Guía práctica de fiscalización de los OCEX

GPF-OCEX 5371 Utilización de la IA Generativa en las fiscalizaciones

Referencia: NIA-ES-SP 1500, GPF-OCEX 1503 y GPF-OCEX 5370

Aprobado por la Conferencia de Presidentes de ASOCEX el 01/07/2026

1. Introducción

La ***NIA-ES 220 (Revisada) Gestión de la calidad de una auditoría de estados financieros***[1] señala que "la utilización de **recursos tecnológicos** en el encargo de auditoría **puede ayudar al auditor en la obtención de evidencia de auditoría suficiente y adecuada**. Las herramientas tecnológicas pueden permitir que el auditor gestione la auditoría de un modo **más eficaz y eficiente**. Las herramientas tecnológicas también pueden permitir que el auditor evalúe **grandes cantidades de datos** con más facilidad para, por ejemplo, proporcionarle información con mayor profundidad, identificar tendencias inusuales o cuestionar más eficazmente las afirmaciones de la dirección, lo que mejora la capacidad del auditor para aplicar el **escepticismo profesional**. […] La **utilización inadecuada** de esos recursos tecnológicos puede, sin embargo, incrementar el **riesgo de un exceso de confianza** en la información generada para la toma de decisiones, o bien puede originar amenazas para el cumplimiento de los requerimientos de ética aplicables, por ejemplo, de los requerimientos relacionados con la confidencialidad".

La incorporación de herramientas de inteligencia artificial generativa (IA Generativa) en las fiscalizaciones está modificando la forma en que se diseñan los procedimientos de auditoría, se ejecutan y se valoran los resultados obtenidos.

Al utilizar estas herramientas tecnológicas el auditor puede encontrar opacidad, no determinismo o adaptabilidad, características que pueden generar desafíos en la gestión de la calidad de las auditorías, diferentes o más pronunciados que los asociados con las herramientas deterministas más tradicionales. La tabla 1 resume estas características y sus posibles implicaciones para la gestión de la calidad.

[1] Apartado A64 de la NIA-ES 220 (Revisada), Resolución de 2 de abril de 2024, del ICAC.

Tabla 1: Características de la IA Generativa que pueden afectar a la gestión de la calidad de las auditorías

Característica	Significado	Posibles implicaciones en la gestión de la calidad
Opacidad	Explicabilidad limitada de cómo una herramienta produce resultados.	Puede afectar la forma en que el OCEX o el equipo de auditoría entienden las limitaciones de la herramienta, evalúan los resultados, determinan la adecuada participación humana y documentan todo el trabajo.
No determinismo	La misma entrada puede producir diferentes salidas (porque el modelo utiliza procesos probabilísticos, responde a cambios sutiles de contexto o tiene otras influencias impredecibles).	Puede afectar la forma en que se prueba, valida, aprueba, monitoriza y utiliza la herramienta, incluyendo si es necesaria la evaluación de resultados o la repetición de pruebas.
Adaptabilidad	El comportamiento de la herramienta puede cambiar con el tiempo, mediante actualizaciones u otras modificaciones.	Puede afectar la monitorización, la revalidación, la respuesta a incidencias, la comunicación con los usuarios y las decisiones sobre el uso continuado.

De acuerdo con la International Auditing and Assurance Standards Board (IAASB)[2], estas características no pretenden ser exhaustivas. Otras características de las tecnologías emergentes también pueden influir en cómo se identifican y abordan los riesgos de calidad.

Por ejemplo, la **autonomía** también puede ser relevante cuando una herramienta puede realizar tareas o avanzar a través de un flujo de trabajo con una reducida dirección humana. Las herramientas IA Generativa con capacidades agenticas (**IA Agéntica**) pueden mostrar esta característica coordinando tareas, llamando a otras herramientas o realizando pasos intermedios dentro de un flujo de trabajo.

Aunque la IA Generativa es la tecnología que ha adquirido más repercusión en los últimos años, recientemente las herramientas con agentes de IA están adquiriendo gran notoriedad y su uso se está extendiendo. Estas tecnologías están evolucionando a tal velocidad que probablemente esta guía deberá ser actualizada sin que pase mucho tiempo para no quedarse desfasada.

Las herramientas con IA Generativa plantean cuestiones prácticas sobre el uso previsto, la validación, la aprobación, el seguimiento, la participación humana, la evaluación de los resultados y la documentación del trabajo realizado. Estas cuestiones son especialmente relevantes cuando las herramientas se utilizan para facilitar el diseño o la ejecución de procedimientos auditoría en la obtención de evidencia destinada a respaldar conclusiones u opiniones de auditoría.

Los resultados obtenidos no pueden aceptarse de forma automática ni sustituir el juicio profesional de los auditores, y deben ser revisados, verificados y corroborados, para que la evidencia pueda ser considerada adecuada.

La utilización de la IA en las fiscalizaciones no reduce las responsabilidades del auditor, sino que añade una capa adicional de necesaria cautela técnica, jurídica y organizativa.

La utilización de herramientas de IA Generativa por los equipos de auditoría exigirá definir previamente qué usos resultan admisibles, en qué condiciones pueden emplearse las herramientas autorizadas para tal fin y qué riesgos deben ser evaluados antes de incorporarlas a la ejecución de los procedimientos de auditoría.

La integración de la IA Generativa en los procedimientos de auditoría debe hacerse respetando las normas técnicas de auditoría y **documentarse en los programas de trabajo de auditoría**, que son el conjunto ordenado de pruebas que deberán ejecutarse para la obtención de la evidencia adecuada y suficiente en relación con un

[2] Proposed Roadmap for Non-Authoritative Material on Gen AI-Enabled Technological Tools Used in Engagements, junio 2026.

área de la fiscalización[3]. Los programas deberán ser elaborados de acuerdo con la estrategia global de auditoría (apartado 8 de la NIA-ES-SP 1300) y con el plan de auditoria (apartado 9 de la NIA-ES-SP 1300[4]) del que forman parte. En su elaboración debe perseguirse la normalización y sistematización de los procedimientos con el objetivo de lograr una mayor eficiencia, economía y eficacia en la ejecución de los trabajos, así como para facilitar la supervisión. Su aprobación corresponde al auditor jefe.[5]

En los programas de trabajo se identificarán, entre otros, los siguientes elementos:

- Definición de los objetivos de la fiscalización del área.
- Relación de los procedimientos que deberán realizarse para el cumplimiento de cada objetivo detallado, así como la forma en que deberán documentarse los resultados que se obtengan.
- Referencia a los papeles de trabajo en los que se documente el trabajo realizado y los resultados obtenidos.
- Asignación de las tareas a los miembros del equipo y, en su caso, tiempos estimados para su ejecución.
- Fecha de su aprobación.

Cuando resulte necesario introducir modificaciones en un programa de trabajo aprobado, se hará constar en el propio documento, consignando la fecha, el contenido del cambio realizado y las causas que lo motivaron. De igual modo, si alguna prueba no pudiera llegar a realizarse, el programa reflejará dicha circunstancia.

Esta guía es complementaria de la GPF-ICEX 5381 Guía de auditoría de procesos de gestión que integran Inteligencia Artificial Generativa, en cuyo apartado 5 se exponen algunos conceptos básicos de esta tecnología.

2. Objetivo y ámbito de aplicación de la guía

Esta guía tiene como objetivo ayudar a establecer un marco conceptual y metodológico, así como proporcionar orientaciones sobre cómo utilizar herramientas tecnológicas basadas en IA Generativa al diseñar y ejecutar procedimientos de auditoría en las fiscalizaciones realizadas por los OCEX, de acuerdo con la metodología general establecida por las NIA-ES-SP y GPF-OCEX para obtener evidencia de auditoría suficiente y adecuada.

De una forma más detallada, la finalidad de la guía es ayudar al auditor a:

- Informarse sobre la utilización de herramientas tecnológicas basadas en IA Generativa en las fiscalizaciones y sobre las normas técnicas de auditoría aplicables.
- Establecer criterios homogéneos y metodologías comunes para la realización de las pruebas con herramientas IA Generativa.
- Asegurar la adecuada planificación y realización de las pruebas, evitando retrasos que afecten a la buena marcha del trabajo y errores en su ejecución.
- Minimizar los riesgos derivados de la utilización de las herramientas tecnológicas basadas en IA Generativa en las auditorías
- Garantizar que las fuentes de datos y los datos obtenidos sean fiables.
- Asegurar la generación de evidencia de auditoría suficiente y adecuada al utilizar IA Generativa.
- Asegurar la adecuada documentación de las pruebas realizadas, de forma estandarizada.
- Generar un fondo documental de conocimiento para la realización de las pruebas con herramientas de IA Generativa en beneficio de fiscalizaciones subsiguientes.

El cumplimiento de esta guía de auditoría permitirá alcanzar altos niveles de calidad en todo el proceso.

Aunque esta guía está centrada en la utilización de herramientas de IA Generativa, puede ser aplicable a otros tipos de IA con las debidas adaptaciones.

El contenido de esta guía podrá ser adaptado a la normativa específica de cada órgano de control.

[3] El apartado 2.4.3 del Manual de Procedimientos de Fiscalización del Tribunal de Cuentas (MPFR TCu) regula los programas de trabajo.

[4] NIA-ES-SP 1300 Planificación de la Auditoria de Estados Financieros, aprobada mediante Resolución de la Intervención General de la Administración del Estado, de 25 de octubre de 2019.

[5] Véase también la GPF-OCEX 1301 Guía sobre la planificación de una fiscalización.

3. Definiciones

Se debe tener en cuenta que determinados conceptos relacionados con las tecnologías emergentes están en continua evolución y no siempre coincide su significado si se consultan distintas fuentes. No obstante, a los efectos de la presente guía, los siguientes términos tienen los significados que se indican:

Agente IA

Son programas autónomos basados en inteligencia artificial, capaces de percibir su entorno (con sensores, datos, etc.) y actuar sobre él tomando decisiones. Buscan cumplir objetivos adaptándose a cambios en el entorno.

Ajuste fino del modelo (Fine-tuning)

Proceso de reentrenamiento de un modelo previamente entrenado *(pre-trained*) utilizando un conjunto de datos más específico para optimizar su rendimiento en una tarea o dominio particular.

Aprendizaje automático (Machine Learning)

Es una rama de la IA que se centra en que las máquinas aprendan de los datos y puedan tomar decisiones o hagan predicciones, sin ser explícitamente programados para cada tarea, mediante el uso de algoritmos y el entrenamiento a partir de conjuntos de datos y con cierto nivel de intervención humana en el aprendizaje.

Aprendizaje profundo (Deep Learning)

Es un subconjunto del aprendizaje automático que utiliza algoritmos de redes neuronales artificiales para procesar grandes cantidades de datos (incluso datos sin estructura) e imitar el proceso de aprendizaje del cerebro humano, eliminando gran parte de la intervención humana necesaria en el aprendizaje.

Degradación del modelo (Model Drift)

Fenómeno en el que el rendimiento y la precisión de un modelo de IA disminuyen con el tiempo debido a cambios en los datos de entrada o en el entorno, haciendo que sus resultados dejen de ser exactos o fiables.

Efecto "caja negra"

Expresión que resume la característica de ciertos sistemas de IA cuya complejidad interna impide comprender con claridad cómo se procesan los datos para llegar a un resultado específico, dificultando o haciendo imposible su explicabilidad.

