CAPÍTULO 1

Inteligencia artificial generativa en la evaluación de la formación docente universitaria: retos éticos e innovación

Generative artificial intelligence in university teacher education assessment: ethical challenges and innovation

1 **Oscar Antonio Martínez Molina**
Doctor en Educación
Universidad Nacional de Educación, Ecuador
oscar.martinez@unae.edu.ec
https://orcid.org/0000-0003-1123-5553

2 **Odalys Fraga Luque**
Doctora en Ciencias Pedagógicas
Universidad Nacional de Educación, Ecuador
ofragaluque@gmail.com
https://orcid.org/0000-0002-4657-8276

3 **Ana Mari Pimentel Garriga**
Magíster en Ciencias de la Educación
Universidad Nacional de Educación, Ecuador
apimentelgarriga@yahoo.es
https://orcid.org/0000-0003-4368-4863

4 **Richard Antonio Martínez Villegas**
Magíster en Pedagogía con mención en Docencia e Innovación Educativa
Instituto Superior Tecnológico Particular Sudamericano, Ecuador
rimartinez@sudamericano.edu.ec
https://orcid.org/0000-0001-6655-9972

DOI https://doi.org/10.64644/9789907956313_1 RECIBIDO 05/08/2026 | ACEPTADO 16/10/2026

RESUMEN

La inteligencia artificial generativa (IAG) ha cambiado con rapidez la manera de escribir, retroalimentar y evaluar en la educación superior. Este avance abre posibilidades valiosas, aunque también plantea dudas sobre autoría, privacidad, sesgos y autonomía profesional. El capítulo analiza la evidencia reciente sobre su uso en la evaluación de la formación docente universitaria, con especial atención a América Latina. Se realizó una revisión integrativa de 21 documentos (17 artículos revisados por pares y cuatro marcos o informes institucionales) publicados entre 2023 y 2026, localizados en ERIC, SciELO, Redalyc y DOAJ, además de portales editoriales y rastreo de referencias. La síntesis temática permitió reconocer cinco núcleos: transformación evaluativa, retroalimentación formativa, competencias docentes, integridad académica y gobernanza institucional. Los resultados muestran beneficios para diseñar instrumentos, mejorar la oportunidad de los comentarios y favorecer la reflexión sobre el desempeño; sin embargo, la utilidad pedagógica disminuye cuando la tecnología reemplaza el razonamiento, oculta la participación del autor o interviene en decisiones sin control profesional. A partir de esta evidencia se formula el marco CUIDA, una propuesta teórica que reúne criterio pedagógico, uso transparente, integridad académica, datos protegidos y agencia humana, cuya validez y eficacia deberán comprobarse empíricamente. El capítulo concluye con recomendaciones para las instituciones y líneas de investigación futura.

Palabras clave: inteligencia artificial generativa, evaluación educativa, formación docente, ética digital, educación superior

ABSTRACT

Generative artificial intelligence (GenAI) has rapidly changed writing, feedback, and assessment practices in higher education. These developments create valuable opportunities, but they also raise concerns about authorship, privacy, bias, and professional autonomy. This chapter examines recent evidence on the use of GenAI in university teacher education assessment, with particular attention to Latin America. An integrative review covered 21 documents (17 peer-reviewed articles and four institutional frameworks or reports) published between 2023 and 2026, identified through ERIC, SciELO, Redalyc, DOAJ, publishers' portals, and reference tracking. Thematic synthesis revealed five areas: assessment transformation, formative feedback, teacher competencies, academic integrity, and institutional governance. The evidence suggests that GenAI may support assessment design, provide more timely feedback, and encourage reflection on performance. Its educational value declines, however, when it replaces reasoning, conceals authorship, or influences decisions without professional oversight. Based on this evidence, the chapter proposes the CUIDA framework, a theoretical proposal that brings together pedagogical judgement, transparent use, academic integrity, data protection, and human agency; its validity and effectiveness remain to be tested empirically. These principles are particularly relevant in institutions with unequal access, limited infrastructure, and diverse cultural and linguistic contexts. The chapter closes with recommendations for institutions and directions for future research.

Keywords: generative artificial intelligence, educational assessment, teacher education, digital ethics, higher education

1. Introducción

Los sistemas de inteligencia artificial generativa (IAG) capaces de producir textos, imágenes, códigos y respuestas complejas en pocos segundos modificaron prácticas universitarias que parecían estables. Los datos recientes de adopción permiten dimensionar el cambio. En una encuesta del Digital Education Council (2024) a más de 3800 estudiantes de 16 países, el 86 % declaró usar inteligencia artificial en sus estudios y el 54 % lo hacía cada semana. En el Reino Unido, el 94 % de los estudiantes de pregrado encuestados empleó IAG en trabajos evaluados y el 12 % incorporó texto generado de forma directa, frente al 3 % registrado en 2024 (Stephenson & Armstrong, 2026).

Estas cifras proceden de encuestas de autorreporte y no informan por sí solas sobre la calidad del aprendizaje, pero muestran que la IAG ya forma parte de la vida académica. En el campo de la evaluación, el cambio resulta evidente: una tarea escrita ya no prueba por sí sola que el estudiante comprendió el tema, tomó decisiones y construyó un argumento propio. Por esa razón, la discusión no puede reducirse a permitir o prohibir la IAG. La cuestión central consiste en definir qué aprendizaje se espera, qué apoyo tecnológico resulta legítimo y qué evidencias permiten reconocer la intervención intelectual de cada persona (Weng et al., 2024; Xia et al., 2024).

En un primer momento, muchas instituciones respondieron con advertencias, prohibiciones y programas de detección. Esa reacción era comprensible, aunque pronto mostró límites técnicos y pedagógicos, porque puso el énfasis en controlar el producto y no en comprender el proceso. Luo (2024) señala que varias políticas universitarias tratan la inteligencia artificial como un problema individual de originalidad, pese a que el reto también afecta el diseño de las tareas, la

alfabetización institucional y el acompañamiento docente. Por ello, la evaluación requiere normas explícitas y actividades que obliguen a interpretar, contrastar, decidir y defender el trabajo presentado.

Este debate tiene un peso especial en la formación docente. Quien se prepara para enseñar necesita dominar contenidos, pero también observar desempeños, ofrecer retroalimentación, justificar decisiones y responder ante la diversidad. Una incorporación sin orientación puede reducir el trabajo intelectual a la aceptación de respuestas verosímiles. En cambio, la IAG puede aportar cuando funciona como objeto de análisis, apoyo provisional o contraparte crítica. Bajo esas condiciones, abre espacios para ensayar, revisar y reflexionar sobre la práctica profesional, sin desplazar el criterio pedagógico de quien aprende ni de quien evalúa (Muñoz-Sepúlveda et al., 2026; Ossa & Willatt, 2023).

La dimensión ética atraviesa toda esta discusión. Los modelos generativos procesan grandes volúmenes de datos, pero sus respuestas no garantizan exactitud, neutralidad ni pertinencia cultural. Asimismo, cargar trabajos estudiantiles, casos escolares o información sensible en plataformas comerciales puede comprometer la privacidad y la propiedad intelectual. UNESCO (2023) propone una orientación centrada en las personas. De forma complementaria, su marco de competencias docentes reúne agencia humana, ética, fundamentos de inteligencia artificial, pedagogía y aprendizaje profesional como dimensiones que deben avanzar de manera articulada (UNESCO, 2024).

En América Latina, el problema adquiere matices propios. La región concentra el 14 % de las visitas mundiales a soluciones de inteligencia artificial, frente al 11 % de los usuarios de internet, pero recibe solo el 1,12 % de la inversión global en este campo, aunque representa el 6,6 % del producto interno bruto mundial (CEPAL, 2025). A esa asimetría se suman desigualdades en conectividad, licencias, dominio lingüístico y soporte institucional, que determinan quién utiliza herramientas avanzadas y quién recibe formación oportuna. Betti Galasso (2026) advierte que las competencias emergentes deben comprenderse como prácticas situadas, de modo que la justicia epistémica y la pertinencia cultural son condiciones para una integración legítima.

La producción reciente aporta resultados útiles, pero dispersos: revisiones sobre inteligencia artificial en educación superior, análisis de políticas, escalas para declarar su uso, experiencias de retroalimentación y programas de capacitación docente. Bond et al. (2024) reclaman mayor rigor, colaboración y sensibilidad ética en este campo, y Qian (2025) identifica una tensión persistente entre la eficiencia de estas herramientas y el riesgo de trasladar fuera del estudiante habilidades que debería desarrollar.

Falta, entonces, una síntesis que articule las posibilidades técnicas de la IAG con los principios de una evaluación formativa, válida y responsable en la formación profesional del profesorado. La rapidez de la automatización no garantiza calidad, y una calificación aparentemente precisa puede resultar injusta cuando el sistema desconoce la lengua, el contexto o las condiciones reales en las que se produjo el trabajo.

El aporte original de este capítulo es el marco CUIDA, una propuesta teórica de cinco dimensiones (criterio pedagógico, uso transparente, integridad académica, datos protegidos y agencia humana) derivada de la comparación crítica de las categorías documentales de la revisión. No procede de una intervención ni de una validación empírica, y su eficacia deberá comprobarse en investigaciones posteriores.

El vacío al que responde es de articulación. Las escalas de niveles de uso (Perkins et al., 2024), los marcos de competencias docentes (UNESCO, 2024) y los marcos de política institucional (Chan, 2023) abordan partes del problema, pero en el corpus revisado no se identificó un instrumento que reúna esas exigencias para diseñar y examinar evaluaciones concretas en la formación docente universitaria. CUIDA propone una matriz de revisión previa a la aplicación de la tarea, con preguntas de diseño, evidencias esperadas y señales de alerta, y presta atención a contextos con desigualdades de acceso y diversidad cultural.

Con base en este panorama, el capítulo tiene como objetivo analizar de manera crítica la evidencia científica y orientadora sobre el uso de la IAG en la evaluación de la formación docente universitaria, con atención a sus aportes pedagógicos, riesgos éticos y posibilidades de innovación en América Latina. La revisión responde a una pregunta central: ¿qué condiciones pedagógicas, éticas e institucionales permiten incorporar la IAG en la evaluación de futuros docentes sin debilitar la autoría, el aprendizaje verificable ni la responsabilidad profesional?

El capítulo se organiza en cinco apartados. Después de esta introducción, la sección 2 describe la metodología; la sección 3 presenta la síntesis por núcleos temáticos; la sección 4 discute las convergencias, el contraste entre la evidencia favorable y la crítica, el marco CUIDA y sus implicaciones institucionales; y la sección 5 reúne las conclusiones, las recomendaciones y las líneas futuras.

2. Metodología de la revisión

El estudio se desarrolló como una revisión integrativa, modalidad que permite reunir investigaciones empíricas, revisiones y documentos orientadores mediante procedimientos explícitos de búsqueda, valoración y síntesis (Whittemore & Knafl, 2005). Esta elección se ajustó al carácter reciente y multidimensional del problema, en el cual convergen evaluación, formación docente, ética y gobernanza.

A diferencia de una revisión narrativa abierta, el proceso definió criterios de elegibilidad, una matriz de valoración, un registro de extracción y una síntesis temática que permite seguir el recorrido desde las fuentes hasta las interpretaciones. Para documentar el procedimiento se tomó como guía de reporte la declaración PRISMA 2020 (Page et al., 2021), con las adaptaciones que exige un diseño integrativo. Las limitaciones de registro que subsisten se declaran en los apartados 2.6, 2.7 y 4.5.

