NUEVOS MODELOS DE EVALUACIÓN AUTÉNTICA

Cómo evaluar el aprendizaje cuando la inteligencia artificial puede generar respuestas y soluciones



Rommel Triana, Yolby Crespo, Marjori Abad, Maryury Molina & Rosa Azanza

EDITORIAL SAGA

Nuevos modelos de evaluación auténtica

Cómo evaluar el aprendizaje cuando la inteligencia artificial puede generar respuestas y soluciones

Rommel Joseph Triana Teran
Yolby María Crespo Blacio
Marjori Susana Abad Jumbo
Maryury Alina Molina Sabando
Rosa María Azanza Jiménez

EDITORIAL
SAGA



Datos bibliográficos

ISBN:	**978-9907-803-96-9**
Título del libro:	Nuevos modelos de evaluación auténtica Cómo evaluar el aprendizaje cuando la inteligencia artificial puede generar respuestas y soluciones
Autores:	Triana Teran, Rommel Joseph Crespo Blacio, Yolby María Abad Jumbo, Marjori Susana Molina Sabando, Maryury Alina Azanza Jiménez, Rosa María
Editorial:	SAGA
Materia:	370 - Educación
Público objetivo:	Profesional / académico
Publicado:	2026-10-25
Número de edición:	1
Tamaño:	5Mb
Soporte:	Libro digital descargable
Formato:	Pdf (.pdf)
Idioma:	Español
DOI:	https://doi.org/10.63415/saga.2026.144

Hecho En el Ecuador / Made in Ecuador

Autores

Rommel Joseph Triana Teran
Unidad Educativa San Carlos
rommel.triana@educacion.gob.ec
https://orcid.org/0000-0002-4059-9201
Vinces, Ecuador

Yolby María Crespo Blacio
Ministerio de Educación, Deporte y Cultura
yolby.crespo@educacion.gob.ec
https://orcid.org/0009-0005-3583-3920
Machala, Ecuador

Marjori Susana Abad Jumbo
Ministerio de Educación, Deporte y Cultura
marjorie.abad@educacion.gob.ec
https://orcid.org/0009-0004-8834-0428
Machala, Ecuador

Maryury Alina Molina Sabando
Ministerio de Educación, Deporte y Cultura
maryury.molina@educacion.gob.ec
https://orcid.org/0009-0009-2710-222X
Machala, Ecuador

Rosa María Azanza Jiménez
Ministerio de Educación, Deporte y Cultura
rosa.azanza@educacion.gob.ec
https://orcid.org/0009-0001-0798-145X
Machala, Ecuador

EDITORIAL
SAGA

El contenido y las ideas expuestas en esta obra se encuentran protegidos por la normativa vigente en materia de propiedad intelectual y constituyen derechos exclusivos de su(s) autor(es)

Todos los derechos reservados © 2026

Sinopsis

La inteligencia artificial generativa ha alterado las condiciones bajo las cuales docentes e instituciones interpretan las evidencias de aprendizaje. Nuevos modelos de evaluación auténtica propone una revisión académica de los criterios, instrumentos y prácticas utilizados para valorar conocimientos, competencias, razonamiento y desempeño cuando una respuesta técnicamente correcta puede provenir de sistemas automatizados. El libro desplaza la atención desde el producto entregado hacia la trazabilidad del pensamiento, la argumentación, la toma de decisiones, la transferencia y la capacidad de justificar procesos intelectuales. A lo largo de cinco capítulos se articulan modelos de tareas auténticas, problemas abiertos, simulaciones, defensas orales, portafolios de progresión, rúbricas dinámicas y evidencias multimodales. La inteligencia artificial adquiere una función pedagógica delimitada: puede apoyar la retroalimentación, contrastar perspectivas y ampliar procesos de revisión sin reemplazar el juicio educativo. La obra también examina la autoría académica, la transparencia en el uso de herramientas generativas y los niveles autorizados de asistencia artificial. Su propuesta orienta el diseño de sistemas evaluativos centrados en evidencias verificables de comprensión y desempeño. Docentes, investigadores y responsables institucionales encontrarán criterios para renovar la evaluación, preservar la integridad académica y reconocer la contribución intelectual humana dentro de entornos educativos mediados por inteligencia artificial con rigor académico.

Palabras clave: evaluación de la educación, evaluación del estudiante, inteligencia artificial, aplicación informática, sistema experto

Synopsis

Generative artificial intelligence has altered the conditions under which educators and institutions interpret evidence of learning. New Models of Authentic Assessment presents an academic reconsideration of the criteria, instruments, and practices used to assess knowledge, competencies, reasoning, and performance when a technically correct response may originate from automated systems. The book shifts attention from the submitted product toward the traceability of thought, argumentation, decision-making, transfer, and the ability to justify intellectual processes. Across five chapters, it articulates models based on authentic tasks, open-ended problems, simulations, oral defenses, developmental portfolios, dynamic rubrics, and multimodal evidence. Artificial intelligence is assigned a clearly delimited pedagogical role: it can support feedback, contrast perspectives, and extend revision processes without replacing educational judgment. The book also examines academic authorship, transparency in the use of generative tools, and authorized levels of artificial assistance. Its proposal guides the design of assessment systems centered on verifiable evidence of understanding and performance. Teachers, researchers, and institutional leaders will find criteria for renewing assessment, preserving academic integrity, and recognizing human intellectual contribution within AI-mediated educational environments while maintaining academic rigor.

Keywords: educational evaluation, student evaluation, artificial intelligence, computer applications, expert systems

Índice General

Sinopsis ..... **v**

Índice General **7**

Índice de Tablas **10**

Índice de Figuras **11**

Introducción **13**

Capítulo I Repensar la evaluación cuando la IA también responde **17**

1.1. Del producto entregado a la evidencia de aprendizaje 21

1.2. La crisis de la respuesta correcta como indicador de aprendizaje 24

1.3. Autoría académica en escenarios de colaboración humano-IA. 28

1.4. Validez de las evidencias en la evaluación contemporánea 31

1.5. De evaluar conocimientos declarados a evaluar actuaciones intelectuales 34

1.6. Autenticidad evaluativa en contextos mediados por inteligencia artificial 38

Capítulo II Arquitecturas de evaluación auténtica para la era de la IA **43**

2.1. Tareas auténticas resistentes a la sustitución cognitiva 47

2.2. Evaluación basada en problemas abiertos y contextos situados 50

2.3. Evaluación multimodal y producción de evidencias complejas 54

2.4. Evaluaciones adaptativas basadas en decisiones del estudiante 57

2.5. Evaluación por escenarios, simulaciones y desempeño profesional 61

2.6. Ecosistemas híbridos de evaluación humana e inteligencia artificial 64

Capítulo III Evidencias del pensamiento: evaluar procesos y no solamente productos 69

3.1. Trazabilidad del razonamiento y construcción de respuestas ... 73

3.2. Evaluación del juicio crítico frente a respuestas generadas por IA 76

3.3. La defensa oral como evidencia complementaria de comprensión 80

3.4. Evaluación de la transferencia mediante variaciones del problema 83

3.5. Metacognición documentada y conciencia del propio aprendizaje 86

3.6. Evidencias longitudinales y portafolios de progresión cognitiva 89

Capítulo IV Evaluar con IA sin delegar el juicio educativo 95

4.1. Alfabetización evaluativa en inteligencia artificial 99

4.2. IA como interlocutor crítico del estudiante 103

4.3. Retroalimentación aumentada y ciclos de mejora del aprendizaje 106

4.4. Rúbricas dinámicas para desempeños mediados por IA 109

4.5. Evaluación de prompts, iteraciones y decisiones de interacción 113

4.6. Supervisión humana y límites de la automatización evaluativa 116

Capítulo V Hacia un nuevo contrato evaluativo entre estudiante, docente e IA 121

5.1. Transparencia del uso de IA como principio de integridad académica 125

5.2. De la prohibición a los niveles autorizados de asistencia artificial 128

5.3. Nuevas rúbricas para valorar la contribución intelectual humana 132

5.4. Evaluación personalizada sin fragmentar los estándares académicos 135

5.5. Integridad académica basada en evidencias y no en vigilancia 138

5.6. Diseño de un sistema institucional de evaluación auténtica con IA 142

Conclusiones 147

Referencias Bibliográficas 151

Índice de Tablas

Tabla 1 *Fundamentos para repensar la evaluación del aprendizaje ante la inteligencia artificial generativa* 41
Tabla 2 *Arquitecturas de evaluación auténtica para el aprendizaje en la era de la inteligencia artificial* 67
Tabla 3 *Dimensiones y criterios para evaluar las evidencias del pensamiento en entornos educativos mediados por inteligencia artificial* 93
Tabla 4 *Dimensiones pedagógicas y criterios para evaluar el aprendizaje mediado por inteligencia artificial* 119
Tabla 5 *Principios y componentes del nuevo contrato evaluativo entre estudiante, docente e inteligencia artificial* 145

Índice de Figuras

Figura 1 *Fundamentos de la evaluación ante la inteligencia artificial* 19
Figura 2 *Limitaciones de la respuesta correcta como evidencia de aprendizaje*.....25
Figura 3 *Criterios de validez de las evidencias de aprendizaje*.....33
Figura 4 *Arquitecturas de evaluación auténtica en la era de la IA*.....45
Figura 5 *Evaluación mediante problemas abiertos y situaciones reales* 53
Figura 6 *Evaluación adaptativa y personalización del aprendizaje*.....59
Figura 7 *Dimensiones de evaluación del pensamiento*.....71
Figura 8 *Valoración crítica de respuestas generadas por IA*.....79
Figura 9 *Transferencia del aprendizaje ante problemas variables* 85
Figura 10 *Dimensiones de la evaluación educativa mediada por IA* ...97
Figura 11 *Diálogo crítico entre estudiante e inteligencia artificial* ...105
Figura 12 *Criterios dinámicos para valorar desempeños con IA*.....111
Figura 13 *Principios del nuevo contrato evaluativo*.....123
Figura 14 *Niveles autorizados de asistencia artificial*.....131
Figura 15 *Personalización y equidad en la evaluación académica*....137

Introducción

La evaluación educativa ha experimentado transformaciones vinculadas con la expansión de las tecnologías digitales y la incorporación de sistemas capaces de producir contenidos académicos. Durante décadas, los ensayos y las pruebas escritas permitieron valorar conocimientos mediante respuestas elaboradas por los estudiantes. La inteligencia artificial generativa ha modificado las condiciones de producción de tales evidencias. Williams (2024) examinó el desempeño de herramientas generativas en evaluaciones universitarias de ciencias biomédicas y aportó elementos para reconsiderar la relación entre calidad de las respuestas y aprendizaje demostrado.

La presencia de inteligencia artificial en las actividades universitarias plantea interrogantes sobre la interpretación de los resultados académicos. Una respuesta correcta puede reflejar comprensión disciplinar o proceder de una herramienta automatizada. La distinción adquiere relevancia cuando las instituciones necesitan acreditar competencias individuales. Kizilcec et al. (2024) analizaron las percepciones de docentes y estudiantes de Australia, Chipre y Estados Unidos sobre las repercusiones de la inteligencia artificial generativa en la evaluación. Sus aportaciones permiten reconocer diferencias de perspectiva que influyen en las prácticas educativas.

La discusión académica requiere revisar los fundamentos utilizados para atribuir validez a las evidencias de aprendizaje. La autenticidad de una tarea depende de su capacidad para representar actuaciones intelectuales significativas y permitir inferencias justificadas sobre las competencias adquiridas. Kulgemeyer et al. (2023) estudiaron las relaciones entre autenticidad y validez en procedimientos evaluativos vinculados con la formación docente. Desde una perspectiva complementaria,

Vlachopoulos y Makri (2024) examinaron prácticas de evaluación auténtica en educación superior y sus relaciones con el desarrollo de competencias contemporáneas.

La necesidad de renovar los procedimientos evaluativos adquiere particular relevancia ante la disponibilidad de herramientas generativas para resolver actividades académicas. Mantener instrumentos centrados exclusivamente en productos terminados puede limitar la observación del razonamiento y la toma de decisiones. Bower et al. (2024) investigaron las respuestas docentes ante el desarrollo de inteligencia artificial generativa cada vez más potente. Sus hallazgos contribuyen a fundamentar la revisión de las prácticas de enseñanza y evaluación. La cuestión exige considerar tanto las posibilidades tecnológicas como la responsabilidad pedagógica de las instituciones.

El propósito general del libro consiste en examinar fundamentos y modelos de evaluación auténtica capaces de reconocer aprendizajes verificables en entornos educativos mediados por inteligencia artificial. Se busca identificar criterios para valorar razonamiento, argumentación y transferencia mediante evidencias complementarias. Xia et al. (2024) revisaron investigaciones sobre las transformaciones de la evaluación universitaria asociadas con inteligencia artificial generativa. Su trabajo ofrece una base académica para analizar nuevas modalidades evaluativas y delimitar las condiciones bajo las cuales pueden contribuir a una valoración fundamentada del desempeño estudiantil.

La investigación se articula alrededor de varias preguntas. ¿Qué evidencias permiten distinguir la comprensión humana de una producción automatizada? ¿Qué características deben reunir las tareas para valorar competencias transferibles? ¿De qué manera pueden incorporarse herramientas generativas sin desplazar el juicio profesional docente? También interesa determinar qué criterios permiten reconocer la contribución intelectual del

estudiante. Foung et al. (2024) abordaron oportunidades de renovación evaluativa mediante ChatGPT y otras herramientas digitales. Sus aportaciones permiten examinar alternativas para diseñar actividades que demanden participación intelectual verificable.

El primer capítulo examina los cambios que experimenta la evaluación cuando la inteligencia artificial puede producir respuestas académicamente satisfactorias. Se revisan las limitaciones del producto entregado como prueba de aprendizaje y las dificultades para establecer autoría en procesos de colaboración humano-máquina. También se analizan la validez de las evidencias y la necesidad de observar actuaciones intelectuales. Fischer et al. (2024) contrastaron las afirmaciones estudiantiles con las valoraciones de evaluadores ante el uso de inteligencia artificial generativa. Su investigación aporta elementos para comprender las tensiones que acompañan la atribución del aprendizaje.

El segundo capítulo desarrolla arquitecturas de evaluación auténtica mediante tareas resistentes a la sustitución cognitiva, problemas abiertos y evidencias multimodales. Se consideran evaluaciones adaptativas y simulaciones que requieren decisiones fundamentadas ante situaciones variables. El tercer capítulo profundiza en la documentación del razonamiento mediante registros de proceso, defensas orales y portafolios de progresión. Pentland et al. (2026) abordaron las evaluaciones orales asincrónicas como alternativa relacionada con la integridad académica y la comunicación. Su planteamiento permite ampliar la discusión sobre procedimientos que hacen observable la comprensión del estudiante.

El cuarto capítulo estudia las posibilidades de utilizar inteligencia artificial como apoyo evaluativo bajo supervisión pedagógica. Se examinan la alfabetización en inteligencia artificial y la retroalimentación aumentada junto con rúbricas dinámicas y criterios para valorar interacciones con sistemas generativos. La

interpretación docente conserva una función determinante. Delcker et al. (2025) desarrollaron un instrumento orientado a evaluar las percepciones del profesorado sobre sus competencias en inteligencia artificial. Por su parte, Langer et al. (2025) analizaron la supervisión humana de sistemas automatizados y la detección de resultados inexactos o injustos.

El quinto capítulo aborda la construcción de acuerdos institucionales sobre transparencia, autoría y niveles autorizados de asistencia artificial. También examina criterios para reconocer contribuciones intelectuales humanas y promover evaluaciones personalizadas con estándares académicos compartidos. Torres-Diaz et al. (2025) investigaron las relaciones entre inteligencia artificial e integridad académica en universidades ecuatorianas. Su aportación permite incorporar una perspectiva latinoamericana a la discusión. El libro propone integrar políticas institucionales y prácticas docentes mediante procedimientos que permitan documentar el aprendizaje y fundamentar decisiones evaluativas responsables.

Capítulo I

Repensar la evaluación cuando la IA también responde

La evaluación educativa atraviesa un momento de transformación que obliga a reconsiderar algunas de sus certezas más arraigadas. Durante generaciones, las instituciones han utilizado exámenes, informes y trabajos escritos para reconocer los conocimientos adquiridos por sus estudiantes. La expansión de la inteligencia artificial generativa introduce nuevas condiciones en esa práctica. Hoy es posible producir respuestas académicamente elaboradas mediante sistemas automatizados. Ante tal realidad, resulta necesario revisar los fundamentos que permiten atribuir significado educativo a los desempeños observados.

El cambio no comenzó con la aparición de herramientas generativas. Desde hace décadas, la educación superior debate las limitaciones de las pruebas centradas en recordar información y reproducir procedimientos. Sin embargo, la capacidad tecnológica para elaborar respuestas complejas ha dado mayor urgencia a preguntas anteriormente conocidas. Williams (2024) examinó el rendimiento de sistemas generativos en evaluaciones escritas de ciencias biomédicas. Su investigación permite reconocer la necesidad de revisar las interpretaciones académicas construidas a partir de producciones convencionales.

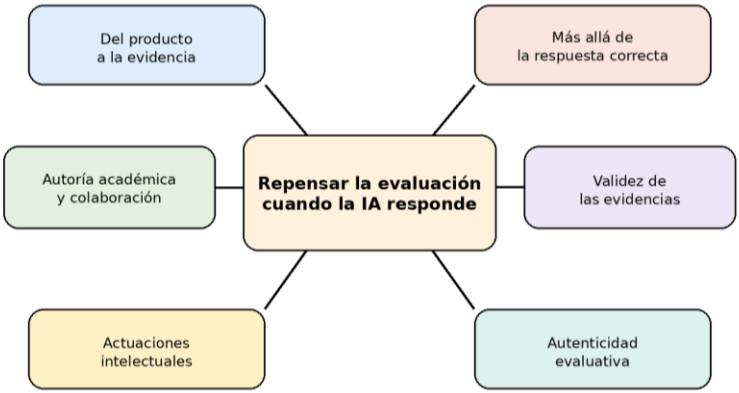
En las aulas, la transformación se manifiesta mediante situaciones que todavía generan incertidumbre. Un docente recibe trabajos cuidadosamente argumentados y necesita determinar qué capacidades representan. Los estudiantes, por su parte, encuentran recursos que modifican sus formas habituales de estudiar y resolver actividades. Kizilcec et al. (2024) investigaron las percepciones de docentes y estudiantes en Australia, Chipre y Estados Unidos. Su aproximación permite comprender que las expectativas sobre la evaluación pueden diferir entre quienes participan directamente en ella.

La discusión alcanza también las relaciones entre producción académica y responsabilidad intelectual. Las herramientas generativas intervienen en actividades que

anteriormente requerían una participación humana más directa. Su presencia modifica las condiciones de elaboración y plantea preguntas sobre las contribuciones reconocibles de cada participante. Fischer et al. (2024) abordaron las diferencias entre declaraciones estudiantiles y valoraciones de evaluadores ante la inteligencia artificial. Tal perspectiva permite apreciar las tensiones interpretativas que acompañan la transformación de las prácticas universitarias contemporáneas.

Figura 1

Fundamentos de la evaluación ante la inteligencia artificial



A medida que cambian las condiciones de elaboración académica, también resulta necesario revisar los fundamentos de la confianza evaluativa. Una actividad puede parecer exigente y ofrecer información insuficiente sobre las capacidades que pretende valorar. Kulgemeyer et al. (2023) estudiaron las relaciones entre autenticidad y validez en la formación docente, considerando los efectos de la digitalización. Su planteamiento permite comprender que la calidad evaluativa depende de las relaciones entre propósitos educativos, actuaciones observadas e interpretaciones justificadas.

La transformación también invita a reconsiderar qué significa demostrar conocimiento en situaciones donde la

información circula con extraordinaria facilidad. Recordar conceptos conserva importancia, aunque las exigencias profesionales demandan utilizarlos para interpretar problemas y tomar decisiones. Zhao y Dang (2026) examinan el rediseño de evaluaciones escritas orientadas al pensamiento crítico de orden superior ante la inteligencia artificial. Su trabajo aporta una perspectiva pertinente para discutir las capacidades intelectuales que merecen atención en la formación universitaria contemporánea.

El debate adquiere una dimensión adicional cuando se consideran las relaciones entre aprendizaje y vida profesional. Las actividades universitarias preparan para escenarios donde las decisiones suelen depender de circunstancias cambiantes y necesidades humanas concretas. Vlachopoulos y Makri (2024) revisaron investigaciones sobre evaluación auténtica en educación superior y su relación con habilidades contemporáneas. Sus aportaciones permiten reconocer la importancia de vincular las experiencias académicas con prácticas significativas y responsabilidades que trascienden los espacios institucionales de enseñanza.

Sin embargo, trasladar tales preocupaciones al trabajo cotidiano requiere algo más que modificar instrucciones o incorporar plataformas digitales. Las decisiones evaluativas forman parte de tradiciones institucionales y expectativas construidas durante años. Cambiarlas implica revisar criterios compartidos y reconocer diferencias entre disciplinas. También exige considerar las condiciones materiales de estudiantes y docentes. Una transformación pedagógica adquiere sentido cuando sus fundamentos pueden comprenderse y aplicarse de manera razonable dentro de las realidades educativas donde se desarrolla.

En medio de tales cambios, el profesorado conserva una responsabilidad intelectual y pedagógica que ninguna herramienta puede asumir plenamente. Interpretar las necesidades formativas

de un grupo requiere conocimiento disciplinar y sensibilidad frente a las circunstancias del aprendizaje. La inteligencia artificial puede ampliar recursos disponibles, pero también introduce incertidumbres que demandan criterio profesional. La discusión evaluativa necesita reconocer esa complejidad sin reducirla a posiciones de aceptación entusiasta o rechazo preventivo frente a las tecnologías emergentes.

El capítulo aborda las transformaciones que atraviesan los fundamentos de la evaluación educativa cuando las respuestas pueden producirse mediante colaboración humano-IA. La reflexión parte de las relaciones entre producciones académicas y aprendizaje, para examinar después la corrección, la autoría y la validez de las evidencias. Posteriormente considera las actuaciones intelectuales y la autenticidad de las experiencias evaluativas. El propósito es ofrecer fundamentos para comprender qué merece valorarse y bajo qué condiciones pueden formularse juicios académicos responsables.

1.1. Del producto entregado a la evidencia de aprendizaje

Durante décadas, buena parte de la evaluación educativa ha descansado en una premisa aparentemente razonable: el trabajo presentado por un estudiante permite reconocer aquello que ha aprendido. Un ensayo bien organizado, una solución matemática correcta o un informe técnicamente elaborado funcionaban como manifestaciones visibles de sus capacidades. Sin embargo, la expansión de la inteligencia artificial generativa ha debilitado tal correspondencia. La calidad del producto conserva valor académico, pero resulta insuficiente para comprender el aprendizaje que realmente ocurrió durante su elaboración.

La dificultad aparece cuando una respuesta académicamente aceptable puede producirse sin que medie una comprensión equivalente por parte de quien la entrega. Los

sistemas generativos redactan argumentos, resuelven problemas y organizan información mediante procedimientos que permanecen parcialmente ajenos al estudiante. Williams (2024) examinó el desempeño de la inteligencia artificial generativa en evaluaciones escritas de ciencias biomédicas de pregrado y posgrado. Su investigación aporta elementos para cuestionar la confianza depositada en el producto escrito como indicador suficiente del dominio disciplinar.

El problema adquiere mayor profundidad cuando se reconoce que dos estudiantes pueden presentar trabajos semejantes después de experiencias intelectuales completamente diferentes. Uno quizá haya investigado, contrastado fuentes y revisado sus argumentos durante varias jornadas. Otro pudo obtener una respuesta mediante instrucciones breves dirigidas a una herramienta generativa. Ambos productos podrían satisfacer los criterios formales de una rúbrica convencional. La diferencia educativa reside en las operaciones cognitivas realizadas y en la capacidad de explicar las decisiones que dieron origen al resultado presentado.

Desplazar la atención hacia la evidencia de aprendizaje exige revisar aquello que el profesorado considera observable y digno de valoración. Una respuesta correcta informa sobre el resultado alcanzado, aunque ofrece indicios limitados acerca de las capacidades movilizadas. Conviene examinar también las decisiones tomadas, los errores reconocidos y las razones que justifican determinadas elecciones. Desde tal perspectiva, evaluar implica reconstruir una trayectoria intelectual mediante manifestaciones verificables. La evidencia adquiere sentido cuando permite relacionar una actuación concreta con los aprendizajes previstos en cada actividad.

Una evidencia valiosa tampoco equivale a acumular borradores o registrar cada movimiento del estudiante. Su utilidad depende de la pregunta pedagógica que permite responder. Un

cambio entre versiones puede revelar revisión conceptual si quien aprende explica por qué modificó su planteamiento. Una conversación breve puede mostrar comprensión cuando exige defender una decisión frente a una objeción inesperada. Lo importante es que las huellas recogidas permitan interpretar el desempeño sin confundir actividad visible con elaboración intelectual.

La investigación de Williams (2024) permite situar la discusión en un terreno: las tareas escritas universitarias pueden ser abordadas por sistemas generativos con resultados que obligan a reconsiderar su función evaluativa. Sin embargo, reconocer tal capacidad tecnológica no autoriza a desestimar la escritura académica. Redactar continúa siendo una práctica formativa cuando requiere seleccionar argumentos, valorar información y sostener posiciones fundamentadas. La cuestión reside en determinar qué aprendizaje puede atribuirse al estudiante que presenta el documento.

También cambia el significado del error. En una evaluación centrada en el producto, una equivocación suele interpretarse como deficiencia que reduce la calificación. Cuando interesa comprender el aprendizaje, los errores permiten reconocer hipótesis iniciales, límites de comprensión y ajustes. Un estudiante que identifica una respuesta engañosa de la inteligencia artificial y fundamenta su corrección ofrece información relevante sobre su criterio. La valoración requiere distinguir entre equivocaciones persistentes y decisiones reflexivas que conducen a una comprensión precisa.

La participación de herramientas generativas introduce otra exigencia: reconocer con claridad las contribuciones humanas y tecnológicas. Declarar su uso puede aportar información, pero una declaración aislada difícilmente demuestra aprendizaje. Resulta más esclarecedor solicitar que el estudiante justifique qué aceptó, qué descartó y bajo qué criterios revisó las respuestas

obtenidas. La transparencia adquiere valor educativo cuando permite examinar la responsabilidad intelectual asumida durante el trabajo y evita que la apariencia de autonomía sustituya a su ejercicio.

El profesorado necesita interpretar las evidencias con prudencia. Ninguna intervención, anotación de proceso o producción escrita garantiza por sí misma una comprensión profunda. Cada manifestación ofrece información y puede estar condicionada por la familiaridad con la tarea, la presión del momento o las oportunidades de preparación. Por ello, la valoración gana consistencia cuando distintas actuaciones permiten reconocer capacidades semejantes. La coherencia entre lo que el estudiante explica, decide y realiza aporta fundamentos para emitir juicios.

En consecuencia, el producto entregado deja de ocupar una posición privilegiada como prueba de aprendizaje. Permanece como parte de una experiencia educativa cuya interpretación demanda conocer las decisiones que lo hicieron posible. Williams (2024) contribuye a comprender por qué la competencia de los sistemas generativos vuelve insuficiente una lectura basada en respuestas escritas. La tarea evaluativa consiste en reconocer actuaciones atribuibles al estudiante y fundamentar juicios sobre aquello que comprende y puede hacer con autonomía.

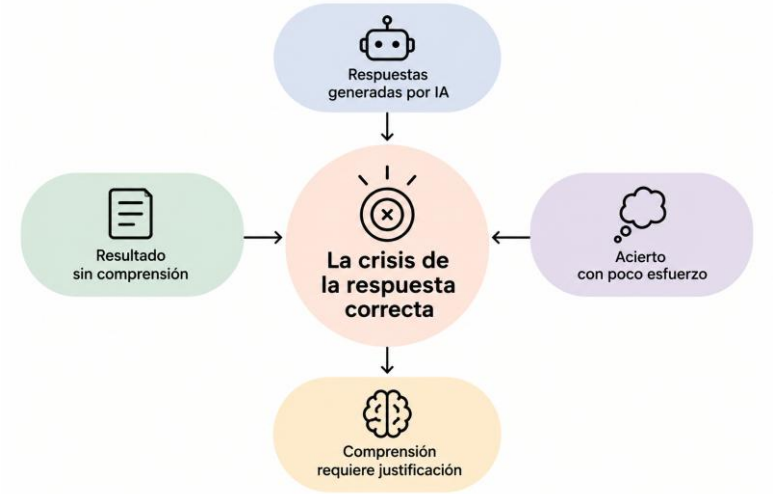
1.2. La crisis de la respuesta correcta como indicador de aprendizaje

La respuesta correcta ha ocupado históricamente un lugar privilegiado en la evaluación educativa. Su presencia parecía confirmar que el estudiante había comprendido un contenido y podía aplicarlo adecuadamente. Tal asociación funcionaba mientras resolver una tarea académica exigía movilizar conocimientos propios. Hoy convence menos. Una respuesta precisa puede proceder de operaciones automatizadas que el

estudiante apenas comprende. La dificultad obliga a revisar el significado pedagógico de acertar y las inferencias construidas a partir del acierto.