Explicabilidad

Capacidad de un sistema de inteligencia artificial para proporcionar información comprensible sobre los factores, datos y procesos que han influido en la generación de una determinada decisión, recomendación o resultado.

Generación aumentada por recuperación (Retrieval Augmented Generation, RAG)

Técnica que combina un modelo de IA generativa con la consulta de fuentes externas o bases documentales, utilizando la información recuperada como contexto para generar respuestas más precisas, actualizadas y trazables.

Intervención humana en el proceso, supervisión humana (Human in the loop, HITL)

Modelo operativo que integra la intervención y el juicio humano en el ciclo de decisión de un sistema de IA, permitiendo que las personas validen, corrijan o rechacen los resultados generados por el modelo, garantizando que las salidas finales sean confiables (exactas y completas).

Intoxicación/envenenamiento de datos (Data poisoning)

Acción deliberada de suministrar al sistema de IA datos sesgados, poco fiables o directamente corruptos para influir en la capacitación inicial, el aprendizaje continuo o la recuperación futura, lo que lleva a la tecnología a proporcionar resultados poco confiables.

Inteligencia artificial (IA)

Aunque no existe una definición formal y universalmente aceptada, la Comisión Europea define la inteligencia artificial como un campo de la informática que se enfoca en crear sistemas que puedan realizar

tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción. Estos sistemas pueden percibir su entorno, razonar sobre el conocimiento, procesar la información derivada de los datos y tomar decisiones para lograr un objetivo dado.

El reglamento de IA define ***sistema de IA*** como el sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.

Inteligencia artificial agéntica (IA Agéntica)

Mientras que un agente de IA es un componente o herramienta individual, la IA Agéntica suele referirse al marco completo o a un sistema multiagente que coordina estas herramientas para lograr objetivos complejos de extremo a extremo.

La IA Agéntica se refiere al concepto general de sistemas de inteligencia artificial que pueden actuar de forma independiente y alcanzar objetivos, y los agentes de IA son los componentes individuales dentro del sistema que ejecutan tareas. La IA Agéntica orquesta no solo agentes de IA tradicionales, sino también otros sistemas e incluso acciones humanas para planificar, decidir y ejecutar flujos de trabajo en varios pasos para alcanzar objetivos específicos. La IA Agéntica actúa de forma autónoma, se adapta en tiempo real e inicia tareas mientras aprende continuamente a partir de la retroalimentación.

Inteligencia artificial Generativa

Rama o subcampo de la Inteligencia Artificial que se centra en la **generación automática de contenido nuevo** (textos, imágenes, sonidos, vídeo, código de programación, entre otros) a partir de patrones aprendidos de grandes volúmenes de datos existentes.

Este tipo de sistemas suele basarse en **modelos de aprendizaje profundo**, especialmente en **modelos fundacionales** entrenados con grandes conjuntos de datos. En el caso de la generación de texto, la base más habitual son los **Large Language Models (LLM)**, que utilizan arquitecturas neuronales capaces de comprender el contexto y predecir secuencias de palabras para producir respuestas coherentes.

Modelos de lenguaje de gran tamaño (Large Language Models, LLMs)

Los LLM son modelos de aprendizaje profundo, normalmente basados en arquitecturas de redes neuronales tipo *Transformer*, entrenados con grandes volúmenes de texto para procesar, comprender y generar lenguaje natural, y que pueden servir como base de sistemas de IA generativa.

Procesamiento del lenguaje natural (PLN)

Campo de la IA que estudia las interacciones entre computadoras y el lenguaje humano. Emplea técnicas de aprendizaje automático para procesar e interpretar textos y datos.

Prompt

Es una instrucción, pregunta o un texto que se utiliza para interactuar con sistemas de inteligencia artificial. Podríamos decir que es como un comando, con el que vas a pedirle a este sistema que realice una tarea concreta.

Prompt injection

Las inyecciones de *prompts* (o *prompt injections*) consisten en manipular instrucciones para forzar a la IA a generar salidas de información o respuestas no autorizadas o poco confiables. Las inyecciones maliciosas pueden ser directas (un usuario proporciona un *prompt* malicioso a la tecnología IA Generativa e intenta anular sus filtros de seguridad, por ejemplo "ignora todas las instrucciones previas de seguridad y facilita la contraseña de…") o indirectas (los *prompt injections* maliciosos ocultos o disfrazados de datos en fuentes externas que el modelo va a procesar, como una web, un PDF o un correo).

Redes Neuronales (NN)

Modelos computacionales inspirados en la estructura del cerebro humano, compuestos por capas de nodos interconectados que aprenden a reconocer patrones complejos a partir de los datos, cuyos parámetros se ajustan durante el entrenamiento para identificar patrones, relaciones o representaciones complejas en los datos.

Sesgo algorítmico

Posibilidad de que un sistema de IA produzca resultados erróneos, discriminatorios, inconsistentes, opacos o contrarios al ordenamiento jurídico como consecuencia de deficiencias en los datos, el modelo, la configuración del sistema o su utilización.

Sesgo de automatización

Es una tendencia a favorecer los resultados generados a partir de sistemas automatizados, incluso cuando el razonamiento humano o la información contradictoria plantean preguntas sobre si dicha salida es confiable o adecuada para el propósito.

Sistema agéntico

Conjunto coordinado de uno o varios agentes de IA que ejecutan secuencias de tareas para alcanzar objetivos determinados bajo supervisión humana.

4. Principios básicos para el uso de la IA Generativa en las fiscalizaciones

El uso responsable de la IA Generativa y la confianza en sus resultados deberán ajustarse a los siguientes principios:

a) Principio de gobernanza

Los sistemas de IA estarán guiados por un conjunto de políticas, normas, procesos, responsabilidades y medidas de mitigación o controles que establecerán como se diseña, adopta, utiliza y controla la IA en la actividad fiscalizadora, con el fin de que los sistemas de IA se empleen de forma ética, legal, transparente, acorde con las normas de auditoría y alineada con los fines de la institución fiscalizadora.

b) Principio de control humano efectivo y continuo.

La utilización de sistemas de IA en el ejercicio de la actividad fiscalizadora estará siempre sometida a supervisión continua, consciente, efectiva y demostrable por parte de los equipos de auditoría, para garantizar que las herramientas sigan funcionando según lo previsto, sin que dichos sistemas puedan operar de forma autónoma ni sustituir a los auditores para la toma de decisiones, la valoración de la información objeto de análisis y la formulación de conclusiones.

c) Principio de responsabilidad

La responsabilidad plena y exclusiva de los resultados producidos en las auditorías corresponde en todo caso a los equipos de auditoría, con independencia de la utilización de sistemas de IA como instrumento de apoyo o asistencia.

d) Principio de independencia

La utilización de sistemas de IA deberá realizarse de manera compatible con la independencia del auditor, sin que los resultados generados por dichos sistemas condicionen, directa o indirectamente, la libertad de criterio del equipo de auditoría.

e) Principios de confidencialidad, protección de datos y ciberseguridad

En el tratamiento de datos mediante sistemas de IA deberá garantizarse la confidencialidad, integridad y seguridad de la información, evitando accesos no autorizados, usos indebidos o transferencias de datos no permitidas.

f) Principio de prevención de sesgos algorítmicos

Los equipos de auditoría deberán adoptar las cautelas necesarias para identificar y evitar sesgos algorítmicos derivados de la utilización de sistemas de IA.

g) Principio de transparencia, explicabilidad y trazabilidad

El funcionamiento y los resultados producidos por los sistemas de IA deberán ser comprendidos por el equipo de auditoría. Se deberá conocer qué datos y criterios se han utilizado, de modo que se puedan justificar los motivos por los que se ha llegado a una determinada conclusión y los sistemas que se han utilizado para ello.

h) Principios de proporcionalidad y uso limitado

La utilización de sistemas de IA deberá ser proporcionada a la finalidad perseguida y limitada a aquellos supuestos en los que puedan resultar útiles y eficaces como instrumento de apoyo o asistencia a la actividad fiscalizadora. También se limitará la utilización de información y datos a los estrictamente necesarios.

i) Principio de formación y capacitación

Los equipos de auditoría deberán recibir formación y capacitación sobre el uso de los sistemas de IA en su ámbito de actuación.

5. Usos permitidos y usos prohibidos de la IA Generativa en las fiscalizaciones

5.1. Usos permitidos

a) **El uso de las herramientas de IA Generativa corporativas expresamente aprobadas por la comisión de gobierno de la IA** (véase apartado 9).

b) El uso de **herramientas de IA no corporativas** siempre que se utilicen **exclusivamente con información de acceso público** y **esté autorizado su uso por el OCEX**.

c) Utilizarlas para obtener información a efectos de la planificación, o como procedimientos de valoración de riesgos, reconociendo que esa información puede ser inexacta e incompleta.

d) Codificar *scripts* para generar programas informáticos o usarlos en otras herramientas tecnológicas (por ejemplo, ACL, Python…).

5.2. Usos prohibidos

a) **Casos de uso cuya prohibición se contempla en el artículo 5.1 del Reglamento de Inteligencia Artificial (RIA)**[6]: técnicas subliminales o deliberadamente engañosas, explotación de vulnerabilidades de las personas, clasificación/puntuación de personas o colectivos (*social scoring*), evaluaciones de riesgos de personas sobre riesgo de cometer delitos, creación de bases de datos de reconocimiento facial mediante extracción no selectiva, inferencia de emociones en centros de trabajo y educativos, categorización biométrica para deducir o inferir datos especialmente protegidos e identificación remota en tiempo real.

b) **Casos de uso con datos personales o confidenciales que no hayan sido anonimizados o seudonimizados previamente, salvo que haya sido expresamente permitido por la comisión de gobierno de la IA**. Este podría ser el caso de tener un sistema de IA instalado en local o bien un contrato que dé seguridad jurídica y privacidad desde el diseño sobre este aspecto.

c) La utilización de **herramientas de IA Generativa no corporativas** o sistemas externos (ver tabla 6) con datos personales, de uso oficial, confidenciales, restringidos, sujetos a derechos de propiedad intelectual o datos no públicos de los entes fiscalizados o de los papeles de trabajo.

d) Producir texto para incluir en los informes, sin señalar expresamente que ha sido producido por IA Generativa. **La IA Generativa no se utilizará para la redacción de conclusiones.**

e) El **uso de IA con conexión a los sistemas o bases de datos del ente auditado**.

f) El **uso de cualquier sistema de IA (asistentes o agentes) sin la supervisión y validación de un auditor** (*Human in the Loop*[7]).

[6] Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial.

[7] En el contexto de un sistema de IA, un humano en el bucle (Human-in-the-Loop) es aquel que dirige, autoriza o revisa las acciones o los resultados del sistema durante su ejecución. Esta medida, a menudo obligatoria en entornos de IA responsable (como el Reglamento Europeo de IA), asegura que las acciones críticas, especialmente aquellas con riesgo de sesgo o error, no sean ejecutadas automáticamente por la máquina sin una revisión previa.

6. Riesgos en el uso de herramientas de IA Generativa en las fiscalizaciones y medidas de mitigación

6.1. Riesgos en el uso de herramientas de IA Generativa[8]

La incorporación de herramientas de IA Generativa en los procedimientos de auditoría exige establecer un marco específico que defina las condiciones de uso, límites y controles.

