universidades: una revisión crítica

Mauricio Rojas-Contreras¹; Laura Villamizar Carrillo¹; Leydi Carolina Quintana Suárez¹; José Liviston Mendoza Bejarano¹; Oscar Albeiro Villamizar Peña¹

¹Universidad de Pamplona, Colombia, {mrojas, lauravillamizar, leydi.quintana, jliviston, oscar.villamizar3}@unipamplona.edu.co

Resumen— La inteligencia artificial (IA) se ha convertido en una tecnología transformadora que redefine los perfiles profesionales y las condiciones de empleabilidad, haciendo imperativo que las instituciones de educación superior (IES) preparen a los graduados para enfrentar estos cambios. Esta revisión crítica examina la literatura sobre la evaluación de la alfabetización en IA (AIL) en el ámbito universitario, soportada en un enfoque multimétodo que integra una revisión crítica y triangulación hermenéutica. Se analizan los marcos conceptuales existentes, como los de la UNESCO y Ng et al., que convergen en dimensiones clave: conocimiento técnico, aplicación, evaluación crítica y ética. Se evidencia una brecha crítica entre la alta dependencia de instrumentos de autoinforme, que pueden estar sesgados por la sobreconfianza del individuo, y la emergente necesidad de evaluaciones basadas en el desempeño, que miden la competencia real. Se destacan herramientas validadas como el cuestionario AILQ (Artificial Intelligence Literacy Questionnaire) y la prueba de rendimiento GLAT (Generative AI Literacy Assessment Test), analizando su estructura y aplicabilidad. Se exploran las respuestas institucionales, desde la integración curricular "AI Across the Curriculum" hasta el desarrollo de políticas éticas. Los hallazgos indican que, si bien existe un consenso sobre las dimensiones de la AIL, persisten desafíos en la estandarización de la evaluación, la integridad académica y el sesgo algorítmico. Se concluye subrayando la urgencia de adoptar un enfoque multimodal que combine autoinformes con pruebas de rendimiento para obtener una medida fiable de la AIL, garantizando que los graduados estén verdaderamente preparados para un mercado laboral impulsado por la IA.

Palabras clave— Evaluación alfabetización inteligencia artificial, inteligencia artificial, alfabetización inteligencia artificial universidades, inteligencia artificial en universidades.

I. INTRODUCCIÓN

La Inteligencia Artificial (IA) ha trascendido el ámbito de la tecnología especializada para convertirse en una fuerza transformadora en prácticamente todos los sectores de la sociedad. En la educación superior, esta transición se ha acelerado drásticamente con la llegada de herramientas de IA Generativa (GenAI) accesibles al público, como los grandes modelos de lenguaje [1], [2]. Estas tecnologías no solo están redefiniendo las prácticas pedagógicas y de investigación, sino que también están alterando las competencias requeridas para la participación académica y profesional en el siglo XXI. En este contexto, la "alfabetización en inteligencia artificial" (AIL) ha surgido como una competencia indispensable para estudiantes y personal universitario, generando una necesidad urgente de comprenderla, fomentarla y evaluarla de manera efectiva [3], [4].

La alfabetización en IA es un constructo multidimensional que va más allá de la simple capacidad para usar herramientas tecnológicas. Las conceptualizaciones fundacionales la definen como un conjunto de competencias

que permiten a los individuos comprender, utilizar, evaluar críticamente y comunicarse sobre la IA [5], [6]. Investigaciones más recientes han consolidado este concepto en un conjunto recurrente de dimensiones que incluyen: 1) el conocimiento conceptual de los fundamentos, capacidades y limitaciones de la IA; 2) las habilidades operacionales para utilizar herramientas de IA; 3) la capacidad de evaluación crítica para analizar los resultados generados por la IA, identificando errores y sesgos; y 4) la comprensión de las implicaciones éticas, sociales y de responsabilidad asociadas a su uso [7], [8], [9]. Esta visión amplia subraya que ser "alfabetizado en IA" implica tanto una competencia técnica como una conciencia crítica y ética.

En respuesta a la necesidad de medir esta competencia, la comunidad investigadora ha desarrollado una notable cantidad de instrumentos de evaluación en un corto período, especialmente entre 2023 y 2025 [8], [10], [11]. Estos instrumentos varían considerablemente en su enfoque y metodología. Por un lado, se encuentran escalas de autoinforme validadas psicométricamente, como la "Multidimensional AI Literacy Competency Scale" (MAIL-CS) [8], la "Critical Artificial Intelligence Literacy Scale" (CAIL) [7] y la escala vietnamita de alfabetización en IA [8], que evalúan la percepción de los estudiantes sobre sus propias habilidades. Por otro lado, han surgido herramientas específicas para la IA Generativa, como el "Generative AI Literacy Assessment Test" (GLAT), un test de opción múltiple diseñado para predecir el desempeño en tareas apoyadas por IA [12]. Sin embargo, la gran mayoría de estos instrumentos se basan en el autoinforme, lo que puede introducir sesgos de sobreconfianza y no reflejar necesariamente la competencia real [13].

A pesar de estos avances, una revisión de la literatura revela brechas significativas y persistentes en el panorama de la evaluación de la AIL. Primero, las dimensiones socioculturales y cívicas —que abordan temas como la gobernanza de la IA, el poder, el impacto laboral y la justicia global— son reconocidas en revisiones teóricas [14], [15], [16], pero raramente se operacionalizan como subescalas distintas en los instrumentos empíricos. Segundo, las evaluaciones basadas en el desempeño y los tests de conocimiento objetivo, que podrían ofrecer una medida más directa de la competencia, siguen siendo escasos y a menudo se limitan a aspectos conceptuales [17], [18], [19]. Tercero, la mayor parte de la investigación se ha centrado en los estudiantes, con una atención comparativamente menor a la

evaluación sistemática de la alfabetización en IA del personal docente, especialmente en sus dimensiones pedagógicas [10], [20]. Finalmente, aunque existen múltiples escalas validadas en contextos nacionales específicos, la invarianza de la medición entre diferentes culturas y disciplinas no ha sido establecida sistemáticamente, lo que limita la validez de las comparaciones globales [4], [21].

