

De la vigilancia al juicio crítico en la evaluación universitaria frente a la IA generativa

*Eduardo Tovar Martínez y José Daniel Sánchez Vásquez**

Resumen

En este ensayo de análisis conceptual, basado en la revisión crítica de literatura reciente y de marcos emergentes de diseño de evaluación, examinamos los desafíos y oportunidades que la inteligencia artificial generativa (GenAI) plantea para la evaluación universitaria. El objetivo es responder a la pregunta de si la evaluación debe reforzar modelos centrados en la vigilancia y la detección, o bien transitar hacia enfoques orientados al juicio crítico, la transparencia y la responsabilidad en la interacción humano-máquina. Se sostiene que las estrategias centradas en vigilancia y detección son insuficientes y contraproducentes, pues debilitan la confianza, la autonomía y la justicia cognitiva. Con base en la evidencia disponible, se propone desplazar el foco de los productos finales hacia los procesos de aprendizaje, valorando la trazabilidad, la procedencia y el juicio crítico. Asimismo, se analizan las brechas tecnológicas, las tensiones entre integridad académica, justicia cognitiva y preparación profesional, así como los riesgos asociados a la dependencia de plataformas corporativas. Finalmente, se argumenta que la universidad pública latinoamericana puede convertir la GenAI en un recurso pedagógico equitativo y humanista, reafirmando su misión de democratizar el conocimiento y formar ciudadanía crítica.

Palabras clave

Inteligencia artificial ¶ Evaluación ¶ Educación

Abstract

This conceptual essay, based on a critical review of recent literature and emerging assessment design frameworks, examines the challenges and opportunities that generative artificial intelligence (GenAI) poses for university assessment. The objective is to address the central question of whether assessment should continue to rely on surveillance and detection-based approaches, or instead shift toward models oriented to critical judgment, transparency, and responsibility in human-machine interaction. It is argued that strategies centered on surveillance and detection are insufficient and

* Profesor de Asignatura. Divisiones de Manufactura Avanzada y Ciencias, Universidad Politécnica de San Luis Potosí, México (eduardo.tovar@upslp.edu.mx). ¶ Profesor-Investigador. Departamento de Físico-Matemáticas, Universidad Autónoma de San Luis Potosí (UASLP), México (daniel.vasquez@uaslp.mx).

counterproductive, as they undermine trust, autonomy, and cognitive justice. Drawing on the available evidence, the paper proposes shifting the focus from final products to learning processes, emphasizing traceability, disclosure of AI use, and critical judgment. It also analyzes technological gaps, the tensions between academic integrity, cognitive justice, and professional preparation, as well as the risks associated with dependence on corporate platforms. Finally, the paper argues that Latin American public universities can turn GenAI into an equitable and humanistic pedagogical resource, reaffirming their mission to democratize knowledge and foster critical citizenship.

Key words

Artificial Intelligence  Evaluation  Education

Introducción

DESDE FINALES de 2022, los sistemas de inteligencia artificial generativa (GenAI) de propósito general —como ChatGPT, GitHub Copilot y Gemini de Google— han pasado de ser una novedad para convertirse en compañeros cotidianos en las universidades. Su capacidad para redactar textos, programar código, analizar datos y producir artefactos multimodales ha acelerado su adopción tanto en estudiantes como entre docentes. Dentro de las ventajas más citadas se encuentran la retroalimentación inmediata, el incremento de la productividad y la posibilidad de personalización del aprendizaje.

Al mismo tiempo, estas fortalezas plantean un desafío fundamental: si un modelo puede generar en segundos un ensayo bien estructurado, un programa funcional o una solución correcta, ¿qué se está evaluando realmente cuando se califican dichos productos? Estudios recientes evidencian que la GenAI puede superar o acercarse a superar evaluaciones tradicionales en ingeniería, redactar ensayos de física con estándares de excelencia e incluso alcanzar resultados comparables a los de estudiantes de medicina en reactivos de respuesta corta. Ello plantea riesgos de validez en contextos de alta exigencia (Morjaria *et al.*, 2023; Nikolic *et al.*, 2023; Yeadon *et al.*, 2023). Estas evidencias muestran que las evaluaciones tradicionales se ven amenazadas por dos factores clásicos: la subrepresentación de constructos, cuando las tareas pueden delegarse a la IA, y la varianza irrelevante, cuando el acceso desigual a estas herramientas introduce ruido en los resultados (Weng *et al.*, 2024; Xia *et al.*, 2024).

La cuestión, sin embargo, trasciende lo técnico y se vuelve filosófica y humanista: ¿qué lugar ocupa la enseñanza cuando una máquina puede realizar gran parte de las tareas que antes se atribuían al estudiantado?, ¿qué significa aprender cuando el error, la duda y la interpretación —elementos centrales de la formación universitaria— parecen externalizarse en sistemas automatizados? Estos interrogantes devuelven la atención a la universidad como espacio de construcción de juicio crítico, más que de mera producción de respuestas.

La brecha de acceso entre versiones gratuitas y de pago, la dependencia de plataformas corporativas y la limitada infraestructura digital promueven formas de pobreza digital —entendida como la restricción estructural de acceso, uso y soberanía tecnológica—, que agravan las tensiones en torno a la equidad y la autonomía universitaria (Lugo y Barrera Rojas, 2024). Ello obliga a pensar en la GenAI no solo como un recurso pedagógico, sino también como un fenómeno que refleja y amplifica desigualdades estructurales, tensionando la misión social de la universidad en la región.

Este ensayo parte de la premisa de que la evaluación universitaria debe replantear su sentido y significado —y, a partir de ello, rediseñarse, más que reforzar esquemas de vigilancia— en un contexto mediado por la inteligencia artificial generativa (GenAI). Su objetivo es analizar los dilemas que esta tecnología introduce en términos de validez, autoría y del papel formativo del error en la educación superior, con especial atención a los campos STEAM (Ciencia, Tecnología, Ingeniería, Artes y Matemáticas, por sus siglas en inglés), entendidos aquí como una aproximación que integra explícitamente las artes y las humanidades en la formación científico-tecnológica, con implicaciones directas para el diseño de tareas, criterios de evaluación y evidencias de aprendizaje. A partir de una revisión crítica de literatura reciente, estudios empíricos y marcos emergentes de evaluación —como AIAS, FACT y la evaluación multimodal-interactiva—, se proponen patrones de diseño para tareas con IA restringida, permitida y requerida, rúbricas centradas en procesos y prácticas de procedencia, así como enfoques programáticos de aseguramiento del aprendizaje alineados con políticas institucionales y estándares profesionales, cuya pertinencia se discute a la luz del contexto de la educación superior en México. La pregunta que articula el análisis es si la irrupción de la GenAI conduce a un vaciamiento de la experiencia universitaria o, por el contrario, abre la posibilidad de reafirmar su misión histórica como espacio de juicio crítico, creatividad y justicia cognitiva.

Desde una perspectiva explícitamente normativa y humanista de la evaluación, este trabajo sitúa la agencia del estudiantado y el papel público de la universidad como ejes del análisis. En coherencia con ello, se privilegian

marcos conceptuales y evidencias que problematizan los enfoques centrados en la vigilancia y el control, y que promueven diseños de evaluación orientados al juicio crítico y a la responsabilidad en contextos mediados por IA. Reconocer este posicionamiento constituye un ejercicio de vigilancia epistemológica sobre los supuestos que orientan el estudio, asumiendo que aproximaciones de carácter más instrumental o centradas en la eficiencia tecnológica pueden priorizar objetivos distintos.