El acierto importa cuando expresa dominio conceptual y permite resolver situaciones relevantes. Sin embargo, pierde fuerza interpretativa cuando diferentes caminos conducen al mismo resultado sin esfuerzos equivalentes. Una fórmula aplicada mecánicamente puede producir el número esperado. También puede hacerlo una herramienta generativa con una instrucción. En ambos casos aparece la solución solicitada, pero permanece abierta la pregunta sobre la comprensión. La evaluación necesita distinguir la coincidencia con una respuesta esperada del conocimiento disponible para justificarla.

Figura 2
Limitaciones de la respuesta correcta como evidencia de aprendizaje



La incertidumbre alcanza a tareas consideradas exigentes por su complejidad. Un argumento aparentemente coherente puede parecer señal de pensamiento crítico aunque su autor

desconozca las razones que lo sostienen. La fluidez del lenguaje añade una dificultad particular: convierte afirmaciones discutibles en explicaciones persuasivas. Cuando el profesorado identifica corrección formal con comprensión, corre el riesgo de valorar una apariencia de competencia. Conviene recordar que escribir con seguridad y razonar con fundamento son capacidades relacionadas, aunque diferentes.

Las percepciones de quienes enseñan y aprenden permiten observar la transformación desde posiciones diferentes. Kizilcec et al. (2024) compararon perspectivas de docentes y estudiantes sobre el impacto de la IA generativa en evaluaciones de Australia, Chipre y Estados Unidos. Su trabajo muestra la pertinencia de considerar ambas voces al examinar cambios educativos asociados con tales herramientas. Las expectativas sobre una respuesta aceptable pueden variar según la experiencia de uso y la responsabilidad asumida ante la evaluación.

Otra consecuencia afecta al valor del esfuerzo cognitivo. Si obtener la solución demanda pocos segundos, el estudiante puede interpretar que comprender el procedimiento constituye una inversión innecesaria. Esa percepción no surge inevitablemente del uso tecnológico; depende también de aquello que la institución reconoce y recompensa. Cuando la calificación privilegia acertar sin atender a las razones, termina fortaleciendo una relación instrumental con el conocimiento. Aprender queda reducido entonces a conseguir resultados verificables por cualquier recurso disponible.

La respuesta correcta también puede ocultar una comprensión conceptual frágil. Alguien reproduce una definición impecable y, sin embargo, encuentra dificultades para reconocer el concepto en situaciones inesperadas. Otro resuelve un ejercicio previamente conocido, pero no advierte que sus condiciones han cambiado. Tales discrepancias existían antes de la inteligencia artificial. La novedad reside en que los sistemas generativos

amplían las oportunidades de producir desempeños convincentes sin superar las limitaciones conceptuales que todavía permanecen detrás de ellos.

Kizilcec et al. (2024) compararon percepciones estudiantiles y docentes en varios países, ofreciendo un punto de partida para reconocer que la evaluación involucra expectativas compartidas y discrepancias. Aquello que un profesor interpreta como demostración de competencia puede recibir otro significado entre estudiantes habituados a consultar sistemas generativos. La discusión requiere atender a tales diferencias sin convertirlas en juicios apresurados sobre honestidad o compromiso. El problema afecta a las reglas mediante las cuales se reconoce el saber.

Frente a tal escenario, cobra relevancia una pregunta: qué permite afirmar que alguien sabe cuando dispone de respuestas inmediatas. La precisión sigue siendo necesaria en muchos campos, especialmente cuando un error compromete decisiones posteriores. Pero conocer también exige advertir límites, reconocer excepciones y explicar por qué una solución resulta pertinente. Una respuesta puede ser correcta bajo determinadas condiciones y dejar de serlo cuando cambia el problema. Comprender significa percibir esa dependencia y actuar con criterio.

El cambio plantea dificultades para la interpretación de las calificaciones. Dos resultados idénticos pueden expresar niveles distintos de independencia intelectual, mientras una respuesta incompleta puede revelar razonamientos valiosos. Tal posibilidad obliga a evitar equivalencias automáticas entre puntuación y aprendizaje. Kizilcec et al. (2024) sitúan las percepciones sobre la inteligencia artificial en un debate que involucra a docentes y estudiantes de distintos entornos universitarios. La legitimidad del juicio académico depende de que sus criterios resulten comprensibles.

Recuperar el significado educativo de acertar requiere reconocer su alcance y sus límites. La respuesta correcta sigue siendo una señal útil cuando puede relacionarse con comprensión y capacidad de juicio. Fuera de esa relación, ofrece una información estrecha para sostener conclusiones sobre el aprendizaje. La presencia de inteligencia artificial vuelve más visible una fragilidad anterior de muchas prácticas evaluativas. Acertar importa, desde luego, pero importa todavía más comprender por qué una respuesta merece considerarse correcta.

1.3. Autoría académica en escenarios de colaboración humano-IA

La autoría académica ha estado vinculada con la responsabilidad de quien firma un trabajo y responde por sus afirmaciones. La inteligencia artificial generativa introduce una dificultad particular en esa relación. Un estudiante puede formular preguntas, seleccionar propuestas y modificar textos producidos mediante sistemas automatizados. El resultado conserva decisiones humanas, aunque parte de su elaboración proceda de una herramienta. Determinar quién es autor exige examinar la naturaleza de las intervenciones y el compromiso intelectual personal asumido.

La colaboración humano-IA no presenta una distribución uniforme de tareas. En ocasiones, la herramienta facilita una búsqueda inicial de enfoques o propone una estructura que después recibe cambios sustanciales. En otras, produce argumentos completos que apenas son revisados antes de su entrega. Ambas situaciones pueden parecer semejantes desde el documento presentado. Sin embargo, difieren en la participación intelectual del estudiante. Reconocer tal diferencia requiere atender al grado de intervención humana y a su significado académico.

La discusión adquiere relevancia cuando las declaraciones estudiantiles sobre el uso tecnológico se comparan con las

valoraciones docentes. Fischer et al. (2024) abordaron precisamente esa relación entre afirmaciones de estudiantes y evaluaciones de quienes califican trabajos académicos. El enfoque permite comprender que la atribución de autoría también depende de interpretaciones sobre la participación legítima de herramientas generativas. Una misma práctica puede recibir valoraciones diferentes cuando faltan acuerdos institucionales claros sobre sus límites y responsabilidades.

Firmar un trabajo implica hacerse cargo de sus afirmaciones, incluso cuando parte del contenido procede de una aplicación. Quien incorpora una explicación generada automáticamente debe poder reconocer sus fundamentos y advertir posibles errores. La responsabilidad académica no desaparece por haber delegado una operación de escritura. Tampoco queda resuelta mediante una declaración genérica de uso. Exige que el estudiante mantenga capacidad de decisión sobre aquello que presenta como conocimiento y acepta defender públicamente de manera razonada.

Existe además una dimensión formativa de la autoría que suele pasar inadvertida. Elaborar un texto propio permite descubrir vacíos conceptuales y modificar posiciones durante la escritura. Si una herramienta resuelve anticipadamente tales dificultades, algunas oportunidades de aprendizaje pueden perderse de manera significativa. Pero también puede ofrecer alternativas que estimulen nuevas preguntas. La diferencia depende del papel asignado al estudiante: aceptar propuestas sin examinarlas o convertirlas en materiales para ejercer su propio juicio crítico.

Fischer et al. (2024) examinaron las diferencias entre lo que estudiantes declaran y lo que evaluadores consideran respecto de la inteligencia artificial generativa. Tal planteamiento invita a mirar la autoría como una cuestión de interpretación institucional además de conducta individual. Las normas poco precisas dejan espacio para desacuerdos sobre prácticas habituales. Por ello, establecer

expectativas comprensibles resulta necesario para que la colaboración tecnológica pueda ser discutida sin depender de sospechas o criterios cambiantes en cada institución.

La atribución responsable también requiere distinguir tareas instrumentales de decisiones sustantivas. Corregir una errata mediante software no tiene el mismo alcance que delegar la construcción de una interpretación histórica. Entre ambos extremos aparecen actividades difíciles de clasificar, como reorganizar argumentos o proponer ejemplos. Una regulación académica razonable debe considerar la finalidad de cada actividad y el aprendizaje comprometido. De otro modo, corre el riesgo de tratar intervenciones muy diferentes como conductas equivalentes en términos pedagógicos.

La transparencia puede contribuir a resolver parte de esas tensiones cuando se practica con precisión. Informar qué herramienta intervino y para qué propósito permite comprender mejor las condiciones de elaboración. Aun entonces, declarar un uso no demuestra que las decisiones hayan sido reflexivas. La información adquiere mayor valor cuando el estudiante explica sus criterios de aceptación y revisión. Tal explicación facilita una conversación académica sobre responsabilidad ante sus profesores sin reducirla a la vigilancia del proceso.

En los entornos universitarios, la autoría también expresa pertenencia a una comunidad que comparte reglas de reconocimiento y atribución. Fischer et al. (2024) ofrecen una perspectiva pertinente al contrastar posiciones estudiantiles y evaluadoras ante la IA generativa. La convivencia de expectativas diferentes puede afectar la confianza si no existen criterios discutidos previamente. De ahí que las instituciones deban precisar de forma explícita qué contribuciones pueden delegarse y cuáles permanecen bajo responsabilidad directa de quien aprende.

La autoría académica adquiere entonces un significado menos ligado a la producción material de cada frase y más relacionado con la responsabilidad por las decisiones intelectuales. Esa ampliación no elimina la necesidad de reconocer aportes externos ni permite atribuirse razonamientos ajenos. Obliga a establecer límites defendibles para cada actividad formativa. En una colaboración humano-IA, ser autor implica comprender lo presentado, justificar las elecciones realizadas y responder por el conocimiento que se comunica ante otros investigadores.

1.4. Validez de las evidencias en la evaluación contemporánea

La validez de una evidencia depende de la interpretación que permite sostener sobre el aprendizaje. Un registro puede ser auténtico y estar cuidadosamente documentado sin representar la capacidad que pretende evaluar. En escenarios mediados por inteligencia artificial, esa distinción adquiere especial relevancia. La presencia de resultados convincentes obliga a examinar las condiciones bajo las cuales fueron producidos. Importa determinar si la actuación observada guarda una relación defendible con el conocimiento que la institución desea reconocer.

Durante mucho tiempo, la calidad de una tarea se asoció con su semejanza respecto de actividades profesionales. Sin embargo, reproducir una situación laboral no garantiza que la evaluación permita inferir competencias pertinentes. Kulgemeyer et al. (2023) analizan las relaciones entre autenticidad y validez en la formación docente, considerando la digitalización de los métodos evaluativos. Su planteamiento invita a examinar qué aspectos de una práctica profesional quedan representados y cuáles desaparecen al convertirla en actividad evaluable.

Una evidencia puede perder validez cuando la tarea exige habilidades distintas de aquellas que se desean valorar. Por ejemplo, una actividad destinada a examinar razonamiento

científico podría terminar premiando principalmente la destreza para redactar instrucciones dirigidas a un sistema generativo. Esa habilidad puede resultar pertinente en determinados programas, pero no reemplaza automáticamente el conocimiento disciplinar. La dificultad reside en identificar qué capacidades intervienen en el desempeño y qué peso corresponde atribuirles dentro del juicio académico.

La interpretación también depende de las condiciones de realización. Resolver un problema con acceso a herramientas digitales constituye una actuación diferente de resolverlo sin asistencia. Ninguna modalidad resulta intrínsecamente superior para todos los propósitos educativos. La elección debe responder a la competencia que se pretende observar. Cuando las condiciones permanecen de manera indefinida, una misma calificación puede reunir desempeños difícilmente comparables. La validez exige reconocer esas diferencias antes de atribuir significado a los resultados obtenidos.

Kulgemeyer et al. (2023) relacionan la autenticidad de las evaluaciones con las inferencias que pueden formularse sobre capacidades profesionales. Tal relación resulta especialmente útil cuando las tareas incorporan tecnologías habituales fuera del aula. Permitir herramientas semejantes a las utilizadas en el trabajo puede aumentar la pertinencia de una actividad. Aun entonces, conviene comprobar si el diseño ofrece oportunidades reales para manifestar las competencias buscadas. La semejanza superficial con la práctica profesional no basta para justificarla.

Otro problema aparece cuando una evidencia representa una fracción demasiado estrecha del aprendizaje. Una intervención oral breve puede mostrar claridad conceptual, pero quizá no permita valorar la capacidad de investigar durante periodos prolongados. Un proyecto extenso ofrece información distinta y también presenta limitaciones. La decisión evaluativa requiere considerar qué dimensiones quedan visibles y cuáles permanecen

fuera de observación. Ningún instrumento debería recibir una autoridad interpretativa mayor que la información que efectivamente proporciona sobre el desempeño.

La digitalización añade posibilidades de registro que pueden resultar atractivas para las instituciones. Plataformas educativas conservan respuestas, tiempos de participación y secuencias de interacción. Sin embargo, disponer de abundantes datos no equivale a contar con evidencias válidas. Una conexión prolongada no demuestra atención sostenida y numerosas intervenciones tampoco acreditan comprensión. La interpretación necesita fundamentos pedagógicos explícitos. De otro modo, la facilidad para medir determinados comportamientos puede terminar definiendo aquello que se considera aprendizaje.

Figura 3
Criterios de validez de las evidencias de aprendizaje



La investigación de Kulgemeyer et al. (2023) aborda los efectos de digitalizar métodos de evaluación dentro de un modelo orientado a la formación docente. Su contribución permite advertir que modificar el medio de aplicación puede alterar las

características de la actuación observada. Trasladar una actividad presencial a una plataforma no conserva necesariamente sus exigencias originales. Por ello, cada transformación tecnológica requiere revisar las inferencias originalmente previstas y las condiciones que hacen posible sostenerlas con suficiente fundamento académico.

La validez también reclama atención a las diferencias entre estudiantes. Una tarea puede favorecer a quienes dominan determinadas herramientas aunque el propósito académico sea valorar otra competencia. Las restricciones de acceso, la familiaridad tecnológica y las condiciones de participación influyen en las oportunidades de demostrar conocimientos. Reconocer tales factores no implica reducir las exigencias. Significa procurar que los resultados expresen realmente aquello que se busca evaluar y no ventajas circunstanciales ajenas al propósito formativo establecido.

Evaluar con validez exige mantener una relación explícita entre propósitos, actuaciones observables e interpretaciones académicas. La inteligencia artificial vuelve más necesaria esa vigilancia porque modifica las condiciones de producción y amplía las formas posibles de participación. El juicio docente conserva importancia, aunque necesita apoyarse en criterios justificables y evidencias pertinentes. Una evaluación contemporánea merece confianza cuando sus resultados permiten formular afirmaciones prudentes sobre el aprendizaje, reconociendo tanto sus alcances como sus límites ante situaciones diversas.

1.5. De evaluar conocimientos declarados a evaluar actuaciones intelectuales

La evaluación de conocimientos declarados ha permitido comprobar durante generaciones qué conceptos, principios y procedimientos pueden recordar los estudiantes. Su utilidad permanece cuando interesa reconocer aprendizajes fundamentales. Sin embargo, disponer de información no garantiza utilizarla con

pertinencia ante situaciones complejas. La inteligencia artificial generativa acentúa esa diferencia al facilitar explicaciones elaboradas sobre prácticamente cualquier materia. El interés educativo se desplaza hacia las actuaciones intelectuales mediante las cuales una persona interpreta problemas y toma decisiones fundamentadas con conocimiento propio.

Una actuación intelectual comienza cuando el estudiante necesita decidir qué conocimientos resultan pertinentes para comprender una situación. No basta con recuperar definiciones almacenadas en la memoria. Debe reconocer relaciones, identificar información insuficiente y establecer criterios para avanzar. Tales operaciones adquieren significado cuando el problema admite interpretaciones diversas. La evaluación puede observar entonces la capacidad de seleccionar recursos conceptuales y utilizarlos razonadamente. El conocimiento permanece como fundamento indispensable, aunque su valor aparece vinculado con las decisiones adoptadas.

Zhao y Dang (2026) examinan la transformación del diseño de evaluaciones escritas ante la incorporación de inteligencia artificial, con especial atención al pensamiento crítico de orden superior. Su planteamiento permite reconsiderar tareas centradas en reproducir información disponible. Cuando una actividad requiere valorar alternativas o fundamentar decisiones, el estudiante enfrenta exigencias intelectuales diferentes. La dificultad académica deja de depender principalmente de la cantidad de contenidos recordados y adquiere relación con la profundidad del razonamiento desarrollado.

Conviene distinguir entre ejecutar una instrucción y comprender las decisiones que exige su cumplimiento. Un estudiante puede aplicar correctamente un procedimiento conocido sin reconocer las condiciones que justifican su empleo. En cambio, elegir entre métodos alternativos demanda examinar restricciones y anticipar consecuencias. La actuación intelectual

aparece precisamente en esa deliberación. Su evaluación requiere problemas que permitan observar elecciones justificadas. También necesita criterios capaces de valorar la pertinencia del razonamiento, incluso cuando existen soluciones diferentes igualmente defendibles.

El pensamiento crítico adquiere una expresión concreta cuando las afirmaciones disponibles deben someterse a examen. Una explicación generada por inteligencia artificial puede contener argumentos plausibles y omisiones difíciles de advertir. El estudiante necesita contrastar fundamentos, reconocer contradicciones y determinar qué información merece confianza. Zhao y Dang (2026) vinculan el rediseño de evaluaciones escritas con la promoción de capacidades críticas superiores. Tal orientación permite valorar el examen razonado de afirmaciones por encima de su aceptación inmediata.

La transferencia del conocimiento constituye otra manifestación relevante. Comprender un principio permite reconocer su utilidad fuera del ejemplo utilizado durante la enseñanza. Ante una situación desconocida, el estudiante debe identificar semejanzas pertinentes sin ignorar diferencias importantes. Esa operación revela flexibilidad intelectual y capacidad de adaptación. Para valorarla, las tareas necesitan introducir variaciones que impidan respuestas rutinarias. El propósito consiste en observar si los saberes adquiridos orientan decisiones razonables cuando cambian las nuevas condiciones del problema.

También importa la capacidad de formular preguntas. Muchas actividades académicas presentan problemas perfectamente delimitados y entregan toda la información necesaria para resolverlos. La práctica intelectual suele comenzar antes, cuando alguien identifica una dificultad y decide qué merece investigarse. Reconocer supuestos discutibles o precisar una pregunta confusa requiere dominio conceptual. Una evaluación

orientada hacia actuaciones intelectuales puede atender a esa labor inicial. Formular adecuadamente un problema representa ya una intervención significativa sobre el conocimiento realmente disponible.

Zhao y Dang (2026) sitúan el pensamiento crítico de orden superior entre las preocupaciones del rediseño de evaluaciones escritas. Desde esa perspectiva, argumentar requiere algo más que organizar afirmaciones convincentes. El estudiante necesita establecer relaciones entre pruebas y conclusiones, considerar objeciones y reconocer incertidumbres. La calidad de una actuación intelectual depende también de su disposición a revisar posiciones. Cambiar de criterio ante razones mejores puede revelar mayor comprensión que defender una respuesta inicial sin examinarla.

Observar actuaciones intelectuales plantea exigencias particulares al profesorado. Las consignas deben ofrecer oportunidades reales para interpretar, decidir y justificar sin convertir cada actividad en una prueba excesivamente extensa. También conviene distinguir la complejidad del razonamiento de la dificultad accidental de la tarea. Una pregunta confusa puede obstaculizar la expresión de conocimientos valiosos. Diseñar experiencias evaluables requiere precisión pedagógica y atención a las condiciones en que cada estudiante puede demostrar efectivamente su propia capacidad de pensar.

El tránsito hacia actuaciones intelectuales modifica el significado de aprender en una cultura de respuestas disponibles. Recordar información continúa siendo necesario para razonar con autonomía, pero adquiere sentido al intervenir en decisiones fundamentadas. La evaluación puede reconocer ese vínculo mediante problemas que demanden interpretación y juicio. Lo relevante es apreciar la capacidad de actuar intelectualmente ante situaciones que no ofrecen instrucciones completas. Allí se manifiesta un aprendizaje capaz de sostenerse más allá de una respuesta.

1.6. Autenticidad evaluativa en contextos mediados por inteligencia artificial

La autenticidad evaluativa adquiere matices cuando las actividades académicas transcurren en entornos donde la inteligencia artificial participa de manera habitual. Durante años, aproximar las tareas universitarias a situaciones profesionales parecía suficiente para fortalecer su pertinencia. Hoy resulta necesario examinar también las condiciones tecnológicas de esas prácticas. Una actividad auténtica debe representar decisiones, restricciones y responsabilidades reconocibles para quienes ejercen una profesión. Su valor educativo depende de esa correspondencia y de las capacidades que permite desarrollar.

El carácter auténtico de una tarea no procede de incorporar herramientas novedosas ni de reproducir superficialmente un ambiente laboral. Aparece cuando el estudiante afronta demandas comparables con aquellas que encontrará fuera de la institución. Interpretar necesidades de personas concretas, negociar prioridades o responder ante información incompleta modifica el sentido del trabajo académico. La inteligencia artificial puede intervenir en tales situaciones, pero su presencia requiere una finalidad realmente reconocible dentro del problema que se intenta resolver.

Vlachopoulos y Makri (2024) revisan la literatura sobre evaluación auténtica en educación superior y consideran prácticas orientadas al desarrollo de habilidades del siglo XXI. Su trabajo permite situar la autenticidad dentro de decisiones pedagógicas que trascienden el formato de las actividades. La cercanía con la práctica profesional necesita traducirse también en oportunidades de participación significativa. También exige atender a las políticas institucionales que determinan qué experiencias reciben reconocimiento académico y bajo qué condiciones se desarrollan.

En una actividad profesional rara vez aparecen problemas completamente ordenados. Los datos pueden resultar contradictorios y las necesidades de los destinatarios cambian durante el trabajo. Incorporar cierta incertidumbre a la evaluación permite representar mejor esa realidad, siempre que las exigencias sean razonables. Un futuro docente podría adaptar una propuesta didáctica ante dificultades inesperadas de aprendizaje. Un estudiante de ingeniería podría justificar modificaciones frente a restricciones materiales. Ambos necesitan responder a circunstancias que alteran sus decisiones.

La mediación tecnológica plantea además una cuestión de pertinencia profesional. En determinadas ocupaciones, utilizar sistemas generativos forma parte de las prácticas emergentes; en otras, su aplicación permanece limitada por exigencias de confidencialidad o seguridad. Una tarea auténtica debe reconocer tales diferencias. Prohibir cualquier asistencia digital puede alejar la actividad de ciertas condiciones laborales. Permitirla indiscriminadamente también puede distorsionar responsabilidades profesionales. El diseño requiere establecer condiciones de uso acordes con la situación representada y sus consecuencias.

Vlachopoulos y Makri (2024) vinculan la evaluación auténtica con el desarrollo de capacidades relevantes para la educación superior contemporánea. Esa orientación invita a considerar la colaboración como parte de determinadas actuaciones profesionales. Muchos problemas requieren escuchar perspectivas ajenas, distribuir responsabilidades y negociar decisiones. La inteligencia artificial puede aportar información durante el intercambio, aunque no reemplaza directamente las relaciones humanas implicadas. Evaluar la colaboración exige observar compromisos compartidos y aportaciones reconocibles dentro de una actividad común.

La autenticidad también depende de los destinatarios del trabajo. Elaborar una propuesta para una comunidad educativa concreta exige atender necesidades que una consigna abstracta puede pasar por alto. Cuando existe un interlocutor reconocible, las decisiones adquieren consecuencias que deben explicarse con cuidado. El estudiante necesita ajustar lenguaje, alcance y recursos disponibles. No siempre será posible trabajar con organizaciones externas. Aun dentro del aula pueden construirse situaciones verosímiles que demanden efectivamente responsabilidad ante otros participantes reales.

Otro aspecto corresponde a los criterios empleados para valorar el desempeño. En una práctica auténtica, la calidad no se reduce al cumplimiento de instrucciones escolares. Importan la pertinencia de las decisiones y su viabilidad frente a restricciones concretas. También cuenta la atención prestada a quienes recibirán sus efectos. Los criterios deben reflejar tales exigencias sin perder claridad pedagógica. De poco serviría proponer situaciones profesionales complejas si después se calificaran mediante indicadores ajenos a sus propósitos.

La revisión de Vlachopoulos y Makri (2024) incorpora consideraciones sobre políticas relacionadas con la evaluación auténtica. Tal dimensión resulta pertinente ante la expansión de herramientas generativas. Las instituciones necesitan ofrecer condiciones para diseñar actividades vinculadas con prácticas profesionales cambiantes. Eso requiere formación docente y acuerdos sobre recursos disponibles. También demanda reconocer diferencias entre disciplinas. Una política uniforme puede facilitar procedimientos administrativos, pero difícilmente representa todas las formas de actuación profesional que conviven en la universidad.

La autenticidad evaluativa conserva su sentido cuando aproxima el aprendizaje a responsabilidades que existen más allá de la calificación. La inteligencia artificial transforma algunas condiciones de trabajo y obliga a revisar qué experiencias merecen

reproducirse en la enseñanza. No toda innovación tecnológica mejora una tarea ni toda práctica tradicional pierde pertinencia. La decisión pedagógica consiste en construir situaciones creíbles donde los estudiantes deban responder ante problemas relevantes, personas afectadas y límites de su campo profesional.

Tabla 1

Fundamentos para repensar la evaluación del aprendizaje ante la inteligencia artificial generativa

Eje de análisis	Implicaciones para la evaluación educativa
Del producto entregado a la evidencia de aprendizaje	La valoración académica requiere examinar las decisiones intelectuales y los procesos de elaboración que permiten reconocer aprendizajes atribuibles al estudiante.
La crisis de la respuesta correcta como indicador de aprendizaje	La precisión de una respuesta pierde capacidad explicativa cuando puede obtenerse mediante sistemas generativos. La comprensión requiere justificación y capacidad de razonamiento.
Autoría académica en escenarios de colaboración humano-IA	La responsabilidad intelectual exige reconocer las contribuciones tecnológicas y justificar las decisiones humanas que intervienen en la producción académica.
Validez de las evidencias en la evaluación contemporánea	Las evidencias deben guardar correspondencia con las competencias evaluadas y permitir interpretaciones fundamentadas sobre el aprendizaje demostrado.
De evaluar conocimientos declarados a evaluar actuaciones intelectuales	La evaluación debe valorar la aplicación del conocimiento mediante interpretación de problemas y argumentación crítica ante situaciones complejas.
Autenticidad evaluativa en entornos mediados por inteligencia artificial	Las actividades evaluativas deben representar situaciones profesionales significativas que demanden decisiones pertinentes y responsabilidad frente a problemas reales.

Nota. Elaboración propia a partir de los fundamentos conceptuales desarrollados en el capítulo y de las fuentes citadas.

Capítulo II

Arquitecturas de evaluación auténtica para la era de la IA

-

La expansión de la inteligencia artificial generativa ha modificado las condiciones bajo las cuales las instituciones educativas reconocen y valoran el aprendizaje. Las respuestas elaboradas mediante sistemas automatizados pueden alcanzar niveles de precisión y coherencia que dificultan identificar la participación intelectual del estudiante. Ante tal realidad, la evaluación necesita revisar sus fundamentos y procedimientos. El problema trasciende la incorporación de herramientas tecnológicas y alcanza las decisiones pedagógicas que determinan qué conocimientos merecen valorarse y mediante cuáles evidencias pueden reconocerse.

Durante décadas, buena parte de las prácticas evaluativas descansó en la confianza depositada en los productos académicos entregados por los estudiantes. La disponibilidad de sistemas capaces de redactar textos y resolver problemas obliga a examinar los alcances de tal confianza. Peters y Angelov (2025) abordan la redefinición de las tareas evaluativas desde la creatividad y la integridad académica. Su planteamiento permite reconocer la necesidad de construir experiencias donde el aprendizaje conserve una expresión intelectual verificable y pedagógicamente significativa.

La transformación de las prácticas evaluativas requiere comprender que los conocimientos adquieren significado cuando permiten interpretar situaciones y orientar actuaciones. Una respuesta formalmente correcta puede resultar insuficiente para reconocer la profundidad de una comprensión. Foung et al. (2024) examinan oportunidades de renovación evaluativa vinculadas con ChatGPT y otras herramientas digitales. Desde esa perspectiva, resulta pertinente ampliar las formas de participación académica y considerar experiencias donde las preguntas planteadas demanden deliberación y atención a circunstancias particulares.

