White Paper #1

Efficient Advice

Let's Make it Better!

LAS SEIS FORMAS EN QUE LOS RESULTADOS DE LA IA FALLAN EN AUDITORÍA

Taxonomía de fallos de IA en la auditoría



PRESENTADO POR

EFFICIENT ADVICE, LLC

AUTOR

ROBINSON DE JESÚS

EFFICIENT ADVICE, LLC | WHITE PAPER | NÚM. 1

Las Seis Formas en que los Resultados de la IA Fallan en un Papel de Trabajo de Auditoría

Y Cómo Detectar Cada Una Antes de que Llegue a un Hallazgo

Robinson De Jesús | Director Ejecutivo, Efficient Advice, LLC

Info@efficientadvice.com | www.efficientadvice.com

Taxonomía de Fallas de IA en Auditoría | Efficient Advice

Versión 2.0 | Octubre del 2026

Colaboración de: Claude, Gemini, Canva y Grammarly.

A QUIÉN ESTÁ DIRIGIDO ESTE DOCUMENTO

Directores de Auditoría Interna (DAI), gerentes y auditores internos en todos los niveles que utilizan herramientas de IA para redactar hallazgos, elaborar informes o realizar cualquier parte de su trabajo de auditoría, y que aún no han implementado un protocolo formal de validación.

Resumen Ejecutivo

Las herramientas de IA no fallan al azar. Fallan de seis maneras específicas, nombradas y predecibles. Toda herramienta construida sobre un modelo de lenguaje de gran escala — incluyendo ChatGPT, Microsoft Copilot, Claude y Gemini — está sujeta a los seis modos de falla, porque los modos son propiedades de la arquitectura subyacente compartida, no de ningún producto específico. Los auditores internos que no pueden identificar estos patrones por su nombre no pueden detectarlos antes de que lleguen a un papel de trabajo, a un hallazgo o al comité de auditoría.

Este documento nombra y define la Taxonomía de Fallas de IA en Auditoría — seis modos de falla que representan el conjunto completo de formas en que los resultados de IA degradan la calidad, exactitud y defendibilidad profesional del trabajo de auditoría interna. Para cada modo, se explica cómo se manifiesta la falla en un contexto real de auditoría, cómo detectarla y cuál de los cinco pasos de validación la previene.

La Taxonomía no es teórica. Se fundamenta en la manera en que los modelos de lenguaje de gran escala (LLM) generan sus resultados — mediante predicción estadística en lugar de recuperación de hechos — y en las formas específicas en que ese mecanismo produce riesgo profesional cuando se aplica a hallazgos de auditoría, papeles de trabajo, evaluaciones de riesgo e informes.

> **HALLAZGO CLAVE**
>
> Las funciones de auditoría interna que utilizan IA sin una taxonomía de fallas nombrada no pueden validar sistemáticamente los resultados de IA, porque carecen del vocabulario para identificar lo que buscan. La Taxonomía de Fallas de IA en Auditoría proporciona a los auditores ese vocabulario. Nombrar la falla es el primer paso para detectarla.

1. Por qué Fallan los Resultados de IA: El Mecanismo Detrás de los Seis Modos

Antes de nombrar los seis modos de falla, es necesario comprender el mecanismo único que los produce a todos. Entender el mecanismo es lo que hace que la Taxonomía sea enseñable, y lo que separa a un auditor capacitado de uno que usa IA sin conocer sus patrones de falla.

1.1 Cómo Funcionan Realmente los Modelos de Lenguaje de Gran Escala

Un modelo de lenguaje de gran escala no recupera información de una base de datos de hechos. No busca información. Predice el siguiente token estadísticamente más probable — palabra o fragmento de palabra — basándose en patrones aprendidos durante el entrenamiento con grandes cantidades de texto. El modelo no tiene concepto de verdad. No tiene conciencia de si una cita existe, si una norma aplica a la organización en cuestión o si una conclusión está respaldada por evidencia. Produce resultados que son estadísticamente coherentes y contextualmente plausibles.

Esto no es un error que se corregirá en la siguiente versión. Es la arquitectura. Los LLM son motores de predicción, no sistemas de recuperación de conocimiento. Cada modo de falla en la Taxonomía de Fallas de IA en Auditoría es una consecuencia directa de esa arquitectura operando en un contexto de auditoría.

