

Modelo de detección de textos generados con inteligencia artificial y generación de propuestas de retroalimentación pedagógica.

Model for detecting artificial intelligence-generated texts and generating pedagogical feedback proposals.

Mario Alberto Camarillo Cabrera 

Maestría en Ingeniería en Sistemas Computacionales, Tecnológico de Estudios Superiores de Ecatepec, 55210, Ecatepec de Morelos, Estado de México, México.

Resumen

La expansión de la inteligencia artificial generativa en la educación superior ha intensificado la necesidad de distinguir entre escritura académica humana, escritura asistida y textos producidos automáticamente. Sin embargo, los detectores disponibles presentan limitaciones técnicas, sesgos lingüísticos y riesgos éticos cuando se usan como instrumentos sancionadores. Este artículo propone un modelo conceptual de detección y retroalimentación pedagógica orientado a textos académicos en español. La propuesta integra procesamiento del lenguaje natural, rasgos estilométricos, clasificación algorítmica e interpretación pedagógica para convertir señales de posible generación automática en recomendaciones formativas. El trabajo no reporta todavía un corpus experimental cerrado ni métricas empíricas de desempeño; por ello, sus resultados se presentan como diseño metodológico, arquitectura funcional y protocolo de validación. El aporte principal consiste en articular la detección técnica con retroalimentación descriptiva, diagnóstica y prescriptiva, de modo que la herramienta apoye la autoría académica, la revisión de textos y la integridad educativa sin sustituir el juicio docente.

Palabras clave: inteligencia artificial generativa, detección de textos, retroalimentación formativa, escritura académica, procesamiento del lenguaje natural.

Abstract

The expansion of generative artificial intelligence in higher education has increased the need to distinguish between human academic writing, AI-assisted writing, and automatically generated texts. However, current detectors show technical limitations, linguistic bias, and ethical risks when used as punitive instruments. This paper proposes a conceptual model for detecting AI-generated academic texts in Spanish and producing pedagogical feedback. The proposal integrates natural language processing, stylistometric features, algorithmic classification, and pedagogical interpretation to transform signals of possible automatic generation into formative recommendations. The study does not yet report a closed experimental corpus or empirical performance metrics; therefore, the results are presented as a methodological design, functional architecture, and validation protocol. The main contribution is the articulation of technical detection with descriptive, diagnostic, and prescriptive feedback, so the tool may support academic authorship, text revision, and educational integrity without replacing teachers' professional judgment.

Keywords: generative artificial intelligence, text detection, formative feedback, academic writing, natural language processing.

1. Introducción

La inteligencia artificial generativa ha modificado de manera acelerada las condiciones de producción, revisión y evaluación de textos académicos. En el ámbito universitario, los modelos de lenguaje extensos permiten generar borradores, resúmenes, explicaciones, paráfrasis y respuestas argumentativas con un nivel de fluidez que dificulta distinguir cuándo un texto expresa una construcción autoral propia y cuándo reproduce una salida automatizada. Esta situación no

solo introduce un problema técnico de clasificación textual, sino también un problema pedagógico: si la detección se usa únicamente para sancionar, se corre el riesgo de producir acusaciones injustas, debilitar la confianza institucional y desaprovechar una oportunidad para enseñar escritura académica, pensamiento crítico y uso responsable de la tecnología.

La literatura reciente advierte que la detección automática de textos generados por IA no debe asumirse como evidencia definitiva. Existen métodos basados en probabilidad,

*Autor para la correspondencia: 202510890@tese.edu.mx

Correo electrónico: 202510890@tese.edu.mx (Mario-Alberto-Camarillo-Cabrera).

Historial del manuscrito: recibido el 11/04/2026, última versión-revisada recibida el 29/05/2026, aceptado el 01/06/2026, publicado el 05/10/2026. DOI: <https://doi.org/10.5281/zenodo.23127130>

curvatura, rasgos estilométricos y clasificadores supervisados, pero su desempeño depende del idioma, el dominio, la longitud del texto, el modelo generador, la edición humana posterior y las condiciones del corpus. Además, los detectores comerciales suelen operar con baja transparencia metodológica, lo cual complica la reproducibilidad y la interpretación de sus resultados en contextos educativos.