Desafíos de equidad, ética e inclusión en la evaluación mediada por IA

En este trabajo, la noción de justicia cognitiva se entiende como la garantía de que estudiantes con trayectorias, contextos socioculturales, recursos materiales y formas de aprendizaje diversas puedan participar en condiciones equitativas en los procesos de enseñanza, aprendizaje y evaluación, sin que las decisiones de diseño evaluativo introduzcan sesgos estructurales que favorezcan determinados perfiles tecnológicos, lingüísticos o cognitivos. Desde esta perspectiva, la justicia cognitiva no se reduce al acceso a herramientas, sino que implica reconocer y legitimar diferentes formas de producir, interpretar y validar el conocimiento en contextos mediados por tecnologías digitales.

El aumento en el uso de la inteligencia artificial generativa (GenAI) en la educación superior no solo ha revelado desigualdades históricas en el acceso y en las relaciones de poder, sino que también ha creado nuevas formas de exclusión y dependencia, específicamente, bajo las circunstancias educativas en México, donde la pobreza digital se ve marcada por las brechas de conectividad y la infraestructura desigual entre instituciones públicas y privadas. En la medida en que la GenAI redefine quién produce qué y en qué condiciones, las decisiones institucionales arbitrarias sobre restringir, permitir o incluso exigir su uso en los procesos de evaluación no pueden considerarse neutras: dichas decisiones pueden abrir o cerrar oportunidades, garantizar o vulnerar la privacidad, y cultivar —o limitar— una agencia estudiantil responsable con capacidad de actuar con juicio crítico y asumir las implicaciones éticas y académicas del uso de estas tecnologías.

Las revisiones recientes y los análisis de lineamientos institucionales coinciden en que limitar la discusión de la evaluación mediada por IA a la integridad académica es insuficiente (Blake Angulo, 2026; Matindingue *et al.*,

2025). Las respuestas centradas en mecanismos punitivos —como detectores automáticos o esquemas de vigilancia intensiva— ofrecen solo una ilusión de control: pueden mitigar algunas conductas, pero a menudo lo hacen al costo de sacrificar dimensiones esenciales de la educación, como la confianza, la autonomía y la justicia cognitiva, es decir, la posibilidad de que estudiantes con distintos contextos, saberes y recursos participen en condiciones equitativas en los procesos de aprendizaje y evaluación (Luo, 2024; Moorhouse *et al.*, 2023; Weng *et al.*, 2024; Xia *et al.*, 2024). Más que resolver los dilemas, desplazan la atención hacia la sospecha y el control, dejando en segundo plano las preguntas centrales sobre qué significa aprender y cómo reconocerlo en un entorno mediado por IA.

En este sentido, la equidad y la inclusión no son márgenes accesorios de la evaluación intervenida por la IA, sino pilares de su validez y de su legitimidad académica. Este reconocimiento exige tres desplazamientos. Primero, entender que el acceso desigual a herramientas, lenguas y apoyos institucionales se traduce en sesgos educativos y cognitivos que perpetúan y acentúan las brechas sociales ya existentes. Segundo, reconocer que la vigilancia intensificada genera daños colaterales en términos de privacidad, libertad académica y justicia, mientras ofrece una evidencia pobre sobre el aprendizaje real. Y tercero, asumir que las universidades están llamadas a reconciliar la honestidad académica con la justicia cognitiva y la preparación profesional, redefiniendo qué debe contar como evidencia de aprendizaje en un tiempo en el que la colaboración con la IA forma parte de las prácticas genuinas de producción y adquisición del conocimiento.

En última instancia, este debate remite a una pregunta más amplia sobre la misión de la universidad: ¿cómo puede la evaluación evitar convertirse en un mecanismo de control que reproduzca desigualdades y, en cambio, constituirse en un espacio crítico que reconozca la diversidad de formas de aprender y pensar en un mundo atravesado por inteligencias autómatas?

I. Brechas tecnológicas y acceso desigual como amenazas a la validez

El acceso material e infraestructural determina, en buena medida, lo que los estudiantes pueden hacer con la GenAI. Análisis institucionales y testimonios de docentes revelan diferencias notables según el tipo de licencia (gratuita o

premium), las limitaciones de ancho de banda y dispositivos, y las políticas locales; factores que inciden directamente en la forma en que se diseñan y experimentan las evaluaciones (Shishavan, 2024; Khlaif *et al.*, 2025; White & Huber, 2024). Las versiones premium ofrecen contextos ampliados, herramientas avanzadas, prioridad de respuesta y funciones multimodales superiores. Cuando las tareas valoran productos sofisticados, razonamientos extensos o iteración rápida, quienes acceden a esas versiones disfrutan de ventajas sistemáticas frente a sus pares en entornos restringidos. Desde la perspectiva de la medición, estas diferencias introducen varianza irrelevante al constructo: lo que se evalúa deja de ser el conocimiento o la competencia prevista y pasa a ser el acceso a un recurso, socavando así tanto la validez como la equidad (Weng *et al.*, 2024; Xia *et al.*, 2024). En el contexto de las universidades públicas, donde una parte importante del estudiantado depende de conectividad intermitente, dispositivos compartidos o licencias limitadas, esta distorsión implica que las calificaciones reflejen condiciones materiales más que logros académicos, comprometiendo la legitimidad misma de la evaluación como mecanismo de reconocimiento del aprendizaje.

