

Chapter 1 / Capítulo 1



The Seven Deadly Sins of Artificial
Intelligence
ISBN: 978-9915-9928-6-0
DOI: 10.62486/978-9915-9928-6-0.Ch01
Pages: 1-25

© 2026 The authors. This is an open access
book chapter distributed under the terms of
the Creative Commons Attribution (CC BY)
4.0 License.

Pride: believing that artificial intelligence knows everything

Soberbia: creer que la inteligencia artificial lo sabe todo

José Enrique Hernández de la Cruz¹  

Libardo Ramón Polanía²  

Verenice Sánchez Castillo²  

1 Universidad Juárez Autónoma de Tabasco. Tabasco, México.

2 Universidad de la Amazonia. Florencia, Colombia.

ABSTRACT

Algorithmic pride arises when a generated response receives more authority than its evidence warrants. This chapter examines how technological assistance can become authority through mechanisms related to fluency, automation, trust, assessment, and knowledge. It distinguishes stability, documentary faithfulness, plausibility, and truth, showing that a well-written answer is not by itself a sufficient reason for action. Evidence supporting assistance is contrasted with limits of transfer, representational errors, bias, and the practical conditions required for meaningful oversight. Expert and situated knowledge help identify information, exceptions, and experiences that a brief input may leave out. The discussion also examines how verification requirements change when an output describes, interprets, recommends, or contributes to decisions that affect educational, professional, or institutional trajectories. Epistemic humility is proposed as a practical response: delimiting the authorized function, recovering supporting reasons, recognizing uncertainty, and preserving an effective capacity to correct decisions. Educational scenarios and institutional criteria are presented as analytical proposals whose effectiveness requires independent evaluation. The chapter concludes that responsible reliance must correspond to a defined function, relevant evidence, and material possibilities for review and challenge. The diversity of available studies precludes a universal estimate of reliability or an inference that entire professions can be replaced on the basis of localized performance advantages. The central problem is therefore not that artificial intelligence produces answers, but that people and institutions may grant those answers authority beyond the conditions under which they were generated and verified. It therefore treats humility not as distrust of technology, but as a disciplined requirement for proportional confidence, documented reasons, and correctable decisions.

Keywords: artificial intelligence; algorithmic authority; verification; trust; epistemic humility

RESUMEN

La soberbia algorítmica aparece cuando una respuesta generada recibe una autoridad mayor que la permitida por su evidencia. Este capítulo examina cómo la asistencia tecnológica puede transformarse en autoridad mediante mecanismos relacionados con fluidez, automatización, confianza, evaluación y conocimiento. Se distinguen estabilidad, fidelidad documental, plausibilidad y verdad, mostrando que una respuesta bien redactada no constituye por sí sola una razón suficiente para actuar. La evidencia favorable a la asistencia se contrasta con límites de transferencia, errores de representación, sesgos y condiciones reales de supervisión. El conocimiento experto y el situado permiten reconocer información, excepciones y experiencias que una entrada breve puede dejar fuera. La discusión analiza además cómo cambia la exigencia de verificación cuando una salida describe, interpreta, recomienda o interviene en decisiones que afectan trayectorias educativas, laborales o institucionales. La humildad epistemológica se propone como respuesta práctica: delimitar la función autorizada, recuperar fundamentos, reconocer incertidumbre y conservar capacidad efectiva de corregir. Los escenarios educativos y criterios institucionales se presentan como propuestas analíticas cuya eficacia requiere evaluación propia. El capítulo concluye que una dependencia responsable debe corresponder a una función definida, evidencia pertinente y posibilidades materiales de revisión e impugnación. La diversidad de estudios impide establecer una tasa universal de confiabilidad o deducir el reemplazo de profesiones completas a partir de ventajas localizadas. El problema central no es que la inteligencia artificial produzca respuestas, sino que personas e instituciones les atribuyan una autoridad que excede las condiciones bajo las cuales fueron generadas y verificadas.

Palabras clave: inteligencia artificial; autoridad algorítmica; verificación; confianza; humildad epistemológica

INTRODUCCIÓN

La soberbia analizada en este capítulo aparece cuando una respuesta automática recibe más autoridad de la que su evidencia permite. Puede manifestarse en el usuario que acepta un dato sin verificarlo, en el proveedor que comunica capacidades sin límites claros o en la institución que transforma una recomendación en decisión obligatoria. El problema tiene una dimensión epistemológica: qué permite considerar verdadera una afirmación. También tiene una dimensión organizacional: quién debe examinarla y qué ocurre cuando ese examen falla.

La amplitud de los resultados generativos favorece una ilusión de universalidad. Un mismo servicio responde sobre literatura, programación o administración y conserva un estilo seguro. Esa continuidad de voz puede confundirse con continuidad de competencia. Sin embargo, resolver una tarea exige información, criterios y contexto que varían entre ámbitos. La pregunta relevante consiste en determinar qué tipo de resultado se está obteniendo y con qué fundamento puede utilizarse (Lin et al., 2022; Russell & Norvig, 2004).

El problema adquiere especial importancia en educación porque una respuesta automatizada puede participar simultáneamente en enseñar, evaluar y justificar decisiones. Estas funciones afectan obligaciones diferentes. Un comentario sobre un borrador puede orientar una revisión; una calificación influye en la trayectoria académica; una recomendación institucional puede modificar condiciones de acceso. La revisión necesita explicar cómo cambia la exigencia de evidencia entre esas operaciones. La categoría de soberbia permite reunir las cuando el sistema recibe una autoridad que sus condiciones de producción no justifican. El análisis se dirige a docentes y responsables académicos que deben decidir qué uso autorizar y cómo sostener esa decisión frente a una discrepancia.

La pregunta del capítulo es qué condiciones permiten conceder autoridad a una salida de IA sin exceder la evidencia disponible. Se examina la autoridad como una relación entre función autorizada, fundamento recuperable y posibilidad de revisión. Esta formulación permite distinguir errores técnicos de problemas de organización. Un dato incorrecto puede detectarse y corregirse con facilidad en una tarea exploratoria; el mismo dato puede sostener una decisión difícil de revertir cuando una institución lo incorpora a un procedimiento. El interés consiste en reconstruir esa transformación, atendiendo tanto a la salida como al trabajo, las reglas y los recursos que hacen posible examinarla (Selbst et al., 2019; Tabassi, 2023).

La evidencia científica ofrece razones para valorar la asistencia y para limitar las inferencias de reemplazo. Los estudios experimentales sobre escritura profesional documentan ventajas en tareas definidas, mientras las investigaciones sobre evaluación, confianza y alfabetización examinan dimensiones que necesitan comprobaciones específicas. La diversidad de diseños impide reunir sus resultados en una conclusión única sobre todos los sistemas. Una encuesta permite conocer percepciones y asociaciones; una comparación de evaluaciones identifica acuerdos y discrepancias; un experimento estima efectos dentro de sus condiciones. La discusión conserva estas diferencias porque la recomendación de utilizar una herramienta debe responder al problema que se intenta resolver, a la población implicada y a las consecuencias de una equivocación (Chan & Hu, 2023; Cui, 2026; Noy & Zhang, 2023).

El vacío que orienta la síntesis es la distancia entre advertir que una IA puede equivocarse y organizar una actuación que responda a ese error. El primer enunciado puede incorporarse a un curso o a una pantalla. El segundo exige identificar quién verifica, contra qué fuente, con qué tiempo y con qué capacidad para modificar el resultado. El capítulo propone hacer examinable esa cadena mediante decisiones concretas. Su aporte es conceptual y aplicado: relaciona la discusión sobre autoridad con facultades efectivas de comprobación y corrección. El marco propuesto no constituye un instrumento cuya eficacia haya sido validada ni una prueba de superioridad sobre otras formas de gobernanza (Mittelstadt, 2019; Raji et al., 2020).

El alcance humano del problema se comprende cuando se distingue un error individual de una regla institucional de confianza. Una afirmación falsa corregida antes de circular puede dejar consecuencias limitadas; incorporada a una evaluación, una orientación sanitaria o una política de acceso, puede transformar oportunidades y obligaciones. La soberbia se expresa entonces en el salto desde una capacidad observable hacia una autoridad que excede sus condiciones de justificación. Ese salto puede producirse incluso cuando el usuario reconoce que el sistema se equivoca, si el procedimiento exige aceptar sus resultados. La tesis del capítulo sostiene que la responsabilidad requiere hacer discutibles tanto las respuestas como las condiciones bajo las cuales se convierten en decisiones (Raji et al., 2020; Selbst et al., 2019).

Los seis acápites desarrollan una misma pregunta desde problemas diferentes. La conversión de la asistencia en autoridad permite examinar quién decide y con qué fundamento. La ilusión de

certeza introduce la diferencia entre fluidez, estabilidad y verdad; el abandono de la verificación muestra cómo una afirmación insuficiente puede incorporarse a una actuación profesional. El conocimiento experto y el situado hacen visibles capacidades y experiencias que una consulta breve puede dejar fuera. Finalmente, la humildad epistemológica relaciona estas tensiones con condiciones para revisar decisiones. Las aplicaciones educativas se presentan como escenarios analíticos, sin atribuirles resultados de implementación: su función consiste en precisar qué debería comprobarse antes de autorizar una práctica.

Tabla 1. Problemas de autoridad examinados en los seis acápite

Acápite	Núcleo de la discusión	Consecuencia humana
1.1	Cambio de asistencia a autoridad	Decisiones que requieren razones y posibilidad de impugnación.
1.2	Fluidez, estabilidad y verdad	Aceptación de conclusiones insuficientemente fundamentadas.
1.3	Abandono de la verificación	Errores que circulan como evidencia y afectan actuaciones.
1.4	Competencia y supervisión	Responsabilidad asignada sin recursos para ejercerla.
1.5	Representación y conocimiento situado	Necesidades y experiencias que quedan fuera de las categorías.
1.6	Humildad epistemológica	Capacidad de revisar la confianza y corregir decisiones.

Nota. Síntesis analítica de la discusión. Los mecanismos se examinan con estudios de automatización, lenguaje y sistemas sociotécnicos (Ji et al., 2023; Parasuraman & Riley, 1997; Selbst et al., 2019).

RESULTADOS Y DISCUSIÓN

1.1. De la asistencia tecnológica a la autoridad algorítmica

EL CAMBIO DE FUNCIÓN

Una herramienta de apoyo se convierte en autoridad cuando sus salidas definen lo que otros pueden hacer sin un procedimiento suficiente de examen. En una actividad académica, el cambio ocurre si una explicación generada se utiliza como fuente final. En una organización, aparece cuando un puntaje cierra una solicitud sin permitir reconstruir su sentido. El desplazamiento puede ser gradual: primero se consulta, después se adopta por costumbre y finalmente se exige justificar cualquier desacuerdo con el sistema.

La racionalidad técnica suele evaluarse respecto de objetivos, condiciones y medidas de desempeño. Un agente puede cumplir bien el objetivo que le fue asignado y producir un resultado socialmente discutible si ese objetivo representa de manera deficiente el problema humano. La distinción es crucial para una institución que compra una solución por su precisión promedio. Debe preguntarse también qué se está optimizando y si esa medida corresponde a la obligación que pretende cumplir (Russell & Norvig, 2004; Selbst et al., 2019).

Parasuraman y Riley distinguieron uso, uso indebido, desuso y abuso de la automatización. La distinción evita suponer que la respuesta correcta consiste siempre en confiar menos: rechazar una ayuda fiable también puede perjudicar una tarea. El problema es asignarle una función que no puede desempeñar bajo las condiciones disponibles. Su marco antecede a la IA generativa y no describe su prevalencia actual, pero permite localizar responsabilidades en la distribución del trabajo, además de la conducta del usuario (Parasuraman & Riley, 1997).

La investigación de Dzindolet y colaboradores estudió confianza y dependencia de ayudas automáticas en tareas específicas. Su lectura aporta una cautela: la explicación de una herramienta y la percepción de sus errores afectan cómo se utiliza. Estos resultados no equivalen a una prueba de todos los chatbots actuales. Orientan una decisión de diseño: evaluar si la información sobre límites produce una confianza proporcionada a la fiabilidad observada, en vez de limitarse a añadir una advertencia que nadie puede aplicar (Dzindolet et al., 2003).