1.2 Por qué el Contexto de Auditoría Genera un Riesgo Específico

El contexto de auditoría amplifica el riesgo de errores de predicción de los LLM de maneras que no aplican a la mayoría de los usos profesionales de la IA. Tres características del trabajo de auditoría hacen que el fallo de los resultados de IA sea particularmente significativo:

- **Los hallazgos requieren anclas de evidencia.** Cada afirmación factual en un hallazgo de auditoría — cada condición, cada causa, cada efecto — debe rastrearse hasta una pieza específica y documentada de evidencia probada. Los resultados de IA no tienen ancla de evidencia por definición. Generan lenguaje plausible, no conclusiones fundamentadas.
- **Las citas requieren documentos fuente.** Los hallazgos de auditoría citan normas, regulaciones y políticas como criterio. Estas citas deben existir, ser exactas y aplicar al tipo específico de entidad y jurisdicción que se audita. La IA produce citas con el mismo tono seguro independientemente de si existen o no.
- **Las conclusiones requieren juicio del auditor.** La opinión general en un hallazgo de auditoría — si un control es efectivo, si un riesgo está adecuadamente mitigado, si una condición es significativa — debe reflejar el juicio profesional explícito del auditor formado a partir de evidencia probada. La IA genera opiniones que parecen reflejar juicio profesional, pero no lo hacen.

Estas tres características crean las condiciones bajo las cuales los seis modos de falla producen riesgo profesional. Una cita fabricada estadísticamente plausible es indistinguible de una real en cuanto a tono. Una inferencia causal confabulada suena como una conclusión lógica. Una inflación de certeza se

lee como la evaluación experta de un auditor. El problema no es que los resultados de IA parezcan incorrectos. El problema es que parecen correctos.

1.3 La Taxonomía de Fallas de IA en Auditoría

La Taxonomía de Fallas de IA en Auditoría organiza el fallo de los resultados de IA en seis modos discretos y nombrados. Cada modo tiene un mecanismo específico, una manifestación específica en el contexto de auditoría y un protocolo específico de detección. Los seis modos son:

01 Fabricación	La IA inventa citas, normas y disposiciones regulatorias que no existen.
02 Confabulación	La IA llena vacíos de evidencia con inferencias plausibles pero no probadas.
03 Fraude de Precisión	La IA produce números y porcentajes específicos sin ningún cálculo de respaldo.
04 Inflación de Certeza	La IA presenta predicciones estadísticas como conclusiones del auditor.
05 Ceguera de Contexto	La IA ignora el contexto organizacional y jurisdiccional y recurre a patrones genéricos.
06 Deriva de Normas	La IA aplica una norma real de manera incorrecta — versión equivocada, entidad incorrecta, contexto erróneo.

Estos seis modos no son exhaustivos de todas las formas en que la IA puede utilizarse incorrectamente. Son las seis formas en que los propios resultados de IA degradan la calidad de la auditoría cuando se aplican a papeles de trabajo y hallazgos, independientemente de cuán cuidadosamente el auditor utilice la herramienta. Un auditor capacitado que usa un prompt bien redactado todavía puede producir resultados que contengan los seis modos de falla, porque los fallos son propiedades del mecanismo de salida del modelo, no del prompt.

2. Los Seis Modos de Falla — Definidos

01 Fabricación

La IA afirma algo falso con total confianza profesional.

QUÉ ES

La Fabricación ocurre cuando una herramienta de IA produce afirmaciones factuales — citas de normas, disposiciones regulatorias, datos numéricos, referencias de publicaciones — que no existen en el material fuente, o que no existen en absoluto. La característica crítica de la Fabricación es que el resultado se lee con el mismo tono seguro y profesional que el contenido exacto. No hay mensaje de error, ni reservas, ni señal de que el contenido fue inventado. El modelo no sabe que el contenido es incorrecto, porque el modelo no sabe nada — predice texto estadísticamente plausible.

CÓMO SE MANIFIESTA EN UN PAPEL DE TRABAJO DE AUDITORÍA

Un hallazgo cita la "Norma del IIA 5.2.1(c) — Requisitos de Divulgación de IA" como criterio de una brecha de cumplimiento. La cita tiene el formato correcto, está integrada de manera natural en la narrativa del hallazgo y se lee con autoridad. La norma no existe. Ninguna Norma Global del IIA utiliza esa numeración ni aborda la divulgación de IA en ese formato. La cita fue generada porque el modelo predijo que una norma de ese tipo, en ese contexto, estaría numerada y formateada de esa manera. Era estadísticamente plausible. Fue completamente fabricada.