Este artículo presenta una revisión crítica de la literatura existente sobre la evaluación de la alfabetización en inteligencia artificial en la educación superior. El objetivo es sintetizar los enfoques actuales, identificar las fortalezas y debilidades de los marcos conceptuales e instrumentos de medición, y destacar las brechas críticas que requieren atención futura. Para ello, se examinan los estudios empíricos que definen, miden y analizan la AIL en poblaciones universitarias. El artículo se estructura de la siguiente manera: la Sección II especifica el diseño metodológico de la investigación, la sección III presentará los resultados de la revisión crítica, detallando los instrumentos y las dimensiones evaluadas. La Sección IV discutirá los patrones empíricos, los desafíos metodológicos y las brechas conceptuales identificadas. Finalmente, la Sección V ofrecerá conclusiones y recomendaciones para guiar futuras investigaciones, el desarrollo de políticas institucionales y el diseño de intervenciones curriculares destinadas a fomentar una alfabetización en IA robusta, crítica y holística en la comunidad universitaria.

II. MATERIALES Y MÉTODOS

Para dar alcance al objetivo del estudio, se utilizó el enfoque multimétodo [4],[5], en el cual se integraron los métodos de revisión crítica de literatura [6] y el método de triangulación hermenéutica [7],[8] como se especifica en la Fig. 1.

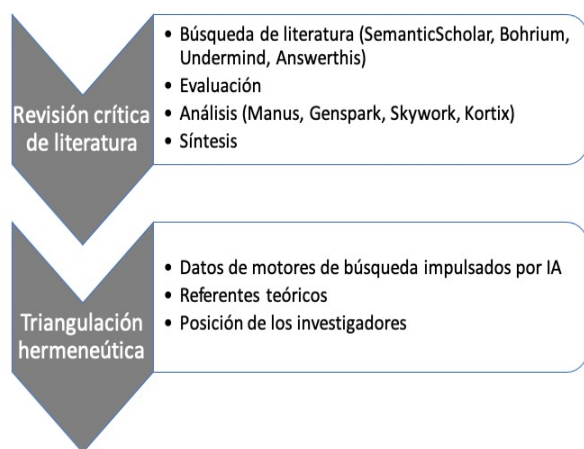


Fig. 1 Diseño metodológico-enfoque multimétodo

En forma específica, una revisión crítica se define como un examen de la literatura actualizada sobre un tema con el objetivo de sintetizar lo que se conoce y ofrecer nuevas perspectivas [9]. En contraste con las revisiones sistemáticas,

las revisiones críticas no buscan revisar y agregar exhaustivamente hallazgos previos. Más bien buscan identificar y resumir hallazgos clave e integrarlos para proporcionar nuevas perspectivas teóricas [10]. En forma complementaria, el método de triangulación hermenéutica adaptado al uso de la IA y particularmente del uso de motores de búsqueda impulsados por IA complementados con superagentes de IA, tiene por alcance analizar los datos obtenidos de la búsqueda en los motores de IA, luego se compara con los referentes teóricos y la posición de los investigadores con el fin de potencializar el pensamiento crítico en este tipo de estudio.

A. Materiales

La investigación se soportó en la búsqueda realizada en los motores impulsados por IA como son SemanticScholar, Bohrium, Undermind, Answerthis; en forma complementaria, se recurrió a superagentes de IA como Manus, Skywork, GenSpark y Kortix. En cuanto a la generación de la ecuación de búsqueda utilizada en los motores impulsados por IA se utilizó el LLM Claude en su versión Sonnet 4.5.

B. Recopilación de datos

Los datos se recopilaron mediante el uso de los motores de búsqueda impulsados por IA SemanticScholar, Bohrium, Undermind, Answerthis, a los cuales se les alimentó con la siguiente ecuación de búsqueda generada con la ayuda del LLM Claude Sonnet 4.5:

("artificial intelligence literacy" OR "AI literacy" OR "artificial intelligence competenc" OR "AI competenc*" OR "digital literacy" AND "artificial intelligence")*
AND
("assessment" OR "evaluation" OR "measurement" OR "assessment tool" OR "evaluation framework*" OR "measuring" OR "testing")*
AND
("university" OR "universities" OR "higher education" OR "tertiary education" OR "undergraduate" OR "college*" OR "academic" OR "postsecondary")*

Los motores impulsados por IA realizan el proceso de búsqueda en las bases de datos de mayor reconocimiento y credibilidad, en los repositorios de las mejores universidades del mundo y en los repositorios de preprints más reconocidos.

En forma complementaria, para soportar el proceso de triangulación hermenéutica se utilizaron los superagentes de IA Manus, Skywork, Genspark, Kortix, a los cuales se les solicitó que:

Genere un informe profundo y detallado acerca de la evaluación de la alfabetización en inteligencia artificial en universidades con citas y referencias en formato IEEE.

C. Análisis hermenéutico

Los datos recopilados se analizaron utilizando el método de triangulación hermenéutica, el cual permitió tensar los datos recopilados a través de los motores de IA, con los informes generados por los superagentes de IA y los ejercicios de verificación y validación del conocimiento a través de la estrategia HITL (human-in-the-loop) permitiendo integrar la posición de los investigadores en el proceso de análisis.

III. RESULTADOS

La revisión crítica revela un campo de investigación en rápida maduración, centrado en definir, medir y fomentar la alfabetización en Inteligencia Artificial (IA) en la educación superior. Los resultados se estructuran en torno a tres ejes principales: la conceptualización y las dimensiones de la alfabetización en IA, los instrumentos de medición desarrollados y los patrones empíricos observados en poblaciones universitarias.