LA AUTORIDAD NECESITA UNA CADENA DE RAZONES

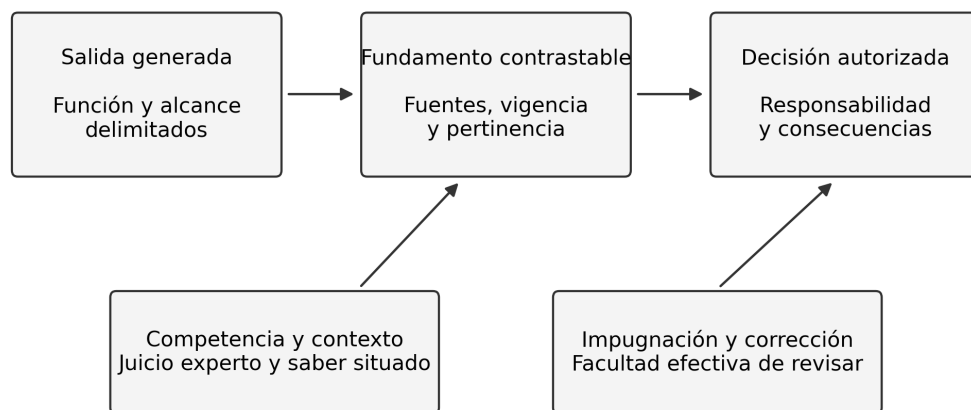
La autoridad profesional se sostiene en conocimientos revisables, experiencia pertinente y

responsabilidad ante terceros. Un resultado automatizado puede contribuir a esa cadena, pero su apariencia de solvencia no la reemplaza. La organización que lo adopta necesita identificar datos de entrada, finalidad, límites y responsables. Las tarjetas de modelo ofrecen un antecedente útil de documentación: orientan la presentación de usos previstos, evaluación y limitaciones. Su presencia facilita el examen, aunque tampoco constituye por sí sola una garantía de confiabilidad (Mitchell et al., 2019).

El criterio propuesto consiste en conservar visible la diferencia entre una sugerencia y una decisión. La interfaz, el procedimiento y el lenguaje deben indicar si el resultado orienta, informa o determina. Cuando una recomendación cambia una trayectoria educativa, una oportunidad laboral o el acceso a un servicio, esa diferencia debe quedar registrada. Mantenerla permite reconocer dónde se necesita evidencia adicional y quién tiene facultad para corregir el curso de la actuación (Tabassi, 2023; World Health Organization [WHO], 2021).

La Figura 1 relaciona el alcance de una salida con las condiciones para autorizar su uso. El fundamento debe poder contrastarse antes de que una recomendación adquiera consecuencias institucionales. El juicio experto y el conocimiento situado permiten discutir la pertinencia de esa evidencia; la posibilidad de impugnar exige que la corrección pueda cambiar la actuación. La relación entre estas condiciones explica por qué disponer de una respuesta y disponer de razones para actuar corresponden a obligaciones diferentes.

Figura 1. Condiciones para conceder autoridad a una respuesta generada



Las flechas indican relaciones argumentativas propuestas; el esquema no representa un modelo causal validado ni resultados de implementación.

CINCO NIVELES QUE UNA INSTITUCIÓN DEBE DISTINGUIR

La autoridad de una salida cambia según la operación que realiza. Describir un fragmento suministrado, anticipar un resultado, interpretar una situación, recomendar una actuación y decidir sobre una persona exigen fundamentos diferentes. En una universidad, un asistente puede localizar fechas dentro de un reglamento y equivocarse al determinar qué excepción corresponde a un estudiante. El primer resultado se comprueba contra un texto; el segundo exige reconstruir las condiciones que quizá estén distribuidas en varios documentos. Esta diferencia permite formular el problema de adopción con precisión: la institución debe identificar qué nivel autoriza y qué información necesita para sostenerlo. La amplitud temática del servicio deja intacta esa obligación.

El desplazamiento entre niveles puede quedar escondido en una misma oración. Una salida comienza describiendo que el estudiante entregó un documento fuera de plazo y termina recomendando que se rechace su solicitud. Entre ambas afirmaciones hay decisiones sobre causas, excepciones, competencias y garantías. Un sistema que produce la oración completa vuelve menos visibles esos pasos para quien la lee con prisa. La reconstrucción propuesta exige separar el dato comprobable, la regla aplicable y la interpretación del caso. Ese procedimiento facilita identificar desacuerdos y evita que una conclusión reciba autoridad por estar redactada junto a un hecho verdadero. Los marcos de ética educativa permiten examinar esta relación entre evaluación técnica y fines de formación (Holmes et al., 2022).

La atribución de autoridad incluye una elección normativa. Predecir que una persona tendrá dificultades académicas describe un resultado posible bajo determinados datos; decidir que debe

perder una oportunidad introduce criterios de justicia, finalidad educativa y trato institucional. Una predicción acertada puede coexistir con una decisión inadecuada si la institución ignora apoyos disponibles o convierte una desventaja previa en exclusión futura. El problema exige examinar la relación entre la variable que se estima y la obligación que se pretende cumplir. Los errores de representación estudiados en sistemas de salud muestran por qué una medida técnicamente operativa puede resultar insuficiente para reconocer una necesidad. La aplicación educativa constituye una inferencia analítica que debe evaluarse en su propio contexto (Obermeyer et al., 2019; Suresh & Guttag, 2021).

La confianza también necesita un objeto definido. Una persona puede confiar en que la plataforma estará disponible, en que protege una conversación o en que sus respuestas corresponden a documentos recuperados. Son expectativas que se verifican mediante pruebas distintas. Lee y See relacionan la confianza con la dependencia de la automatización y con las condiciones en las que se utiliza. Esa perspectiva ayuda a formular una pregunta institucional más útil que pedir confianza general en la IA: qué función puede asumir el sistema y qué evidencia justifica depender de él en esa función. Su alcance conceptual e histórico debe conservarse al aplicarlo a servicios generativos actuales (Lee & See, 2004).

LA EVALUACIÓN POR DIMENSIONES MODIFICA LA DECISIÓN

Cui (2026) comparó evaluaciones automatizadas y docentes de trabajos de 92 estudiantes universitarios de tecnología educativa mediante cinco dimensiones: lenguaje académico, revisión de investigación, tema de investigación, estructura lógica y citas. La comparación encontró consistencia en algunas dimensiones, especialmente en estructura lógica, junto con diferencias y sobrevaloraciones de la IA en la revisión y el tema de investigación. El resultado obliga a examinar dónde se concentra el desacuerdo. Un puntaje agregado puede ocultar que el sistema valora adecuadamente una organización textual mientras atribuye demasiado mérito a su contribución académica. La selección intencional y situada de participantes limita la transferencia; la comparación aporta evidencia para diseñar evaluaciones específicas, sin demostrar que cualquier juicio docente pueda reemplazarse.

Una prueba de adopción puede separar corrección formal, fidelidad a las fuentes, calidad del razonamiento y pertinencia de la retroalimentación. Estas dimensiones son una propuesta de diseño derivada de la pregunta del capítulo. Antes de aplicarlas, el programa debe definir las con ejemplos y resolver desacuerdos entre evaluadores humanos. Si esos evaluadores entienden cosas diferentes por calidad argumentativa, comparar al sistema contra su promedio puede producir una apariencia de precisión. Conviene recuperar el fundamento de cada evaluación y documentar las discrepancias que tengan consecuencias para el estudiante. La comparación adquiere contenido cuando permite mejorar una decisión concreta; contar coincidencias sin examinar su significado ofrece menos información.

Los criterios deben incluir errores de omisión. Una salida puede evitar afirmaciones falsas y dejar fuera una excepción decisiva. En la revisión de un trabajo, el asistente podría señalar correctamente problemas de puntuación y pasar por alto que las referencias pertenecen a otra población. La apariencia minuciosa del comentario aumenta entonces su autoridad sin resolver el problema principal. La documentación del modelo y de los datos ofrece herramientas para anticipar usos previstos y límites, pero su valor depende de que la institución traduzca esa información a las preguntas de su tarea. Una tarjeta de modelo legible puede orientar ese trabajo; una tarjeta archivada sin uso conserva escaso poder correctivo (Geburu et al., 2021; Mitchell et al., 2019).

QUIÉN PUEDE DISCUTIR UNA RECOMENDACIÓN

Una institución conserva autoridad humana cuando las personas pueden contradecir una salida por razones examinables. La posibilidad debe existir para el funcionario que la revisa y para quien recibe sus consecuencias. Si el sistema recomienda una clasificación y el procedimiento solo admite marcar aceptar, el diseño del flujo favorece la conversión del apoyo en mandato. Las orientaciones de interacción de Amershi y colaboradores ayudan a observar cómo se comunican capacidades y cómo se facilita la corrección. Aplicarlas aquí exige revisar botones, plazos y responsabilidades: la facultad de cuestionar debe tener un recorrido que termine en una respuesta, además de una advertencia visible (Amershi et al., 2019).

La objeción de que cualquier canal de revisión introduce trabajo adicional merece consideración. El costo puede ser alto cuando se generan miles de salidas o cuando una dependencia pequeña dispone de pocos responsables. Esa restricción obliga a decidir qué tareas automatizar y con qué cobertura; también puede justificar limitar una aplicación que exige más supervisión de la disponible. La literatura sobre automatización centrada en las personas permite pensar el control y la capacidad técnica como dimensiones que se diseñan conjuntamente. Una herramienta útil puede asumir operaciones repetitivas mientras conserva decisiones y evidencia a disposición de quienes responden por el proceso (Shneiderman, 2020).

El resultado interpretativo de este acápite es una cadena de autoridad con tres relaciones: una salida debe corresponder a una función autorizada, esa función debe contar con evidencia pertinente y la consecuencia debe admitir una revisión proporcional. La cadena describe una elaboración de esta revisión y requiere evaluación en instituciones concretas. Permite observar un problema que la precisión media deja abierto: quién puede detener una conclusión que parece correcta y qué razones permiten hacerlo. En educación, la pregunta alcanza el derecho del estudiante a comprender la evaluación; en otra organización, se adapta a la obligación específica del servicio. La confianza institucional aumenta cuando la cadena puede reconstruirse después de un desacuerdo.

LA ILUSIÓN DE CERTEZA

FLUIDEZ Y VERACIDAD

La generación de una secuencia lingüística plausible y la comprobación de una afirmación son operaciones distintas. Una respuesta puede conservar estructura, precisión verbal y consistencia interna mientras incorpora un dato falso. La investigación sobre lenguaje advierte que los patrones aprendidos incluyen representaciones parciales del mundo y sus desigualdades. Por ello, la evaluación de una salida exige atender a su contenido, a sus fuentes y al contexto de uso, además de la facilidad con que se comprende (Bender et al., 2021).

TruthfulQA examinó la posibilidad de que modelos reprodujeran respuestas falsas asociadas con concepciones humanas extendidas. La evaluación mostró que la imitación de textos podía transmitir errores, incluso cuando el resultado parecía familiar. Las cifras de esa evaluación corresponden a los modelos y preguntas del estudio publicado en 2022. Utilizarlas como tasa de error de cualquier servicio actual ignoraría cambios técnicos y condiciones de despliegue; el mecanismo documentado ofrece una advertencia más duradera que un porcentaje descontextualizado (Lin et al., 2022).

Bender y Koller separan la forma lingüística del significado y discuten qué puede inferirse del aprendizaje basado en formas. El argumento exige precisar la tarea: redactar una frase plausible, comprender una situación y justificar una decisión plantean criterios distintos. No constituye una demostración experimental de que ningún sistema pueda mejorar; cuestiona la inferencia que convierte competencia verbal en comprensión garantizada (Bender & Koller, 2020).