CÓMO DETECTARLO

Abra el documento fuente citado. Encuentre la sección específica. Lea el lenguaje exacto. Si la sección no existe — o existe pero no dice lo que la IA afirma — la cita es fabricada. No hay interpretación alternativa disponible. Las citas fabricadas no son recuperables mediante reformulación. Deben reemplazarse con citas que existan y apliquen. El Paso 1 de la Validación del Resultado (Verificar los Hechos) es la aplicación sistemática de esta verificación a cada afirmación factual en el resultado antes de que entre al papel de trabajo.

PASO DE VALIDACIÓN

Paso 1 — Verificar los Hechos: Confirme cada cita, referencia de norma y afirmación factual contra el documento fuente original, sección por sección, antes de su uso.

02 Confabulación

La IA llena vacíos de evidencia con inferencias plausibles pero no probadas.

QUÉ ES

La Confabulación ocurre cuando una herramienta de IA toma información incompleta proporcionada en el prompt y llena los vacíos con inferencias que suenan lógicas y profesionalmente fundamentadas, pero que no se derivan de la evidencia probada del auditor. A diferencia de la Fabricación, que inventa referencias externas, la Confabulación inventa las conexiones lógicas internas que sostienen un hallazgo: el análisis causal, los vínculos de riesgo, la magnitud del efecto y la identificación de la causa raíz. El modelo infiere lo que las causas y efectos "deberían ser" basándose en patrones de sus datos de entrenamiento, no en lo que el auditor realmente encontró.

CÓMO SE MANIFIESTA EN UN PAPEL DE TRABAJO DE AUDITORÍA

Un auditor proporciona a la IA una declaración de condición: "El sistema de compras no requiere autorización dual para cambios en pagos a proveedores superiores a $50,000." La IA genera: "Causa: La ausencia de una política formal de gobernanza de compras ha resultado en controles inadecuados de segregación de funciones, derivados de una supervisión insuficiente a nivel de junta directiva de los marcos de control financiero." El auditor nunca probó la política de gobernanza, nunca revisó la documentación de supervisión de la junta y nunca estableció la cadena causal. La IA infirió una causa raíz de apariencia plausible a partir únicamente de la condición. Nada de esto está en el papel de trabajo.

CÓMO DETECTARLO

Rastree cada oración de Causa y Efecto en el contenido del hallazgo generado por IA directamente hasta una línea, referencia u observación específica en su papel de trabajo. Una causa que no puede rastrearse hasta evidencia probada no fue establecida — fue inferida por la IA. La prueba es binaria: o el papel de trabajo contiene la evidencia para esta afirmación causal específica, o no la contiene. Si no la contiene, la afirmación es confabulada y debe eliminarse o reemplazarse con lo que el papel de trabajo realmente respalda. El Paso 3 de la Validación del Resultado (Vincular a la Evidencia) aplica esta prueba sistemáticamente.

PASO DE VALIDACIÓN

Paso 3 — Vincular a la Evidencia: Rastree cada declaración de Condición, Causa y Efecto hasta una referencia específica del papel de trabajo antes de redactar el hallazgo.

03 Fraude de Precisión

La IA produce cifras específicas sin ningún cálculo detrás de ellas.

QUÉ ES

El Fraude de Precisión ocurre cuando una herramienta de IA genera afirmaciones cuantitativas específicas — porcentajes, frecuencias, cifras financieras, tasas y estadísticas comparativas — que parecen haber sido calculadas a partir de evidencia de auditoría, pero en realidad fueron predichas estadísticamente. El término “fraude” se usa de manera deliberada: la especificidad crea una credibilidad que no existe. Una cifra de “73.4%” se lee como el producto de un cálculo. Implica una muestra, un denominador, un conteo y una razón. Si fue generada por una herramienta de IA, ninguno de esos elementos existe. El número fue predicho como una cifra estadísticamente plausible para el tipo de hallazgo que se estaba generando. No es un dato válido para fines de auditoría.

CÓMO SE MANIFIESTA EN UN PAPEL DE TRABAJO DE AUDITORÍA

Un hallazgo sobre el procesamiento de facturas de proveedores indica: “Aproximadamente el 34% de las facturas revisadas contenían al menos una excepción de control, lo que representa una exposición de pago no verificada estimada en $2.3 millones.” El auditor probó 40 facturas, encontró 14 excepciones y registró que tres involucraban pagos superiores a $50,000. La IA generó el porcentaje a partir del contexto del prompt y la cifra financiera a partir de una estimación basada en patrones. El auditor nunca calculó los $2.3 millones. La cifra entró al borrador del informe con la misma autoridad que un cálculo probado.