A. Conceptualización y Dimensiones de la Alfabetización en IA

La investigación actual converge en un modelo multidimensional de la alfabetización en IA, superando las nociones iniciales de simple competencia digital. A través de múltiples estudios y el desarrollo de instrumentos psicométricos, se han consolidado consistentemente entre cuatro y cinco dimensiones fundamentales [2], [22]. Estas dimensiones recurrentes son:

1. Conocimiento Conceptual y Técnico: Tiene por alcance la comprensión de los conceptos básicos de la IA y el aprendizaje automático (ML), sus capacidades, limitaciones y aplicaciones comunes [3], [23-24].

2. Habilidades Operacionales y de Aplicación: Se refiere a la capacidad de utilizar herramientas de IA, integrarlas en tareas académicas o profesionales y resolver problemas específicos de un dominio [3], [25-26]. En el contexto de la IA Generativa (GenAI), esta dimensión se ha expandido para incluir explícitamente la ingeniería de prompts, la configuración de herramientas y la integración en flujos de trabajo [27], [28].

3. Pensamiento Crítico y Evaluación: Implica la habilidad para verificar los resultados generados por la IA, identificar errores, sesgos o limitaciones, y evitar una dependencia excesiva [3], [22], [29]. Esta competencia se considera un puente clave entre la comprensión técnica y la conciencia ética [17].

4. Ética y Responsabilidad: Abarca la comprensión de las implicaciones éticas, sociales y legales de la IA, incluyendo la equidad, el sesgo algorítmico, la privacidad, la rendición de

cuentas y la integridad académica [3], [7], [23-24]. La mayoría de los instrumentos de medición recientes incluyen subescalas dedicadas a esta dimensión [1], [8], [22].

Adicionalmente, la literatura identifica dimensiones emergentes. La alfabetización pedagógica en IA, específica para docentes y futuros profesores, se centra en la capacidad de diseñar actividades de aprendizaje con IA y gestionar su uso en la evaluación [10-11]. Por otro lado, la alfabetización sociocultural y cívica —que aborda temas de poder, regulación, impacto laboral y justicia global— es reconocida conceptualmente en revisiones teóricas [14-15], pero raramente se operacionaliza como una subescala distinta en los instrumentos empíricos existentes.

Se observa una distinción clara entre la alfabetización en IA general y la específica para GenAI. Mientras que la primera se enfoca en conceptos estables e independientes de la herramienta, la segunda es marcadamente práctica y centrada en tareas, abordando nuevas preocupaciones epistémicas y éticas como la autenticidad, la agencia humana y la automatización en el trabajo académico [7], [12].

B. Instrumentos de Medición y Enfoques de Evaluación

El avance en la conceptualización ha ido acompañado del desarrollo de un conjunto robusto de instrumentos de evaluación, predominantemente escalas de autoinforme validadas psicométricamente. Entre las más destacadas para poblaciones estudiantiles se encuentran:

La Escala Multidimensional de Competencia en Alfabetización en IA (MAIL-CS), desarrollada en China, que evalúa cuatro factores (Conocimiento Fundacional y Ética, Habilidades Operacionales, Evaluación Crítica, y Aplicación e Innovación) con alta fiabilidad y validez convergente [29].

La Escala de Alfabetización Crítica en Inteligencia Artificial (CAIL), originada en Italia, que mide cuatro dimensiones (conocimiento, operación, crítica y ética) y está fuertemente anclada en un marco de escrutinio crítico [3].

Una escala desarrollada y validada en Vietnam con 26 ítems que evalúa tres dimensiones (comprensión, aplicación, y ética y responsabilidad), demostrando coeficientes alpha de Cronbach superiores a 0.886 [8].

La Escala de Alfabetización en IA para Estudiantes Universitarios Chinos (AILS-CCS), un instrumento conciso de 15 ítems que cubre conciencia, uso, evaluación y ética [30].

Para el personal docente y futuros profesores, se han creado instrumentos específicos que incorporan la dimensión pedagógica, como las escalas desarrolladas en Corea y Tailandia [10-11].

En paralelo, han surgido instrumentos enfocados en la GenAI. El Test de Evaluación de la Alfabetización en IA Generativa (GLAT) es un test de opción múltiple de 20 preguntas que ha demostrado predecir el desempeño de los estudiantes en tareas apoyadas por IA mejor que las medidas de autoinforme [7]. Otros marcos se centran en habilidades específicas para tareas, como la escritura académica,

evaluando dimensiones como la autenticidad, la rendición de cuentas y la agencia [17].

A pesar del predominio del autoinforme, están comenzando a aparecer medidas objetivas y basadas en el desempeño. Ejemplos notables incluyen un test de conocimiento de 25 ítems para docentes no técnicos en EE. UU. [31] y la Escala Objetiva de Competencia en IA (AICOS) [32]. Sin embargo, estos enfoques son todavía incipientes y representan una clara área de oportunidad para futuras investigaciones.

C. Patrones Empíricos y Factores Asociados

Los estudios empíricos multinacionales revelan que la mayoría de los estudiantes universitarios poseen un nivel de alfabetización en IA que puede calificarse de "fundamental" o "moderado" [12], [17]. Generalmente, muestran altos niveles de autoeficacia, interés y actitudes positivas hacia el uso de la IA [17]. Sin embargo, un hallazgo recurrente es la existencia de una brecha significativa entre la percepción funcional de la IA (saber usarla) y una comprensión técnica o ética profunda de su funcionamiento [33].

Se han identificado disparidades consistentes en los niveles de alfabetización. Los estudiantes de disciplinas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas) tienden a obtener puntuaciones más altas que sus pares de humanidades y ciencias sociales [4], [18]. Asimismo, se observan variaciones significativas entre países, lo que subraya la importancia del contexto cultural y educativo [18], [33]. Por ejemplo, un estudio con estudiantes ecuatorianos no encontró diferencias significativas en la alfabetización en IA entre las modalidades de estudio presencial y virtual [9].