La diversidad de lenguajes mediante los cuales se expresa el conocimiento también plantea nuevas posibilidades para la

evaluación. La escritura mantiene su relevancia, aunque comparte espacio con representaciones visuales y otras formas de comunicación académica. Zhuang et al. (2025) investigan la retroalimentación generativa en actividades de escritura guiadas por imágenes dentro del aprendizaje de lenguas. Su trabajo permite apreciar el interés educativo de relacionar diferentes recursos expresivos y examinar las posibilidades de revisión que ofrecen las tecnologías generativas.

Figura 4

Arquitecturas de evaluación auténtica en la era de la IA



La atención a las diferencias individuales constituye otra dimensión de las transformaciones evaluativas contemporáneas. Los estudiantes avanzan mediante experiencias diversas y manifiestan necesidades que difícilmente pueden comprenderse mediante procedimientos uniformes. Yang et al. (2022) estudian un sistema de evaluación formativa adaptativa que relaciona pruebas informatizadas con ciclos de memoria del aprendizaje. Su investigación aporta elementos para considerar arquitecturas capaces de responder a trayectorias particulares y mantener criterios académicos compartidos durante los procesos de formación.

La preparación para el ejercicio profesional exige considerar situaciones donde los conocimientos deben traducirse

en actuaciones responsables. Las instituciones educativas necesitan reconocer capacidades comunicativas y criterios de intervención que adquieren significado durante la práctica. Pentland et al. (2026) examinan las evaluaciones orales asincrónicas en relación con la participación y la integridad académica. Tal aproximación amplía las posibilidades de observar desempeños mediante registros comunicativos y permite reconsiderar la relación entre las actividades evaluativas y las responsabilidades profesionales.

La incorporación de inteligencia artificial en los procedimientos de evaluación también modifica las responsabilidades institucionales. Los sistemas automatizados pueden contribuir al tratamiento de información académica, aunque sus resultados requieren interpretación y supervisión. Ahangama (2026) desarrolla un marco de evaluación adaptado a la educación superior en tecnologías de información y comunicación ante la inteligencia artificial generativa. Su propuesta permite examinar la necesidad de establecer criterios de diseño vinculados con las particularidades disciplinares y las decisiones pedagógicas.

En medio de tales transformaciones, el papel docente conserva una responsabilidad intelectual y ética que merece atención. Evaluar implica interpretar evidencias y reconocer circunstancias que pueden escapar a los procedimientos automatizados. La experiencia profesional permite advertir matices relevantes y fundamentar decisiones cuando aparecen resultados contradictorios. Por ello, cualquier arquitectura evaluativa necesita establecer relaciones comprensibles entre los recursos tecnológicos y el juicio académico. La confianza institucional depende de que las decisiones permanezcan abiertas a revisión y explicación.

La renovación evaluativa también plantea interrogantes sobre equidad y participación. Las diferencias de acceso tecnológico pueden condicionar las oportunidades de demostrar aprendizajes, especialmente cuando las actividades requieren recursos digitales especializados. Conviene reconocer que la innovación pedagógica

demanda condiciones institucionales capaces de sostener su aplicación responsable. La claridad de los criterios y la accesibilidad de los procedimientos adquieren relevancia para evitar desigualdades. Una evaluación académicamente exigente necesita ofrecer oportunidades razonables de participación sin desconocer las circunstancias de cada estudiante.

Desde tal horizonte, el capítulo examina arquitecturas de evaluación auténtica capaces de responder a las condiciones educativas creadas por la inteligencia artificial. La reflexión avanza desde el diseño de tareas intelectualmente exigentes hacia problemas abiertos y producciones multimodales. Posteriormente aborda mecanismos adaptativos y experiencias de desempeño profesional antes de considerar la cooperación entre docentes y sistemas automatizados. El propósito es comprender las relaciones pedagógicas que permiten construir evaluaciones pertinentes y preservar la responsabilidad académica ante nuevas posibilidades tecnológicas.

2.1. Tareas auténticas resistentes a la sustitución cognitiva

La capacidad de la inteligencia artificial para producir respuestas académicamente aceptables obliga a reconsiderar la naturaleza de las tareas evaluativas. Una actividad pierde valor formativo cuando el estudiante puede delegar íntegramente su resolución sin comprender las decisiones implicadas. Frente a ello, las tareas auténticas resistentes a la sustitución cognitiva requieren actuaciones intelectuales vinculadas con situaciones particulares. Su diseño demanda identificar aquellos procesos de interpretación y juicio cuya realización permita reconocer una participación genuina del estudiante durante el aprendizaje.

La resistencia a la sustitución cognitiva depende de la relación entre las exigencias de una actividad y las capacidades que pretende desarrollar. Peters y Angelov (2025) plantean la necesidad

de redefinir las tareas evaluativas para favorecer la creatividad y la integridad académica ante la inteligencia artificial generativa. Desde tal perspectiva, una tarea adquiere mayor consistencia cuando exige formular criterios propios, examinar alternativas y justificar decisiones que difícilmente pueden resolverse mediante respuestas genéricas previamente elaboradas por sistemas automatizados.

Una tarea auténtica comienza con una situación suficientemente abierta para admitir distintas interpretaciones, aunque delimitada por condiciones verificables. El estudiante debe reconocer información relevante, distinguir restricciones y establecer prioridades antes de proponer una respuesta. Por ejemplo, analizar un problema institucional concreto exige comprender las relaciones entre sus participantes y las consecuencias de cada decisión. La calidad del desempeño depende entonces de la pertinencia del razonamiento empleado y de su correspondencia con las particularidades del caso examinado.

La singularidad de las situaciones constituye un recurso pedagógico para reducir la delegación irreflexiva del pensamiento. Cuando una actividad incorpora observaciones realizadas por el estudiante, testimonios obtenidos directamente o datos procedentes de experiencias específicas, aumenta la necesidad de interpretar información situada. Sin embargo, incorporar elementos particulares carece de sentido si permanecen ajenos al problema central. Su valor reside en exigir conexiones razonadas entre evidencias disponibles, conocimientos disciplinares y decisiones justificadas durante la elaboración de una respuesta académica.

Otro componente relevante consiste en introducir decisiones cuyo fundamento deba permanecer explícito. Peters y Angelov (2025) vinculan la redefinición de las tareas académicas con oportunidades para ejercer la creatividad y preservar la integridad del trabajo estudiantil. Bajo tal orientación, resulta

pertinente solicitar propuestas acompañadas de explicaciones sobre las alternativas descartadas. La atención se desplaza hacia las razones que sostienen cada elección y permite examinar la profundidad del pensamiento sin depender exclusivamente de la presentación del producto elaborado.

La resistencia cognitiva también puede fortalecerse mediante restricciones que obliguen a adaptar conocimientos a condiciones cambiantes. Una solución inicialmente razonable podría perder pertinencia ante nuevas evidencias, recursos limitados o necesidades imprevistas. Pedir al estudiante que revise sus decisiones permite observar su flexibilidad intelectual y su capacidad para reconocer errores. Conviene que tales modificaciones respondan a situaciones plausibles. Las dificultades artificiales, introducidas para complicar una actividad, pueden desviar la atención de los aprendizajes verdaderamente relevantes.

La dimensión creativa adquiere particular importancia cuando las tareas admiten respuestas diferentes sin renunciar al rigor disciplinar. Peters y Angelov (2025) sitúan la creatividad entre los propósitos que deben orientar la renovación de las actividades evaluativas. Ello permite valorar propuestas originales mediante criterios relacionados con su viabilidad, pertinencia y fundamentación. Una producción novedosa carece de mérito académico cuando desconoce las condiciones del problema. La originalidad cobra significado cuando amplía las posibilidades de comprensión y actuación responsable.

Las tareas resistentes requieren también una relación reconocible entre la actividad académica y las prácticas propias de cada campo profesional. En educación, por ejemplo, diseñar una intervención exige interpretar necesidades de aprendizaje y anticipar respuestas posibles. En ingeniería, resolver un problema demanda ponderar limitaciones técnicas y consecuencias operativas. Tales exigencias permiten apreciar conocimientos aplicados mediante decisiones disciplinariamente justificadas. La

autenticidad depende de que el trabajo reproduzca responsabilidades intelectuales significativas y no meras apariencias de desempeño profesional.

La participación de herramientas generativas plantea una distinción necesaria entre recibir asistencia y transferir el juicio académico. Una tarea bien concebida puede admitir recursos automatizados para organizar información o contrastar alternativas, siempre que el estudiante conserve responsabilidad sobre sus decisiones. Peters y Angelov (2025) relacionan la transformación evaluativa con la protección de la integridad académica. En consecuencia, importa establecer exigencias que hagan reconocible la contribución intelectual personal y permitan examinar críticamente las aportaciones producidas mediante inteligencia artificial.

Diseñar tareas resistentes a la sustitución cognitiva exige comprender que ninguna actividad permanece inmune a las transformaciones tecnológicas. Su fortaleza reside en requerir operaciones intelectuales pertinentes y evidencias de participación que permitan apreciar decisiones propias. La autenticidad se construye mediante problemas significativos cuya resolución compromete conocimientos, experiencias y responsabilidades concretas. Desde una perspectiva pedagógica, el propósito consiste en preservar oportunidades reales para pensar, decidir y aprender, incluso cuando las herramientas disponibles facilitan considerablemente la producción de respuestas.

2.2. Evaluación basada en problemas abiertos y contextos situados

La evaluación basada en problemas abiertos desplaza la atención hacia situaciones que admiten respuestas diversas y requieren decisiones fundamentadas. Su valor pedagógico reside en la incertidumbre que acompaña al análisis de fenómenos reales. Ante un problema sin solución predeterminada, el estudiante

necesita delimitar sus alcances y reconocer qué información resulta pertinente. La inteligencia artificial puede contribuir al proceso, pero la calidad del aprendizaje depende de la interpretación de las condiciones particulares que orientan cada respuesta.

Los problemas abiertos adquieren sentido cuando sus condiciones proceden de ámbitos reconocibles para quienes aprenden. Una controversia comunitaria, una dificultad organizativa o una necesidad educativa permiten relacionar conocimientos disciplinares con circunstancias concretas. Foung et al. (2024) examinan oportunidades para renovar las prácticas evaluativas mediante herramientas digitales e inteligencia artificial generativa. Tal planteamiento invita a diseñar actividades donde la tecnología amplíe las posibilidades de indagación sin determinar de antemano las preguntas relevantes ni los criterios de actuación.

El carácter situado de una evaluación exige atender las condiciones sociales y materiales que influyen en el problema. Los recursos disponibles, las normas institucionales y las expectativas de los participantes modifican las decisiones posibles. Por ello, una respuesta técnicamente correcta puede resultar poco adecuada para una comunidad determinada. Examinar tales diferencias demanda sensibilidad interpretativa. También obliga a reconocer que aplicar conocimientos implica negociar tensiones entre principios generales y necesidades concretas que no siempre coinciden plenamente.

Formular un problema abierto requiere precisión pedagógica. La amplitud excesiva puede producir desconcierto, mientras que una consigna demasiado cerrada reduce las oportunidades de investigación. Conviene establecer una situación inicial comprensible y ofrecer información suficiente para comenzar el análisis. Después, corresponde al estudiante identificar vacíos y formular preguntas adicionales. El docente conserva la responsabilidad de delimitar los aprendizajes esperados sin

anticipar las conclusiones. Esa combinación permite sostener una indagación exigente y viable para distintos niveles formativos.

En ambientes mediados por inteligencia artificial, la búsqueda de información deja de constituir por sí misma una evidencia suficiente de aprendizaje. Foung et al. (2024) analizan posibilidades de transformación evaluativa asociadas al empleo de ChatGPT y otras herramientas digitales. Desde tal perspectiva, conviene valorar la capacidad para contrastar fuentes y detectar contradicciones. También importa determinar cuándo una respuesta automatizada desconoce circunstancias relevantes. El estudiante demuestra comprensión al revisar críticamente la información antes de incorporarla a sus decisiones.

Los problemas situados pueden desarrollarse mediante escenarios que evolucionan durante la actividad. Una decisión inicial modifica las condiciones posteriores y obliga a reconsiderar prioridades. Por ejemplo, una propuesta de mejora institucional puede enfrentar restricciones presupuestarias o desacuerdos entre sus participantes. El interés evaluativo radica en observar la revisión razonada de las alternativas. No se trata de multiplicar obstáculos arbitrarios. Cada variación debe aportar información relevante para comprender mejor las consecuencias de actuar en circunstancias cambiantes.

La evaluación de respuestas abiertas necesita criterios capaces de reconocer diferencias legítimas entre propuestas. Resulta pertinente examinar la consistencia argumentativa, la utilización de evidencias y la atención prestada a las condiciones del caso. Foung et al. (2024) abordan la renovación de las prácticas de evaluación en entornos donde intervienen herramientas generativas. Esa discusión permite considerar procedimientos que valoren el razonamiento documentado y no dependan de comparar cada producto con una respuesta previamente establecida como modelo.

Una dificultad frecuente aparece cuando los escenarios denominados reales presentan situaciones desprovistas de voces humanas. Los problemas sociales involucran intereses y experiencias que rara vez pueden reducirse a datos cuantificables. Incorporar testimonios, perspectivas divergentes o necesidades expresadas por participantes concretos permite enriquecer la interpretación. El estudiante debe reconocer a quién beneficia una propuesta y quién podría resultar perjudicado. De ese modo, el análisis disciplinar adquiere una dimensión ética vinculada con las consecuencias de las decisiones.

Figura 5

Evaluación mediante problemas abiertos y situaciones reales



El trabajo con problemas abiertos también modifica la intervención docente durante la evaluación. Las preguntas de seguimiento pueden ayudar a precisar argumentos sin conducir al estudiante hacia una respuesta determinada. Es importante distinguir entre acompañar la indagación y resolver las dificultades en su lugar. La retroalimentación adquiere utilidad cuando señala inconsistencias o evidencias insuficientes. Al mismo tiempo, debe preservar espacio para que cada estudiante reformule su

planteamiento y asuma las consecuencias intelectuales de sus elecciones.

La incorporación de problemas abiertos y situaciones particulares permite construir evaluaciones sensibles a la complejidad del conocimiento aplicado. Su diseño exige equilibrio entre libertad de respuesta y exigencia académica. La inteligencia artificial puede facilitar consultas y ofrecer alternativas, aunque ninguna herramienta elimina la necesidad de comprender las circunstancias examinadas. El aprendizaje se vuelve visible cuando el estudiante interpreta información incompleta, fundamenta decisiones y reconoce los límites de sus propuestas ante necesidades concretas de intervención educativa.

2.3. Evaluación multimodal y producción de evidencias complejas

La evaluación multimodal amplía las posibilidades de reconocer aprendizajes mediante evidencias expresadas en distintos lenguajes. Un informe escrito puede convivir con registros sonoros, representaciones visuales o demostraciones prácticas. Cada formato permite observar aspectos diferentes del desempeño y plantea exigencias particulares de interpretación. Su incorporación requiere identificar qué aporta cada medio al propósito evaluativo. La diversidad de soportes adquiere sentido cuando permite comprender procesos intelectuales que una producción exclusivamente textual dejaría parcialmente fuera de observación académica.

La elección de modalidades debe responder a la naturaleza del conocimiento examinado. En determinadas disciplinas, un esquema permite apreciar relaciones que resultarían difíciles de explicar mediante párrafos extensos. En otras, una exposición oral revela decisiones argumentativas y dominio conceptual. Conviene evitar que la variedad formal se convierta en una exigencia ornamental innecesaria. El criterio pedagógico consiste en

seleccionar representaciones capaces de hacer visibles dimensiones pertinentes del aprendizaje y establecer relaciones comprensibles entre las evidencias producidas.

La combinación de imágenes y escritura ofrece un ejemplo concreto de articulación entre modalidades. Zhuang et al. (2025) estudian la retroalimentación generativa aplicada a tareas de escritura guiadas por imágenes en el aprendizaje de lenguas. Su investigación permite considerar el potencial pedagógico de vincular estímulos visuales con producciones verbales. En términos evaluativos, interesa observar la interpretación del material presentado y la capacidad para transformar sus elementos en una elaboración discursiva coherente, pertinente y comunicativamente eficaz.

Una evidencia compleja reúne componentes que permiten examinar diferentes momentos de una misma actuación. El producto elaborado adquiere mayor significado cuando puede relacionarse con bocetos, registros de revisión o explicaciones sobre decisiones expresivas. Tales materiales requieren una organización que facilite comprender sus vínculos internos. Acumular archivos sin una finalidad definida dificulta el juicio evaluativo. La selección razonada de evidencias permite apreciar transformaciones en el trabajo y distinguir aportaciones relevantes de documentos meramente accesorios o repetitivos.

La retroalimentación automatizada introduce posibilidades de revisión que merecen atención dentro de las producciones multimodales. Zhuang et al. (2025) investigan su aplicación en actividades de escritura vinculadas con imágenes. Tal línea de trabajo abre preguntas sobre el aprovechamiento educativo de observaciones generadas por inteligencia artificial. Para evaluar el aprendizaje importa examinar qué modificaciones realiza el estudiante y qué razones ofrece para aceptarlas o rechazarlas. La revisión documentada puede revelar comprensión lingüística y mayor autonomía editorial.

El diseño de evidencias multimodales también exige considerar las condiciones de acceso a los recursos. La disponibilidad de dispositivos, conectividad y programas de edición varía entre estudiantes e instituciones. Una evaluación que ignore tales diferencias corre el riesgo de confundir competencia académica con dominio instrumental. Resulta conveniente ofrecer alternativas equivalentes y establecer requisitos técnicos proporcionados. El formato elegido debe facilitar la expresión del aprendizaje sin introducir barreras ajenas a los propósitos formativos de la actividad.

La interpretación de producciones multimodales demanda criterios atentos a la función comunicativa de cada recurso. La claridad de una narración sonora no se juzga exactamente igual que la organización de una infografía. Sin embargo, ambas pueden examinarse por su precisión conceptual y pertinencia comunicativa. Una rúbrica adecuada distingue aspectos compartidos y cualidades específicas de cada modalidad. Tal diferenciación ayuda a evitar valoraciones dominadas por la apariencia estética o por preferencias personales del propio evaluador.

La participación de inteligencia artificial en la creación de imágenes, textos y sonidos vuelve necesario reconocer las decisiones humanas involucradas. Una producción técnicamente cuidada puede ocultar intervenciones automatizadas difíciles de identificar mediante su apariencia. Por ello, adquieren relevancia las explicaciones sobre selección de materiales y transformación de resultados. El interés académico reside en determinar qué comprensión orientó la composición y qué aportaciones fueron examinadas críticamente durante el trabajo, sin atribuir calidad intelectual al acabado tecnológico.

Las evidencias complejas pueden favorecer intercambios evaluativos entre estudiantes cuando sus componentes resultan accesibles y comprensibles. Escuchar una explicación acompañada

de representaciones visuales permite formular observaciones distintas de las que suscita un texto aislado. La discusión colectiva necesita criterios explícitos para mantener su orientación académica y pedagógica. Además, conviene reconocer que ciertas modalidades requieren tiempos diferentes de interpretación. La comparación responsable atiende al aprendizaje demostrado y respeta las particularidades expresivas de cada producción.

La evaluación multimodal plantea una ampliación del repertorio disponible para reconocer conocimientos y capacidades. Su pertinencia depende de la correspondencia entre medios expresivos, objetivos formativos y procedimientos de valoración. Los avances generativos incrementan las opciones de producción y revisión, pero también exigen examinar con cuidado la autoría intelectual. Una arquitectura evaluativa bien articulada permite relacionar evidencias heterogéneas y comprender mejor el desempeño académico, sin confundir abundancia documental con profundidad efectiva del aprendizaje alcanzado.

2.4. Evaluaciones adaptativas basadas en decisiones del estudiante

La evaluación adaptativa introduce una lógica de respuesta pedagógica ante las decisiones que toma cada estudiante durante una actividad. Su interés reside en observar el aprendizaje mientras se desarrolla y ajustar las exigencias según las evidencias disponibles. A diferencia de una secuencia uniforme, permite que distintas elecciones conduzcan a nuevas preguntas o actuaciones. El diseño requiere conservar propósitos comunes para que las variaciones del itinerario no comprometan la interpretación de los desempeños ni su valoración.

Una arquitectura adaptativa necesita definir qué decisiones estudiantiles tendrán consecuencias dentro de la evaluación. Elegir una estrategia, interpretar un resultado o solicitar información adicional puede activar actividades diferentes. La adaptación debe

responder a criterios pedagógicos explícitos y no a cambios arbitrarios de dificultad. Yang et al. (2022) estudian un sistema de evaluación formativa que articula pruebas adaptativas informatizadas y ciclos de memoria del aprendizaje para apoyar procesos personalizados. Su enfoque aporta fundamentos para organizar ajustes.

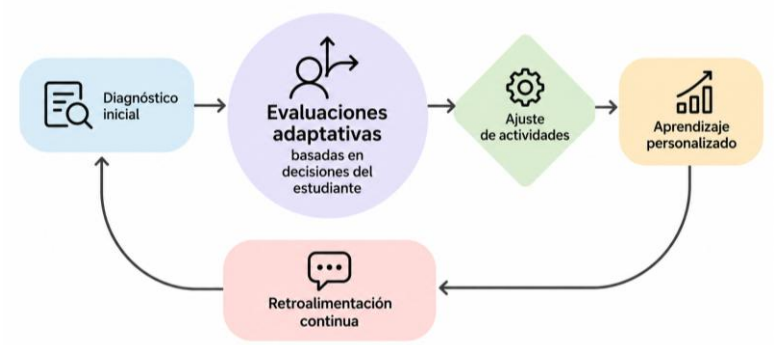
El punto de partida consiste en identificar conocimientos previos mediante actuaciones breves que permitan distinguir necesidades específicas de apoyo. Una respuesta incorrecta puede revelar una confusión conceptual, una interpretación apresurada o una estrategia inadecuada. Por ello, la adaptación exige atender al tipo de error y no únicamente a la puntuación obtenida. Dos estudiantes con resultados equivalentes podrían necesitar intervenciones diferentes. La evaluación gana precisión cuando sus decisiones posteriores responden a evidencias interpretadas con suficiente cuidado.

Las rutas de evaluación pueden organizarse mediante decisiones encadenadas cuya dificultad varía según el desempeño observado. Si el estudiante justifica adecuadamente una elección, recibe una situación que exige transferir el criterio empleado. Cuando aparecen inconsistencias, conviene ofrecer una actividad que permita revisar el fundamento conceptual. Ninguna ruta debería convertirse en una clasificación permanente de capacidades individuales. La adaptación mantiene sentido formativo cuando permite recuperar avances y modificar gradualmente el nivel de exigencia durante el proceso.

Yang et al. (2022) relacionan la evaluación formativa adaptativa con un ciclo de memoria del aprendizaje orientado a la personalización. Tal aproximación permite considerar la importancia de recuperar conocimientos en momentos posteriores y observar su permanencia. Una decisión acertada durante una actividad inmediata no garantiza comprensión duradera. Por ello, pueden incorporarse nuevas oportunidades de recuperación

espaciada que examinen la estabilidad del aprendizaje y permitan ajustar las intervenciones sin depender de una única actuación puntual.

Figura 6
Evaluación adaptativa y personalización del aprendizaje



La inteligencia artificial amplía las posibilidades técnicas para analizar respuestas y proponer variaciones en tiempo oportuno. Sin embargo, una recomendación automatizada puede interpretar de manera insuficiente una explicación ambigua o una solución poco convencional. El docente necesita conocer los criterios utilizados por el sistema y conservar capacidad de intervención. También resulta necesario comunicar al estudiante por qué recibe determinada actividad. La transparencia pedagógica favorece que los ajustes sean comprensibles y puedan discutirse cuando resulte pertinente.

La autonomía estudiantil adquiere un papel particular cuando las decisiones modifican el desarrollo de la evaluación. Elegir entre alternativas exige anticipar consecuencias y reconocer necesidades personales de aprendizaje. El sistema puede ofrecer información sobre el progreso sin imponer continuamente la siguiente acción. Conviene distinguir una elección fundamentada de otra realizada por conveniencia inmediata. Las explicaciones breves sobre los motivos de cada decisión permiten valorar la

autorregulación y comprender mejor las estrategias empleadas durante la actividad.

La personalización plantea dificultades para garantizar comparabilidad entre estudiantes que siguen rutas distintas. Evaluar mediante actividades diferentes requiere establecer referentes compartidos de logro y procedimientos consistentes de interpretación. Yang et al. (2022) aportan un antecedente sobre la articulación de pruebas adaptativas y evaluación formativa personalizada. Desde una perspectiva institucional, resulta necesario documentar las reglas de adaptación y revisar posibles diferencias injustificadas en las oportunidades de demostrar aprendizaje entre quienes participan en una misma evaluación académica.

La información acumulada durante una evaluación adaptativa puede revelar patrones de avance que pasan inadvertidos en una prueba convencional. Los cambios de estrategia, las solicitudes de apoyo y las revisiones sucesivas ofrecen indicios sobre el desarrollo de la comprensión. Sin embargo, registrar cada interacción no garantiza interpretaciones pertinentes. Es necesario seleccionar indicadores relacionados con objetivos académicos y proteger la información personal. Una arquitectura responsable evita convertir el seguimiento pedagógico en vigilancia permanente del comportamiento estudiantil.

La evaluación adaptativa basada en decisiones requiere equilibrar flexibilidad pedagógica y estabilidad de los criterios académicos. Su aporte consiste en permitir que las actuaciones estudiantiles orienten nuevas oportunidades de aprendizaje y produzcan evidencias progresivas de comprensión. La tecnología facilita ciertos ajustes, aunque su pertinencia depende de decisiones educativas justificadas. Cuando las rutas permanecen abiertas a la revisión, la evaluación puede reconocer trayectorias diferentes sin renunciar a exigencias compartidas ni a responsabilidades docentes claramente definidas.

2.5. Evaluación por escenarios, simulaciones y desempeño profesional

La evaluación mediante escenarios y simulaciones permite examinar el desempeño estudiantil ante situaciones próximas al ejercicio profesional. Su interés reside en reproducir responsabilidades y condiciones de actuación que difícilmente aparecen en una prueba convencional. El estudiante debe intervenir dentro de una situación definida, atender sus consecuencias y comunicar decisiones comprensibles. La calidad del diseño depende de la correspondencia entre las exigencias de la simulación y las competencias propias del ámbito laboral representado en cada disciplina.

Un escenario profesional necesita participantes, recursos y reglas que otorguen sentido a las actuaciones esperadas. La descripción inicial debe ofrecer información suficiente para comprender las responsabilidades asignadas sin anticipar cada movimiento. En una simulación clínica, por ejemplo, importa reconocer prioridades asistenciales y comunicarse con precisión. En una negociación institucional adquieren relevancia los compromisos asumidos. La fidelidad pedagógica procede de reproducir decisiones significativas y no de acumular detalles superficiales que distraigan del desempeño profesional esperado.

La representación de papeles profesionales introduce exigencias comunicativas que pueden observarse durante la actuación. Pentland et al. (2026) examinan las evaluaciones orales asincrónicas en relación con la integridad académica, la participación y la comunicación en la era de la inteligencia artificial. Su planteamiento resulta pertinente para valorar explicaciones grabadas vinculadas con escenarios profesionales. Tales registros permiten apreciar la claridad expresiva y la fundamentación de decisiones, siempre que respondan a propósitos evaluativos previamente definidos con claridad.

Las simulaciones pueden incorporar incidentes que alteren las condiciones iniciales del ejercicio profesional. Una demora inesperada, una solicitud contradictoria o una información incompleta obligan a reconsiderar la actuación. Sin embargo, introducir contingencias requiere prudencia y suficiente planificación pedagógica. El propósito consiste en observar respuestas ante dificultades propias de la práctica y no en sorprender gratuitamente al participante. Una secuencia verosímil permite examinar prioridades y responsabilidades sin convertir la incertidumbre en una prueba de resistencia emocional.

El desempeño profesional involucra acciones observables que deben distinguirse de las explicaciones posteriores. Un estudiante puede describir correctamente un procedimiento y encontrar dificultades al ejecutarlo. También puede actuar con eficacia sin expresar con claridad los fundamentos empleados. La evaluación necesita atender ambas dimensiones mediante registros adecuados y verificables. Las demostraciones prácticas ofrecen información sobre ejecución y oportunidad, mientras que las explicaciones permiten examinar criterios. Su articulación evita reducir la competencia profesional a declaraciones de conocimiento.

La modalidad asincrónica ofrece alternativas cuando las condiciones institucionales dificultan reunir simultáneamente a evaluadores y estudiantes. Pentland et al. (2026) abordan las evaluaciones orales asincrónicas como una vía para atender la comunicación y la integridad académica. En una simulación, una respuesta grabada puede documentar instrucciones, recomendaciones o informes profesionales. Su pertinencia depende de establecer límites temporales razonables y condiciones de preparación que permitan interpretar justamente el desempeño registrado por cada participante durante la experiencia evaluativa.