En este marco, el presente artículo propone un modelo conceptual para analizar textos académicos en español y generar retroalimentación pedagógica. El énfasis no está en sustituir la evaluación docente ni en emitir veredictos automáticos, sino en diseñar una arquitectura que permita identificar señales lingüísticas relevantes y traducirlas en orientaciones útiles para la mejora de la escritura. La propuesta se ubica en una fase de diseño metodológico y validación preliminar conceptual; por tanto, no se presentan métricas experimentales definitivas ni se afirma la existencia de un sistema entrenado en producción.

Enunciado del problema

Las instituciones de educación superior enfrentan dificultades para responder al uso de inteligencia artificial generativa en tareas escritas. Por un lado, necesitan proteger la integridad académica y promover la autoría auténtica; por otro, deben evitar prácticas de vigilancia algorítmica que generen falsos positivos o decisiones sancionadoras sin evidencia suficiente. El problema central consiste en que la mayoría de las soluciones de detección se concentran en clasificar un texto como “humano” o “generado por IA”, pero no explican de manera pedagógica qué rasgos del texto requieren revisión ni cómo el estudiante puede mejorar su escritura.

En el caso de textos en español, el reto se amplía debido a la menor disponibilidad de corpus documentados, la variabilidad regional del idioma, la diversidad de géneros académicos y la falta de validaciones públicas comparables. La ausencia de un enfoque integrado provoca que la detección sea técnicamente frágil y pedagógicamente limitada. En consecuencia, se requiere un modelo que combine análisis lingüístico, interpretación contextual y retroalimentación formativa, sin presentar la clasificación algorítmica como prueba absoluta.

Objetivo y alcance

El objetivo del trabajo es diseñar un modelo conceptual de detección de textos generados con inteligencia artificial y generación de propuestas de retroalimentación pedagógica para textos académicos en español. El alcance se limita a la definición de arquitectura, módulos funcionales, flujo de datos, criterios de análisis lingüístico, lógica de retroalimentación y protocolo de evaluación futura. No se reporta entrenamiento final de un clasificador, repositorio de código, corpus validado ni resultados estadísticos concluyentes; esos elementos quedan definidos como líneas de validación empírica posterior.

2. Estado del arte

La detección de texto generado por IA se ha desarrollado mediante aproximaciones diversas. Un primer grupo utiliza clasificadores supervisados entrenados con textos humanos y

textos sintéticos; su ventaja es la posibilidad de aprender patrones específicos del corpus, pero su principal debilidad es la dependencia del conjunto de entrenamiento. Cuando el dominio, idioma o modelo generador cambian, el desempeño puede deteriorarse. Un segundo grupo emplea señales probabilísticas del propio modelo de lenguaje, como la predictibilidad, la entropía o la curvatura de la probabilidad. DetectGPT, por ejemplo, plantea que los textos generados por modelos tienden a ubicarse en regiones de curvatura negativa de la función de log-probabilidad, lo que permite una detección sin entrenar un clasificador independiente.

A pesar de estos avances, la literatura muestra límites relevantes. Los detectores pueden fallar ante paráfrasis, edición humana ligera, traducción, textos breves o estilos de escritura muy estandarizados. También se han reportado sesgos contra escritores no nativos del inglés, lo cual es especialmente importante para comunidades académicas multilingües. En educación, este riesgo implica que un sistema de detección no debe convertirse en juez automático, porque una clasificación errónea puede afectar la trayectoria de un estudiante.

Además, investigaciones recientes sobre integridad académica y evaluación de detectores advierten que la IA generativa puede producir textos plausibles pero difíciles de verificar, y que las herramientas de detección presentan resultados variables según el modelo generador, el género textual y las condiciones de evaluación (Cotton et al., 2024; Elkhatat et al., 2023; Gao et al., 2023; Perkins, 2023). Por ello, el estado del arte coincide en que los detectores no deben emplearse como única evidencia sancionadora, sino como instrumentos de apoyo sujetos a revisión humana y criterios pedagógicos explícitos.

Desde la perspectiva educativa, la discusión no puede reducirse a identificar fraude. La IA generativa puede apoyar procesos de planeación, explicación y revisión, pero también exige criterios de transparencia, alfabetización digital y responsabilidad académica. Las recomendaciones internacionales insisten en una visión centrada en las personas, con reglas institucionales claras, protección de datos y formación docente y estudiantil. Por ello, una herramienta de detección útil para educación debe explicar sus límites, evitar decisiones punitivas automáticas y contribuir al aprendizaje.