La investigación también ha comenzado a modelar los predictores y los resultados de la alfabetización en IA. La competencia digital previa y la experiencia en codificación son predictores positivos consistentes [33]. Factores motivacionales y de autorregulación del aprendizaje también juegan un papel crucial [21].

Más importante aún, la alfabetización en IA ha demostrado ser un predictor significativo de resultados educativos y conductuales clave. Se correlaciona positivamente con el rendimiento académico y el bienestar estudiantil [34], un mejor desempeño en tareas de escritura asistida por IA [21], una mayor conciencia ética sobre sesgos y privacidad [35-36], y una adopción más reflexiva de las herramientas de IA [37]. Por ejemplo, un estudio encontró una correlación positiva moderada-alta ($Rho=0.639$) entre el uso de herramientas de IA y el desarrollo de competencias digitales generales [38]. Estos hallazgos validan que la alfabetización en IA no es un constructo aislado, sino una competencia con consecuencias educativas tangibles.

Finalmente, la literatura señala desafíos persistentes que las universidades deben abordar, tales como la integridad académica, la dependencia tecnológica, los sesgos algorítmicos inherentes a las herramientas y la frecuente

ausencia de políticas institucionales claras que guíen el uso ético y efectivo de la IA [39], [40].

IV. DISCUSIÓN

Esta revisión crítica de la literatura sobre la evaluación de la alfabetización en Inteligencia Artificial (IA) en la educación superior revela un campo de investigación en rápida maduración, pero con brechas significativas que requieren atención urgente. Los resultados indican una notable convergencia hacia modelos conceptuales multidimensionales y un aumento en el desarrollo de instrumentos psicométricamente robustos para medir la alfabetización en IA de los estudiantes. Sin embargo, la evaluación sigue dependiendo en gran medida del autoinforme, y existen deficiencias críticas en la medición de las competencias del personal docente, así como en la operacionalización de las dimensiones socioculturales y cívicas de la alfabetización en IA. Esta sección discute las implicaciones de estos hallazgos, las limitaciones persistentes en el campo y las direcciones futuras para la investigación y la práctica institucional.

A. Convergencia Conceptual y la Distinción entre IA General y GenAI

Uno de los hallazgos más significativos es la consolidación de un marco conceptual para la alfabetización en IA. A través de múltiples estudios y contextos geográficos, emerge un consenso en torno a un conjunto de dimensiones centrales que definen esta competencia. Estas dimensiones incluyen consistentemente: 1) el conocimiento conceptual o técnico (saber qué es la IA, sus capacidades y limitaciones); 2) las habilidades operacionales o de aplicación (usar herramientas de IA); 3) el pensamiento crítico o de evaluación (analizar y cuestionar los resultados de la IA); y 4) la conciencia ética y la responsabilidad (comprender las implicaciones sociales y los sesgos) [3],[7] [9], [10]. Esta convergencia es un signo de madurez, permitiendo que la investigación avance desde la definición del constructo hacia su medición estandarizada y la evaluación de intervenciones pedagógicas.

Instrumentos como la Multidimensional AI Literacy Competency Scale (MAIL-CS) [41], la Critical Artificial Intelligence Literacy Scale (CAIL) [3] y la escala desarrollada en Vietnam por Do et al. [14] ejemplifican este avance, demostrando una alta fiabilidad y validez de constructo en poblaciones estudiantiles. Estos instrumentos proporcionan a las instituciones herramientas validadas para diagnosticar las necesidades de sus estudiantes y diseñar currículos basados en evidencia.

Paralelamente, la irrupción de la IA generativa (GenAI) ha creado una subárea de especialización. La alfabetización en GenAI, si bien se basa en las dimensiones generales, añade un fuerte énfasis en competencias prácticas y orientadas a tareas, como la ingeniería de prompts, la integración en flujos de

trabajo creativos y la comunicación [41-42]. Además, introduce nuevas consideraciones éticas y epistémicas relacionadas con la autenticidad, la agencia y la responsabilidad en el trabajo académico mediado por IA [43]. Aunque los marcos para la alfabetización en GenAI son más recientes y, en general, menos maduros psicométricamente que los de IA general, su enfoque práctico responde directamente a los desafíos y oportunidades que los estudiantes y docentes enfrentan actualmente en las aulas.

B. Brechas Persistentes en la Metodología de Evaluación

A pesar de los avances, la revisión identifica tres brechas metodológicas fundamentales que limitan la comprensión completa de la alfabetización en IA en el entorno universitario.

En primer lugar, existe una dependencia excesiva de las escalas de autoinforme. Si bien son escalables y útiles para medir percepciones y actitudes, los autoinformes pueden no reflejar la competencia real, siendo susceptibles a sesgos de sobreconfianza o deseabilidad social [44]. Los resultados de estudios multinacionales que muestran actitudes positivas y alta autoeficacia [5] contrastan con otros que revelan una brecha notable entre la percepción funcional y la comprensión técnica o ética profunda [5]. La escasez de medidas de evaluación objetivas o basadas en el desempeño es una debilidad crítica. Instrumentos como el Generative AI Literacy Assessment Test (GLAT) [12] y las pruebas de conocimiento objetivo para docentes [18] son excepciones notables que señalan un camino a seguir, pero se necesita un esfuerzo mucho mayor para desarrollar y validar tareas auténticas que midan la capacidad demostrada de los estudiantes para usar, criticar y crear con IA de manera responsable.

