REVISTA ECUATORIANA DE DESARROLLO SOCIAL Y AMBIENTAL
Julio - Septiembre, 2026 | Vol. 1 | Núm. 3
ISSN: 3151-8303

La evaluación educativa ecuatoriana ante la inteligencia artificial generativa: Desafíos para la autenticidad del aprendizaje

Ecuadorian Educational Assessment in the Face of Generative Artificial Intelligence: Challenges for Learning Authenticity

1 **Henry Fabricio Carrillo Chacón**
Investigador independiente; Salcedo - Ecuador;
henry_carrillo@outlook.es
https://orcid.org/0009-0008-9897-8168

2 **Fátima Lourdes Vega Barberán**
Investigadora independiente; Santo Domingo - Ecuador;
fatilour1695@gmail.com
https://orcid.org/0009-0007-1800-3436

3 **Diego Alexander Cevallos Torres**
Investigador independiente; Latacunga - Ecuador;
diegocevallos2026@gmail.com
https://orcid.org/ 0009-0008-3961-6984

4 **María Lorena Torres Meza**
Investigador independiente; Latacunga - Ecuador;
lorenatorres2026t@gmail.com
https://orcid.org/0009-0000-5795-3433

Enviado: 21/07/2026 **Aceptado:** 29/07/2026 **Publicado:** 31/07/2026

Tipo de Investigación: Artículo de revisión **Páginas:** 125 - 142

DOI: https://doi.org/10.66646/redsa-2026/xx5c4v36

Conflicto de Intereses Ninguno que declarar.

Correspondencia: **Henry Fabricio Carrillo Chacón**
Investigador independiente; Salcedo - Ecuador;
henry_carrillo@outlook.es

Esta obra se distribuye bajo la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0).
https://creativecommons.org/licenses/by-nc-sa/4.0/

RESUMEN

La expansión de la inteligencia artificial generativa desafía la capacidad de las evaluaciones tradicionales para acreditar la autoría, la comprensión y el razonamiento del estudiantado. El objetivo fue analizar los desafíos que estas tecnologías plantean a la autenticidad del aprendizaje en la evaluación educativa ecuatoriana y derivar orientaciones pedagógicas compatibles con la normativa nacional. Se realizó una revisión integrativa con búsqueda sistematizada y análisis documental normativo. El corpus conservado por el equipo autoral reunió 30 documentos académicos, institucionales y normativos, seleccionados de 130 registros mediante criterios explícitos de pertinencia, actualidad y acceso al texto completo. La información se codificó temáticamente en seis categorías. Los resultados muestran que las tareas centradas exclusivamente en productos finales son especialmente vulnerables a la delegación cognitiva; que los detectores automáticos no ofrecen evidencia suficiente para atribuir autoría; y que la respuesta pedagógica más consistente combina trazabilidad del proceso, desempeño situado, diálogo oral, retroalimentación y declaración transparente del uso de inteligencia artificial. La normativa ecuatoriana proporciona bases formativas y contextualizadas para este rediseño, aunque la literatura disponible no permite describir la frecuencia real de tales prácticas en las instituciones del país. Se concluye que la autenticidad no depende de prohibir toda asistencia tecnológica, sino de alinear el nivel de ayuda permitido con los resultados de aprendizaje y de recoger evidencias complementarias del proceso y del desempeño.

Palabras clave:

inteligencia artificial generativa; evaluación educativa; evaluación auténtica; integridad académica; evaluación formativa; alfabetización en inteligencia artificial.

ABSTRACT

The expansion of generative artificial intelligence challenges the capacity of traditional assessment to establish student authorship, understanding, and reasoning. This study analyzed the challenges these technologies pose to learning authenticity in Ecuadorian educational assessment and derived pedagogical guidelines consistent with national regulations. An integrative review with a systematized search and normative document analysis was conducted. The corpus preserved by the authors comprised 30 academic, institutional, and regulatory documents selected from 130 records using explicit criteria of relevance, recency, and full-text availability. The evidence was coded into six thematic categories. Results show that tasks focused exclusively on final products are particularly vulnerable to cognitive delegation; automated detectors do not provide sufficient evidence to attribute authorship; and the most consistent pedagogical response combines process traceability, situated performance, oral dialogue, feedback, and transparent disclosure of artificial-intelligence use. Ecuadorian regulations provide formative and contextual foundations for this redesign, although available literature does not establish how frequently these practices occur in the country's institutions. Learning authenticity therefore depends less on prohibiting all technological assistance than on aligning permitted assistance with learning outcomes and collecting complementary evidence of process and performance.

Keywords:

generative artificial intelligence; educational assessment; authentic assessment; academic integrity; formative assessment; artificial intelligence literacy.

INTRODUCCIÓN

La inteligencia artificial generativa (IAG) modificó las condiciones bajo las cuales se producen textos, imágenes, códigos, explicaciones y soluciones académicas. A diferencia de herramientas digitales orientadas principalmente a recuperar o presentar información, los grandes modelos de lenguaje pueden generar productos completos a partir de instrucciones breves. Esta capacidad amplía las oportunidades de apoyo, personalización y retroalimentación, pero también debilita la inferencia tradicional según la cual un producto bien elaborado demuestra, por sí mismo, que quien lo entrega domina el conocimiento evaluado (Kasneci et al., 2023; Lo, 2023; Tlili et al., 2023).

El problema central no es la mera presencia de una tecnología, sino la validez de la inferencia evaluativa. Una calificación es defendible cuando la evidencia recogida representa el aprendizaje que se pretende valorar. Si una herramienta puede redactar un ensayo, resolver un cuestionario o elaborar una síntesis sin que el estudiante explique las decisiones tomadas, el producto final deja de ser una evidencia suficiente de comprensión. La cuestión relevante pasa de preguntar si hubo uso de IA a determinar qué parte del desempeño corresponde al estudiante, qué asistencia estaba autorizada y qué evidencias adicionales permiten sostener el juicio docente.