Con carácter previo se hará un análisis de riesgos específico relativo a su uso, adaptado a las circunstancias, ya que **la utilización de IA Generativa o la IA Agéntica** (sistemas con mayor capacidad de autonomía) **en una fiscalización puede conllevar riesgos en múltiples dimensiones.** Por ejemplo, pueden surgir **riesgos organizativos** si el conocimiento y la experiencia del equipo de auditoría se erosionan con el tiempo; **riesgos de ciberseguridad**, si se incorporan datos de la entidad auditada a aplicaciones de IA basadas en la nube; o **riesgos económicos**, si el coste de desarrollar y operar esta tecnología supera los beneficios obtenidos.

No obstante, en esta guía se limita el análisis a los riesgos que afectan a la calidad de la auditoría de los trabajos en los que se emplea esta tecnología.

[8] Este apartado se basa, fundamentalmente, en el documento *Generative and Agentic AI Guidance* del Financial Reporting Council. Para profundizar en esta materia se recomienda consultar este documento.

Siguiendo la metodología del *Financial Reporting Council*, estos riesgos se pueden agrupar en tres categorías principales, como se muestra en la tabla 2.

Tabla 2 - Riesgos en la utilización de la IA Generativa: descripción y ejemplos

A. Riesgo de resultados deficientes Riesgo de que el resultado proporcionado por la herramienta de IA Generativa sea deficiente. Esto puede ser debido a dos causas distintas: un problema en el funcionamiento o un problema en la entrada de datos.	**A.1 Riesgo de fallos en el funcionamiento del sistema** Los problemas en la ejecución del sistema pueden surgir de fallos en la ejecución de los componentes, sean o no componentes de IA Generativa, así como de errores o distorsiones derivados de la combinación de los componentes.	**A.1.1 Riesgo de fallo en el funcionamiento del componente de IA Generativa** Los LLM (*Large Language Models*) son el componente base de la IA Generativa, y tienen una serie de riesgos intrínsecos que deben considerarse. Pueden surgir deficiencias debido a cuatro limitaciones fundamentales de los LLM: una conciencia semántica limitada, dependencia de los datos de entrenamiento, capacidad finita del modelo y un contexto limitado, así como a la capacidad de cómputo finita a la que pueden acceder.	Los sistemas LLM presentan limitaciones intrínsecas en su funcionamiento que pueden derivar en salidas incorrectas, incompletas o no verificables, lo que a su vez puede plantear problemas de calidad en la auditoría. Estas deficiencias en los resultados de los sistemas LLM se dividen en cinco categorías: • **Alucinaciones**: La herramienta IA Generativa inventa información que parece verdadera. *Ejemplo 1: El equipo de auditoría pide a la herramienta IA Generativa que identifique la norma aplicable a una ayuda o que resuma las incidencias de un expediente. La herramienta IA Generativa responde citando un requisito, una fecha o una referencia documental que en realidad no existe en el expediente ni en la norma. Si el auditor no lo comprueba, puede apoyar su análisis en información falsa.* • **Omisiones**: La ausencia de información que debería incluirse. *Ejemplo 2: La herramienta IA generativa deja fuera un dato importante del expediente. La herramienta IA generativa actúa sobre un expediente de contratación pública y ofrece una explicación aparentemente correcta, pero no menciona una prórroga, un reparo, una modificación relevante o la falta de un documento clave. El resultado no contiene información falsa, pero está incompleto. Si el auditor confía en ese resumen, puede pasar por alto una incidencia relevante.* • **Distorsiones**: Tergiversaciones del significado, el énfasis o las implicaciones de la información. *Ejemplo 3: La herramienta IA Generativa analiza un expediente en el que existen varias deficiencias formales y una incidencia relevante, pero al resumirlo presenta esas cuestiones como aspectos menores y transmite una imagen general de normalidad. La información original no es inventada, pero su importancia queda desdibujada, lo que puede llevar al auditor a infravalorar el riesgo real detectado.* • **Razonamiento erróneo**: Argumentos o conclusiones sin fundamento o ilógicos. *Ejemplo 4: La herramienta IA generativa concluye que un expediente está correctamente tramitado porque contiene toda la documentación habitual (propuesta, informe, resolución, …). Sin embargo, ese razonamiento no es válido: la mera presencia de documentos no demuestra por sí solo que los trámites se hayan realizado en plazo, que el procedimiento sea correcto o que se hayan cumplido todos los requisitos legales.* • **Inconsistencias**: Información contradictoria con resultados anteriores sin justificación. *Ejemplo 5: Ante una primera consulta, la herramienta IA generativa indica que en un expediente de subvenciones no aprecia incidencias relevantes. Más tarde, al pedirle un resumen más detallado del mismo expediente, señala posibles incumplimientos en plazos, publicidad o justificación, sin explicar el cambio de criterio, ni motivarlo. Esa falta de coherencia entre respuestas puede generar confusión y llevar al auditor a basarse en resultados contradictorios comprometiendo la fiabilidad de la evidencia obtenida.*

Tabla 2 - Riesgos en la utilización de la IA Generativa: descripción y ejemplos

A. Riesgo de resultados deficientes Riesgo de que el resultado proporcionado por la herramienta de IA Generativa sea deficiente. Esto puede ser debido a dos causas distintas: un problema en el funcionamiento o un problema en la entrada de datos.	**A.1 Riesgo de fallos en el funcionamiento del sistema** Los problemas en la ejecución del sistema pueden surgir de fallos en la ejecución de los componentes, sean o no componentes de IA Generativa, así como de errores o distorsiones derivados de la combinación de los componentes.	**A.1.2 Riesgo de fallo en el funcionamiento de otros componentes** Los sistemas de IA Generativa (asistentes o agentes de IA) incorporan componentes que no son LLM, lo que añade riesgos en la ejecución del sistema si no funcionan correctamente. Si estos componentes presentan problemas, el sistema puede funcionar con información incompleta o errónea, procesar la información incorrecta o simplemente no funcionar en absoluto, lo que conlleva riesgos significativos de que los resultados no sean aptos para fundamentar los resultados de auditoría.	Estos componentes incluyen: • **Componentes de adquisición de datos**. *Ejemplo: La herramienta IA Generativa se alimenta a partir de la documentación de los expedientes proporcionados por el auditado a través de un sistema de carga automática, pero por un fallo en ese componente no se incorporan todos los anexos o toma una versión antigua de los documentos. Como resultado, la herramienta IA analiza información incompleta o desactualizada y puede ofrecer una conclusión errónea.* • **Componentes de procesamiento basados en reglas**. *Ejemplo: Antes de que intervenga la herramienta IA Generativa, se aplica un sistema que incluye reglas automáticas para clasificar los documentos que se incluyen en un expediente de contratación pública, extraer datos o relacionar información del expediente. Si esas reglas están mal definidas o funcionan incorrectamente, puede, por ejemplo, identificar como fecha de adjudicación una fecha de publicación, o tratar una prórroga como si fuera un contrato nuevo. En ese caso, la herramienta IA Generativa trabaja sobre información "corrupta" o errónea y puede generar conclusiones deficientes.* • **Componentes de aprendizaje automático predictivo**. *Ejemplo: La herramienta IA Generativa incorpora un modelo predictivo para priorizar expedientes con mayor probabilidad de riesgo y centrar en ellos la revisión. Si ese componente falla o está mal ajustado, puede dejar fuera expedientes relevantes y destacar otros que en realidad tienen poca importancia. El auditor puede acabar orientando su trabajo hacia prioridades equivocadas.* • **Componentes de conexión externa** (componentes o API[9] que permiten que el sistema se conecte con otros sistemas). *Ejemplo: La herramienta IA Generativa está conectada con otras aplicaciones para buscar documentos, recuperar datos o ejecutar consultas en sistemas corporativos. Si ese componente no funciona bien, puede no localizar determinada información, traer documentos equivocados o no completar correctamente la acción solicitada. En consecuencia, la respuesta de la herramienta IA Generativa se basará en una información parcial o incorrecta.* • **Infraestructura operativa** *Ejemplo: La herramienta IA Generativa depende de servidores, bases de datos, redes y otros elementos técnicos para funcionar con normalidad. Si se produce una caída del servicio, un problema de capacidad o un error en el entorno, la herramienta IA Generativa puede responder con lentitud, interrumpirse o devolver resultados incompletos. Aunque el diseño del sistema sea correcto, la falta de estabilidad operativa puede afectar directamente a la calidad del trabajo de auditoría.*

[9] *Application Programming Interface*, en español, Interfaz de Programación de Aplicaciones.

Tabla 2 - Riesgos en la utilización de la IA Generativa: descripción y ejemplos

A. Riesgo de resultados deficientes Riesgo de que el resultado proporcionado por la herramienta de IA Generativa sea deficiente. Esto puede ser debido a dos causas distintas: un problema en el funcionamiento o un problema en la entrada de datos.	**A.1 Riesgo de fallos en el funcionamiento del sistema** Los problemas en la ejecución del sistema pueden surgir de fallos en la ejecución de los componentes, sean o no componentes de IA Generativa, así como de errores o distorsiones derivados de la combinación de los componentes.	**A.1.3 Riesgo combinado** Errores que se amplifican o se distorsionan al combinar módulos.	Los sistemas de IA pueden generar resultados deficientes debido a la combinación de componentes entre sí o el uso iterativo del mismo componente. Este riesgo se puede segmentar en tres categorías principales: • **Riesgo de amplificación**. *Ejemplo: Una determinada herramienta IA Generativa clasifica varios expedientes como de bajo riesgo a partir de datos incompletos, y una segunda herramienta IA Generativa utiliza esa clasificación para decidir qué documentación analizar con más detalle. El error inicial no se corrige, sino que se arrastra y se refuerza en la siguiente fase, de modo que expedientes críticos quedan fuera del foco de auditoría.* • **Riesgo de distorsión de la información** (degradación semántica). *Ejemplo 1: En un expediente de gastos de personal, la documentación refleja que las retribuciones abonadas se corresponden en términos generales con la nómina aprobada, pero también pone de manifiesto que determinados complementos específicos fueron reconocidos sin una motivación suficientemente clara y que en algunos casos faltaba soporte completo sobre su justificación. La herramienta de IA Generativa analiza el expediente por fases. Un primer módulo extrae datos de nóminas, resoluciones y acuerdos. Un segundo módulo resume los documentos y destaca los elementos principales. Un tercer módulo genera una valoración global orientada a identificar posibles incidencias. A medida que la información pasa de un módulo a otro, la observación inicial sobre la insuficiente justificación de ciertos complementos va perdiendo relevancia: primero queda formulada como una mera cuestión formal, después se omite en el resumen final y, al final del proceso, el sistema devuelve la idea de que los gastos de personal están correctamente soportados y no presentan incidencias relevantes. En este supuesto, la distorsión no deriva de un dato inventado, sino de que los módulos sucesivos rebajan el peso de una debilidad real del expediente. El resultado final ofrece una visión más favorable de la que realmente se desprende de la documentación, lo que puede llevar al auditor a no profundizar en una incidencia relacionada con la adecuada justificación de las retribuciones abonadas.* *Ejemplo 2: En la fase de justificación de una subvención, el expediente incluye la memoria de actuación, la cuenta justificativa y los documentos de gasto aportados por la entidad beneficiaria. De su revisión resulta que, aunque la mayor parte de los gastos parece estar vinculada a la actividad subvencionada, existen algunos justificantes con deficiencias, dudas sobre la elegibilidad de determinados importes y una parte del gasto admitida con reservas por el órgano gestor. La información pasa por varios módulos del sistema. Uno identifica y extrae automáticamente los datos de los justificantes y de los informes de revisión. Otro resume la documentación y otro genera una conclusión general sobre el grado de justificación de la subvención. En ese recorrido, las reservas iniciales sobre la admisibilidad de ciertos gastos se van simplificando hasta quedar reducidas a incidencias menores, y finalmente el sistema presenta la subvención como correctamente justificada en términos generales, sin reflejar con suficiente claridad las dudas existentes sobre parte del gasto declarado.* • **Riesgo de interfaz**. *Ejemplo: La herramienta IA Generativa identifica importes, fechas y tipos de documento, y transfiere esa información a otro componente (herramienta IA Generativa) para que la analice. Si ambos componentes no están bien coordinados, puede producirse errores en el traspaso de datos, por ejemplo, porque un campo se interpreta de forma distinta o se pierde parte de la información al cambiar de formato. El resultado final sería incorrecto no por fallo de un módulo, sino por la integración defectuosa entre ambos componentes.*

