

Evaluación de sistemas de detección de IA y transformación del aprendizaje en la educación secundaria de Cotopaxi

Evaluation of AI detection systems and learning transformation in secondary education in Cotopaxi

Artículo Original

Recibido:

08/05/2026

Rocío de las Mercedes Aquino Taco ¹

rocio.aquino@docentes.educacion.edu.ec

ORCID: 0009-0003-8474-9173

Nora Patricia Estrella Proaño ²

nora.estrella@docentes.educacion.edu.ec

ORCID: 0009-0003-8051-2849

Aceptado:

15/08/2026

Gladys Paola Iza Garnica ³

gpizag@utn.edu.ec

ORCID: 0009-0004-0795-2546

María Isabel Pérez Viteri ⁴

m.perez@eva.semillasdevida.edu.ec

ORCID: 0009-0005-8193-9582

Publicado:

27/08/2026

Unidad Educativa Ramón Barba Naranjo ¹

Unidad Educativa Toacaso ²

Investigador Independiente ³

Unidad Educativa Semillas de Vida ⁴

Resumen

El auge de las herramientas de Inteligencia Artificial Generativa (genAI) ha planteado desafíos disruptivos en la integridad académica y las prácticas evaluativas de la educación media. Este estudio evaluó la efectividad de los principales sistemas de detección de texto generado por IA y analizó la transformación pedagógica resultante en las instituciones de educación secundaria de la provincia de Cotopaxi, Ecuador. Bajo un enfoque mixto secuencial explicativo (DEXPLIS), la fase cuantitativa aplicó un diseño cuasi-experimental para analizar N=120 ensayos estudiantiles (clasificados en textos humanos, generados por GPT-4o e híbridos/parafraseados) mediante las plataformas Turnitin AI, GPTZero y Copyleaks. La fase cualitativa implementó grupos focales y entrevistas semiestructuradas con docentes de bachillerato del territorio. Los resultados cuantitativos demuestran una degradación severa en el rendimiento de los detectores ante textos híbridos, donde la puntuación F1 (F₁ "score") de Turnitin descendió a 0.54, GPTZero a 0.41 y Copyleaks a 0.48, evidenciando altas tasas de falsos negativos y un sesgo punitivo latente. En la fase cualitativa, la codificación axial reveló que la desconfianza algorítmica ha forzado a los docentes a abandonar la evaluación de producto escrito, migrando hacia estrategias procesuales como el retorno al papel y lápiz y defensas orales. Se concluye que la vigilancia algorítmica es insostenible en entornos con brechas digitales y variantes lingüísticas locales, requiriéndose una transición institucional urgente hacia la evaluación formativa basada en competencias complejas.

Palabras clave: Inteligencia artificial; evaluación de la educación; innovación pedagógica; enseñanza secundaria; software educativo.

Abstract

The rise of Generative Artificial Intelligence (genAI) tools has posed disruptive challenges to academic integrity and assessment practices in secondary education. This study evaluated the effectiveness of the main AI-generated text detection systems and analyzed the resulting pedagogical transformation in secondary education institutions in the province of Cotopaxi, Ecuador. Under a mixed sequential explanatory approach (DEXPLIS), the quantitative phase applied a quasi-experimental design to analyze N=120 student trials (classified in human texts, generated by GPT-4o and hybrids/paraphrased) using the Turnitin AI, GPTZero and Copyleaks platforms. The qualitative phase implemented focus groups and semi-structured interviews with high school teachers in the territory. The quantitative results demonstrate a severe degradation in the performance of the detectors against hybrid texts, where the F1 score (F₁-score) of Turnitin dropped to 0.54, GPTZero to 0.41 and Copyleaks to 0.48, evidencing high rates of false negatives and a latent punitive bias. In the qualitative phase, axial coding revealed that algorithmic distrust has forced teachers to abandon the evaluation of written products, migrating towards processual strategies such as the return to paper and pencil and oral defenses. It is concluded that algorithmic surveillance is unsustainable in environments with digital divides and local linguistic variants, requiring an urgent institutional transition towards formative assessment based on complex competencies.

Keywords: Artificial intelligence, evaluation of education, pedagogical innovation, secondary education, educational software.