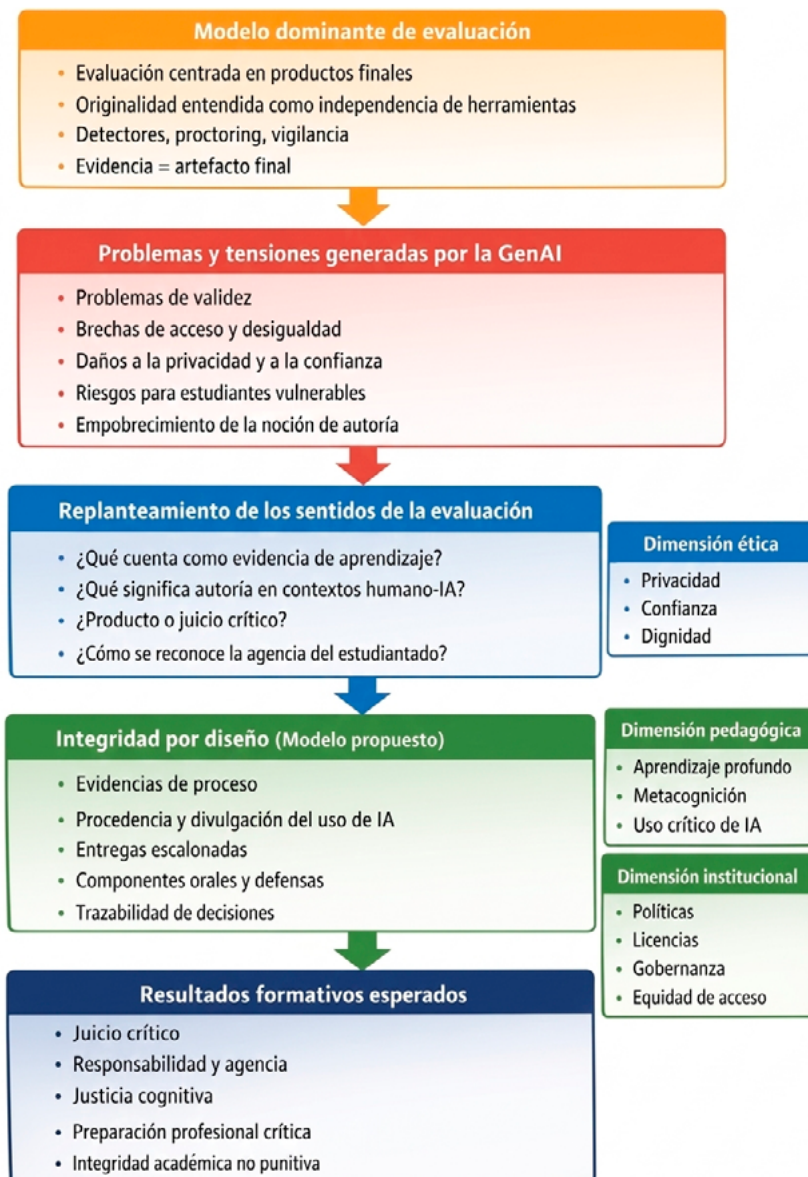
La accesibilidad para el estudiantado en contextos de diversidad funcional, lingüística y socioeconómica entrelaza esta discusión con las políticas institucionales. En particular, las prohibiciones generales sobre la “asistencia con IA” pueden inhibir el uso legítimo de tecnologías de apoyo —como lectores de pantalla o sistemas de dictado por voz—, penalizando precisamente a quienes enfrentan mayores barreras para participar en condiciones de equidad (Shishavan, 2024; Lye & Lim, 2024; White & Huber, 2024). En el extremo opuesto, permitir funciones generativas sin criterios claros de divulgación y verificación puede colocar en desventaja a quienes no pueden —o no deben— usar determinadas herramientas por motivos de discapacidad, carga cognitiva, privacidad o restricciones jurídicas. Las revisiones recientes insisten en que transformar la evaluación implica replantear los resultados de aprendizaje (por ejemplo, la autorregulación o el uso responsable de la IA) y reconocer nuevas formas de evidencia basadas en procesos y verificación (Xia *et al.*, 2024; Zhao *et al.*, 2024). La equidad, en consecuencia, exige mecanismos accesibles y equivalentes para demostrar dichas competencias, acompañados de políticas transparentes que distingan entre tecnologías de asistencia y tecnologías generativas.

Las respuestas de diseño, tanto en los planes de estudio como en el desarrollo de los cursos, se derivan de este diagnóstico. Más que decisiones técnicas, estas medidas expresan una toma de postura ética y axiológica sobre el lugar

de la persona en el proceso educativo, orientada por principios de equidad, inclusión y justicia educativa, y reconociendo al estudiantado como sujeto de derechos y no como mero usuario de tecnologías. En primer lugar, si el uso de IA está permitido o es requerido, las instituciones deben garantizar licencias institucionales y ofrecer alternativas de bajo consumo de datos o acceso *offline* en contextos con conectividad limitada (Shishavan, 2024; Khlaif *et al.*, 2025). En segundo lugar, el permiso debe ir de la mano de transparencia: estudiantes y docentes deberían declarar herramientas, versiones, indicaciones y pasos de verificación, para que la evaluación valore la calidad de la supervisión y no solo el acabado final (Feng *et al.*, 2025; Perkins *et al.*, 2024; Wilson & Nishimoto, 2024). En tercer lugar, los diseños deben asegurar rutas accesibles: en tareas con IA restringida para validar la competencia conceptual, las tecnologías de apoyo deben mantenerse disponibles; y en tareas con IA permitida o requerida, debe ofrecerse formación diferenciada para estudiantes con trayectorias previas diversas (Elshall & Badir, 2025; Lye & Lim, 2024). En cuarto lugar, es indispensable diversificar los portafolios de evaluación para que ninguna condición de acceso determine por sí sola la inferencia sobre el aprendizaje. Modelos como FACT ejemplifican cómo es posible equilibrar de manera operativa el rigor académico, la autenticidad de las tareas y la accesibilidad en la enseñanza de STEAM, al proponer criterios de diseño curricular, actividades y evaluación centrados en problemas reales, en la participación del estudiantado y en la reducción de barreras tecnológicas y contextuales. En este sentido, FACT no solo funciona como una guía pedagógica, sino como un referente para el cambio institucional hacia prácticas de enseñanza más inclusivas y formativamente coherentes (Elshall & Badir, 2025).

En el fondo, estas brechas de acceso no son meramente técnicas: expresan tensiones estructurales entre la universidad pública y las plataformas corporativas, así como entre los derechos educativos y las economías de suscripción. Afrontarlas no implica únicamente ampliar el acceso a licencias, sino afirmar un principio de justicia educativa según el cual la validez de la evaluación no debe comprarse en el mercado, sino garantizarse como un bien común. Estas desigualdades no solo inciden en los resultados académicos, sino que interpelan directamente el lugar de la universidad frente a las plataformas privadas y los modelos de negocio que organizan el acceso al conocimiento. Este desplazamiento conceptual se sintetiza en la Figura 1, que muestra el tránsito desde un modelo de evaluación centrado en la vigilancia hacia un enfoque de integridad por diseño orientado al juicio crítico.

Figura 1. Esquema del tránsito conceptual desde un modelo de evaluación universitaria centrado en la vigilancia hacia un enfoque de integridad por diseño



Fuente: Elaboración propia.

II. Vigilancia, detección y privacidad: ¿integridad a qué costo?

La irrupción de la GenAI ha trastocado la noción de autoría y, ante esta incertidumbre, la detección se ha convertido en una respuesta tentadora. Sin embargo, la evidencia disponible es clara: los detectores de productos generados por IA son poco fiables, fáciles de evadir y proclives a falsos positivos y negativos, lo que los convierte en una base endeble para tomar decisiones de alto impacto académico (Shishavan, 2024; Luo, 2024; Moorhouse *et al.*, 2023). Lejos de fortalecer la integridad, la proliferación de navegadores bloqueados, analíticas de tecleo o registros ampliados degrada la experiencia estudiantil, introduce barreras de acceso (limitaciones de hardware, viviendas compartidas, precariedad digital) y ofrece un valor probatorio marginal sobre lo que realmente importa: qué sabe y qué puede hacer el estudiante (White & Huber, 2024). La vigilancia intensificada, además, inhibe la divulgación honesta del uso de herramientas, socavando la transparencia que permitiría evaluar el juicio de los estudiantes en su interacción con la IA.

Las limitaciones de este enfoque son tanto epistémicas como éticas. En lo epistémico, ningún detector puede demostrar la autoría humana en contextos generales: se confunden con paráfrasis, traducciones o borradores híbridos, y resultan incapaces de sustentar juicios individualizados o argumentos sólidos de validez programática (Shishavan, 2024; Luo, 2024). En lo ético, las tecnologías de vigilancia erosionan la privacidad, amplifican desigualdades —particularmente para estudiantes sin entornos adecuados para la supervisión remota— y aumentan la probabilidad de acusaciones erróneas. El costo de estos mecanismos no se distribuye equitativamente, y el daño reputacional o académico suele recaer con mayor peso sobre los más vulnerables. Al insistir en la detección como salvaguarda de la “originalidad”, muchas instituciones perpetúan una visión estrecha de la autoría que desconoce la naturaleza distribuida y mediada de la producción de conocimiento en la era digital (Luo, 2024). Esta concepción dialoga, además, con el marco jurídico mexicano, donde la Ley Federal del Derecho de Autor reconoce la creación únicamente como obra atribuible a personas físicas, reforzando una idea individualista de autoría que resulta cada vez menos adecuada para describir prácticas académicas híbridas, colaborativas y asistidas por IA (Ley Federal del Derecho de Autor [LFDA], 1996).