En segundo lugar, se observa una brecha significativa entre la evaluación de estudiantes y la del personal docente. La literatura se ha centrado predominantemente en la alfabetización de los estudiantes, mientras que la medición de las competencias del profesorado, especialmente la alfabetización pedagógica en IA, sigue estando subdesarrollada. Aunque existen algunas escalas prometedoras para futuros docentes [9-11], la investigación sistemática sobre cómo los docentes en ejercicio integran la IA en su pedagogía, gestionan su uso en la evaluación y apoyan el desarrollo de la alfabetización de sus estudiantes es limitada. Sin una comprensión clara de las competencias y necesidades del profesorado, cualquier esfuerzo por fomentar la alfabetización en IA a nivel institucional será incompleto.

Finalmente, la brecha más profunda es la débil operacionalización de la dimensión sociocultural y cívica. La mayoría de los instrumentos actuales, aunque incluyen subescalas de "ética", tienden a centrarse en cuestiones a nivel individual como la privacidad, la integridad académica y el sesgo en los resultados. Sin embargo, las revisiones conceptuales más amplias abogan por una "alfabetización crítica en IA" que aborde explícitamente las estructuras de

poder, el impacto en el mercado laboral, la justicia global, la gobernanza y la regulación [15], [44-45]. Esta dimensión cívica es crucial para preparar a los estudiantes no solo como usuarios competentes, sino como ciudadanos críticos capaces de participar en el debate público sobre el papel de la IA en la sociedad. La ausencia de esta dimensión en las herramientas de evaluación representa una oportunidad perdida para fomentar una comprensión verdaderamente holística.

C. Implicaciones para la Educación Superior

Los hallazgos de esta revisión tienen implicaciones directas para las políticas y prácticas de las instituciones de educación superior. La evidencia de que los niveles de alfabetización en IA son generalmente moderados [17] y varían significativamente según la disciplina [4], [19] refuta la idea de que los estudiantes, por ser "nativos digitales", desarrollarán estas competencias de forma orgánica. Se requiere una acción institucional proactiva y estructurada.

Desde una perspectiva pedagógica, las universidades deben ir más allá de la simple adopción de herramientas de IA y diseñar programas de formación explícitos. Los instrumentos validados discutidos en esta revisión [3], [8], [46] pueden servir como herramientas de diagnóstico para identificar las fortalezas y debilidades de diferentes cohortes de estudiantes, permitiendo el diseño de intervenciones curriculares específicas por disciplina. Por ejemplo, los estudiantes de humanidades pueden necesitar un mayor énfasis en la comprensión técnica, mientras que los de STEM pueden requerir un desarrollo más profundo del pensamiento crítico y la conciencia ética [19].

A nivel institucional, la falta de políticas claras y marcos éticos sólidos es un desafío recurrente [47-48]. Las universidades deben desarrollar directrices integrales que aborden no solo la integridad académica, sino también el uso responsable de los datos, la equidad en el acceso a las herramientas y el desarrollo profesional continuo del personal docente. La evidencia de que la formación formal en IA, aunque rara, se asocia con una mayor alfabetización [49] subraya la necesidad de invertir en programas de capacitación tanto para estudiantes como para profesores.

D. Limitaciones y Direcciones Futuras para la Investigación

Esta revisión, aunque exhaustiva, está limitada por el rápido ritmo de desarrollo del campo. La literatura sobre la alfabetización en GenAI, en particular, está en su infancia. No obstante, las brechas identificadas señalan un claro camino a seguir para la investigación futura. Se proponen las siguientes direcciones prioritarias:

1. Desarrollo y validación de evaluaciones basadas en el desempeño: Es imperativo complementar las medidas de autoinforme con tareas auténticas que evalúen la competencia demostrada. Esto podría incluir la evaluación de la calidad de los prompts, la capacidad para identificar errores y sesgos en

textos generados por IA, o la creación de artefactos utilizando herramientas de IA de manera innovadora y ética.

2. Operacionalización de la dimensión sociocultural: La investigación futura debe centrarse en la creación y validación de subescalas que midan explícitamente la comprensión de los estudiantes sobre el impacto de la IA en la gobernanza, el trabajo, la equidad global y la justicia social, en línea con los marcos de alfabetización crítica en IA [15], [44].

3. Estudios longitudinales y de intervención: La mayoría de los estudios actuales son de corte transversal. Se necesitan estudios longitudinales para rastrear cómo se desarrolla la alfabetización en IA a lo largo de un programa de grado y estudios experimentales o cuasiexperimentales para evaluar rigurosamente la eficacia de diferentes intervenciones pedagógicas.

4. Investigación centrada en el personal docente: Es crucial un mayor enfoque en el desarrollo de instrumentos robustos para medir la alfabetización pedagógica en IA y en la evaluación de programas de desarrollo profesional para el profesorado.

5. Validación transcultural de instrumentos: Para permitir comparaciones globales significativas, es necesario establecer la invarianza de medida de los instrumentos clave a través de diferentes contextos culturales, lingüísticos y disciplinarios.

En conclusión, el campo de la evaluación de la alfabetización en IA ha logrado avances considerables en la conceptualización y medición de las competencias estudiantiles. Sin embargo, para que las universidades puedan preparar eficazmente a los estudiantes para un futuro mediado por la IA, la investigación y la práctica deben abordar las brechas metodológicas existentes, ampliando el enfoque para incluir evaluaciones basadas en el desempeño, una comprensión más profunda de las dimensiones socioculturales y un compromiso sostenido con el desarrollo profesional del personal docente.

V. CONCLUSIONES

Esta revisión crítica ha sintetizado el estado actual de la investigación sobre la evaluación de la alfabetización en inteligencia artificial (IA) en la educación superior, revelando un campo en rápida maduración pero con brechas significativas. La literatura demuestra un consenso emergente en la conceptualización de la alfabetización en IA como un constructo multidimensional, que abarca consistentemente el conocimiento conceptual, las habilidades operativas, el pensamiento crítico y la competencia ética [7], [9], [50]. Este marco conceptual se ha materializado en una notable proliferación de instrumentos de autoinforme psicométricamente robustos, validados en diversos contextos geográficos y disciplinarios, como el MAIL-CS en China [46], la escala CAIL en Italia [3] y otros instrumentos específicos para Vietnam [8], Corea [51] y Tailandia [10]. Estos esfuerzos han permitido establecer que, en general, los estudiantes universitarios poseen un nivel de alfabetización en IA de

moderado a fundamental, con actitudes mayoritariamente positivas hacia la tecnología [17].