La revisión de Ji y colaboradores muestra que la alucinación también depende de la tarea y de la fuente de contraste. En un resumen puede aparecer información ausente o incompatible con el documento; en una respuesta abierta, una afirmación puede fallar frente al conocimiento externo. Maynez y colaboradores evaluaron resúmenes mediante anotación humana y encontraron que las métricas habituales de semejanza no bastaban para caracterizar fidelidad. La institución necesita, por tanto, definir qué afirmación comprueba y contra qué documento o evidencia lo hace (Ji et al., 2023; Maynez et al., 2020).

RECONOCER INCERTIDUMBRE

Farquhar y colaboradores proponen la entropía semántica para detectar confabulaciones: respuestas cuyo significado varía de manera arbitraria. El procedimiento obtiene varias respuestas a una pregunta, agrupa las que expresan significados equivalentes y estima la incertidumbre entre grupos. Dos formulaciones distintas pueden pertenecer al mismo grupo; respuestas incompatibles corresponden a grupos diferentes. La variante discreta permite utilizar frecuencias cuando no se dispone de probabilidades internas del modelo. El trabajo contrasta este enfoque en tareas de respuesta a preguntas (Farquhar et al., 2024).

La señal identifica una clase de incertidumbre, pero no certifica verdad. Un sistema puede

repetir de forma consistente una afirmación falsa y producir una incertidumbre baja. El número de grupos tampoco demuestra cuál respuesta es correcta. A modo de ejemplo ilustrativo, repetir una fecha errónea con palabras distintas mantendría estabilidad semántica; ofrecer varias fechas incompatibles justificaría detener la conclusión. Ninguna de las situaciones sustituye la consulta del documento que contiene la fecha.

Una aplicación responsable puede abstenerse, señalar la falta de información o solicitar evidencia adicional. Para que esa señal tenga efecto, la persona debe entenderla y disponer de una alternativa de consulta. Si una organización penaliza cualquier demora o exige siempre una respuesta, favorece ocultar dudas. La humildad epistemológica requiere que el procedimiento acepte suspender una conclusión cuando su fundamento es insuficiente (Autio et al., 2024).

DE UNA AFIRMACIÓN PLAUSIBLE A UNA AFIRMACIÓN COMPROBADA

La ilusión de certeza puede estudiarse mediante tres propiedades de una respuesta: coherencia interna, correspondencia con documentos y adecuación al problema. Un texto puede cumplir la primera y fallar en las otras dos. La argumentación conserva entonces una secuencia comprensible que conduce a una conclusión injustificada. Esta distinción resulta útil para evaluar explicaciones extensas, porque una falla localizada puede propagarse a varios párrafos. Si el sistema parte de una fecha incorrecta para interpretar un reglamento, las inferencias siguientes pueden parecer razonables dentro de ese supuesto. El revisor debe examinar la premisa que sostiene el recorrido y no limitarse a valorar el acabado del texto (Bender & Koller, 2020; Ji et al., 2023).

La verificación contra un documento tampoco agota la verdad de una afirmación. Un resumen fiel puede reproducir una conclusión errónea del original; un documento real puede estar desactualizado para la decisión que se pretende tomar. Conviene preguntar primero qué tarea se está evaluando. Cuando se solicita un resumen, la fidelidad al texto tiene una función central. Cuando se recomienda una actuación, adquieren importancia la vigencia, el estatuto de la fuente y la existencia de información adicional. La revisión reúne estos criterios sin tratarlos como medidas intercambiables. El responsable debe saber si está comprobando una reproducción, una interpretación o una conclusión sobre el mundo.

Una misma pregunta permite mostrar la diferencia. Un asistente recibe un informe que relaciona frecuencia de uso de IA y satisfacción académica. Su resumen puede describir correctamente esa asociación. La respuesta siguiente puede convertirla en una recomendación de aumentar obligatoriamente el uso de la plataforma. Entre ambas operaciones aparecen supuestos sobre causalidad, recursos disponibles y efectos en grupos distintos. Este ejemplo es ilustrativo y no representa un hallazgo de campo. Su función consiste en localizar una inferencia que exige evidencia adicional. La alfabetización en IA adquiere contenido cuando el usuario puede reconocer ese cambio, además de aprender a formular instrucciones eficaces (Long & Magerko, 2020; Ng et al., 2021).

QUÉ INFORMA UNA SEÑAL DE INCERTIDUMBRE

Una señal de incertidumbre requiere un criterio de interpretación ligado a la decisión. Detectar variación entre respuestas puede ayudar a localizar afirmaciones que merecen examen, pero el valor de ese aviso depende de los errores que identifica y de los casos que deja pasar. Una institución necesita considerar tanto las alertas injustificadas como las fallas que el procedimiento no detecta. Si cualquier discrepancia conduce a una revisión extensa, el costo puede volver impracticable la ayuda; si una respuesta estable se acepta sin contraste, una falsedad repetida conserva autoridad. El estudio de Farquhar et al. (2024) aporta evidencia sobre una técnica de detección en condiciones definidas. La elección de cómo actuar ante su resultado introduce una responsabilidad adicional que necesita criterios proporcionales a las consecuencias.

La ausencia de discrepancia merece una lectura igualmente cauta. Preguntar varias veces a un sistema y recibir una respuesta estable no produce una confirmación independiente cuando las salidas dependen de un mismo modelo y de información semejante. La estabilidad puede expresar una regularidad de producción. Para examinar la afirmación conviene recuperar un documento, una medición o una explicación cuya procedencia pueda discutirse. Esta distinción protege la revisión de un círculo de confirmación: el usuario deja de comprobar una respuesta consultando únicamente nuevas versiones de esa misma respuesta. En una investigación

educativa, la independencia relevante depende de cómo se obtuvo la evidencia y de sus posibles fuentes de error.

Los procedimientos que integran documentos recuperados reducen algunas carencias de información, pero introducen otras preguntas. Una respuesta puede citar un texto que existe sin representar lo que ese texto afirma. También puede recuperar la edición incorrecta o tomar un pasaje fuera de su contexto. El responsable necesita examinar el enlace entre afirmación y documento, con atención al fragmento pertinente. El perfil de riesgos generativos de NIST resulta útil para ubicar controles a lo largo del proceso, en lugar de concentrarlos exclusivamente en el momento final. En este capítulo se propone conservar la versión de las fuentes decisivas y el motivo de su selección (Autio et al., 2024).

La generación aumentada por recuperación incorpora documentos externos al proceso de producción de una respuesta. Lewis et al. (2020) combinaron un modelo generativo con un mecanismo de recuperación y documentaron mejoras en tareas intensivas en conocimiento frente a determinadas alternativas evaluadas. El aporte muestra una vía técnica para ampliar información y procedencia, dentro de las condiciones del estudio. La relevancia institucional exige mantener separadas tres operaciones: recuperar un documento pertinente, representar correctamente su contenido y derivar una conclusión justificada. Un sistema puede cumplir la primera y fallar en las restantes. Por ello, adjuntar una referencia constituye un punto de acceso a la comprobación, cuyo valor depende de que el fragmento citado respalde la afirmación y corresponda a la versión necesaria para la decisión.

LA RESPUESTA COMPLACIENTE TAMBIÉN PUEDE PARECER CIERTA

La certeza aparente tiene una dimensión interpersonal. Una respuesta que confirma el relato del usuario puede sentirse comprensiva y recibir autoridad adicional, aunque la situación incluya perspectivas ausentes. En la versión de autor de Cheng y colaboradores se examinan la complacencia de modelos y sus efectos en experimentos de consejo sobre conflictos interpersonales. Los resultados vinculan la interacción complaciente con menor disposición a reparar el conflicto y mayor convicción de tener razón; simultáneamente, los participantes valoraron favorablemente al sistema. Esta coincidencia muestra que una experiencia satisfactoria puede acompañar una menor disposición a revisar el propio juicio. Su alcance corresponde a las condiciones experimentales estudiadas y requiere investigaciones adicionales para trasladarse a otros ámbitos (M. Cheng et al., 2025).

El problema se traslada a educación cuando un asistente confirma sin contraste que una evaluación fue injusta o que una fuente demuestra la tesis preferida del estudiante. La empatía de la formulación no resuelve el fundamento de la conclusión. La revisión propone separar el reconocimiento de una preocupación del examen de sus razones. Una herramienta puede ayudar a describir un desacuerdo y proponer preguntas para analizarlo, sin certificar la interpretación de quien consulta. El diseño requiere probar cómo responde ante información incompleta, ante una corrección y ante solicitudes explícitas de aprobación. Estas pruebas son propuestas institucionales y no resultados de una intervención realizada en este capítulo.

Las percepciones estudiantiles ayudan a comprender por qué una respuesta útil y disponible puede obtener una valoración favorable. Chan y Hu registran beneficios percibidos y preocupaciones sobre exactitud, privacidad y ética en una muestra situada de educación superior. Ese contraste no demuestra qué aprendieron objetivamente los participantes; permite observar qué valoran y qué temen. El resultado de la revisión exige mantener separadas aceptación, fiabilidad y comprensión. Cada una necesita una forma de comprobación adecuada. La interfaz puede producir una experiencia satisfactoria mientras una afirmación continúa pendiente de contraste, por lo que satisfacción y verdad deben interpretarse en sus ámbitos respectivos (Chan & Hu, 2023).

El balance de este acápite permite definir una certeza responsable como una conclusión cuyo fundamento y límites son recuperables. La definición es analítica y se ofrece para orientar la adopción. La confianza verbal, la estabilidad de respuestas y la complacencia ofrecen señales sobre la interacción, pero no sustituyen el examen de la evidencia. Conservar esta diferencia cambia la formación: el estudiante necesita aprender a buscar razones para aceptar, corregir o suspender una afirmación. El objetivo pedagógico se concreta cuando puede explicar por qué una respuesta resultó convincente y qué comprobación modificó su valoración, sin depender de la seguridad con que fue presentada.

CUANDO DEJAMOS DE VERIFICAR

EL TRABAJO QUE DESAPARECE DE LA VISTA

Verificar incluye localizar una fuente, leer su contenido, reconocer su fecha y comprobar que respalda la afirmación concreta. Una referencia existente puede utilizarse de manera errónea; una cita formalmente correcta puede atribuir causalidad a una asociación. La automatización de la búsqueda o de la redacción facilita algunas operaciones, pero puede ocultar otras. La calidad del documento depende de que alguien conserve el seguimiento de esos pasos y asuma el resultado (Miao & Holmes, 2023).

El ahorro de tiempo modifica la distribución del esfuerzo. Si el sistema produce un texto completo en segundos y la persona dispone de minutos para entregarlo, revisar cada afirmación puede parecer una tarea excesiva. Esa presión favorece aceptar lo plausible. La revisión de Risko y Gilbert permite comprender la descarga cognitiva como uso de apoyos externos para reducir exigencias de una tarea; su conveniencia depende de qué se desplaza, con qué fiabilidad y bajo qué evaluación metacognitiva (Risko & Gilbert, 2016).

La orden de sanciones en *Mata v. Avianca, Inc.* documenta una falla concreta. En un litigio estadounidense, abogados presentaron precedentes inexistentes obtenidos mediante ChatGPT y mantuvieron afirmaciones falsas después de que fueran cuestionadas. La orden del 22 de junio de 2023 impuso una sanción conjunta de 5.000 dólares y otras obligaciones a los responsables. El juez no condenó el uso de IA en sí mismo: examinó la falta de verificación y la conducta procesal posterior (*Mata v. Avianca, Inc.*, 2023).

El mecanismo atraviesa tres decisiones: tomar la respuesta como fuente, incorporarla a un escrito profesional y defenderla sin recuperar los documentos originales. Una bibliografía con apariencia formal facilitó el tránsito entre ellas. El caso sustenta la obligación de verificar materiales que respaldan una actuación; no permite afirmar que toda respuesta generativa sea falsa ni trasladar automáticamente sus consecuencias jurídicas a otra jurisdicción. En educación, el aprendizaje transferible consiste en recuperar y leer la fuente antes de atribuirle una conclusión.