Frente a este panorama, la alternativa no es más control, sino otro concepto de integridad. Las prácticas más prometedoras desplazan la vigilancia hacia un diseño rico en evidencias: declaraciones de procedencia, registros de *prompts* y

respuestas, historiales de repositorios, pruebas unitarias diseñadas por el propio alumnado y defensas orales dirigidas a resultados clave (Feng *et al.*, 2025; Furze *et al.*, 2024; Perkins *et al.*, 2024; Petrovska *et al.*, 2024; Wilson & Nishimoto, 2024). Estos mecanismos hacen más difícil la mala práctica sin criminalizar de entrada al estudiantado y, sobre todo, generan evidencia de proceso —planificación, *prompting*, depuración, verificación— que sí respalda inferencias válidas sobre el aprendizaje.

En el plano institucional, los marcos de gobernanza colaborativa —entendida como la distribución explícita y coordinada de responsabilidades entre profesorado, estudiantado y autoridades para diseñar, implementar y evaluar las políticas de evaluación— reparten funciones entre profesorado, estudiantado y autoridades: mientras el profesorado diseña tareas que capturan procesos, el estudiantado documenta la transparencia y las autoridades establecen políticas que garanticen la transparencia y la rendición de cuentas sin caer en una carrera armamentista de detección (Ilieva *et al.*, 2025). En última instancia, las guías sectoriales coinciden en que rediseñar es más eficaz que vigilar: la integridad no puede sostenerse a costa de la privacidad ni de la confianza (Shishavan, 2024; Doorsamy *et al.*, 2025).

La cuestión de fondo, por tanto, no es si debemos perfeccionar los mecanismos de detección, sino qué significa garantizar la integridad en un entorno en el que la autoría y la producción de conocimiento han dejado de ser prácticas estrictamente individuales para convertirse en procesos crecientemente distribuidos y mediados por tecnologías digitales. En este contexto, la integridad no se sostiene mediante la sospecha y el control, sino a través de diseños evaluativos que hagan visible el juicio humano, la trazabilidad de los procesos y la responsabilidad compartida en la construcción del conocimiento.

III. Integridad, justicia cognitiva y preparación para el trabajo: una tensión triple

La evaluación mediada por IA se ubica en la encrucijada de tres exigencias inseparables: salvaguardar la integridad académica como marco institucional de valores, normas y prácticas que orientan el uso responsable de la tecnología, y promover la honestidad académica como compromiso ético individual con la autoría, la transparencia y la responsabilidad; garantizar la justicia cognitiva; y preparar a los egresados para un mundo donde la colaboración con sistemas automatizados ya es parte constitutiva de la práctica profesional. Esta triple tensión no es meramente

técnica, sino profundamente filosófica, pues redefine qué significa aprender, producir conocimiento y ejercer juicio en la universidad contemporánea.

En términos de integridad académica —no concebida como un control de plagio, sino como un marco ético que oriente y valide el conocimiento—, la cuestión ya no es si la GenAI puede resolver pruebas tradicionales —eso ha quedado suficientemente documentado—, sino qué significa para la validez seguir equiparando “respuesta correcta” con “aprendizaje”. El dilema central no está en la capacidad técnica de la IA, sino en la fragilidad conceptual de nuestras inferencias evaluativas cuando reducimos la evidencia a un artefacto terminado.

La justicia cognitiva cuestiona, en cambio, la noción reduccionista de “originalidad” entendida como independencia absoluta de herramientas. Políticas universitarias continúan concibiendo el “trabajo propio” como aquello no influenciado por IA (Luo, 2024), ignorando que la autoría en la producción de conocimiento ha sido siempre relacional y distribuida. Un enfoque más justo reconoce el uso razonado de la IA como parte de la agencia estudiantil: lo central no es si se usó o no la herramienta, sino, cómo se usó, con qué justificación y bajo qué criterios de verificación. La exigencia filosófica aquí es recuperar la noción de virtud epistémica: saber no consiste únicamente en producir respuestas, sino en tener buenas razones, detectar errores, considerar alternativas y calibrar la confianza. Evaluar la autoría mediada por IA exige, por tanto, valorar la trazabilidad y el juicio, no el acabado superficial de un artefacto (Feng *et al.*, 2025; Perkins *et al.*, 2024).

El tercer vector es la preparación para el trabajo, donde se juega la relevancia social de la universidad. La cuestión no es si los egresados utilizarán IA, sino si lo harán con criterio, ética y responsabilidad. Líderes educativos subrayan que el futuro profesional exige dominar habilidades como diseñar *prompts* estratégicos, ejecutar pruebas sistemáticas, visibilizar sesgos en datos y modelos, interpretar salidas ambiguas, detectar alucinaciones y diagnosticar modos de falla (Khlaif *et al.*, 2025; Petrovska *et al.*, 2024). Sin embargo, estudios muestran que la dependencia acrítica a la IA se correlaciona con un bajo desempeño en tareas de transferencia y razonamiento independiente (Bouamor *et al.*, 2025; Huang *et al.*, 2024). Esto sugiere que prohibir la IA en nombre de la integridad puede producir egresados desactualizados, mientras que permitirla sin estructura genera competencias frágiles. El verdadero reto, por tanto, no solo es convivir con la IA sin diluir su agencia intelectual, sino replantear las competencias profesionales mismas: formar profesionistas capaces de usarla de manera crítica, responsable y situada, comprendiendo sus limitantes, riesgos y efectos sociales.

La evidencia emergente apunta hacia diseños pedagógicos de alta modalidad e interactividad: tareas que integran simulación, razonamiento verbal, manipulación

de datos en vivo o defensa argumentativa de resultados. Estas experiencias no solo dificultan la delegación a la máquina, sino que visibilizan los procesos de respuesta que constituyen el núcleo del juicio experto. Son, en última instancia, dispositivos que restituyen a la evaluación su carácter humanista: no solo porque hacen visibles los procesos cognitivos, sino porque invitan al estudiante a posicionarse como responsable frente a la GenAI, a pensar con y contra la herramienta, en lugar de sustituir su criterio por automatismos (Gao *et al.*, 2025).

IV. Contrapesos y consecuencias no previstas

La evaluación centrada en la equidad en la era de la GenAI no es un terreno de consensos fáciles, sino un campo de contrapesos inevitables. Cada decisión implica costos y consecuencias no siempre visibles, que, de no examinarse con cuidado, reproducen inequidades bajo el disfraz de soluciones técnicas.