La literatura reciente identifica una tensión doble. Por una parte, la IAG puede ofrecer explicaciones adaptadas, ejemplos, retroalimentación preliminar y apoyo a la ideación. Por otra, puede producir información falsa, reproducir sesgos, facilitar la delegación del esfuerzo intelectual y generar incertidumbre sobre la autoría (Chan & Hu, 2023; Crompton & Burke, 2023; UNESCO, 2023). Por ello, los enfoques centrados únicamente en la prohibición resultan insuficientes: no distinguen entre apoyo legítimo y sustitución cognitiva, son difíciles de aplicar fuera del aula y pueden desplazar el debate pedagógico hacia una vigilancia técnicamente frágil.

La detección automática tampoco resuelve el problema. Los clasificadores de texto generado por IA han mostrado variaciones de desempeño, falsos positivos y sesgos contra escritores no nativos de inglés (Liang et al., 2023; Weber-Wulff et al., 2023). En consecuencia, una puntuación emitida por un detector no debería tratarse como prueba autónoma de falta académica. Una respuesta institucional proporcional requiere evidencias múltiples, debido proceso, diálogo con el estudiante y tareas cuyo diseño haga visible el aprendizaje.

Estas transformaciones reactivan debates previos sobre evaluación formativa y auténtica. La evaluación formativa utiliza la evidencia para ajustar la enseñanza y ayudar al estudiante a cerrar la distancia entre su desempeño actual y el esperado (Black & Wiliam, 1998; Sadler, 1989). La evaluación auténtica, por su parte, demanda desempeños intelectualmente significativos, contextualizados y semejantes a los usos reales del conocimiento (Gulikers et al., 2004; Wiggins, 1990). Ambas perspectivas desplazan la atención desde el producto aislado hacia la calidad del proceso, el juicio, la aplicación y la capacidad de explicar.

La validez constituye un criterio decisivo para comprender el problema. Evaluar implica interpretar una muestra de actuación y utilizarla para formular una conclusión sobre conocimientos

o capacidades que no pueden observarse de manera total. Esa conclusión pierde solidez cuando factores ajenos al aprendizaje objetivo explican una parte sustancial del resultado. En una tarea de escritura autónoma, por ejemplo, la generación automática del texto introduce una fuente de varianza que impide saber si la organización, el vocabulario y la argumentación proceden del estudiante. En una actividad destinada a evaluar el análisis crítico de respuestas de IA, en cambio, emplear la herramienta forma parte legítima de la competencia. La misma tecnología puede, por tanto, invalidar una tarea o constituir su objeto de evaluación.

La autenticidad tampoco debe confundirse con la ausencia de tecnología. En las prácticas profesionales contemporáneas se utilizan buscadores, correctores, traductores, sistemas de análisis y asistentes generativos. Una evaluación auténtica puede incorporar estas herramientas cuando el propósito es valorar la capacidad de seleccionar, verificar, adaptar y justificar sus resultados. Lo indispensable es que la ayuda disponible no oculte la competencia evaluada. De ahí la necesidad de describir el nivel de asistencia autorizado y de establecer qué decisiones deben permanecer bajo control del estudiante.

En este artículo se utiliza el concepto de delegación cognitiva para describir la transferencia a un sistema automatizado de operaciones intelectuales que la actividad pretendía desarrollar o comprobar. La delegación no es necesariamente negativa: liberar tareas rutinarias puede permitir mayor atención a problemas complejos. Se vuelve problemática cuando reemplaza la práctica deliberada necesaria para aprender, cuando impide atribuir el desempeño o cuando el estudiante acepta resultados sin verificarlos. La diferencia entre apoyo y sustitución depende del objetivo, del momento formativo, del grado de control y de la capacidad para explicar y corregir el producto.

La trazabilidad se entiende como la posibilidad de reconstruir razonablemente cómo se elaboró una respuesta. No exige vigilancia permanente ni almacenamiento invasivo de cada interacción. Puede lograrse mediante propuestas iniciales, borradores, decisiones comentadas, fuentes consultadas, versiones sucesivas y una declaración breve de herramientas empleadas. Su finalidad es pedagógica: ofrecer al estudiante oportunidades para reflexionar sobre su proceso y al docente evidencia para retroalimentarlo. Cuando se convierte en vigilancia indiscriminada, la trazabilidad puede afectar la privacidad y deteriorar la confianza educativa.

En Ecuador, el marco educativo concibe la evaluación como un proceso continuo orientado a observar, valorar, registrar y retroalimentar el aprendizaje. El Instructivo de evaluación estudiantil plantea actividades contextualizadas, pertinentes y flexibles, mientras que la Ley Orgánica de Educación Intercultural y su Reglamento sostienen principios de calidad, inclusión y atención a las necesidades del estudiantado (Asamblea Nacional del Ecuador, 2021; Ministerio de Educación del Ecuador, 2022; Presidencia de la República del Ecuador, 2023). Estas disposiciones ofrecen un punto de apoyo para rediseñar la evaluación ante la IAG, aunque por sí solas no demuestran cómo evalúan efectivamente todas las instituciones del país.

La brecha entre orientación normativa y práctica constituye, por tanto, una pregunta empírica aún abierta. Gran parte de la evidencia sobre IAG y evaluación proviene de educación superior

y de contextos internacionales (Bittle & El-Gayar, 2025; Xia et al., 2024). Sus hallazgos son conceptualmente útiles para educación básica y bachillerato, pero su transferencia debe realizarse con cautela debido a diferencias de edad, autonomía, acceso tecnológico, formación docente y regulación institucional.