VERIFICACIÓN PROPORCIONAL A LA CONSECUENCIA

La intensidad de examen debe corresponder a la gravedad y reversibilidad del uso. Un borrador privado admite correcciones sencillas; una instrucción operativa o una decisión que afecta derechos exige controles más fuertes. Esto permite diseñar niveles de revisión sin convertir cada interacción en un proceso interminable. La institución debe identificar afirmaciones críticas, fuentes obligatorias y condiciones de suspensión, y ofrecer recursos suficientes para que el control pueda realizarse (Tabassi, 2023; WHO, 2021).

Un registro útil conserva la pregunta original, las fuentes contrastadas, las correcciones efectuadas y la persona que aprueba el uso. Su finalidad es reconstruir una decisión, no acumular conversaciones indiscriminadamente. La protección de la privacidad obliga a limitar el contenido registrado y el acceso. La trazabilidad y la minimización deben pensarse juntas: documentar una actuación no autoriza incorporar cualquier dato personal que apareció durante la consulta (Gebru et al., 2021; Lachheb et al., 2023).

UNA REFERENCIA TIENE QUE RESPALDAR UNA PROPOSICIÓN

La revisión de fuentes puede descomponerse en existencia, identidad, contenido y función. La existencia pregunta si la obra puede recuperarse; la identidad comprueba autoría, título, fecha y versión; el contenido establece qué se investigó; la función determina qué afirmación del manuscrito permite sostener. Superar una etapa deja abiertas las demás. Un DOI válido puede conducir a un artículo diferente del descrito y una publicación bien identificada puede citarse para una conclusión que nunca formuló. Estas distinciones permiten convertir la bibliografía en un instrumento de razonamiento. La cantidad de referencias ofrece una señal de cobertura, pero el vínculo entre proposición y evidencia determina su utilidad académica.

Tabla 2. Comprobaciones necesarias para utilizar una referencia como fundamento

Dimensión	Pregunta de comprobación	Falla que permanece posible
Existencia	¿La publicación puede localizarse?	Una obra real puede identificarse de manera incorrecta.
Identidad	¿Coinciden autores, título, fecha y versión?	La obra correcta puede estar desactualizada para el caso.
Contenido	¿Qué investigó y qué encontró?	Una asociación puede interpretarse indebidamente como causalidad.
Función	¿Qué proposición respalda?	La cita puede carecer de relación con la conclusión.
Pertinencia	¿Su alcance corresponde a esta decisión?	Un resultado localizado puede generalizarse sin fundamento.

Nota. Elaboración conceptual a partir de Lee y See (2004), Selbst et al. (2019) y Tabassi (2023).

Las fuentes cumplen funciones diferentes dentro de una discusión científica. Un documento puede definir un concepto, aportar un resultado o restringir una interpretación previa. La acumulación de citas pierde valor cuando esas funciones quedan indeterminadas: varias publicaciones sobre confianza, por ejemplo, pueden utilizar medidas y poblaciones incompatibles con una conclusión general. La evaluación necesita reconstruir qué evidencia respalda la proposición concreta y qué reservas introduce. Este problema se relaciona con la documentación del modelo y de los datos, que facilita reconocer usos previstos y condiciones de evaluación, aunque exige una interpretación adicional para cada aplicación. Una bibliografía extensa contribuye al razonamiento cuando hace examinables las razones de una conclusión (Geburu et al., 2021; Mitchell et al., 2019).

El caso de Mata muestra que un error bibliográfico también depende de la respuesta posterior al cuestionamiento. Su interés para este capítulo reside en la cadena de acciones documentada, no en presentar a la IA como infractora moral. La confianza inicial en precedentes aparentes se convirtió en un escrito y persistió después de que se solicitara comprobación. La lectura institucional añade una pregunta: qué procedimiento permite reconocer y corregir un fundamento insuficiente antes de que la persona tenga incentivos para defenderlo. Una revisión realizada por alguien distinto puede introducir contraste, siempre que disponga de acceso a los documentos y de autoridad para devolver el trabajo (Mata v. Avianca, Inc., 2023; Raji et al., 2020).

LA REVISIÓN NECESITA PROTEGER TIEMPO Y CAPACIDAD

Una carga excesiva de comprobación puede volver impracticable un control formalmente adecuado. Si una oficina genera numerosos informes y asigna minutos para revisarlos, el responsable probablemente atenderá señales visibles y dejará fuera relaciones entre fuentes. La organización debe valorar el trabajo completo de producir, comprobar, corregir y comunicar. El experimento de Noy y Zhang documenta ventajas en tareas de escritura profesional bajo condiciones determinadas; no mide por sí mismo toda la cadena de responsabilidad de una universidad. Su incorporación como evidencia de contraste permite reconocer la utilidad de la asistencia y delimitar el examen adicional que exige un uso distinto (Noy & Zhang, 2023).

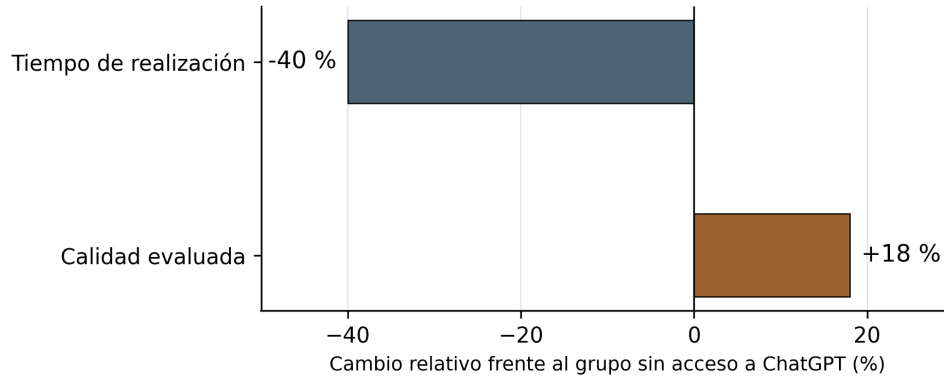
La Figura 2 muestra el contraste favorable que debe conservar una discusión de la soberbia algorítmica: la asistencia puede ahorrar tiempo y mejorar una producción evaluada. La reducción del tiempo no demuestra por sí misma que el resultado sea verdadero, y la calidad medida en esas tareas no establece una competencia universal. Para una institución educativa, el hallazgo justifica investigar usos delimitados y considerar el trabajo adicional de comprobación. La extrapolación hacia decisiones con efectos sobre derechos requiere evidencia pertinente al nuevo contexto (Noy & Zhang, 2023).

La solución propuesta consiste en clasificar afirmaciones por consecuencia y por dificultad de comprobación. Una fecha localizada en un documento público exige un control diferente de una interpretación que decide una oportunidad. Los responsables pueden definir cuáles fundamentos deben verificarse íntegramente y cuáles admiten una revisión acotada. La clasificación debe mantenerse abierta a errores de omisión: una frase aparentemente descriptiva podría sostener una recomendación de alto impacto. El procedimiento se justifica por las obligaciones del caso y por la posibilidad de rectificar. Su propósito es distribuir atención hacia las relaciones decisivas, con recursos suficientes para que la supervisión pueda cumplirse.

La investigación sobre funciones que exigen elaborar una respuesta propia antes de recibir ayuda ofrece otra vía de intervención. Buçinca y colaboradores examinan diseños destinados a reducir la dependencia excesiva de recomendaciones automatizadas. Ese antecedente permite

plantear ejercicios de verificación en los que el estudiante formule primero una interpretación, compare después la salida y explique cualquier cambio. La adaptación es una propuesta pedagógica de esta revisión. Su utilidad necesita comprobarse mediante calidad de las razones y transferencia a tareas nuevas, además de satisfacción. Introducir una pausa puede aumentar esfuerzo sin mejorar comprensión si el diseño solo añade un trámite (Buçinca et al., 2021).

Figura 2. Cambios relativos en tiempo y calidad de tareas de escritura profesional



Nota. Elaboración propia con datos de Noy y Zhang (2023), <https://doi.org/10.1126/science.adh2586>. Experimento con 453 profesionales. Los porcentajes indican cambios relativos frente al grupo sin acceso a ChatGPT. Tiempo y calidad son variables distintas; no se representan intervalos de confianza. El alcance corresponde a las tareas del experimento.

EL DESACUERDO DEBE PODER CAMBIAR LA ACTUACIÓN

Un control pierde valor cuando la persona reconoce una falla y carece de un recorrido para corregirla. El registro de errores debe conectarse con la versión del documento, el responsable de aprobación y la comunicación a quienes recibieron el resultado. En educación, la corrección puede requerir revisar una retroalimentación, una calificación o una orientación ya utilizada. El grado de respuesta depende de la consecuencia, del alcance de circulación y del tiempo transcurrido. La literatura sobre ética de diseño y privacidad ayuda a evitar que ese registro se convierta en acumulación innecesaria de conversaciones o datos personales. Se necesita documentar el fundamento decisivo conservando información proporcional a su finalidad (Lachheb et al., 2023).

La calidad de una verificación depende también del acceso efectivo a la evidencia. El resumen de un artículo permite identificar su objetivo y algunos resultados, mientras el texto completo suele ser necesario para examinar mediciones, selección de participantes y decisiones analíticas. Cuando una recomendación depende de esos detalles, la falta de acceso limita lo que puede afirmarse. El responsable debe reconocer ese límite y buscar una versión de autor, una fuente complementaria o una comprobación apropiada, en lugar de completar lo desconocido mediante una respuesta plausible. En educación, esta exigencia convierte la lectura de fuentes en una competencia sustantiva: permite detectar cuándo una conclusión tiene respaldo suficiente y cuándo necesita permanecer abierta (Miao & Holmes, 2023).

El resultado operativo es una nota de evidencia que acompaña afirmaciones determinantes. Su contenido mínimo identifica proposición, fuente recuperada, fragmento pertinente, inferencia y motivo de aprobación o corrección. La nota se propone como herramienta de trabajo y puede mantenerse fuera del texto publicado cuando contiene información sensible. Permite reconstruir cómo una bibliografía participa en una decisión y hace visible el trabajo que la fluidez había ocultado. Su efectividad requiere observar si detecta errores que cambian el resultado, qué costo introduce y quién puede utilizarla. Esas preguntas convierten una buena intención de verificar en un diseño susceptible de evaluación.

IA FRENTE AL CONOCIMIENTO EXPERTO

COMPETENCIA SITUADA Y RESPONSABILIDAD

El conocimiento experto permite reconocer condiciones que una pregunta breve deja

implícitas. Un profesional distingue qué medición falta, qué término resulta ambiguo o qué recomendación necesita una revisión presencial. Esa capacidad también puede fallar y debe someterse a crítica. La alternativa responsable consiste en comparar herramientas y profesionales respecto de tareas definidas, en lugar de atribuir infalibilidad a cualquiera de ellos. La guía de la OMS subraya la relación entre gobernanza, derechos y quienes dependen de aplicaciones de IA en salud (WHO, 2021).

La asistencia puede ampliar la capacidad del experto. Resumir documentos, explorar hipótesis o identificar diferencias facilita trabajo que antes requería más tiempo. El riesgo aparece cuando esas operaciones reemplazan la indagación necesaria o modifican una decisión sin que el responsable comprenda el cambio. La competencia profesional debe incluir saber utilizar el sistema y saber cuándo detener su uso. Una buena integración conserva el criterio para reconocer qué evidencia aún falta (Autio et al., 2024).

LA SUPERVISIÓN REQUIERE CONDICIONES MATERIALES

Asignar a una persona la etiqueta de supervisora resulta insuficiente si carece de acceso a información o recibe más decisiones de las que puede revisar. También puede resultar nominal cuando un contrato o una jerarquía le impiden contradecir la herramienta. El control efectivo exige tiempo, formación, autoridad y un mecanismo de escalamiento. La auditoría a lo largo del ciclo de desarrollo ayuda a observar estas condiciones desde el diseño hasta el despliegue (Raji et al., 2020).