Acceso versus autenticidad: Incluir herramientas específicas de IA en la evaluación puede aumentar la autenticidad y la alineación con la práctica laboral contemporánea. Sin embargo, al depender de licencias premium o funciones avanzadas, se corre el riesgo de institucionalizar un esquema de pago por desempeño, difícilmente sostenible para universidades públicas, donde quienes pueden costearlo obtienen ventajas sistemáticas (White & Huber, 2024). Prohibir la IA, en sentido contrario, puede homogeneizar el acceso, pero a costa de excluir a los estudiantes de un entrenamiento formativo que refleja el mundo profesional (Khlaif *et al.*, 2025). La literatura sugiere un compromiso más responsable: acceso financiado institucionalmente, evaluaciones diseñadas para herramientas de capacidades conocidas y alternativas viables para quienes no puedan —o no deban— usarlas (Shishavan, 2024). De este modo, la tecnología opera como un medio intercambiable y no como una condición de posibilidad para el aprendizaje.

Vigilancia versus garantía: El despliegue de detectores, navegadores bloqueados y sistemas de registro invasivo genera una ilusión de control, pero aporta evidencia débil para la toma de decisiones académicas y conlleva costos éticos significativos. Luo (2024) muestra que numerosas universidades continúan equiparando la “originalidad” con la independencia total de la IA, lo que refuerza un enfoque punitivo que erosiona la confianza, empobrece la noción de autoría y debilita la agencia del estudiantado al someter sus producciones a sistemas presentados como objetivos, pese a estar diseñados por personas y a reproducir sus mismos sesgos. De manera consistente, White y Huber (2024) evidencian que las plataformas de supervisión intensiva deterioran la experiencia estudiantil y amplifican

barreras de acceso, mientras que Shishavan (2024) advierte que, aun cuando los marcos institucionales reconocen riesgos específicos para estudiantes vulnerables o con discapacidad, no ofrecen garantías equivalentes de validez evaluativa. Más allá de un dilema técnico, la cuestión es formativa y filosófica: ¿qué precio estamos dispuestos a pagar en términos de privacidad, confianza y agencia del estudiantado por una integridad que la propia tecnología no puede asegurar?, y ¿cómo transitar desde un modelo formativo basado en la vigilancia hacia una garantía de integridad centrada en la persona, el aprendizaje y el diseño pedagógico?

Ganancias inmediatas versus aprendizaje duradero: El uso no estructurado de IA eleva el desempeño en tareas rutinarias, sobre todo en estudiantes con menor preparación, pero compromete la transferencia cuando no se exige verificación y reflexión (Bouamor *et al.*, 2025; Huang *et al.*, 2024). Las prohibiciones, en contraste, pueden preservar parte de esa transferencia, pero con el costo de agrandar las brechas a corto plazo. Diseños escalonados —predecir → consultar la IA → criticar → revisar—, evaluados por su metacognición más que por la corrección final, representan una vía para transformar la dependencia en aprendizaje profundo (Elshall & Badir, 2025; Feng *et al.*, 2025).

Blindaje de ítems versus aseguramiento programático: La obsesión con diseñar reactivos “a prueba de IA” conduce con frecuencia a consignas artificiosas o excesivamente onerosas, que terminan perjudicando al estudiantado más que garantizando validez. La evidencia sugiere que la garantía de aprendizaje debe trasladarse al nivel programático: diversificar evidencias a lo largo del plan de estudios, exigir defensas orales en proyectos terminales o consolidar portafolios acumulativos (Jongkind *et al.*, 2025; Lloyd *et al.*, 2024; Verdicchio, 2025). Dicho de otro modo, no se trata de blindar reactivos, sino de asegurar trayectorias formativas.

Simplicidad normativa versus justicia cognitiva: Los marcos binarios —permitir o prohibir— son administrativamente atractivos, pero conceptualmente empobrecen la discusión. Reproducen una idea estrecha de “originalidad” como producción aislada y desatienden la naturaleza distribuida de la autoría en la era digital (Luo, 2024; Moorhouse *et al.*, 2023). Una política cognitivamente justa exige reconocer la contribución legítima mediada por herramientas, exigir trazabilidad del proceso y distinguir entre tecnologías de apoyo y tecnologías generativas (Shishavan, 2024). Es un camino más exigente, pero también más coherente con la realidad del trabajo académico y con una concepción de la autoría que implica responsabilidad personal. En última instancia, la cuestión central no es si el estudiantado utiliza o no sistemas de IA generativa, sino si la universidad está formando personas capaces de asumir responsabilidad por sus decisiones, sus producciones y sus actos, con y sin mediación tecnológica.

Cada contrapeso revela que las decisiones de diseño no son neutras, sino que redistribuyen oportunidades y riesgos. Exigir funciones premium en nombre de la “autenticidad” puede profundizar la exclusión; intensificar la vigilancia para “restaurar la confianza” puede inhibir la transparencia y socavar los mismos procesos que se busca evaluar. Por el contrario, ignorar la GenAI por temor a conductas indebidas expone el riesgo de graduar estudiantes desprovistos de las competencias éticas y técnicas que sus profesiones demandan. Lo que está en juego, más allá de la técnica, es el sentido mismo de la universidad como espacio de justicia, confianza y formación de juicio autocrítico.

Repensar la evidencia de aprendizaje en la era de la IA generativa

La inteligencia artificial generativa ha obligado a la educación superior a enfrentar una pregunta simple pero crucial: si un sistema puede producir, bajo demanda, un ensayo pulido, un código ejecutable o una respuesta correcta, ¿qué cuenta, con precisión, como evidencia de aprendizaje? La investigación reciente ha mostrado la fragilidad de muchos artefactos antes considerados indicadores válidos de comprensión; lo relevante aquí no es reiterar esa demostración, sino asumir su consecuencia: si la evaluación se limita al producto final, corre el riesgo de confundir fluidez automatizada con comprensión humana (Morjaria *et al.*, 2023; Nikolic *et al.*, 2023; Sharma *et al.*, 2025).

Las revisiones convergen en un giro de enfoque que no se limita al desplazamiento desde los productos hacia los procesos, sino que implica también la redefinición de los propios productos académicos y de los criterios con los que estos son evaluados, en coherencia con prácticas de aprendizaje mediadas por IA. La evidencia debe desplazarse hacia las huellas observables mediante las cuales los estudiantes planifican, verifican e interpretan su trabajo (Feng *et al.*, 2025; Gao *et al.*, 2025; Weng *et al.*, 2024; Xia *et al.*, 2024). Este replanteamiento se alinea con la noción de virtud epistémica: saber no es solo afirmar, sino justificar, reconocer objeciones y calibrar la confianza. En clave práctica, “tener buenas razones” significa diseñar planes de verificación, ejecutar pruebas y usar fuentes con rigor. Un estudiante que identifica y corrige errores de la IA ofrece una evidencia más sólida que quien entrega una respuesta pulida pero acrítica.

Esto no implica abandonar los productos, si no rediseñarlos y reinterpretarlos a la luz de los procesos que los sostienen: informes acompañados de bitácoras,

modelos con trazabilidad de decisiones, resultados defendidos oralmente o contextualizados críticamente. Así, la evaluación y la evidencia dejan de centrarse en el artefacto terminado como prueba de autosuficiencia y se redefine como ejercicio rastreable y explicable de agencia y juicio, donde el producto es parte necesaria, pero nunca suficiente.