A partir de este problema, el objetivo general fue analizar los desafíos que plantea la inteligencia artificial generativa a la autenticidad del aprendizaje en la evaluación educativa ecuatoriana, mediante una revisión integrativa de literatura y normativa.

Los objetivos específicos fueron:

- **a.** identificar los aportes teóricos y empíricos sobre IAG y evaluación;
- **b.** examinar los riesgos para la autoría, la integridad y la validez de las evidencias;
- **c.** relacionar la evaluación formativa y auténtica con la normativa ecuatoriana; y
- **d.** proponer orientaciones pedagógicas aplicables y verificables. La pregunta orientadora fue: ¿qué desafíos plantea la IAG a la autenticidad del aprendizaje y qué criterios pueden guiar el rediseño de la evaluación educativa ecuatoriana?

METODOLOGÍA

Se desarrolló una revisión integrativa con búsqueda sistematizada y análisis documental normativo. Esta denominación reconoce que el corpus combinó estudios empíricos, revisiones, propuestas conceptuales, documentos institucionales y normativa, por lo que no correspondía aplicar un metaanálisis ni atribuir homogeneidad metodológica a las fuentes. La revisión integrativa permite reunir evidencia heterogénea para construir explicaciones y orientaciones, siempre que se expliciten la selección, extracción y síntesis (Snyder, 2019). El flujo de identificación y selección se informó mediante una adaptación de PRISMA 2020; no se presenta el estudio como una revisión sistemática de efectos ni como un protocolo registrado (Page et al., 2021).

El corpus documental conservado por el equipo autoral provino de búsquedas en Scopus, Web of Science, ERIC, SciELO y Google Scholar, complementadas con consulta dirigida en portales editoriales y repositorios institucionales. Se emplearon combinaciones en español e inglés de los términos “inteligencia artificial generativa”, “evaluación educativa”, “evaluación auténtica”, “evaluación formativa”, “integridad académica”, “ChatGPT”, “generative artificial intelligence”, “educational assessment”, “authentic assessment” y “academic integrity”. Las combinaciones centrales fueron: (“generative artificial intelligence” OR ChatGPT) AND (“educational assessment” OR “authentic assessment” OR “academic integrity”), y sus equivalentes en español. La documentación original no conservó el rendimiento desagregado de cada base ni la fecha exacta de cada consulta; esta carencia se reconoce como limitación y es una razón adicional para definir el estudio como revisión integrativa con búsqueda sistematizada.

Se incluyeron publicaciones en español o inglés vinculadas directamente con IAG, evaluación, autenticidad o integridad académica; estudios y revisiones con texto completo disponible; documentos institucionales de organismos reconocidos; y normativa ecuatoriana vigente relacionada con evaluación. El periodo prioritario

fue 2022-2026, coincidente con la expansión pública de la IAG. Se conservaron trabajos anteriores cuando eran indispensables para el marco metodológico o para los conceptos de evaluación formativa y auténtica. Se excluyeron duplicados, resúmenes sin texto completo, artículos sobre IA no generativa sin relación evaluativa, textos de opinión sin desarrollo argumental o metodológico y fuentes que no respondían a la pregunta de revisión.

El registro documental informó 130 resultados iniciales: 116 académicos y 14 institucionales o normativos. Se eliminaron 40 registros antes del cribado —22 duplicados y 18 fuera del periodo o del alcance—. Se revisaron 90 títulos y resúmenes, de los cuales 48 fueron excluidos por baja pertinencia. Se buscaron 42 textos completos; cuatro no fueron recuperados y 38 se evaluaron en elegibilidad. Ocho se excluyeron: tres no abordaban evaluación, dos se centraban en IA no generativa, dos eran textos de opinión sin desarrollo suficiente y uno no aportaba resultados pertinentes. El corpus final quedó integrado por 30 documentos.

Extracción y análisis

Se utilizó una matriz con autoría, año, tipo de fuente, nivel educativo, objeto, metodología, hallazgos, limitaciones y utilidad para la pregunta. La síntesis siguió un análisis temático de seis momentos: familiarización, codificación inicial, agrupación, revisión, definición y redacción interpretativa (Braun & Clarke, 2006). Los códigos se organizaron en seis categorías: fundamentos metodológicos; posibilidades y riesgos de la IAG; integridad y detección; transformación de la evaluación; evaluación formativa y auténtica; y normativa ecuatoriana. La valoración fue de pertinencia y suficiencia informativa, no una puntuación uniforme de riesgo de sesgo, porque el corpus combinó diseños no comparables.