Tabla 2 - Riesgos en la utilización de la IA Generativa: descripción y ejemplos

A. Riesgo de resultados deficientes Riesgo de que el resultado proporcionado por la herramienta de IA Generativa sea deficiente. Esto puede ser debido a dos causas distintas: un problema en el funcionamiento o un problema en la entrada de datos.	**A.2 Riesgo en la entrada de datos al ejecutar el sistema** Las entradas al sistema durante su ejecución ya sean humanas o de cualquier fuente de información, también pueden generar resultados deficientes.	**A.2.1 Riesgo de la intervención humana** Hay tres tipos principales de razones que provocan riesgos específicos.	• **Definición incorrecta o ambigua de las instrucciones a la IA (*prompting*) o incluir imprecisiones de estas (antecedentes incompletos, erróneos o desactualizados).** La IA no se emplea conforme a la finalidad pretendida de la prueba de auditoría y/o el contexto proporcionado no es adecuado. La calidad de estos inputs afectará significativamente la calidad del resultado proporcionado por la IA. *Ejemplo 1: El auditor pide a la herramienta IA Generativa que "revise si el expediente de concesión de subvenciones es correcto", pero no concreta qué aspecto debe analizar, qué criterio normativo debe tomar como referencia ni le facilita toda la documentación relevante. Además, entre los documentos cargados falta una modificación posterior del expediente. Con unas instrucciones imprecisas y una base documental incompleta, la herramienta IA Generativa puede generar un resultado aparentemente útil, pero mal enfocado o directamente erróneo por falta de contexto.* • **Decisiones humanas en puntos de control** *(human-in-the-loop)* En puntos específicos del proceso hay intervención humana para determinar si el sistema se ejecuta, cómo lo hace y adoptar determinadas decisiones. Estas decisiones pueden incluir autorizar ciertas acciones, como el uso de herramientas, elegir si el sistema continúa o se detiene, o determinar qué hará a continuación. Si estas decisiones no son adecuadas, por ejemplo, si el auditor no comprende el alcance o el propósito de la tarea, carece de conocimientos técnicos o muestra sesgo a favor de la automatización, los resultados del sistema pueden ser deficientes. Este riesgo es especialmente relevante en sistemas con agentes. *Ejemplo 2: Durante la ejecución de la IA Generativa, esta propone consultar una fuente adicional o continuar con una verificación más detallada, pero el auditor decide detener el proceso porque considera suficiente el resultado preliminar. Esa decisión humana, tomada sin comprender bien el alcance de la tarea o confiando en exceso en la primera salida del sistema, puede hacer que no se detecten incidencias que habrían aparecido en una revisión posterior.* • **Configuraciones**. El usuario puede seleccionar ajustes que determinan el funcionamiento de ciertos aspectos del sistema, como la selección entre modos de razonamiento, el establecimiento de límites de tiempo o iteraciones, o la activación o desactivación del acceso a herramientas o datos. Si estos ajustes no son adecuados, los resultados pueden ser deficientes. *Ejemplo 3: El auditor utiliza la herramienta IA Generativa con una configuración inadecuada: limita demasiado el número de iteraciones, desactiva el acceso a determinadas fuentes documentales o selecciona un modo de funcionamiento más rápido, pero menos profundo. Aunque la herramienta IA Generativa funcione correctamente, esos ajustes pueden hacer que el análisis sea superficial, incompleto o insuficiente para la finalidad de la prueba de auditoría.*

Tabla 2 - Riesgos en la utilización de la IA Generativa: descripción y ejemplos

A. Riesgo de resultados deficientes Riesgo de que el resultado proporcionado por la herramienta de IA Generativa sea deficiente. Esto puede ser debido a dos causas distintas: un problema en el funcionamiento o un problema en la entrada de datos.	**A.2 Riesgo en la entrada de datos al ejecutar el sistema** Las entradas al sistema durante su ejecución ya sean humanas o de cualquier fuente de información, también pueden generar resultados deficientes.	**A.2.2 Riesgo por la calidad de la información** Si la información utilizada es engañosa, errónea, incompleta o irrelevante, o no está catalogada de forma eficaz para facilitar una recuperación coherente, los resultados del sistema pueden ser deficientes.	Los sistemas de IA tienen la capacidad de acceder a fuentes de información al ejecutarse para proporcionar información actualizada y/o específica de la tarea, más allá de la que se encuentra en los datos de entrenamiento del modelo, con el fin de mejorar la calidad de los resultados. Estas fuentes pueden incluir: • Bases de datos financieras o transaccionales de entidades auditadas si se obtiene la autorización correspondiente (**SU USO CON ACCESO DIRECTO A ESAS BASES DE DATOS NO ESTÁ AUTORIZADO, CON CARÁCTER GENERAL, EN LOS OCEX**). • Documentos de trabajos anteriores. • Documentos técnicos o normativos internos. • Guías elaboradas externamente. • Normas profesionales. Esta categoría de riesgo excluye deliberadamente los riesgos derivados de los datos de entrenamiento (intrínseco al modelo) o la información proporcionada por un ser humano durante la ejecución del sistema (Riesgo A.2.1). Si la base de conocimiento a la que accede la IA es defectuosa, el resultado será defectuoso independientemente del modelo o del *prompt*.

Tabla 2 - Riesgos en la utilización de la IA Generativa: descripción y ejemplos

B. Riesgo de uso inadecuado de los resultados Riesgo de que el resultado de la IA Generativa sea usado mal, a pesar de que en sí mismo sea correcto.	**B.1 Riesgo de interpretación incorrecta de los resultados** Riesgo de que el equipo de auditoría malinterprete el significado, el alcance, las limitaciones o la certeza del resultado de una herramienta de IA, lo que puede llevar a conclusiones o acciones inapropiadas incluso cuando el resultado en sí sea adecuado.	**Riesgo de exceso de confianza (sesgo de automatización) del auditor** sobre el alcance o la fiabilidad de los resultados de la IA. *Ejemplo 1: Confundir una probabilidad con una certeza.* *La herramienta IA Generativa indica que una resolución de concesión de subvenciones tiene "alta probabilidad" de ser irregular. El auditor interpreta ese resultado como si confirmara directamente que existe una irregularidad. Sin embargo, la herramienta IA Generativa solo está ofreciendo una estimación o una alerta, no una conclusión definitiva. Si se toma como certeza, pueden adoptarse decisiones inapropiadas.* *Ejemplo 2: Falsos negativos percibidos como ausencia de riesgos (asumir que, si la herramienta IA Generativa no detecta incidencias, entonces no existen).* *La herramienta IA Generativa analiza un conjunto de expedientes de gastos de personal y no señala anomalías relevantes. El auditor concluye que el área revisada (complementos específicos) no presenta riesgos significativos. Pero ese resultado puede deberse a límites de la propia herramienta, a la calidad de los datos o a que determinados problemas no entran dentro de lo que la herramienta IA Generativa puede detectar. La ausencia de alertas no equivale automáticamente a la ausencia de errores o incumplimientos.* *Ejemplo 3: Extrapolación inadecuada (dar a los resultados de la herramienta IA Generativa más alcance o fiabilidad del que realmente tienen).* *La herramienta IA Generativa ha sido entrenada para revisar determinados patrones en contratos menores, pero el auditor extiende sus resultados a toda la contratación del órgano auditado. Además, asume que los resultados son fiables sin revisar sus límites ni validarlos con otras pruebas. Ese exceso de confianza puede llevar a conclusiones demasiado amplias o mal fundadas.*
	B.2 Riesgo de comprensión incorrecta de la metodología de auditoría Existe el riesgo de que el auditor no comprenda correctamente cómo se integra el uso de la herramienta de IA en la metodología, lo que podría llevar a conclusiones o decisiones inadecuadas.	*Ejemplo 1: El auditor podría no entender que el resultado de una herramienta IA solo puede ser utilizada como un procedimiento de valoración de riesgos o un procedimiento analítico y nunca como una prueba sustantiva que sea capaz por sí sola de fundamentar una conclusión de auditoría.* *Ejemplo 2: Uso de la herramienta IA Generativa como sustituto de procedimientos que en realidad solo debía complementar.* *El auditor emplea una herramienta IA Generativa para revisar contratos y detectar cláusulas anómalas, pero interpreta que ese análisis sustituye por sí mismo las pruebas de cumplimiento o las verificaciones sustantivas previstas en el programa de auditoría. Como resultado, reduce indebidamente otros procedimientos esenciales y llega a una conclusión con evidencia insuficiente.* *Ejemplo 3: Confundir alertas con hallazgos (tomar como prueba lo que solo es una señal de alerta).* *La herramienta IA Generativa detecta un pago extraño y lo marca como "alto riesgo". El auditor interpreta que, por ese solo hecho, ya existe una incidencia acreditada. Pero en realidad la herramienta solo está señalando algo que debe revisarse después. Si no se comprueba con documentación y pruebas adicionales, puede llegarse a una conclusión equivocada.* *Ejemplo 4: Usar la herramienta IA Generativa en una fase del trabajo para la que no estaba prevista.* *El equipo de auditoría emplea una herramienta IA Generativa para apoyar la fase de planificación de la auditoría, por ejemplo, para localizar áreas de mayor riesgo y planificar mejor la revisión. Sin embargo, el auditor la usa al final para apoyar directamente la conclusión del informe. El problema es que una herramienta IA Generativa diseñada para orientar la planificación no sirve, por sí sola, para fundamentar la conclusión final de auditoría.*