En el terreno práctico, ya se observan rediseños que vuelven legible el proceso. En ciencias ambientales, el portafolio FACT equilibra fundamentos con IA restringida, proyectos aplicados con IA asistida, exámenes conceptuales sin IA y casos críticos habilitados por IA (Elshall & Badir, 2025). En cursos de computación, se incorporan registros de *prompts*, historiales de versiones, suites de prueba y defensas orales que desplazan el énfasis hacia planificación, depuración y justificación; el resultado es aprendizaje más profundo y menor dependencia acrítica de la GenAI (Feng *et al.*, 2025; Petrovska *et al.*, 2024; Wilson & Nishimoto, 2024). En coherencia, los argumentos de validez deben reconstruirse sobre huellas verificables de trabajo, reequilibrando los constructos hacia interpretación, verificación y uso responsable de herramientas (Weng *et al.*, 2024; Xia *et al.*, 2024; Zhao *et al.*, 2024).

Aunque las respuestas institucionales han tendido primero a controles (prohibiciones, detectores, vigilancia), sus límites técnicos y éticos son conocidos y se desarrollan en la sección dedicada a ese tema. Aquí basta con extraer la lección operativa: la vía prometedoras no es vigilar más, sino redefinir los sentidos de la evaluación—qué cuenta como evidencia de aprendizaje, cómo se reconoce la autoría, y qué tipo de juicio formativo se espera en entornos donde la IA puede participar en la producción textual. En esa clave, la integridad deja de perseguirse como detección *ex post* y pasa a construirse por diseño: procedencia obligatoria, entregas escalonadas, componentes orales y contextos novedosos que hagan visible el proceso y reduzcan la rentabilidad de la delegación mecánica (Elshall & Badir, 2025; Gao *et al.*, 2025; Wilson & Nishimoto, 2024).

Las implicaciones varían por disciplina, por ejemplo, en STEAM, donde corrección y reproducibilidad son centrales, el riesgo de externalización coexiste con oportunidades de evaluar prácticas auténticas: pruebas unitarias, análisis dimensional, validación de modelos (Nikolic *et al.*, 2023). Por otro lado, en humanidades, la GenAI puede imitar la fluidez retórica, pero no sustituye la lectura atenta, la contextualización crítica ni la interpretación argumentada; aquí, anotaciones, borradores iterativos y defensas orales aportan la evidencia de proceso decisiva (Luo, 2024; Moorhouse *et al.*, 2023). Pese a las diferencias, la lección converge: desplazar el foco del artefacto al proceso devuelve a la evaluación su carácter humanista, centrado en discernir cómo se aprende, no solo qué se entrega.

Finalmente, repensar la evidencia comporta un imperativo ético. Las decisiones de diseño distribuyen oportunidades y riesgos: la equidad exige atender brechas de acceso, límites de la detección y tensiones entre integridad y calidad de la retroalimentación (Khlaif *et al.*, 2025). De ahí se siguen compromisos mínimos: transparencia en el uso de herramientas (Perkins *et al.*, 2024); portafolios mixtos que combinen tareas con IA restringida y habilitada (Elshall & Badir, 2025); garantías programáticas que sustituyan la vigilancia generalizada (Lloyd *et al.*, 2024); y justicia cognitiva, cultivando capacidades metacognitivas y éticas para gobernar el propio uso de la IA (Xia *et al.*, 2024). Recentrada así, la evidencia de aprendizaje deja de ser un artefacto aislado para convertirse en un registro de discernimiento y juicio en contextos donde producir es cada vez más fácil, pero comprender sigue siendo lo esencial.

La universidad pública latinoamericana frente a la IA generativa

La inteligencia artificial generativa irrumpe en la educación superior latinoamericana en un terreno marcado por la desigualdad. Se encuentra con déficits crónicos de infraestructura, un financiamiento público desigual y un mosaico lingüístico y cultural que históricamente ha quedado en la periferia del desarrollo de los sistemas dominantes de IA. En este contexto, la evaluación se convierte a la vez en un espacio de riesgo —donde las brechas existentes pueden reproducirse a gran escala— y en un espacio de posibilidad, donde las universidades pueden re-imaginar la evaluación como una práctica pública y democrática. La literatura empírica en y sobre la región aún es incipiente, pero ya apunta a presiones sobre el diseño evaluativo, riesgos claros de equidad y la necesidad de repensar la originalidad y la autoría en formas compatibles con la misión pública de la universidad, así como de revisar los marcos éticos y axiológicos que orientan la formación profesional y la responsabilidad social del egresado. Al mismo tiempo, deja abiertas cuestiones sustantivas de gobernanza, en particular para el sector público mexicano, que demandan un juicio institucional cuidadoso y políticas con arraigo local (Aquino Zúñiga *et al.*, 2020; Mateus *et al.*, 2024; Murcia *et al.*, 2025; Salinas-Navarro *et al.*, 2024; Valdivieso & González, 2025).

La brecha digital no es una abstracción. Evidencias de El Salvador muestran que estudiantes de contextos de bajos ingresos acceden de forma desproporcionada a la GenAI mediante teléfonos inteligentes y versiones gratuitas, mientras que sus

pares con mayores recursos emplean laptops y funciones premium, con diferencias correlativas en confianza y en preocupaciones éticas (Valdivieso & González, 2025). Es razonable inferir que una estratificación semejante se presentará en universidades públicas mexicanas, dado el perfil de vulnerabilidad documentado en ese sector —discapacidad, desventaja socioeconómica, etnicidad y género— que antecede a la GenAI pero inevitablemente interactuará con ella (Aquino Zúñiga *et al.*, 2020). Estudios de actores en contextos latinoamericanos afines (p. ej., programas de lenguas en Colombia) también señalan preocupaciones sobre autenticidad y confianza en la evaluación, junto con un llamado a políticas democráticas e inclusivas para gobernar su uso (Murcia *et al.*, 2025). Investigaciones cualitativas iniciales en México sugieren que el profesorado ya está rediseñando tareas en respuesta a la GenAI, tratándola alternativamente como aliada o como rival y ajustando *prompts* y criterios de evaluación en consecuencia (Mateus *et al.*, 2024). En conjunto, estos estudios sugieren que cualquier estrategia de evaluación que presuponga banda ancha estable, laptops y acceso a servicios de pago favorecerá a quienes ya están en ventaja y penalizará al resto.

Las limitaciones de infraestructura complican la equidad incluso allí donde existe acceso. Evaluaciones sincrónicas de alta vigilancia (como sistemas de supervisión que demandan gran ancho de banda) exacerban déficits de conectividad rural y periurbana, además de afectar la accesibilidad de estudiantes con discapacidad —presiones ya documentadas en la literatura sobre equidad pre-GenAI en universidades públicas mexicanas (Aquino Zúñiga *et al.*, 2020)—. A ello se suman las barreras lingüísticas y culturales: las variedades del español, las lenguas indígenas y las prácticas de alternancia de códigos pueden producir diferencias de desempeño en modelos entrenados con registros dominantes. Aunque la evidencia empírica aún no cuantifica este sesgo, reportes a nivel de aula y ensayos de diseño destacan tanto oportunidades como riesgos: la IA puede facilitar retroalimentación y evaluación basada en rúbricas en español, pero sin una validación explícita por variedad lingüística, también puede imponer un estándar estrecho y penalizar silenciosamente a quienes quedan fuera de él (Niño Muñoz, 2025; Salinas-Navarro *et al.*, 2024). En suma, las desigualdades estructurales no son mero telón de fondo: son determinantes formativos de si la evaluación mediada por IA fortalece o debilita la justicia educativa.