En el campo de la retroalimentación, la evidencia pedagógica destaca que los comentarios más efectivos son específicos, oportunos, comprensibles y orientados a la mejora. La retroalimentación formativa debe ayudar al estudiante a responder tres preguntas: hacia dónde va, cómo está avanzando y qué acciones puede realizar después. Desde el aprendizaje autorregulado, esta información permite revisar estrategias, reconocer errores recurrentes y planear mejoras. En consecuencia, el vacío que atiende esta propuesta no es únicamente técnico; consiste en vincular la detección lingüística con una intervención pedagógica explícita.

Tabla 1: Síntesis de trabajos relacionados y aportación de la propuesta.

Línea de trabajo	Aporte principal	Limitación para este estudio	Respuesta propuesta
Detectores probabilísticos y de curvatura	Identifican señales estadísticas asociadas a texto generado por LLM.	Pueden depender del modelo generador y de condiciones específicas de prueba.	Usar la señal técnica como indicio, no como veredicto final.
Clasificadores supervisados	Permiten entrenar modelos con rasgos lingüísticos y corpus etiquetados.	Requieren corpus documentado y validación por idioma y dominio.	Definir protocolo de corpus en español antes de reportar métricas.
Críticas a detectores comerciales	Evidencian fragilidad, sesgo y baja transparencia.	Dificultan replicar resultados y justificar decisiones educativas.	Exigir explicabilidad, revisión humana y comunicación de incertidumbre.
Retroalimentación formativa	Orienta al estudiante hacia acciones concretas de mejora.	No siempre se integra con análisis computacional.	Convertir rasgos detectados en comentarios descriptivos y prescriptivos.

Fuente: elaboración propia a partir del alcance metodológico del estudio.

3. Materiales y Método

El estudio se plantea como una investigación aplicada de diseño y desarrollo tecnológico en fase conceptual. La metodología se organiza en cuatro etapas: revisión documental, diseño de arquitectura funcional, especificación de criterios lingüísticos y pedagógicos, y propuesta de protocolo de validación. Esta reformulación es necesaria porque el manuscrito no documenta todavía un corpus experimental cerrado, un modelo entrenado verificable ni resultados reproducibles. En lugar de presentar hallazgos estadísticos no comprobados, el artículo delimita una arquitectura evaluable y un plan de validación futuro.

Diseño general de la solución propuesta

La solución propuesta opera como un sistema modular. El texto académico ingresa al sistema, se normaliza, se segmenta y se representa mediante características lingüísticas. Posteriormente, un módulo de clasificación o puntuación estima señales de posible generación automática. El resultado técnico se interpreta junto con información contextual del género académico, la longitud del texto y el propósito de escritura. Finalmente, un módulo pedagógico convierte los rasgos observados en recomendaciones de mejora. La Figura 1 resume esta arquitectura.

Arquitectura conceptual del sistema



Figura 1: Arquitectura conceptual del sistema propuesto. Fuente: elaboración propia.

Módulos funcionales

El primer módulo corresponde al ingreso y preprocesamiento. Su función es recibir el texto, normalizar caracteres, eliminar ruido mínimo, segmentar oraciones y preparar unidades de análisis. El segundo módulo extrae rasgos lingüísticos y estilométricos, tales como longitud promedio de oración, diversidad léxica, repetición de conectores, densidad nominal, presencia de afirmaciones genéricas, patrones de cohesión y variación sintáctica. Estos rasgos no deben interpretarse de manera aislada, porque un texto humano también puede mostrar regularidad y un texto generado puede ser editado posteriormente.

El tercer módulo corresponde a la clasificación o puntuación. En una futura implementación, este componente podría entrenarse con aprendizaje supervisado o calibrarse mediante reglas ponderadas y revisión experta. En el estado actual del artículo, se define como componente metodológico, no como modelo empírico ya validado. El cuarto módulo realiza la interpretación contextual para evitar decisiones binarias. El quinto módulo genera retroalimentación pedagógica en tres niveles: descriptivo, al explicar qué se observa; diagnóstico, al señalar por qué puede afectar la calidad académica; y prescriptivo, al proponer acciones de reescritura.

Tabla 2: Componentes técnicos y pedagógicos del modelo conceptual.