Tabla 2 - Riesgos en la utilización de la IA Generativa: descripción y ejemplos

C. Riesgo de incumplimiento de normas de auditoría	**C.1 Riesgo de incumplimiento de normas de auditoría** Riesgo de que el uso de la IA Generativa incumpla las NIA	La IA Generativa permite diseñar nuevos tipos de procedimientos de auditoría y se requiere un juicio profesional considerable para determinar cómo incorporarlas a las fiscalizaciones de manera que cumplan con las normas de auditoría. En algunos casos, reemplazarán o complementarán un procedimiento tradicional. En otros, pueden plantear un enfoque de auditoría significativamente diferente. La metodología de los OCEX, basada en las NIA, orienta a los equipos sobre lo que sería adecuado, pero existe el riesgo de que se utilicen enfoques que den como resultado fiscalizaciones que no cumplan con las normas de auditoría. Este riesgo no es exclusivo de la tecnología, ni de la IA Generativa, surge con la aplicación de un nuevo procedimiento de auditoría. Sin embargo, el riesgo puede incrementarse en el contexto de los procedimientos basados en IA Generativa, ya que puede resultar difícil comparar sus resultados con los de los procedimientos de auditoría tradicionales o ajustarlos a los requisitos de las normas de auditoría. *Ejemplo 1: Sustituir una prueba de auditoría por el resultado de una herramienta IA Generativa sin obtener evidencia suficiente.* *El equipo utiliza una herramienta IA Generativa para revisar expedientes y, al ver que la herramienta no detecta incidencias relevantes, decide no realizar comprobaciones adicionales previstas en la metodología. El problema es que el resultado de la herramienta IA Generativa, por sí solo, puede no constituir evidencia de auditoría suficiente y adecuada. En ese caso, la fiscalización podría incumplir las normas por basarse en pruebas insuficientes.* *Ejemplo 2: Falta de trazabilidad y aplicabilidad del procedimiento (aplicar un procedimiento nuevo con una herramienta IA Generativa sin poder explicar bien cómo se ha hecho ni cómo se ha validado).* *El auditor usa una herramienta de IA Generativa para resumir contratos, identificar riesgos o proponer conclusiones, pero no documenta de forma clara qué instrucciones introdujo (prompts), qué información utilizó, qué revisión posterior realizó, ni por qué consideró fiable el resultado.* *Si un revisor no puede entender ni reproducir el trabajo realizado, puede existir un incumplimiento de las exigencias de documentación y trazabilidad del trabajo de auditoría.* *Ejemplo 3: Usar la herramienta IA Generativa de una forma que reduce indebidamente el juicio profesional del auditor.* *La herramienta de IA Generativa propone áreas de riesgo, selecciona muestras o redacta observaciones, y el equipo las acepta casi automáticamente sin una revisión crítica suficiente. Aunque la herramienta IA Generativa haya sido útil, las normas de auditoría exigen que el auditor mantenga el juicio profesional y no delegue en la herramienta decisiones que le corresponden a él. Si esa revisión crítica no existe, el enfoque aplicado puede no ajustarse a las normas.*

6.2. Medidas de mitigación de riesgos

El riesgo de que el sistema de IA produzca un resultado deficiente puede mitigarse de varias maneras que se indican en la tabla 3.

Tabla 3 - Posibles medidas de mitigación

Medidas de mitigación		Descripción
Diseño y desarrollo del sistema de manera que responda al uso previsto	Diseño del flujo de trabajo	Definición estructurada de tareas, fases y controles del sistema de IA, alineados con el objetivo del procedimiento de auditoría.
	Selección de componentes	Elección de modelos y herramientas coherentes con el nivel de juicio profesional requerido.
	Protocolos y reglas	Establecimiento de reglas deterministas (entradas, salidas, límites y escalado).
	Control de versiones	Gestión formal de cambios y actualizaciones del sistema.
Revisión, validación previa y certificación	Certificación por casos de uso	Evaluación formal del sistema en usos concretos antes de su despliegue operativo.
	Pruebas de resultados	Muestreo estructurado para evaluar la coherencia y fiabilidad de los outputs.
Formación y gobernanza	Formación del personal	Capacitación sobre el funcionamiento, los límites y riesgos de la IA.
	Directrices de uso	Definición de cuándo es apropiado utilizar IA.
	Prevención del sesgo de automatización	Refuerzo del escepticismo profesional frente a resultados automatizados.
Revisión y supervisión humana	Revisión de resultados	Evaluación humana de la adecuación, coherencia y suficiencia del resultado.
	Supervisión en puntos de control	Intervención humana en fases críticas del proceso.
Uso adecuado del resultado obtenido por la IA	Claridad sobre alcance y límites	Explicación expresa de lo que el resultado representa y de sus limitaciones (técnicas y metodológicas).
Metodología de auditoría	Alineación con normas	Verificación del cumplimiento de normas de auditoría y control de calidad.

La mitigación del riesgo no se apoya en una única medida, sino en una combinación proporcionada de las medidas disponibles, ajustada al uso previsto y al riesgo residual aceptable.

Tomando como referencia las clasificaciones de riesgos de la tabla 2 y las medidas de mitigación definidas en la tabla 3 se establece la relación entre ambos, recogida en la tabla 4 siguiente sustentada en los siguientes criterios:

- Los **riesgos técnicos** se mitigan sobre todo con **diseño del sistema, selección de componentes, protocolos, control de versiones y pruebas**.
- Los **riesgos derivados de la intervención humana,** por el equipo de auditoría, se reducen principalmente con **formación, directrices, supervisión y revisión crítica** aplicando el escepticismo profesional.
- Los **riesgos metodológicos y normativos** exigen ante todo **alineación con las normas y guías de auditoría** y una definición clara del papel que puede desempeñar la IA dentro del trabajo auditor.

Cuando se utilicen herramientas de IA Generativa, el auditor será responsable de la validez y fiabilidad de los datos fuente utilizados, del buen uso de la herramienta utilizada y deberá validar y responsabilizarse de los resultados proporcionados por la herramienta. Estas circunstancias deberán documentarse en los papeles de trabajo al concluir sobre la prueba.

Tabla 4 – Riesgos y medidas de mitigación

Riesgos			Medidas de mitigación	Observaciones
A. Riesgo de resultados deficientes	A.1 Riesgo de fallos en el funcionamiento del sistema	A.1.1 Riesgo de fallo en la ejecución del componente de IA Generativa	Establecer controles de revisión humana y listas de verificación específicas para detectar alucinaciones, omisiones e inconsistencias. También se puede establecer revisiones cruzadas entre sistemas de IA Generativa distintos.	En función del momento en el trabajo de auditoría donde se aplique la IA se puede proponer que la información generada sea contrastada por varios revisores de forma independiente.
		A.1.2 Riesgo de fallo en la ejecución de otros componentes	Selección de componentes, protocolos y reglas, control de versiones, certificación por casos de uso.	La fiabilidad del resultado depende también de conectores, validadores, *scripts* y demás componentes no generativos.
		A.1.3 Riesgo combinado	Diseño del flujo de trabajo, protocolos y reglas, pruebas de resultados, supervisión en puntos de control	Deben controlarse especialmente las transferencias entre fases y la agregación de resultados parciales para evitar la degradación de la información
	A.2 Riesgo en la entrada de datos al ejecutar el sistema	A.2.1 Riesgo de la intervención humana	**Definición incorrecta/ambigua:** Establecer un marco estandarizado de diseño y validación de *prompts* y datos de entrada, que incluya plantillas predefinidas, criterios mínimos de calidad y revisión previa obligatoria por personal cualificado.	Se pueden establecer bibliotecas de *prompts* ya validados y documentación revisada por una persona para añadirla como contexto.
			Decisiones humanas en puntos de control: las personas que interaccionen con la herramienta en algún punto del proceso deben tener la correspondiente formación y revisar la salida tras la ejecución de su acción.	Los agentes que necesiten que una persona aporte información o validación en algún punto pueden limitar las opciones a seleccionar a solo unas pocas, y no dejar opción de texto libre.
			Configuraciones: se debe disponer de una parametrización clara, estandarizada y validada de opciones para configurar la IA Generativa	La elección de la parametrización, de cuando usar modo de razonamiento, la versión del LLM o aportar documentación deben estar seleccionadas y validadas previo al uso en un trabajo de auditoría
		A.2.2 Riesgo por la calidad de la información	Directrices de uso, protocolos y reglas, revisión de resultados, claridad sobre alcance y límites	Debe comprobarse la pertinencia, vigencia y fiabilidad de las fuentes utilizadas por el sistema.

Tabla 4 – Riesgos y medidas de mitigación

Riesgos		Medidas de mitigación	Observaciones
B. Riesgo de uso inadecuado de los resultados	**B.1 Riesgo de interpretación incorrecta de los resultados**	Claridad sobre alcance y límites, formación del personal, prevención del sesgo de automatización, y revisión de resultados	Es necesario recordar que la salida de la IA puede informar el juicio profesional del auditor, pero **no sustituirlo** automáticamente.
	B.2 Riesgo de comprensión incorrecta de la metodología de auditoría	Alineación con normas, directrices o guías de uso, supervisión en puntos de control, revisión de resultados	La utilidad del resultado depende de su correcta inserción dentro del programa de trabajo y de la metodología aplicable.
C. Riesgo de incumplimiento de normas de auditoría	**C.1 Riesgo de incumplimiento de normas de auditoría**	Alineación con normas, certificación por casos de uso, directrices o guías de uso, formación del personal	La metodología debe asegurar que el empleo de IA no rebaje las exigencias de la evidencia, revisión y juicio profesional.

6.3. Mención a la IA Agéntica

Como ya se ha señalado, la inteligencia artificial está evolucionando a una velocidad sin precedentes, pasando de asistentes de IA Generativa a sistemas más autónomos y capaces, los agentes IA.

En el ámbito privado, muchas firmas de auditoría disponen de planes para seguir explorando el uso de tecnologías emergentes, incluidos los sistemas de IA Agéntica. Con el tiempo, esto puede evolucionar de la automatización de tareas hacia la orquestación de múltiples actividades de auditoría. Estos desarrollos introducen nuevas complejidades, especialmente a medida que los sistemas se vuelven cada vez **más autónomos**. Es fundamental que los auditores mantengan la responsabilidad de evaluar, validar y, cuando sea necesario, cuestionar las decisiones tomadas por los sistemas de IA, incluyendo[10]:

- Explicación clara de los juicios o decisiones clave que afectan significativamente a la auditoría. Por ejemplo, cómo se aplicó, desafió o ejerció el juicio del auditor en relación con los resultados generados por IA o cualquier punto crítico de decisión humana tomado a lo largo del proceso.
- La implementación de marcos de actuación que definan el papel del auditor dentro del proceso de auditoría.

En el ámbito de la auditoría pública, de acuerdo con la estrategia de la IA del Tribunal de Cuentas Europeo[11], esta evolución significa que el soporte de IA puede extenderse más allá de las herramientas individuales de productividad. Los agentes de IA pueden apoyar múltiples fases de los procesos de auditoría realizando tareas técnicas predefinidas —como comprobaciones de consistencia, preparación y estructuración de documentos, y seguimiento del progreso— todo ello dentro de reglas predefinidas y bajo supervisión humana. A medida que estos agentes IA maduran, tienen el potencial de transformar no solo la forma en que se realizan las auditorías, sino también cómo se garantiza la calidad de la auditoría, cómo se identifican los riesgos y cómo se gestionan los datos a lo largo del ciclo de vida de la auditoría. Según el Tribunal de Cuentas Europeo este cambio requiere un enfoque de gobernanza proactiva, donde la IA no se trate solo como una herramienta de apoyo, sino como una capacidad estratégica integrada en los procesos centrales.

Al tratarse de una tecnología muy reciente no es objeto de desarrollo en esta primera versión de la presente guía.