La observación de simulaciones requiere instrumentos capaces de distinguir actuaciones competentes de respuestas

aparentemente convincentes. Una rúbrica puede atender la precisión técnica, el cumplimiento de protocolos y la comunicación con los participantes. También debe considerar la oportunidad de las decisiones cuando el tiempo constituye una condición relevante. Conviene describir conductas reconocibles y evitar categorías vagas. El juicio evaluativo gana consistencia cuando diferentes observadores pueden fundamentar sus apreciaciones mediante evidencias obtenidas durante la actuación profesional simulada.

Las simulaciones mediadas por inteligencia artificial pueden representar interlocutores y generar respuestas variables durante una interacción profesional. Tales posibilidades requieren revisar la coherencia de los personajes simulados y la pertinencia de sus intervenciones. Un comportamiento automatizado inconsistente puede alterar injustamente la dificultad del ejercicio. Además, corresponde establecer qué información reciben los estudiantes y qué recursos pueden consultar. La supervisión docente mantiene la responsabilidad de garantizar condiciones razonables y valorar las actuaciones con criterios transparentes.

La devolución posterior a una simulación permite recuperar aspectos del desempeño que pasan inadvertidos durante la acción. Revisar una grabación o contrastar observaciones facilita reconocer decisiones apresuradas y oportunidades de mejora. Pentland et al. (2026) vinculan las evaluaciones orales asincrónicas con dimensiones de participación y comunicación relevantes para la evaluación contemporánea. La revisión dialogada puede aprovechar tales registros, siempre que la retroalimentación se centre en actuaciones verificables y respete plenamente la dignidad del estudiante evaluado.

La evaluación por escenarios y simulaciones exige reconocer los límites de toda representación del trabajo profesional. Ningún ejercicio reproduce plenamente las responsabilidades ni las consecuencias de una intervención real. Su

valor académico depende de seleccionar situaciones pertinentes y establecer condiciones de observación suficientemente rigurosas. Cuando la simulación permite apreciar acciones, comunicaciones y decisiones responsables, ofrece evidencias relevantes para juzgar la preparación profesional sin confundir una actuación ensayada con experiencia laboral plenamente consolidada y acreditada.

2.6. Ecosistemas híbridos de evaluación humana e inteligencia artificial

Los ecosistemas híbridos de evaluación articulan el juicio docente con capacidades automatizadas para recoger e interpretar evidencias de aprendizaje. Su funcionamiento exige distribuir responsabilidades entre personas y sistemas según la naturaleza de cada decisión. La inteligencia artificial puede identificar patrones o preparar devoluciones preliminares, mientras el profesorado examina su pertinencia educativa. El interés de la arquitectura reside en establecer relaciones de cooperación y evitar que la eficiencia técnica determine por sí misma el valor académico.

Una arquitectura híbrida comienza por definir qué operaciones pueden automatizarse sin comprometer la calidad del juicio evaluativo. Clasificar entregas, detectar omisiones formales o reunir antecedentes son actividades diferentes de valorar originalidad y comprensión. Ahangama (2026) propone un marco integral pertinente de evaluación adaptado a la educación superior en tecnologías de información y comunicación ante la inteligencia artificial generativa. Su enfoque permite considerar la importancia de ajustar el diseño evaluativo a las particularidades disciplinares e institucionales.

El profesorado necesita intervenir en los puntos donde una interpretación automatizada puede alterar consecuencias académicas relevantes. Una calificación desfavorable, por ejemplo, requiere siempre fundamentos plenamente verificables que puedan explicarse y revisarse. Para ello conviene establecer umbrales de

supervisión y procedimientos de contraste entre evaluadores. La intervención humana pierde eficacia cuando se limita a aprobar resultados sin examinarlos. Su responsabilidad implica reconocer errores del sistema y atender circunstancias que los indicadores disponibles no consiguen representar adecuadamente.

La coordinación entre participantes requiere reglas institucionales compartidas claras y explícitas sobre el tránsito de información durante la evaluación. El estudiante aporta evidencias, el sistema procesa determinados registros y el docente interpreta resultados bajo criterios conocidos. Cada intervención debería dejar constancia suficiente para reconstruir decisiones controvertidas. Tal trazabilidad permite identificar discrepancias y corregir procedimientos. También facilita distinguir una observación generada automáticamente de una valoración académica asumida por una persona con responsabilidad institucional sobre sus consecuencias.

La integración tecnológica necesita atender diferencias entre disciplinas y formas de desempeño. Ahangama (2026) aborda el diseño de evaluaciones para la educación superior en tecnologías de información y comunicación mediante un marco disciplinar específico. Esa orientación invita a evitar soluciones uniformes para actividades con exigencias distintas. Una herramienta apropiada para revisar código puede resultar insuficiente al valorar deliberaciones éticas. La selección tecnológica debe responder a las evidencias requeridas y a los límites de cada procedimiento.

El estudiante ocupa una posición activa dentro del ecosistema cuando puede conocer y cuestionar las valoraciones recibidas. La transparencia exige comunicar qué intervenciones corresponden a sistemas automatizados y cuáles proceden del profesorado. También requiere mecanismos plenamente accesibles para solicitar aclaraciones o corregir información equivocada. La posibilidad de apelación no debería depender de conocimientos técnicos especializados. Una relación evaluativa legítima necesita

procedimientos comprensibles que permitan discutir resultados sin convertir cada desacuerdo en una controversia administrativa prolongada.

El tratamiento de datos representa una responsabilidad compartida que necesita límites institucionales muy precisos. Los registros de actividad pueden contener información personal y revelar hábitos que exceden los objetivos académicos. Antes de incorporarlos conviene justificar su necesidad y definir plazos de conservación. También importa restringir accesos y examinar los riesgos de reutilización. La disponibilidad técnica de información no autoriza automáticamente su empleo evaluativo, especialmente cuando podría afectar la privacidad o la confianza de los participantes.

La formación docente adquiere relevancia cuando las herramientas automatizadas intervienen en decisiones de evaluación. Comprender sus márgenes de error permite interpretar recomendaciones sin atribuirles una autoridad indebida. Ahangama (2026) plantea un marco de diseño evaluativo frente a las transformaciones introducidas por la inteligencia artificial generativa. Desde esa perspectiva, la preparación profesional de los docentes debe abarcar también criterios pedagógicos y capacidades de revisión técnica. Ambas dimensiones favorecen decisiones fundamentadas ante resultados contradictorios o insuficientemente explicados.

Tabla 2

Arquitecturas de evaluación auténtica para el aprendizaje en la era de la inteligencia artificial

Arquitectura de evaluación	**Características y aportaciones pedagógicas**
Tareas auténticas resistentes a la sustitución cognitiva	Plantean actividades que requieren interpretación personal y decisiones fundamentadas. Favorecen la creatividad y la integridad académica mediante evidencias de participación intelectual que permiten reconocer el aprendizaje del estudiante.
Evaluación basada en problemas abiertos y situaciones particulares	Propone problemas vinculados con realidades educativas y profesionales. Promueve el análisis de información y la formulación de alternativas pertinentes. Valora la capacidad para fundamentar decisiones ante condiciones complejas.
Evaluación multimodal y producción de evidencias complejas	Integra producciones escritas y recursos visuales junto con registros orales y demostraciones prácticas. Amplía las posibilidades de reconocer conocimientos mediante diferentes lenguajes y permite examinar la coherencia entre las evidencias presentadas.
Evaluaciones adaptativas basadas en decisiones del estudiante	Ajustan las actividades según las respuestas y necesidades de aprendizaje. Facilitan trayectorias formativas diferenciadas mediante retroalimentación y recuperación progresiva de conocimientos. Mantienen referentes académicos compartidos.
Evaluación por escenarios, simulaciones y desempeño profesional	Reproduce situaciones vinculadas con responsabilidades profesionales. Permite observar actuaciones y capacidades comunicativas ante condiciones verosímiles. Favorece la valoración de competencias aplicadas mediante criterios de desempeño verificables.
Ecosistemas híbridos de evaluación humana e inteligencia artificial	Articulan herramientas automatizadas con la interpretación y supervisión docente. Facilitan el tratamiento de evidencias y la retroalimentación académica. Requieren transparencia y protección de datos para garantizar decisiones evaluativas justificables.

Nota. Elaboración propia a partir de los enfoques de Peters y Angelov (2025), Foung et al. (2024), Zhuang et al. (2025), Yang et al. (2022), Pentland et al. (2026) y Ahangama (2026).

Los ecosistemas híbridos requieren mecanismos periódicos de revisión para detectar efectos no previstos. Conviene comparar cuidadosamente valoraciones humanas y automatizadas, examinar desacuerdos recurrentes y escuchar también las experiencias estudiantiles. Una coincidencia elevada entre resultados tampoco demuestra por sí misma que el procedimiento sea justo. Puede reflejar criterios compartidos pero deficientes. La revisión institucional necesita considerar la calidad de las evidencias y las consecuencias educativas de los ajustes antes de ampliar el uso de una herramienta.

La sostenibilidad de una arquitectura híbrida depende de acuerdos institucionales sobre responsabilidades, recursos y procedimientos de mejora. Incorporar inteligencia artificial exige tiempo para verificar resultados y atender incidencias. Una reducción aparente de trabajo puede trasladar cargas hacia tareas de supervisión poco visibles. Por ello, la evaluación del sistema debe considerar sus costos pedagógicos y organizativos. La cooperación entre docentes y tecnologías adquiere sentido cuando mantiene decisiones justificables y protege la confianza académica de quienes participan.

Capítulo III

Evidencias del pensamiento: evaluar procesos y no solamente productos

La evaluación del aprendizaje atraviesa una transformación que obliga a reconsiderar aquello que verdaderamente merece ser observado. Durante mucho tiempo, las producciones académicas han ocupado un lugar privilegiado como manifestaciones del conocimiento adquirido. Sin embargo, la inteligencia artificial generativa ha modificado las condiciones bajo las cuales se elaboran respuestas y soluciones. Un trabajo bien presentado puede ofrecer información limitada sobre la comprensión de quien lo entrega. Ante tal situación, adquiere relevancia examinar las manifestaciones del pensamiento que acompañan la actividad intelectual.

La dificultad no reside únicamente en determinar quién produjo determinado contenido. Existe una cuestión pedagógica más profunda relacionada con la capacidad de reconocer el aprendizaje cuando intervienen recursos capaces de ejecutar operaciones intelectuales complejas. Coates et al. (2025) examinan la integridad académica desde la necesidad de preservar la autenticidad del pensamiento de orden superior. Su planteamiento invita a revisar los criterios mediante los cuales las instituciones educativas atribuyen valor a los desempeños y reconocen responsabilidades intelectuales.

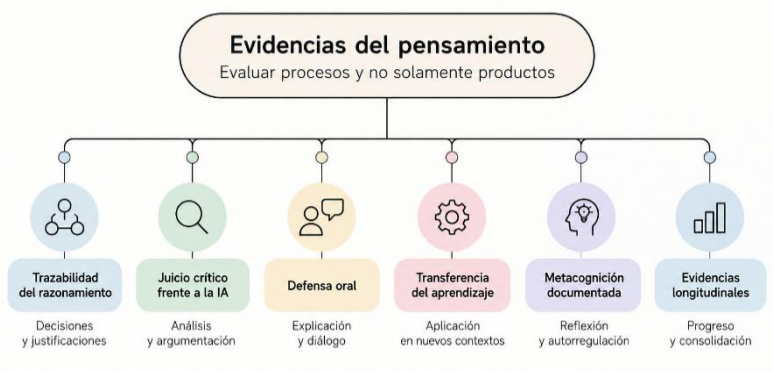
Examinar el pensamiento durante su desarrollo requiere aceptar que el aprendizaje deja manifestaciones diversas y frecuentemente incompletas. Las decisiones adoptadas, las interpretaciones revisadas y las dificultades reconocidas pueden aportar información que permanece ausente en una producción terminada. Ninguna evidencia aislada permite representar toda la complejidad de una experiencia formativa. Por ello, la evaluación necesita articular observaciones que permitan comprender las relaciones entre lo que el estudiante produce y aquello que efectivamente logra comprender.

La presencia de inteligencia artificial en las actividades académicas también modifica las formas de participación

intelectual. El estudiante puede recibir explicaciones elaboradas y propuestas que amplían sus posibilidades de trabajo. Sin embargo, aprovechar tales recursos exige reconocer la procedencia de las ideas y mantener criterios propios de valoración. Katsampoxaki-Hodgetts (2026) aborda la visibilidad epistémica en experiencias educativas mediadas por inteligencia artificial. Su perspectiva permite considerar la necesidad de hacer reconocibles las contribuciones humanas durante la construcción del conocimiento.

Figura 7

Dimensiones de evaluación del pensamiento



La evaluación adquiere una responsabilidad particular cuando debe distinguir entre la aceptación de información convincente y el ejercicio de un criterio fundamentado. La fluidez de una explicación generada puede despertar confianza incluso cuando presenta debilidades conceptuales. Oates y Johnson (2025) examinan el papel de ChatGPT en actividades orientadas a la evaluación crítica de contenidos académicos. Su investigación permite reconocer el interés pedagógico de valorar la capacidad para examinar afirmaciones y establecer juicios sustentados mediante conocimientos disciplinares.

Ahora bien, comprender una producción académica también exige considerar las posibilidades de interacción entre

estudiantes y docentes. El intercambio de preguntas y explicaciones permite reconocer matices que difícilmente aparecen en un documento escrito. Una conversación puede revelar seguridad conceptual o dificultades todavía pendientes de resolución. Su valor depende de condiciones respetuosas y criterios académicos transparentes. La evaluación necesita evitar que las diferencias expresivas oculten conocimientos relevantes o atribuyan mayor competencia a quienes poseen facilidad comunicativa.

La comprensión también se manifiesta cuando los conocimientos conservan utilidad ante circunstancias distintas de aquellas en que fueron aprendidos. Las situaciones educativas pueden incorporar restricciones imprevistas y condiciones que obligan a reconsiderar decisiones anteriores. König et al. (2023) examinan la transferencia de herramientas digitales en la educación superior y destacan diferentes aproximaciones para abordar su aplicación. Desde el ámbito evaluativo, dicha preocupación permite reconocer la importancia de observar aprendizajes capaces de responder a exigencias cambiantes.

Otra dimensión relevante corresponde a la conciencia que desarrolla cada estudiante sobre sus capacidades y necesidades formativas. Aprender implica reconocer incertidumbres y advertir cuándo una estrategia deja de resultar adecuada. Combrinck y Loubser (2025) estudian la autorreflexión como recurso para gestionar el uso de inteligencia artificial generativa en evaluaciones de grupos numerosos. Su aportación permite considerar el valor de las experiencias reflexivas cuando ayudan a interpretar decisiones personales y favorecen una participación académica más consciente.

El tiempo constituye igualmente una dimensión necesaria para comprender el desarrollo del conocimiento. Los aprendizajes pueden mostrar avances irregulares y transformaciones que una evaluación puntual difícilmente permite reconocer. Tempelaar et al. (2024) vinculan la analítica disposicional del aprendizaje con la

evaluación formativa y destacan la relación entre ambas perspectivas. Tal aproximación permite considerar el valor de las evidencias acumuladas durante periodos prolongados, siempre que su interpretación responda a propósitos pedagógicos definidos y respete las particularidades individuales.

Desde tal perspectiva, evaluar el pensamiento requiere una mirada capaz de reconocer manifestaciones complementarias de comprensión sin convertirlas en procedimientos de vigilancia académica. La responsabilidad docente consiste en interpretar evidencias mediante criterios explícitos y atender a las condiciones que influyen en cada desempeño. El presente capítulo aborda diferentes posibilidades para ampliar la comprensión evaluativa del aprendizaje mediado por inteligencia artificial. Su orientación responde a una preocupación educativa persistente: reconocer con justicia aquello que cada estudiante aprende.

3.1. Trazabilidad del razonamiento y construcción de respuestas

La trazabilidad del razonamiento permite examinar el camino intelectual mediante el cual un estudiante construye una respuesta. Su valor educativo reside en hacer visibles las decisiones que intervienen entre la comprensión inicial de un problema y la formulación de una solución. Cuando la inteligencia artificial participa en las actividades académicas, conocer dicho proceso adquiere especial relevancia. Una respuesta correcta puede ocultar interpretaciones débiles, decisiones poco fundamentadas o una dependencia tecnológica que limita la autonomía intelectual del estudiante durante su aprendizaje.

Reconstruir el razonamiento exige prestar atención a las operaciones intelectuales que sostienen cada decisión. El estudiante identifica información pertinente, establece relaciones y contrasta posibles explicaciones antes de adoptar una postura. Tales operaciones rara vez siguen una secuencia perfectamente

ordenada. Existen retrocesos, dudas y reformulaciones que revelan la complejidad del aprendizaje. Desde una perspectiva evaluativa, reconocer dichas variaciones permite comprender el desarrollo de las ideas sin reducir la calidad del desempeño a una sucesión rígida de pasos previamente establecidos.

La presencia de sistemas generativos introduce nuevas dificultades para reconocer la procedencia de las ideas. Una explicación bien redactada puede combinar conocimientos previos, aportaciones automatizadas y decisiones personales difíciles de distinguir. Katsampoxaki-Hodgetts (2026) aborda la visibilidad epistémica mediante experiencias de diseño educativo que integran la interacción humana con inteligencia artificial. Su planteamiento permite considerar la necesidad de identificar las contribuciones del estudiante y las intervenciones tecnológicas durante la elaboración de respuestas académicas, preservando la responsabilidad intelectual sobre el trabajo producido.

La documentación del proceso adquiere sentido cuando permite reconocer decisiones significativas y comprender sus fundamentos. Registrar una pregunta inicial, conservar una hipótesis descartada o explicar la elección de determinada evidencia puede revelar aspectos que desaparecen en el producto entregado. Sin embargo, acumular registros indiscriminadamente tampoco garantiza una evaluación rigurosa. El interés pedagógico recae en la pertinencia de las huellas conservadas y en su capacidad para mostrar transformaciones del pensamiento que puedan examinarse mediante criterios académicos explícitos y compartidos.

Una cuestión particularmente delicada aparece cuando el estudiante acepta una respuesta generada sin examinar sus fundamentos. La fluidez verbal del sistema puede producir una impresión de certeza que desalienta la revisión crítica. Frente a ello, la trazabilidad permite reconocer si hubo contrastación de fuentes, identificación de inconsistencias o reconsideración de

afirmaciones. Importa conocer qué razones llevaron a conservar una explicación y cuáles motivaron su modificación. Ese examen aporta información sobre el juicio académico ejercido durante la interacción con herramientas generativas.

En la investigación de Katsampoxaki-Hodgetts (2026), las interacciones registradas durante el rediseño de clases permiten observar momentos de planificación, revisión y rechazo argumentado de propuestas automatizadas. La autora examina conjuntamente las consultas, las respuestas obtenidas y las justificaciones de quienes participan. Esa articulación ofrece una vía para interpretar decisiones que permanecerían invisibles al revisar únicamente el material elaborado. También muestra que una intervención tecnológica adquiere significado educativo según el criterio empleado para aceptarla o cuestionarla.

Desde el trabajo docente, reconstruir tales decisiones requiere preguntas capaces de abrir el razonamiento sin imponer una explicación retrospectiva artificial. Conviene indagar por qué se eligió determinada estrategia, qué evidencia modificó una interpretación o qué alternativa perdió fuerza durante la resolución. Las respuestas pueden mostrar vacilaciones e incluso contradicciones. Su lectura exige prudencia: una narración convincente sobre el proceso no demuestra necesariamente que cada operación descrita haya ocurrido en el momento señalado por el estudiante.

La trazabilidad también plantea una diferencia entre registrar acciones y comprender su significado. Una secuencia de instrucciones enviadas a un sistema puede documentar actividad abundante sin revelar aprendizaje profundo. Para interpretar su valor resulta necesario vincular cada intervención con un propósito reconocible y con las modificaciones posteriores del trabajo. Katsampoxaki-Hodgetts (2026) destaca el examen de las justificaciones que acompañan la aceptación, adaptación o

exclusión de propuestas generadas durante tareas de diseño pedagógico en formación docente.

El tratamiento de las evidencias demanda criterios proporcionados y respeto por la intimidad intelectual del estudiante. Ninguna actividad educativa necesita convertir cada duda privada en un registro obligatorio. Puede bastar una selección razonada de momentos decisivos que permita discutir el fundamento de las elecciones realizadas. Importa distinguir entre una corrección promovida por indicaciones externas y una revisión nacida del propio juicio. Ambas pueden contribuir al aprendizaje, aunque expresan grados diferentes de iniciativa y apropiación conceptual.

En consecuencia, la construcción de respuestas puede examinarse como una práctica de elaboración cuyo valor depende de las razones que sostienen cada cambio. La evidencia adquiere profundidad cuando permite relacionar preguntas, decisiones y revisiones sin exigir una reconstrucción perfecta del pensamiento. Para el profesorado, esa lectura demanda interpretar indicios y reconocer los límites de cualquier registro. Para el estudiante, ofrece oportunidades de explicar sus elecciones y asumir responsabilidad por aquello que presenta como conocimiento propio.

3.2. Evaluación del juicio crítico frente a respuestas generadas por IA

El juicio crítico adquiere una dimensión particular cuando el estudiante debe valorar respuestas producidas por inteligencia artificial. La facilidad con que tales sistemas elaboran explicaciones coherentes obliga a distinguir entre claridad expresiva y validez del conocimiento. Evaluar dicha capacidad requiere observar la disposición para interrogar afirmaciones aparentemente convincentes. También interesa reconocer si el estudiante identifica aquello que necesita comprobar antes de aceptar una

explicación. La confianza en una respuesta debe construirse mediante razones verificables y suficientemente fundamentadas.

Una dificultad frecuente consiste en atribuir autoridad intelectual a contenidos cuya presentación transmite seguridad. Los sistemas generativos pueden formular argumentos persuasivos incluso cuando presentan imprecisiones o interpretaciones discutibles. El estudiante necesita reconocer que la corrección gramatical carece de valor probatorio por sí misma. Oates y Johnson (2025) encontraron que los textos generados por ChatGPT mostraban buena organización y precisión lingüística sin producir mejoras generales en el rendimiento académico de los participantes de su investigación universitaria realizada.

La evaluación del juicio crítico puede organizarse alrededor de afirmaciones que exijan contrastación disciplinar. Ante una explicación generada, el estudiante debe identificar proposiciones verificables y examinar las evidencias disponibles. El interés académico reside en la calidad de las razones utilizadas para confirmar, matizar o rechazar cada planteamiento. Una objeción fundada demuestra mayor comprensión que una descalificación genérica. Del mismo modo, aceptar una afirmación correctamente sustentada puede expresar discernimiento cuando responde a un examen cuidadoso de su contenido.

El estudio de Oates y Johnson (2025) ofrece una referencia pertinente para comprender el valor evaluativo de tales actividades. Su investigación con estudiantes de posgrado en ciencias biomédicas combinó la producción de ensayos mediante ChatGPT con su posterior evaluación crítica mediante literatura científica revisada por pares. Los resultados mostraron mejores calificaciones en las tareas de valoración crítica. Aunque participaron diez estudiantes, el hallazgo permite considerar alternativas pedagógicas centradas en examinar argumentos generados mediante inteligencia artificial.

El juicio crítico también se manifiesta en la capacidad para reconocer omisiones relevantes. Una respuesta puede presentar información correcta y mantener una interpretación insuficiente del problema. Detectar perspectivas ausentes exige conocer los límites de la explicación ofrecida y comprender las exigencias particulares de cada disciplina. Durante la evaluación conviene atender a la pertinencia de las observaciones realizadas. Identificar una ausencia secundaria tiene menor alcance intelectual que advertir una omisión capaz de modificar sustancialmente la interpretación académica propuesta.

La comparación entre respuestas generadas y fuentes académicas permite examinar la calidad del criterio utilizado por el estudiante. No basta con encontrar diferencias textuales. Resulta necesario determinar si representan errores, discrepancias interpretativas o niveles realmente distintos de profundidad conceptual. El docente puede valorar la precisión de los argumentos y la capacidad para reconocer incertidumbres legítimas. Una evaluación cuidadosa evita premiar la sospecha permanente y reconoce también las decisiones de aceptación respaldadas por evidencia pertinente.

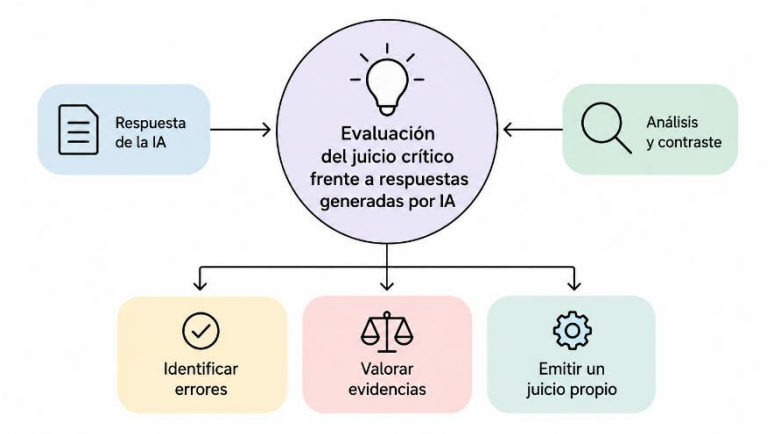
Otro aspecto evaluable corresponde a la independencia del juicio frente a la aparente autoridad tecnológica. Algunos estudiantes modifican sus respuestas cuando el sistema presenta una alternativa formulada con seguridad. Otros mantienen posiciones iniciales sin considerar argumentos nuevos. Ambas conductas merecen examen. La autonomía intelectual requiere apertura para revisar creencias y firmeza para defender conclusiones justificadas. Su valoración puede apoyarse en situaciones donde existan interpretaciones rivales y sea necesario explicar por qué una merece mayor confianza.

Oates y Johnson (2025) observaron que la interacción del estudiante con ChatGPT influía en la utilidad de la herramienta. Los participantes señalaron dificultades relacionadas con referencias y

comprobación de información. Tales observaciones permiten comprender que el juicio crítico exige esfuerzo intelectual incluso cuando la producción textual resulta más rápida. En términos evaluativos, interesa reconocer la calidad de la comprobación realizada y evitar que la velocidad de respuesta se confunda con dominio conceptual o rigor académico.

Figura 8

Valoración crítica de respuestas generadas por IA



Las tareas de valoración crítica necesitan criterios comprensibles para quienes participan. La exactitud factual constituye una dimensión, pero también importan la pertinencia de las pruebas y la consistencia de las inferencias. Una rúbrica puede distinguir observaciones descriptivas de objeciones justificadas y revisiones conceptuales profundas. Conviene evitar que la calificación dependa de la cantidad de errores detectados. Una respuesta generada con pocas fallas puede demandar mayor discernimiento que otra plagada de equivocaciones evidentes en cada caso.

Evaluar el juicio crítico frente a la inteligencia artificial implica reconocer la responsabilidad del estudiante al atribuir validez a una afirmación. El aprendizaje se aprecia cuando puede

defender una interpretación con argumentos propios y admitir límites razonables en sus conclusiones. La intervención docente adquiere valor al examinar la fuerza de tales razones sin imponer acuerdos automáticos. Queda entonces abierta una pregunta pedagógica exigente: qué comprensión demuestra quien decide confiar en determinada respuesta y puede justificarlo.

3.3. La defensa oral como evidencia complementaria de comprensión

La defensa oral ofrece una oportunidad para reconocer la comprensión que sostiene un trabajo académico. Durante el diálogo, el estudiante debe explicar decisiones, precisar conceptos y responder preguntas que no siempre anticipó al preparar su entrega. Su intervención permite apreciar la relación entre las ideas presentadas y el conocimiento que puede movilizar personalmente. La calidad de dicha evidencia depende de preguntas pertinentes y de condiciones que favorezcan una expresión reflexiva sin convertir la conversación en interrogatorio.

Su valor complementario aparece cuando el producto escrito admite distintas interpretaciones sobre el dominio alcanzado. Un informe cuidadosamente elaborado puede expresar comprensión profunda o reproducir formulaciones cuya lógica permanece ajena a quien las presenta. La conversación académica permite indagar las relaciones conceptuales que sustentan el documento. Conviene evitar preguntas destinadas a descubrir una falta mediante sorpresa. Resulta más fecundo solicitar explicaciones de decisiones relevantes y escuchar atentamente las razones antes de formular nuevas preguntas.

Coates et al. (2025) sitúan la autenticidad del pensamiento de orden superior entre las preocupaciones de la integridad académica ante la inteligencia artificial generativa. Desde esa perspectiva, la defensa oral puede contribuir a valorar la autoría intelectual sin reducirla a una declaración de uso tecnológico. La

atención recae en la capacidad de argumentar sobre el trabajo realizado. También exige reconocer que una intervención verbal breve ofrece indicios parciales y requiere una interpretación pedagógica especialmente prudente.