La unidad de análisis estuvo formada por documentos académicos y técnicos que abordaron de manera sustantiva la IAG en educación superior, evaluación o formación docente. El periodo cubrió

desde enero de 2023 hasta el 27 de julio de 2026, fecha fijada como cierre documental. Este recorte respondió a un cambio histórico concreto: la difusión pública de ChatGPT a finales de 2022 abrió una etapa distinta en la producción científica. Así, la revisión se concentró en modelos generativos contemporáneos y excluyó literatura previa sobre inteligencia artificial educativa que no examinaba sus capacidades actuales.

2.1. Estrategia de búsqueda

La búsqueda documental cerró el 27 de julio de 2026. Se emplearon descriptores en español e inglés, combinados con los operadores AND y OR. En español, las cadenas reunieron los términos "inteligencia artificial generativa", "IA generativa" y ChatGPT con "evaluación", "retroalimentación", "integridad académica", "formación docente" y "educación superior". Para la búsqueda en inglés se usaron generative artificial intelligence, generative AI, assessment, feedback, academic integrity, teacher education y higher education. Esta selección respondió de forma directa a la pregunta y al objetivo del capítulo.

Cada ecuación se ajustó a la sintaxis y a las opciones de filtrado de la fuente correspondiente. Además, se revisaron las referencias de los trabajos más pertinentes y las publicaciones posteriores que los citaron. Esta doble ruta permitió localizar experiencias latinoamericanas que no siempre aparecen en los primeros resultados de las bases internacionales. Los términos relacionados con educación básica solo se admitieron cuando el documento examinó competencias docentes con aplicación posible en la educación superior. De este modo, la amplitud de la búsqueda no debilitó la coherencia temática del corpus.

Para asegurar reproducibilidad, la ficha de búsqueda debe conservar, por cada fuente, la ecuación literal ejecutada, la fecha, el intervalo temporal, los filtros aplicados y el número de registros recuperados. En la documentación disponible se conservan descriptores, fuentes, fecha de cierre y corpus definitivo, pero no las consultas exportadas ni los conteos iniciales desagregados. En consecuencia, estos datos no se atribuyen retrospectivamente a búsquedas que no pueden auditarse; su reconstrucción exige recuperar el historial original de las plataformas (véase el apartado 2.6).

2.2. Bases de datos consultadas

La consulta incluyó ERIC, SciELO, Redalyc y DOAJ. También se revisaron los portales de revistas y editoriales para confirmar la versión de registro, los datos bibliográficos y el DOI. La combinación de fuentes internacionales e iberoamericanas buscó evitar una lectura concentrada exclusivamente en el Norte global. A este conjunto se sumaron documentos de la UNESCO y de la OCDE, debido a su aporte sobre agencia humana, privacidad, competencias docentes y gobernanza educativa. Su inclusión tuvo una función orientadora y se distinguió de la evidencia empírica.

El corpus temático reunió 21 documentos: 17 artículos científicos sometidos a revisión por pares y cuatro informes o marcos institucionales. Todas las publicaciones se sitúan entre 2023 y 2026, por lo que los artículos representan cerca del 81 % del conjunto. Las referencias

metodológicas utilizadas para organizar el diseño y la síntesis se registraron por separado y no forman parte del conteo del corpus. Esta diferenciación permitió mantener trazabilidad entre la evidencia analizada, los documentos orientadores y las fuentes empleadas exclusivamente para sustentar el método.

Además del corpus, el capítulo emplea literatura complementaria. Para responder a las observaciones de la revisión por pares y contrastar los hallazgos, la introducción y las secciones 3.4, 3.5 y 4 incorporan encuestas de adopción, estudios latinoamericanos recientes, comunicados institucionales, una norma y trabajos sobre sesgos, dependencia y validación de contenido.

Estas fuentes se identificaron de manera dirigida durante la revisión posterior a la evaluación por pares, no pasaron por la matriz de valoración, no forman parte de los 21 documentos ni de la síntesis temática, y se señalan expresamente como literatura complementaria cada vez que se emplean. Page et al. (2021) y Malterud et al. (2016) se suman a las referencias metodológicas registradas por separado y tampoco integran el corpus.

2.3. Criterios de inclusión

Se admitieron estudios en español o inglés con texto completo, o con información metodológica suficiente para valorar su aporte. Los documentos debían examinar usos evaluativos de la IAG, cambios en la retroalimentación, integridad académica, competencias docentes o políticas universitarias. También se incluyeron revisiones con procedimientos explícitos y documentos institucionales de alcance internacional. La pertinencia latinoamericana se estableció a partir de la población, la afiliación, el contexto de aplicación o una discusión específica sobre desigualdad, cultura, regulación y formación del profesorado en la región.

2.4. Criterios de exclusión

Se descartaron duplicados, noticias, entradas de blogs, materiales promocionales, documentos sin autoría identificable y publicaciones que solo mencionaban la inteligencia artificial de manera incidental. También quedaron fuera los estudios centrados en predicción administrativa, reconocimiento facial o analítica institucional sin relación con la evaluación del aprendizaje. Los preprints sirvieron únicamente para localizar versiones publicadas. Cuando no se encontró una versión sometida a revisión por pares, el texto no se incorporó al corpus.

2.5. Procedimiento de selección y análisis

La selección avanzó en tres momentos: lectura de títulos y resúmenes, revisión del texto completo y valoración de la calidad. En la depuración se aplicaron los criterios de exclusión definidos previamente, con atención específica a duplicados, materiales no académicos, documentos sin autoría identificable, menciones incidentales y trabajos sin relación sustantiva con la evaluación del aprendizaje o la formación docente.

La valoración final consideró cuatro criterios: pertinencia para la pregunta, claridad metodológica, trazabilidad de los hallazgos y calidad de la fuente. Se aplicó una matriz ad hoc con

valores de cero a dos por criterio y un mínimo de seis puntos sobre ocho. La puntuación no buscó comparar diseños diferentes, sino excluir documentos con información insuficiente o sin un aporte claramente distinguible.

La matriz de extracción registró autoría, año, país o alcance, tipo de estudio, población, uso de IAG, modalidad evaluativa, beneficios, riesgos, recomendaciones y limitaciones. Luego se aplicó una síntesis temática de orientación crítica, con procesos de codificación, agrupación y revisión de categorías (Braun & Clarke, 2021). Los códigos se reunieron en cinco núcleos: transformación de la evaluación, retroalimentación formativa, competencias docentes, integridad académica y gobernanza institucional. El análisis buscó coincidencias, diferencias y vacíos. La cantidad de menciones no se interpretó como prueba de efectividad.

La revisión no estima un efecto agregado, ya que los estudios presentan poblaciones, herramientas, diseños y resultados muy diversos. Su propósito consiste en integrar conocimiento útil para el diseño de evaluaciones y para la formación del profesorado. Esta delimitación evita atribuir relaciones causales a experiencias exploratorias. También permite separar con claridad los resultados empíricos, las orientaciones institucionales y las propuestas conceptuales que aparecen en el corpus (véase la Tabla 1).

Tabla 1

Decisiones metodológicas de la revisión integrativa

Componente	Decisión metodológica	Aplicación en la revisión	Garantía de rigor
Diseño	Revisión integrativa	Artículos empíricos, revisiones y documentos institucionales	Procedimiento explícito y síntesis trazable
Alcance temporal	Enero de 2023 a 27 de julio de 2026	Cobertura del periodo de expansión pública de la IAG	Actualidad y delimitación justificadas
Fuentes	ERIC, SciELO, Redalyc, DOAJ y portales editoriales	Búsqueda bilingüe y rastreo de referencias	Cobertura internacional e iberoamericana
Elegibilidad	Pertinencia, claridad, trazabilidad y calidad de la fuente	Matriz ad hoc de 0 a 2 por criterio; umbral de 6/8	Exclusión de documentos insuficientes o incidentales
Corpus	21 documentos	17 artículos y 4 marcos o informes institucionales	Predominio de evidencia revisada por pares
Reporte	PRISMA 2020 como guía adaptada (Page et al., 2021)	Fases documentadas en la Tabla 2 y la Figura 1	Transparencia sobre lo registrado y lo no conservado
Análisis	Síntesis temática crítica	Cinco núcleos analíticos y contraste de convergencias	Diferenciación entre evidencia, orientación y propuesta
Literatura complementaria	Fuentes adicionales fuera del corpus	Contraste en las secciones 3 y 4, con rótulo expreso	Separación respecto de la síntesis del corpus

Nota. IAG = inteligencia artificial generativa; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Elaboración propia a partir del protocolo de revisión.

2.6. Trazabilidad de la selección documental

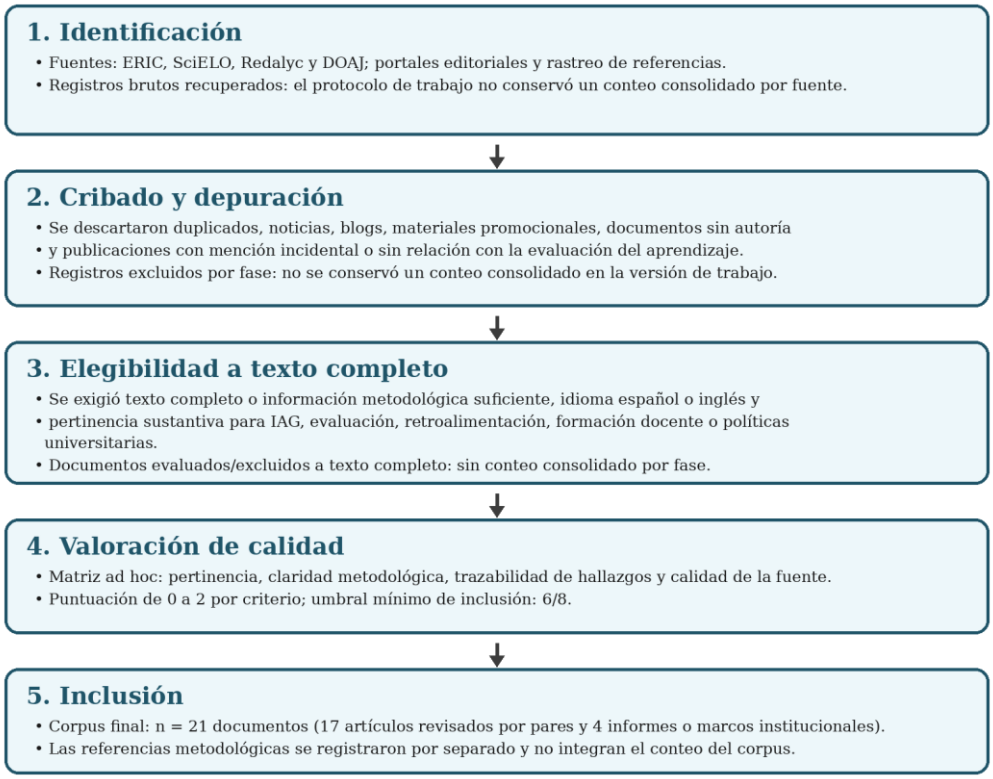
La trazabilidad del corpus se organizó mediante un esquema equivalente a un flujo PRISMA, adaptado al carácter integrativo de la revisión (Page et al., 2021). El proceso distingue identificación, cribado y depuración, elegibilidad a texto completo, valoración de calidad e inclusión. Los motivos de exclusión se derivan de los criterios definidos en los apartados 2.3 y 2.4, y la decisión final se apoya en la matriz de calidad descrita en el apartado 2.5.

Los pares evaluadores solicitaron consignar el número de documentos recuperados en cada base, los excluidos por duplicidad, los descartados tras la lectura de títulos y resúmenes y los excluidos en texto completo, con sus motivos. La documentación conservada permite reportar la composición final del corpus y los criterios aplicados en cada fase, pero no los conteos intermedios, porque no se preservaron las consultas exportadas ni los totales desagregados por base y etapa. Por rigor metodológico, estas cifras no se reconstruyen de memoria ni se estiman de forma retrospectiva.