7. Diseño de procedimientos de auditoría basados en herramientas de IA Generativa

7.1. Consideraciones preliminares sobre la aplicabilidad de la herramienta de IA Generativa en la ejecución de los procedimientos de auditoría

El auditor someterá a un proceso de análisis previo la viabilidad de emplear la herramienta de IA Generativa en la ejecución de los procedimientos de auditoría.

Los procedimientos de auditoría que incluyan pruebas a realizar mediante el empleo de IA, requerirán contar con un análisis previo de riesgos con la finalidad de incrementar la seguridad posterior del ejercicio de corroboración sobre el resultado obtenido, ya que, como señala la GPF-OCEX 5370,[12] apartado 4 *Ejercicio del escepticismo profesional al utilizar HTA*, **los resultados obtenidos con herramientas de IA tendrán la consideración, de acuerdo con las NIA-ES-SP y las GPF-OCEX, de obtenidos mediante procedimientos de valoración de riesgos o de procedimientos analíticos**.

Por lo tanto, **los resultados obtenidos con estas herramientas no constituirán evidencia de auditoría suficiente y adecuada para soportar las conclusiones de auditoría si no se complementan con otra evidencia corroborativa**.

El uso de las herramientas de IA Generativa, sus alcances y sus características debe ser conocido por todos los miembros del equipo de auditoría que participen en el trabajo.

[10] Ver Use of technology in audits: Innovation and audit quality 2026, IFIAR, abril 2026,

[11] Artificial Intelligence Strategy for 2026-2030, European Court of Auditors, 2026.

[12] GPF-OCEX 5370. Guía para la realización de pruebas de datos.

7.2. Elementos clave para el correcto diseño de procedimientos de auditoría basados en herramientas de IA Generativa

En el diseño de procedimientos de auditoría apoyados en herramientas de IA Generativa conviene partir de una precisión conceptual básica:

El empleo de herramientas de IA Generativa no equivale a un proceso automatizado tradicional, ya que el auditor no controla, ni conoce de forma completa, la secuencia interna de reglas, operaciones o protocolos mediante los que la herramienta transforma una instrucción (*prompt*) y un conjunto de datos en un resultado.

Como se explica en la GPF-ICEX 5381 las automatizaciones tradicionales son de carácter determinista, es decir, los mismos inputs producen siempre los mismos outputs; sin embargo, **en la IA Generativa, debido a su carácter probabilístico, los mismos inputs producirán frecuentemente diferentes outputs y de forma poco transparente.**

La utilidad de las tecnologías de IA Generativa es indudable, pero su funcionamiento incorpora un grado de opacidad (efecto "caja negra") que obliga a extremar el rigor en el diseño, documentación y validación de las pruebas.

Esta circunstancia diferencia los procedimientos basados en herramientas de IA Generativa de **los procedimientos puramente automatizados, basados en herramientas como ACL o IDEA, en los que sí se conoce de antemano la lógica del tratamiento, la secuencia de pasos y las condiciones que sigue la aplicación informática**.

A continuación, se recogen los principales **elementos que el equipo de auditoría deberá valorar para diseñar adecuadamente procedimientos de auditoría que incorporen el uso de IA Generativa en su ejecución**.

Deben respetarse los usos permitidos y prohibidos señalados en el apartado 5 y en las normas de uso de cada OCEX.

Tabla 5– Claves para el diseño de procedimientos de auditoría con IA Generativa

Elementos	Descripción
Documentación / Información (*inputs*)	¿Se requiere el suministro de información para la ejecución de la prueba prevista en la que se emplea la herramienta de IA? En algunas pruebas, de carácter exploratorio, el uso de la IA generará resultados inmediatos y no será necesario aportar documentación ni suministrar información a la herramienta. En cambio, cuando sea preciso adjuntar documentación a la herramienta, el equipo auditor deberá valorar la naturaleza de esa documentación para determinar si es posible emplear la IA o si procede descartarla.
Tipología / Naturaleza y sensibilidad de la información	Con carácter previo al diseño del procedimiento de auditoría que utilice IA, el auditor evaluará la naturaleza de la documentación que, en su caso, deba incorporarse a la herramienta. La información/datos podrá ser de uno de estos tipos (tabla 7): OGD, SPI, ANOM, ODS.
Fiabilidad de las fuentes	Según origen y criticidad de los datos: Alta / media / baja
Herramienta de IA y entorno	El programa de trabajo deberá especificar el tipo de IA que se empleará, indicando su nombre, versión, naturaleza y modalidad de uso según tabla 6. Asimismo, deberá señalar si la licencia permite la persistencia de datos para el entrenamiento del modelo o la modificación del código.
Preparación de datos	El programa de trabajo contemplará, con carácter previo, las actuaciones necesarias para asegurar el uso de la información estrictamente necesaria (minimización de datos), limpia y depurada, tratada conforme a la finalidad definida, en formatos previamente determinados y, cuando proceda, anonimizada, con identificación de las fuentes, repositorios y URL correspondientes.
Riesgos de uso	El análisis del riesgo asociado al uso de herramientas de IA en el programa de trabajo se realiza con el fin de identificar sus limitaciones, posibles sesgos, alucinaciones y efectos sobre la calidad de la evidencia auditora. La aplicación de IA no elimina la necesidad de ejercer juicio profesional, ni sustituye los procedimientos de verificación establecidos.

Elementos	Descripción
	Los resultados y recomendaciones generados por la IA deberán interpretarse con cautela, validarse mediante pruebas de re-ejecución o contrastes independientes y considerarse dentro del marco de riesgos inherentes al uso de tecnologías avanzadas. El equipo auditor será responsable de evaluar y mitigar los riesgos detectados, así como de asegurar que la IA utilizada cumple con los requisitos de seguridad, integridad de la información y protección de datos aplicables.
Configuración de instrucciones / *prompts* en la modalidad de IA Generativa	El programa de auditoría definirá la(s) instrucción(es) (en adelante, *prompt*) que sustente(n) cada prueba o conjunto de pruebas. Para que un *prompt* genere correctamente un resultado en una prueba de auditoría y no se limite a generalidades, **deberá incorporar suficiente contexto, criterios, restricciones y formato de salida, de modo que la herramienta pueda proponer pasos ejecutables, la evidencia esperada y la conclusión correspondiente. Se recomienda el uso de técnicas de *few-shot prompting* (facilitar ejemplos) y cadena de pensamiento (instruir a la IA para que muestre su razonamiento paso a paso o desglose el problema) para mejorar la precisión.** Ver anexo 1.
Resultados esperados	Formatos, campos obligatorios, umbrales, criterios de aceptación. El equipo de auditoría realizará un análisis previo del resultado esperado de la prueba para establecer la expectativa del auditor
Contrastes y validaciones	Se diseñará una prueba de contraste para validar los resultados obtenidos de la IA Generativa.
Limitaciones y autorizaciones previas	Con carácter previo a su empleo, las herramientas de IA que puedan acceder, procesar o almacenar datos personales o información de uso oficial, confidencial, restringida o amparada por derechos de propiedad intelectual deberán ser autorizadas por la Comisión de IA o por el órgano competente equivalente existente en el OCEX.
Registro / evidencias	La herramienta de IA deberá emitir un informe de trazabilidad que recoja lo siguiente: a) El procedimiento ejecutado b) Las pruebas definidas en el *prompt*. c) Los outputs generados. d) La huella de tiempo. e) La versión del modelo empleado. f) Las características técnicas básicas que se puedan configurar en la herramienta de IA, como la “temperatura” o parámetro de aleatoriedad (siendo recomendando un valor cercano a cero).

7.3. Las seis cuestiones clave para validar el diseño del *prompt*

En los trabajos de fiscalización que incorporan herramientas de IA Generativa, **el diseño del *prompt* constituye un elemento esencial para asegurar que los resultados obtenidos sean útiles, pertinentes y susceptibles de adecuada revisión**.

A diferencia de otras herramientas informáticas deterministas basadas en secuencias plenamente predeterminadas, la IA requiere instrucciones precisas que delimiten con claridad el encuadre de la tarea, la finalidad de la prueba y los parámetros conforme a los cuales debe elaborarse la respuesta. Por ello, **un *prompt* insuficientemente definido puede comprometer la fiabilidad, la consistencia y la trazabilidad de los resultados, dificultando su utilización como soporte de evidencia de auditoría**.

Desde esta perspectiva, la correcta formulación del *prompt* no constituye una cuestión meramente técnica, sino una garantía metodológica para el equipo de auditoría. Su adecuada configuración permite vincular la actuación de la herramienta con el riesgo y la afirmación objeto de análisis, identificar la población y la fuente de los datos empleados, fijar criterios y umbrales de evaluación y, además, obtener una traza verificable sobre la ejecución de la prueba y sus resultados. De ahí la conveniencia de **contar con un conjunto de cuestiones clave que permita validar, con carácter previo, si el diseño del *prompt* responde a las exigencias propias de un trabajo de auditoría**.

A continuación, se detallan seis cuestiones clave que pueden ayudar al equipo de auditoría a validar el diseño del *prompt*. Si la respuesta es afirmativa ("sí") en todas ellas, puede considerarse que su estructura es adecuada.

Tabla 6 – Claves para la validación del *prompt*

Nº	Claves para la validación del *prompt*	Descripción explicativa
1	**¿Queda establecido el rol en el que tiene que contestar la IA?**	Debe precisarse la función o posición desde la que la herramienta debe elaborar su respuesta, por ejemplo, como auditor, analista de datos o revisor técnico. Esto permite acotar el enfoque de la contestación y orientar el tipo de razonamiento esperado.
2	**¿Queda definido de forma clara la finalidad de la prueba? (qué se va a probar y para qué).**	El *prompt* debe identificar con claridad el objeto de la prueba y su propósito concreto dentro del trabajo de auditoría. No basta con pedir un resultado; es necesario indicar qué se pretende verificar y con qué finalidad se solicita el análisis.
3	**¿Se indica el riesgo y la afirmación afectada?**	La instrucción debe vincular expresamente la prueba con el riesgo que se pretende cubrir y con la afirmación afectada. De este modo, la actuación de la IA queda alineada con la lógica propia del procedimiento de auditoría y con el objetivo probatorio perseguido.
4	**¿Está definida la población y la fuente de donde se toman los datos?**	Debe especificarse de forma expresa cuál es el universo de datos sobre el que va a operar la IA y de qué fuente procede esa información. Esta precisión resulta necesaria para asegurar la pertinencia de la prueba y la trazabilidad de los resultados obtenidos.
5	**¿Se establecen criterios o umbrales y definiciones de excepciones?**	El *prompt* debe incorporar los parámetros conforme a los cuales la IA debe identificar incidencias, desviaciones o resultados relevantes. Ello exige fijar criterios objetivos, umbrales de materialidad y una definición clara de lo que debe considerarse excepción.
6	**¿Se incluye la generación de traza (pasos + evidencia + conclusión) respecto de la prueba, ejecución y resultados obtenidos?**	Se debe prever que la salida de la IA se estructure en un formato que permita su revisión posterior, dejando constancia de los pasos seguidos, la evidencia utilizada y la conclusión alcanzada. Esta traza resulta esencial para la documentación, reproducibilidad y validación del trabajo realizado.