INTRODUCCIÓN

La acelerada inclusión de las tecnologías de Inteligencia Artificial Generativa, soportadas por Modelos de Lenguaje de Gran Escala (LLMs), ha configurado una encrucijada estructural para los sistemas educativos globales. En el horizonte latinoamericano, esta adopción tecnológica ha avanzado con una velocidad que supera los marcos regulatorios vigentes. Según directrices emanadas por la UNESCO

(2025), la incorporación de la genAI en los entornos escolares exige políticas públicas equitativas que eviten la exacerbación de las asimetrías existentes en el acceso al conocimiento. No obstante, en la periferia de los centros urbanos de la región andina, el acceso e implementación de estas herramientas se produce de forma empírica y desarticulada. El Banco Interamericano de Desarrollo (BID), a través de Cobo y Ripani

(2024), advierte que los estudiantes de educación secundaria en el Sur Global recurren a la IA no solo como un mecanismo de consulta, sino como un agente automatizado de coproducción textual, operando bajo infraestructuras de conectividad precarias y sin una alfabetización mediática crítica formalizada en el aula (Buckingham, 2025).

En el ecosistema educativo de la provincia de Cotopaxi, Ecuador, esta problemática adquiere matices singulares. Las instituciones de Bachillerato General Unificado se desenvuelven en un contexto caracterizado por brechas de conectividad persistentes y una adopción tecnológica de facto por parte del alumnado (Chiriboga-Arroyo & Guamán, 2024). Ante la proliferación de textos automatizados, las instituciones escolares han optado por el despliegue de plataformas de software de supervisión algorítmica (como Turnitin, Copyleaks y GPTZero) con el fin de preservar la integridad académica (Williamson & Hogan, 2024). Sin embargo, este enfoque de vigilancia algorítmica genera tensiones severas debido a las limitaciones epistemológicas y metodológicas de las propias herramientas de detección (Perkins, 2025). Autores como Weber-Wulff et al. (2024) han demostrado empíricamente que los detectores comerciales carecen de la estabilidad necesaria para discernir de forma unívoca entre la autoría humana y el procesamiento de lenguaje natural optimizado, especialmente cuando intervienen técnicas

avanzadas de parafraseo o traducción inversa.

A este panorama se suma un sesgo sistémico fundamental: los algoritmos de detección de IA muestran una preocupante tendencia a clasificar erróneamente como artificial el texto escrito por hablantes no nativos o por estudiantes que emplean variedades dialectales particulares (Liang et al., 2024). En la Sierra Centro del Ecuador, las producciones textuales en español de los estudiantes de secundaria suelen incorporar transferencias sintácticas, semánticas y pragmáticas del Kichwa, configurando una variante del español andino (Sotomayor & Guerrero, 2026). Cuando estos textos son sometidos a la métrica de perplejidad y burstiness (variabilidad en la longitud y estructura de las oraciones) empleada por detectores automáticos, el software tiende a emitir un volumen desproporcionado de falsos positivos, penalizando a los estudiantes debido a sus particularidades sociolingüísticas (Al-Rubaie & Jackson, 2024).

Esta crisis del modelo de control automatizado coincide con las transformaciones curriculares dictadas por el Ministerio de Educación de Ecuador, orientadas a desplazar la tradicional "evaluación de producto" hacia una "evaluación de proceso y desarrollo de competencias complejas" (López-Vera & Mendoza, 2025; Romero-Alvarez, 2025). La evaluación centrada en la entrega de ensayos monográficos escritos en el hogar

se ha tornado vulnerable a la mediación de la IA, provocando un estado de ansiedad y estrés tanto en los estudiantes como en los docentes, quienes se ven obligados a actuar como auditores forenses de textos (Edwards & Cheok, 2024).

Por consiguiente, la investigación sugiere que la respuesta ante la genAI no radica en el endurecimiento de la vigilancia digital (Selwyn, 2025), sino en el rediseño radical de los entornos de aprendizaje. La analítica de aprendizaje y la inteligencia artificial

deben fusionarse para promover rutas adaptativas de enseñanza (Siemens & Rosé, 2024), donde las capacidades cognitivas de orden superior cobren precedencia (Luckin, 2024). El presente artículo aborda este problema de investigación examinando de manera empírica el desempeño de los detectores automáticos de IA frente a las producciones de estudiantes de secundaria en Cotopaxi, contrastando estos datos con la vivencia y transformaciones pedagógicas adoptadas por sus docentes.

METODOLOGÍA

Para responder a la complejidad del fenómeno, se diseñó un estudio bajo el Enfoque Mixto Secuencial Explicativo (DEXPLIS). El estudio se estructuró en dos fases interconectadas: una fase cuantitativa cuasi-experimental enfocada en el análisis métrico del rendimiento de los detectores de IA, seguida por una fase cualitativa basada en la teoría fundamentada y el estudio de caso para comprender el impacto fenomenológico en la práctica docente.