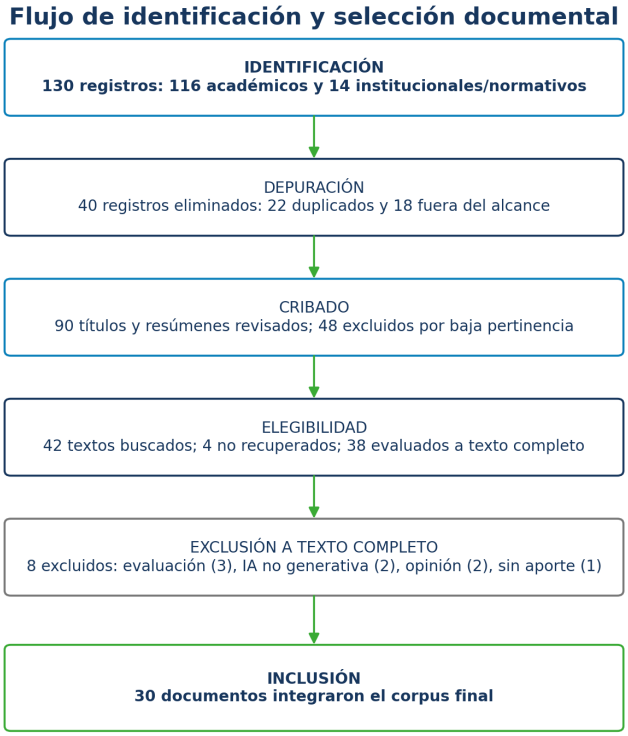
La interpretación se desarrolló con tres reglas. Primero, se distinguieron hallazgos empíricos de argumentos conceptuales y orientaciones institucionales. Segundo, se consideró convergencia cuando varias fuentes de diferente tipo respaldaban una misma dirección interpretativa, sin convertir el número de menciones en una medida de efecto. Tercero, las recomendaciones se formularon únicamente cuando podían relacionarse con un problema identificado y con principios de evaluación formativa, autenticidad, equidad o integridad. Debido a la diversidad del corpus, no se calcularon tamaños de efecto, ponderaciones estadísticas ni niveles agregados de certeza.

Consideraciones éticas y alcance

No se recopilaron datos personales ni se intervino con participantes, por lo que no fue necesaria aprobación de un comité de ética. Se respetaron la atribución de fuentes y el principio de no presentar inferencias documentales como mediciones de la realidad escolar. Los resultados describen tendencias del corpus y posibilidades normativas; no estiman prevalencias, efectos causales ni frecuencia de prácticas en instituciones ecuatorianas. La síntesis también evitó presentar los detectores automáticos como instrumentos diagnósticos y preservó la distinción entre sospecha tecnológica, evidencia académica y decisión institucional.

FIGURA 1

Flujo de identificación y selección documental



Nota. Adaptación del flujo PRISMA 2020 a partir de los registros conservados por el equipo autoral.

RESULTADOS

Los 30 documentos se distribuyeron en seis categorías analíticas. La síntesis mostró que la amenaza a la autenticidad no procede de una propiedad aislada de la IAG, sino de la interacción entre la capacidad de automatizar productos, el diseño de la tarea, el grado de transparencia y la diversidad de evidencias utilizadas para emitir el juicio evaluativo.

TABLA 1

Síntesis temática del corpus documental

Categoría	n	Hallazgo convergente	Implicación
Fundamentos metodológicos	5	Revisión integrativa, selección transparente y análisis temático.	La heterogeneidad exige síntesis cualitativa y límites explícitos.
IAG en educación	10	Apoyo a explicación, ideación y retroalimentación; riesgos de error, sesgo y dependencia.	La utilidad depende del diseño pedagógico y la alfabetización en IA.
Integridad y detección	4	Autoría incierta; falsos positivos y desempeño variable de detectores.	La detección automática no debe ser prueba única.
Evaluación con IAG	4	Rediseño de tareas y niveles explícitos de asistencia tecnológica.	Debe alinearse el uso permitido con el resultado de aprendizaje.

Evaluación formativa y auténtica	4	Proceso, desempeño situado, retroalimentación y metacognición.	Estas perspectivas reducen la dependencia del producto aislado.
Normativa ecuatoriana	3	Evaluación continua, contextualizada, flexible y orientada a la mejora.	Existe base normativa para innovar, pero no evidencia de implementación nacional.

La primera convergencia del corpus fue la vulnerabilidad de tareas cuyo juicio depende exclusivamente de un producto elaborado fuera de observación docente. Ensayos genéricos, resúmenes, respuestas declarativas y cuestionarios domiciliarios pueden ser producidos o reformulados con IAG. El riesgo no significa que todo producto asistido sea inválido; significa que el docente no puede inferir automáticamente comprensión a partir de su calidad superficial. Kasneci et al. (2023), Lo (2023) y Tlili et al. (2023) coinciden en que los modelos pueden generar explicaciones plausibles y lingüísticamente correctas incluso cuando contienen errores o cuando el usuario no comprende el contenido.

De esta convergencia emergió un cambio conceptual: la autenticidad se relaciona con la coherencia entre producto, proceso y desempeño. Un producto puede considerarse evidencia parcial cuando el estudiante conserva borradores, explica sus decisiones, contrasta las respuestas de la herramienta, reconoce sus limitaciones y demuestra el aprendizaje en otro formato. La defensa oral, la resolución de una variante del problema o la aplicación a un contexto local funcionan como evidencias complementarias, no como castigos por usar tecnología.

La segunda categoría mostró que la integridad académica no puede reducirse a una dicotomía entre uso y no uso. La IAG participa en grados diferentes: corrección lingüística, lluvia de ideas, generación de ejemplos, búsqueda de alternativas, elaboración parcial o producción total. La relevancia ética de cada grado depende de la consigna. Una ayuda que es legítima en una actividad destinada a evaluar argumentación crítica puede invalidar otra destinada a valorar escritura autónoma. Perkins et al. (2024) y Furze et al. (2024) proponen hacer explícitos los niveles de asistencia permitida y relacionarlos con los resultados esperados.

El corpus respalda tres condiciones de transparencia: informar antes de la tarea qué usos están permitidos; solicitar una declaración breve sobre herramientas, propósito y modificaciones; y exigir responsabilidad del estudiante sobre la exactitud y autoría del producto final. Esta solución evita equiparar toda asistencia con fraude y, al mismo tiempo, impide que una declaración de uso sustituya la demostración del aprendizaje.