La Tabla 2 muestra, fase por fase, el procedimiento aplicado, los motivos de exclusión documentados y el estado del registro numérico. La Figura 1 representa esas fases y reporta únicamente el dato cuantitativo verificable de inclusión: 21 documentos.

Figura 1

Esquema de selección documental y conformación del corpus



Nota. El único conteo cuantitativo conservado de manera verificable corresponde al corpus final (n = 21). Elaboración propia a partir del protocolo de revisión.

Tabla 2

Fases de la selección documental, motivos de exclusión y registro disponible

Fase	Procedimiento aplicado	Motivos de exclusión documentados	Registro numérico
Identificación	Búsqueda bilingüe en ERIC, SciELO, Redalyc y DOAJ; revisión de portales editoriales; rastreo de referencias y de citas posteriores (apartados 2.1 y 2.2)	No aplica	No conservado por base
Duplicados	Eliminación de registros repetidos entre bases y rutas de localización (apartado 2.4)	Duplicidad	No conservado
Cribado de títulos y resúmenes	Lectura con los criterios de inclusión y exclusión (apartados 2.3 y 2.4)	Noticias, entradas de blogs, materiales promocionales, documentos sin autoría identificable, menciones incidentales de la inteligencia artificial, predicción administrativa, reconocimiento facial o analítica institucional sin relación con la evaluación del aprendizaje	No conservado
Elegibilidad a texto completo	Revisión del texto completo o de información metodológica suficiente (apartado 2.3)	Preprints sin versión publicada revisada por pares; ausencia de relación sustantiva con la evaluación del aprendizaje o la formación docente	No conservado
Valoración de calidad	Matriz ad hoc de cuatro criterios, de 0 a 2 puntos, con mínimo de 6 sobre 8 (apartado 2.5)	Información insuficiente o sin aporte claramente distinguible	No conservado
Inclusión	Corpus final: 17 artículos revisados por pares y 4 marcos o informes institucionales	No aplica	21 documentos

Nota. "No conservado" indica que el dato no pudo recuperarse de la documentación disponible y no se estima de forma retrospectiva. La asignación de cada motivo a una fase sigue el orden lógico de los apartados 2.3 a 2.5 y no procede de un registro por fase. Elaboración propia a partir del protocolo de revisión.

La trazabilidad horizontal de la revisión se establece del siguiente modo: la pregunta y el objetivo determinan los criterios de elegibilidad; la extracción sistemática permite reconocer cinco núcleos; la comparación de beneficios, riesgos y condiciones institucionales sustenta la discusión; y esa interpretación conduce a los componentes conceptuales de CUIDA (véanse la Figura 3 y la Tabla 5). Por tanto, el modelo sintetiza la evidencia y no debe confundirse con un resultado experimental.

2.7. Suficiencia del corpus y saturación temática

La suficiencia de los 21 documentos se interpreta como adecuación temática al objetivo y a los cinco núcleos de síntesis, no como representatividad estadística de la producción mundial. En las revisiones documentales, la saturación temática suele entenderse como el punto en que nuevos documentos dejan de aportar categorías distintas a las ya codificadas. Este capítulo no puede

afirmar saturación formal, porque no se conservaron los registros de las iteraciones de codificación que permitirían demostrarla.

Como criterio alternativo se recurre, por analogía, al concepto de poder de información (Malterud et al., 2016).

Según este concepto, se necesita menos material cuando es específico para el propósito del estudio, se apoya en teoría establecida, ofrece información de calidad y se analiza con una estrategia focalizada. En este caso, el objetivo es acotado, todos los documentos tratan de manera sustantiva alguno de los cinco núcleos, el análisis se apoya en marcos reconocidos como el de competencias de la UNESCO (2024) y una matriz de valoración fijó un umbral de calidad. El concepto procede de la investigación cualitativa con participantes, por lo que se aplica aquí como orientación y no como prueba.

La muestra responde a la pregunta de investigación en distinto grado según el núcleo (véase la Tabla 5). Los núcleos de transformación evaluativa (seis documentos), competencias docentes (cinco), integridad académica (tres) y gobernanza institucional (dos) cubren las dimensiones pedagógica, ética e institucional de la pregunta, aunque con una base desigual. La retroalimentación formativa descansa en dos estudios sobre formación inicial docente, y tres documentos son transversales. Esa distribución obliga a formular las conclusiones con prudencia.

El tamaño del corpus también puede introducir sesgos. Con pocos documentos por núcleo, la ausencia de perspectivas discrepantes no prueba su inexistencia, y la selección por pertinencia y claridad pudo favorecer trabajos con resultados alentadores. Para atenuar este riesgo, la discusión contrasta los hallazgos con literatura complementaria de carácter crítico, claramente señalada, y las conclusiones se formulan como síntesis argumentativa y no como generalización de eficacia.

3. Desarrollo de la revisión

La síntesis se organiza en cinco núcleos: transformación de la evaluación, retroalimentación formativa, competencias docentes, integridad académica y gobernanza institucional. El contexto latinoamericano aparece como una dimensión transversal, porque condiciona el acceso, la pertinencia cultural y la capacidad de respuesta de cada institución. El último apartado reúne los vacíos de investigación identificados y prepara el paso hacia la discusión. Esta organización permite leer los hallazgos como partes de un mismo problema, sin perder las particularidades de cada tema.

La Tabla 3 resume el país o ámbito, el diseño, el núcleo principal, el aporte y el riesgo señalado en cada uno de los 21 documentos del corpus, de modo que el lector pueda ubicar cada fuente antes de seguir su análisis.

La procedencia es desigual. Cinco de los 17 artículos se centran en contextos latinoamericanos (dos de Chile, dos de Ecuador y uno de Brasil con alcance regional); cinco se vinculan con instituciones de Hong Kong, algunos en colaboración con equipos de China continental, Singapur o Taiwán; dos con Australia y Vietnam; dos con Estados Unidos; uno con Finlandia; uno con España y México; y uno con equipos de Australia, Reino Unido, Noruega y Suecia.

Los cuatro documentos restantes son marcos o informes de la UNESCO y de la OCDE. Predominan, por tanto, las revisiones, los análisis de políticas y los estudios exploratorios, como se examina en el apartado 3.7.

Tabla 3

Corpus de la revisión: procedencia, diseño, núcleo, aporte y riesgo por documento

Fuente	País o ámbito	Diseño o tipo	Núcleo principal	Aporte principal	Riesgo o limitación señalado
Betti Galasso (2026)	Brasil (alcance latinoamericano)	Revisión de alcance	Competencias docentes	Competencias situadas y justicia epistémica	Desigualdad y modelos importados
Bond et al. (2024)	Australia, Reino Unido, Noruega y Suecia	Metarrevisión sistemática	Transversal	Reclama mayor ética, colaboración y rigor en el campo	Rigor y sensibilidad ética insuficientes
Celik (2023)	Finlandia	Estudio empírico con docentes	Competencias docentes	Incorpora la ética al conocimiento profesional docente (TPACK inteligente)	Adopción guiada por la novedad, sin juicio profesional
Chan (2023)	Hong Kong (China)	Marco de política institucional	Gobernanza institucional	Articula dimensiones pedagógicas, éticas y operativas	Orientaciones generales insuficientes para usos por asignatura
Furze et al. (2024)	Australia, Singapur y Vietnam	Implementación piloto de la AIAS	Transformación evaluativa	Una política comprensible reduce dudas y promueve el rediseño de tareas	La escala no reemplaza el análisis pedagógico
Gallent-Torres et al. (2023)	España y México	Análisis desde la ética y la integridad académica	Integridad académica	La IAG amplía problemas de plagio, autoría y veracidad	Presentar como propio un razonamiento no comprendido
Liang et al. (2023)	Estados Unidos	Evaluación de detectores de texto	Integridad académica	Evidencia de sesgo contra escritores no nativos de inglés	Falsos positivos en decisiones de integridad
López-Vasco et al. (2025)	Ecuador y Venezuela	Cuasiexperimento con 299 docentes	Competencias docentes	Mejora técnica y actitudinal tras 80 horas de formación	Persisten dudas sobre originalidad
Luo (2024)	Hong Kong (China)	Revisión crítica de políticas universitarias	Transformación evaluativa	Cuestiona una noción estrecha de originalidad	Enfoque punitivo y vigilancia
Méndez Méndez et al. (2025)	Ecuador	Intervención formativa preexperimental en Cuenca	Competencias docentes	Avances en dimensiones TPACK tras la formación	Diseño sin comparación ni seguimiento
Moorhouse et al. (2023)	Hong Kong y Singapur	Análisis de directrices de universidades de alto ranking	Transformación evaluativa	Muestra orientaciones muy diversas entre universidades	Diferencias en claridad, alcance y responsabilidades

Fuente	País o ámbito	Diseño o tipo	Núcleo principal	Aporte principal	Riesgo o limitación señalado
Muñoz-Sepúlveda et al. (2026)	Chile	Sistematización con simulación y videoanálisis	Retroalimentación formativa	Retroalimentación basada en desempeño	Dependencia de infraestructura y diseño
OECD (2023)	Organismo internacional	Informe internacional	Gobernanza institucional	Destaca la necesidad de reglas flexibles ante una tecnología cambiante	Rigidez normativa ante el cambio rápido
OECD (2026)	Organismo internacional	Síntesis internacional de evidencia	Transversal	La IAG apoya cuando existe intención pedagógica	Mejora del desempeño sin aprendizaje real
Ossa y Willatt (2023)	Chile	347 estudiantes y 578 respuestas a casos	Retroalimentación formativa	Retroalimentación descriptiva a escala con validación docente	Imprecisiones que requieren revisión
Perkins et al. (2024)	Vietnam y Australia	Marco de cinco niveles de uso (AIAS)	Transformación evaluativa	Clarifica grados de integración de la IAG	Aplicación mecánica sin alineación curricular
Qian (2025)	Estados Unidos	Revisión sistemática	Transversal	Identifica la tensión entre eficiencia y riesgo para el aprendizaje	Traslado fuera del estudiante de habilidades que debe desarrollar
UNESCO (2023)	Organismo internacional	Orientación internacional	Integridad académica	Orientación centrada en las personas, con límites de edad, protección de datos y validación pedagógica	Privacidad y sesgos
UNESCO (2024)	Organismo internacional	Marco de competencias docentes	Competencias docentes	Agencia humana, ética y pedagogía	Reducción de la formación a habilidades técnicas
Weng et al. (2024)	Hong Kong y Singapur	Revisión de alcance sobre evaluación y resultados	Transformación evaluativa	Mayor aprendizaje cuando la IAG amplía la construcción de conocimiento	Externalización del esfuerzo cognitivo
Xia et al. (2024)	China, Hong Kong y Taiwán	Revisión de alcance de 32 estudios	Transformación evaluativa	Transformación en los niveles del estudiante, el docente y la institución	Respuestas parciales y políticas fragmentadas

Nota. País o ámbito según la afiliación institucional de la autoría registrada en Crossref y OpenAlex; en autorías con varias afiliaciones se indican los países registrados. El núcleo principal es una asignación reconstruida a partir del contenido de cada documento. AIAS = Artificial Intelligence Assessment Scale; TPACK = conocimiento tecnológico, pedagógico y disciplinar; IAG = inteligencia artificial generativa. Elaboración propia.