Sin embargo, los hallazgos empíricos también exponen una desafío central: a pesar de la autoeficacia y el uso funcional de herramientas de IA, existe una brecha persistente entre la competencia operativa y una comprensión profunda de los mecanismos técnicos, las implicaciones éticas y el impacto sociocultural de la IA [14]. Esta disonancia es particularmente pronunciada en el contexto de la IA generativa (GenAI), donde la facilidad de uso puede enmascarar una falta de pensamiento crítico sobre la fiabilidad, los sesgos y la procedencia de la información generada [52]. La investigación ha comenzado a abordar esta especificidad con marcos y herramientas de evaluación centrados en la GenAI, como el GLAT [12], que se enfocan en competencias como la ingeniería de prompts, la evaluación crítica de resultados y la gestión de la agencia humana en la colaboración con la IA [2].

Las implicaciones de esta revisión para las instituciones de educación superior son profundas. En primer lugar, la disponibilidad de instrumentos validados ofrece una base sólida para que las universidades realicen diagnósticos institucionales, identifiquen brechas de competencia por disciplina [4] y diseñen intervenciones curriculares específicas. La evidencia de que la alfabetización en IA predice positivamente el rendimiento académico [38], la adopción responsable de herramientas [53] y una mayor conciencia ética [54] subraya la urgencia de integrar esta formación de manera transversal. Las políticas institucionales deben trascender la simple regulación del uso de la IA para abordar proactivamente el desarrollo de competencias, estableciendo marcos éticos claros y fomentando una cultura de integridad académica adaptada a la era de la IA [47].

A pesar de los avances, la investigación presenta limitaciones claras que delinean una agenda para el futuro. La principal es la predominancia de las escalas de autoinforme. Si bien son escalables, son susceptibles a sesgos de sobreconfianza y deseabilidad social. Es imperativo avanzar hacia una triangulación metodológica, incorporando más evaluaciones basadas en el desempeño y pruebas de conocimiento objetivo, como los escasos ejemplos existentes [18-19], para obtener una medida más precisa de la competencia real. En segundo lugar, la dimensión sociocultural y cívica de la alfabetización en IA —que abarca la comprensión de la gobernanza, el poder, el impacto laboral y la justicia algorítmica— está conceptualmente reconocida en revisiones teóricas [15], [55] pero raramente operacionalizada en los instrumentos de evaluación empírica. Integrar esta dimensión es crucial para formar ciudadanos capaces de participar críticamente en el debate público sobre la IA.

Finalmente, la investigación futura debe priorizar la evaluación rigurosa de las intervenciones pedagógicas. La mayoría de los estudios actuales son de naturaleza diagnóstica y transversal. Se necesitan con urgencia estudios longitudinales y experimentales que midan la eficacia de diferentes enfoques curriculares (integrados vs. cursos

dedicados) y modalidades pedagógicas para fomentar una alfabetización en IA profunda y duradera. Asimismo, la investigación sobre la alfabetización en IA del personal docente, aunque incipiente [11], [20], es fundamental. El desarrollo profesional de los educadores no solo es clave para su propia competencia, sino también para modelar y guiar eficazmente el desarrollo de la alfabetización en IA de sus estudiantes. En conclusión, si bien se han sentado las bases para medir y comprender la alfabetización en IA en la universidad, el próximo desafío consiste en pasar del diagnóstico a la acción transformadora, garantizando que tanto estudiantes como educadores estén equipados no solo para usar la IA, sino para comprenderla, criticarla y darle forma de manera responsable.