Los estudios de Liang et al. (2023) y Weber-Wulff et al. (2023) convergen en que los detectores presentan errores y no ofrecen certeza suficiente para adjudicar autoría. Su salida depende del modelo evaluado, la longitud del texto, el idioma, las modificaciones posteriores y el estilo de quien escribe. Los falsos positivos pueden afectar de manera desproporcionada a quienes utilizan estructuras lingüísticas previsibles o escriben en una lengua distinta de la materna.

El hallazgo implica que un porcentaje de probabilidad no debe convertirse automáticamente en sanción. Ante una duda razonable, corresponde revisar evidencias de proceso, conversar con el estudiante, solicitar explicación del contenido y aplicar procedimientos institucionales. El detector, si se utiliza, solo puede operar como señal inicial y nunca como prueba autosuficiente.

Evaluación formativa y auténtica como respuesta

La literatura clásica y reciente coincidió en desplazar el centro de la evaluación hacia evidencias que permitan actuar sobre el aprendizaje. Black y Wiliam (1998) y Sadler (1989) sustentan la retroalimentación y el uso pedagógico de la evidencia; Wiggins (1990) y Gulikers et al. (2004) destacan tareas significativas y desempeño contextualizado. Frente a la IAG, estas perspectivas adquieren una función adicional: permiten observar razonamiento, transferencia, criterio y autorregulación.

Las estrategias más consistentes fueron portafolios de proceso, entregas escalonadas, anotación de cambios, preguntas orales, proyectos locales, estudios de caso, diarios metacognitivos, coevaluación y tareas presenciales breves. Ninguna estrategia garantiza por sí sola la autenticidad. Su fuerza surge de combinar evidencias y asegurar que la tarea exija decisiones que no puedan resolverse mediante reproducción genérica.

Los documentos ecuatorianos revisados definen la evaluación como continua y orientada a la mejora, y promueven actividades contextualizadas y retroalimentación oportuna (Ministerio de Educación del Ecuador, 2022). Estos principios son compatibles con el seguimiento del proceso, la diversificación de evidencias y el rediseño de tareas. La normativa no necesita ser interpretada como una defensa de productos estandarizados; permite justificar prácticas centradas en desempeño y acompañamiento.

No obstante, el corpus ecuatoriano fue normativo y no empírico. Por ello, el estudio no permite afirmar cuántas instituciones disponen de políticas de IAG, qué prácticas predominan ni cómo varían entre zonas urbanas, rurales o niveles educativos. El resultado defendible es que existe compatibilidad normativa para innovar, no que dicha innovación ya se encuentre implementada.

El análisis transversal permitió identificar cinco condiciones que aparecen de manera recurrente en las alternativas de rediseño: alineación, transparencia, trazabilidad, triangulación y proporcionalidad. La alineación exige que la regla de uso de IAG responda al resultado de aprendizaje. La transparencia comunica antes de la tarea qué herramientas y acciones se permiten. La trazabilidad conserva evidencias razonables del proceso. La triangulación combina dos o más formas de demostrar el aprendizaje. La proporcionalidad evita controles más invasivos o sanciones más severas que el riesgo observado.

La alineación fue la condición estructural. Cuando el objetivo es producir un texto con independencia, la generación automática debe restringirse y la evidencia puede incluir escritura supervisada. Cuando el objetivo es evaluar fuentes o argumentar, la IAG puede generar un insumo que el estudiante deba cuestionar. Cuando la competencia consiste en colaborar responsablemente con sistemas de IA, su uso debe integrarse de manera explícita. La claridad del objetivo evita reglas generales

incapaces de responder a la diversidad de tareas.

La triangulación fortalece la inferencia sin convertir cada evaluación en un procedimiento extenso. Un proyecto puede combinar producto, bitácora y defensa breve; un ensayo, esquema inicial, versión final y comentario metacognitivo; un problema, solución escrita y explicación de una variante. Si las evidencias convergen, el juicio gana solidez. Si se contradicen, la discrepancia orienta nuevas preguntas y retroalimentación antes de asumir una falta.

La proporcionalidad también apareció vinculada con la edad y el momento de aprendizaje. Estudiantes que comienzan a desarrollar una habilidad necesitan oportunidades de ejecución autónoma y acompañamiento más directo. En etapas posteriores, el uso crítico de herramientas puede aproximarse a prácticas profesionales. Por ello, una política institucional debe permitir que los docentes ajusten la asistencia según nivel, propósito y condiciones de acceso, manteniendo criterios comunes de transparencia y responsabilidad.

Riesgos pedagógicos y oportunidades de aprendizaje

La comparación de las categorías permitió organizar los efectos de la IAG en cuatro dimensiones: cognitiva, evaluativa, ética e institucional. En la dimensión cognitiva, las oportunidades incluyen acceso a explicaciones alternativas, generación de ejemplos y apoyo para revisar ideas; los riesgos comprenden aceptación acrítica, reducción de la práctica necesaria y dependencia. En la dimensión evaluativa, la herramienta facilita retroalimentación y creación de escenarios, pero debilita la atribución cuando la consigna no delimita su uso.

En la dimensión ética, el corpus destacó autoría, atribución, sesgos, privacidad y responsabilidad. Declarar la herramienta no libera al estudiante de comprobar la información ni convierte automáticamente el uso en legítimo. La transparencia debe indicar para qué se utilizó, qué partes fueron modificadas y quién asume la decisión final. También corresponde distinguir entre una herramienta que procesa información proporcionada por el usuario y otra que expone datos a servicios externos con condiciones de almacenamiento poco conocidas.

En la dimensión institucional, la principal oportunidad es transformar la irrupción de la IAG en un proceso de revisión de la calidad evaluativa. La principal amenaza es responder mediante reglas fragmentadas: prohibiciones diferentes entre asignaturas, sanciones sin procedimiento y dependencia de detectores. Las políticas efectivas requieren principios comunes, márgenes de decisión docente, ejemplos por área, mecanismos de apelación y actualización periódica.