3.1. Transformación de la evaluación universitaria

La evidencia revisada coincide en un punto: la IAG no es solo una herramienta adicional, pues también cuestiona la validez de numerosas evaluaciones tradicionales. Cuando un sistema redacta ensayos, resuelve ejercicios o produce argumentos plausibles, las tareas centradas en repetir información pierden capacidad para distinguir entre desempeño propio y apoyo externo. Xia et al. (2024) sostienen que la transformación debe alcanzar al estudiante, al profesorado y a la institución. Ningún cambio aislado ofrece garantías suficientes de aprendizaje responsable.

Weng et al. (2024) encontraron mayor valor educativo cuando la tecnología amplía la construcción de conocimiento en lugar de reemplazar el esfuerzo cognitivo. Esta diferencia obliga a revisar la relación entre tarea, criterio y resultado de aprendizaje. Una evaluación puede admitir el uso de IAG y conservar un alto nivel de exigencia si solicita justificar decisiones, contrastar fuentes, presentar versiones sucesivas y reconocer los límites de la respuesta obtenida. En ese caso, la herramienta forma parte del proceso, pero la comprensión sigue bajo responsabilidad del estudiante.

Las políticas institucionales también muestran un cambio gradual, desde la prohibición hacia niveles diferenciados de uso. Moorhouse et al. (2023) observaron orientaciones muy diversas entre universidades de reconocido prestigio, con diferencias en claridad, alcance y distribución de responsabilidades. Luo (2024) agrega que las normas centradas solo en la originalidad pueden mantener una idea limitada de la autoría. El reto consiste en precisar qué aportes debe realizar la persona, qué apoyos se consideran legítimos y qué registro permitirá reconstruir la interacción entre el estudiante y el sistema.

La Artificial Intelligence Assessment Scale ofrece una respuesta práctica mediante cinco niveles, desde la ausencia de inteligencia artificial hasta su uso pleno y declarado (Perkins et al., 2024). La experiencia piloto mostró que una política comprensible puede reducir dudas y promover el rediseño de tareas (Furze et al., 2024). Sin embargo, la escala no reemplaza el análisis pedagógico. Cada nivel debe corresponderse con los resultados de aprendizaje, las características de la disciplina, el acceso tecnológico y el grado de autonomía del estudiantado.

En la formación docente, la evaluación asume una responsabilidad adicional, pues anticipa prácticas que después pueden reproducirse en escuelas y universidades. Evaluar a futuros profesores mediante productos fáciles de automatizar transmite una visión instrumental del conocimiento profesional. Por el contrario, las simulaciones, los análisis de casos, los portafolios comentados, las defensas orales y las observaciones de desempeño permiten apreciar cómo cada estudiante interpreta situaciones, adapta principios y responde ante dilemas concretos. Estas evidencias ofrecen una base más cercana a la complejidad real de la práctica educativa.

3.2. Retroalimentación formativa y aprendizaje verificable

Una evaluación rediseñada exige también repensar la manera de acompañarla. La posibilidad de ofrecer retroalimentación inmediata y adaptada constituye uno de los aportes más visibles de la IAG. Esta ventaja resulta atractiva en grupos numerosos, donde el tiempo del profesorado suele ser limitado. Sin embargo, la calidad de los comentarios depende de la información disponible, los criterios de evaluación y la revisión humana. Un texto extenso no siempre orienta mejor. Cuando la respuesta es genérica, puede aumentar la carga cognitiva y dejar al estudiante sin una ruta clara para mejorar.

Ossa y Willatt (2023) utilizaron IAG para retroalimentar 578 respuestas a casos pedagógicos elaboradas por 347 estudiantes de formación inicial docente. El proceso se apoyó en criterios definidos y mantuvo supervisión humana, por lo que la herramienta funcionó como asistencia y no como evaluador autónomo. La experiencia muestra que la tecnología adquiere valor cuando existe

una rúbrica clara, ejemplos de calidad y una revisión profesional capaz de identificar imprecisiones o lecturas inadecuadas. Sin estas condiciones, la rapidez de respuesta puede ocultar errores relevantes.

Muñoz-Sepúlveda et al. (2026) sistematizaron un modelo que reúne simulación, videoanálisis, retroalimentación colaborativa e IAG. La propuesta resulta valiosa porque la evidencia principal no se limita a un texto, sino que incluye una actuación profesional observada desde distintas perspectivas. La tecnología ayuda a ordenar comentarios y reconocer patrones. La interpretación, en cambio, permanece vinculada con el diálogo, la experiencia y el criterio docente. Esta distribución de funciones evita que el sistema ocupe el lugar de quien evalúa.

A partir de estas experiencias, la retroalimentación mediada por IAG debería cumplir cuatro condiciones. En primer lugar, necesita criterios visibles y comprensibles. En segundo término, debe señalar evidencias concretas del desempeño. Además, ha de proponer acciones posibles, no recomendaciones vagas. Por último, debe abrir un espacio para que el estudiante responda, cuestione y explique qué hizo con la orientación. Cuando se cumplen estas condiciones, la retroalimentación deja de ser una entrega unilateral y adquiere el carácter de una conversación formativa.

El aprendizaje verificable exige conservar huellas del proceso. Los borradores, las decisiones registradas, los comentarios sobre salidas generadas, las fuentes contrastadas y las defensas breves permiten reconocer la evolución intelectual del estudiante. Estas evidencias no buscan vigilar cada movimiento. Su propósito es ofrecer una base razonable para valorar comprensión, autonomía y responsabilidad. Si la evaluación conserva únicamente el producto final, pierde información decisiva sobre la forma en que se construyó el conocimiento.

3.3. Competencias docentes para una evaluación mediada por inteligencia artificial

3.3.1. Marcos de competencias para la docencia

Ofrecer retroalimentación de calidad depende, en última instancia, de las competencias de quien enseña. Una integración responsable requiere más que destreza para redactar instrucciones. UNESCO (2024) organiza las competencias docentes en cinco dimensiones: perspectiva centrada en las personas, ética, fundamentos y aplicaciones, pedagogía y aprendizaje profesional. Esta estructura evita reducir la alfabetización en inteligencia artificial al manejo técnico de una plataforma.

El docente necesita comprender cómo se producen las respuestas, cuáles son sus límites y en qué momento el uso de la herramienta contradice el propósito educativo. Solo así puede decidir con criterio cuándo incorporarla y cuándo prescindir de ella.

Celik (2023) amplía el conocimiento tecnológico, pedagógico y disciplinar hacia un TPACK inteligente, en el que la ética forma parte de la competencia profesional. Desde esta perspectiva, saber que una herramienta genera rúbricas o comentarios resulta insuficiente. El profesorado debe valorar su adecuación al nivel educativo, reconocer sesgos, proteger datos, revisar la calidad de la

evidencia y anticipar efectos sobre motivación, autonomía y equidad. Esta capacidad de juicio distingue un uso profesional de una adopción guiada solo por la novedad.

La formación docente latinoamericana muestra capacidad de adaptación, aunque también necesidades particulares. Betti Galasso (2026) identifica dimensiones técnico-operacionales, pedagógico-didácticas, ético-críticas, evaluativas y colaborativas. La dimensión evaluativa merece especial atención, porque obliga a redefinir qué significa demostrar aprendizaje cuando una máquina puede producir respuestas semejantes a las humanas. A la vez, la colaboración entre pares ofrece una vía concreta para compartir criterios, revisar errores y construir soluciones ajustadas al contexto.

3.3.2. Evidencia sobre programas de formación

Los programas formativos obtienen mejores resultados cuando combinan conocimiento técnico, práctica guiada y discusión ética. López-Vasco et al. (2025) registraron mejoras en conocimiento y predisposición después de un programa de 80 horas con docentes de universidades ecuatorianas. La percepción sobre originalidad, en cambio, no presentó cambios significativos. Este resultado sugiere que la capacitación operativa no resuelve por sí sola los dilemas de autoría, confianza y valoración académica. La formación necesita abordar esos conflictos de manera explícita.

Méndez Méndez et al. (2025) también observaron avances en dimensiones TPACK tras una intervención formativa en Cuenca. Los resultados son alentadores, aunque proceden de diseños preexperimentales y todavía requieren comparaciones y seguimiento. Una competencia sostenible se reconoce cuando el docente modifica sus instrumentos, explica las decisiones adoptadas y examina el efecto de la tecnología sobre el aprendizaje. Declarar mayor confianza para usar una herramienta no constituye, por sí solo, una evidencia suficiente de cambio profesional.

Por esta razón, la formación debería incluir talleres de rediseño evaluativo, análisis de respuestas erróneas, construcción de rúbricas, revisión de sesgos, protección de datos y elaboración de declaraciones de uso. También conviene abrir espacios donde el profesorado discuta casos reales y observe cómo una misma herramienta cambia sus respuestas según la disciplina, el idioma o el contexto. Esta preparación fortalece el juicio profesional y reduce la dependencia de recetas generales que no siempre responden a las necesidades de cada institución.

3.4. Ética, integridad académica y autoría

3.4.1. Detectores, debido proceso y autoría

Estas competencias incluyen decisiones éticas que afectan de manera directa a los estudiantes. La integridad académica no puede reducirse al uso de detectores. Los clasificadores que buscan identificar textos generados por IAG producen falsos positivos y resultados variables, sobre todo en textos breves, revisados o escritos por personas cuya primera lengua no es el inglés (Liang et al., 2023). Por ello, una acusación basada únicamente en una puntuación automática puede vulnerar el debido proceso y afectar con mayor intensidad a determinados grupos de estudiantes. El juicio humano y la revisión contextual son indispensables.

Gallent-Torres et al. (2023) muestran que la IAG amplía problemas ya conocidos de plagio, autoría y veracidad. Al mismo tiempo, obliga a revisar prácticas institucionales que antes de su aparición ya presentaban fragilidad. El riesgo no se limita a copiar una respuesta; también incluye presentar como propio un razonamiento que no se comprende. Frente a esta situación, la respuesta educativa debe priorizar la declaración del apoyo recibido, la verificación de las fuentes y la capacidad de explicar las decisiones tomadas durante el proceso.

Como literatura complementaria, Navarrete-Cazales et al. (2026) revisaron investigaciones publicadas entre 2019 y 2024 y concluyen que la IAG redefine la autoría humana, dificulta la detección del plagio automatizado y puede ampliar la brecha digital entre instituciones con distinto acceso tecnológico. Estos hallazgos coinciden con los del corpus sobre la fragilidad de los detectores y sobre la necesidad de políticas institucionales que regulen el uso académico de estas herramientas.

La autoría puede entenderse como responsabilidad sobre el producto y sobre el camino seguido para construirlo. Un estudiante conserva esa responsabilidad cuando selecciona el problema, formula criterios, contrasta información, corrige errores y responde por las consecuencias de su trabajo. La IAG puede colaborar, pero no puede asumir responsabilidad moral ni profesional. Esta distinción resulta decisiva en las profesiones educativas, donde las decisiones afectan a otras personas y deben sostenerse en conocimiento pedagógico, evidencia contextual y razones que puedan explicarse.

3.4.2. Privacidad, sesgos y equidad

La privacidad constituye otro punto crítico. Cargar datos de estudiantes, informes de prácticas, historias escolares o grabaciones en servicios externos puede trasladar información a sistemas con políticas poco transparentes. UNESCO (2023) recomienda límites de edad, protección de datos y validación pedagógica. En la formación docente, una regla mínima debería impedir el ingreso de información identificable. Cuando el análisis de casos auténticos resulte necesario, los datos deben anonimizarse antes de utilizar cualquier plataforma externa.