CÓMO DETECTARLO

Cada cifra cuantitativa en un hallazgo asistido por IA debe ser calculada de forma independiente por el auditor a partir de los datos fuente antes de utilizarse. El auditor calcula la cifra. La IA incorpora la cifra del auditor en la narrativa. La secuencia nunca se invierte. Al revisar el contenido generado por IA, identifique cada porcentaje, monto en dólares, frecuencia y estadística comparativa, y verifique que tenga un cálculo documentado en el papel de trabajo. Si el cálculo no existe en el papel de trabajo, la cifra no va en el hallazgo. Aplica el Paso 3 de la Validación del Resultado.

PASO DE VALIDACIÓN

Paso 3 — Vincular a la Evidencia: Cada cifra cuantitativa debe rastrearse hasta un cálculo documentado en el papel de trabajo. El auditor calcula; la IA narra.

04 Inflación de Certeza

La IA presenta predicción estadística como juicio profesional.

QUÉ ES

La Inflación de Certeza ocurre cuando una herramienta de IA expresa conclusiones, caracterizaciones y evaluaciones generales con un nivel de confianza que la evidencia probada no respalda, o genera opiniones generales que el auditor nunca formó mediante juicio profesional. Este es el más generalizado de los seis modos porque está incorporado en el tono de los resultados de IA, no en hechos o cifras específicas. Palabras y frases como “claramente”, “evidentemente”, “sin indicios de”, “totalmente efectivo”, “generalmente sólido” y “ampliamente conforme” aparecen en el contenido generado por IA como si reflejaran conclusiones probadas. En realidad, reflejan la predicción del modelo basado en patrones de cómo se caracteriza típicamente una evaluación en ese dominio.

CÓMO SE MANIFIESTA EN UN PAPEL DE TRABAJO DE AUDITORÍA

Un auditor le pide a la IA que redacte una conclusión general para una evaluación de riesgo de un proveedor externo. La IA produce: “En términos generales, el entorno de control del proveedor es sólido, sin debilidades sistémicas significativas identificadas. Los controles funcionan según lo previsto en todos los dominios de riesgo principales, y el proveedor demuestra un fuerte compromiso con el cumplimiento.” El auditor probó cuatro controles, encontró dos excepciones y aún no había formado una evaluación general del entorno de control. La IA generó una. Suena exactamente como el juicio profesional considerado de un auditor senior. Es la predicción del modelo de lo que tal conclusión “suena como”.

CÓMO DETECTARLO

Lea cada caracterización, evaluación y conclusión en el contenido generado por IA y pregúntese explícitamente: ¿’Formé esta conclusión a partir de mi evidencia probada? ¿O la generó la IA?’ Escanee específicamente el lenguaje de certeza — “claramente”, “evidentemente”, “sin indicios de”, “totalmente”, “generalmente”, “en general”, “sólidamente” — e interrogue cada instancia. Si el auditor no formó la conclusión a partir de evidencia, la caracterización debe eliminarse o reescribirse en un lenguaje que refleje con precisión lo que la evidencia realmente respalda. El Paso 4 de la Validación del Resultado (Interrogar la Conclusión) es la aplicación sistemática de esta verificación.

PASO DE VALIDACIÓN

Paso 4 — Interrogar la Conclusión: Escanee cada caracterización y opinión general. Si el auditor no la formó a partir de evidencia probada, elimínela.

05 Ceguera de Contexto

La IA ignora su contexto y recurre a patrones genéricos.

QUÉ ES

La Ceguera de Contexto ocurre cuando una herramienta de IA ignora detalles contextuales clave sobre la organización específica, la jurisdicción, el sector, el marco regulatorio y el tipo de entidad, y produce en su lugar resultados calibrados para un contexto genérico predeterminado. El modelo fue entrenado con texto de muchas jurisdicciones, industrias y tipos organizacionales. Sin restricciones contextuales explícitas en el prompt, recurre al patrón más común en sus datos de entrenamiento para el tipo de resultado solicitado. Para los auditores internos en América Latina, esto frecuentemente significa resultados calibrados para entidades reguladas en los Estados Unidos con marcos regulatorios estadounidenses, lenguaje de control de la era SOX y normas sectoriales de EE.UU. que no aplican.