Los resultados no respaldan una sustitución general de las evaluaciones escritas. La escritura sigue siendo una práctica esencial para organizar el pensamiento, construir argumentos y comunicar conocimiento. Lo que cambia es la suficiencia del texto final como evidencia exclusiva. Mantener la escritura y complementarla con proceso, reflexión o desempeño permite proteger su valor educativo sin ignorar las condiciones tecnológicas actuales.

TABLA 2

Matriz de decisiones para una evaluación auténtica con inteligencia artificial generativa

Resultado evaluado	Regla de uso de IAG	Evidencias complementarias
Escritura autónoma	No permitida o limitada a corrección formal	Borradores, escritura supervisada y explicación oral
Análisis crítico	Permitida para generar una respuesta que será auditada	Identificación de errores, contraste de fuentes y justificación
Proyecto aplicado	Permitida con declaración de uso	Bitácora, producto contextualizado y defensa del proyecto
Aprendizaje de IA	Integrada como objeto y herramienta	Calidad del prompt, verificación, reflexión ética y producto
Retroalimentación	Permitida como apoyo preliminar	Cambios argumentados y decisión final del estudiante

DISCUSIÓN

Los resultados permiten sostener que la IAG no anula la evaluación, pero sí reduce la validez de diseños que confunden fluidez del producto con aprendizaje. Esta interpretación coincide con Kasneci et al. (2023) y Cotton et al. (2024): la tecnología ofrece oportunidades educativas y, simultáneamente, facilita formas de delegación que las consignas tradicionales no hacen visibles. El desafío consiste en definir qué capacidad humana se evalúa y construir evidencia suficiente para atribuir el desempeño.

La principal contribución conceptual del estudio es proponer una autenticidad relacional. La autenticidad no reside únicamente en que una tarea sea "realista", sino en la relación verificable entre el estudiante, el proceso, las ayudas empleadas y el desempeño demostrado. En entornos con IAG, una tarea contextualizada también puede ser automatizada; por ello, la contextualización debe acompañarse de trazabilidad, explicación y transferencia. Esta ampliación complementa los marcos de Wiggins (1990) y Gulikers et al. (2004) sin abandonar sus fundamentos.

La segunda contribución es diferenciar asistencia de sustitución cognitiva. El uso de IA puede apoyar el aprendizaje cuando el estudiante compara alternativas, detecta errores, mejora una explicación o recibe retroalimentación que luego evalúa críticamente. Se aproxima a la sustitución cuando la herramienta ejecuta la capacidad que constituía el objetivo de aprendizaje y el estudiante entrega el resultado sin comprenderlo. Esta distinción evita políticas absolutas y permite reglas proporcionales a cada tarea.

La escala de evaluación con IA constituye una referencia útil porque traduce esa proporcionalidad en niveles comunicables (Furze et al., 2024; Perkins et al., 2024). Sin embargo, su adopción no debe reducirse a colocar una etiqueta en la consigna. El docente necesita explicar qué acciones están permitidas, cómo se declararán y qué evidencia será evaluada. También debe prever condiciones de equidad para estudiantes sin acceso estable a herramientas pagadas o conectividad.

La evidencia sobre detectores obliga a separar

prevención, investigación y sanción. Un detector puede alertar sobre una posible anomalía, pero carece de la estabilidad necesaria para demostrar falta por sí solo (Liang et al., 2023; Weber-Wulff et al., 2023). Una política responsable debe prohibir decisiones automáticas, proteger la presunción de buena fe, documentar el procedimiento y permitir que el estudiante explique su trabajo. Esta cautela es especialmente relevante en contextos multilingües y con desigualdades de alfabetización académica.

Esta separación tiene consecuencias para el debido proceso institucional. La prevención corresponde al diseño de consignas, la formación y la comunicación de reglas. La investigación comienza cuando existen inconsistencias concretas y debe reunir información contextual: versiones, fuentes, explicación del contenido y oportunidades de respuesta. La sanción solo procede cuando la evidencia alcanza el estándar previsto en la normativa interna. Mezclar estas etapas conduce a respuestas improvisadas en las que una señal probabilística se interpreta como veredicto. Además de injusto, ese procedimiento puede debilitar la relación pedagógica y generar incentivos para ocultar el uso de herramientas.

Para Ecuador, el hallazgo central es la compatibilidad entre el rediseño propuesto y el enfoque formativo de la normativa. Seguimiento, retroalimentación, contextualización y flexibilidad pueden sostener portafolios, proyectos y evaluaciones dialogadas. No obstante, la revisión no autoriza a afirmar que las prácticas escolares actuales sean homogéneas ni que exista una política nacional específica sobre IAG. Convertir esa ausencia de evidencia en una afirmación negativa reproduciría el mismo problema de validez que el artículo critica.

En educación básica y bachillerato, la mediación docente requiere una atención particular. Los estudiantes todavía construyen hábitos de autorregulación, criterios de autoría y capacidades para comprobar información. Pedir una declaración de uso sin enseñar a reconocer qué debe declararse tendría un efecto limitado. Las actividades pueden modelar el proceso: comparar una respuesta generada con el texto curricular, localizar afirmaciones no respaldadas, explicar por qué se acepta o rechaza una sugerencia y reconstruir una solución sin apoyo. Así, la alfabetización en IA se integra al aprendizaje disciplinar en lugar de convertirse en un contenido aislado.