Los sesgos también comprometen la calidad de la evaluación. Un modelo construido con datos dominantes puede reproducir estereotipos relacionados con lengua, género, cultura, capacidad o desempeño. En América Latina, la escasez de contenidos locales y las variaciones del español aumentan el riesgo de respuestas fuera de contexto. La revisión profesional debe buscar errores factuales, pero también supuestos culturales, omisiones y criterios de normalidad que podrían perjudicar a grupos específicos (UNESCO, 2023). El apartado 4.2.1 amplía este punto.

La equidad, por último, depende de las condiciones de acceso. Autorizar herramientas de pago sin ofrecer alternativas institucionales puede convertir la evaluación en una competencia de recursos. Las universidades deben indicar qué versiones están permitidas, qué apoyos se ofrecerán y cómo se evaluará a quienes no puedan o no deseen usarlas. Una innovación pierde legitimidad cuando amplía desigualdades o convierte una tecnología en requisito tácito para alcanzar un mejor desempeño.

3.5. Contexto latinoamericano y transformación digital

3.5.1. Condiciones de acceso y evidencia del corpus

Estas cuestiones éticas y pedagógicas se viven de manera desigual según el contexto. La transformación digital en América Latina ocurre en condiciones muy distintas. Algunas universidades urbanas cuentan con infraestructura sólida, mientras otras enfrentan conectividad irregular, dispositivos compartidos y poco soporte técnico. Esta realidad obliga a evitar modelos que suponen acceso permanente, licencias costosas o dominio avanzado del inglés. Una política responsable debe partir de las capacidades reales de la comunidad y ofrecer alternativas equivalentes para quienes no disponen de los mismos recursos (Betti Galasso, 2026; UNESCO, 2023).

La región también posee fortalezas propias para orientar la innovación. Las tradiciones de pedagogía crítica, educación popular y formación situada ofrecen herramientas para cuestionar la aparente neutralidad tecnológica. Betti Galasso (2026) sitúa la justicia epistémica entre las competencias docentes emergentes. Esto lleva a preguntar qué conocimientos privilegia el sistema, qué voces deja fuera y de qué forma pueden incorporarse experiencias locales en la evaluación. Estas preguntas aportan una mirada que no siempre está presente en los marcos internacionales.

Las experiencias de Chile y Ecuador muestran que la IAG puede apoyar la retroalimentación, las simulaciones y el desarrollo profesional cuando existe mediación docente (López-Vasco et al., 2025; Muñoz-Sepúlveda et al., 2026; Ossa & Willatt, 2023). Sin embargo, estas iniciativas aún son recientes y se concentran en instituciones específicas. Hacen falta estudios multicéntricos que comparen disciplinas, modalidades, regiones y grupos con condiciones de acceso diferentes. Esa evidencia permitiría distinguir los resultados vinculados con la herramienta de aquellos que dependen del diseño o del contexto.

3.5.2. Literatura complementaria sobre universidades latinoamericanas

Para profundizar en la realidad regional se incorporaron estudios recientes sobre universidades latinoamericanas que no forman parte del corpus. Roblez Torres et al. (2026) aplicaron un cuestionario de 30 ítems a 482 estudiantes y 121 docentes de una universidad ecuatoriana y observaron correlaciones positivas y significativas entre evaluación auténtica, trazabilidad del aprendizaje e integridad académica. La trazabilidad, sin embargo, obtuvo la media más baja de los cinco dominios. Este resultado respalda la importancia de conservar huellas del proceso, aunque el diseño correlacional y la muestra por conveniencia no permiten inferir causalidad.

Vallejos-Parás y Galindo-Cuesta (2025) analizaron las preocupaciones éticas de 133 docentes de educación superior de nueve instituciones lasallianas de Brasil, Colombia y México. Los participantes expresaron opiniones mixtas, con experiencias positivas y negativas en el aula, y mantuvieron incertidumbre sobre la aplicación ética de la IAG, aun en países que han impulsado el debate público y marcos regulatorios. Este resultado converge con el de López-Vasco et al. (2025), quienes observaron que la capacitación operativa no resolvió por sí sola los dilemas de originalidad.

Los datos regionales de adopción permiten situar estas experiencias. La CEPAL (2025) señala que la adopción de inteligencia artificial se concentra en pocos países y se orienta sobre todo al consumo de herramientas listas para el usuario final, que la formación avanzada es insuficiente y que muchas estrategias nacionales carecen de financiamiento, mecanismos de implementación y sistemas de evaluación de impacto. El comunicado no aporta cifras de infraestructura ni de acceso en el ámbito universitario, de modo que las brechas de conectividad, dispositivos y licencias siguen documentadas de forma indirecta y requieren estudios propios.

En el plano normativo, en octubre de 2025 la Cámara de Diputadas y Diputados de Chile despachó al segundo trámite legislativo un proyecto marco que regula los sistemas de inteligencia artificial según niveles de riesgo (Cámara de Diputadas y Diputados de Chile, 2025). El comunicado consultado no alude de forma específica a la educación superior, lo que sugiere que las reglas de evaluación académica seguirán dependiendo, en buena medida, de la normativa institucional.

3.5.3. Pertinencia lingüística y transformación digital

La pertinencia lingüística añade otro desafío. Los sistemas pueden producir un español correcto en apariencia y, aun así, desconocer expresiones, normas, marcos curriculares o referencias propias de cada país. La revisión profesional debe comprobar el vocabulario y su relación con la normativa educativa local. En la formación docente, un ejemplo fuera de contexto puede conducir a decisiones equivocadas, aunque su redacción parezca convincente. La corrección formal no sustituye la comprensión cultural ni el conocimiento del sistema educativo.

Por estas razones, la transformación digital no debería medirse por el número de herramientas incorporadas. Su calidad depende de la capacidad institucional para mejorar el aprendizaje, reducir barreras y sostener decisiones transparentes. La IAG cobra sentido cuando amplía el acompañamiento y fortalece la reflexión. Pierde valor cuando reemplaza vínculos educativos o traslada responsabilidades hacia proveedores externos.

3.6. Gobernanza institucional y rediseño evaluativo

Las decisiones individuales de los docentes necesitan un marco institucional que las sostenga. Las universidades necesitan una gobernanza capaz de conectar política, currículo, infraestructura y desarrollo profesional. Las orientaciones generales resultan insuficientes porque cada asignatura enfrenta usos diferentes. Una tarea de escritura, una simulación clínica y una práctica docente requieren criterios propios. Chan (2023) propone marcos de política educativa que articulan dimensiones pedagógicas, éticas y operativas. La OCDE, por su parte, destaca la necesidad de mantener reglas flexibles ante una tecnología que cambia con rapidez (OECD, 2023).

La política institucional debe diferenciar el uso prohibido, opcional, requerido y sujeto a declaración. También debe aclarar responsabilidades frente a errores, protección de datos, propiedad intelectual y mecanismos de apelación. Estas definiciones necesitan aparecer en el sílabo y en cada consigna, no solo en un reglamento general. La claridad reduce conflictos y permite que el estudiante organice su trabajo con expectativas conocidas desde el inicio.

El rediseño evaluativo puede apoyarse en evidencias complementarias. Un producto elaborado con ayuda de IAG debería incluir una nota metodológica, fuentes verificadas, versiones del proceso y una defensa breve. En tareas profesionales, también conviene incorporar escenarios que cambien y exijan adaptar la respuesta. Estas estrategias no agregan trámites sin propósito. Su función es recuperar señales de comprensión que el producto final, por sí solo, ya no ofrece.

La supervisión humana debe permanecer en decisiones de alto impacto, como calificaciones finales, aprobación de prácticas o sanciones. La IAG puede sugerir patrones, resumir evidencias o comparar una respuesta con ciertos criterios. Aun así, el docente debe revisar el contexto y asumir la decisión. Esta responsabilidad requiere tiempo, formación y soporte institucional. Una supervisión solo nominal, sin condiciones de trabajo adecuadas, se convierte en una formalidad y no protege de manera efectiva a quienes participan en la evaluación.

Las políticas también requieren evaluación periódica. Conviene registrar incidentes, dificultades, percepciones, cambios en el rendimiento y efectos sobre la carga docente. Una reducción de tiempo no siempre representa una mejora, del mismo modo que toda resistencia no equivale a rechazo irracional. La evidencia institucional debe servir para ajustar reglas, retirar usos que producen daño y ampliar aquellos que muestran valor formativo. Sin este seguimiento, las políticas corren el riesgo de quedar desactualizadas o responder solo a impresiones.

3.7. Vacíos y desafíos de investigación

El balance de la evidencia muestra, además, áreas que aún no se han investigado. La revisión encontró una presencia alta de estudios exploratorios, encuestas de percepción y revisiones, frente a un número menor de investigaciones longitudinales. Todavía se sabe poco sobre la permanencia de los aprendizajes, la transferencia a la práctica profesional y los efectos de una exposición prolongada a sistemas generativos. También faltan comparaciones entre retroalimentación humana, automatizada y combinada, realizadas con criterios comunes de calidad. Este vacío limita la posibilidad de establecer qué modalidad funciona mejor y en qué condiciones.

Otro vacío se concentra en América Latina. Las experiencias disponibles muestran potencial, pero proceden de pocos países e instituciones. Se requieren estudios multicéntricos, participación de universidades rurales e interculturales y análisis de lenguas distintas al español dominante. La investigación también debe examinar si las herramientas reproducen desigualdades, cómo afectan la autonomía docente y qué costos institucionales acompañan su adopción. Sin esa evidencia, existe el riesgo de trasladar conclusiones obtenidas en contextos muy diferentes.

Aún faltan instrumentos validados para valorar alfabetización crítica, transparencia de uso y calidad del juicio humano en evaluaciones mediadas por IAG. Medir la frecuencia o la intención de uso no permite saber si la integración mejora el aprendizaje. Los próximos estudios deberían combinar indicadores de desempeño, análisis cualitativo del proceso y seguimiento de decisiones profesionales. Esta combinación ayudaría a comprender no solo qué resultados se obtienen, sino también cómo se producen y qué consecuencias tienen para estudiantes y docentes. Estos vacíos orientan las líneas de investigación del apartado 5.3.

4. Discusión

4.1. Convergencias y tensiones en la evidencia

La pregunta que orientó esta revisión permite reconocer un hallazgo central: la IAG hace visibles debilidades que ya existían en la evaluación universitaria. Antes de 2022, las tareas basadas en repetición, los criterios implícitos y la retroalimentación tardía ya generaban dificultades. La nueva tecnología las expone con mayor claridad, porque produce respuestas aceptables con poco esfuerzo. Por ello, la solución no consiste en recuperar el pasado mediante prohibiciones. El desafío real es construir evaluaciones en las que el pensamiento del estudiante resulte indispensable.

La defensa de la agencia humana constituye la principal coincidencia entre las fuentes. UNESCO (2023, 2024), la OECD (2026) y los estudios académicos sostienen que la tecnología debe ampliar capacidades sin sustituir la responsabilidad. En evaluación, un sistema puede colaborar con ejemplos, comentarios o preguntas. No debería, sin embargo, decidir de forma autónoma sobre logros, sanciones o aptitud profesional. Esas decisiones requieren contexto, criterio y una persona que pueda explicar y asumir sus consecuencias.

La transparencia representa una segunda coincidencia. Las políticas ambiguas generan inseguridad y respuestas defensivas, mientras una declaración clara convierte una práctica oculta en objeto de aprendizaje. Aun así, declarar el uso no es suficiente. El estudiante debe indicar qué herramienta empleó, con qué propósito, qué partes revisó y de qué manera verificó la información. La transparencia adquiere sentido cuando permite reconstruir el razonamiento y valorar la calidad de las decisiones, no cuando se limita a completar un formulario.