Una pregunta que se debe formular es: ¿Quién se beneficia de la GenAI en la evaluación y quién queda excluido? La evidencia inicial apunta a una respuesta conocida: los beneficios recaen primero en quienes cuentan con dispositivos, conectividad y habilidades para gestionar *prompts* y criticar salidas, mientras que las exclusiones se multiplican para quienes dependen de *smartphones* y versiones

gratuitas o carecen de orientación institucional (Huallpa *et al.*, 2023; Murcia *et al.*, 2025; Valdivieso y González, 2025). La formación docente puede modificar este panorama: una amplia iniciativa regional de capacitación mejoró la competencia técnica y la disposición del profesorado para utilizar la GenAI en la enseñanza, pero no dispó las preocupaciones sobre autenticidad y originalidad (López-Vasco *et al.*, 2025). En este sentido, la equidad no depende solo del acceso o de la capacitación instrumental, sino también de generar una literacidad en GenAI —entendida como la capacidad de comprender críticamente el funcionamiento, los sesgos, los límites y las implicaciones éticas de estos sistemas—. Cuando el profesorado no tiene claridad sobre qué constituye un uso legítimo de IA, la aplicación discrecional llena ese vacío, lo que suele afectar con mayor dureza a los estudiantes ya marginados.

La respuesta de política que se ha mencionado en párrafos anteriores, sugerida tanto por análisis pedagógicos como normativos, es desplazar la evaluación del énfasis exclusivo en el producto hacia evidencias de proceso, defensas orales y divulgación explícita del uso de IA, evaluando el juicio y la justificación en lugar de la mera calidad del resultado. Este replanteamiento se alinea con las críticas a las políticas de originalidad que tratan cualquier intervención de IA como contaminación; en cambio, reconoce que el conocimiento se produce cada vez más en ensamblajes distribuidos humano-máquina y exige transparencia y compromiso crítico más que pureza. Implementados de manera equitativa, estos diseños pueden mitigar riesgos de sustitución sin recurrir a regímenes de alta vigilancia que penalizan de forma desproporcionada.

Apuntes finales

La GenAI en la evaluación no es solo un problema técnico o legal; es una cuestión de quién puede conocer y ser reconocido como sujeto de conocimiento. En continuidad con la noción de justicia cognitiva desarrollada en las secciones anteriores, y desde una perspectiva de justicia epistémica, los sistemas de IA entrenados en corpus que subrepresentan la producción académica latinoamericana, las epistemologías indígenas o las variedades del español corren el riesgo de reproducir injusticias testimoniales —al devaluar la credibilidad de ciertas voces— o injusticias hermenéuticas —al invisibilizar experiencias locales que los modelos no logran interpretar—. La pedagogía crítica exige que la evaluación cultive la concientización más que mera conformidad; un diseño que requiera a los estudiantes divulgar, criticar y situar su uso de la IA puede convertir al modelo en objeto de indagación,

en lugar de en un oráculo invisible. Las críticas a la “originalidad” en documentos actuales de educación superior refuerzan este giro al cuestionar una concepción individualista y estrecha de la autoría, e invitar a criterios más inclusivos y orientados al proceso. En términos prácticos, esto significa valorar la calidad en la formulación de problemas, la selección de evidencias y el razonamiento ético sobre el uso de herramientas —capacidades coherentes con una visión humanista de la educación superior— aun cuando parte del trabajo textual sea asistido por IA.

Históricamente, la universidad pública latinoamericana ha tenido el mandato de democratizar el conocimiento, formar ciudadanía crítica y atender a la justicia social en contextos de desigualdad. La GenAI incide en esta misión en dos direcciones. Por un lado, puede normalizar relaciones extractivas de datos y atar a las universidades a dependencias epistemológicas y técnicas que estrechen el currículo a lo que los sistemas propietarios consideran seguro producir. Por otro, puede ampliar el acceso a retroalimentación de calidad y apoyos multilingües, reducir barreras lingüísticas y de discapacidad, y favorecer evaluaciones auténticas a gran escala —siempre que se diseñen con criterios de equidad y soberanía—.

El futuro no está determinado por la tecnología, sino por las decisiones de diseño y gobernanza. Si las universidades públicas optan por evaluaciones centradas solo en productos, regímenes punitivos de originalidad y dependencia opaca de plataformas, amplificarán la exclusión y debilitarán su autonomía. Si, en cambio, implementan provisión equitativa, marcos de divulgación y crítica, y acuerdos de gobernanza soberana, la GenAI puede ponerse al servicio de la misión pública: democratizar el conocimiento, cultivar ciudadanía crítica y enfrentar la injusticia social.

La pregunta de fondo es filosófica antes que técnica: ¿será la evaluación un ritual de exclusión gobernado por infraestructuras privadas, o una práctica pública de indagación colectiva donde la IA se convierte en herramienta a ser usada, examinada y, cuando sea necesario, cuestionada?

Referencias

- Aquino, S. P., García, V., y Palmeros, G. (2020). Atención a grupos vulnerables e indicadores de equidad en educación superior: Caso de una universidad pública en el sureste mexicano. *Dilemas contemporáneos: Educación, Política y Valores*. <https://doi.org/10.46377/dilemas.v32i1.1995>
- Blake Angulo, J. (2026). Docencia e inteligencia artificial generativa en educación superior en América Latina: Focos y vacíos de la investigación emergente (revisión sistemática 2023-2025). *Revista Realidad Educativa*, 6(1), 153–189. <https://doi.org/10.38123/rre.v6i1.600>
- Bouamor, H., Gongora-Svartzman, G., Heimann, L., & Huang, S. (2025). Evaluating GenAI's Effectiveness for Students with Varied Programming Backgrounds in a Software Development Course. *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 2*, 1395–1396. <https://doi.org/10.1145/3641555.3705175>
- Doorsamy, W., Padayachee, K., & Cornell, A. S. (2025). Generative Artificial Intelligence and Encounters with Knowledge in STEM Higher Education Curricula. *2025 IEEE Global Engineering Education Conference (EDUCON)*, 1–5. <https://doi.org/10.1109/EDUCON62633.2025.11016322>
- Elshall, A. S., & Badir, A. (2025). Balancing AI-assisted learning and traditional assessment: The FACT assessment in environmental data science education. *Frontiers in Education*, 10, 1596462. <https://doi.org/10.3389/educ.2025.1596462>
- Feng, T. H., Luxton-Reilly, A., Wünsche, B. C., & Denny, P. (2025). From Automation to Cognition: Redefining the Roles of Educators and Generative AI in Computing Education. *Proceedings of the 27th Australasian Computing Education Conference*, 164–171. <https://doi.org/10.1145/3716640.3716658>
- 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). <https://doi.org/10.14742/ajet.9434>
- Gao, Y., Zhai, X., Li, M., Lee, G., & Liu, X. (2025). A Multimodal Interactive Framework for Science Assessment in the Era of Generative Artificial Intelligence. *Journal of Research in Science Teaching*, tea.70009. <https://doi.org/10.1002/tea.70009>