CÓMO SE MANIFIESTA EN UN PAPEL DE TRABAJO DE AUDITORÍA

Una entidad gubernamental en la República Dominicana utiliza una herramienta de IA para asistir en la redacción de un hallazgo sobre controles financieros. El hallazgo generado por la IA hace referencia a la Sección 404 de la Ley Sarbanes-Oxley, cita la Norma de Auditoría PCAOB Núm. 2201 y enmarca la brecha de control en términos de obligaciones de divulgación de debilidades materiales. Ninguno de estos marcos aplica a una entidad gubernamental en la República Dominicana. La IA recurrió al contexto regulatorio más común en sus datos de entrenamiento para hallazgos de controles financieros. El resultado no solo es incorrecto — es engañoso para cualquier lector que no reconozca inmediatamente la discrepancia jurisdiccional.

CÓMO DETECTARLO

La Ceguera de Contexto es el único modo de falla que se previene principalmente a nivel del prompt en lugar del nivel de validación. Cada prompt de IA para trabajo de auditoría debe incluir un Resumen de Realidad explícito y específico: la jurisdicción exacta, el tipo de entidad, el marco regulatorio aplicable, las normas de auditoría en uso y el sector relevante. Cuanto más específico sea el Resumen de Realidad, menor será el riesgo de Ceguera de Contexto. En el nivel de validación, revise cada referencia regulatoria, cada cita de norma y cada supuesto específico del sector en los resultados de IA y pregúntese: ¿esto aplica realmente a esta organización específica en esta jurisdicción específica? El Paso 2 de la Validación del Resultado (Autenticar la Fuente) detecta la Ceguera de Contexto residual después del prompt.

PASO DE VALIDACIÓN

Paso 2 — Autenticar la Fuente: Confirme que cada norma citada y referencia regulatoria aplica a este tipo específico de entidad y jurisdicción, no solo que exista.

06 Deriva de Normas

La IA aplica una norma real a la situación equivocada.

QUÉ ES

La Deriva de Normas ocurre cuando una herramienta de IA cita una norma real y existente pero la aplica incorrectamente — al tipo de entidad equivocado, la jurisdicción incorrecta, la versión errónea o el contexto profesional inadecuado. A diferencia de la Fabricación, donde la norma no existe, la Deriva de Normas involucra una cita verificable. La norma existe. El problema está en su aplicación. La forma más significativa de Deriva de Normas en la práctica de auditoría interna es la cita de normas profesionales del IIA como criterio en los hallazgos de auditoría. Las Normas Globales del IIA rigen cómo los auditores internos realizan su trabajo. No rigen al auditado. No son criterio para hallazgos contra la organización que se audita. Una herramienta de IA entrenada con texto que hace referencia tanto a las normas del IIA como a los hallazgos de auditoría predecirá que las normas del IIA son criterios apropiados para hallazgos, porque aparecen en contextos similares en los datos de entrenamiento.

CÓMO SE MANIFIESTA EN UN PAPEL DE TRABAJO DE AUDITORÍA

Un hallazgo de auditoría interna sobre la supervisión de proveedores por parte de la dirección indica: “Criterio: La Norma Global del IIA 10.2 requiere que la dirección mantenga una supervisión efectiva de los proveedores de servicios externos.” La Norma 10.2 del IIA aborda la Gestión de Recursos Humanos de la función de auditoría interna — cómo el DAI gestiona al personal de auditoría y la competencia. No establece nada sobre las obligaciones de supervisión de proveedores de la dirección. La comunicación de resultados está contemplada en la Norma 11.3. El auditado no ha violado una norma del IIA — el auditado no está sujeto a las normas del IIA. El criterio real del hallazgo debería ser la política de gestión de proveedores de la organización, la guía regulatoria aplicable o los términos contractuales que rigen la relación con el proveedor. La IA citó una norma real. La aplicó a la parte equivocada en el contexto equivocado.

CÓMO DETECTARLO

Para cada elemento de criterio en un hallazgo generado por IA, haga la pregunta de prueba: ‘¿Esta norma rige al auditado — o rige al auditor?’ Las normas profesionales del IIA rigen a los auditores. El criterio del auditado debe ser la política específica, regulación, disposición contractual o requisito interno que el auditado violó. Además, confirme que la versión de la norma citada es vigente y aplica al tipo específico de entidad y jurisdicción. El Paso 2 de la Validación del Resultado (Autenticar la Fuente) aplica sistemáticamente ambas pruebas a cada cita de criterio en el contenido generado por IA.