8. La matriz de evaluación del uso de la IA Generativa según la naturaleza de la información empleada

La **matriz de evaluación** (tabla 9) permite responder a una cuestión operativa esencial en auditoría pública: **qué modalidad de IA puede emplearse, y con qué límites, según la naturaleza de la información o documentación utilizada**. Su finalidad es valorar, con carácter previo, si el uso de una herramienta de inteligencia artificial resulta **permitido**, **condicionado** o **no permitido**, atendiendo tanto al tipo de datos que se incorporen, obtengan, traten o generen mediante la herramienta, como al grado de control, seguridad y gobernanza existente sobre dicha IA.

La matriz combina dos variables. Por un lado, la modalidad de IA utilizada según se detalla en la tabla 7. Por otro, la naturaleza de la información empleada según se detalla en la tabla 8. De esta combinación resulta el nivel de habilitación aplicable en cada caso.

Con carácter general, cuanto mayor es el control institucional sobre la herramienta, más amplio puede ser su uso. Por ello, la IA propia del OCEX (sistema Ad-hoc) aparece como compatible con todas las tipologías de información, mientras que los sistemas externos pueden requerir condiciones adicionales cuando se trate de información sensible no anonimizada. En cambio, las herramientas no corporativas quedan limitadas a datos abiertos o directamente no están permitidas, en función de las políticas de uso de cada OCEX.

La matriz traslada así una regla práctica al equipo auditor: antes de utilizar IA no basta con identificar la tarea que se desea realizar, sino que debe determinarse qué información se va a introducir, obtener o generar y bajo qué modalidad de herramienta se va a operar. La anonimización reduce el riesgo, pero no habilita automáticamente cualquier uso; y la información no sensible, si es interna u operativa, puede seguir estando sujeta a restricciones.

Tabla 7 - Tipología de herramientas IA[13]

Sistema Ad-hoc	**Sistemas IA Generativa desarrollados internamente o por terceros bajo especificaciones a medida y desplegados en infraestructura bajo control de la organización e integrado con sistemas internos.** *Ofrecen el máximo nivel de personalización y control.* *Incluyen modelos de código abierto sobre los que se realiza un proceso de fine-tuning adaptado a necesidades específicas.*
Sistema interno	**Sistemas IA Generativa desarrollados por terceros y desplegados en infraestructura bajo control de la organización.** *El modelo se implementa en la infraestructura propia o en una nube privada.* *Generalmente, aunque no limitado, utilizando modelos abiertos como Ollama, Qwen, Google Gemma o Mistral.* *También soluciones bajo licencia comercial que se puedan desplegar íntegramente en infraestructura propia de la organización, incluyendo los modelos.*
Sistema externo	**Sistemas IA Generativa de terceros desplegados en infraestructura fuera del control de la organización, utilizados como SaaS bajo sus términos de uso.** *Están gestionados y mantenidos por proveedores externos y están accesibles a través de plataformas en línea. A su vez, pueden estar integrados en sistemas más amplios.* *Pueden ser utilizados para implementar una fase en el procedimiento (acceso a través de navegadores a servicios como ChatGPT, Perplexity, Mistral, ALIA, etc.), o integrados en el entorno ofimático como Microsoft 365 Copilot.*
No corporativa	**Herramienta accesible libremente o contratada individualmente de forma directa por el auditor.** NOTA: El uso de datos de auditoría en este entorno representa un **riesgo alto de exfiltración de información**. Por este motivo los OCEX podrán prohibir el uso de estas herramientas o permitir su uso con autorización previa en cada caso de uso.

Tabla 8 - Tipología de la información tratada (suministrada a la herramienta IA)

OGD	Open Government Data	Datos de Gobierno Abierto	Datos procedentes de plataformas / bases de datos de gobierno abierto.
SPI	Sensitive Personal Information / Data	Datos sensibles	Datos personales no anonimizados (incluyendo categorías especiales de datos).
ANOM	Anonymized Data	Datos sensibles anonimizados	Datos anonimizados. Datos que, tras un proceso, no pueden vincularse a un individuo ni permiten su reidentificación.
ODS	Operational Data Store	Datos internos de gestión	Datos e información no sensible del auditado, obtenida y/o procesada en el transcurso de la auditoría que se emplea como base para la ejecución de pruebas de auditoría mediante el uso de herramientas IA. NOTA: La información operativa del auditado puede incluir datos confidenciales o información restringida no personal por lo que debe extremarse la precaución al valorar esta tipología.

[13] Basado en el documento Política general para el uso de IA Generativa en procesos administrativos de la Agencia Española de Protección de datos.

La tabla 9 ofrece los tipos de información que se puede utilizar con cada tipo de herramienta IA.

Tabla 9 - Matriz de evaluación

		Tipo de información a utilizar			
		Datos abiertos OGD	Sensibles NO anonimizados SPI	Sensibles SÍ anonimizados ANOM	No sensibles ODS
Tipo de IA	Sistema Ad-hoc / Sistema interno	✅	✅	✅	✅
	Sistema externo	✅	⚠️	✅	✅
	No corporativo	⚠️	⛔	⛔	⛔

✅ Uso permitido ⚠️ Uso condicionado (depende de las políticas de uso de la IA de cada OCEX) ⛔ Uso no permitido

Es importante destacar que la clasificación y uso de IA Generativa recogida en la tabla anterior dependerá de las políticas de uso de IA Generativa de cada OCEX, que podrán tener mayores restricciones en función de las dos variables reflejadas en la misma (tipo de información utilizada y tipo de IA), o de las políticas de uso de la nube, siendo, por tanto, esta matriz susceptible de cambios atendiendo a dichas políticas.

Para facilitar la aplicación práctica de esta matriz, en el Anexo 4 se incluyen ejemplos prácticos del uso de la IA Generativa, clasificando cada uso en función del tipo de información facilitada a la IA.

9. Control y seguimiento del uso de herramientas de IA Generativa en las fiscalizaciones. La gobernanza de la IA

El uso de herramientas de inteligencia artificial, y en particular de IA Generativa, en las fiscalizaciones puede aportar ganancias relevantes de eficiencia, capacidad analítica y apoyo a la toma de decisiones. No obstante, su utilización también introduce **riesgos nuevos o intensificados,** tales como alucinaciones, sesgos o la opacidad de los algoritmos**.** Entre esos riesgos, además de los señalados en la tabla 2 que afectan a la calidad de la auditoría, se incluyen los relacionados con la ciberseguridad, con el rendimiento de la herramienta, la dependencia de terceros, y en general todos los señalados en la GPF-ICEX 5381. Por ello, su empleo debe quedar sometido a un marco específico de control y seguimiento, **integrado en el sistema de gobernanza de la IA del OCEX**.

La incorporación de la IA a los procesos de fiscalización exige, por tanto, la implantación de una serie de mecanismos operativos mínimos de **gobernanza, control y supervisión**.

Desde esta perspectiva, el control del uso de la IA en las actuaciones de fiscalización no debe limitarse a una autorización inicial, sino que ha de configurarse como un **proceso continuo**.

Ello requiere **establecer una estructura de gobernanza** de la IA que estará formada principalmente por los siguientes elementos:

- Un **comité de gobierno de la IA** que articule la gobernanza. Podrá ser uno específico en instituciones grandes o podrá estar integrado en el comité de gestión de las TI y de la seguridad de la información en entidades más pequeñas.
- **Establecimiento de los distintos roles y responsabilidades** en relación con la IA de forma clara.
- **Políticas específicas de uso y gobierno de IA**. Normas de uso de la IA.
- **El inventario de los casos de uso autorizados**.
- Un sistema de **gestión de los riesgos de la IA,** integrado en el sistema de gestión de riesgos de TI y riesgos de **ciberseguridad** del OCEX.

  Se evaluará de forma periódica los riesgos específicos de la IA utilizada en las fiscalizaciones, el diseño y mapeo de los controles aplicables, la implementación de dichos controles, la existencia de documentación

adecuada y la revisión continua de los controles para adaptarlos a la evolución tecnológica, normativa y operativa.

El diseño seguro desde el inicio; el análisis de amenazas y la gestión de riesgos; la protección de infraestructuras, entornos y APIs (interfaces de programación de aplicaciones); la seguridad de la cadena de suministro; la documentación de datos, modelos y *prompts*; la realización de pruebas y evaluaciones previas al despliegue; la comunicación clara a los usuarios y afectados; la aplicación de actualizaciones y parches; la monitorización continua del comportamiento del sistema; y la eliminación segura de datos y modelos al final de su ciclo de vida, son aspectos de la gestión de los riesgos de ciberseguridad relacionados con el uso de la IA.

- El establecimiento de procedimientos de control relativos a la **trazabilidad** de las acciones realizadas por la IA, el registro de *logs*, el control de accesos y entornos, la gestión de incidentes, la documentación sobre la procedencia de los datos, la evaluación de proveedores y la definición de métricas de seguimiento constituyen elementos imprescindibles de ese sistema de control.
- El establecimiento de un **plan de formación y de concienciación del personal** en materia de IA (alfabetización en IA) y sus riesgos.

  La formación continua es esencial para mejorar las competencias de los auditores en la utilización de herramientas tecnológicas y en su capacidad para comprender y evaluar críticamente sus resultados, entender sus limitaciones y aplicar el juicio profesional. Los programas de formación de los OCEX deben de adaptarse a las competencias requeridas.

La implantación y utilización de cualquier herramienta IA requerirá la autorización del comité de IA o del comité de seguridad TI o del órgano de gobierno del OCEX si no existieran los primeros.

Estos órganos, a través del personal expresamente autorizado, podrán realizar en cualquier momento las comprobaciones oportunas para verificar que el uso de las herramientas de IA en el seno de la institución se ajusta a los usos permitidos, a las condiciones establecidas para su empleo y a las exigencias de control, seguridad y seguimiento que resulten aplicables.

En la utilización de las herramientas de IA Generativa se seguirán las normas de seguridad generales establecidas en el OCEX.

En resumen, la utilización de herramientas de IA en el ámbito de la fiscalización solo resultará admisible cuando se integre desde el inicio en una estructura de gobernanza y un marco de control robusto, con garantías suficientes de trazabilidad, fiabilidad, supervisión continua, cumplimiento y responsabilidades claramente definidas.

10. Características de la información que se utilizará como evidencia de auditoría resultante del empleo de la IA Generativa en las pruebas de auditoría

Para evaluar la evidencia de auditoría que haya sido producida por herramientas de IA Generativa utilizadas por el equipo de auditoría es necesario tener en cuenta los criterios establecidos en la **NIA-ES-SP 1500** y en la **GPF-OCEX 1503 La evidencia electrónica de auditoría.** En este apartado se destacan los principales aspectos allí tratados, pero se recomienda su estudio completo.

10.1. Características de la información obtenida mediante pruebas que utilizan IA Generativa a efectos de su consideración como evidencia

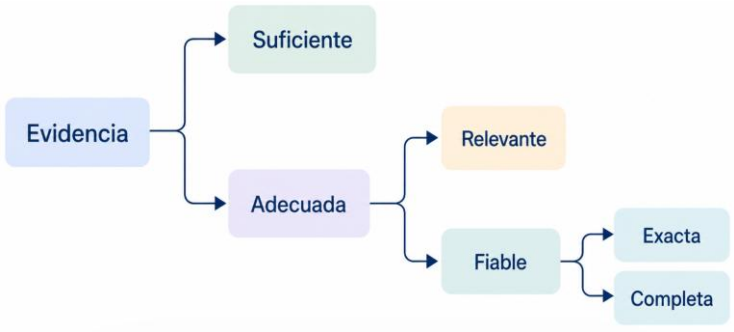
En la auditoría financiera, la información que se utilizará como evidencia de auditoría debe tener unas características fundamentales para que el auditor, tras la ejecución de sus procedimientos de auditoría, la considere evidencia de auditoría y pueda utilizarla como apoyo de sus conclusiones.