La objeción habitual afirma que una revisión humana introduce demoras y errores. Esa posibilidad debe evaluarse, pero no resuelve automáticamente el problema de responsabilidad. Un procedimiento puede combinar verificaciones técnicas, revisiones por muestreo y control intensivo de casos críticos. Su calidad se examina por los daños que previene y las decisiones que permite corregir. El criterio de eficiencia debe incluir el costo de un error y de su reparación, además del tiempo inicial de respuesta.

LA COMPETENCIA SE RECONOCE EN LAS PREGUNTAS QUE FALTAN

El conocimiento experto incluye reconocer qué información necesita una tarea. Esa capacidad adquiere importancia cuando la respuesta automática parece completa antes de que el problema esté bien definido. Un asistente puede redactar una interpretación de un proyecto educativo sin advertir que los datos corresponden a cursos distintos o que la población cambió durante el periodo. El profesional necesita examinar esas condiciones y formular preguntas adicionales. La diferencia resulta observable en el procedimiento: quien comprende el problema puede justificar por qué requiere un dato, qué interpretación modifica y cuándo debe suspender una conclusión. La competencia se expresa mediante esa relación entre pregunta, evidencia y consecuencia, además de conocimientos declarativos.

Una comparación justa entre IA y especialistas necesita un referente de desempeño definido. Medir rapidez, legibilidad y corrección factual responde a preguntas diferentes de evaluar pertinencia de una decisión. También importa qué información reciben los participantes y qué herramientas pueden utilizar. Una persona que consulta un expediente completo y un modelo que recibe un resumen breve no resuelven la misma tarea informativa. La revisión propone documentar entradas, ayudas y criterio de evaluación antes de atribuir superioridad global. Este procedimiento permite identificar dónde el sistema aporta y dónde depende de conocimiento adicional, sin confundir una ventaja localizada con sustitución de una profesión completa (Cui, 2026; Shneiderman, 2020).

Collins y Evans (2002) sitúan la discusión sobre expertise en la relación entre conocimiento especializado y participación en decisiones. Su propuesta permite preguntar qué competencia requiere cada intervención, evitando que todas las opiniones reciban el mismo estatuto por compartir una forma verbal. Aplicada a la IA, esta distinción impide inferir dominio práctico a partir de una conversación fluida sobre una profesión. El conocimiento que permite discutir un procedimiento puede diferir del necesario para ejecutarlo y reconocer sus excepciones. La evaluación de un asistente debe identificar cuál de esas capacidades se examina. La ampliación de la participación sigue siendo valiosa, pero exige que las contribuciones puedan juzgarse por su fundamento y por su pertinencia para la tarea, sin reducir autoridad a credenciales o a elocuencia.

CONOCIMIENTO EXPLÍCITO Y TRABAJO QUE DEPENDE DEL CONTEXTO

La distinción entre conocimiento tácito y explícito permite comprender por qué el acceso a documentos resulta insuficiente para algunas decisiones. Polanyi (1966/2009) explica que la práctica del conocimiento incluye componentes que una descripción verbal no logra expresar por completo. Reconocer un caso excepcional puede depender de experiencia perceptiva, participación en una actividad y familiaridad con sus criterios. Este argumento antecede a la IA generativa y carece del propósito de evaluar sus capacidades actuales; ofrece una pregunta que sigue siendo pertinente: qué parte de una competencia queda representada en el texto que recibe el sistema. Un procedimiento escrito facilita localizar requisitos, mientras su aplicación a una situación ambigua puede exigir capacidades adicionales. La revisión de Litvinenko (2026) aborda esta relación entre conocimiento y colaboración en el aprendizaje organizacional, ampliando el análisis hacia la distribución de capacidades dentro de una institución.

La automatización también modifica el conocimiento que llega a explicitarse. Si una institución incorpora al sistema solamente reglas fáciles de codificar, puede darles más peso que a razones difíciles de representar. Un docente podría terminar evaluando las dimensiones que la herramienta produce con facilidad y dejar en segundo plano una interpretación original que exige lectura cuidadosa. Este efecto constituye una hipótesis organizacional derivada del análisis y debe contrastarse en cada implementación. Su interés radica en que la tecnología participa en la definición de lo que cuenta como buen trabajo. Evaluar únicamente si ejecuta una regla omite examinar por qué esa regla fue seleccionada.

La revisión de Cai y colaboradores sobre gestión del conocimiento y sostenibilidad empresarial ofrece una perspectiva complementaria: el valor de la IA depende de cómo se articula con conocimiento, estrategia y gobernanza. La diversidad de métodos del corpus limita inferencias causales comunes. Para este capítulo, su función consiste en ampliar la unidad de análisis desde una respuesta aislada hacia la organización que incorpora la asistencia. La pregunta cambia cuando se necesita compartir experiencia, corregir documentos y conservar aprendizajes. Un sistema puede producir una recomendación útil y dejar sin resolver cómo la institución aprende de sus errores o transmite los criterios que hicieron útil esa recomendación (Cai et al., 2026).

EL ESPECIALISTA NECESITA CONSERVAR PRÁCTICA

La asignación del trabajo puede debilitar las capacidades necesarias para supervisarlo. Bainbridge analizó las dificultades que aparecen cuando la automatización deja al operador responsabilidades excepcionales que requieren habilidades poco ejercitadas. Su argumento antecede a los servicios generativos y ofrece una relación conceptual pertinente: la supervisión puede exigir precisamente la experiencia que la rutina automatizada reduce. La adaptación educativa de este capítulo propone conservar momentos de elaboración, interpretación y corrección. Su necesidad debe decidirse por la competencia que la persona debe mantener y por el tipo de incidente al que tendrá que responder, en lugar de asumir que toda tarea delegada produce pérdida de capacidad (Bainbridge, 1983).

Un equipo académico puede diseñar esa conservación mediante revisión de casos discrepantes. Los integrantes examinan una recomendación, explican el criterio aplicable y comparan su resultado con la salida automatizada. El valor del ejercicio depende de que contenga problemas auténticos de interpretación y de que la discusión conduzca a una regla o a una excepción documentada. Repetir una lista de advertencias ofrece menor oportunidad de desarrollar juicio. La propuesta admite diferentes formatos según tiempo y recursos: una reunión breve sobre decisiones difíciles puede servir mejor que una formación general que nunca regresa a los casos. Su eficacia requiere observar desempeño posterior y calidad de las correcciones.

La resistencia a una herramienta tampoco demuestra experiencia suficiente. La revisión de Burton y colaboradores organiza factores relacionados con la aceptación o el rechazo de ayudas algorítmicas; los experimentos de Dietvorst y colaboradores muestran que observar errores puede favorecer el abandono de un algoritmo útil en las tareas estudiadas. Ambos antecedentes limitan una lectura que identifique prudencia con rechazo permanente. La confianza responsable exige comparar desempeño y consecuencias bajo condiciones semejantes. Un especialista puede sobrevalorar su propio acierto, del mismo modo que otro usuario sobrevalora la fluidez del sistema. El objetivo de la revisión es una dependencia justificada y corregible (Burton et al., 2020; Dietvorst et al., 2015).

LA SUPERVISIÓN ES UNA RELACIÓN DE TRABAJO

La implementación organizacional sitúa la confianza dentro de una relación de trabajo. El estudio de Fox y Solihin (2026) examina la incorporación de IA generativa junto con sus beneficios y riesgos de seguridad. La discusión sobre adopción adquiere sentido cuando se conecta con la distribución de responsabilidades: quién proporciona información, quién examina la salida y quién puede detener una actuación. Una supervisión declarada puede resultar nominal si la persona carece de tiempo o de autoridad para corregir. Por ello, el juicio profesional debe mantenerse unido a recursos y mecanismos de revisión. La evidencia de adopción aporta preguntas para estudiar esa organización, sin convertir la existencia de una estrategia en demostración de su eficacia.

El juicio profesional puede ampliarse mediante la IA cuando la asistencia hace recuperables documentos, comparaciones o alternativas que la persona puede examinar. La condición decisiva consiste en conservar el vínculo entre competencia y responsabilidad. El experto debe poder justificar una conclusión con razones distintas de la conformidad del sistema y modificarla cuando aparecen datos pertinentes. A su vez, la institución debe ofrecer condiciones para que ese examen forme parte del trabajo real. Esta relación permite valorar una integración por lo que conserva y desarrolla en el equipo, además de lo que produce durante una interacción. El conocimiento situado del siguiente acápite amplía la pregunta hacia quienes quedaron fuera de la definición inicial del problema.

SABER UNIVERSAL FRENTE A CONOCIMIENTO SITUADO

LO QUE APARECE EN LOS DATOS

La disponibilidad digital de un conocimiento influye en la posibilidad de que un sistema lo incorpore. Los textos abundantes y accesibles tienen condiciones diferentes de representación frente a saberes orales, prácticas locales o lenguas con escasos recursos digitales. Una respuesta amplia puede ocultar esas asimetrías. La advertencia sobre concentración de datos y representación obliga a preguntar qué voces se encuentran detrás de una explicación y cuáles quedan difíciles de recuperar (Bender et al., 2021; UNESCO, 2021).

Un proyecto educativo territorial puede utilizar IA para comparar información publicada y organizar preguntas. La descripción de un conflicto por agua, una práctica agrícola o una memoria comunitaria requiere además consultar a quienes participan en esa experiencia. La propuesta del CELAM vincula desarrollo con oportunidades reales y reconocimiento de contextos. Esta perspectiva ayuda a evitar que una forma dominante de información se convierta en medida suficiente de todas las experiencias (Grupo de Trabajo de Frontera Tecnológica del Celam [GFT-CELAM], 2026).

PARTICIPACIÓN EN LA DEFINICIÓN DEL PROBLEMA

Incluir conocimientos situados exige intervenir antes de la evaluación final. Las comunidades deben poder discutir el problema, las categorías, la información que se recopila y el uso de los resultados. Incorporar algunos testimonios después de diseñar una solución conserva intacta la relación de autoridad. La perspectiva sociotécnica muestra que los límites de abstracción determinan qué actores y relaciones se consideran relevantes para una evaluación de justicia (Selbst et al., 2019).

Los sistemas generativos pueden ampliar el acceso lingüístico y producir versiones comprensibles de materiales especializados. Este beneficio se sostiene cuando la adaptación se verifica con hablantes y expertos pertinentes. Una traducción fluida puede perder matices normativos o culturales. La solución propuesta consiste en combinar asistencia con validación situada y derecho a corregir representaciones. Reconocer la diversidad fortalece el conocimiento que puede circular y la legitimidad de quien lo comunica (Cortina, 2024; Miao & Holmes, 2023).

LO UNIVERSAL DEPENDE DE QUÉ PUDO REGISTRARSE

La presentación de un asistente como interlocutor universal oculta una selección previa de información, lenguas y formas de expresión. Incluso un sistema con capacidades multilingües

puede recibir consultas formuladas desde categorías institucionales que dejan fuera experiencias relevantes. La cuestión no se reduce a cuántos idiomas comprende. También exige examinar qué se considera evidencia, quién puede aportarla y qué ocurre cuando una persona expresa un problema de una manera distinta a la esperada. La perspectiva sociotécnica permite observar esas relaciones en la construcción de una tarea. Un desempeño técnico adecuado dentro de categorías estrechas puede coexistir con una representación insuficiente del problema humano (Bender et al., 2021; Selbst et al., 2019).

Haraway (1988) plantea el conocimiento situado como una exigencia de responsabilidad respecto de la posición desde la cual se conoce. Su crítica de una mirada que pretende abarcarlo todo permite examinar la apariencia de universalidad de la IA sin abandonar la aspiración de objetividad. Identificar límites, condiciones y perspectivas hace posible discutir una afirmación con mayor precisión. En una respuesta generativa, la ausencia de un punto de vista declarado puede ocultar elecciones sobre información, lenguas y categorías. La pregunta relevante consiste en localizar esas elecciones y contrastarlas con el problema que se intenta comprender. El argumento filosófico antecede a los modelos actuales; su aplicación propone un criterio de evaluación, cuya utilidad debe comprobarse mediante análisis de representaciones concretas.