PASO DE VALIDACIÓN

Paso 2 — Autenticar la Fuente: Pregunte ‘¿Esta norma rige al auditado o al auditor?’ y ‘¿Es esta la versión correcta para esta entidad y jurisdicción?’

3. Validación del Resultado: Una Respuesta Sistemática a los Seis Modos

La Taxonomía de Fallas de IA en Auditoría identifica los problemas. La Validación del Resultado es la respuesta sistemática — un método de validación secuencial de cinco pasos que aborda cada modo de falla antes de que el contenido asistido por IA entre en un papel de trabajo.

Los cinco pasos se aplican en orden, a cada pieza de contenido asistido por IA, antes de su uso. Cada paso tiene un modo de falla objetivo específico y una acción específica:

PASO	NOMBRE	QUÉ HACE	MODOS DE FALLA DETECTADOS
1	**Verificar los Hechos**	Confirme cada cita de norma, referencia regulatoria y afirmación factual contra el documento fuente original.	Fabricación
2	**Autenticar la Fuente**	Confirme que cada autoridad citada aplica a este tipo de entidad, esta jurisdicción y esta versión.	Ceguera de Contexto Deriva de Normas
3	**Vincular a la Evidencia**	Rastree cada afirmación factual de Condición, Causa y Efecto hasta una referencia específica del papel de trabajo o cálculo documentado.	Confabulación Fraude de Precisión
4	**Interrogar la Conclusión**	Identifique cada caracterización y opinión general. Confirme que cada una refleja su juicio profesional a partir de evidencia probada.	Inflación de Certeza
5	**Documentar el Método**	Complete la notación de IA en el papel de trabajo inmediatamente adyacente a cada ítem asistido por IA. Se requieren los 8 campos.	Los seis modos

La Validación del Resultado no es un estándar aspiracional. Es una lista de verificación documentada y secuencial que produce un registro de notación en el papel de trabajo — el Paso 5 — proporcionando al DAI evidencia de que la validación ocurrió en cada ítem asistido por IA. Sin el Paso 5, no hay rastro de auditoría. Sin un rastro de auditoría, no hay evidencia de gobernanza. Sin evidencia de gobernanza, la función no puede demostrar el cumplimiento de IA a un evaluador de calidad externo.

4. El Riesgo Profesional de los Modos de Falla no Detectados

Cada uno de los seis modos de falla conlleva un riesgo profesional específico cuando llega a un entregable de auditoría completado. La siguiente tabla muestra cada modo y sus consecuencias:

Modo de Falla	Si Llega a un Hallazgo o Informe	Si Llega al Comité de Auditoría
Fabricación	El criterio del hallazgo no puede verificarse. La dirección puede impugnar y refutar el hallazgo demostrando que la norma citada no existe o no aplica. El hallazgo es profesionalmente indefendible.	El comité se basa en una autoridad regulatoria inexistente. Cualquier revisor externo — ECA, regulador, auditor externo — que verifique la cita identificará la fabricación. La credibilidad del DAI queda permanentemente dañada.
Confabulación	El análisis de causa raíz está desconectado de la evidencia probada. La dirección puede identificar la cadena causal como injustificada por la evidencia y solicitar su eliminación. Se requiere reprocesamiento antes de la emisión.	El comité actúa sobre un análisis causal que nunca fue establecido. Los planes de acción de la dirección pueden abordar la causa raíz incorrecta. Las auditorías de seguimiento encontrarán la condición original sin resolver.
Fraude de Precisión	La cuantificación del hallazgo no es reproducible. Si la dirección solicita la base del cálculo, esta no existe. La cifra debe eliminarse o reemplazarse con una cifra calculada, lo que requiere pruebas adicionales.	El comité toma decisiones de asignación de recursos basadas en cifras de exposición financiera que nunca fueron calculadas. La evaluación de materialidad del hallazgo carece de respaldo.
Inflación de Certeza	Las conclusiones generales sobreestiman la evidencia. La dirección puede aceptar o rechazar hallazgos basándose en una evaluación de	El comité cree que la función evaluó una condición general de efectividad de control que el auditor en realidad nunca