Una defensa bien planteada comienza por preguntas vinculadas con el propósito del trabajo. Pedir al estudiante que explique por qué eligió determinada perspectiva permite examinar su comprensión del problema. Después pueden introducirse variaciones que obliguen a reconsiderar las conclusiones. La conversación gana profundidad cuando las preguntas responden a lo escuchado y no obedecen a un cuestionario invariable. De ese intercambio pueden surgir matices importantes que el texto escrito no alcanzó a expresar con claridad.

La capacidad de responder inmediatamente no equivale necesariamente a comprender mejor. Algunos estudiantes necesitan tiempo para ordenar sus ideas y otros dominan recursos retóricos que disimulan vacíos conceptuales. Por ello, la valoración debe distinguir entre soltura comunicativa y fundamento académico. Permitir breves pausas o reformular una pregunta puede mejorar considerablemente la calidad de la evidencia. El propósito consiste en reconocer el conocimiento realmente disponible sin convertir la rapidez verbal en criterio predominante de rendimiento.

En el marco de la integridad académica, Coates et al. (2025) plantean la necesidad de gobernar institucionalmente la autenticidad del pensamiento superior. La defensa oral puede integrarse en ese esfuerzo cuando existen criterios transparentes y procedimientos proporcionales. Su aplicación requiere explicar previamente qué capacidades serán valoradas. También demanda evitar que el encuentro se utilice como mecanismo informal de sospecha. La confianza educativa se fortalece mediante reglas conocidas y oportunidades razonables para demostrar comprensión del aprendizaje.

Las preguntas de transferencia amplían el alcance de la defensa oral. Presentar una situación distinta permite observar si el estudiante reconoce principios aplicables fuera del ejemplo trabajado. Puede pedírsele que adapte una propuesta ante una restricción nueva o que examine una consecuencia inesperada. La respuesta revela hasta qué punto comprende las relaciones que organizan su planteamiento. No se trata de exigir soluciones perfectas, sino de atender a la pertinencia del razonamiento ante cambios verdaderamente significativos.

La interacción también permite advertir la disposición para reconocer límites en el propio conocimiento. Cuando una pregunta excede lo trabajado, admitir incertidumbre puede expresar madurez académica. Lo relevante es distinguir una respuesta evasiva de una delimitación razonada del problema. El docente puede solicitar que se indiquen datos necesarios para avanzar o criterios para resolver la duda. Tales intervenciones permiten apreciar prudencia intelectual y evitan confundir seguridad expresiva con comprensión verdaderamente fundamentada en conocimientos disciplinares pertinentes.

La incorporación de defensas orales exige considerar diferencias lingüísticas, emocionales y comunicativas entre estudiantes. Una evaluación justa puede ofrecer preguntas accesibles, tiempos proporcionados y alternativas de apoyo compatibles con los objetivos formativos. También conviene registrar criterios y observaciones para reducir decisiones basadas en impresiones personales. El desempeño debe interpretarse junto con otras evidencias disponibles. Ninguna conversación aislada permite representar por completo la amplitud del aprendizaje desarrollado durante una experiencia académica compleja de formación universitaria.

La defensa oral adquiere mayor sentido cuando genera una conversación intelectualmente exigente y respetuosa. Su aporte reside en permitir que el estudiante explique relaciones, examine

objeciones y reconozca consecuencias de sus decisiones. Frente a producciones mediadas por inteligencia artificial, ofrece una ocasión para apreciar la apropiación del conocimiento sin reclamar certeza absoluta sobre su origen. La evaluación conserva entonces una responsabilidad pedagógica esencial: escuchar con atención y juzgar mediante razones explícitas y criterios previamente compartidos.

3.4. Evaluación de la transferencia mediante variaciones del problema

La transferencia del aprendizaje se reconoce cuando un estudiante utiliza conocimientos adquiridos para resolver situaciones que presentan condiciones diferentes. En la evaluación auténtica, modificar un problema permite examinar la flexibilidad de su comprensión y la pertinencia de sus decisiones. La variación introduce exigencias que una respuesta previamente elaborada quizá no resuelva. Su valor reside en observar qué principios conserva el estudiante, cuáles reconsidera y qué argumentos emplea para ajustar su actuación ante nuevas circunstancias académicas.

Diseñar variaciones significativas exige identificar primero los elementos que definen la estructura intelectual del problema. Cambiar nombres, cantidades o escenarios puede producir una apariencia de novedad sin alterar las operaciones necesarias para resolverlo. Una modificación relevante afecta restricciones, relaciones o criterios de decisión. El estudiante necesita reconocer qué permanece vigente y qué requiere una estrategia distinta. La evaluación puede apreciar entonces su capacidad para distinguir rasgos superficiales de condiciones que verdaderamente transforman la tarea propuesta.

König et al. (2023) examinan la transferencia de herramientas digitales de enseñanza, aprendizaje y evaluación en la educación superior desde diferentes enfoques y prácticas. Su

contribución permite situar la transferencia dentro de procesos educativos que requieren adaptación a condiciones particulares. En la valoración del aprendizaje, esa preocupación adquiere otra escala: interesa determinar si los conocimientos desarrollados conservan utilidad cuando cambian las demandas. La semejanza entre situaciones puede facilitar el desempeño sin garantizar comprensión suficientemente flexible.

Una estrategia consiste en presentar inicialmente un problema familiar y modificar después una condición determinante. En una actividad de planificación, por ejemplo, puede reducirse considerablemente el presupuesto disponible o introducirse una necesidad imprevista. El estudiante debe revisar prioridades y explicar las consecuencias de sus ajustes. La atención evaluativa se dirige hacia la consistencia de las decisiones tomadas. Mantener una solución anterior sin reconsiderar sus fundamentos puede revelar dependencia de procedimientos aprendidos de manera mecánica.

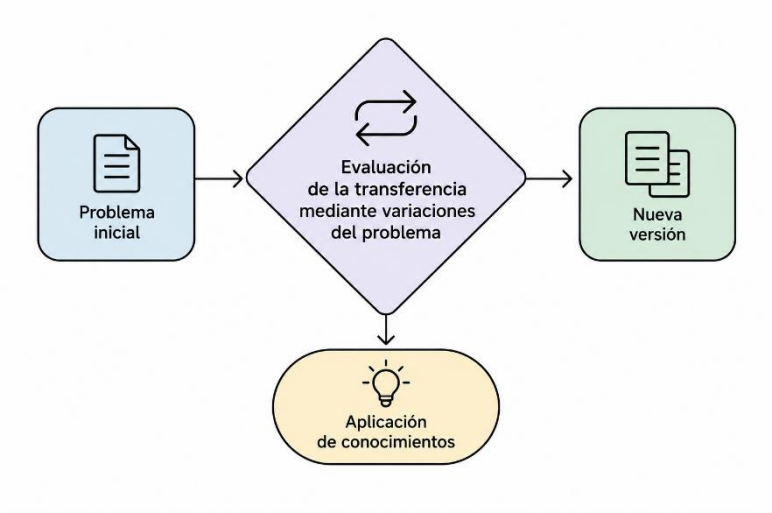
Las variaciones también pueden alterar la información disponible. Un problema resuelto con datos completos adquiere otra dificultad cuando aparecen evidencias contradictorias o faltan antecedentes relevantes para decidir. Ante tal situación, el estudiante necesita reconocer incertidumbres y determinar qué información adicional permitiría avanzar. La evaluación puede distinguir entre decisiones justificadas y conjeturas presentadas como certezas. Importa valorar la prudencia con que se delimitan las conclusiones y la pertinencia de las preguntas formuladas durante el análisis académico.

La inteligencia artificial generativa puede resolver numerosas versiones de una misma tarea con notable fluidez verbal. Por ello, presentar variaciones no garantiza por sí mismo una evidencia de transferencia personal. Es necesario examinar las decisiones del estudiante frente a cambios que exijan seleccionar conocimientos pertinentes y justificar adaptaciones. El interés

recae en la correspondencia entre las nuevas condiciones y las modificaciones propuestas. Una respuesta diferente carece de valor si no responde realmente al problema transformado.

Figura 9

Transferencia del aprendizaje ante problemas variables



La distancia entre la situación aprendida y la nueva tarea merece atención especial. Cuando ambas comparten una estructura evidente, la adaptación puede requerir ajustes menores. En cambio, trasladar un principio a un ámbito poco familiar demanda reconocer relaciones menos visibles. Conviene graduar esa distancia según los objetivos formativos y la experiencia del grupo. Una variación excesivamente ajena al aprendizaje previo puede medir conocimientos adicionales antes que la capacidad de transferir aquello efectivamente enseñado en clase.

König et al. (2023) abordan, desde el ámbito de la educación superior, diferentes aproximaciones vinculadas con la transferencia de herramientas digitales. Tal perspectiva permite reconocer que las condiciones de aplicación importan para interpretar los resultados. En las tareas evaluativas, conviene

considerar recursos disponibles, restricciones institucionales y características del escenario planteado. Una solución pertinente en determinada situación puede resultar inadecuada en otra. Reconocer esa dependencia constituye una manifestación valiosa de comprensión disciplinar y sensibilidad práctica.

La valoración de las respuestas requiere criterios que permitan apreciar adaptaciones justificadas sin exigir uniformidad. Dos estudiantes pueden proponer soluciones diferentes y demostrar un dominio comparable si ambos reconocen las restricciones relevantes. Las rúbricas deberían atender a la identificación de cambios, la selección de principios y la coherencia de las decisiones. También importa distinguir errores conceptuales de elecciones razonables ante información incompleta. Esa distinción evita penalizar respuestas originales por apartarse de modelos previamente establecidos.

Evaluar la transferencia mediante variaciones del problema permite reconocer la capacidad de utilizar conocimientos con discernimiento ante circunstancias cambiantes. Su aplicación demanda tareas cuidadosamente diseñadas y criterios atentos a las relaciones que el estudiante logra establecer. La evidencia más significativa aparece cuando una modificación inesperada provoca una reconsideración pertinente de la estrategia. En una educación atravesada por herramientas generativas, esa flexibilidad razonada ayuda a valorar aprendizajes que mantienen sentido más allá de una respuesta determinada.

3.5. Metacognición documentada y conciencia del propio aprendizaje

La metacognición documentada permite reconocer la manera en que el estudiante interpreta su propio aprendizaje y regula sus decisiones. Su interés evaluativo reside en la capacidad de advertir avances, dificultades y cambios de criterio mientras desarrolla una tarea. Cuando intervienen herramientas generativas,

dicha conciencia adquiere especial relevancia porque las respuestas disponibles pueden acelerar el trabajo sin aclarar qué se ha comprendido. Documentar esa experiencia exige expresar razones personales y reconocer los límites del conocimiento alcanzado.

Un registro reflexivo adquiere valor cuando supera la descripción de actividades realizadas. Anotar que se consultó una herramienta o que se corrigió un texto informa poco sobre el aprendizaje. Resulta más revelador explicar qué dificultad motivó la consulta, qué expectativa existía y qué cambió después. La escritura metacognitiva permite observar la interpretación que cada estudiante hace de sus decisiones. Su calidad depende de la precisión de esa interpretación y de su relación con experiencias concretas.

Combrinck y Loubser (2025) examinan la autorreflexión estudiantil como recurso para gestionar el uso de inteligencia artificial generativa en evaluaciones de grupos numerosos. Su trabajo aporta una perspectiva pertinente para considerar el registro reflexivo dentro de prácticas evaluativas habituales. En lugar de tratar cada declaración como prueba concluyente, conviene analizar qué permite conocer acerca de la participación del estudiante. La reflexión escrita abre una vía de indagación pedagógica que necesita criterios claros para ser interpretada.

La conciencia del propio aprendizaje también requiere distinguir entre familiaridad y comprensión. Después de leer una explicación convincente, el estudiante puede experimentar una sensación de dominio que desaparece cuando intenta aplicar las ideas sin ayuda. Un registro metacognitivo puede recoger esa diferencia mediante preguntas sobre aquello que logró explicar autónomamente y aquello que todavía necesita revisar. Reconocer una dificultad concreta tiene mayor utilidad formativa que afirmar de manera general que una actividad resultó plenamente satisfactoria.

El momento en que se documenta la reflexión modifica la información disponible. Una anotación realizada durante la tarea puede conservar incertidumbres que después se olvidan. La revisión posterior permite interpretar decisiones con mayor distancia, aunque también favorece reconstrucciones demasiado ordenadas. Ambas modalidades ofrecen perspectivas diferentes sobre la experiencia. Su combinación puede enriquecer la evaluación cuando responde a propósitos definidos y evita transformar la reflexión en una obligación repetitiva que interrumpe constantemente el trabajo académico.

En grupos numerosos, la reflexión documentada plantea una cuestión práctica: obtener información significativa sin exigir una lectura imposible de sostener. Combrinck y Loubser (2025) sitúan precisamente la autorreflexión dentro de la gestión evaluativa del uso de inteligencia artificial generativa en clases amplias. Esa orientación invita a diseñar consignas breves y específicas. La extensión del registro importa menos que su capacidad para mostrar decisiones reconocibles y ofrecer al docente indicios útiles sobre necesidades de acompañamiento pedagógico.

La sinceridad constituye una condición delicada de cualquier práctica metacognitiva. Si el estudiante percibe que admitir confusión perjudicará su calificación, puede redactar una historia de progreso continuo. Por ello, conviene separar el reconocimiento de dificultades de los juicios punitivos sobre el desempeño. Una reflexión valiosa puede describir errores persistentes o estrategias que fracasaron. Evaluar su calidad exige atender a la profundidad del examen personal y evitar premiar relatos artificialmente optimistas sobre el aprendizaje alcanzado.

La documentación también permite observar cambios en las creencias del estudiante sobre sus propias capacidades. Al principio puede atribuir sus resultados a la facilidad de una herramienta o a la dificultad del contenido. Con el tiempo, quizá identifique decisiones que dependen de su preparación y esfuerzo.

Tales desplazamientos merecen atención cuando aparecen vinculados con experiencias verificables. El propósito pedagógico consiste en favorecer explicaciones más precisas sobre el aprendizaje sin exigir confesiones personales innecesarias durante la evaluación.

Una práctica metacognitiva bien orientada incorpora la planificación de acciones posteriores. Reconocer una dificultad resulta insuficiente cuando no se examinan alternativas para abordarla. El estudiante puede identificar qué necesita estudiar, qué procedimiento debería modificar o qué ayuda convendría solicitar. La evaluación atiende entonces a la correspondencia entre el diagnóstico personal y las medidas elegidas. También puede revisar si dichas medidas produjeron cambios observables, evitando confundir la formulación de buenas intenciones con regulación efectiva del aprendizaje.

La metacognición documentada ofrece una mirada particular sobre la relación entre aprendizaje, autonomía y responsabilidad intelectual. Su aportación aparece cuando el estudiante logra interpretar sus decisiones sin ocultar incertidumbres ni atribuir automáticamente sus avances a recursos externos. Para el docente, tales registros constituyen evidencias que deben leerse junto con el desempeño observado. Su valor no depende de presentar una trayectoria impecable, sino de mostrar una conciencia cada vez más precisa del trabajo intelectual propio.

3.6. Evidencias longitudinales y portafolios de progresión cognitiva

Las evidencias longitudinales permiten examinar el aprendizaje como una transformación observable durante periodos prolongados. Frente a producciones aisladas, ofrecen información sobre la evolución de las interpretaciones y la estabilidad de los conocimientos adquiridos. Un portafolio de progresión cognitiva reúne muestras seleccionadas que permiten reconocer diferencias

entre momentos formativos. Su utilidad depende de establecer relaciones significativas entre los trabajos conservados y los propósitos educativos. La continuidad temporal aporta profundidad cuando revela cambios que merecen explicación académica.

La selección de evidencias requiere atender a momentos representativos del desarrollo intelectual. Conservar todas las actividades puede dificultar la lectura del progreso y convertir el portafolio en un archivo acumulativo. Conviene elegir producciones que permitan apreciar cambios en la complejidad de los argumentos o en el manejo de conocimientos disciplinares. Una primera aproximación incompleta puede adquirir especial valor al compararse con trabajos posteriores. La diferencia necesita interpretarse mediante criterios compartidos y no mediante impresiones generales.

Tempelaar et al. (2024) relacionan la analítica del aprendizaje disposicional con la evaluación formativa y destacan su estrecha vinculación. Su planteamiento invita a considerar que los datos sobre el aprendizaje adquieren mayor utilidad cuando orientan decisiones pedagógicas. En una perspectiva longitudinal, los registros de desempeño pueden dialogar con información sobre las disposiciones del estudiante. Sin embargo, cualquier interpretación requiere especial cautela para evitar atribuir causas personales a cambios observados sin evidencias suficientes que los respalden.

La progresión cognitiva rara vez mantiene un ritmo uniforme. Un estudiante puede avanzar rápidamente en determinados conceptos y experimentar dificultades persistentes en otros. También aparecen periodos de aparente estancamiento que preceden a reorganizaciones importantes del conocimiento. El portafolio permite reconocer tales irregularidades sin convertirlas automáticamente en deficiencias. Su lectura exige examinar la naturaleza de cada tarea y las condiciones en que fue realizada.

Comparar desempeños heterogéneos requiere criterios que respeten sus diferencias de exigencia intelectual.

Cuando interviene inteligencia artificial generativa, la comparación temporal necesita considerar los recursos utilizados en cada actividad. Una producción posterior puede presentar mayor sofisticación verbal sin evidenciar un desarrollo equivalente de las capacidades personales. El portafolio permite contrastar trabajos realizados bajo condiciones distintas y examinar la permanencia de determinados aprendizajes. Importa establecer qué conocimientos logra movilizar el estudiante en situaciones comparables. La calidad del progreso no debería confundirse con mejoras atribuibles principalmente a herramientas tecnológicas disponibles.

Los registros longitudinales también permiten observar la consolidación de conocimientos que inicialmente aparecen fragmentados. Una idea formulada con vacilaciones puede adquirir precisión mediante experiencias sucesivas de aprendizaje. Reconocer tal evolución exige atender a relaciones conceptuales y no únicamente a resultados numéricos. Tempelaar et al. (2024) vinculan la evaluación formativa con información sobre disposiciones de aprendizaje. Esa relación abre posibilidades para interpretar trayectorias individuales considerando tanto el desempeño como diversos factores asociados con la participación académica.

La organización del portafolio debe facilitar comparaciones comprensibles sin imponer una trayectoria idéntica a todos los estudiantes. Las muestras pueden agruparse por competencias o problemas recurrentes y acompañarse de fechas claramente verificables. También conviene conservar las consignas originales para interpretar las exigencias de cada producción. Un trabajo temprano no adquiere significado por su antigüedad. Su relevancia aparece cuando permite reconocer diferencias sustantivas respecto

de desempeños posteriores y comprender mejor la naturaleza de las transformaciones observadas.

La valoración longitudinal requiere distinguir progreso individual y nivel de dominio alcanzado. Un estudiante puede experimentar avances considerables sin satisfacer todavía los criterios establecidos para una competencia. Otro puede mantener desempeños elevados con variaciones menores durante el periodo observado. Ambas trayectorias aportan información relevante, aunque responden a preguntas evaluativas diferentes. Conviene reconocer el crecimiento sin perder de vista las expectativas académicas comunes. La equidad depende de comunicar claramente el peso asignado a cada dimensión evaluativa.

El seguimiento prolongado plantea responsabilidades relacionadas con la interpretación y conservación de información estudiantil. Los registros pueden revelar hábitos de trabajo o dificultades que merecen tratamiento reservado. Tempelaar et al. (2024) abordan la relación entre analítica disposicional y evaluación formativa, una vinculación que invita a examinar cuidadosamente el uso pedagógico de los datos. La acumulación de información necesita propósitos definidos y límites razonables. Ninguna trayectoria debería convertirse en una etiqueta permanente sobre las capacidades personales.

Un portafolio de progresión cognitiva adquiere sentido cuando permite reconocer aprendizajes que permanecen y se transforman durante experiencias sucesivas. Su lectura ofrece al docente una perspectiva temporal que difícilmente proporciona una prueba aislada. También permite valorar diferencias entre desempeños inmediatos y conocimientos que resisten nuevas exigencias académicas. En entornos mediados por inteligencia artificial, la continuidad de las evidencias ayuda a interpretar el desarrollo intelectual con mayor prudencia y a fundamentar decisiones educativas sobre trayectorias documentadas.

Tabla 3

Dimensiones y criterios para evaluar las evidencias del pensamiento en entornos educativos mediados por inteligencia artificial

Dimensión de evaluación	Criterios y evidencias de aprendizaje
Trazabilidad del razonamiento y construcción de respuestas	Reconstrucción de decisiones intelectuales durante la resolución de problemas. Valoración de las justificaciones y modificaciones que permiten reconocer la participación del estudiante en la elaboración del conocimiento.
Juicio crítico frente a respuestas generadas por IA	Capacidad para examinar afirmaciones y contrastar información con fuentes académicas. Identificación de errores y omisiones mediante argumentos fundamentados que demuestren autonomía intelectual.
Defensa oral como evidencia complementaria de comprensión	Explicación verbal de conceptos y decisiones académicas. Valoración de la consistencia argumentativa y de la capacidad para responder preguntas que permitan reconocer el dominio personal del conocimiento.
Transferencia mediante variaciones del problema	Aplicación de conocimientos ante modificaciones de las condiciones iniciales. Reconocimiento de principios pertinentes y adaptación justificada de estrategias frente a nuevas exigencias.
Metacognición documentada y conciencia del propio aprendizaje	Reconocimiento de avances y dificultades mediante registros reflexivos. Valoración de la capacidad para identificar necesidades formativas y regular decisiones durante el aprendizaje.
Evidencias longitudinales y portafolios de progresión cognitiva	Análisis de producciones académicas elaboradas en diferentes momentos. Identificación de transformaciones conceptuales y consolidación de aprendizajes mediante comparaciones fundamentadas en criterios previamente definidos.

Nota. Elaboración propia

Capítulo IV

Evaluar con IA sin delegar el juicio educativo

La presencia de inteligencia artificial en los procesos educativos ha modificado las condiciones bajo las cuales se produce, interpreta y valora el aprendizaje. Las herramientas generativas participan en actividades que anteriormente dependían de intercambios directos entre docentes y estudiantes. Pueden formular observaciones, analizar producciones y ofrecer respuestas con notable rapidez. Sin embargo, su incorporación plantea una pregunta pedagógica fundamental: quién responde por las decisiones que afectan la trayectoria formativa de una persona cuando intervienen sistemas capaces de emitir valoraciones automatizadas.

La evaluación educativa adquiere sentido mediante la interpretación de evidencias vinculadas con propósitos formativos. Ninguna herramienta tecnológica puede establecer por cuenta propia el significado pedagógico de una actuación sin recurrir a criterios definidos previamente. El profesorado necesita comprender las posibilidades de la inteligencia artificial y reconocer las condiciones que limitan su aplicación. Delcker et al. (2025) estudian las percepciones docentes sobre sus competencias en inteligencia artificial, aportando una perspectiva relevante para examinar las necesidades de preparación profesional.

La transformación de las prácticas evaluativas también modifica la experiencia del estudiante. Las herramientas generativas permiten consultar dudas, contrastar argumentos y recibir observaciones durante la elaboración de trabajos académicos. Cada interacción puede abrir oportunidades para reconsiderar decisiones intelectuales. Heil et al. (2025) investigan las percepciones estudiantiles sobre el impacto de la inteligencia artificial generativa en el aprendizaje y la evaluación universitaria, atendiendo al papel de la competencia individual. Su aproximación invita a examinar las condiciones que favorecen una participación académica reflexiva.

En medio de tales transformaciones, la retroalimentación ocupa un lugar especialmente significativo. La posibilidad de recibir comentarios durante el desarrollo de una actividad modifica los tiempos habituales de revisión y permite intervenir antes de entregar una producción terminada. Mekheimer (2025) estudia la retroalimentación asistida por inteligencia artificial generativa en la escritura en inglés como lengua extranjera, considerando el dominio lingüístico, la frecuencia de revisión y la calidad textual. Su investigación aporta elementos para comprender la mejora como proceso observable.

Figura 10

Dimensiones de la evaluación educativa mediada por IA



La disponibilidad de comentarios inmediatos también obliga a reconsiderar el significado de aprender mediante revisiones sucesivas. Un estudiante puede modificar repetidamente su trabajo sin comprender las razones que justifican cada cambio. En otras ocasiones, una observación puntual desencadena una reconsideración profunda del problema. La diferencia merece atención pedagógica. Evaluar el aprendizaje implica reconocer la calidad de las transformaciones intelectuales y las decisiones que

permiten sostenerlas mediante argumentos propios, incluso cuando intervienen recursos tecnológicos durante la elaboración.

Los instrumentos utilizados para valorar desempeños necesitan responder a condiciones educativas que cambian con rapidez. Las rúbricas tradicionales ofrecen referentes compartidos, aunque pueden resultar insuficientes cuando una actividad incorpora interacciones complejas con sistemas generativos. Norgaila et al. (2026) investigan la optimización de rúbricas mediante ingeniería reflexiva de instrucciones y examinan la alineación entre personas e inteligencia artificial. Su trabajo permite considerar la revisión de criterios como una práctica que requiere deliberación pedagógica y consistencia interpretativa.

La formulación de preguntas dirigidas a sistemas inteligentes representa otra manifestación del pensamiento académico. Cada instrucción revela decisiones sobre aquello que merece investigarse y las condiciones consideradas relevantes para obtener una respuesta. Nguyen y Hayward (2024) examinan la aplicación de inteligencia artificial generativa a la crítica de evaluaciones científicas. Su investigación ofrece una referencia para reflexionar sobre el valor de interrogar y revisar propuestas mediante herramientas tecnológicas, especialmente cuando la calidad de las preguntas condiciona el trabajo posterior.

A medida que aumentan las posibilidades de automatización, también crece la necesidad de establecer responsabilidades claramente reconocibles. La aparente precisión de una respuesta tecnológica puede favorecer su aceptación sin suficiente examen. Langer et al. (2025) analizan la supervisión humana de sistemas basados en inteligencia artificial desde una perspectiva centrada en detectar resultados inexactos e injustos. Su planteamiento adquiere relevancia educativa porque las valoraciones académicas pueden afectar oportunidades de aprendizaje y decisiones institucionales con consecuencias personales duraderas.

La responsabilidad evaluativa requiere reconocer que cada estudiante desarrolla su aprendizaje mediante experiencias y condiciones particulares. Una producción académica puede expresar dificultades transitorias, avances significativos o decisiones originales que demandan interpretación cuidadosa. La inteligencia artificial puede aportar información para examinar tales manifestaciones, aunque su utilidad depende de la lectura pedagógica realizada posteriormente. Preservar espacios de deliberación permite atender circunstancias que difícilmente quedan representadas mediante puntuaciones y categorías generales utilizadas en procedimientos de valoración automatizada.

El capítulo aborda la incorporación de inteligencia artificial a la evaluación desde una preocupación compartida por la calidad del juicio educativo. La preparación docente, la interacción crítica del estudiante y la revisión de evidencias forman parte de una práctica que necesita criterios comprensibles y responsabilidades definidas. También adquieren relevancia los instrumentos de valoración y los límites de las decisiones automatizadas. El propósito es examinar condiciones que permitan aprovechar recursos tecnológicos preservando la deliberación humana sobre el aprendizaje.

4.1. Alfabetización evaluativa en inteligencia artificial

La alfabetización evaluativa en inteligencia artificial constituye una competencia profesional que permite al docente comprender, examinar y utilizar críticamente los sistemas inteligentes durante la valoración del aprendizaje. Su desarrollo exige reconocer las posibilidades técnicas de dichas herramientas y las decisiones pedagógicas que intervienen en su aplicación. Evaluar mediante inteligencia artificial requiere interpretar sus resultados desde criterios educativos explícitos. La responsabilidad

docente permanece vinculada con la calidad de las evidencias, la pertinencia de las valoraciones y el sentido formativo del proceso.

La incorporación de inteligencia artificial a las prácticas evaluativas demanda conocimientos que exceden el manejo instrumental de plataformas digitales. El profesorado necesita comprender las condiciones bajo las cuales un sistema produce respuestas, clasifica desempeños o formula recomendaciones. Delcker et al. (2025) abordan la competencia docente en inteligencia artificial mediante el desarrollo de un instrumento fundamentado en evidencias para examinar las percepciones profesionales. Su investigación aporta una referencia para considerar la preparación del profesorado como una dimensión susceptible de análisis y fortalecimiento.

Una dimensión central de dicha alfabetización reside en distinguir entre la información generada automáticamente y las evidencias que permiten formular juicios educativos fundamentados. Un sistema puede identificar regularidades en una producción estudiantil sin comprender plenamente las circunstancias de su elaboración. La interpretación pedagógica exige atender las intenciones del estudiante, sus decisiones y la evolución de sus aprendizajes. Por ello, los resultados automatizados adquieren significado cuando el docente los contrasta con observaciones y criterios previamente establecidos.