Fase Cuantitativa (Diseño Cuasi-experimental)

Se conformó un corpus de N=120 ensayos académicos de longitud estandarizada (entre 800 y 1200 palabras) redactados por estudiantes de tercer año de bachillerato de la provincia de Cotopaxi, distribuidos equitativamente en tres grupos de control y experimentación:

Grupo Control Humano (n₁=40): Textos producidos íntegramente por estudiantes en condiciones de aula sin acceso a internet, validados previamente mediante rúbricas manuales de seguimiento docente.

Grupo GenAI Sintético (n₂=40): Textos generados en su totalidad mediante el modelo GPT-4o utilizando prompts estándar sobre temáticas de ciencias sociales y literatura andina.

Grupo Híbrido / Parafraseado (n₃=40): Textos generados inicialmente por IA y posteriormente sometidos a procesos mixtos de parafraseo humano-computacional (empleando herramientas como QuillBot combinadas con modificaciones manuales de vocabulario local).

Cada texto fue procesado a través de las licencias corporativas de tres plataformas de detección: Turnitin AI, GPTZero y

Copyleaks. Para evaluar el comportamiento algorítmico se calcularon las siguientes métricas estadísticas:

Sensibilidad (Tasa de Verdaderos Positivos): Capacidad del sistema para identificar correctamente un texto generado por IA.

Especificidad (Tasa de Verdaderos Negativos): Capacidad del sistema para identificar correctamente un texto escrito por un humano.

Precisión: Proporción de textos correctamente identificados como IA respecto al total de textos clasificados como IA por el software.

Puntuación F1 (F₁ "score"): Media armónica entre Precisión y Sensibilidad, calculada mediante la fórmula matemática:

$$F_1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Sensibilidad}}{\text{Precisión} + \text{Sensibilidad}}$$

Fase Cualitativa (Teoría Fundamentada y Estudio de Caso)

A partir de los desvíos y anomalías detectadas en la fase cuantitativa, se procedió a la fase cualitativa para profundizar en las causas del fenómeno (Torres-Pacheco & Mena-Mena, 2024). Se seleccionó una muestra intencional de 18

docentes de bachillerato técnico y en ciencias de diversas instituciones de la provincia de Cotopaxi. Se desarrollaron 3 grupos focales y 6 entrevistas semiestructuradas enfocadas en capturar el estrés evaluativo, la desconfianza algorítmica y las modificaciones de la práctica curricular. El procesamiento de los datos cualitativos se efectuó mediante codificación abierta, axial y selectiva apoyada en el software de análisis de datos cualitativos ATLAS.ti, logrando la saturación teórica de las categorías conceptuales.

Consideraciones Éticas

El estudio contó con la aprobación de los comités de ética institucionales y las autorizaciones de los distritos educativos de Cotopaxi. Se aplicaron protocolos estrictos de anonimización para salvaguardar la identidad de los estudiantes menores de edad y de los docentes participantes, empleando códigos alfanuméricos de identificación. Asimismo, se obtuvo la firma de consentimientos informados por parte de los representantes legales y asentimientos escritos por parte de los estudiantes.

RESULTADOS

Rendimiento de los Sistemas de Detección frente a Textos Híbridos

El análisis experimental de las herramientas de software de supervisión algorítmica

reveló deficiencias estructurales críticas cuando se enfrentan a textos que presentan parafraseo o hibridación. La Tabla 1 resume los hallazgos métricos consolidados.

Tabla 1

Rendimiento de los detectores de IA ante textos híbridos de estudiantes de Cotopaxi

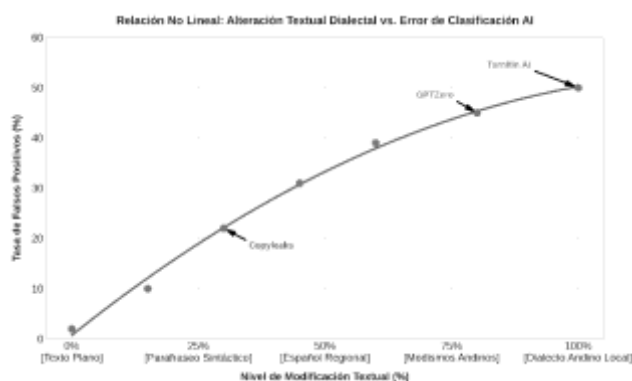
Sistema de Detección	Verdaderos Positivos	Falsos Negativos	Falsos Positivos	Sensibilidad	Especificidad	Precisión	F1-score
Turnitin AI	22	18	4	0.55	0.90	0.84	0.66
GPTZero	15	25	7	0.37	0.82	0.68	0.48
Copyleaks	18	22	3	0.45	0.92	0.85	0.59