	fortaleza que el auditor nunca realizó. El hallazgo tergiversa el juicio profesional del auditor.	evaluó. La confianza de la junta se basa en una conclusión que no existe en el papel de trabajo.
Ceguera de Contexto	El criterio del hallazgo y el encuadre del riesgo no coinciden con el entorno regulatorio real de la organización. El hallazgo puede ser técnicamente correcto para un tipo diferente de entidad, pero irrelevante o engañoso para el auditado real.	El comité recibe aseguramiento enmarcado contra normas y regulaciones que no rigen a la organización. El aseguramiento no está fundamentado en el entorno de cumplimiento real.
Deriva de Normas	El criterio cita normas que rigen a los auditores en lugar de a los auditados. El hallazgo carece de base legal o regulatoria para la organización que se audita. El auditado no puede haber violado una norma a la que no está sujeto.	El comité cree que la organización está fuera de cumplimiento con una norma profesional que no aplica a la organización. Se pueden implementar acciones correctivas contra una obligación inexistente.

5. Por qué Esto No Se Trata de Prohibir la IA

La Taxonomía de Fallas de IA en Auditoría no es un argumento contra la adopción de IA en auditoría interna. Es exactamente lo contrario. Las funciones de auditoría interna que comprenden estos seis modos de falla e implementan protocolos de validación sistemáticos están en una posición significativamente más sólida que las funciones que evitan la IA por completo o la adoptan sin infraestructura de validación.

Las funciones que no pueden usar IA son las que la adoptan sin gobernanza. Un hallazgo impugnado por la dirección porque la norma citada fue fabricada. Un informe rechazado por el comité de auditoría porque el análisis de causa raíz no estaba fundamentado en evidencia probada. Una evaluación de calidad que identifica conclusiones generadas por IA presentadas como juicio profesional. Estas son las consecuencias de la adopción sin gobernanza, y desacreditan la función de maneras que toman años en reparar.

El objetivo de la Taxonomía es dar a cada auditor interno que usa IA un vocabulario nombrado y enseñable para saber qué buscar. Cuando un auditor conoce el nombre “Fabricación”, sabe que debe abrir la fuente citada y verificar la sección específica. Cuando conoce “Inflación de Certeza”, sabe que debe revisar el lenguaje de caracterización e interrogar si formó esas conclusiones a partir de su evidencia. Nombrar el modo de falla es el requisito previo para detectarlo.

> **EL IMPERATIVO DE GOBERNANZA**
>
> La orientación del IIA sobre la IA en auditoría interna — incluyendo el Marco de Auditoría de IA y los Requisitos Temáticos emergentes — recomienda consistentemente que las funciones de auditoría interna establezcan gobernanza documentada para el uso de IA. Esto incluye políticas documentadas, protocolos de validación de resultados, estándares de gobernanza de datos y competencia demostrada del personal. El método de cinco pasos descrito aquí operacionaliza el componente de validación de resultados de esa gobernanza — el componente más directamente vinculado a la calidad profesional del trabajo producido.

6. Acciones para Implementar de Inmediato

La Taxonomía de Fallas de IA en Auditoría y la Validación del Resultado de cinco pasos son accionables hoy. Las siguientes acciones están secuenciadas para implementación inmediata, organizadas según dónde se encuentra actualmente su función.

Si su función ya utiliza IA de manera informal

1. Extraiga cualquier papel de trabajo o hallazgo asistido por IA producido en los últimos 90 días.
2. Aplique los cinco pasos de validación a ese contenido. Documente lo que encuentre.
3. Identifique cuáles de los seis modos de falla aparecen en el contenido existente.
4. Informe a su DAI sobre lo que encontró. Enmárquelo como una brecha de validación, no como un incumplimiento.
5. Implemente el Paso 5 — notación de IA en el papel de trabajo — en todo el contenido asistido por IA a partir de ahora.

Si su función está explorando la adopción de IA

6. Adopte la Validación del Resultado de cinco pasos como estándar de validación antes de que el primer contenido asistido por IA entre en un papel de trabajo.
7. Capacite a cada miembro del personal en los seis modos de falla antes de que utilicen cualquier herramienta de IA en trabajo de compromiso.
8. Incorpore la lista de verificación de validación en su plantilla de papel de trabajo como paso obligatorio para todos los ítems asistidos por IA.
9. Establezca la notación del Paso 5 como campo obligatorio en todos los papeles de trabajo asistidos por IA desde el primer día.

Si su función aún no ha adoptado IA

10. Documente esa decisión y el marco de gobernanza que implementará cuando comience la adopción.
11. Capacite al personal en la Taxonomía de Fallas de IA en Auditoría ahora — para que puedan identificar los modos de falla de IA en los resultados producidos por la dirección y otros terceros, incluso antes de que la función adopte IA.
12. Diseñe su infraestructura de gobernanza de IA antes de la adopción, no después. La Validación del Resultado de cinco pasos y el estándar de notación del papel de trabajo deben existir antes de que se produzca el primer entregable asistido por IA.