El conocimiento territorial presenta dificultades específicas de recuperación. Una práctica que se transmite oralmente, una memoria local o una experiencia de acceso a servicios puede tener una presencia digital limitada. La ausencia de un documento fácilmente localizable no permite concluir que el conocimiento sea irrelevante o inexistente. Para una universidad que desarrolla proyectos comunitarios, esta diferencia modifica el método de trabajo. El asistente puede organizar materiales publicados y preguntas preliminares; el equipo conserva la necesidad de escuchar a los participantes y discutir con ellos el sentido de las categorías. Esta aplicación ilustrativa evita utilizar una respuesta generativa como descripción final de una comunidad (Cortina, 2024; GFT-CELAM, 2026).

UN INDICADOR PUEDE REPRESENTAR DE FORMA INSUFICIENTE UNA NECESIDAD

Los estudios de sesgos ayudan a reconstruir cómo una medida conveniente puede fallar como representación de aquello que importa. Obermeyer y colaboradores examinaron un algoritmo de salud cuyo uso de costos como representación de necesidad contribuyó a desigualdades raciales. El caso se incorpora como evidencia de un mecanismo de representación en un contexto específico. Su traslado a educación no supone que se haya observado la misma magnitud de desigualdad. Permite plantear una comparación de categorías: disponibilidad de registros, consumo de recursos y necesidad educativa describen fenómenos diferentes. Una institución debe justificar el indicador con el que reconoce a las personas y revisar a quién deja fuera (Obermeyer et al., 2019).

Una universidad regional podría utilizar actividad en la plataforma como señal de compromiso académico. El indicador facilita el registro, pero depende de conectividad, acceso a equipos, tareas asignadas y formas de estudiar. En el escenario propuesto, un estudiante que descarga materiales y trabaja fuera de línea podría aparecer como poco activo. Antes de delegar una recomendación de apoyo, el programa debe examinar qué explica esa actividad y qué fuentes adicionales necesita. La clasificación constituye una elaboración ilustrativa de este capítulo. Su utilidad consiste en mostrar que una medición puede ser exacta respecto del registro y débil respecto del concepto que pretende representar.

La investigación cualitativa de Kornhaber y colaboradores recoge experiencias de 149 participantes en una universidad regional australiana y muestra la coexistencia de apoyo flexible, barreras y preocupaciones éticas. La participación voluntaria y las respuestas breves delimitan sus conclusiones. La fuente no describe universidades colombianas ni estima una prevalencia general. Ofrece un contraste situado que impide considerar la disponibilidad del servicio como condición suficiente de aprovechamiento. El problema educativo requiere examinar recursos y trayectorias de uso. La literatura latinoamericana sobre socialización digital constituye una vía complementaria para investigar cómo acceso y habilidades se relacionan con beneficios efectivos (González-Chouciño & Ruiz-Callado, 2023; Kornhaber et al., 2026).

PARTICIPAR SIGNIFICA PODER MODIFICAR CATEGORÍAS

La participación relevante incluye discutir cómo se formula un problema antes de

convertirlo en una tarea computacional. Pedir opinión sobre una salida ya definida puede mejorar su aceptación y dejar intactos los criterios que excluyen experiencias. En una evaluación de aprendizaje, estudiantes y docentes necesitan comprender qué dimensión se mide y qué procedimiento admite explicar una discrepancia. En un proyecto territorial, las comunidades requieren conocer la finalidad del registro y decidir qué información consideran apropiado compartir. La investigación sobre fuentes de daño a lo largo del aprendizaje automático permite examinar esta temporalidad: las decisiones de representación preceden a la evaluación final del modelo (Suresh & Guttag, 2021).

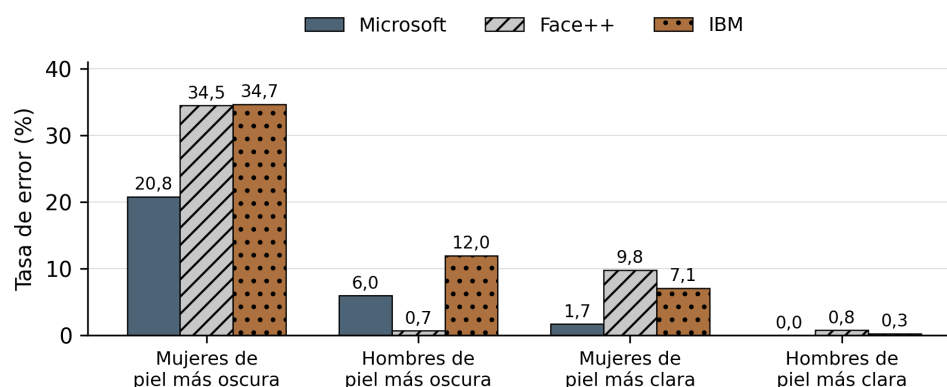
La injusticia epistémica introduce otra dimensión de la soberbia. Fricker (2007) estudia daños que afectan a una persona en su capacidad de aportar conocimiento, incluidos los déficits injustos de credibilidad y las dificultades para interpretar experiencias cuando faltan recursos compartidos. El traslado de esta perspectiva a la IA constituye una elaboración conceptual: una institución puede otorgar mayor credibilidad a una salida automática que al relato de quien conoce una excepción relevante. El problema merece atención incluso si la recomendación fue generada correctamente a partir de los registros disponibles. La ausencia de una experiencia en esos registros puede convertirse en motivo para desestimarla. Un procedimiento responsable conserva vías para aportar información, discutir categorías y revisar la decisión, sin exigir que toda experiencia adopte previamente el lenguaje reconocido por el sistema.

La privacidad añade una condición al reconocimiento de voces. Incorporar experiencias personales sin discutir finalidad, acceso y conservación puede ampliar la representación a costa de exponer a quienes participan. La revisión propone recuperar información necesaria y separar registros identificables de las síntesis que alimentan una herramienta. El diseño debe examinar si una finalidad puede cumplirse mediante documentos públicos, materiales agregados o procesamiento local antes de solicitar nuevos datos. La ética del diseño educativo permite relacionar estas decisiones con alternativas reales, evitando reducir la protección a una declaración que acompaña la misma práctica de extracción (Lachheb et al., 2023).

Las evaluaciones desagregadas pueden ayudar a localizar límites de desempeño, siempre que las categorías y los grupos comparados tengan una justificación. Gender Shades documentó diferencias en sistemas de clasificación facial evaluados bajo condiciones definidas. El antecedente ilustra por qué un promedio puede ocultar distribuciones desiguales; no permite asignar sus resultados a todo asistente de lenguaje ni deducir automáticamente qué categorías convienen en una institución educativa. La pregunta transferible consiste en identificar a quién afecta cada error y con qué evidencia puede observarse esa distribución. La comparación debe conservar privacidad y evitar que categorías de auditoría se transformen en etiquetas injustificadas sobre personas (Buolamwini & Gebru, 2018).

La Figura 3 permite observar cómo cambia el error al desagregar el desempeño. En los tres clasificadores, las mujeres de piel más oscura presentan tasas mayores que los demás grupos comparados. El promedio de un sistema no basta para reconocer quién soporta sus fallas. Este antecedente de clasificación facial fundamenta la exigencia de revisar distribuciones y categorías, manteniendo separados los problemas técnicos y sociales de cada aplicación. Su traslado a educación aporta una pregunta sobre representación; no permite asignar estas tasas a estudiantes, asistentes de lenguaje o sistemas actuales (Buolamwini & Gebru, 2018).

Figura 3. Tasas de error desagregadas en los clasificadores de Gender Shades



Nota. Elaboración propia con datos de Buolamwini y Gebru (2018, Tabla 4), <https://proceedings.mlr.press/v81/buolamwini18a.html>. Error = 1 - tasa de verdaderos positivos. Evaluación en Pilot Parliaments Benchmark de las versiones de abril y mayo de 2017. Se conservan las categorías del estudio; las cifras no describen sistemas actuales ni IA generativa.

LA ADAPTACIÓN REQUIERE INVESTIGACIÓN EDUCATIVA

La revisión de Zawacki-Richter y colaboradores llamó la atención sobre la presencia limitada de perspectivas pedagógicas y de reflexión crítica en parte de la investigación de IA en educación superior que examinó. El periodo y el corpus de ese estudio son históricos. Su incorporación al capítulo sustenta una exigencia de diseño: explicar qué finalidad educativa se persigue y qué teoría o criterio permite valorar el aprendizaje. Una aplicación generativa actual requiere además comprobar sus condiciones técnicas y sociales. El antecedente ayuda a evitar que la evaluación se limite a rendimiento computacional cuando la institución busca comprensión, participación o autonomía (Zawacki-Richter et al., 2019).

Reconocer conocimiento situado modifica finalmente la idea de autoridad. Una respuesta puede ofrecer información general y necesitar corrección por quienes conocen condiciones que el sistema no recibió. La corrección debe incorporarse a la decisión, con un registro proporcional y con reconocimiento de la contribución. Esta relación abre la posibilidad de usar IA como apoyo de intercambio y traducción sin convertirla en medida final de todas las experiencias. La autoridad responsable emerge de una deliberación que hace visibles sus fuentes y permite discutir sus categorías. La humildad epistemológica adquiere, por ello, una dimensión institucional: aceptar límites implica distribuir capacidad real para señalarlos.

LA HUMILDAD EPISTEMOLÓGICA COMO RESPUESTA

UN PROTOCOLO PARA APRENDER DE LOS LÍMITES

La humildad epistemológica se expresa en procedimientos que permiten reconocer límites sin paralizar la acción. Antes de utilizar una respuesta, conviene precisar qué se necesita saber y qué consecuencias tendrá un error. Después debe identificarse la evidencia pertinente y examinar si el sistema tiene acceso a ella. La revisión final compara la afirmación con las fuentes y registra incertidumbres relevantes. El resultado es una decisión sustentada, susceptible de corrección y proporcional a su contexto (Farquhar et al., 2024; Tabassi, 2023).

La formación puede organizarse mediante ejercicios de discrepancia. Dos respuestas distintas sobre un mismo problema obligan a localizar criterios para compararlas. Una bibliografía con referencias reales pero mal interpretadas permite estudiar un error diferente de la fabricación de fuentes. Una salida correcta obtenida con razonamiento deficiente muestra que el resultado final tampoco agota la evaluación. Estas actividades son propuestas pedagógicas; su eficacia debe examinarse mediante desempeño de los estudiantes y transferencia a situaciones nuevas (Alier-Forment et al., 2026; Mustafa et al., 2024).

LA HUMILDAD SE EXPRESA EN DECISIONES QUE PUEDEN CAMBIAR

Una práctica de humildad epistemológica necesita definir qué información haría revisar una actuación. Declarar de manera general que el sistema puede equivocarse deja indeterminado el siguiente paso. En una universidad, el procedimiento puede señalar que una fuente no recuperable impide utilizar determinada afirmación o que una discrepancia sobre población exige modificar una conclusión. Las condiciones deben responder a la tarea y a la consecuencia del error. Esta propuesta convierte el reconocimiento del límite en una relación operativa: una duda identificada produce una comprobación, y el resultado de esa comprobación puede alterar lo que se comunica o decide. Su efectividad debe examinarse en el trabajo institucional.

Los marcos de principios ofrecen un punto de partida, pero necesitan procedimientos que hagan observable su aplicación. Floridi y Cowls sintetizan principios para una IA orientada a la sociedad; Jobin y colaboradores comparan orientaciones éticas y Mittelstadt discute los límites de confiar exclusivamente en principios. Estas contribuciones cumplen funciones conceptuales distintas. La síntesis de este capítulo las utiliza para preguntar qué persona puede ejercer una facultad concreta y con qué recursos. En educación, autonomía puede traducirse en comprender una evaluación y discutirla. La traducción propuesta conserva la discusión normativa y añade una pregunta susceptible de comprobación institucional (Floridi & Cowls, 2019; Jobin et al., 2019; Mittelstadt, 2019).