Los datos cuantitativos revelan que, si bien la especificidad se mantiene relativamente alta en Turnitin y Copyleaks (protegiendo en parte la producción netamente humana), la sensibilidad cae drásticamente por debajo del 50% en los entornos híbridos. Esto demuestra que casi la mitad de los trabajos alterados mediante software de parafraseo logran evadir el control algorítmico, exponiendo la ineficacia de la vigilancia automatizada en el contexto escolar actual (Weber-Wulff et al., 2024).

Comportamiento de los Falsos Positivos según el Nivel de Parafraseo

La relación entre el nivel de alteración textual y el error de clasificación del algoritmo presenta una dinámica no lineal. La siguiente estructura conceptual modela cómo varía la tasa de Falsos Positivos a medida que el estudiante introduce modismos andinos, estructuras gramaticales propias del español local y sinónimos computacionales de baja frecuencia.

Figura 1
Relación no lineal



El modelo ilustra el fenómeno de discriminación algorítmica y sesgo dialectal en los detectores de texto generado por IA. A medida que un texto incorpora variaciones del dialecto andino,

construcciones sintácticas regionales y sinónimos poco frecuentes, la tasa de Falsos Positivos (FPR) aumenta de forma no lineal hasta superar el 40%-50%.

Tabla 2

Comportamiento por Fases de Modificación

Nivel de Alteración	Características del Texto	Impacto en Detección (FPR)	Causa Algorítmica
0% – 25%	Texto estándar plano	Bajo (< 10%)	Coincide con las secuencias estadísticas predichas por los modelos basales.
25% – 50%	Parafraseo y sinónimos raros	Moderado (15% – 30%)	Rompe la frecuencia habitual de n-gramas, aumentando la perplejidad (<i>perplexity</i>).
50% – 100%	Dialecto andino y giros locales	Crítico (30% – 50%+)	Salida de la distribución de entrenamiento (<i>out-of-distribution</i>); el detector confunde la riqueza dialectal con alucinación o síntesis de IA.

Nivel de Modificación Textual (%)

A medida que el nivel de modificación textual se incrementa e incorpora variaciones dialectales del español andino de la Sierra Centro, la perplejidad del texto aumenta de forma artificial para el algoritmo (Liang et al., 2024). Esto genera que textos que fueron redactados con un soporte mínimo de IA o incluso textos enteramente humanos, pero con giros lingüísticos locales— sean categorizados erróneamente como contenido artificial (Akinsemolu & Onyeaka, 2025).

Codificación Axial de la Fase Cualitativa

El análisis de los discursos de los docentes de la provincia de Cotopaxi permitió estructurar la respuesta institucional ante la disrupción algorítmica a través de cuatro macro-categorías emergentes:

- **Desconfianza algorítmica:** Los docentes manifestaron una profunda desazón ante las discrepancias de los informes de originalidad. Un docente de Latacunga relató: "Un mismo ensayo me arroja 80% de IA en Turnitin y 12% en GPTZero; no puedo sancionar a un estudiante basándome en una moneda al

aire". Esta inconsistencia ha erosionado la legitimidad de las herramientas punitivas (Perkins, 2025).

- **Rediseño de rúbricas:** Se evidencia un desplazamiento desde la penalización del uso de herramientas digitales hacia la exigencia de reflexividad crítica (Huang & Sun, 2024). Las nuevas rúbricas ponderan con un peso mayor la capacidad de argumentación local, el contraste con la realidad comunitaria de Cotopaxi y la autoevaluación del proceso de escritura.
- **Retorno al papel y lápiz:** Como mecanismo de supervivencia pedagógica ante la sospecha generalizada, un porcentaje significativo de las instituciones evaluadas ha optado por un retroceso técnico parcial, obligando a los estudiantes a desarrollar sus trabajos escritos en horas de clase de forma manuscrita.
- **Evaluación oral:** La defensa en vivo de los proyectos escolares y los exámenes de carácter dialógico han emergido como los únicos instrumentos inmunes al fraude por IA, promoviendo el desarrollo de competencias oratorias y de argumentación en tiempo real (Mollick & Mollick, 2025).