El dominio evaluativo también requiere reconocer las limitaciones de los modelos generativos al interpretar producciones académicas. Una respuesta lingüísticamente convincente puede contener afirmaciones inexactas, inferencias débiles o valoraciones incompatibles con los objetivos de aprendizaje. La fluidez verbal del sistema tampoco garantiza precisión conceptual. Frente a tales situaciones, el profesorado necesita examinar la correspondencia entre los criterios solicitados y las respuestas obtenidas. Tal ejercicio demanda atención disciplinar y una disposición

permanente a cuestionar resultados aparentemente satisfactorios sin evidencia suficiente.

La alfabetización evaluativa adquiere profundidad cuando el docente examina sus propias capacidades para trabajar con inteligencia artificial. Las percepciones de competencia influyen en la disposición para experimentar con herramientas y reconocer necesidades de formación. El trabajo de Delcker et al. (2025) ofrece una aproximación investigativa a dicha dimensión profesional. Conviene recordar que sentirse competente no equivale necesariamente a demostrar un desempeño adecuado. La formación requiere contrastar la confianza personal con actuaciones observables ante situaciones evaluativas concretas.

Otro componente relevante corresponde a la capacidad de formular instrucciones que orienten el análisis automatizado de evidencias educativas. La precisión de una solicitud influye en la pertinencia de la respuesta generada. Sin embargo, una instrucción detallada tampoco elimina las posibilidades de error. El docente necesita establecer propósitos verificables, delimitar los criterios de valoración y revisar las interpretaciones producidas. Aprender a interrogar críticamente al sistema forma parte de una práctica evaluativa consciente de sus alcances y limitaciones.

La lectura crítica de los resultados automatizados exige prestar atención a los sesgos que pueden afectar las valoraciones del aprendizaje. Determinadas expresiones culturales, formas de argumentación o características lingüísticas podrían recibir interpretaciones inadecuadas cuando el sistema opera con patrones insuficientemente representativos. La alfabetización evaluativa permite advertir tales riesgos y examinar sus consecuencias pedagógicas. Una valoración responsable considera la diversidad de trayectorias estudiantiles y evita convertir las regularidades estadísticas en parámetros incuestionables de calidad académica.

La dimensión ética atraviesa las decisiones relacionadas con el tratamiento de información estudiantil mediante sistemas inteligentes. Antes de incorporar producciones académicas, registros de desempeño o comentarios personales, corresponde examinar las condiciones de privacidad y protección de datos. También resulta necesario determinar quién puede acceder a los resultados y bajo qué criterios serán utilizados. La transparencia adquiere relevancia cuando los estudiantes conocen la intervención tecnológica y comprenden las razones pedagógicas que fundamentan las valoraciones recibidas durante su formación.

El fortalecimiento de la alfabetización evaluativa requiere experiencias de formación vinculadas con problemas habituales de la práctica docente. Analizar respuestas generadas, comparar interpretaciones y discutir discrepancias permite desarrollar criterios profesionales compartidos. Desde la perspectiva de Delcker et al. (2025), el estudio de las competencias docentes en inteligencia artificial ofrece elementos para orientar procesos de diagnóstico profesional. La reflexión colegiada puede ampliar dicha preparación mediante el intercambio de experiencias y el examen de decisiones evaluativas tomadas en situaciones reales.

En términos pedagógicos, la alfabetización evaluativa representa una capacidad que se construye mediante el estudio, la práctica reflexiva y la revisión constante de decisiones. Su desarrollo permite incorporar inteligencia artificial con mayor discernimiento profesional y comprender las consecuencias de utilizarla para valorar aprendizajes. La autoridad educativa del docente se expresa en su capacidad para justificar cada juicio mediante evidencias pertinentes. Allí reside el fundamento de una evaluación tecnológicamente informada que conserva su responsabilidad humana y formativa.

4.2. IA como interlocutor crítico del estudiante

La inteligencia artificial puede desempeñar una función pedagógica relevante cuando participa como interlocutor crítico durante la elaboración de respuestas y soluciones académicas. Su intervención permite al estudiante confrontar argumentos, examinar alternativas y reconocer aspectos insuficientemente desarrollados en sus razonamientos. La calidad de dicha interacción depende de la disposición para cuestionar las respuestas recibidas. En una evaluación auténtica, el intercambio adquiere valor cuando favorece decisiones intelectuales propias y hace visible el proceso mediante el cual se construye conocimiento.

Concebir la inteligencia artificial como interlocutor implica modificar la relación habitual entre pregunta y respuesta. El estudiante deja de buscar una solución inmediata y aprende a sostener conversaciones orientadas al examen de sus ideas. Puede solicitar objeciones, identificar contradicciones o contrastar interpretaciones ante un problema determinado. La interacción adquiere profundidad cuando cada respuesta abre nuevas posibilidades de indagación. Interesa especialmente observar las razones que llevan al estudiante a aceptar, modificar o rechazar los planteamientos generados.

La participación de sistemas generativos en el aprendizaje plantea interrogantes acerca de las competencias necesarias para aprovechar sus posibilidades. Heil et al. (2025) investigan la percepción estudiantil sobre el impacto de dichas herramientas en el aprendizaje y la evaluación universitaria, considerando el papel de la competencia individual en inteligencia artificial. Su planteamiento permite reconocer que la experiencia educativa con sistemas generativos merece examinarse atendiendo a las capacidades del usuario y a las condiciones que orientan sus interacciones académicas.

Una conversación pedagógicamente significativa requiere preguntas capaces de provocar razonamientos más elaborados. Cuando el estudiante solicita explicaciones alternativas o plantea situaciones que contradicen una respuesta inicial, transforma el intercambio en una oportunidad para examinar sus propias ideas. La inteligencia artificial puede contribuir mediante contraargumentos y observaciones que amplían las perspectivas disponibles. Sin embargo, corresponde al estudiante determinar cuáles resultan pertinentes. La evaluación puede atender precisamente la calidad de las preguntas formuladas y las decisiones adoptadas durante el diálogo.

El valor formativo del intercambio también aparece cuando las respuestas generadas producen desacuerdos razonados. Una interpretación discutible puede impulsar al estudiante a consultar fuentes, revisar procedimientos o defender una posición mediante evidencias. Tales situaciones permiten observar capacidades que difícilmente se manifiestan en productos académicos terminados. La discrepancia adquiere utilidad educativa cuando conduce a una revisión fundamentada. Por consiguiente, el diálogo con inteligencia artificial puede ofrecer oportunidades para documentar procesos de deliberación y argumentación personal.

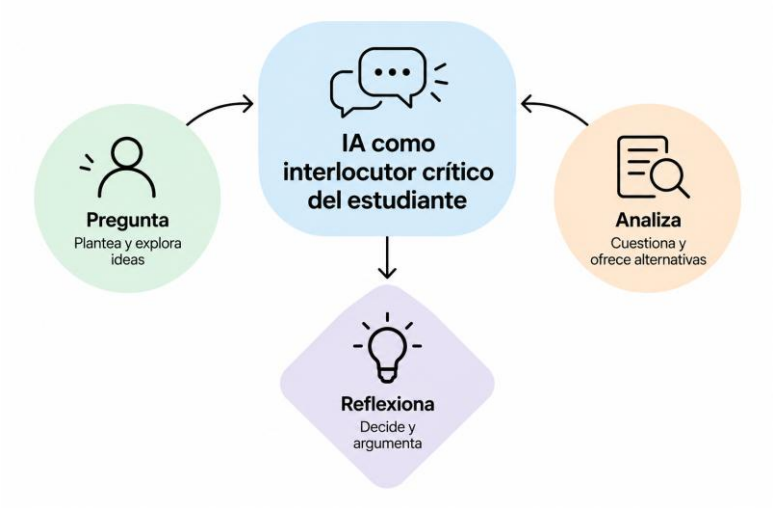
La autonomía intelectual constituye una condición necesaria para aprovechar críticamente las conversaciones con sistemas generativos. Heil et al. (2025) sitúan la competencia individual en inteligencia artificial dentro del análisis de las percepciones estudiantiles sobre aprendizaje y evaluación. Tal enfoque permite considerar la preparación del estudiante al examinar sus experiencias con herramientas generativas. Desde una perspectiva educativa, conviene observar si las interacciones fortalecen la capacidad de formular juicios propios o favorecen una aceptación poco reflexiva de respuestas automatizadas.

El intercambio crítico puede incorporarse a actividades evaluativas mediante registros breves de conversación

acompañados de comentarios reflexivos. El estudiante identifica preguntas relevantes, explica las modificaciones realizadas y fundamenta las decisiones que orientaron su trabajo. Tales registros permiten examinar la relación entre interacción tecnológica y elaboración intelectual. No interesa acumular transcripciones extensas. Resulta más significativo seleccionar momentos donde una objeción, una duda o una explicación alternativa haya contribuido a revisar el razonamiento desarrollado inicialmente.

Figura 11

Diálogo crítico entre estudiante e inteligencia artificial



La intervención docente adquiere especial relevancia al establecer condiciones que favorezcan conversaciones intelectualmente exigentes. Una consigna puede solicitar al estudiante que someta su argumento a críticas y examine posteriormente la consistencia de las observaciones recibidas. También puede requerir la identificación de afirmaciones discutibles mediante referencias académicas verificables. La calidad del aprendizaje se aprecia en la justificación de las decisiones

posteriores. Cada intercambio ofrece posibilidades distintas según la disciplina y la naturaleza del problema estudiado.

La experiencia de dialogar con inteligencia artificial también requiere reconocer las diferencias entre una interacción lingüística convincente y una comprensión auténtica del problema. Heil et al. (2025) examinan las percepciones estudiantiles acerca de las herramientas generativas en educación superior, una dimensión relevante para estudiar su incorporación académica. Desde la práctica evaluativa, corresponde distinguir entre satisfacción con las respuestas recibidas y capacidad para examinarlas críticamente. La confianza tecnológica necesita acompañarse de criterios intelectuales que permitan reconocer errores.

Cuando la inteligencia artificial participa como interlocutor crítico, la evaluación puede prestar mayor atención a la evolución del pensamiento estudiantil. Las preguntas reformuladas, los argumentos revisados y las decisiones justificadas permiten reconocer aprendizajes desarrollados durante la interacción. El valor educativo reside en la capacidad de sostener una posición razonada frente a respuestas que pueden resultar persuasivas o insuficientes. Aprender a dialogar críticamente con sistemas generativos amplía las oportunidades de ejercer autonomía intelectual ante problemas académicos complejos.

4.3. Retroalimentación aumentada y ciclos de mejora del aprendizaje

La retroalimentación aumentada mediante inteligencia artificial amplía las oportunidades de acompañar el aprendizaje mientras una tarea permanece abierta a revisión. Su aporte pedagógico depende de la relación entre las observaciones recibidas y las acciones que emprende el estudiante. Una devolución inmediata puede orientar ajustes puntuales, pero adquirir significado requiere interpretar sus fundamentos. La evaluación auténtica encuentra aquí un espacio para observar decisiones

sucesivas y reconocer avances que una entrega aislada difícilmente permitiría apreciar con claridad.

El ciclo de mejora comienza cuando el estudiante comprende qué aspecto de su desempeño necesita atención y por qué merece revisarlo. Las herramientas generativas pueden ofrecer comentarios sobre organización, precisión o coherencia según las características de la actividad. Conviene evitar indicaciones demasiado generales que apenas invitan a mejorar. Una devolución útil identifica un problema reconocible y abre posibilidades de intervención. La siguiente versión del trabajo permite examinar si aquella orientación produjo cambios pertinentes y comprensibles.

En el ámbito de la escritura en lengua extranjera, Mekheimer (2025) estudia la retroalimentación asistida por inteligencia artificial generativa en relación con el dominio lingüístico, la frecuencia de revisión y la calidad textual. La articulación de dichas dimensiones ofrece una referencia para pensar la mejora como proceso. Desde una perspectiva evaluativa, interesa atender tanto a las modificaciones realizadas como a las razones que las justifican. Revisar reiteradamente carece de valor cuando los cambios permanecen superficiales.

La oportunidad temporal de la retroalimentación influye en su aprovechamiento. Un comentario recibido durante la elaboración permite reconsiderar decisiones antes de que el estudiante dé por concluida la tarea. Sin embargo, la rapidez puede estimular correcciones automáticas sin suficiente examen. Por ello, resulta conveniente reservar momentos para interpretar las observaciones y decidir cuáles atender. La secuencia entre producir, recibir comentarios y volver a trabajar necesita tiempos de reflexión compatibles con la complejidad del aprendizaje real.

La retroalimentación aumentada también permite diferenciar necesidades dentro de un mismo grupo. Algunos estudiantes requieren orientaciones sobre aspectos fundamentales

mientras otros pueden beneficiarse de preguntas que profundicen sus argumentos. La personalización exige cautela: un sistema puede adaptar la formulación de sus comentarios sin comprender plenamente las dificultades individuales. El docente conserva la responsabilidad de reconocer cuándo una indicación resulta pertinente. La diversidad de apoyos educativos adquiere sentido cuando favorece progresos verificables y evita dependencias innecesarias.

La frecuencia de revisión merece una interpretación cuidadosa dentro de los ciclos de mejora. Mekheimer (2025) incorpora dicha variable al estudio de la retroalimentación generativa y la calidad de la escritura en inglés como lengua extranjera. Su consideración invita a distinguir cantidad de intervenciones y profundidad de los cambios. Un estudiante puede realizar numerosas correcciones menores mientras otro reorganiza sustancialmente su argumento. La evaluación necesita examinar la naturaleza de las transformaciones antes de atribuirles valor formativo.

Para documentar el aprendizaje conviene conservar versiones sucesivas acompañadas de explicaciones breves sobre las decisiones adoptadas. Una comparación entre borradores revela modificaciones visibles, aunque no siempre permite conocer sus motivos. El estudiante puede señalar qué comentario recibió, qué alternativa consideró y por qué eligió determinado cambio. Dicho registro ofrece evidencias de autorregulación sin convertir cada interacción en una obligación burocrática. Importa preservar las huellas relevantes del pensamiento y mantener proporcionada la carga de documentación académica.

No toda retroalimentación automatizada merece incorporarse al trabajo académico. Algunas observaciones pueden resultar imprecisas o privilegiar convenciones formales frente a decisiones expresivas justificadas. Aprender a descartar una recomendación también forma parte del ciclo de mejora. La

revisión responsable requiere contrastar cada comentario con los propósitos de la tarea y con conocimientos disciplinares pertinentes. Cuando el estudiante explica debidamente por qué mantiene una elección original, hace visible una capacidad evaluativa propia que también merece reconocimiento pedagógico.

El seguimiento docente puede centrarse en patrones de mejora que aparecen a lo largo de varias producciones. Mekheimer (2025) vincula en su investigación la retroalimentación generativa con variables relativas al desempeño escrito y la revisión textual. Tal aproximación permite plantear preguntas sobre la continuidad del progreso sin reducirlo a una puntuación. Interesa conocer qué dificultades persisten, cuáles disminuyen y qué apoyos dejan de ser necesarios. La interpretación de tales cambios requiere evidencias comparables y atención.

Una cultura de retroalimentación aumentada se consolida cuando revisar deja de percibirse como reparación de errores y pasa a formar parte del trabajo intelectual habitual. Las herramientas generativas pueden ampliar las ocasiones de recibir observaciones, aunque el aprendizaje depende de lo que cada estudiante hace con ellas. El ciclo cobra sentido al identificar una necesidad, intervenir deliberadamente y comprobar los resultados. Evaluar dicho proceso permite reconocer avances graduales sin confundir asistencia tecnológica con desarrollo autónomo.

4.4. Rúbricas dinámicas para desempeños mediados por IA

Las rúbricas dinámicas permiten valorar desempeños académicos que evolucionan durante la interacción con inteligencia artificial. Su diseño reconoce que una producción puede atravesar decisiones sucesivas antes de adquirir una forma suficientemente elaborada. Los criterios necesitan atender tanto a la calidad del resultado como a las acciones intelectuales que lo hicieron posible. La flexibilidad no implica modificar

arbitrariamente las exigencias. Requiere establecer referentes claramente comprensibles que acompañen diferentes momentos del trabajo y mantengan expectativas académicas compartidas.

Una rúbrica concebida para tareas mediadas por inteligencia artificial necesita distinguir dimensiones observables del desempeño. La formulación del problema, la selección de procedimientos y la justificación de decisiones pueden adquirir relevancia según los propósitos formativos. Conviene describir cada dimensión mediante conductas reconocibles y niveles diferenciados de calidad. Cuando los descriptores resultan ambiguos, aumentan las interpretaciones contradictorias. La precisión permite que estudiantes y docentes comprendan qué evidencias respaldan una valoración y dónde existen posibilidades de mejora.

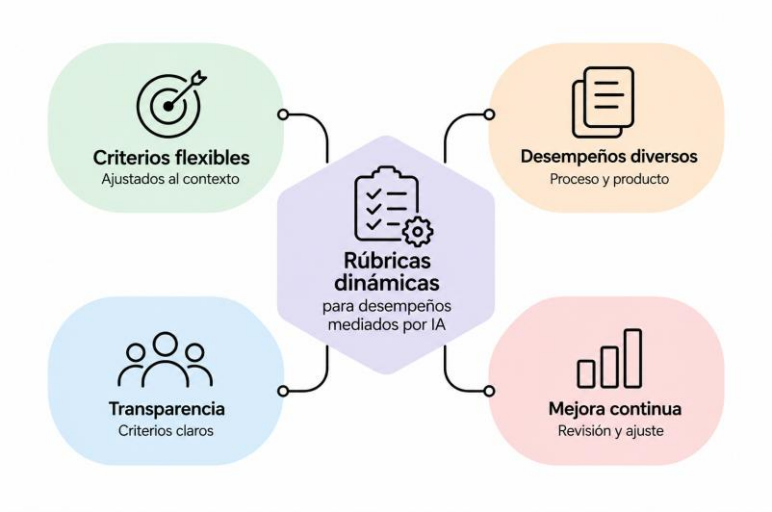
Norgaila et al. (2026) investigan la ingeniería reflexiva de instrucciones para optimizar rúbricas de evaluación y examinan la alineación entre personas e inteligencia artificial. Su trabajo sitúa la formulación y revisión de criterios dentro de un problema pedagógico relevante. La colaboración tecnológica puede facilitar propuestas de descriptores, pero cada formulación requiere examen profesional. La pertinencia de una rúbrica depende de su correspondencia con los aprendizajes esperados y con las características concretas de la actividad evaluada.

El carácter dinámico aparece cuando una rúbrica admite ajustes razonados durante el diseño y la revisión de una actividad. Tales modificaciones deben responder a evidencias pedagógicas y quedar documentadas para evitar diferencias injustificadas entre estudiantes. No conviene alterar los criterios después de conocer un resultado con la intención de favorecerlo o penalizarlo. La adaptabilidad exige reglas transparentes y acuerdos previos. De ese modo, el instrumento puede evolucionar sin comprometer la consistencia de las decisiones evaluativas.

En tareas desarrolladas con apoyo generativo, los niveles de desempeño pueden describir distintas formas de intervención estudiantil. Un nivel inicial podría evidenciar aceptación acrítica de propuestas automatizadas. Otros niveles reconocerían decisiones justificadas, comprobaciones pertinentes y transformaciones conceptuales relevantes. La progresión debe evitar premiar mecánicamente una mayor cantidad de interacciones. Lo importante es establecer diferencias realmente cualitativas entre actuaciones. Una rúbrica bien construida permite reconocer la autonomía demostrada mediante evidencias verificables dentro de cada producción académica.

Figura 12

Criterios dinámicos para valorar desempeños con IA



La revisión de descriptores mediante instrucciones reflexivas ofrece una vía para detectar inconsistencias antes de aplicar el instrumento. Norgaila et al. (2026) estudian precisamente la optimización de rúbricas mediante procedimientos orientados a la alineación humana con inteligencia artificial. Desde una aplicación pedagógica prudente, el profesorado puede comparar formulaciones alternativas y discutir sus implicaciones. Cada ajuste

necesita comprobar que los niveles mantienen diferencias comprensibles y que las exigencias corresponden a aquello que verdaderamente se pretende valorar.

La participación estudiantil en el análisis de rúbricas puede contribuir a comprender mejor las expectativas académicas. Antes de comenzar una actividad, conviene discutir ejemplos de desempeño y examinar por qué determinadas evidencias corresponden a distintos niveles. La inteligencia artificial puede generar casos hipotéticos para estimular dicha conversación. Sin embargo, los acuerdos requieren interpretación colectiva y orientación docente. Comprender los criterios permite anticipar decisiones durante el trabajo y reconocer posteriormente las razones de una valoración recibida.

Una dificultad frecuente consiste en confundir precisión formal con calidad del razonamiento. Los sistemas generativos pueden producir textos ordenados que aparentan cumplir numerosos descriptores sin demostrar una elaboración intelectual suficiente. Por ello, las rúbricas necesitan diferenciar presentación, fundamentación y contribución personal cuando la actividad lo requiera. La ponderación específica de cada dimensión debe responder a objetivos explícitos. Un trabajo expresivamente correcto puede presentar también debilidades conceptuales que merecen reconocimiento diferenciado dentro de la valoración académica.

La consistencia del instrumento puede examinarse mediante ejercicios de valoración compartida entre docentes. Ante una misma evidencia, las discrepancias permiten identificar descriptores poco precisos o niveles insuficientemente diferenciados. Norgaila et al. (2026) abordan la alineación humana con inteligencia artificial como objeto de investigación sobre optimización de rúbricas. Tal perspectiva permite considerar la comparación crítica de interpretaciones automatizadas y profesionales. El acuerdo mutuo razonado fortalece la confiabilidad

del instrumento sin eliminar la necesidad de deliberación pedagógica.

Las rúbricas dinámicas adquieren sentido cuando sus ajustes permiten representar mejor aquello que el estudiante demuestra en actividades mediadas por inteligencia artificial. Su utilidad depende de criterios públicos, descriptores pertinentes y procedimientos de revisión justificables. La tecnología puede colaborar en la elaboración del instrumento y facilitar comparaciones, pero la valoración exige interpretar evidencias educativas. Mantener abierta la revisión de criterios favorece instrumentos sensibles a nuevas prácticas académicas sin renunciar a exigencias compartidas de equidad educativa.

4.5. Evaluación de prompts, iteraciones y decisiones de interacción

La evaluación de instrucciones dirigidas a inteligencia artificial permite examinar una dimensión invisible del trabajo académico: las decisiones que orientan la producción de respuestas. Cada prompt expresa una interpretación del problema y delimita aquello que el estudiante considera necesario preguntar. Su calidad depende de la claridad del propósito, la pertinencia de las restricciones y la capacidad para provocar información examinable. Valorar tales elecciones ayuda a comprender el razonamiento anterior al producto entregado en actividades educativas.

Un prompt pertinente nace de comprender la tarea antes de acudir al sistema. El estudiante necesita reconocer qué información posee, qué incertidumbres enfrenta y qué clase de ayuda resulta adecuada. Una instrucción extensa puede ser confusa mientras una pregunta breve puede revelar un planteamiento riguroso. Por ello, la evaluación debe atender la relación entre intención y formulación. Interesa identificar si la consulta responde a una necesidad intelectual y permite avanzar sin desdibujar la responsabilidad personal.

Nguyen y Hayward (2024) estudian la aplicación de inteligencia artificial generativa a la crítica de evaluaciones en ciencias. Su investigación ofrece un punto de referencia para considerar el examen de preguntas y propuestas evaluativas mediante sistemas generativos. Trasladada al trabajo estudiantil, dicha preocupación invita a observar la calidad de las consultas planteadas. Una instrucción bien orientada puede solicitar la identificación de debilidades conceptuales y abrir una discusión rigurosa y fundamentada sobre criterios de validez disciplinar.

Las iteraciones revelan la manera en que una persona responde a resultados inesperados o insuficientes. Reformular una consulta puede indicar comprensión creciente del problema, aunque también puede expresar una búsqueda mecánica de respuestas convenientes. La diferencia aparece en las razones del cambio. Conviene examinar qué aspecto de la respuesta anterior motivó una nueva intervención y qué resultado se esperaba obtener. La secuencia adquiere significado evaluativo cuando permite reconstruir decisiones conscientes durante todo el proceso académico.

La evaluación de interacciones requiere distinguir entre ajustes superficiales y modificaciones sustantivas. Cambiar palabras para obtener una redacción más agradable no equivale a replantear una hipótesis o precisar una condición experimental. Ambas acciones pueden tener utilidad, pero representan exigencias intelectuales diferentes. El estudiante debería explicar qué aprendió de cada respuesta y qué criterio orientó su siguiente consulta. De tal manera, las iteraciones dejan de interpretarse mediante su cantidad y adquieren valor por su fundamento intelectual.

En actividades científicas, una interacción puede utilizarse para someter preguntas a examen antes de aplicarlas. Nguyen y Hayward (2024) abordan el empleo de inteligencia artificial generativa para criticar instrumentos de evaluación en ciencias. Tal

enfoque permite plantear ejercicios donde el estudiante analice observaciones automatizadas y determine cuáles merecen consideración. La valoración puede centrarse en la capacidad de reconocer ambigüedades, detectar exigencias desproporcionadas y defender modificaciones mediante conocimientos del campo disciplinar estudiado y sus exigencias específicas.

Las decisiones de interacción también comprenden el momento de detener una conversación. Continuar solicitando versiones puede producir cambios mínimos e incluso debilitar una respuesta previamente adecuada. Reconocer cuándo existe evidencia suficiente constituye una manifestación de criterio académico. El estudiante necesita establecer condiciones para cerrar la búsqueda y asumir la elaboración posterior. La evaluación puede considerar si dicha decisión responde a razones verificables o a una aceptación precipitada de la propuesta generada por el propio sistema.

Documentar el proceso no exige conservar indiscriminadamente todas las conversaciones. Una selección razonada de instrucciones iniciales, reformulaciones significativas y decisiones posteriores permite reconstruir el trabajo sin multiplicar registros innecesarios. Cada fragmento debería acompañarse de una explicación sobre su función dentro de la tarea. La evidencia adquiere mayor fuerza cuando muestra relaciones entre preguntas, respuestas y acciones. Un historial extenso carece de utilidad pedagógica cuando permanece desconectado de los aprendizajes que pretende demostrar ante el docente.

El análisis de instrucciones puede incorporarse a la revisión de actividades diseñadas por los propios estudiantes. Nguyen y Hayward (2024) examinan la crítica de evaluaciones científicas asistida por inteligencia artificial generativa. Desde esa línea temática, resulta pertinente solicitar que cada estudiante contraste una pregunta inicial con observaciones obtenidas durante la

interacción. La valoración debe atender la pertinencia de los cambios realizados y la capacidad para justificar por qué determinadas recomendaciones fueron rechazadas durante la revisión.

Evaluar prompts e iteraciones permite reconocer la autoría intelectual en decisiones que rara vez aparecen en una entrega terminada. La formulación de consultas, la interpretación de respuestas y la interrupción deliberada del intercambio constituyen acciones susceptibles de examen pedagógico. Su valoración requiere evidencias seleccionadas y criterios vinculados con la tarea. Cuando el estudiante puede explicar sus elecciones, la interacción tecnológica deja una huella comprensible de su razonamiento y de su responsabilidad sobre su propio aprendizaje.

4.6. Supervisión humana y límites de la automatización evaluativa

La supervisión humana de la evaluación automatizada comienza al reconocer que toda calificación afecta oportunidades educativas. Un sistema puede ordenar evidencias y producir valoraciones consistentes en apariencia, pero carece de responsabilidad institucional ante sus consecuencias. El docente necesita conservar autoridad para examinar resultados, solicitar aclaraciones y corregir decisiones improcedentes. Dicha responsabilidad adquiere especial relevancia cuando una valoración influye en la promoción, el acceso a apoyos o la percepción que un estudiante construye sobre sus capacidades.

Supervisar exige algo más que revisar ocasionalmente las respuestas producidas por una herramienta. La intervención debe contar con momentos definidos, información accesible y posibilidades reales de modificar resultados. Cuando el profesorado recibe una calificación automatizada sin conocer sus fundamentos, la revisión pierde profundidad. Una supervisión efectiva requiere identificar qué datos fueron considerados y cuáles quedaron fuera

del análisis. También demanda tiempo institucional para contrastar las valoraciones con evidencias del desempeño estudiantil antes de comunicarlas oficialmente.

Langer et al. (2025) analizan la supervisión humana de sistemas basados en inteligencia artificial desde una perspectiva de detección de señales, centrada en identificar resultados inexactos e injustos. Su planteamiento permite examinar una dificultad frecuente: advertir errores dentro de conjuntos de respuestas aparentemente aceptables. En evaluación educativa, la vigilancia humana necesita condiciones que faciliten reconocer anomalías. La mera presencia de una persona durante el procedimiento no garantiza que las decisiones automatizadas reciban un examen suficiente.

La confianza excesiva en las recomendaciones automatizadas puede reducir la atención dedicada a casos particulares. Cuando varias valoraciones parecen razonables, aumenta el riesgo de aceptar una respuesta problemática sin comprobarla. También puede ocurrir lo contrario: desconfiar indiscriminadamente de resultados pertinentes y multiplicar revisiones innecesarias. El equilibrio requiere procedimientos de contraste proporcionados al riesgo educativo. Una calificación con consecuencias importantes merece controles más exigentes que una orientación provisional destinada a mejorar un ejercicio durante su elaboración.