La evidencia sobre retroalimentación ofrece resultados concretos, en especial para grupos numerosos. Su valor depende, sin embargo, de la arquitectura pedagógica. Ossa y Willatt (2023) y Muñoz-Sepúlveda et al. (2026) sitúan la tecnología alrededor de casos, rúbricas, simulaciones y revisión profesional. No la presentan como sustituto del criterio docente. Esta diferencia separa una automatización orientada solo al ahorro de tiempo de una innovación que busca mejorar el aprendizaje y la reflexión sobre la práctica.

La relación entre eficiencia y profundidad mantiene una tensión difícil de resolver. Qian (2025) advierte que la IAG puede aumentar la productividad y, al mismo tiempo, trasladar fuera del estudiante habilidades que necesita desarrollar. La evaluación debe aprovechar la eficiencia para liberar tiempo destinado a interpretar, debatir y crear. Si la herramienta elimina precisamente la actividad que constituye el aprendizaje, la mejora operativa se transforma en una pérdida pedagógica.

En América Latina, la innovación requiere una perspectiva de justicia. Traducir marcos internacionales no basta. Las universidades deben considerar conectividad, costos, repertorios lingüísticos, diversidad cultural y autonomía institucional. La competencia docente también incluye saber cuándo no conviene utilizar IAG, cuándo es necesario ofrecer una alternativa y cómo reconocer conocimientos locales que el modelo puede omitir o representar de forma deficiente (Betti Galasso, 2026).

4.2. Equilibrio analítico: evidencia favorable y evidencia crítica

La evidencia reunida tiende a destacar los beneficios de la IAG, y esa inclinación merece discusión. Parte del corpus describe experiencias piloto con resultados alentadores, como las de Ossa y Willatt (2023), López-Vasco et al. (2025) y Méndez Méndez et al. (2025). Otra parte advierte sobre sesgos, opacidad, dependencia y límites de medición. Los estudios favorables suelen ser exploratorios y proceden de instituciones específicas, mientras que los críticos incluyen revisiones y análisis de políticas, pero también evidencia empírica, como la de Liang et al. (2023) sobre los detectores. Esta sección contrasta ambas lecturas y distingue lo que procede del corpus de lo que se incorpora como literatura complementaria.

4.2.1. Sesgos algorítmicos y transparencia de los contenidos generados

Los sesgos de los modelos no constituyen un riesgo hipotético. Liang et al. (2023) mostraron que los detectores de texto clasifican con mayor frecuencia como generados por máquina los textos de escritores no nativos de inglés, y UNESCO (2023) advierte que los modelos pueden reproducir estereotipos presentes en sus datos. Bender et al. (2021), como literatura complementaria, sostienen que los modelos entrenados con grandes volúmenes de texto pueden codificar visiones dominantes y producir lenguaje fluido sin comprensión, por lo que reclaman documentar con cuidado los datos de entrenamiento. En evaluación, esto implica que una retroalimentación bien redactada puede encubrir criterios de normalidad ajenos al contexto del estudiante.

La transparencia presenta una segunda dificultad. Quien recibe un texto, una rúbrica o un comentario generado automáticamente difícilmente puede reconstruir en qué fuentes, supuestos o criterios se apoyó el sistema, de modo que verificar su exactitud depende de la revisión humana, y las respuestas fluidas dificultan detectar errores, omisiones o simplificaciones.

Kasneci et al. (2023), como literatura complementaria, señalan entre los retos de estos modelos el posible sesgo de las respuestas y la necesidad de supervisión humana continua, así como de estrategias de verificación de datos y pensamiento crítico. Por ello, la declaración de uso (apartados 3.4 y 4.1) es una condición mínima de transparencia: permite saber qué parte del trabajo se apoyó en el sistema y cuál fue revisada por una persona.

4.2.2. Dependencia tecnológica

En el corpus, la preocupación por la dependencia aparece de forma indirecta. Weng et al. (2024) asocian el mayor valor educativo con usos que amplían la construcción de conocimiento, y Qian (2025) advierte sobre el traslado de habilidades fuera del estudiante. Bastani et al. (2025), como literatura complementaria, realizaron un experimento de campo con casi mil estudiantes de matemáticas de secundaria y hallaron que el acceso irrestricto a GPT-4 mejoró el desempeño durante la práctica, pero redujo las calificaciones en un 17 % cuando se retiró el acceso. Un tutor con salvaguardas pedagógicas mitigó ese efecto.

Los resultados de Bastani et al. corresponden a otro nivel educativo y no pueden trasladarse sin más a la educación superior, pero ilustran el riesgo de usar la herramienta como apoyo permanente.

Respecto del profesorado, la evidencia del corpus es limitada. Méndez Méndez et al. (2025) y López-Vasco et al. (2025) reportan mejoras en conocimiento, predisposición o dimensiones TPACK tras programas formativos, pero ninguno examina si esa confianza se traduce en autonomía profesional sostenida.

Más de la mitad de los estudiantes encuestados por el Digital Education Council (2024) cree que la dependencia excesiva de la inteligencia artificial perjudicará su rendimiento académico. La dependencia de servicios comerciales afecta, además, la continuidad y la autonomía institucional. Conviene, por tanto, comprobar en estudios futuros si los docentes conservan la capacidad de elaborar rúbricas, retroalimentar y decidir sin la mediación del sistema, y no solo si declaran mayor confianza para usarlo.

4.2.3. Competencias complejas y uso no supervisado

Los resultados generados por IAG constituyen insumos revisables y no mediciones suficientes del desempeño en competencias complejas, como el juicio pedagógico, la deliberación ética, la atención a la diversidad y la decisión frente a la incertidumbre. Muñoz-Sepúlveda et al. (2026) sitúan la tecnología como apoyo para ordenar comentarios y reconocer patrones, mientras la interpretación permanece en el diálogo y el criterio docente, y Ossa y Willatt (2023) mantuvieron supervisión humana sobre las respuestas generadas. En los estudios descritos en este capítulo, la tecnología asiste la valoración, pero no la sustituye.

El uso no supervisado tiene consecuencias éticas concretas. Una sanción fundada solo en la puntuación de un detector puede afectar de manera desigual a estudiantes cuya primera lengua no es el inglés (Liang et al., 2023), y una calificación automatizada puede desconocer la lengua, el contexto o las condiciones de producción del trabajo. Por ello, cualquier implementación debe prever revisión humana, transparencia, resguardo de información y alternativas no automatizadas.

El Reglamento (UE) 2024/1689 clasifica como sistemas de alto riesgo, con requisitos que incluyen la supervisión humana, los destinados a evaluar resultados de aprendizaje, entre otros usos educativos (Parlamento Europeo y Consejo de la Unión Europea, 2024). Esa clasificación ilustra cómo otras jurisdicciones tratan estas decisiones.

4.2.4. Desafíos regulatorios futuros

Los desafíos regulatorios son de tres tipos. El primero es la heterogeneidad normativa: mientras la Unión Europea adoptó un reglamento con niveles de riesgo, las iniciativas latinoamericanas son diversas y se encuentran en etapas distintas, como el proyecto chileno despachado en 2025 (Cámara de Diputadas y Diputados de Chile, 2025). El segundo es la brecha entre estrategia y capacidad, pues muchas estrategias nacionales carecen de financiamiento, mecanismos de implementación y evaluación de impacto (CEPAL, 2025). El tercero es la falta de especificidad sectorial: el comunicado chileno consultado no alude a la educación superior, mientras que el reglamento europeo sí incluye usos educativos.

Para las universidades, esto significa que la regulación futura probablemente exigirá demostrar supervisión humana, trazabilidad y protección de datos en las decisiones evaluativas.

Las instituciones que ya registren usos, declaraciones, incidentes y apelaciones estarán en mejor posición para adaptarse. Esta previsión es una inferencia a partir de los marcos revisados y no una predicción sobre normas aún no aprobadas.

4.3. El marco CUIDA como propuesta teórica integradora

A partir de estas coincidencias y tensiones, el capítulo propone el marco CUIDA. Se trata de una propuesta teórica derivada de la revisión integrativa: resulta de la comparación crítica de categorías documentales y no de una intervención, de un estudio empírico ni de un proceso de validación. Su nombre reúne cinco condiciones: criterio pedagógico, uso transparente, integridad académica, datos protegidos y agencia humana. Su propósito es orientar decisiones antes, durante y después de la evaluación, y no convertirse en una lista burocrática. Cada dimensión obliga a justificar la incorporación tecnológica desde el aprendizaje, la equidad y la responsabilidad profesional, por lo que las cinco deben revisarse de manera conjunta.

El criterio pedagógico relaciona el uso de IAG con los resultados de aprendizaje y la evidencia solicitada. El uso transparente establece permisos, límites y formas de declaración. La integridad académica vincula autoría, comprensión y responsabilidad. Los datos protegidos incluyen privacidad, seguridad, sesgo y acceso equitativo. La agencia humana, por último, reserva al profesorado y al estudiantado aquellas decisiones que exigen contexto, juicio profesional y responsabilidad moral. Cada dimensión protege un aspecto distinto, pero ninguna funciona de manera aislada.

El aporte del marco consiste en reunir elementos que a menudo se analizan por separado. Una tarea puede ser pedagógicamente interesante y, al mismo tiempo, inaceptable si expone datos sensibles. Una política puede defender la integridad y resultar injusta si depende de detectores poco confiables. CUIDA invita a revisar el conjunto y ayuda a reconocer dónde se concentra el riesgo antes de aplicar una evaluación. Esta mirada evita soluciones parciales que resuelven un problema mientras crean otro.

La propuesta mantiene una relación directa con los ejes transversales del libro. Examina una disrupción científica, analiza la transformación digital, plantea una innovación evaluativa y responde a retos globales de ética, inclusión y gobernanza. Su aporte específico conecta estos debates con la formación docente, un campo estratégico porque las decisiones actuales influirán en las prácticas educativas de los próximos años. La Figura 2 representa esta arquitectura. La Tabla 4, a su vez, convierte cada dimensión en preguntas de diseño, evidencias esperadas y señales de alerta para la aplicación institucional.

Figura 2

Marco CUIDA para una evaluación formativa mediada por inteligencia artificial generativa (IAG)



Nota. Las cinco dimensiones se revisan de manera articulada antes, durante y después de la evaluación. Elaboración propia.