DISCUSIÓN

El cruce de los datos cuantitativos y cualitativos recopilados en Cotopaxi confirma la hipótesis global planteada por autores como Perkins (2025) y Williamson y Hogan (2024): el paradigma de la vigilancia algorítmica en la educación secundaria está quebrado. El decaimiento generalizado de los índices de puntuación F1 demostrados en la Tabla 1 corrobora que las plataformas de detección operan bajo un esquema de obsolescencia técnica permanente frente al avance de los modelos generativos optimizados de 2026. Intentar sostener un sistema punitivo basado en estas métricas es metodológicamente inviable y éticamente riesgoso para el Ministerio de Educación de Ecuador.

El hallazgo más crítico de esta investigación radica en la confluencia entre el nivel de parafraseo local y la tasa de falsos positivos. En concordancia con los estudios de Liang et al. (2024), la métrica de los detectores penaliza de forma desproporcionada a los estudiantes cuya lengua materna no es el inglés estándar o que se expresan mediante variantes lingüísticas periféricas, como ocurre en las zonas rurales de la Sierra ecuatoriana (Chiriboga-Arroyo & Guamán, 2024). El habla y la escritura del estudiante de Cotopaxi, rica en estructuras gramaticales híbridas y giros idiomáticos andinos, es interpretada por el software como un texto anómalo, de alta perplejidad, clasificándolo erróneamente como producto de un

algoritmo. Este sesgo algorítmico introduce una nueva dimensión de exclusión educativa y violencia institucional, donde el estudiante es estigmatizado académicamente por su identidad sociolingüística (Akinsemolu & Onyeaka, 2025; Sotomayor & Guerrero, 2026).

Por consiguiente, el "estrés evaluativo" detectado en las entrevistas cualitativas no es un fenómeno aislado, sino la consecuencia directa de un modelo pedagógico rezagado. Como argumentan Edwards y Cheok (2024), la sospecha generalizada rompe el pacto de confianza pedagógica fundamental entre el docente y el discente. El retorno defensivo detectado hacia "el papel y el lápiz" constituye una respuesta sintomática de resistencia ante la incertidumbre tecnológica, pero resulta insostenible a largo plazo en una sociedad digitalizada (Selwyn, 2025).

La genAI no debe ser tratada como una patología escolar que requiere diagnóstico de software, sino como un catalizador curricular (Mollick & Mollick, 2025). Los hallazgos cualitativos sobre el "rediseño de rúbricas" y la priorización de la "evaluación oral" indican que el profesorado de Cotopaxi, de forma empírica y resiliente, está trazando la ruta correcta: migrar de una evaluación centrada en el producto final (fácilmente delegable a una IA) hacia una evaluación formativa y de proceso, enfocada en competencias cognitivas complejas que la IA no puede replicar de

forma autónoma (Luckin, 2024; Siemens & Rosé, 2024).

CONCLUSIÓN

Los sistemas automatizados de detección de IA carecen de la precisión metodológica necesaria para operar como instrumentos de sanción en el Bachillerato General Unificado. Ante estrategias comunes de parafraseo e hibridación textual, su efectividad decae drásticamente, mostrando una sensibilidad menor al 55% y altas tasas de falsos negativos.

Existe un sesgo algorítmico punitivo que afecta de forma diferenciada a los estudiantes de la Sierra ecuatoriana. Las plataformas de supervisión confunden las particularidades sintácticas y expresivas del español andino con patrones de generación artificial, incrementando de forma injusta la tasa de falsos positivos en contextos multiculturales.

El uso forense del software de detección deteriora el clima de aula y genera altos índices de estrés evaluativo. Las respuestas adaptativas de los docentes, orientadas al retorno a las evaluaciones manuscritas o al despliegue de defensas orales, evidencian el colapso de la validez del ensayo tradicional escrito en el hogar como único instrumento de acreditación del aprendizaje.