Módulo	Entrada	Proceso	Salida
Preprocesamiento	Texto del estudiante	Normalización, segmentación y tokenización	Texto preparado para análisis
Extracción de rasgos	Texto segmentado	Cálculo de rasgos léxicos, sintácticos y discursivos	Vector de características
Clasificación o puntuación	Vector de características	Estimación de señales de posible generación automática	Puntuación interpretable, no concluyente
Interpretación contextual	Puntuación y metadatos académicos	Análisis por género, longitud y propósito del texto	Nivel de riesgo y advertencias
Retroalimentación	Rasgos y nivel de riesgo	Generación de comentario descriptivo, diagnóstico y prescriptivo	Reporte formativo para revisión

Fuente: elaboración propia a partir del alcance metodológico del estudio.

Representación formal del análisis

Para evitar una ecuación ambigua, el modelo se expresa como una función de puntuación que recibe una representación lingüística del texto. Sea t el texto analizado y sea $\varphi(t)$ el vector de características lingüísticas y estilométricas extraídas. La función $f\theta$ representa el clasificador o mecanismo de puntuación que deberá calibrarse en la fase empírica. La salida pIA no debe leerse como certeza absoluta, sino como una estimación condicionada al modelo, al corpus y a los criterios de validación.

$$pIA = f\theta(\varphi(t)) \quad (1)$$

En la ecuación (1), t representa el texto de entrada, $\varphi(t)$ corresponde al vector de rasgos calculados, θ denota los parámetros del modelo o los criterios de calibración, y pIA indica una puntuación de probabilidad o riesgo de generación

automática. Dado que el artículo no documenta un entrenamiento empírico final, θ y pIA se presentan como parte de la arquitectura metodológica, no como resultados medidos.

Descripción algorítmica

Algoritmo 1: Proceso conceptual de detección y retroalimentación.

Entrada: texto académico t
Salida: reporte formativo R
1. Normalizar t y segmentarlo en oraciones y unidades léxicas.
2. Calcular $\phi(t)$: rasgos léxicos, sintácticos, discursivos y estilométricos.
3. Estimar $pIA = f\theta(\phi(t))$ o una puntuación de riesgo equivalente.
4. Identificar rasgos relevantes: repetición, generalidad, baja evidencia, cohesión débil.
5. Interpretar la puntuación según género textual, extensión y propósito académico.
6. Generar retroalimentación descriptiva: “qué se observa”.
7. Generar retroalimentación diagnóstica: “por qué afecta la calidad del texto”.
8. Generar retroalimentación prescriptiva: “qué puede reescribirse y cómo”.
9. Devolver R sin emitir sanción automática ni reemplazar la revisión docente.

Fuente: elaboración propia. El algoritmo describe el flujo metodológico y no implica una implementación ya validada en producción.

Protocolo de evaluación propuesto

La validación futura debe separar la evaluación técnica de la evaluación pedagógica. La evaluación técnica requiere un corpus documentado con textos humanos, textos generados por modelos de lenguaje, textos asistidos por IA y textos editados por humanos. Cada muestra debe registrar idioma, género textual, longitud, fuente, condiciones de generación, criterios éticos y consentimiento cuando corresponda. Sobre ese corpus podrán calcularse exactitud, precisión, exhaustividad y F1; sin embargo, estos indicadores solo deberán reportarse cuando exista una matriz de confusión verificable.

La evaluación pedagógica debe revisar si la retroalimentación es clara, específica, útil y no punitiva. Para ello se propone una revisión por expertos en escritura académica y educación superior, además de una prueba piloto con estudiantes bajo consentimiento informado. La Figura 2 presenta el flujo metodológico previsto para esa validación.



Figura 2: Flujo metodológico de validación futura. Fuente: elaboración propia.

4. Resultados

El resultado principal del trabajo es una arquitectura conceptual evaluable que integra detección lingüística y retroalimentación pedagógica sin presentar la clasificación algorítmica como evidencia definitiva. Esta contribución es de diseño metodológico y no de validación experimental final.

Resultado 1: arquitectura funcional trazable

La arquitectura propuesta permite trazar cada salida del sistema hacia un módulo específico. Esta trazabilidad es relevante porque evita que el reporte final sea una etiqueta

opaca. En lugar de afirmar únicamente “texto generado por IA”, el modelo identifica rasgos observables y los relaciona con acciones de revisión. La Figura 3 muestra el ciclo pedagógico esperado después del análisis técnico.