Los límites de la automatización aparecen con especial claridad ante producciones que admiten interpretaciones legítimas diferentes. Un ensayo puede desarrollar una perspectiva original sin ajustarse a patrones habituales de organización. Una solución matemática puede utilizar un procedimiento poco frecuente y conservar validez. La valoración automatizada podría pasar por alto tales particularidades. Corresponde al docente reconocer cuándo la singularidad del desempeño exige una lectura disciplinar detenida

y una justificación que considere las características de cada respuesta.

La detección de resultados injustos plantea exigencias adicionales porque algunas diferencias de tratamiento pueden permanecer ocultas tras puntuaciones aparentemente coherentes. Langer et al. (2025) consideran conjuntamente la identificación de inexactitudes y de resultados injustos al estudiar la supervisión humana. En educación, dicha preocupación invita a revisar valoraciones que afecten de manera desigual a estudiantes con formas diversas de expresión. La revisión necesita apoyarse en evidencias y procedimientos verificables, evitando atribuir intenciones al sistema sin fundamento.

Una supervisión responsable necesita establecer vías para impugnar resultados y solicitar una segunda valoración. El estudiante debe conocer los criterios aplicados y disponer de oportunidades para presentar evidencias adicionales. La revisión humana adquiere legitimidad cuando puede modificar una decisión y explicar los motivos del cambio. Sin mecanismos de reclamación, la automatización corre el riesgo de convertir errores corregibles en decisiones difíciles de cuestionar. La institución debe garantizar procedimientos accesibles y plazos compatibles con sus consecuencias.

La distribución de responsabilidades también merece atención cuando intervienen docentes, equipos técnicos y autoridades académicas. Quien configura una herramienta puede desconocer particularidades curriculares que resultan decisivas para interpretar sus respuestas. Por su parte, el docente puede necesitar apoyo especializado para comprender fallos recurrentes. La coordinación requiere definir quién valida criterios, quién examina incidencias y quién autoriza decisiones con efectos académicos. Ninguna distribución debería diluir la obligación institucional de responder ante valoraciones perjudiciales o insuficientemente fundamentadas.

Tabla 4

Dimensiones pedagógicas y criterios para evaluar el aprendizaje mediado por inteligencia artificial

Dimensión de evaluación	Orientaciones pedagógicas y criterios de aplicación
Alfabetización evaluativa en inteligencia artificial	Desarrollar competencias docentes para comprender las capacidades y limitaciones de los sistemas inteligentes. Examinar críticamente sus resultados y fundamentar las valoraciones mediante evidencias educativas pertinentes.
IA como interlocutor crítico del estudiante	Promover interacciones orientadas al cuestionamiento de argumentos y la revisión de ideas. Valorar la capacidad del estudiante para contrastar respuestas y justificar sus decisiones mediante razonamientos propios.
Retroalimentación aumentada y ciclos de mejora	Aprovechar las observaciones generadas mediante IA para orientar revisiones sucesivas. Examinar la profundidad de los cambios y reconocer avances mediante evidencias del proceso de aprendizaje.
Rúbricas dinámicas para desempeños mediados por IA	Diseñar criterios flexibles y transparentes que permitan valorar procesos y productos académicos. Revisar los descriptores según los objetivos formativos y garantizar interpretaciones coherentes entre diferentes desempeños.
Evaluación de prompts, iteraciones y decisiones de interacción	Examinar la pertinencia de las instrucciones formuladas y las razones que motivan sus modificaciones. Reconocer la calidad del razonamiento expresado durante las interacciones con sistemas generativos.
Supervisión humana y límites de la automatización evaluativa	Mantener la responsabilidad docente sobre las decisiones académicas. Establecer procedimientos de revisión que permitan detectar errores y corregir valoraciones injustas. Garantizar transparencia y posibilidades de reclamación.

Nota. Elaboración propia. La tabla organiza las dimensiones del capítulo 4 y sus principales criterios de aplicación pedagógica.

La eficacia de la supervisión depende asimismo de las condiciones de trabajo. Langer et al. (2025) proponen examinar la detección de resultados problemáticos mediante un enfoque que permite atender las exigencias de vigilancia humana. En el ámbito

educativo, revisar grandes volúmenes de calificaciones bajo presión temporal puede dificultar la identificación de errores. Conviene diseñar controles manejables y priorizar casos de mayor repercusión. La capacidad real de revisión importa tanto como la existencia de protocolos institucionales.

Establecer límites a la automatización evaluativa implica decidir qué actuaciones requieren deliberación profesional indelegable. La tecnología puede colaborar en tareas de organización y análisis preliminar, mientras las decisiones de mayor trascendencia necesitan siempre revisión fundamentada. La supervisión cobra valor cuando dispone de información comprensible, autoridad efectiva y mecanismos de rectificación. En una evaluación auténtica, responder por cada juicio exige reconocer al estudiante como persona cuyo aprendizaje merece una interpretación cuidadosa y una decisión académica justificable.

Capítulo V

Hacia un nuevo contrato evaluativo entre estudiante, docente e IA

La incorporación de inteligencia artificial generativa en la educación superior ha modificado las condiciones bajo las cuales estudiantes y docentes construyen sus relaciones académicas. Las actividades que antes permitían reconocer determinados aprendizajes ahora pueden resolverse mediante sistemas capaces de producir respuestas elaboradas. Frente a tal transformación, la evaluación necesita recuperar su sentido formativo y su legitimidad institucional. El problema trasciende la disponibilidad tecnológica y alcanza los acuerdos que permiten reconocer el mérito intelectual dentro de una comunidad educativa.

Durante mucho tiempo, las instituciones educativas han organizado sus prácticas evaluativas alrededor de una relación relativamente estable entre producción académica y desempeño individual. La expansión de herramientas generativas introduce nuevas mediaciones en esa relación y obliga a reconsiderar expectativas compartidas. Xia et al. (2024) examinan las transformaciones que la inteligencia artificial generativa plantea para la evaluación universitaria. Su revisión permite comprender que las respuestas institucionales requieren una mirada amplia sobre los procesos educativos y sus fundamentos pedagógicos.

La construcción de un nuevo contrato evaluativo comienza por reconocer que la confianza académica necesita acuerdos comprensibles para quienes participan en ella. El estudiante debe conocer las responsabilidades vinculadas con su producción intelectual. El docente requiere criterios que permitan interpretar actuaciones desarrolladas mediante recursos tecnológicos diversos. La institución tiene la obligación de proporcionar condiciones de aplicación coherentes. Ninguna de tales responsabilidades puede ejercerse adecuadamente cuando las reglas permanecen implícitas o dependen de interpretaciones individuales contradictorias.

La integridad académica adquiere particular relevancia dentro de esa reorganización de responsabilidades. Las posibilidades de generar contenidos mediante inteligencia artificial

amplían las preguntas sobre autoría y honestidad intelectual. Torres-Diaz et al. (2025) estudian las relaciones entre inteligencia artificial y plagio en universidades ecuatorianas, situando una preocupación directamente vinculada con las prácticas contemporáneas de producción académica. El debate exige considerar las condiciones educativas que favorecen comportamientos responsables y permiten reconocer la participación humana en los trabajos presentados.

Figura 13

Principios del nuevo contrato evaluativo



Las respuestas centradas en restricciones generales encuentran dificultades cuando las herramientas generativas intervienen en actividades habituales de aprendizaje. Gruenhagen et al. (2024) investigan las percepciones estudiantiles y el empleo de chatbots como apoyo en evaluaciones académicas. Su aproximación permite reconocer la importancia de comprender las prácticas reales antes de formular disposiciones institucionales. La discusión requiere examinar qué responsabilidades permanecen bajo control del estudiante y qué formas de colaboración tecnológica pueden resultar compatibles con determinados propósitos educativos.

La transformación también alcanza los criterios mediante los cuales se reconoce la calidad del desempeño académico. La

presentación de un producto bien elaborado puede resultar insuficiente para comprender el trabajo intelectual desarrollado durante su preparación. Gustilo et al. (2024) analizan prácticas y percepciones docentes relacionadas con la escritura mediada por algoritmos y la integridad académica. Tal problemática invita a reconsiderar el significado de la autoría y la pertinencia de las evidencias utilizadas para fundamentar decisiones evaluativas.

Las diferencias entre estudiantes plantean otra cuestión que merece atención institucional. Una educación comprometida con la equidad necesita reconocer trayectorias diversas sin debilitar las exigencias propias de cada formación profesional. Bower et al. (2024) examinan las perspectivas docentes sobre los cambios educativos asociados con sistemas generativos de creciente capacidad. La discusión permite considerar oportunidades de adaptación pedagógica y sus límites. Mantener expectativas académicas compartidas requiere distinguir entre condiciones diferenciadas de participación y competencias efectivamente acreditadas.

La confianza educativa también puede verse afectada cuando la incertidumbre sobre la autoría conduce a prácticas de supervisión excesiva. Waltzer et al. (2024) investigan la identificación de textos generados por inteligencia artificial en ensayos universitarios. Su trabajo sitúa una cuestión relevante para el juicio académico: las dificultades de atribuir procedencia a una producción escrita. Ante tal escenario, las instituciones necesitan procedimientos prudentes que respeten los derechos estudiantiles y permitan fundamentar sus decisiones mediante información pertinente.

El alcance de tales transformaciones reclama una participación institucional que vaya más allá de iniciativas individuales. Las decisiones sobre evaluación comprometen programas académicos y mecanismos de acreditación cuya consistencia necesita preservarse. Xia et al. (2024) sitúan la

inteligencia artificial generativa dentro de un panorama amplio de cambios en la educación superior. Desde esa perspectiva, la coordinación entre actores educativos adquiere relevancia para sostener acuerdos compartidos y evitar que cada asignatura responda de manera aislada ante problemas semejantes.

El nuevo contrato evaluativo requiere comprender la relación entre estudiante, docente e inteligencia artificial como una realidad educativa sometida a deliberación permanente. Sus fundamentos descansan en responsabilidades reconocibles y criterios académicos capaces de conservar legitimidad frente a condiciones cambiantes. Las decisiones adoptadas deberán responder tanto a las posibilidades tecnológicas como a los compromisos formativos de la institución. Desde esa preocupación se plantea una revisión de los acuerdos que orientan la evaluación y sostienen la confianza universitaria.

5.1. Transparencia del uso de IA como principio de integridad académica

La transparencia en el uso de inteligencia artificial constituye una condición necesaria para preservar la integridad académica en los procesos educativos contemporáneos. Su importancia radica en reconocer abiertamente la participación de herramientas automatizadas durante la elaboración de actividades y productos de aprendizaje. Cuando el estudiante declara dicha intervención, permite distinguir sus aportaciones intelectuales de aquellas generadas mediante sistemas tecnológicos. Tal reconocimiento fortalece la confianza institucional y favorece una relación educativa basada en la responsabilidad personal y la honestidad.

La incorporación de inteligencia artificial en las prácticas universitarias ha modificado las condiciones bajo las cuales se interpreta la autoría académica. Una producción escrita puede reunir razonamientos personales y contenidos generados mediante

aplicaciones automatizadas sin que resulte evidente su procedencia. Ante tal circunstancia, Torres-Diaz et al. (2025) examinan las relaciones entre inteligencia artificial, plagio e integridad académica en universidades ecuatorianas. Su investigación permite situar la declaración del uso tecnológico dentro de una discusión institucional más amplia.

Declarar la participación de inteligencia artificial exige reconocer que las decisiones intelectuales permanecen vinculadas con quien presenta un trabajo académico. La transparencia adquiere sentido cuando el estudiante identifica las tareas realizadas mediante herramientas automatizadas y explica su intervención personal en los resultados obtenidos. Conviene distinguir entre utilizar una aplicación para revisar aspectos lingüísticos y delegarle la elaboración de argumentos sustantivos. Ambas prácticas requieren criterios de declaración diferenciados porque representan grados distintos de participación tecnológica en la producción académica.

La integridad académica también depende de la claridad con que las instituciones comuniquen sus expectativas respecto del uso de inteligencia artificial. Las declaraciones ambiguas pueden provocar interpretaciones contradictorias entre estudiantes y docentes, especialmente cuando una misma herramienta cumple funciones diversas. Torres-Diaz et al. (2025) sitúan el problema del plagio relacionado con inteligencia artificial dentro del ámbito universitario ecuatoriano. Desde tal perspectiva, establecer orientaciones comprensibles resulta pertinente para evitar que las diferencias de interpretación comprometan la confianza académica compartida.

Una declaración transparente adquiere mayor valor cuando describe acciones verificables y evita formulaciones genéricas. Expresiones como haber utilizado inteligencia artificial durante una actividad ofrecen información insuficiente para comprender su verdadera participación. Resulta más esclarecedor

indicar si la herramienta intervino en la búsqueda de ideas, la organización preliminar del contenido o la revisión de un borrador. La precisión permite reconocer decisiones intelectuales y delimitar responsabilidades sin convertir cada interacción tecnológica en un procedimiento administrativo innecesariamente complejo para todos.

El compromiso con la transparencia requiere condiciones institucionales que favorezcan declaraciones sinceras. Cuando comunicar el uso de inteligencia artificial despierta automáticamente sospechas de deshonestidad, algunos estudiantes pueden optar por ocultarlo. Una cultura académica orientada hacia la responsabilidad necesita diferenciar la utilización reconocida de herramientas tecnológicas de las prácticas destinadas a presentar aportaciones ajenas como propias. La investigación de Torres-Diaz et al. (2025) aporta un referente para examinar dicha problemática desde la integridad universitaria contemporánea.

La responsabilidad docente adquiere particular relevancia al interpretar las declaraciones de uso tecnológico. Una explicación detallada puede revelar decisiones metodológicas que permanecen invisibles en el producto entregado. También permite identificar diferencias entre estudiantes que emplearon herramientas semejantes con propósitos distintos. El reconocimiento de tales particularidades demanda una lectura cuidadosa de las evidencias disponibles. La transparencia pierde legitimidad cuando las declaraciones se utilizan para formular juicios anticipados sin considerar las condiciones reales de elaboración del trabajo presentado.

La dimensión ética de la transparencia trasciende el cumplimiento de disposiciones institucionales. Declarar la intervención de inteligencia artificial expresa una disposición a responder por las decisiones adoptadas durante el trabajo intelectual. En tal sentido, la honestidad académica se manifiesta mediante prácticas reconocibles que permiten comprender la

procedencia de las aportaciones. Torres-Diaz et al. (2025) abordan una problemática vinculada con las relaciones entre tecnología y plagio universitario, cuya consideración invita a examinar cuidadosamente las responsabilidades de autoría.

La transparencia también plantea exigencias relacionadas con la privacidad y el tratamiento responsable de la información académica. Documentar el empleo de inteligencia artificial no debería obligar a divulgar conversaciones personales ni materiales confidenciales utilizados durante una actividad. El principio de proporcionalidad permite establecer qué información resulta pertinente para reconocer la intervención tecnológica. Una declaración suficientemente precisa puede cumplir su propósito sin exponer datos innecesarios. Preservar dicho equilibrio contribuye a mantener la confianza entre quienes participan en las prácticas educativas.

Consolidar la transparencia como principio de integridad académica requiere reconocer su carácter formativo y su dimensión institucional. La declaración del uso de inteligencia artificial adquiere legitimidad cuando responde a criterios conocidos y se interpreta con justicia. Su aplicación cotidiana permite desarrollar hábitos de responsabilidad intelectual compatibles con las transformaciones tecnológicas del conocimiento. En el nuevo contrato evaluativo, la confianza entre estudiante y docente depende también de reconocer abiertamente las mediaciones tecnológicas presentes en cada producción académica.

5.2. De la prohibición a los niveles autorizados de asistencia artificial

La prohibición general de inteligencia artificial en las actividades evaluativas responde a una preocupación legítima por preservar la autoría del estudiante. Sin embargo, resulta difícil sostenerla cuando las herramientas generativas forman parte de múltiples prácticas de estudio y producción intelectual. El

problema pedagógico consiste en determinar qué formas de asistencia son compatibles con los aprendizajes esperados. Una regulación diferenciada permite abandonar respuestas uniformes y establecer límites vinculados con las capacidades que cada actividad pretende valorar.

El tránsito hacia niveles autorizados de asistencia exige reconocer que las herramientas generativas pueden intervenir en momentos distintos del trabajo académico. Gruenhagen et al. (2024) estudian las percepciones estudiantiles y el uso de chatbots como apoyo en tareas evaluativas, aportando elementos para comprender esa diversidad de prácticas. Una política institucional sensible a tales diferencias puede distinguir entre consultas orientativas, apoyo en la revisión y generación de contenido. Cada modalidad plantea exigencias particulares de autonomía intelectual.

En un nivel inicial, la asistencia artificial puede limitarse a resolver dudas operativas o facilitar la comprensión de instrucciones. El estudiante conserva íntegramente la responsabilidad de interpretar el problema y construir su respuesta. Bajo ese criterio, consultar definiciones o solicitar aclaraciones generales no equivale a delegar el razonamiento exigido. La autorización debe formularse con ejemplos cercanos a la actividad concreta para evitar que una categoría aparentemente sencilla admita usos incompatibles con el propósito formativo previsto.

Un segundo nivel permite utilizar inteligencia artificial durante la preparación del trabajo, especialmente para organizar materiales o contrastar posibles enfoques. La herramienta puede proponer esquemas y preguntas que ayuden a delimitar una investigación. Corresponde al estudiante seleccionar las alternativas pertinentes y justificar sus decisiones. La frontera pedagógica aparece cuando la organización deja de ser un apoyo y comienza a sustituir la elaboración conceptual. Reconocer esa

diferencia requiere atender a la naturaleza de cada disciplina académica.

En actividades que admiten asistencia para revisar borradores, la autorización puede abarcar observaciones sobre claridad argumentativa, coherencia textual o problemas de expresión. Tal posibilidad resulta valiosa cuando el objetivo principal no consiste en medir habilidades de corrección lingüística. Aun entonces, aceptar recomendaciones automáticas demanda juicio propio. El estudiante necesita comprender los cambios incorporados y rechazar aquellos que alteren el significado de sus ideas. La revisión asistida mantiene sentido educativo mientras conserve esa deliberación personal consciente.

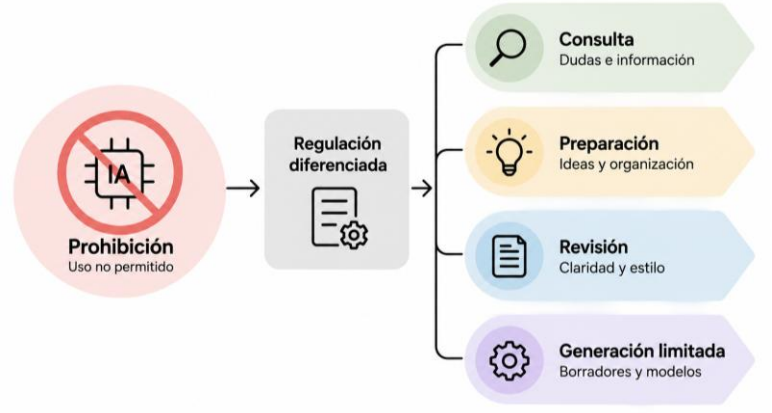
Los niveles más amplios permiten generar fragmentos, modelos o soluciones preliminares que después serán examinados críticamente. Gruenhagen et al. (2024) analizan el empleo de chatbots en evaluaciones desde la perspectiva estudiantil, una aproximación pertinente para discutir permisos diferenciados. En determinadas actividades, valorar una respuesta producida por inteligencia artificial puede constituir parte del aprendizaje. La autorización adquiere coherencia cuando el estudiante debe identificar errores, contrastar evidencias y fundamentar las modificaciones críticas realizadas sobre el material recibido.

La elección del nivel permitido depende de la competencia que se busca observar. Si una actividad pretende valorar escritura autónoma, autorizar la generación completa del texto comprometería la interpretación del desempeño. En cambio, cuando interesa evaluar decisiones profesionales frente a propuestas automatizadas, restringir cualquier asistencia podría empobrecer la tarea. No existe una escala universal aplicable sin ajustes. Cada autorización necesita relacionarse con el propósito de aprendizaje y con las evidencias que permitirán juzgarlo de manera razonada.

La comunicación de los permisos merece tanta atención como su diseño. Una indicación breve puede resultar insuficiente cuando el estudiante desconoce qué acciones quedan comprendidas dentro de cada nivel. Conviene presentar situaciones permitidas y otras que excedan la autorización establecida. Gruenhagen et al. (2024) examinan las percepciones de estudiantes sobre chatbots y evaluaciones, asunto que refuerza la pertinencia de considerar la interpretación estudiantil de las reglas. La claridad reduce conflictos y facilita decisiones académicas informadas.

Figura 14

Niveles autorizados de asistencia artificial



Una regulación graduada también requiere reconocer diferencias materiales entre estudiantes. El acceso a versiones comerciales, la familiaridad con determinadas aplicaciones y las habilidades para formular instrucciones pueden producir ventajas desiguales. Autorizar una herramienta sin considerar tales condiciones puede afectar la equidad evaluativa. Por ello, las instituciones necesitan precisar recursos disponibles y alternativas razonables para quienes no puedan utilizarlos. La igualdad de oportunidades depende de condiciones efectivas de participación, además de reglas comunes acordadas para todos.

Pasar de la prohibición a permisos graduados modifica la responsabilidad compartida dentro de la evaluación. El docente determina qué asistencia resulta admisible según los aprendizajes previstos y el estudiante decide utilizarla dentro de límites conocidos. La institución proporciona criterios consistentes para resolver situaciones controvertidas. Tal distribución evita convertir la tecnología en una excepción permanente. Los niveles autorizados adquieren sentido cuando permiten juzgar el desempeño humano con criterios explícitos y respetan las particularidades de cada actividad.

5.3. Nuevas rúbricas para valorar la contribución intelectual humana

La presencia de inteligencia artificial en la producción académica obliga a revisar aquello que las rúbricas consideran evidencia de aprendizaje. La calidad formal de una respuesta ya no permite inferir directamente el dominio intelectual de quien la presenta. Resulta necesario valorar decisiones que revelen comprensión, criterio y responsabilidad sobre el trabajo realizado. Una rúbrica renovada debe identificar aportaciones humanas observables sin confundir fluidez expresiva con conocimiento construido mediante reflexión durante el proceso de elaboración académica.

El primer criterio corresponde a la formulación del problema. Reconocer una dificultad relevante, delimitar sus alcances y establecer preguntas pertinentes demanda una intervención intelectual que merece valoración específica. Una herramienta puede ofrecer formulaciones alternativas, pero corresponde al estudiante justificar por qué adopta determinada perspectiva. La rúbrica debería distinguir entre reproducir una consigna y convertirla en una cuestión investigable. Tal distinción permite apreciar el razonamiento anterior a cualquier respuesta y su relación con propósitos disciplinares concretos.

La selección de información constituye otra dimensión evaluable. No basta reunir referencias o presentar afirmaciones plausibles cuando los sistemas generativos pueden producir ambas con rapidez. Interesa examinar los motivos por los cuales una fuente fue considerada pertinente y otra quedó descartada. Gustilo et al. (2024) estudian las prácticas y percepciones docentes ante la escritura impulsada por algoritmos. Su trabajo ofrece un punto de partida para reconsiderar criterios de valoración vinculados con la integridad académica contemporánea.

La argumentación requiere indicadores que permitan reconocer relaciones razonadas entre evidencias y conclusiones. Una respuesta convincente en apariencia puede ocultar inferencias débiles o afirmaciones insuficientemente fundamentadas. Por ello, la rúbrica debe atender a la capacidad de explicar conexiones, reconocer objeciones y revisar posiciones cuando aparecen datos contradictorios. Importa menos la extensión del discurso que la consistencia de sus decisiones intelectuales. Tal criterio puede observarse mediante comentarios justificativos y defensas orales breves del trabajo presentado.

También merece atención la capacidad de examinar críticamente las propuestas generadas por inteligencia artificial. Identificar errores conceptuales constituye una manifestación de conocimiento, pero detectar omisiones o presupuestos discutibles exige una lectura todavía más cuidadosa. Una rúbrica puede diferenciar entre aceptar recomendaciones sin examen y someterlas a contrastación fundada. La valoración aumenta cuando el estudiante explica las modificaciones introducidas y demuestra que comprende sus consecuencias. El desacuerdo argumentado adquiere entonces un valor educativo propio y verificable.

La originalidad necesita una interpretación menos dependiente de la novedad superficial del producto. Una aportación intelectual puede manifestarse en la elección de un enfoque, la articulación de conocimientos o la adaptación de una

solución a circunstancias particulares. Gustilo et al. (2024) examinan las respuestas educativas ante la escritura mediada por algoritmos y sus implicaciones para la integridad. Desde esa preocupación, valorar la originalidad requiere atender al juicio humano implicado en decisiones académicas identificables y justificadas.

Las rúbricas también deben reconocer la revisión reflexiva como parte del desempeño. Un borrador corregido adquiere significado cuando las modificaciones responden a dificultades identificadas y a razones comprensibles. Registrar versiones sucesivas puede aportar evidencias útiles, aunque su cantidad no garantiza aprendizaje. Conviene valorar la calidad de las decisiones de mejora y la capacidad de explicar aquello que cambió. El criterio evita premiar una acumulación mecánica de correcciones y atiende al progreso intelectual efectivamente demostrado individualmente.

Para aplicar tales criterios, los descriptores necesitan expresar diferencias observables entre niveles de desempeño. Calificativos generales como excelente o insuficiente ofrecen poca orientación si carecen de comportamientos reconocibles. Una formulación precisa puede distinguir entre justificar una decisión mediante evidencias pertinentes y limitarse a afirmar preferencias. Gustilo et al. (2024) abordan las prácticas docentes y las políticas relacionadas con escritura algorítmica. Su objeto de estudio recuerda la importancia de traducir principios académicos en procedimientos evaluativos comprensibles.

La ponderación de los criterios debe responder a las exigencias particulares de cada actividad. En un estudio de caso puede predominar la calidad del diagnóstico, mientras una investigación documental exige mayor atención al tratamiento de fuentes. Aplicar porcentajes idénticos a tareas diferentes reduciría la sensibilidad de la rúbrica. Además, conviene evitar que la presentación lingüística absorba una proporción desmedida de la

calificación cuando el propósito principal reside en el razonamiento disciplinar demostrado por cada estudiante.

Valorar la contribución intelectual humana exige instrumentos capaces de sostener juicios explicables y abiertos a revisión. Una rúbrica bien construida permite al estudiante reconocer qué decisiones cuentan como aprendizaje y al docente fundamentar sus apreciaciones. También facilita discutir discrepancias sin depender de impresiones generales sobre la autoría. Dentro del nuevo contrato evaluativo, su función consiste en hacer visible el trabajo intelectual que permanece bajo responsabilidad humana, incluso cuando intervienen herramientas de generación automatizada de contenidos.

5.4. Evaluación personalizada sin fragmentar los estándares académicos

La personalización evaluativa adquiere un significado particular cuando la inteligencia artificial permite adaptar actividades a necesidades individuales. Sin embargo, modificar las condiciones de una tarea no debería alterar las competencias que una institución se compromete a acreditar. El problema consiste en ofrecer oportunidades diferenciadas para demostrar aprendizajes equivalentes. Una evaluación académicamente consistente necesita distinguir entre las vías utilizadas por cada estudiante y los referentes comunes que permiten interpretar su desempeño con justicia y rigor institucional.

La diversidad de trayectorias educativas justifica ofrecer alternativas en los formatos de respuesta y en las situaciones propuestas. Un estudiante puede analizar un problema mediante un informe escrito mientras otro desarrolla una exposición argumentada. Ambas opciones resultan comparables cuando exigen operaciones intelectuales realmente equivalentes y responden a criterios compartidos. La personalización pierde legitimidad si determinadas modalidades reducen

inadvertidamente la complejidad requerida. Por ello, cada alternativa necesita una revisión previa de sus demandas cognitivas y disciplinares.

La inteligencia artificial amplía las posibilidades de diseñar casos con distintos grados de familiaridad para quienes aprenden. Bower et al. (2024) examinan las opiniones docentes sobre cambios necesarios en la enseñanza y la evaluación ante sistemas generativos cada vez más potentes. Su investigación permite situar la adaptación de tareas dentro de una revisión pedagógica institucional contemporánea más amplia. La variedad de escenarios puede enriquecer la participación siempre que permanezcan constantes los resultados de aprendizaje exigidos.