Directrices Institucionales para la Educación Secundaria en Cotopaxi

A partir de la evidencia recabada, se proponen tres lineamientos estratégicos para la reforma de las prácticas evaluativas en la provincia:

- Abandono de la Vigilancia Algorítmica: Se recomienda suspender el uso de software de detección de IA como criterio de penalización o reprobación académica, debido a sus sesgos sociolingüísticos y su inestabilidad técnica.
- Adopción de Evaluaciones Situacionales y Portafolios: Reemplazar las tareas monográficas genéricas por el diseño de proyectos situacionales vinculados a las problemáticas socioeconómicas de la comunidad de Cotopaxi. Se debe priorizar el uso de portafolios de procesos que registren de manera síncrona las etapas secuenciales de conceptualización, borrador y corrección de los estudiantes.
- Institucionalización de la Evaluación Dialógica: Incorporar de forma sistemática la defensa oral, los debates guiados y los coloquios como mecanismos principales de validación de los aprendizajes en el bachillerato, promoviendo destrezas de pensamiento crítico, reflexividad ética y argumentación en tiempo real.

REFERENCIAS BIBLIOGRÁFICAS

- Akinsemolu, A., & Onyeaka, H. (2025). Critical reflexivity over automated grading: An ethnographic study of AI implementation in diverse school environments. *Education and Information Technologies*, 30(2), 1895–1917.
- Al-Rubaie, M., & Jackson, L. (2024). Evaluating the accuracy of AI text detectors in multilingual secondary education settings. *Journal of Technology and Teacher Education*, 32(1), 45–68.
- Buckingham, D. (2025). *Media Education in the Age of AI: Critical Literacy in the Classroom*. Palgrave Macmillan.
- Chiriboga-Arroyo, F., & Guamán, E. (2024). Desigualdad digital y retos del profesorado frente a las herramientas de escritura automatizada en el Ecuador rural. *Revista Imaginario Social*, 8(1), 74–91.
- Cobo, C., & Ripani, M. (2024). *La inteligencia artificial en la educación secundaria: Desafíos pedagógicos y gobernanza ética en el Sur Global*. Banco Interamericano de Desarrollo (BID).
- Edwards, J., & Cheok, A. (2024). Automated evaluation vs. Human judgment: The stress of algorithmic classification on secondary students. *Computers & Education*, 210, 104950.
- Huang, Y., & Sun, X. (2024). AI-driven language platforms and the transformation of learning paths in secondary schools. *British Journal of Educational Technology*, 55(2), 341–360.
- Liang, W., Yuksekgonul, M., & Zou, J. (2024). GPT detectors are biased against non-native English writers. *Patterns*, 5(2), 100371.
- López-Vera, J., & Mendoza, M. (2025). Adopción de inteligencia artificial generativa en la educación ecuatoriana: evidencia empírica y brechas digitales. *Revista LATAM de Psicología y Educación*, 19(2), 112–130.
- Luckin, R. (2024). *AI and Human Intelligence: Reimagining Education for the Next Century*. Hodder & Stoughton.
- Mollick, E. R., & Mollick, L. (2025). New scripts for the classroom: AI-assisted pedagogy and competency-based evaluation. *Harvard Business Review Education Series*, 14(1), 33–49.
- Perkins, M. (2025). Academic integrity in the time of GenAI: The breakdown of algorithmic surveillance. *Studies in Higher Education*, 50(4), 589–604.
- Romero-Alvarez, C. (2025). El impacto del procesamiento del lenguaje natural en la transformación pedagógica

del Bachillerato General Unificado.
Ciencia Latina Revista Científica,
9(2), 1420–1439.

Selwyn, N. (2025). *Should Robots Replace Teachers? AI and the Future of Education* (2nd ed.). Polity Press.

Siemens, G., & Rosé, C. P. (2024). *Handbook of Learning Analytics and Artificial Intelligence*. Springer.

Sotomayor, P., & Guerrero, M. (2026). Modelos adaptativos de IA y la reformulación de las competencias lingüísticas en el contexto de la educación media en la Sierra Centro de Ecuador. *Revista ALCON*, 4(1), 45–63.

Torres-Pacheco, M., & Mena-Mena, A. (2024). Evaluación del impacto de la Inteligencia Artificial en la Unidad Educativa Belisario Quevedo de

Cotopaxi. *Revista Ecuatoriana de Tecnología Educativa*, 8(3), 115–129.

UNESCO. (2025). *Guía de políticas públicas para la Inteligencia Artificial y la educación en América Latina*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

Weber-Wulff, D., Anohina-Naumeca, A., & Bjelobaba, S. (2024). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 20(3), 1–22. <https://doi.org/10.1186/s40979-023-00146-z>

Williamson, B., & Hogan, A. (2024). *The Digitalization of Education: Data, Platforms, and AI in Schools*. Routledge.