REFERENCIAS

- [1] Y. Zhou, S. Curle and J. Zou, "Mapping GenAI Literacy: Disciplinary Differences, Latent Profiles, and Perceptions Among EMI Undergraduates", *J. Inf. Technol. Educ. Res.*, 24, 44, 2025.
- [2] Y. Jin, R. Martínez-Maldonado, D. Gašević and L. Yan, "The Agency Gap: The Role of Generative AI Literacy in Human-AI Co-Writing," en *Proceedings of the 14th Learning Analytics and Knowledge Conference*, 2024, pp. 1–12.
- [3] G. Biagini, M. Ranieri and S. Cuomo, "Developing and validating a critical artificial intelligence literacy scale for higher education," *Computers and Education: Artificial Intelligence*, vol. 8, p. 100233, 2025, doi: 10.1016/j.caeai.2025.100233.
- [4] A. Mansoor, A. A. Al-dhaqm and M. A. Ghaleb, "Artificial intelligence literacy among university students—a comparative transnational survey," *Education and Information Technologies*, vol. 29, no. 10, pp. 13087–13111, 2024.
- [5] D. Long and B. Magerko, "What is AI Literacy? Competencies and Design Considerations," in *Proc. 2020 CHI Conf. Hum. Factors Comput. Syst.*, 2020, pp. 1–16, doi: 10.1145/3313831.3376727.
- [6] D. T. L. Ng, S. K. W. Chu, S. S. Q. Chan, M. S. Q. Lau, and T. K. F. Chiu, "Conceptualizing AI literacy: An exploratory review," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100041, 2021.
- [7] M. Ranieri, G. Biagini and S. Cuomo, "AI Literacy in Higher Education," *International Journal of Digital Literacy and Digital Competence*, vol. 16, no. 1, pp. 1–16, sep. 2025, doi: 10.4018/ijdlc.388469
- [8] H. V. Do, T. T. Bui, D. H. Tran, H. T. Nguyen and L. D. Dinh, "Developing and validating an AI literacy scale for university students in Vietnam," *IFLA Journal*, p. 034003522513850, oct. 2025, doi: 10.1177/03400352251385058.
- [9] N. R. Sabando-García, J. S. Moreira-Choez, A. A. Bazurto-Loor and A. B. Bravo-Madrid, "Alfabetización en inteligencia artificial en modalidades de estudio presencial y virtual entre estudiantes universitarios," *Social & Educational Lens*, vol. 4, no. e4, dic. 2024, doi: 10.56931/sel.2024.e4.
- [10] P. Pingmuang and J. Khlaisang, "Development of an artificial intelligence literacy scale for Thai pre-service teachers," *International Journal of Information and Education Technology*, vol. 15, no. 2, pp. 245–254, 2025, doi: 10.18178/ijiet.2025.15.2.2201.
- [11] J. Lim and H. Lee, "Development of a pre-secondary school teachers' AI literacy scale," *The Journal of Korean Association of Computer Education*, vol. 26, no. 2, pp. 11–25, 2023.
- [12] Y. Jin, R. Martínez-Maldonado, D. Gašević and L. Yan, "GLAT: The Generative AI Literacy Assessment Test," *Computers and Education: Artificial Intelligence*, p. 100251, nov. 2024, doi: 10.1016/j.caeai.2024.100251.
- [13] P. M. Kruger and D. Dunning, "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments," *Journal of Personality and Social Psychology*, vol. 77, no. 6, pp. 1121–1134, 1999.
- [14] M. C. Laupichler, J. Aster, M. Mergen and T. Raupach, "Artificial intelligence literacy in higher and adult education: A scoping literature review," *Computers and Education: Artificial Intelligence*, vol. 3, p. 100101, 2022, doi: 10.1016/j.caeai.2022.100101.
- [15] X. Gu and B. Ericson, "AI Literacy in K-12 and Higher Education in the Wake of Generative AI: An Integrative Review," en *Proceedings of the 2025 ACM Conference on International Computing Education Research V.1*, 2025.
- [16] A. Wolters, A. F. Straussenburg and D.M. Riehle, "AI Literacy in Adult Education - A Literature Review". *Hawaii International Conference on System Sciences*, 2024.
- [17] M. Hornberger, A. Bewersdorff, D. Schiff and C. Nerdel, "A multinational assessment of AI literacy among university students in Germany, the UK, and the US," *Computers in Human Behavior: Artificial Humans*, vol. 2, p. 100132, feb. 2025, doi: 10.1016/j.chbah.2025.100132.
- [18] Y. Ding, J. B. Sturges and I. H. Yeter, "Development and validation of an objective AI literacy assessment for non-technical individuals," *Computers and Education: Artificial Intelligence*, vol. 7, p. 100212, 2024.
- [19] D. Rupnik and S. Avsec, "Toward a Coherent AI Literacy Pathway in Technology Education: Bibliometric Synthesis and Cross-Sectional Assessment," *Education Sciences*, vol. 15, no. 1, p. 100, 2025.
- [20] D. Lérias, A. P. Lopes, and C. Rebelo, "AI Literacy among University Teachers: A Case Study at a Portuguese Polytechnic," *Education Sciences*, vol. 14, no. 4, p. 358, 2024, doi: 10.3390/educsci14040358.
- [21] M. Krause, M. Z. Hossain and S. R. R. Abidin, "Generative AI skills among university students: a comparative study of Croatia, India, and Slovenia," *Issues in Information Systems*, vol. 25, no. 2, pp. 1–14, 2024.
- [22] J.L. Callahan, "Constructing a Manuscript: Distinguishing Integrative Literature Reviews and Conceptual and Theory Articles". *Human Resource Development Review*, 9, 300 – 304, 2010.
- [23] E.D. Rivas Briceño and M.A.Valdivia Pinto O, "El múltimétodo en el quehacer investigativo de las Ciencias Sociales ¿alternativa o paradoja?" *Revista EDUCARE - UPEL-IPB - Segunda Nueva Etapa 2.0*, 2022.
- [24] S.M. Barrios, "El enfoque múltimétodo como opción para abordar la investigación educativa". *GACETA DE PEDAGOGÍA*, 2021.J.-G. Lu, "Title of paper with only the first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [25] M. Gessler and C. Siemer, "Umbrella review: Methodological review of reviews published in peer-reviewed journals with a substantial focus on vocational education and training research". *International Journal for Research in Vocational Education and Training*, 2020.
- [26] H.F. Ávila, M.M. González and S.M. Licea, "La triangulación metodológica como método de la investigación científica: Apuntes para una conceptualización." *Didasc@lia: Didáctica y Educación*, 2019, 10, 137-146.
- [27] L.B. Salinas, "El método de la triangulación para estudiar los efectos de la comunicación sanitaria", 2016.
- [28] S. Hughes, T.E. Davis and S.N. Imenda, "Demystifying Theoretical and Conceptual Frameworks: A Guide for Students and Advisors of Educational Research." *Journal of the Social Sciences*, 58(1–3), 2019, doi: 10.31901/24566756.2019/58.1-3.2188
- [29] Y. Zhou, Y. Liu, Z. Wang and H. H. Yang, "Development and validation of the multidimensional AI literacy competency scale for university students in the generative AI era (MAIL-CS)," *The Internet and Higher Education*, vol. 66, p. 101040, jul. 2025, doi: 10.1016/j.iheduc.2025.101040.
- [30] B. Wang, Y. Yuan and T. Yuan, "Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale," *Behaviour & Information Technology*, vol. 42, no. 13, pp. 2321–2334, oct. 2023, doi: 10.1080/0144929X.2022.2126917.
- [31] Y. Chen, X. Chen and J. Zhang, "Exploring the five dimensions of AI literacy for academic English writing in the context of generative AI," *Journal of English for Academic Purposes*, vol. 75, p. 101490, 2025.
- [32] M. A. Al-Shboul, "AICOS: AI Competency Objective Scale," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 20, no. 01, pp. 219–231, 2025.
- [33] A. Bewersdorff, M. Hornberger, D. Schiff and C. Nerdel, "Mind the gap: University students' functional perception of AI versus their technical and ethical understanding," *Computers and Education: Artificial Intelligence*, vol. 6, p. 100189, 2024