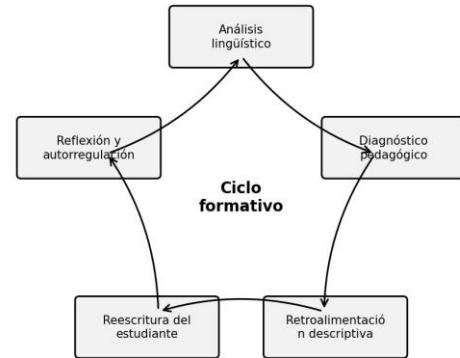


Figura 3: Ciclo de retroalimentación pedagógica. Fuente: elaboración propia.

Resultado 2: criterios de retroalimentación

La propuesta organiza la retroalimentación en tres capas. La capa descriptiva comunica el rasgo observado sin juicio moral; por ejemplo, repetición de conectores de cierre o afirmaciones generales sin evidencia. La capa diagnóstica explica el efecto de ese rasgo en la calidad académica del texto. La capa prescriptiva propone una acción concreta: incorporar evidencia, matizar una afirmación, variar la estructura sintáctica, precisar conceptos o vincular el argumento con fuentes.

Tabla 3: Conversión de rasgos lingüísticos en retroalimentación formativa.

Rasgo observado	Interpretación pedagógica	Retroalimentación sugerida
Repetición de conectores conclusivos	Puede indicar cierre artificial o falta de desarrollo argumentativo.	Revisar la progresión del argumento y conservar solo conectores necesarios.
Afirmaciones generales sin evidencia	Debilita la validez académica del texto.	Agregar ejemplo, dato, cita o justificación vinculada con la disciplina.
Homogeneidad extrema en longitud de oraciones	Puede reducir naturalidad y variedad discursiva.	Alternar estructuras, integrar explicaciones y matices propios.
Vocabulario grandilocuente sin precisión conceptual	Produce apariencia de profundidad sin sustento.	Sustituir términos generales por conceptos definidos y verificables.

Fuente: elaboración propia a partir del alcance metodológico del estudio.

Resultado 3: ejemplo cualitativo de funcionamiento

Para ilustrar la salida esperada se utiliza un fragmento didáctico construido para el artículo. El ejemplo no forma parte de un corpus experimental ni permite calcular métricas. Su función es mostrar cómo una señal lingüística puede transformarse en retroalimentación formativa.

Texto de entrada: “En conclusión, la inteligencia artificial es una herramienta trascendental que revoluciona todos los sectores de la sociedad moderna. Por lo tanto, las nuevas

generaciones deben adoptarla para alcanzar mejores resultados en el futuro”.

Lectura técnica: el fragmento presenta conectores conclusivos acumulados, afirmaciones amplias sin evidencia y léxico valorativo poco preciso. Estos rasgos pueden coincidir con patrones de escritura generada o asistida por IA, pero no constituyen prueba suficiente de autoría automática.

Retroalimentación pedagógica: “El párrafo comunica una idea general, pero necesita evidencia y precisión. Revisa qué sector, contexto o caso quieres explicar. En lugar de afirmar que la IA ‘revolucionará todos los sectores’, incorpora un ejemplo verificable, delimita el alcance de tu afirmación y explica con tus propias palabras por qué ese ejemplo sostiene tu argumento”.

Resultado 4: esquema de evaluación

La propuesta queda lista para una siguiente fase de validación que combine criterios técnicos, pedagógicos y éticos. El diseño evita reportar porcentajes o métricas no comprobadas; en su lugar, define qué evidencias deberán recabarse para sostener una evaluación robusta. La Figura 4 sintetiza los tres planos de evaluación.

Esquema de evaluación propuesto

Validez técnica	Validez pedagógica	Validez ética
• Corpus documentado • Etiquetado verificable • Métricas: precisión, recall, F1	• Revisión de expertos • Claridad de retroalimentación • Utilidad para reescritura	• Transparencia • No sanción automática • Protección de datos

Figura 4: Esquema de evaluación técnica, pedagógica y ética. Fuente: elaboración propia.

5. Discusión

La propuesta coincide con la literatura que cuestiona el uso de detectores automáticos como instrumentos sancionadores. Los estudios sobre sesgo y confiabilidad muestran que la clasificación de autoría generada por IA puede variar según idioma, estilo de escritura, edición posterior y condiciones de evaluación. En consecuencia, el modelo planteado adopta una postura prudente: el análisis técnico se entiende como indicio que debe interpretarse, no como dictamen definitivo.