Tabla 4

Matriz de decisiones del marco CUIDA

Dimensión	Pregunta de diseño	Evidencia esperada	Señal de alerta
Criterio pedagógico	¿La IAG aporta al aprendizaje que se pretende evaluar?	Alineación con resultados, tarea y criterios	Uso por novedad, presión o comodidad
Uso transparente	¿Están claros los permisos, límites y formas de declaración?	Consigna, registro de uso y nota metodológica	Reglas ambiguas, implícitas o cambiantes
Integridad académica	¿Puede verificarse comprensión y autoría responsable?	Borradores, fuentes, defensa y reflexión	Producto final sin huellas del proceso
Datos protegidos	¿Se preservan privacidad, seguridad, sesgo y equidad?	Anonimización, alternativas y acceso equivalente	Datos sensibles o herramientas obligatorias de pago
Agencia humana	¿Quién revisa, decide y responde por el resultado?	Supervisión, explicación y posibilidad de apelación	Calificación, sanción o exclusión automatizada

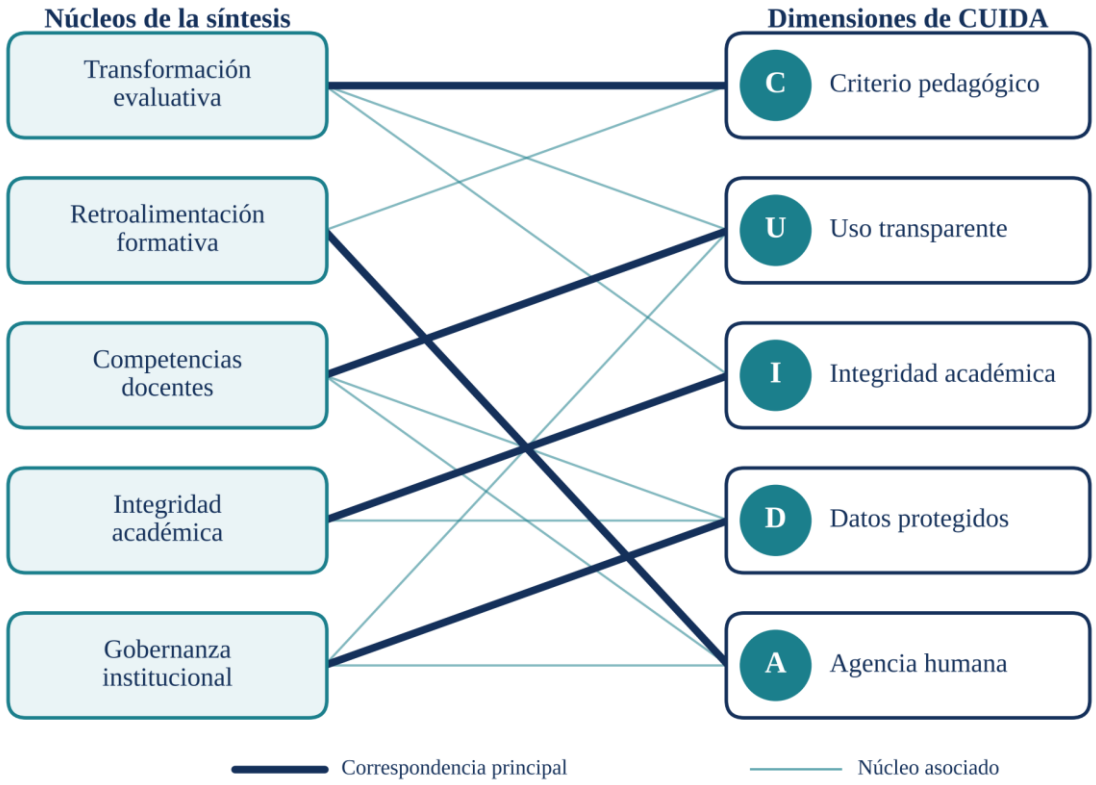
Nota. La matriz sirve para revisar consignas, rúbricas y políticas institucionales antes de su aplicación. IAG = inteligencia artificial generativa. Elaboración propia.

4.3.1. Derivación de CUIDA a partir de los hallazgos

La Figura 3 sintetiza cómo cada núcleo de la síntesis se vincula con las dimensiones de CUIDA. La línea gruesa indica la correspondencia principal, es decir, la dimensión que responde de forma más directa al riesgo o la necesidad dominante en el núcleo. Las líneas delgadas señalan dimensiones asociadas que el análisis de ese núcleo también moviliza. La relación no es biunívoca: un mismo núcleo plantea exigencias a varias dimensiones, y cada dimensión recibe aportes de más de un núcleo.

Figura 3

Correspondencia entre los núcleos de la síntesis y las dimensiones del marco CUIDA



Nota. La línea gruesa indica la correspondencia principal y la línea delgada, el núcleo asociado. Las correspondencias son interpretativas y no han sido validadas empíricamente. Elaboración propia.

La Tabla 5 explica el proceso de derivación y vincula cada núcleo con las fuentes del corpus que lo sustentan. Cada fila se lee de izquierda a derecha: del núcleo y sus fuentes a la dimensión que lo traduce en una exigencia de diseño. La asignación de documentos a núcleos se reconstruyó a partir de su contenido, porque no se conservaron los registros originales de codificación, y debe leerse como una propuesta interpretativa. Las relaciones no constituyen evidencia de eficacia del marco; representan criterios argumentados que deberán contrastarse mediante estudios futuros.

Tabla 5

Correspondencia entre núcleos de hallazgos, literatura del corpus y dimensiones de CUIDA

Núcleo de hallazgos	Fuentes del corpus	Dimensión principal de CUIDA	Dimensiones asociadas	Justificación conceptual
Transformación evaluativa	Furze et al. (2024); Luo (2024); Moorhouse et al. (2023); Perkins et al. (2024); Weng et al. (2024); Xia et al. (2024)	Criterio pedagógico	Uso transparente; integridad académica	Alinear el apoyo automatizado con evidencias observables de aprendizaje.
Retroalimentación formativa	Muñoz-Sepúlveda et al. (2026); Ossa y Willatt (2023)	Agencia humana	Criterio pedagógico	Conservar la deliberación y la revisión del estudiante y del docente.
Competencias docentes	Betti Galasso (2026); Celik (2023); López-Vasco et al. (2025); Méndez Méndez et al. (2025); UNESCO (2024)	Uso transparente	Datos protegidos; agencia humana	Definir permisos, límites, explicación y registro del apoyo de IA.
Integridad académica	Gallent-Torres et al. (2023); Liang et al. (2023); UNESCO (2023)	Integridad académica	Datos protegidos	Documentar autoría, responsabilidad y verificabilidad del proceso.
Gobernanza institucional	Chan (2023); OECD (2023)	Datos protegidos	Uso transparente; agencia humana	Prevenir riesgos de privacidad, sesgos y exclusión por acceso desigual.
Transversal	Bond et al. (2024); OECD (2026); Qian (2025)	Las cinco dimensiones	No aplica	Aportan la exigencia ética, la intención pedagógica y la tensión entre eficiencia y aprendizaje que atraviesan todo el marco.

Nota. Las correspondencias son interpretativas y no han sido validadas empíricamente. IA = inteligencia artificial. Elaboración propia a partir de la síntesis de los 21 documentos.

4.3.2. Alcance científico y líneas de validación

Conviene precisar el alcance de CUIDA. Es una construcción conceptual: sus dimensiones, preguntas y señales de alerta se infirieron de convergencias entre fuentes, y no de datos obtenidos con estudiantes o docentes. Su validez de contenido, su factibilidad, su aceptación institucional y su efecto sobre el aprendizaje, la equidad y las decisiones docentes deberán comprobarse mediante investigaciones futuras. Hasta entonces, el marco debe usarse como guía de revisión y no como instrumento de certificación.

Se propone una validación gradual en cuatro líneas. La primera es la revisión de contenido por especialistas en evaluación, formación docente y ética digital, que puede cuantificarse con la razón

de validez de contenido de Lawshe (1975). La segunda es el pilotaje de consignas y rúbricas elaboradas con el marco en cursos de formación inicial docente de distintas regiones, con registro del uso real.

La tercera línea es el análisis de consistencia de los instrumentos derivados, por ejemplo, mediante el acuerdo entre evaluadores. La cuarta es la comparación de las decisiones pedagógicas antes y después de aplicar CUIDA, en estudios con seguimiento y, cuando sea viable, con grupos de comparación en contextos educativos reales, incluidos aquellos con acceso tecnológico limitado.

4.4. Implicaciones institucionales

Las universidades pueden traducir la síntesis en decisiones concretas. La Tabla 6 organiza cinco ámbitos de acción con las recomendaciones que se derivan de los núcleos analizados, las fuentes que las respaldan y la dimensión de CUIDA con la que se relacionan. Las recomendaciones son propuestas formuladas a partir de la síntesis y de la literatura complementaria, y no han sido evaluadas en su aplicación.

En América Latina estas medidas requieren prever conectividad desigual, costos de licencias, variantes lingüísticas y diversidad cultural, y evitar que una herramienta de pago se convierta en requisito tácito. Los programas de desarrollo profesional, además, deben partir de la infraestructura disponible, pues la CEPAL (2025) describe una formación avanzada insuficiente y concentrada en pocos países.

Tabla 6

Implicaciones y recomendaciones institucionales derivadas de la síntesis

Ámbito	Recomendaciones para la institución	Fundamento	Dimensión de CUIDA
Políticas universitarias de uso responsable	Definir por tarea y resultado de aprendizaje si el uso de la IAG está prohibido, es opcional, requerido o sujeto a declaración; incluir la regla en el sílabo y en cada consigna; precisar responsabilidades ante errores, propiedad intelectual y mecanismos de apelación.	Furze et al. (2024); Luo (2024); Moorhouse et al. (2023); Perkins et al. (2024)	Uso transparente
Protocolos de integridad académica	No fundar una sanción únicamente en la puntuación de un detector; exigir declaración de uso, fuentes verificadas, versiones del proceso y una defensa breve; garantizar el debido proceso y la revisión contextual.	Gallent-Torres et al. (2023); Liang et al. (2023); Navarrete-Cazales et al. (2026), literatura complementaria	Integridad académica
Lineamientos para la evaluación asistida por IA	Publicar criterios y rúbricas visibles; usar la IAG para borradores de retroalimentación que el docente revisa; mantener la decisión humana en calificaciones finales, aprobación de prácticas y sanciones; ofrecer alternativas a quien no use la herramienta.	Muñoz-Sepúlveda et al. (2026); Ossa y Willatt (2023); UNESCO (2024); Parlamento Europeo y Consejo de la Unión Europea (2024), literatura complementaria	Criterio pedagógico; agencia humana

Ámbito	Recomendaciones para la institución	Fundamento	Dimensión de CUIDA
Capacitación docente en competencias digitales avanzadas	Ofrecer programas que combinen conocimiento técnico, práctica guiada y discusión ética, con talleres de rediseño evaluativo, análisis de respuestas erróneas, construcción de rúbricas, revisión de sesgos, protección de datos y declaraciones de uso; dar seguimiento a los cambios en los instrumentos.	Celik (2023); López-Vasco et al. (2025); Méndez Méndez et al. (2025); UNESCO (2024)	Uso transparente; datos protegidos
Gobernanza institucional	Crear una instancia que articule política, currículo, infraestructura y desarrollo profesional; registrar incidentes y percepciones; revisar periódicamente las reglas; anonimizar los datos antes de usar plataformas externas; prever licencias institucionales o alternativas equivalentes.	Chan (2023); OECD (2023); UNESCO (2023); CEPAL (2025), literatura complementaria	Datos protegidos

Nota. Las recomendaciones son propuestas derivadas de la síntesis y no han sido evaluadas en su aplicación. IAG = inteligencia artificial generativa; IA = inteligencia artificial. Elaboración propia.

4.5. Limitaciones de la revisión

Las conclusiones de este capítulo deben leerse con las siguientes limitaciones. La primera deriva del carácter emergente y acelerado del campo analizado. El periodo 2023–2026 se justificó por la difusión pública de ChatGPT, pero también excluye estudios previos sobre inteligencia artificial educativa y concentra una producción reciente en la que predominan revisiones, estudios exploratorios y trabajos basados en percepciones, mientras siguen siendo escasas las investigaciones longitudinales y multicéntricas. Esta composición restringe la posibilidad de establecer efectos sostenidos, relaciones causales o comparaciones estables entre modalidades de evaluación mediadas por IAG.

Una segunda limitación corresponde a la representación regional. La evidencia latinoamericana incluida procede de un número reducido de países e instituciones, con escasa presencia de universidades rurales, interculturales y contextos lingüísticos diversos.

Una tercera limitación se relaciona con las restricciones idiomáticas y con la literatura no indexada. La búsqueda se restringió a documentos en español e inglés, lo que pudo dejar fuera trabajos en portugués u otras lenguas pertinentes para la región. Además, se consultaron cuatro bases y portales editoriales, por lo que pudo omitirse literatura relevante no indexada en ellos, como informes institucionales, tesis o literatura gris local, que no se incluyó de manera sistemática.