La retroalimentación generada por IA merece un tratamiento específico. Puede ampliar la oportunidad de recibir comentarios iniciales, especialmente cuando el docente atiende grupos numerosos. Sin embargo, su utilidad depende de la calidad del criterio utilizado y de la capacidad del estudiante para interpretar las sugerencias. No toda recomendación mejora el texto ni toda corrección respeta el propósito de la tarea. Una práctica formativa consiste en solicitar que el estudiante clasifique las sugerencias, justifique cuáles incorpora y explique cuáles descarta. El objeto evaluado deja de ser la obediencia al sistema y pasa a ser el juicio ejercido sobre la retroalimentación.

La oralidad y el desempeño supervisado tampoco deben presentarse como soluciones universales. Las defensas orales pueden generar barreras para estudiantes con ansiedad, discapacidad comunicativa o necesidades lingüísticas; las actividades presenciales pueden verse afectadas por tiempo, tamaño del grupo y disponibilidad de recursos. La diversificación debe ofrecer modos equivalentes de demostrar comprensión,

con ajustes razonables y rúbricas transparentes. La autenticidad se fortalece cuando aumenta la calidad de la evidencia, no cuando se introduce una dificultad ajena al resultado de aprendizaje.

En términos prácticos, las instituciones pueden avanzar en cuatro niveles. Primero, acordar un vocabulario común sobre asistencia permitida, atribución y sustitución. Segundo, revisar las tareas de mayor vulnerabilidad y combinar productos con evidencia procesual. Tercero, formar a docentes y estudiantes en verificación, privacidad, sesgos y responsabilidad. Cuarto, establecer procedimientos proporcionales para resolver dudas de autoría sin depender de detectores. El marco de competencias docentes de UNESCO organiza esta formación en enfoque humano, ética, fundamentos, pedagogía y desarrollo profesional (Miao & Cukurova, 2024).

El primer nivel puede concretarse mediante una taxonomía institucional sencilla. Uso prohibido indica que la competencia debe demostrarse sin asistencia generativa; apoyo limitado permite funciones especificadas, como revisión ortográfica; colaboración permite ideación o retroalimentación con declaración; integración supone que la IAG forma parte del objetivo. Estas categorías deben acompañarse de ejemplos y revisarse por área, porque una misma acción puede tener significados diferentes en Lengua, Matemáticas, Ciencias o Educación Artística.

El segundo nivel demanda auditar las tareas y no al estudiantado. Una auditoría pedagógica puede preguntar si la consigna admite una respuesta genérica, si exige información contextual, si recoge decisiones intermedias, si permite verificar la comprensión y si ofrece alternativas equitativas. Las tareas vulnerables no siempre deben eliminarse: pueden convertirse en secuencias. Un ensayo domiciliario, por ejemplo, puede iniciar con una discusión, continuar con un esquema comentado, recibir retroalimentación y concluir con una breve defensa o aplicación. La secuencia distribuye la evidencia y mejora la retroalimentación.

El tercer nivel exige desarrollo profesional continuo. La formación técnica sobre plataformas envejece con rapidez; en cambio, los criterios sobre validez, privacidad, sesgo, atribución y diseño conservan mayor estabilidad. Los docentes necesitan experimentar con herramientas, analizar sus errores y adaptar consignas dentro de comunidades profesionales. También requieren respaldo institucional para evitar que cada caso de posible uso indebido se resuelva de manera individual y bajo presión.

El cuarto nivel debe incluir gobernanza de datos. Algunas plataformas almacenan instrucciones, documentos o interacciones, de modo que introducir nombres, calificaciones, diagnósticos o producciones identificables puede comprometer la privacidad. Las instituciones deberían definir qué información no puede cargarse, qué servicios están autorizados, cómo se informa a las familias y qué alternativas existen para quienes no consienten o no tienen acceso. La evaluación auténtica no puede construirse a costa de la protección de datos o de la igualdad de oportunidades.

La equidad debe ser un criterio transversal. Exigir herramientas específicas puede ampliar brechas de acceso; prohibirlas de modo absoluto puede impedir el desarrollo de competencias necesarias; y permitirlas sin acompañamiento puede favorecer a estudiantes con mayor autorregulación. Una política justa ofrece alternativas equivalentes,

informa sobre el tratamiento de datos, evita introducir información personal en plataformas y valora el juicio del estudiante por encima de la sofisticación del sistema utilizado.

La sostenibilidad de las innovaciones dependerá de que no incrementen de forma desproporcionada la carga docente. Conservar cada interacción o realizar defensas extensas para todas las tareas sería inviable. Resultan más realistas las muestras estratégicas, las entregas por hitos, las preguntas breves de comprobación y el uso de rúbricas compartidas. La institución puede concentrar la triangulación intensiva en evaluaciones de mayor peso y mantener procedimientos más ligeros en actividades formativas de bajo riesgo.

El papel de las rúbricas también debe revisarse. Criterios excesivamente generales, como claridad o creatividad, pueden aplicarse a productos generados sin revelar el aprendizaje. Una rúbrica adaptada al contexto de IAG puede valorar precisión conceptual, calidad de la verificación, justificación de decisiones, integración de evidencias, reconocimiento de límites y reflexión sobre el proceso. No se trata de añadir un criterio tecnológico a todas las actividades, sino de representar las capacidades que la consigna realmente pretende desarrollar.

La coevaluación ofrece otra posibilidad formativa. Cuando los estudiantes revisan productos y procesos de sus pares con criterios explícitos, deben identificar fortalezas, inconsistencias y decisiones que requieren explicación. Esta práctica puede desarrollar juicio evaluativo y ayudarles a reconocer respuestas genéricas o poco sustentadas sin convertirlos en vigilantes. Para funcionar, necesita ejemplos, modelado docente, reglas de respeto y protección de la información compartida.