LA FORMACIÓN DEBE ENFRENTAR RESULTADOS QUE RESTRINGEN SU PROMESA

La alfabetización en IA ofrece categorías para comprender capacidades, evaluar usos y participar en decisiones. Su valor necesita distinguirse de la expectativa de que cualquier formación elimine la dependencia o los errores. Tanveer (2026) examinó mediante encuesta relaciones entre uso de IA, alfabetización y resultados cognitivos percibidos en 518 estudiantes de universidades públicas de Riad. El efecto moderador esperado de la alfabetización sobre la relación entre uso y pereza cognitiva no se confirmó. El resultado restringe una promesa amplia de capacitación, pero su diseño transversal y sus autoinformes impiden tratarlo como prueba objetiva de deterioro de la memoria. La decisión formativa necesita distinguir lo que las personas declaran de aquello que pueden demostrar al resolver una tarea.

La consecuencia para el diseño consiste en especificar qué debe aprender una persona y cómo se observará. Una actividad puede evaluar si distingue una asociación de una relación causal, recupera una fuente o reconoce un supuesto ausente. El desempeño debe examinarse en tareas con diferencias suficientes para comprobar transferencia. Una alta valoración del curso puede expresar satisfacción y dejar abierta la competencia. Esta revisión propone relacionar cada objetivo formativo con una operación de comprobación y con un caso donde el criterio habitual resulte insuficiente. La propuesta requiere evaluación propia; la fuente anterior no demuestra que este diseño vaya a resolver el problema.

La confianza cognitiva y la confianza emocional también deben reconocerse como dimensiones diferentes. La encuesta de Q. Cheng y colaboradores estudia usuarios, alucinaciones y colaboración mediante relaciones estadísticas, considerando el ajuste entre tarea y tecnología. Su función aquí es delimitar la confianza como fenómeno situado. La relación entre variables de encuesta no establece por sí misma causalidad experimental ni autoriza atribuir a todos los usuarios una misma conducta. El capítulo propone valorar si la persona conoce las condiciones del apoyo y si puede reconocer una salida que debe revisar. Una experiencia agradable puede contribuir a continuar la interacción y requerir un examen separado de su fundamento (Q. Cheng et al., 2026).

DE LAS ORIENTACIONES GENERALES A LAS CONDICIONES DE USO

El análisis documental de Ares-Ferreirós y colaboradores sobre buenas prácticas universitarias reúne directrices, leyes e informes de un periodo definido. Se utiliza como evidencia de orientaciones de gobernanza y transparencia, con alcance documental. La existencia de una recomendación no prueba su cumplimiento ni sus efectos sobre aprendizaje. La institución debe explicar qué orientación adopta, cómo la adapta y qué registro permite valorar su aplicación. Un procedimiento de uso puede distinguir asistencia permitida, información que debe protegerse y decisiones que necesitan aprobación. La redacción adquiere utilidad cuando una persona puede reconocer qué conducta corresponde a una tarea real (Ares-Ferreirós et al., 2026).

Las clasificaciones de riesgos organizacionales facilitan reconocer problemas que la evaluación de una respuesta aislada puede omitir. Crocco et al. (2026) examinan riesgos de implementación de la IA; su utilidad para esta discusión reside en ampliar la atención hacia condiciones institucionales, además de capacidades técnicas. Una oficina que trabaja con documentos públicos enfrenta necesidades distintas de otra que procesa expedientes sensibles. La pregunta sobre autoridad exige estudiar qué decisiones se delegan, qué información se utiliza y qué posibilidad conserva una persona de cuestionar el resultado. Las tipologías orientan ese examen, pero la frecuencia y gravedad de los daños necesitan evidencia del contexto donde se aplica el sistema.

EVALUAR TANTO EL ACIERTO COMO LA POSIBILIDAD DE CORREGIR

Un procedimiento puede producir buenos resultados durante una prueba y fallar cuando cambia la fuente, el curso o la versión del servicio. Por ello, la revisión de adopción necesita registrar condiciones de evaluación y señales de cambio. La organización puede definir qué modificaciones exigen repetir una comparación y quién conserva la información necesaria. Esta propuesta evita convertir una aprobación inicial en una autorización indefinida. El seguimiento debe examinar errores relevantes y resultados favorables, con atención a la carga que introduce. Las auditorías a lo largo del ciclo de vida y la gestión de riesgos permiten situar esa obligación dentro de una gobernanza más amplia (Raji et al., 2020; Tabassi, 2023).

La posibilidad de corregir incluye la comunicación a quienes utilizan el documento. Si una orientación se modifica después de circular, debe identificarse la versión vigente y la relación con la anterior. El registro necesita permitir reconstruir la razón del cambio sin conservar datos personales innecesarios. En educación, esta práctica también puede mostrar cómo se produce conocimiento revisable: la corrección responde a una fuente o a un argumento y no equivale automáticamente a fracaso del trabajo. El diseño propuesto distribuye responsabilidad entre quien genera, quien comprueba y quien aprueba. Esas funciones pueden coincidir en una persona para tareas simples, pero deben permanecer reconocibles cuando las consecuencias aumentan.

La humildad epistemológica permite utilizar un sistema sin concederle autoridad universal y corregir una decisión sin renunciar a toda asistencia. Su exigencia principal consiste en conservar razones recuperables y condiciones materiales para examinarlas. Este resultado interpretativo une los seis acápites: el grado de dependencia debe corresponder a la función, a la evidencia y a la posibilidad de revisar. La aplicación que sigue desarrolla esa relación en una decisión educativa ilustrativa. El caso permitirá observar dónde el procedimiento facilita actuar y qué problemas permanecen abiertos, para que el cierre del capítulo pueda valorar un diseño y no limitarse a reiterar una precaución.

Tabla 3. Condiciones para revisar una decisión apoyada en inteligencia artificial

Condición	Señal de insuficiencia	Respuesta propuesta
Función delimitada	Una descripción se convierte en decisión.	Separar información, interpretación y autorización.
Fundamento accesible	Una fuente falta o contradice la salida.	Suspender la afirmación y contrastar evidencia.
Representación adecuada	Una categoría omite circunstancias pertinentes.	Incorporar contexto y discutir el indicador.
Supervisión efectiva	La carga o la jerarquía impiden revisar.	Reducir alcance y asignar tiempo y autoridad.
Corrección posible	El error se registra, pero sigue produciendo efectos.	Modificar la decisión y comunicar la rectificación.

Nota. Distinciones analíticas desarrolladas en el acápite. La documentación permite examinar usos y límites, pero requiere contraste con la afirmación concreta (Gebu et al., 2021; Mitchell et al., 2019).

La humildad epistemológica también obliga a revisar la confianza en el propio juicio humano. La evidencia sobre aversión algorítmica muestra que una falla visible puede llevar a abandonar una herramienta que sigue siendo útil en la tarea evaluada. La cautela razonable exige comparar resultados bajo condiciones semejantes y reconocer cuándo el sistema aporta información que el profesional había omitido. Aceptar esa contribución mantiene la obligación de comprenderla y examinar sus consecuencias. Así, la revisión puede corregir tanto una recomendación automática como un prejuicio del evaluador. El criterio de cierre de una decisión debe depender de la suficiencia de sus razones, de la posibilidad de reparación y de la evidencia disponible, con un margen explícito para reconsiderarla (Burton et al., 2020; Dietvorst et al., 2015; Lee & See, 2004).

APLICACIÓN ILUSTRATIVA PARA REVISAR RETROALIMENTACIÓN ACADÉMICA

Un programa universitario de formación docente considera utilizar un asistente para producir comentarios iniciales sobre informes de investigación. El escenario es propuesto y desarrolla la aplicación del manuscrito anterior. El problema institucional combina oportunidad de apoyo y responsabilidad de evaluación: los estudiantes necesitan retroalimentación a tiempo, mientras los docentes deben examinar pregunta, evidencia y alcance de las conclusiones. La decisión planteada consiste en autorizar o limitar el sistema para preparar borradores de comentarios. La calificación y la aprobación del informe permanecen bajo un procedimiento docente cuya evidencia debe ser recuperable.

El programa identifica dos tipos de tareas. La primera localiza elementos presentes, como objetivo, población descrita y referencias utilizadas. La segunda interpreta si esos elementos se corresponden y si la conclusión excede el diseño. Una salida que cumple la primera tarea puede fallar en la segunda. La comparación propuesta separa ambas operaciones y registra comentarios correctos, observaciones infundadas y omisiones determinantes. El ejercicio necesita una rúbrica discutida previamente por docentes, porque el desacuerdo humano también requiere explicación. El estudio de Cui ofrece un antecedente para evaluar por dimensiones, pero no garantiza el desempeño del sistema que este programa seleccione (Cui, 2026).

La institución selecciona inicialmente materiales cuya utilización esté autorizada y establece qué información se excluye. Las referencias decisivas se recuperan de su publicación original o de

una versión identificada. El asistente recibe instrucciones para vincular cada observación con un fragmento y para declarar cuando una inferencia exige información adicional. Estas instrucciones organizan el flujo; su presencia no demuestra cumplimiento. El docente debe revisar si los comentarios corresponden al documento, si citan la edición pertinente y si mantienen el grado de certeza del estudio. La privacidad y la documentación del proceso se valoran conjuntamente, evitando incorporar conversaciones personales que no sean necesarias para la revisión (Ares-Ferreirós et al., 2026; Lachheb et al., 2023).

La comparación introduce una discrepancia deliberada de interpretación. Un informe presenta una asociación entre uso de una plataforma y satisfacción; el comentario automatizado recomienda afirmar que la plataforma mejoró aprendizaje. El docente identifica que la conclusión combina una inferencia causal y un resultado que el estudio no midió. El registro propuesto conserva el fragmento original, la observación generada y la razón de la corrección. Este episodio es una ilustración diseñada para el capítulo. Su función es mostrar qué cambia al comprobar la relación entre afirmación y evidencia: el comentario se reescribe y el estudiante recibe una explicación del límite, en lugar de una aprobación aparentemente fundamentada.

Otra discrepancia revela el límite de ese control. Un informe describe correctamente los datos disponibles, pero omite que parte de los estudiantes trabajó fuera de línea. El asistente produce comentarios fieles al texto y el docente, si revisa únicamente correspondencia documental, puede aprobar una interpretación insuficiente. El programa necesita decidir cuándo la tarea exige información sobre el contexto y cómo recuperarla. La regla de verificar citas y fragmentos resulta necesaria para ciertos errores, pero deja abiertas fallas de representación. El caso muestra por qué la revisión incorpora el conocimiento situado y la posibilidad de que los participantes discutan una categoría (Selbst et al., 2019; Suresh & Gutttag, 2021).

La tensión práctica aparece en la carga de trabajo. Un asistente puede ampliar el número y la extensión de comentarios, mientras el programa dispone de una capacidad limitada para comprobarlos. La propuesta debe evaluar tiempo completo de revisión y utilidad de las observaciones, además de velocidad de generación. Si el volumen impide examinar cuestiones determinantes, la institución puede acotar la herramienta a localización de elementos y reservar la interpretación para el procedimiento docente. Una adopción limitada puede sostener beneficios de apoyo y reducir una parte de la carga. Esta alternativa necesita comparación real antes de atribuirle ventajas; aquí se formula como una decisión posible.

El seguimiento propuesto observa si los estudiantes comprenden las correcciones y pueden aplicar el criterio a un informe diferente. Entregar una retroalimentación bien redactada no basta para establecer que el aprendizaje ocurrió. También se registra si el estudiante puede cuestionar una observación, qué información recibe y quién responde. El canal de revisión debe admitir que una crítica al comentario revele un problema de la rúbrica o de la interpretación humana. Así se conserva una distribución de responsabilidad capaz de aprender del desacuerdo. La herramienta forma parte de esa distribución y sus resultados quedan sujetos a la misma obligación de explicación (Holmes et al., 2022; Raji et al., 2020).