Pueden existir, asimismo, sesgos de publicación. Los estudios con resultados favorables suelen publicarse con mayor frecuencia que los que no encuentran efectos, y la exigencia de revisión por pares para los artículos pudo reforzar esa tendencia. Los cambios rápidos de las herramientas durante el periodo analizado reducen, además, la vigencia de algunos hallazgos.

Otras dos limitaciones afectan la reproducibilidad. La primera es la ausencia de un registro auditable de resultados brutos por base y etapa de exclusión (apartado 2.6). La segunda es que la asignación de documentos a núcleos se reconstruyó a partir de su contenido y no de los registros de

codificaciones originales, lo que impide afirmar saturación formal (apartado 2.7). Estas debilidades no invalidan la síntesis conceptual, pero exigen prudencia al atribuir exhaustividad y justifican una actualización futura del protocolo de búsqueda.

Finalmente, la heterogeneidad de poblaciones, herramientas, diseños y resultados impide estimar un efecto agregado. Esta revisión integrativa prioriza la síntesis crítica y la identificación de convergencias, riesgos y vacíos; por ello, sus conclusiones deben interpretarse como orientaciones fundamentadas en el corpus analizado y no como estimaciones generalizables de eficacia para todos los contextos universitarios. La literatura complementaria, por su parte, se seleccionó de manera dirigida y no pasó por la matriz de valoración, de modo que sirve para contrastar los hallazgos y no para ampliar la síntesis.

5. Conclusiones

5.1. Hallazgos principales

1. La IAG modifica la evaluación porque reduce la capacidad de los productos tradicionales para demostrar aprendizaje individual. La respuesta más sólida consiste en diversificar evidencias, valorar el proceso y solicitar explicaciones que permitan reconocer comprensión, juicio y responsabilidad. Este cambio no elimina la escritura ni los productos finales. Los integra con borradores, decisiones registradas y otras huellas que muestran cómo se elaboró el trabajo.
2. La retroalimentación automatizada puede ampliar la oportunidad y la cobertura del acompañamiento, siempre que cuente con criterios explícitos, supervisión humana y posibilidades de diálogo. Su calidad no depende de la extensión del comentario, sino de su relación con evidencias concretas y acciones posibles de mejora. El valor formativo aparece cuando el estudiante interpreta, cuestiona y utiliza la orientación recibida.
3. La formación docente debe integrar competencias técnicas, pedagógicas, éticas y evaluativas. Los programas centrados únicamente en el manejo de herramientas pueden aumentar la confianza, pero no resuelven dilemas de autoría, privacidad, sesgo y originalidad. La preparación profesional requiere práctica situada, análisis de casos, revisión de respuestas incorrectas y rediseño de tareas relacionadas con contextos reales.
4. En América Latina, la incorporación de IAG debe considerar las desigualdades de acceso, la diversidad cultural, la pertinencia lingüística y las capacidades institucionales. Una política uniforme puede ampliar brechas cuando supone licencias, conectividad o conocimientos que no están disponibles para toda la comunidad. La innovación responsable exige alternativas equivalentes, soporte técnico y participación de los actores educativos.
5. La evidencia favorable es, en su mayoría, exploratoria y procede de contextos específicos, mientras que la evidencia crítica advierte sesgos, opacidad, dependencia y límites para

valorar competencias complejas. Ambas lecturas son compatibles: la IAG aporta cuando la tarea, el criterio y la supervisión humana están definidos, y debilita el aprendizaje cuando el sistema sustituye el razonamiento o decide sin control profesional.

5.2. Implicaciones prácticas para las instituciones de educación superior

El marco CUIDA constituye una propuesta teórica surgida de la revisión integrativa. Sus cinco dimensiones (criterio pedagógico, uso transparente, integridad académica, datos protegidos y agencia humana) ofrecen preguntas orientadoras para revisar prácticas evaluativas, pero su eficacia, validez de contenido, factibilidad y aceptación institucional todavía requieren comprobación empírica. Con esa salvedad, y a partir de la síntesis, se recomienda a las instituciones de educación superior lo siguiente.

1. Establecer políticas diferenciadas por tarea y resultado de aprendizaje, que indiquen si el uso de la IAG está prohibido, es opcional, requerido o sujeto a declaración, y que figuren en el sílabo y en cada consigna.
2. Adoptar protocolos de integridad académica que no basen sanciones en la puntuación de un detector, exijan declaración de uso, fuentes verificadas y evidencias del proceso, y garanticen el debido proceso y la posibilidad de apelación.
3. Definir lineamientos de evaluación asistida por IA con criterios y rúbricas visibles, retroalimentación revisada por el docente y decisión humana en calificaciones finales, aprobación de prácticas y sanciones, junto con alternativas para quien no use la herramienta.
4. Capacitar al profesorado mediante programas que combinen conocimiento técnico, práctica guiada y discusión ética, con talleres de rediseño evaluativo, análisis de respuestas erróneas, revisión de sesgos y protección de datos, y con seguimiento de los cambios en los instrumentos.
5. Constituir instancias de gobernanza que revisen periódicamente las reglas, registren incidentes, protejan los datos del estudiantado y ofrezcan alternativas equivalentes de acceso, con atención a la conectividad, las lenguas y la diversidad cultural de cada contexto latinoamericano.

5.3. Líneas futuras de investigación

1. Desarrollar estudios longitudinales y multicéntricos que comparen formas de retroalimentación, examinen la transferencia a la práctica y valoren efectos sobre autonomía, equidad y carga docente, con participación de universidades rurales e interculturales y de lenguas distintas al español dominante.
2. Validar el marco CUIDA mediante revisión de contenido por especialistas, pilotaje de consignas y rúbricas, análisis de consistencia de los instrumentos y comparación de decisiones pedagógicas antes y después de su uso en contextos educativos reales.

3. Construir instrumentos validados para medir aprendizaje verificable, alfabetización crítica y calidad del juicio humano en evaluaciones mediadas por IAG, y examinar la dependencia tecnológica del profesorado y el efecto de las regulaciones emergentes sobre la evaluación.

Referencias bibliográficas

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences, 122*(26), Article e2422633122. https://doi.org/10.1073/pnas.2422633122

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Betti Galasso, B. J. (2026). Inteligencia artificial generativa y formación docente: Competencias emergentes en América Latina – una revisión de alcance. *Actualidades Investigativas en Educación, 26*(1), 1–43. https://doi.org/10.15517/z4mcs052

Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, Article 4. https://doi.org/10.1186/s41239-023-00436-z

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology, 18*(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Cámara de Diputadas y Diputados de Chile. (2025, 13 de octubre). *Cámara despachó marco regulatorio para uso de la inteligencia artificial (IA)* [Comunicado de prensa]. https://www.camara.cl/cms/2025/10/13/camara-despacho-marco-regulatorio-para-uso-de-la-inteligencia-artificial-ia/

Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior, 138*, Article 107468. https://doi.org/10.1016/j.chb.2022.107468

CEPAL. (2025, 3 de octubre). *América Latina y el Caribe acelera la adopción de la inteligencia artificial, aunque con desafíos en inversión, talento y gobernanza* [Comunicado de prensa]. https://www.cepal.org/es/comunicados/america-latina-caribe-acelera-la-adopcion-la-inteligencia-artificial-aunque-desafios

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*, Article 38. https://doi.org/10.1186/s41239-023-00408-3

Digital Education Council. (2024, 7 de agosto). *What students want: Key results from DEC Global AI Student Survey 2024*. https://www.digitaleducationcouncil.com/post/what-students-want-key-results-from-dec-global-ai-student-survey-2024

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology, 40*(4), 38–55. https://doi.org/10.14742/ajet.9434

Gallent-Torres, C., Zapata-González, A., & Ortego-Hernando, J. L. (2023). El impacto de la inteligencia artificial generativa en educación superior: Una mirada desde la ética y la integridad académica. *RELIEVE*, *29*(2), 50–65. https://doi.org/10.30827/relieve.v29i2.29134

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, *4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

López-Vasco, F. E., Angulo-Álvarez, M. R., & Sosa-Zúñiga, D. I. (2025). Formación docente en IA generativa: Impacto ético y retos en educación superior. *Alteridad*, *20*(2), 166–177. https://doi.org/10.17163/alt.v20n2.2025.01

Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the originality of students' work. *Assessment & Evaluation in Higher Education*, *49*(5), 651–664. https://doi.org/10.1080/02602938.2024.2309963

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, *26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Méndez Méndez, K. F., Delgado Calle, B. B., & Sequera Morales, A. G. (2025). Efectos de la aplicación de un programa de formación docente sobre el nivel de uso de programas de IAG en una institución educativa de Cuenca, Ecuador. *Espacios*, *46*(3), 361–370. https://doi.org/10.48082/espacios-a25v46n03p28

Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, *5*, Article 100151. https://doi.org/10.1016/j.caeo.2023.100151

Muñoz-Sepúlveda, F., Willatt, C., & Bawarshi-Abrigo, Y. (2026). Evaluación formativa en la formación inicial docente: Un modelo de retroalimentación basado en simulación, videoanálisis e inteligencia artificial generativa. *Revista Realidad Educativa*, *6*(2), 133–160. https://doi.org/10.38123/rre.v6i2.969

Navarrete-Cazales, Z., Cruz-Victoria, J. C., & Manzanilla-Granados, H. M. (2026). Autoría e integridad académica ante la inteligencia artificial generativa en educación superior. *Revista Educación Superior y Sociedad (ESS)*, *37*(2), 205–222. https://doi.org/10.54674/ess.v37i2.1174

OECD. (2023). *OECD digital education outlook 2023: Towards an effective digital education ecosystem*. https://doi.org/10.1787/c74f03de-en

OECD. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. https://doi.org/10.1787/062a7394-en

Ossa, C., & Willatt, C. (2023). Uso de inteligencia artificial generativa para retroalimentar escritura académica en procesos de formación inicial docente. *European Journal of Education and Psychology*, *16*(2), 1–16. https://doi.org/10.32457/ejep.v16i2.2412

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Parlamento Europeo y Consejo de la Unión Europea. (2024, 13 de junio). *Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial)*. Diario Oficial de la Unión Europea, L 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, *21*(6), 49–66. https://doi.org/10.53761/q3azde36

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, *69*(5), 1105–1120. https://doi.org/10.1007/s11528-025-01100-1

Roblez Torres, G. V., Andrade Apolo, A. M., & Gonzaga Aguilar, K. E. (2026). Alfabetización en inteligencia artificial generativa y evaluación auténtica en educación superior: Integridad académica. *Educação em Revista*, *42*, Article e67812. https://doi.org/10.1590/0102-469867812

Stephenson, R., & Armstrong, C. (2026). *Student generative artificial intelligence survey 2026* (HEPI Report 199). Higher Education Policy Institute. https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/

UNESCO. (2023). *Guidance for generative AI in education and research*. https://unesdoc.unesco.org/ark:/48223/pf0000386693

UNESCO. (2024). *AI competency framework for teachers*. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Vallejos-Parás, C., & Galindo-Cuesta, J. A. (2025). Ética e inteligencia artificial generativa en la educación: Perspectivas docentes de Brasil, Colombia y México. *Actualidades Pedagógicas*, (85), Article e5234. https://doi.org/10.19052/ap.vol1.iss85.5234

Weng, X., Xia, Q., Gu, M., Rajaram, K., & Chiu, T. K. F. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*, *40*(6), 37–65. https://doi.org/10.14742/ajet.9540

Whittemore, R., & Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, *52*(5), 546–553. https://doi.org/10.1111/j.1365-2648.2005.03621.x

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education, 21*, Article 40. https://doi.org/10.1186/s41239-024-00468-z