La participación estudiantil en la construcción de las normas puede aumentar su legitimidad. Explicar por qué una ayuda está autorizada en una tarea y restringida en otra permite comprender la relación entre regla y aprendizaje. Los acuerdos de aula pueden incluir formas de atribución, preguntas sobre privacidad y procedimientos para solicitar orientación. La participación no elimina la responsabilidad institucional, pero transforma una política percibida como control externo en un marco comprensible de responsabilidad compartida.

Por último, el rediseño debe evaluarse. Implementar portafolios, declaraciones o defensas no garantiza mejoras si estas prácticas se convierten en requisitos burocráticos. Las instituciones pueden analizar la calidad de la evidencia obtenida, la comprensión de las reglas, la carga de trabajo, la accesibilidad y la percepción de justicia. La retroalimentación de docentes y estudiantes permitirá ajustar el sistema y evitar que una respuesta creada para proteger la autenticidad termine reduciendo el tiempo disponible para aprender.

El estudio presenta limitaciones. La documentación disponible no conservó fechas y rendimientos desagregados por base, lo que impide replicar de forma exacta la búsqueda; no hubo protocolo registrado ni evaluación uniforme del riesgo de sesgo; el corpus combinó fuentes académicas, conceptuales y normativas; y gran parte de la evidencia empírica provino de educación superior y contextos no ecuatorianos. Estas limitaciones restringen la generalización, pero no invalidan la síntesis conceptual. Futuras investigaciones deben levantar datos con docentes y estudiantes ecuatorianos, comparar niveles y territorios,

evaluar intervenciones de rediseño y estudiar la validez y equidad de nuevas formas de evidencia.

CONCLUSIONES

La inteligencia artificial generativa cuestiona la suficiencia de evaluaciones centradas exclusivamente en productos finales. La calidad lingüística o formal de una entrega ya no permite inferir, sin evidencias adicionales, comprensión, razonamiento o autoría. En respuesta al objetivo general, el desafío para la evaluación ecuatoriana consiste en fortalecer la relación entre producto, proceso, ayudas declaradas y desempeño demostrado.

Respecto al primer objetivo específico, la literatura muestra un balance entre oportunidades de apoyo y riesgos de error, sesgo, dependencia y delegación cognitiva. Sobre el segundo, los detectores automáticos no constituyen prueba suficiente de autoría o falta académica; la integridad requiere reglas previas, transparencia, evidencia múltiple y procedimientos justos.

En relación con el tercer objetivo, la evaluación formativa y auténtica ofrece fundamentos para recoger borradores, diálogo, aplicación contextualizada, retroalimentación y metacognición. La normativa ecuatoriana es compatible con estas prácticas, aunque el corpus no permite afirmar su grado de implementación en el país.

Finalmente, para cumplir el cuarto objetivo se proponen cuatro orientaciones: alinear el uso permitido de IAG con cada resultado de aprendizaje; solicitar declaraciones breves y verificables de uso; combinar el producto con evidencias de proceso y desempeño; y formar a la comunidad educativa en ética, verificación, privacidad y alfabetización en IA. La autenticidad no exige excluir toda tecnología, sino preservar la agencia y la responsabilidad intelectual del estudiante.

BIBLIOGRAFÍA

Asamblea Nacional del Ecuador. (2021). Ley Orgánica Reformatoria de la Ley Orgánica de Educación Intercultural. Registro Oficial, Suplemento 434.

Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. Information, 16(4), 296. https://doi.org/10.3390/info16040296

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7-74. https://doi.org/10.1080/0969595980050102

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, 43. https://doi.org/10.1186/s41239-023-00411-8

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, 22. https://doi.org/10.1186/s41239-023-00392-8

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

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52, 67-86. https://doi.org/10.1007/BF02504676

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, 102274. https://doi.org/10.1016/j.lindif.2023.102274

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

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410

Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Ministerio de Educación del Ecuador. (2022). Instructivo de evaluación estudiantil de los servicios educativos ordinarios. https://recursos.educacion.gob.ec/wp-content/uploads/2022/curriculares/Instructivo_de_Evaluacion_Estudiantil.pdf

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, n71. https://doi.org/10.1136/bmj.n71

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 & Learning Practice, 21(6). https://doi.org/10.53761/q3azde36

Presidencia de la República del Ecuador. (2023). Reglamento General a la Ley Orgánica de Educación Intercultural. Registro Oficial, Segundo Suplemento 254.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. Instructional Science, 18, 119-144. https://doi.org/10.1007/BF00117714

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x

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

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, 26. https://doi.org/10.1007/s40979-023-00146-z

Wiggins, G. (1990). The case for authentic assessment. Practical Assessment, Research, and Evaluation, 2(1), Article 2. https://doi.org/10.7275/ffb1-mm19

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, 40. https://doi.org/10.1186/s41239-024-00468-z

DERECHOS DE AUTOR

Carrillo Chacón, H. F., Vega Barberán, F. L., Cevallos Torres, D. A., & Torres Meza, M. L. (2026)



Este es un artículo de acceso abierto distribuido bajo la licencia Creative Commons de Atribución No Comercial 4.0, que permite su uso sin restricciones, su distribución y reproducción por cualquier medio, siempre que no se haga con fines comerciales y el trabajo original sea fielmente citado.

COMO CITAR:

Carrillo Chacón, H. F., Vega Barberán, F. L., Cevallos Torres, D. A., & Torres Meza, M. L. (2026). La evaluación educativa ecuatoriana ante la inteligencia artificial generativa: Desafíos para la autenticidad del aprendizaje. *REDSA Revista Ecuatoriana de Desarrollo Social yAmbiental, 1*(3) 125 - 142. https://doi.org/10.66646/redsa-2026/xx5c4v36