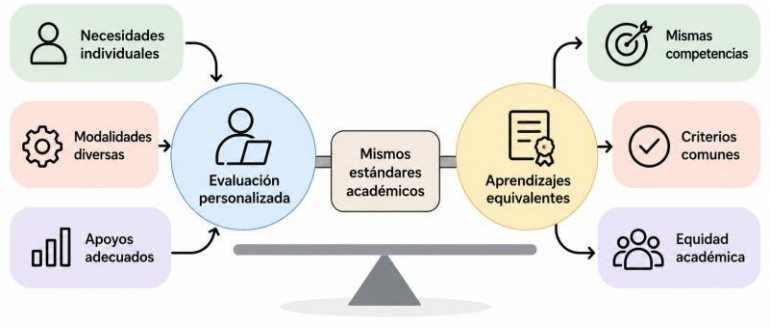
Personalizar una evaluación también implica considerar los apoyos disponibles durante su desarrollo. Algunos estudiantes necesitan instrucciones más detalladas o ejemplos preliminares para comprender qué se espera de ellos. Tales ajustes pueden favorecer el acceso a la actividad sin modificar su exigencia intelectual. Conviene distinguir el apoyo que facilita interpretar una consigna de aquel que resuelve anticipadamente el problema. La equivalencia académica depende de conservar para cada participante las decisiones fundamentales que serán objeto de valoración.

Los itinerarios diferenciados plantean una dificultad adicional cuando las actividades avanzan a ritmos distintos. Permitir nuevas oportunidades de práctica puede resultar pedagógicamente razonable, aunque las evidencias utilizadas para acreditar competencias deben conservar exigencias comparables. La rapidez de ejecución tampoco representa necesariamente un indicador de dominio académico. Un diseño cuidadoso distingue el tiempo destinado al aprendizaje del momento en que se comprueba su alcance. Esa separación evita confundir condiciones personales de trabajo con niveles de logro.

La adaptación automatizada requiere vigilancia sobre las decisiones que orientan la asignación de tareas. Un sistema puede recomendar ejercicios según respuestas anteriores, pero sus clasificaciones no deben convertirse en expectativas permanentes sobre la capacidad individual. Bower et al. (2024) recogen perspectivas docentes acerca de las transformaciones asociadas con la inteligencia artificial generativa. Tal discusión resulta pertinente para examinar la responsabilidad educativa de supervisar las recomendaciones tecnológicas y mantener abiertas oportunidades de aprendizaje exigentes para todos.

Figura 15

Personalización y equidad en la evaluación académica



La comparabilidad exige criterios de calificación suficientemente estables para interpretar productos diferentes. Una matriz común puede valorar precisión conceptual, pertinencia de las evidencias y calidad de las decisiones sin imponer idéntica forma de presentación. No obstante, los descriptores necesitan adaptarse cuidadosamente al medio elegido para evitar ventajas injustificadas. La calibración entre docentes adquiere relevancia cuando varias modalidades conviven en una asignatura. Revisar muestras de desempeño ayuda a detectar diferencias de exigencia antes de consolidar calificaciones.

También conviene atender a las condiciones materiales que intervienen en la personalización. La disponibilidad de dispositivos,

conectividad y herramientas accesibles puede determinar qué opciones resultan viables para cada estudiante. Ofrecer alternativas tecnológicas sin recursos equivalentes corre el riesgo de ampliar desigualdades preexistentes. La institución necesita prever modalidades practicables que conduzcan a los mismos aprendizajes acreditables. La flexibilidad adquiere valor cuando amplía oportunidades reales de participación efectiva sin convertir el acceso tecnológico en una ventaja evaluativa.

La validez de las adaptaciones debe examinarse mediante evidencias obtenidas durante su aplicación. Comparar resultados entre modalidades permite identificar patrones que merecen investigación, aunque las diferencias observadas no demuestren por sí mismas una ventaja indebida. Bower et al. (2024) sitúan la evaluación entre los ámbitos educativos que requieren reconsideración ante la expansión de herramientas generativas. Esa preocupación puede traducirse en procesos periódicos de revisión institucional sistemática para comprobar que las alternativas mantienen expectativas académicas compartidas.

Una evaluación personalizada necesita preservar la posibilidad de reconocer un mismo nivel de competencia mediante experiencias diferentes. El docente conserva la responsabilidad de justificar las adaptaciones y revisar sus efectos. El estudiante debe conocer los criterios que orientan su valoración con independencia del itinerario elegido. La inteligencia artificial puede facilitar la elaboración de alternativas, pero no determinar por cuenta propia el significado de una acreditación. La confianza académica depende de mantener públicamente reconocibles tales compromisos.

5.5. Integridad académica basada en evidencias y no en vigilancia

La integridad académica enfrenta una tensión creciente entre la necesidad de verificar los aprendizajes y la tentación de supervisar permanentemente al estudiante. La disponibilidad de

inteligencia artificial ha intensificado las sospechas sobre la procedencia de los trabajos presentados. Sin embargo, una evaluación centrada en detectar comportamientos irregulares puede deteriorar la confianza educativa. Resulta más pertinente construir procedimientos que permitan reconocer conocimientos demostrables y decisiones intelectuales justificadas mediante evidencias pertinentes obtenidas durante las actividades de aprendizaje.

La vigilancia tecnológica presenta limitaciones cuando pretende establecer la autoría mediante características superficiales de un texto. La fluidez lingüística, la regularidad sintáctica o determinadas estructuras argumentativas pueden encontrarse tanto en producciones humanas como automatizadas. Waltzer et al. (2024) investigan la identificación de escritura generada por inteligencia artificial en ensayos universitarios. Su trabajo permite examinar críticamente la confianza depositada en los juicios de reconocimiento y la necesidad de evitar acusaciones sustentadas en impresiones insuficientes.

Una cultura evaluativa basada en evidencias necesita definir qué manifestaciones permiten atribuir razonablemente determinados aprendizajes al estudiante. Explicar una decisión, interpretar información desconocida o defender una conclusión ante preguntas pertinentes proporciona elementos de juicio vinculados con el dominio disciplinar. Ninguna evidencia aislada garantiza por completo la autoría. Sin embargo, la convergencia de actuaciones coherentes permite construir apreciaciones mejor fundamentadas. El interés docente se desplaza hacia aquello que puede comprobarse mediante el desempeño académico observado directamente.

La documentación del proceso puede aportar información valiosa cuando responde a propósitos pedagógicos definidos. Borradores comentados, anotaciones de lectura y registros de decisiones permiten reconstruir aspectos relevantes de la

elaboración intelectual. Conviene evitar que tales materiales se conviertan en obligaciones exhaustivas que consuman tiempo sin mejorar la valoración. La evidencia adquiere utilidad cuando permite comprender una elección o reconocer una dificultad superada. Su pertinencia depende de la relación con los aprendizajes examinados en cada actividad.

El problema de atribuir autoría mediante indicios tecnológicos merece especial cautela. Waltzer et al. (2024) abordan precisamente la identificación de textos producidos por inteligencia artificial en trabajos universitarios. La investigación sitúa una cuestión relevante para las instituciones: distinguir entre sospechas razonables y pruebas suficientes para adoptar decisiones académicas. Un procedimiento justo necesita considerar explicaciones alternativas y reconocer la incertidumbre existente. La apariencia de un escrito nunca debería sustituir una valoración cuidadosa de evidencias verificables.

Las conversaciones académicas breves ofrecen oportunidades para contrastar la comprensión sin recurrir a interrogatorios intimidantes. Solicitar que un estudiante explique una elección metodológica o interprete una objeción permite observar su razonamiento en condiciones próximas al trabajo profesional. La finalidad debe quedar explícita desde el inicio para evitar que cada intercambio parezca una investigación disciplinaria. Cuando las preguntas se relacionan con criterios conocidos, la interacción fortalece la validez de la evaluación y respeta la dignidad personal.

La interpretación de evidencias exige reconocer que el desempeño humano presenta variaciones legítimas. Una persona puede escribir con precisión y experimentar dificultades al expresarse oralmente. Otra puede necesitar tiempo adicional para ordenar argumentos que comprende adecuadamente. Por ello, las discrepancias entre productos y explicaciones requieren análisis antes de atribuir conductas indebidas. La valoración responsable

considera condiciones de participación y admite oportunidades razonables para aclarar dudas sin convertir diferencias expresivas en presunciones automáticas de fraude académico.

Las instituciones necesitan procedimientos proporcionados para atender posibles incumplimientos. Una sospecha debería iniciar una revisión fundada y permitir al estudiante conocer los elementos que motivan la consulta. Waltzer et al. (2024) estudian la identificación de escritura artificial en ensayos universitarios, un asunto directamente relacionado con la prudencia probatoria. Antes de establecer responsabilidades conviene escuchar explicaciones y contrastar información pertinente. La imparcialidad exige separar los indicios iniciales de las conclusiones que puedan afectar la trayectoria académica.

La protección de datos personales forma parte de una evaluación respetuosa. Registrar cada pulsación o almacenar conversaciones privadas puede generar información abundante sin aportar pruebas pertinentes de aprendizaje. La recopilación de evidencias debería limitarse a materiales necesarios y conocidos por quienes participan. Requiere reglas claras sobre acceso, conservación y eliminación de registros. La confianza institucional se debilita cuando la supervisión invade espacios ajenos a la actividad académica o permanece activa sin una justificación educativa suficiente.

Una política de integridad orientada por evidencias transforma también el papel del docente. Su responsabilidad consiste en diseñar oportunidades de demostración y fundamentar decisiones cuando aparecen dudas razonables. El estudiante participa con derecho a explicar su trabajo y conocer los criterios aplicados. La institución debe garantizar procedimientos consistentes y revisables. Bajo tales condiciones, la integridad deja de depender de una observación permanente y encuentra respaldo en actuaciones académicas verificables, proporcionales y respetuosas de las personas.

5.6. Diseño de un sistema institucional de evaluación auténtica con IA

Diseñar un sistema institucional de evaluación auténtica con inteligencia artificial exige articular decisiones pedagógicas que habitualmente permanecen dispersas entre asignaturas y departamentos. La institución necesita establecer una orientación compartida sobre aquello que considera aprendizaje demostrable y sobre las condiciones para acreditarlo. Tal propósito requiere coordinación académica y capacidad de adaptación disciplinar. El sistema adquiere consistencia cuando sus componentes responden a criterios reconocibles y las decisiones particulares pueden justificarse ante la comunidad universitaria con argumentos educativos.

El punto de partida consiste en diagnosticar las prácticas evaluativas vigentes antes de introducir nuevas herramientas. Conviene examinar qué competencias se acreditan mediante tareas reproducibles por sistemas generativos y cuáles requieren actuaciones situadas. Xia et al. (2024) revisan la transformación de la evaluación universitaria asociada con inteligencia artificial generativa. Su aportación permite considerar ampliamente el cambio. El diagnóstico institucional debe identificar también capacidades docentes y recursos disponibles para evitar decisiones alejadas de la realidad educativa.

La arquitectura del sistema necesita distribuir responsabilidades entre órganos académicos y equipos docentes. Una instancia institucional puede definir principios generales y mecanismos de revisión, mientras las carreras precisan evidencias adecuadas para sus campos profesionales. Los profesores conservan margen para diseñar actividades pertinentes dentro de acuerdos verificables. Tal distribución evita tanto la dispersión de criterios como la imposición de procedimientos rígidos. La coordinación

requiere espacios regulares donde puedan discutirse casos y resolverse discrepancias mediante deliberación colegiada.

El diseño curricular debe conectar resultados de aprendizaje con experiencias que demanden decisiones propias del ejercicio profesional. Analizar situaciones inciertas, justificar intervenciones o revisar soluciones ante restricciones concretas ofrece evidencias distintas de la reproducción informativa. La inteligencia artificial puede intervenir en tales experiencias bajo condiciones definidas por la actividad. Lo decisivo es preservar la relación entre competencia y desempeño observable. Cada programa necesita documentar esa correspondencia para facilitar revisiones posteriores sin perder su identidad disciplinar.

La infraestructura tecnológica requiere decisiones prudentes sobre interoperabilidad, accesibilidad y protección de información. Incorporar plataformas sin examinar sus condiciones de funcionamiento puede generar dependencias costosas y dificultades para recuperar evidencias académicas. Xia et al. (2024) ofrecen una revisión del papel transformador de la inteligencia artificial generativa en la evaluación superior. Esa perspectiva invita a considerar la tecnología como parte de un sistema mayor. Su adopción debe responder a necesidades pedagógicas identificadas y contar con supervisión.

La formación docente constituye una condición operativa del sistema y requiere algo más que aprender instrucciones para aplicaciones generativas. Los equipos académicos necesitan oportunidades para diseñar tareas, contrastar criterios de valoración y examinar trabajos de estudiantes. El intercambio entre disciplinas puede revelar problemas que una capacitación exclusivamente técnica dejaría intactos. También resulta necesario reconocer tiempos de preparación y acompañamiento. Sin apoyo organizativo, las innovaciones dependen del esfuerzo individual y difícilmente alcanzan continuidad entre periodos académicos.

La participación estudiantil debe incorporarse al funcionamiento del sistema mediante canales efectivos de consulta y revisión. Conocer experiencias de aplicación permite detectar instrucciones confusas o cargas desiguales entre asignaturas. Los estudiantes pueden aportar información sobre accesibilidad y dificultades prácticas sin asumir decisiones reservadas a los órganos académicos. Una participación bien organizada fortalece la legitimidad de las reglas. También facilita comunicar modificaciones y recoger observaciones antes de extender nuevos procedimientos a toda una carrera o institución.

La implantación gradual permite examinar el funcionamiento de las propuestas antes de generalizarlas. Conviene seleccionar experiencias piloto con objetivos definidos y criterios para valorar sus resultados. Xia et al. (2024) analizan diversas transformaciones de la evaluación universitaria relacionadas con inteligencia artificial generativa. Esa amplitud temática respalda la conveniencia de observar múltiples dimensiones del cambio. Durante las pruebas interesa registrar dificultades de aplicación y efectos sobre la carga docente sin confundir aceptación tecnológica con calidad educativa.

El seguimiento institucional necesita indicadores que permitan valorar la calidad del sistema sin reducirla a cifras de utilización tecnológica. Resulta pertinente examinar la correspondencia entre tareas y competencias, la consistencia de las calificaciones y las oportunidades de participación. Las revisiones entre pares pueden complementar los datos administrativos mediante análisis de muestras académicas. Cuando aparecen diferencias persistentes, corresponde investigar sus causas. La mejora requiere decisiones documentadas y nuevas comprobaciones que permitan reconocer avances o corregir errores.

Tabla 5

Principios y componentes del nuevo contrato evaluativo entre estudiante, docente e inteligencia artificial

Componente del capítulo	Orientaciones para la evaluación académica
5.1. Transparencia del uso de IA como principio de integridad académica	Promueve la declaración responsable del uso de inteligencia artificial en las actividades académicas. Establece criterios para reconocer la participación tecnológica y preservar la autoría intelectual humana.
5.2. De la prohibición a los niveles autorizados de asistencia artificial	Propone regular la asistencia tecnológica mediante niveles diferenciados según los objetivos de aprendizaje. Define límites de utilización que permiten mantener la autonomía intelectual del estudiante.
5.3. Nuevas rúbricas para valorar la contribución intelectual humana	Plantea criterios de valoración centrados en el razonamiento crítico y la justificación de decisiones. Prioriza evidencias del trabajo intelectual humano frente a la calidad superficial de los productos generados.
5.4. Evaluación personalizada sin fragmentar los estándares académicos	Favorece la adaptación de actividades evaluativas a las necesidades individuales. Preserva criterios comunes de calidad y niveles equivalentes de exigencia para garantizar la equidad académica.
5.5. Integridad académica basada en evidencias y no en vigilancia	Orienta la verificación del aprendizaje hacia evidencias observables y procedimientos justos. Evita depender de mecanismos de detección automatizada y prácticas de supervisión invasivas.
5.6. Diseño de un sistema institucional de evaluación auténtica con IA	Articula políticas académicas y responsabilidades institucionales para integrar la inteligencia artificial en la evaluación. Considera formación docente y mecanismos de seguimiento que permitan revisar la calidad del sistema.

Nota. Elaboración propia a partir de los planteamientos desarrollados en el capítulo.

La sostenibilidad del sistema depende de convertir sus acuerdos en prácticas revisables y no en disposiciones permanentes ajenas al cambio educativo. Los órganos responsables necesitan actualizar orientaciones cuando evolucionan las capacidades tecnológicas o aparecen nuevas evidencias pedagógicas. Al mismo tiempo, deben preservar la estabilidad para que estudiantes y

docentes comprendan sus obligaciones. Un sistema institucional maduro combina continuidad normativa con aprendizaje organizativo. Su legitimidad descansa en decisiones justificadas y responsabilidades reconocidas por la comunidad académica.

Conclusiones

La evaluación del aprendizaje atraviesa una transformación que exige reconsiderar la relación entre los productos académicos y las competencias que pretenden representar. La capacidad de la inteligencia artificial para generar respuestas satisfactorias limita el valor probatorio de determinados instrumentos tradicionales. Una producción correcta no garantiza comprensión. Por ello, la evaluación necesita incorporar procedimientos que permitan reconocer la participación intelectual del estudiante y fundamentar las inferencias sobre sus conocimientos mediante evidencias de razonamiento, interpretación y desempeño verificable.

La validez de las evidencias constituye un principio determinante para renovar las prácticas evaluativas. Su reconocimiento requiere examinar la correspondencia entre las competencias previstas y las actuaciones observadas durante el aprendizaje. Los resultados adquieren significado cuando permiten identificar decisiones fundamentadas y capacidades transferibles. Desde una perspectiva académica, evaluar implica interpretar manifestaciones del conocimiento mediante criterios explícitos. La calidad de una respuesta debe examinarse junto con el proceso intelectual que permitió construirla y justificar su pertinencia disciplinar.

La evaluación auténtica ofrece posibilidades para reconocer competencias mediante actividades vinculadas con situaciones profesionales y problemas de complejidad variable. Su diseño requiere formular tareas que demanden interpretación de información y selección razonada de alternativas. La participación de herramientas generativas puede integrarse sin eliminar la responsabilidad intelectual del estudiante. El valor educativo reside en las decisiones que orientan la resolución del problema. Una tarea

pertinente permite observar capacidades de análisis y adaptación ante condiciones que requieren respuestas justificadas.

Los problemas abiertos y las simulaciones amplían las posibilidades de observar el aprendizaje mediante actuaciones que requieren movilizar conocimientos ante condiciones cambiantes. La ausencia de una respuesta predeterminada favorece la valoración de argumentos y decisiones. Sin embargo, la complejidad de una actividad no garantiza por sí misma su pertinencia evaluativa. Resulta necesario establecer criterios que permitan distinguir niveles de desempeño y reconocer diferentes soluciones fundamentadas. La autenticidad requiere correspondencia entre las exigencias de la tarea y las competencias evaluadas.

La documentación del razonamiento permite comprender la evolución de las decisiones adoptadas durante una actividad académica. Los borradores y registros de revisión proporcionan información sobre las estrategias utilizadas para construir respuestas. También permiten reconocer dificultades y modificaciones conceptuales. Las defensas orales complementan dichas evidencias mediante preguntas que requieren explicar procedimientos y justificar elecciones. La evaluación adquiere mayor profundidad cuando considera la evolución del pensamiento y establece relaciones verificables entre los procesos desarrollados y los resultados alcanzados.

La transferencia del aprendizaje representa una dimensión necesaria para valorar competencias que trascienden la reproducción de información. Su evaluación requiere introducir variaciones en los problemas y examinar la capacidad del estudiante para adaptar conocimientos previamente adquiridos. Una actuación competente implica reconocer diferencias entre situaciones y seleccionar procedimientos pertinentes. La metacognición complementa dicha valoración mediante explicaciones sobre decisiones y estrategias personales. Ambas

dimensiones permiten apreciar la autonomía intelectual y la capacidad de revisar críticamente los propios procesos de aprendizaje.

La incorporación de inteligencia artificial a los procesos evaluativos requiere delimitar las funciones correspondientes a los sistemas automatizados y al profesorado. Las herramientas generativas pueden contribuir a formular preguntas y proporcionar retroalimentación durante las actividades académicas. Su utilidad depende de la interpretación pedagógica de los resultados. El juicio docente permanece vinculado con la valoración de evidencias y la consideración de las particularidades del desempeño. Ningún procedimiento automatizado reemplaza la responsabilidad profesional de fundamentar decisiones educativas.

Las rúbricas orientadas a desempeños mediados por inteligencia artificial necesitan reconocer dimensiones que trascienden la calidad formal del producto entregado. La argumentación y la verificación de información permiten identificar aportaciones intelectuales propias. También resulta pertinente valorar las decisiones adoptadas durante las interacciones con herramientas generativas. Los criterios deben formularse mediante descriptores comprensibles y aplicables a diferentes modalidades evaluativas. Su utilización favorece la transparencia de las expectativas académicas y permite fundamentar valoraciones consistentes entre estudiantes.

La integridad académica requiere acuerdos institucionales que definan condiciones transparentes para utilizar inteligencia artificial durante las actividades formativas. Declarar las herramientas empleadas permite reconocer grados de intervención y responsabilidades individuales. Las restricciones generales resultan insuficientes cuando las tareas demandan competencias digitales diferenciadas. Por ello, conviene establecer niveles autorizados de asistencia vinculados con los resultados de aprendizaje. La integridad adquiere mayor fundamento cuando se

apoya en evidencias coherentes de participación intelectual y criterios compartidos de autoría.

La construcción de sistemas institucionales de evaluación auténtica demanda coherencia entre políticas académicas y prácticas pedagógicas. Su desarrollo requiere formación docente y procedimientos de revisión que permitan valorar la pertinencia de los instrumentos utilizados. La inteligencia artificial continuará modificando las condiciones de producción del conocimiento y las formas de demostrar competencias. Frente a dicha transformación, la responsabilidad educativa consiste en preservar el sentido formativo de la evaluación. Reconocer el aprendizaje humano exige comprender sus procesos y valorar sus manifestaciones con rigor.

Referencias Bibliográficas

Ahangama, N. (2026). Designing assessments in the generative AI era: A tailored assessment framework for ICT tertiary education. International Journal of Educational Technology in Higher Education, 23, Article 9. https://doi.org/10.1186/s41239-026-00582-0

Bower, M., Torrington, J., Lai, J. W. M., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative artificial intelligence? Outcomes of the ChatGPT teacher survey. Education and Information Technologies, 29, 15403–15439. https://doi.org/10.1007/s10639-023-12405-0

Coates, H., Croucher, G., & Calderon, A. (2025). Governing academic integrity: Ensuring the authenticity of higher thinking in the era of generative artificial intelligence. Journal of Academic Ethics, 23, 2015–2028. https://doi.org/10.1007/s10805-025-09639-7

Combrinck, C., & Loubser, N. (2025). Student self-reflection as a tool for managing GenAI use in large class assessment. Discover Education, 4, Article 72. https://doi.org/10.1007/s44217-025-00461-2

Delcker, J., Heil, J., & Ifenthaler, D. (2025). Evidence-based development of an instrument for the assessment of teachers' self-perceptions of their artificial intelligence competence. Educational Technology Research and Development, 73, 115–133. https://doi.org/10.1007/s11423-024-10418-1

Fischer, I., Sweeney, S., Lucas, M., & Gupta, N. (2024). Making sense of generative AI for assessments: Contrasting student claims and assessor evaluations. The International Journal of Management Education, 22(3), 101081. https://doi.org/10.1016/j.ijme.2024.101081

Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. Computers and Education: Artificial Intelligence, 6, 100250. https://doi.org/10.1016/j.caeai.2024.100250

Gruenhagen, J. H., Sinclair, P. M., Carroll, J.-A., Baker, P. R. A., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. Computers and Education: Artificial Intelligence, 7, 100273. https://doi.org/10.1016/j.caeai.2024.100273

Gustilo, L., Ong, E., & Lapinid, M. R. (2024). Algorithmically-driven writing and academic integrity: Exploring educators' practices, perceptions, and policies in AI era. International Journal for Educational Integrity, 20, Article 3. https://doi.org/10.1007/s40979-024-00153-8

Heil, J., Ifenthaler, D., Cooper, M., Mascia, M. L., Conti, R., & Penna, M. P. (2025). Students' perceived impact of GenAI tools on learning and assessment in higher education: The role of individual AI competence. Smart Learning Environments, 12, Article 37. https://doi.org/10.1186/s40561-025-00395-0

Katsampoxaki-Hodgetts, K. (2026). Enacted AI literacy towards epistemic visibility; Embedding human-AI design as assessment-as-learning in pre-service science teachers' lesson redesign. Research in Science Education. https://doi.org/10.1007/s11165-026-10360-7

Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S., & Raduescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. Computers and Education: Artificial Intelligence, 7, 100269. https://doi.org/10.1016/j.caeai.2024.100269

Kulgemeyer, C., Riese, J., Vogelsang, C., Buschhüter, D., Borowski, A., Weißbach, A., Jordans, M., Reinhold, P., & Schecker, H. (2023). How authenticity impacts validity: Developing a model of teacher education assessment and exploring the effects of the digitisation of assessment methods. Zeitschrift für Erziehungswissenschaft. https://doi.org/10.1007/s11618-023-01154-y

König, J., Lindmeier, A., Zlatkin-Troitschanskaia, O., & Borowski, A. (2023). The transfer of digitalized teaching-learning and assessment tools in higher education: Approaches and best practices. Zeitschrift für Erziehungswissenschaft, 26, 579–584. https://doi.org/10.1007/s11618-023-01172-w

Langer, M., Baum, K., & Schlicker, N. (2025). Effective human oversight of AI-based systems: A signal detection perspective on the detection of inaccurate and unfair outputs. Minds and Machines, 35, Article 1. https://doi.org/10.1007/s11023-024-09701-0

Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. Discover Education, 4, Article 170. https://doi.org/10.1007/s44217-025-00602-7

Nguyen, H., & Hayward, J. (2024). Applying generative artificial intelligence to critiquing science assessments. Journal of Science Education and Technology. https://doi.org/10.1007/s10956-024-10177-x

Norgaila, E., Daniela, L., & Kalniņa, D. (2026). Reflective prompt engineering for assessment rubric optimization: An empirical study of human–AI alignment. Technology, Knowledge and Learning, 31, 1023–1038. https://doi.org/10.1007/s10758-026-09979-2

Oates, A., & Johnson, D. (2025). ChatGPT in the classroom: Evaluating its role in fostering critical evaluation skills. International Journal of Artificial Intelligence in Education, 35, 1793–1824. https://doi.org/10.1007/s40593-024-00452-8

Pentland, S. J., Lowenthal, P. R., & Krier, K. (2026). Asynchronous oral assessments: Enhancing integrity, engagement, and communication in the AI era. Educational Technology Research and Development. https://doi.org/10.1007/s11423-026-10677-0

Peters, M., & Angelov, D. (2025). Redefining assessment tasks to promote students' creativity and integrity in the age of generative artificial intelligence. International Journal for Educational Integrity. https://doi.org/10.1007/s40979-025-00201-x

Tempelaar, D., Rienties, B., & Giesbers, B. (2024). Dispositional learning analytics and formative assessment: An inseparable twinship. International Journal of Educational Technology in Higher Education, 21, Article 57. https://doi.org/10.1186/s41239-024-00489-8

Torres-Diaz, J. C., Duart, J., Rivera, D., & Beltran Flandoli, A. (2025). Artificial intelligence and academic integrity: Exploring plagiarism in Ecuadorian universities. International Journal for Educational Integrity, 21, Article 35. https://doi.org/10.1007/s40979-025-00209-3

Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. Studies in Educational Evaluation, 83, 101425. https://doi.org/10.1016/j.stueduc.2024.101425

Waltzer, T., Pilegard, C., & Heyman, G. D. (2024). Can you spot the bot? Identifying AI-generated writing in college essays. International Journal for Educational Integrity. https://doi.org/10.1007/s40979-024-00158-3

Williams, A. (2024). Comparison of generative AI performance on undergraduate and postgraduate written assessments in the biomedical sciences. International Journal of Educational Technology in Higher Education. https://doi.org/10.1186/s41239-024-00485-y

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

Yang, A. C. M., Flanagan, B., & Ogata, H. (2022). Adaptive formative assessment system based on computerized adaptive testing and the learning memory cycle for personalized learning. Computers and Education: Artificial Intelligence, 3, 100104. https://doi.org/10.1016/j.caeai.2022.100104

Zhao, H., & Dang, T. N. Y. (2026). Transforming written assessment design to embrace AI: What needs to be changed to encourage higher-order critical thinking. Education and Information Technologies, 31, 1909–1931. https://doi.org/10.1007/s10639-025-13870-5

Zhuang, Y., Zhao, R., Xie, Z., & Yu, P. L. H. (2025). Enhancing language learning through generative AI feedback on picture-cued writing tasks. Computers and Education: Artificial Intelligence, 9, 100450. https://doi.org/10.1016/j.caeai.2025.100450

Red de Investigación
Científica y Desarrollo
Tecnológico **Del Pacífico**

Repensar la **evaluación** para un **aprendizaje auténtico** en la era de la inteligencia artificial.

EDITORIAL
SAGA

ISBN: 978-9907-803-96-9