- Huallpa, J. J., Flores Arocutipa, J. P., Panduro, W. D., Huete, L. C., Flores Limo, F. A., Herrera, E. E., Alba Callacna, R. A., Ariza Flores, V. A., Medina Romero, M. Á., Quispe, I. M., y Hernández Hernández, F. A. (2023). Exploring the ethical considerations of using Chat GPT in university education. *Periodicals of Engineering and Natural Sciences*, 11(4), 105-115. <https://doi.org/10.21533/pen.v11i4.3770>
- Huang, D., Huang, Y., & Cummings, J. J. (2024). Exploring the integration and utilisation of generative AI in formative e-assessments: A case study in higher education. *Australasian Journal of Educational Technology*, 40(4) <https://doi.org/10.14742/ajet.9467>
- Ilieva, G., Yankova, T., Ruseva, M., & Kabaivanov, S. (2025). A Framework for Generative IA-Driven Assessment in Higher Education. *Information*, 16(6), 472. <https://doi.org/10.3390/info16060472>
- Jongkind, R., Elings, E., Joukes, E., Broens, T., Leopold, H., Wiesman, F., & Meinema, J. (2025). Is your curriculum GenAI-proof? A method for GenAI impact assessment and a case study. *MedEdPublish*, 15, 11. <https://doi.org/10.12688/mep.20815.1>
- Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning Assessments for IA-Enhanced Learning: A Framework for Educators in the Generative IA Era. *Education Sciences*, 15(2), 174. <https://doi.org/10.3390/educsci15020174>
- Ley Federal del Derecho de Autor. Diario Oficial de la Federación, 12 de diciembre de 1996 (México). <https://www.wipo.int/wipolex/es/legislation/details/16108>
- Lloyd, C., Doherty, S., Herb, A., & Warburton, S. (2024). A risk-based approach to mitigating the Gen(AI) challenge to assessment integrity: The Programmatic Assessment Security Project (PASP). *Proceedings ASCILITE*, 53–54. <https://doi.org/10.14742/apubs.2024.1387>
- López, F. E., Angulo, M. R., y Sosa, D. I. (2025). Formación Docente en IA Generativa: Impacto Ético y Retos en Educación Superior. *Alteridad*, 20(2), 166–177. <https://doi.org/10.17163/alt.v20n2.2025.01>
- Lugo, L. J., y Barrera, M. Á. (2024). Actualización sobre el concepto de brecha digital en tiempos de la inteligencia artificial: Hacia una propuesta cualitativa. *Sintaxis*, 13, 49–78. <https://doi.org/10.36105/stx.2024n13.05>
- Luo (Jess), J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the “originality” of students’ work. *Assessment & Evaluation in Higher Education*, 49(5), 651–664. <https://doi.org/10.1080/02602938.2024.2309963>

- Lye, C. Y., & Lim, L. (2024). Generative Artificial Intelligence in Tertiary Education: Assessment Redesign Principles and Considerations. *Education Sciences*, 14(6), 569. <https://doi.org/10.3390/educsci14060569>
- Mateus, J.-C., Cappello, G., Lugo, N., & Guerrero, M. (2024). Communication Educators Facing the Arrival of Generative Artificial Intelligence: Exploration in Mexico, Peru, and Spain. *Digital Education Review*, 45, 106–114. <https://doi.org/10.1344/der.2024.45.106-114>
- Matindingue, A. D., Conte, E., & Duduka, J.. (2025). Navegando nas Fronteiras Éticas: Uma Revisão do Impacto da Inteligência Artificial na EAD. *Avaliação: Revista Da Avaliação Da Educação Superior (campinas)*, 30, e025026. <https://doi.org/10.1590/1982-57652025v30id293990>
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, 100151. <https://doi.org/10.1016/j.caeo.2023.100151>
- Morjaria, L., Burns, L., Bracken, K., Ngo, Q. N., Lee, M., Levinson, A. J., Smith, J., Thompson, P., & Sibbald, M. (2023). Examining the Threat of ChatGPT to the Validity of Short Answer Assessments in an Undergraduate Medical Program. *Journal of Medical Education and Curricular Development*, 10. <https://doi.org/10.1177/23821205231204178>
- Murcia, D., Jaime, M. F., Jaramillo, L. F., & Hoyos, Y. A. (2025). GenAI in Language Teaching, Learning, and Assessment: Stakeholders Insights from Two Undergraduate Language Programs. *Lenguaje*, 53(1S), e20314387. <https://doi.org/10.25100/lenguaje.v53i1S.14387>
- Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., Lyden, S., Neal, P., & Sandison, C. (2023). ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *European Journal of Engineering Education*, 48(4), 559–614. <https://doi.org/10.1080/03043797.2023.2213169>
- Niño, D. O. (2025). Pedagogía: Oportunidades y desafíos para enseñar en la era de la inteligencia artificial. *Voces y Silencios. Revista Latinoamericana de Educación*, 16(2), 151–168. <https://doi.org/10.18175/VyS16.2.2025.8>
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching and Learning Practice*, 21(06). <https://doi.org/10.53761/q3azde36>
- Petrovska, O., Clift, L., Moller, F., & Pearsall, R. (2024). Incorporating Generative AI into Software Development Education. *Proceedings of the*

- 8th Conference on Computing Education Practice, 37–40. <https://doi.org/10.1145/3633053.3633057>
- Salinas, D. E., Vilalta, E., Michel, R., & Montesinos, L. (2024). Using Generative Artificial Intelligence Tools to Explain and Enhance Experiential Learning for Authentic Assessment. *Education Sciences*, 14(1), 83. <https://doi.org/10.3390/educsci14010083>
- Sharma, A., Shailendra, S., & Kadel, R. (2025). Experiences with Content Development and Assessment Design in the Era of GenAI. *2025 6th International Conference on Computer Science, Engineering, and Education (CSEE)*, 1–5. <https://doi.org/10.1109/CSEE64583.2025.00008>
- Shishavan, H.B. (2024). Ai in higher education: Guidelines on assessment design from Australian universities. *ASCILITE Publications*, 118–126. <https://doi.org/10.14742/apubs.2024.1205>
- Valdivieso, T., & González, O. (2025). Generative AI Tools in Salvadoran Higher Education: Balancing Equity, Ethics, and Knowledge Management in the Global South. *Education Sciences*, 15(2), 214. <https://doi.org/10.3390/educsci15020214>
- Verdicchio, M. (2025). Adapting Program Assessment for the Age of Generative AI. *2025 IEEE Engineering Education World Conference (EDUNINE)*, 1–6. <https://doi.org/10.1109/EDUNINE62377.2025.10981409>
- Weng, X., Xia, Q., Gu, M., Rajaram, K., & Chiu, T. K. F. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*. <https://doi.org/10.14742/ajet.9540>
- White, A., & Huber, E. (2024). Evaluating an online assessment framework through the lens of Generative AI. *ASCILITE Publications*, 618–623. <https://doi.org/10.14742/apubs.2024.1423>
- Wilson, S. E., & Nishimoto, M. (2024). Assessing Learning of Computer Programming Skills in the Age of Generative Artificial Intelligence. *Journal of Biomechanical Engineering*, 146(5), 051003. <https://doi.org/10.1115/1.4064364>
- 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(1), 40. <https://doi.org/10.1186/s41239-024-00468-z>
- Yeadon, W., Inyang, O.-O., Mizouri, A., Peach, A., & Testrow, C. P. (2023). The death of the short-form physics essay in the coming AI revolution. *Physics Education*, 58(3). <https://doi.org/10.1088/1361-6552/acc5cf>

Zhao, J., Chapman, E., & Sabet, P. G. P. (2024). Generative AI and Educational Assessments: A Systematic Review. *Education Research and Perspectives*, 51, 124–155. <https://doi.org/10.70953/ERPv51.2412006>