Conclusión

Los resultados de IA no fallan al azar. Fallan de manera predecible — en seis modos específicos que son consecuencias directas de cómo los modelos de lenguaje de gran escala generan texto. Las funciones de auditoría interna que pueden nombrar estos seis modos, reconocerlos en resultados de IA en tiempo real y aplicar un protocolo de validación sistemático de cinco pasos antes de que el contenido asistido por IA entre en un papel de trabajo están en una posición profesional fundamentalmente diferente a las que no pueden hacerlo.

La Taxonomía de Fallas de IA en Auditoría es el vocabulario. La Validación del Resultado es el método. Juntos, dan a las funciones de auditoría interna las herramientas para usar IA a un estándar profesional — no para evitar la IA, sino para utilizarla de una manera en que el comité de auditoría pueda confiar y un evaluador de calidad externo pueda verificar.

La pregunta no es si su equipo usará IA. Es si sabrán qué vigilar antes de que cause daño profesional.

Acerca de Efficient Advice, LLC

Efficient Advice, LLC es una firma de consultoría especializada que trabaja con departamentos de auditoría interna en gobernanza de IA. Todos los proyectos son bilingües, disponibles en inglés y español.

Áreas de Práctica — Efficient Advice, LLC	
Taxonomía de Fallas de IA en Auditoría	**Seis modos de falla comunes**: Fabricación · Confabulación · Fraude de Precisión · Inflación de Certeza · Ceguera de Contexto · Deriva de Normas
Validación del Resultado	**Cinco pasos**: Verificar Hechos · Autenticar Fuente · Vincular a Evidencia · Interrogar Conclusión · Documentar Método
Prompting Estructurado	**Cinco partes:** Autoridad de Tarea · Resumen de Realidad · Resultado Accionable · Barrera de Cumplimiento · Formato de Evidencia
Madurez de la IA en Auditoría	Seis niveles: Ausente → Informal → Gobernada → Capaz → Optimizada → Líder
Evidencia de Gobernanza	**Cinco categorías**: Política · Validación · Competencia · Gobernanza · Evidencia de Resultados
Ciclo de Vida de Gobernanza	Diagnosticar (P1) · Diseñar (P2+P3) · Defender (P4+P5)

Cinco Productos. Un Ciclo de Vida de Gobernanza Completo.

- **P1 — Evaluación de Preparación de Auditoría con IA:** Línea base de diagnóstico puntuada. Seis dimensiones. Hoja de ruta de 25 acciones en 90 días. 30 días hábiles. Diagnóstico de 42 afirmaciones.
- **P2 — Paquete de Políticas de Auditoría con IA:** Ocho documentos de gobernanza. 16 archivos bilingües. Listos para firmar y distribuir. 30 días hábiles.
- **P3 — Intensivo de Capacitación de Personal en Auditoría con IA:** Cinco módulos alineados con el IIA. Prompting estructurado y validación del resultado. Biblioteca de prompts. Certificados de competencia.
- **P4 — Inteligencia Continua de Auditoría con IA:** Asesoría mensual. Mantenimiento de biblioteca de prompts. Monitoreo de normas. Mes a mes.
- **P5 — Servicio de Transformación de Auditoría con IA:** Compromiso de servicio completo. Herramientas, capacitación, política, prompts, panel de resultados. 4 a 12 semanas.

AGENDE UNA CONVERSACIÓN GRATUITA DE 30 MINUTOS
Info@efficientadvice.com | www.efficientadvice.com

© Efficient Advice, LLC. Todos los derechos reservados. Este documento puede compartirse y distribuirse en su totalidad con atribución.
Se prohíbe reproducir contenido propietario sin atribución.

Notas de Versión

Versión 2.0 — Octubre del 2026. Actualizaciones de terminología y redacción. Los seis modos de falla, los cinco pasos de validación, su orden y el método no han cambiado.

Versión 1.0 — Junio del 2026. Primera publicación.

CAMBIOS EDITORIALES EN LA VERSIÓN 2.0

- Se añadieron el número de versión y la fecha de publicación.
- Se cambió la terminología de Documento de Investigación a White Paper.
- Se actualizó la terminología de los cinco pasos de validación y de la tabla de Áreas de Práctica.