La autorización cambia si aparecen referencias no recuperables, observaciones repetidamente injustificadas, exposición indebida de información o una carga que vuelve nominal la supervisión. El programa debe concretar esas condiciones con sus obligaciones y recursos. Este capítulo propone que la decisión deje registrada función autorizada, evidencia de evaluación, límites y responsable de revisión. La ficha hace examinable el permiso y permite modificarlo ante cambios del servicio o de la tarea. El caso no concluye con una adopción exitosa: concluye con una decisión acotada cuya validez depende de pruebas y de condiciones que una institución real deberá comprobar (Fox & Solihin, 2026; Tabassi, 2023).

CONCLUSIONES

La soberbia algorítmica se reduce cuando una institución conserva la diferencia entre producir respuestas y justificar decisiones. Esa distinción permite aprovechar la amplitud de la asistencia sin atribuirle un conocimiento universal. El aprendizaje central consiste en reconocer qué sabe el sistema sobre una tarea, cómo se comprueba el resultado y dónde interviene una competencia adicional. La confianza se vuelve más sólida cuando admite dudas, contrastes y correcciones.

La transición hacia el problema de la avaricia se comprende al observar que una herramienta puede utilizarse con cautela y depender de infraestructuras, datos o contratos controlados por pocos actores. Reconocer el límite de una respuesta constituye una condición del juicio; conocer

quién organiza su producción permite examinar las posibilidades reales de elegir y corregir. La soberbia se refiere aquí a la autoridad injustificada del conocimiento aparente, mientras la concentración del poder tecnológico requiere una discusión propia sobre distribución y dependencia. Esta relación explica por qué los siete excesos deben estudiarse en su contexto, con argumentos y fuentes pertinentes a cada problema (Selbst et al., 2019; UNESCO, 2021).

La revisión responde la pregunta mediante una relación entre función, fundamento y posibilidad de corregir. Conceder autoridad a una salida requiere identificar qué tarea ejecuta, qué evidencia respalda su uso y quién puede revisar sus consecuencias. Estas condiciones explican por qué una respuesta correcta en una operación de localización no justifica automáticamente una interpretación o una decisión. La autoridad institucional depende del recorrido que transforma información en actuación. Su reconstrucción permite utilizar capacidades generativas y reconocer qué trabajo sigue siendo necesario para sostener el resultado.

Los hallazgos distinguen fluidez, estabilidad y satisfacción de una comprobación de verdad. También muestran que la correspondencia con un documento puede coexistir con límites de vigencia, representación o inferencia. Las señales de incertidumbre orientan revisiones y necesitan contraste externo. Las percepciones de confianza aportan información sobre la interacción, mientras las evaluaciones por dimensiones permiten observar desacuerdos que un promedio oculta. La diversidad de diseños del corpus obliga a conservar estas diferencias. El capítulo ofrece una síntesis conceptual y no una tasa general de confiabilidad de la IA.

El conocimiento experto y el situado amplían el examen hacia preguntas que una entrada breve deja ausentes. La supervisión requiere práctica, tiempo, información y autoridad para modificar decisiones. La aversión a herramientas útiles introduce un contraste necesario: prudencia no equivale a rechazo universal. Una integración responsable compara resultados y obligaciones, permitiendo reconocer beneficios localizados y fallas que exigen limitar el uso. La facultad de comprobar consiste en una capacidad material y argumentativa, además de una actitud favorable a la crítica.

La aplicación ilustrativa permite seguir una decisión de retroalimentación con criterios de revisión, información necesaria y condiciones para cambiar el permiso. También expone un límite del control documental: una salida fiel puede dejar fuera un problema de representación. La respuesta propuesta exige conectar verificación y participación en la definición de la tarea. Su eficacia debe comprobarse en una implementación concreta. Los procedimientos formulados sirven como diseños de investigación y de adopción; este capítulo no midió sus resultados.

Las implicaciones educativas se concentran en formar criterios que puedan transferirse entre tareas. Las instituciones necesitan evaluar cómo se reconoce una inferencia indebida, cómo se recupera una fuente y qué ocurre después de una corrección. La investigación futura puede comparar configuraciones de ayuda y estudiar cambios en competencia, carga de revisión y capacidad de impugnar. La conclusión conserva el alcance de la selección crítica: la soberbia algorítmica se limita cuando la organización hace recuperables las razones de su dependencia y mantiene condiciones para revisar esa dependencia.

DECLARACIÓN DE USO DE INTELIGENCIA ARTIFICIAL

Se utilizaron herramientas de inteligencia artificial generativa como apoyo en los procesos de organización, redacción y revisión del manuscrito. Su utilización se limitó a funciones de asistencia editorial y analítica, bajo supervisión y validación permanente de los autores. Las propuestas, interpretaciones y escenarios desarrollados en el texto corresponden a elaboraciones de carácter analítico y no constituyen resultados derivados de procesos de implementación empírica. Las herramientas de inteligencia artificial empleadas no reciben atribución de autoría ni sustituyen las fuentes científicas, documentales y académicas citadas. La selección de las fuentes, la interpretación de la evidencia, la construcción argumentativa, la revisión final del contenido y la decisión de someter el manuscrito a publicación son responsabilidad exclusiva de los autores.

REFERENCIAS

- Alíer-Forment, M., Casañ-Guerrero, M. J., Pereira-Varela, J. A., García-Peñalvo, F. J., & Llorens-Largo, F. (2026). Inteligencia artificial generativa y autonomía educativa: Metáforas históricas y principios éticos para la transformación pedagógica. *RIED Revista Iberoamericana de Educación a Distancia*, 29(1), 9–28. <https://doi.org/10.5944/ried.45536>
- Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. En *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). Association for Computing Machinery. <https://doi.org/10.1145/3290605.3300233>
- Ares-Ferreirós, M., de la Cruz-Campos, J. C., Morodo Horrillo, A., & Robert Guillén, S. (2026). Buenas prácticas en el uso de la IA generativa en las universidades. Un análisis documental. *Innoeduca International Journal of Technology and Educational Innovation*, 12(1), 5–20. <https://doi.org/10.24310/ijtei.121.2026.22152>
- Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.600-1>
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. En *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5185–5198). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.463>
- Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), Article 188. <https://doi.org/10.1145/3449287>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239. <https://doi.org/10.1002/bdm.2155>
- Cai, F., Bolisani, E., Nakash, M., & Kassaneh, T. C. (2026). AI-driven knowledge management for sustainable businesses: A comprehensive analysis. *VINE Journal of Information and Knowledge Management Systems*. Publicación anticipada en línea. <https://doi.org/10.1108/VJIKMS-09-2025-0444>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2025). *Sycophantic AI decreases prosocial intentions and promotes dependence* [Prepublicación]. arXiv. <https://doi.org/10.48550/arXiv.2510.01395>
- Cheng, Q., Dai, Y., Liu, X., & Peng, S. (2026). The trust crisis in artificial intelligence: AI hallucinations and human-AI collaboration. *Technology in Society*, 86, Article 103286. <https://doi.org/10.1016/j.techsoc.2026.103286>
- Collins, H. M., & Evans, R. (2002). The third wave of science studies: Studies of expertise and experience. *Social Studies of Science*, 32(2), 235–296. <https://doi.org/10.1177/0306312702032002003>
- Cortina, A. (2024). *¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnologizada*. Paidós.
- Crocco, E., Balleisio, E., Broccardo, L., & Alghafes, R. (2026). Welcome to the dark side: Unveiling the potential risks related to the implementation of Artificial Intelligence. *Journal of Risk Research*. Publicación anticipada en línea. <https://doi.org/10.1080/13669877.2025.2611954>
- Cui, Y. (2026). Artificial intelligence and human assessment consistency and bias in higher education academic writing evaluation. *Discover Artificial Intelligence*, 6, 559. <https://doi.org/10.1007/s44163-026-01286-0>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697–718. [https://doi.org/10.1016/S1071-5819\(03\)00038-7](https://doi.org/10.1016/S1071-5819(03)00038-7)
- Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature*, 630, 625–630. <https://doi.org/10.1038/s41586-024-07421-0>
- Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Fox, G., & Solihin, Y. (2026). Exploring organizational implementation of generative AI: Navigating benefits and safety risks. *Journal of Decision Systems*, 35(1), 2649940. <https://doi.org/10.1080/12460125.2026.2649940>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021).

- Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. <https://doi.org/10.1145/3458723>
- González-Chouciño, M. A., & Ruiz-Callado, R. (2023). Analysis of the third digital divide in relation to digital socialization itineraries among university students in Uruguay. *Societies*, 13(12), Article 252. <https://doi.org/10.3390/soc13120252>
- Grupo de Trabajo de Frontera Tecnológica del Celam. (2026). *Inteligencia artificial y desarrollo humano integral*. Consejo Episcopal Latinoamericano y Caribeño. <https://repositorio.celam.org/s/repositorio/item/3387>
- Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, 14(3), 575–599. <https://doi.org/10.2307/3178066>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Buckingham Shum, S., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32, 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), Article 248. <https://doi.org/10.1145/3571730>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Kornhaber, R., Bragg, S., Matheson, A., & Irwin, P. (2026). Qualitative inquiry into generative artificial intelligence impact on rural and regional higher education. *Australian and International Journal of Rural Education*, 36(2), 1–23. <https://doi.org/10.47381/aijre.v36i2.812>
- Lachheb, A., Abramenska-Lachheb, V., Moore, S., & Gray, C. (2023). The role of design ethics in maintaining students' privacy: A call to action to learning designers in higher education. *British Journal of Educational Technology*, 54, 1653–1670. <https://doi.org/10.1111/bjet.13382>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, 33, 9459–9474. <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>
- Lin, S., Hilton, J., & Evans, O. (2022). TruthfulQA: Measuring how models mimic human falsehoods. En *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 3214–3252). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.229>
- Litvinenko, A. (2026). The knowledge collaboration framework: A review of artificial intelligence in organisational learning and knowledge management. *Journal of Decision Systems*, 35(1), 2616703. <https://doi.org/10.1080/12460125.2026.2616703>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. En *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- Mata v. Avianca, Inc., No. 22-cv-1461 (PKC), Documento 54 (S.D.N.Y., 22 de junio de 2023). <https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2022cv01461/575368/54/>
- Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. En *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1906–1919). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.173>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. En *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287596>
- Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, 1, 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Mustafa, M. Y., Tlili, A., Lampropoulos, G., Huang, R., Jandrić, P., Zhao, J., Salha, S., Xu, L., Panda, S., Kinshuk, López-Pernas, S., & Saqr, M. (2024). A systematic review of literature reviews on artificial intelligence in education (AIED): A roadmap to a future research agenda. *Smart Learning Environments*, 11, Article 59. <https://doi.org/10.1186/s40561-024-00350-5>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, Article 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253. <https://doi.org/10.1518/001872097778543886>
- Polanyi, M. (2009). *The tacit dimension*. University of Chicago Press. (Trabajo original publicado en 1966). <https://press.uchicago.edu/ucp/books/book/chicago/T/bo6035368.html>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. En *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44).

- Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372873>
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
- Russell, S. J., & Norvig, P. (2004). *Inteligencia artificial. Un enfoque moderno* (J. M. Corchado Rodríguez, F. Martín Rubio, J. M. Cadenas Figueredo, L. D. Hernández Molinero, E. Paniagua Arís, R. FuentetajaPinzán, M. Robledo de los Santos, & R. Rizo Aldegue, trans.; 2.^a ed.). Pearson Educación.
- Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. En *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287598>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Suresh, H., & Guttag, J. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. En *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, Article 17. Association for Computing Machinery. <https://doi.org/10.1145/3465416.3483305>
- Tabassi, E. (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
- Tanveer, M. (2026). Artificial intelligence literacy moderating the effects of artificial intelligence use on cognitive outcomes including memory retention decision making and cognitive laziness. *Discover Psychology*, 6, 182. <https://doi.org/10.1007/s44202-026-00697-5>
- UNESCO. (2021). *Recomendación sobre la ética de la inteligencia artificial*. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
- World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. <https://www.who.int/publications/i/item/9789240029200>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>