- [34] M. A. Ali and S. Mughari, "Effect of AI literacy on online information search competencies among medical students in Pakistan," *Information Development*, p. 026666692412351, 2024.
- [35] M. Syukur, A. R. D. Andriani and A. F. O. Pasaribu, "How does AI literacy affect students' ethical awareness? The mediating role of critical appraisal," *Cogent Education*, vol. 12, no. 1, 2025, doi: 10.1080/2331186X.2024.2421381.
- [36] M. Z. Hossain, M. M. Islam and N. T. Fariha, "Assessing the artificial intelligence literacy level of library and information science students in South Asia," *IFLA Journal*, p. 034003522412921, dic. 2024, doi: 10.1177/03400352241292190.
- [37] Y. Shi, Y. Liu and H. H. Yang, "How AI literacy and self-regulated learning affect students' writing performance and digital well-being in GAI-supported writing," *Assessing Writing*, vol. 63, p. 100898, 2025.
- [38] J. Xiao, S. F. Behzadpoor, y H. H. Yang, "How AI Literacy Affects Students' Educational Attainment in Online Learning: Testing a Structural Equation Model in Higher Education Context," *The International Review of Research in Open and Distributed Learning*, vol. 25, no. 2, pp. 1–19, 2024.
- [39] Y. Zhang, Y. Liu and H. H. Yang, "Integrating AI literacy into the TPB-TAM model to investigate students' GenAI adoption in higher education," *Education and Information Technologies*, 2025.
- [40] A. A. Bazurto-Loor, J. S. Moreira-Choez and N. R. Sabando-García, "Uso de la inteligencia artificial y su correlación con las competencias digitales en estudiantes universitarios," *Revista de Ciencias Sociales*, vol. 30, pp. 343–356, 2024.
- [41] A. Hershkovitz, N. A. Berg and A. Cohen, "A task-centered generative AI literacy framework for higher-education students," *Education and Information Technologies*, nov. 2024, doi: 10.1007/s10639-024-13101-y.
- [42] X. Zhou, Y. Li, and J. Wang, "Mapping the Landscape of Generative AI Literacy in EMI Undergraduates: A Mixed-Methods Study on Disciplinary Differences and Latent Profiles," *Journal of English for Academic Purposes*, 2025.
- [43] Y. Chen, J. Li, and H. Wang, "From Application to Agency: A Five-Dimensional Framework for AI Literacy in Academic English Writing," *System*, 2025.
- [44] M. C. Laupichler, M. C. Wever, and T. Raupach, "Artificial intelligence literacy in higher and adult education: A scoping literature review," *Comput. Educ. Artif. Intell.*, vol. 3, p. 100052, 2022, doi: 10.1016/j.caeai.2022.100052.
- [45] L. Goosen, "AI Literacy Design Matrix (AILDM): A Framework for Curriculum Design in the Age of Artificial Intelligence," *International Journal of Information and Education Technology*, 2025.
- [46] Y. Liu, Y. Zhang, J. Li, and H. Li, "Development and validation of the Multidimensional AI Literacy Competency Scale for university students in the generative AI era (MAIL-CS)," *Computers & Education*, vol. 221, p. 105010, 2025, doi: 10.1016/j.compedu.2024.105010.
- [47] S. Cuomo, G. Biagini, and M. Ranieri, "AI Literacy: Challenges and Opportunities for Higher Education," *QWERTY - Open and Interdisciplinary Journal of Technology, Culture and Education*, 2023.
- [48] M. A. Al-Rabaani and A. M. Al-Maawali, "Concerns of university students about the use of artificial intelligence applications: A descriptive study," *Journal of Educational and Psychological Sciences*, 2024.
- [49] C. Spirgi and S. Seufert, "Generative AI as a learning assistant: How AI competence predicts usage for cognitive and metacognitive learning strategies," *British Journal of Educational Technology*, 2025.
- [50] D. Ng, S. K. Chu, K. F. T. Chan, and M. S. Qiao, "Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach," *Br. J. Educ. Technol.*, vol. 54, no. 5, pp. 1397–1418, 2023, doi: 10.1111/bjet.13339.
- [51] S. H. Hwang, S. Y. Lee and y H. S. Hwang, "Development and validation of an artificial intelligence literacy scale for Korean university students," *Journal of Digital Convergence*, vol. 22, no. 4, pp. 1–10, abr. 2024.
- [52] J. Kim, N. Li, and S. Lee, "Exploring students' perspectives on Generative AI-assisted academic writing," *Education and Information Technologies*, vol. 29, pp. 1–25, 2024, doi: 10.1007/s10639-024-12589-5.
- [53] Y. Zhang, Y. Wang, and H. Wu, "Integrating artificial intelligence literacy into a TPB-TAM framework to explain university students' generative AI adoption," *Heliyon*, vol. 10, no. 10, p. e29988, 2024, doi: 10.1016/j.heliyon.2024.e29988.
- [54] M. Syukur, S. Suwardi, and A. Arifin, "The role of critical appraisal in mediating the effect of AI literacy on students' ethical awareness," *International Journal of Educational Technology in Higher Education*, vol. 22, no. 1, p. 15, 2025.
- [55] L. Goosen, "AI Literacy Design Matrix (AILDM): A Framework for Curriculum Design in the Age of Artificial Intelligence," *International Journal of Information and Education Technology*, 2025.