Frente a enfoques centrados solo en detección, el aporte de este trabajo es integrar una capa pedagógica explícita. Esta capa responde a la necesidad de que el estudiante comprenda qué aspectos del texto debe mejorar. Así, la detección deja de funcionar como una etiqueta cerrada y se convierte en una oportunidad para fortalecer autoría, argumentación y revisión crítica. Esta orientación resulta compatible con modelos de retroalimentación formativa y aprendizaje autorregulado, donde el objetivo es promover acciones de mejora y no únicamente calificar el producto final.

La principal limitación es que el artículo aún no cuenta con una validación empírica completa. Por ello, no se sostienen cifras de precisión, F1, AUC, tasas de falsos positivos ni comparaciones cuantitativas con detectores comerciales. La discusión se limita a la pertinencia del diseño, su coherencia con la literatura y las condiciones necesarias para validar el prototipo. Esta delimitación fortalece la honestidad científica

del manuscrito y reduce el riesgo de sobreafirmar resultados inexistentes.

6. Conclusiones

El artículo presenta un modelo conceptual para detectar señales de textos generados con inteligencia artificial y transformarlas en retroalimentación pedagógica. La contribución central consiste en articular procesamiento del lenguaje natural, rasgos estilométricos, interpretación contextual y comentarios formativos para apoyar la escritura académica en español.

La revisión realizada permite concluir que la detección automática no debe emplearse como prueba única de autoría ni como mecanismo sancionador aislado. Su valor educativo aumenta cuando se comunica con transparencia, reconoce incertidumbre y ofrece al estudiante rutas concretas de mejora. En ese sentido, el modelo propuesto privilegia la formación, la autorregulación y la integridad académica por encima de la vigilancia automática.

Como trabajo futuro, será necesario construir un corpus documentado en español, definir criterios de etiquetado, implementar el prototipo, calcular métricas técnicas verificables y evaluar la utilidad pedagógica de la retroalimentación mediante revisión experta y pruebas piloto con estudiantes. Solo después de esa validación podrán reportarse resultados cuantitativos de desempeño. Esta ruta permitirá avanzar hacia una herramienta responsable, reproducible y útil para contextos universitarios hispanohablantes.

7. Referencias

- Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S., (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610-623. DOI: 10.1145/3442188.3445922
- Boud, D., Molloy, E., (2013). Rethinking models of feedback for learning: The challenge of design. Assessment & Evaluation in Higher Education 38(6), 698-712. DOI: 10.1080/02602938.2012.691462
- 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. DOI: 10.1080/14703297.2023.2190148
- Elkhatat, A. M., Elsaid, K., Almeer, S., (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. International Journal for Educational Integrity 19, 17. DOI: 10.1007/s40979-023-00140-5
- Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., Pearson, A. T., (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. npj Digital Medicine 6, 75. DOI: 10.1038/s41746-023-00819-6
- Hattie, J., Timperley, H., (2007). The power of feedback. Review of Educational Research 77(1), 81-112. DOI: 10.3102/003465430298487
- Hyland, K., Hyland, F., (2006). Feedback on second language students' writing. Language Teaching 39(2), 83-101. DOI: 10.1017/S0261444806003399
- 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., Stadler, M., Weller, J., Kuhn, J., Kasneci, G., (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences 103, 102274. DOI: 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. DOI: 10.1016/j.patter.2023.100779

Miao, F., Holmes, W., (2023). Guidance for generative AI in education and research. UNESCO.

Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., Finn, C., (2023). DetectGPT: Zero-shot machine-generated text detection using probability curvature. *Proceedings of the 40th International Conference on Machine Learning*, 24950-24962.

Nicol, D. J., Macfarlane-Dick, D., (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* 31(2), 199-218. DOI: 10.1080/03075070600572090

Perkins, M., (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice* 20(2). DOI: 10.53761/1.20.02.07

Shute, V. J., (2008). Focus on formative feedback. *Review of Educational Research* 78(1), 153-189. DOI: 10.3102/0034654307313795

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. DOI: 10.1007/s40979-023-00146-z

Zimmerman, B. J., (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice* 41(2), 64-70. DOI: 10.1207/S15430421TIP4102_2