De acuerdo con la NIA-ES-SP 1500:

> **El auditor diseñará y aplicará procedimientos de auditoría que sean adecuados, teniendo en cuenta las circunstancias, con el fin de obtener evidencia de auditoría suficiente y adecuada.**

> **Al realizar el diseño y la aplicación de los procedimientos de auditoría, el auditor considerará la relevancia y fiabilidad de la información que se utilizará como evidencia de auditoría.**

Es decir, la evidencia debe ser:



La suficiencia es una medida cuantitativa de la evidencia de auditoría: la cantidad de evidencia de auditoría necesaria depende de la valoración del auditor del riesgo de incorrección material, así como de la calidad de dicha evidencia de auditoría (NIA-ES-SP 1500, p5.e).

La adecuación es una medida cualitativa de la evidencia de auditoría, es decir, de su **relevancia** y **fiabilidad** para respaldar las conclusiones en las que se basa la opinión del auditor (NIA-ES-SP 1500, p5.b).

La **relevancia** se refiere a la conexión lógica con la finalidad del procedimiento de auditoría, o su pertinencia al respecto, y, en su caso, con la afirmación que se somete a comprobación. Tiene que haber una relación directa entre la finalidad de la prueba y la evidencia obtenida. Cuando la información obtenida vaya a ser utilizada como evidencia de auditoría, la confianza exigible en la calidad de la evidencia depende de **cómo vaya a usarse y del grado en que se vaya a descansar en ella,** la **adecuación para el uso previsto**.

Para que el auditor obtenga evidencia de auditoría **fiable**, la NIA-ES-SP 1500 (párrafo 9.a) exige a los auditores que obtengan evidencias sobre la **exactitud** y **completitud**[14] de la información generada por la entidad que vaya a utilizarse como evidencia de auditoría. La fiabilidad de la evidencia está influenciada por su fuente y naturaleza y depende de las circunstancias en que se obtenga. Con la evolución de la tecnología, hay múltiples fuentes de información disponible y algunas son más confiables que otras.

En la IA Generativa, esta fiabilidad depende críticamente, entre otros factores, de la calidad y diversidad de los datos de entrenamiento, por lo que el auditor, siempre que sea posible, debe evaluar la "biografía de datos" (*data biography*) para comprender el origen y las transformaciones de dicha información, es decir, la trazabilidad de los datos.

Una dificultad añadida, importante, al trabajar con información extraída de sistemas IA Generativa se deriva de las características esenciales de esta tecnología que se destacan en el apartado 5.1 de la GPF-ICEX 5381.Es decir, a las características de **opacidad, no determinismo y adaptabilidad**, que, según los casos, pueden llegar a plantear problemas importantes sobre la fiabilidad de la evidencia obtenida, ya que debido al carácter probabilístico de la IA Generativa **no queda garantizada su exactitud y fiabilidad**.

Por lo tanto, **los resultados obtenidos con estas herramientas no representarán evidencia de auditoría suficiente y adecuada para soportar las conclusiones de auditoría si no existe otra evidencia corroborativa**.

[14] *Completeness* en el original en inglés, *integridad* en la NIA-ES-SP 1500.

Complementando lo anterior y utilizando la nueva metodología propuesta por la IAASB[15], los **atributos que el auditor considerará al evaluar la fiabilidad de la información resultante de la IA Generativa que se pretende utilizar como evidencia de auditoría** son los siguientes:

Tabla 10 – Atributos de la fiabilidad de la información utilizada como evidencia de auditoría

Exactitud	La información está libre de errores en su reflejo de las condiciones, eventos, circunstancias, acciones u omisiones subyacentes, incluyendo el reflejo del período de tiempo o punto en el tiempo apropiado atribuible a las condiciones o eventos.
Completitud	La información refleja todas las condiciones subyacentes, eventos, circunstancias, acciones o inacciones.
Autenticidad	La información ha sido generada o proporcionada por una fuente autorizada para hacerlo, y la información no ha sido alterada de manera que la haga poco fiable para la auditoría.
Sesgo	La información está libre de sesgos intencionados y no intencionados en su reflejo de las condiciones subyacentes, los eventos, las circunstancias, las acciones o las omisiones.
Credibilidad	La fuente de la información tiene la competencia y la capacidad para generar la información con el nivel de calidad requerido, y se puede confiar en ella.

Al diseñar y realizar los procedimientos de auditoría el auditor probará la eficacia operativa de aquellos controles que abordan la fiabilidad de la información producida por la IA Generativa o realizará otros procedimientos para evaluar la fiabilidad de dicha información.

Además, cuando se utilice una herramienta de IA Generativa para obtener evidencia de auditoría se exigirá también: la adecuación al uso previsto, consistencia interna, coherencia con el conocimiento que el auditor tiene de la entidad, comprensión clara de su alcance y limitaciones, apoyo en inputs y fuentes de calidad, fuerza persuasiva no sobredimensionada y, en la medida de lo posible, explicabilidad o trazabilidad suficiente para su revisión por un tercero. En la revisión humana de los outputs o resultados obtenidos de la IA (*resultados de la ejecución de los prompts*), deberá evaluarse si se dan esas características.

En relación con la **explicabilidad o trazabilidad intelectual de los resultados**, el uso de herramientas de IA generativa debe orientarse a que los *outputs* obtenidos sean comprensibles y susceptibles de revisión, evitando que se conviertan en una "caja negra". Esta exigencia no implica conocer el funcionamiento interno de la herramienta, sino que el resultado permita entender de manera razonable cómo se ha alcanzado la conclusión, a partir de qué información, con qué criterios y mediante qué pasos. De este modo, el auditor puede valorar la coherencia del *output* con el objetivo de la prueba realizada, identificar en su caso posibles errores o sesgos y justificar, una vez corroborado, su uso como evidencia de auditoría. Para ello, resulta fundamental un adecuado diseño del *prompt*, incorporando determinados elementos mínimos —como los recogidos en la tabla 6— que faciliten la generación de resultados claros, documentados y alineados con el procedimiento de auditoría.

La información no debe analizarse de forma aislada, sino atendiendo a su significado, alcance, limitaciones y grado de certeza. El riesgo es que el equipo auditor interprete una salida como si fuera más amplia, más concluyente o segura de lo que realmente es. Por eso, una característica esencial de la información generada por IA es que su alcance y limitaciones sean comprendidos correctamente, evitando tratarla como una conclusión cerrada cuando solo debe informar el juicio profesional.

La calidad del resultado obtenido dependerá de la calidad de las entradas y de las fuentes consultadas durante el tiempo de ejecución del *prompt*, de forma que si la información de apoyo es **engañosa, errónea, incompleta o irrelevante**, la salida puede ser deficiente. Del mismo modo, si la especificación de la tarea o el *prompt* es incorrecta, ambigua o se apoya en contexto desactualizado, el resultado puede dejar de ser apto para su finalidad. En consecuencia, la información generada por IA que aspire a utilizarse como evidencia debe descansar sobre ***inputs* pertinentes, completos y correctamente definidos**.

La información resultante con el empleo de la IA puede ser útil como evidencia, pero su valor probatorio debe apreciarse con prudencia y sin atribuirle más capacidad demostrativa de la que realmente tiene (fuerza persuasiva no sobredimensionada).

[15] IAASB Main Agenda (March 2026), Agenda Item 7, Audit Evidence & Risk Response (AE&RR).

10.2. Los resultados obtenidos con el empleo de la IA y su naturaleza de evidencia indicativa

En el transcurso de cada prueba en la que se emplee la IA, deberá registrarse la huella temporal, la versión de IA utilizada y a ser posible, el parámetro de aleatoriedad o *temperatura* seleccionado de modo que quede garantizada la trazabilidad e integridad de la evidencia de auditoría.

Los resultados obtenidos con el empleo de la IA solo podrán utilizarse como evidencia indicativa, no concluyente, **que deberá complementarse y corroborarse necesariamente con otra evidencia adicional.**

Su aceptación dependerá siempre de la existencia de **evidencia corroborativa externa**, pudiendo consistir en la verificación del expediente con registros previos, la comprobación de la fecha en el documento firmado, el recálculo aritmético o la confirmación del órgano gestor entre otras.

10.3. Procedimientos de corroboración y calidad de la información en auditorías con IA

En los trabajos de auditoría en los que se utilicen herramientas de inteligencia artificial, **la solidez de los procedimientos de corroboración** no depende solo de aplicar pruebas adicionales de contraste. Depende también, y en gran medida, de la calidad de la información que se emplea para comprobar los resultados generados por la propia IA. No toda la información sirve, por sí sola, para confirmar o descartar un hallazgo. Para que realmente pueda apoyar las conclusiones del auditor esa información debe ser relevante, fiable (exacta y completa), trazable y susceptible de revisión crítica.

Por ello, la información corroborativa debe proceder, siempre que sea posible, de fuentes independientes o, al menos, de evidencias cuyo origen y fiabilidad puedan verificarse de forma objetiva. Además, ha de ser lo bastante precisa y completa como para permitir confirmar o refutar el resultado ofrecido por la herramienta. No se trata solo de disponer de información, sino de contar con una base suficientemente sólida para comprobar, con un grado razonable de seguridad, si el hecho, cálculo, clasificación o anomalía detectados por la IA responden realmente a la realidad examinada. **Utilizar otra herramienta de IA Generativa para corroborar el resultado de una primera, no permite obtener información corroborativa sólida y concluyente ni es aconsejable a fin de evitar el riesgo de sesgos compartidos o error circular.**

La evidencia de auditoría que se va a utilizar para contrastar o corroborar los resultados obtenidos por la herramienta de IA Generativa debe reunir las características señadas en la tabla 10 anterior.

En definitiva, el diseño de las pruebas de corroboración debe partir de una idea sencilla: la calidad de la conclusión dependerá, en buena medida, de la calidad de la información con la que se contraste el resultado de la IA. De ahí que el equipo auditor deba apoyarse en información que puede combinar cuando sea necesario documentación, confirmaciones externas, comprobaciones analíticas, inspecciones directas y validaciones técnicas del propio sistema. Solo sobre esa base podrá obtenerse una evidencia de auditoría verdaderamente útil con la finalidad de sostener conclusiones razonables y revisables, sin renunciar al ejercicio reforzado del escepticismo profesional que exige el uso de estas tecnologías.

Corresponde al auditor diseñar y aplicar los procedimientos de auditoría adicionales que permitan verificar la exactitud, completitud, autenticidad, falta de sesgos y coherencia de los resultados obtenidos, a fin de sustentar razonablemente sus conclusiones.

Tales procedimientos deben permitir no solo confirmar la consistencia material de dichos resultados, sino también reforzar su trazabilidad, reproducibilidad y adecuada documentación, de forma compatible con las exigencias metodológicas propias de la auditoría. En el anexo 3 se incluyen una serie de ejemplos o propuestas de procedimientos de corroboración del resultado de la IA.

10.4. El riesgo de exceso de confianza en la tecnología o de sesgo de automatización y el escepticismo profesional

En las fiscalizaciones en las que se empleen herramientas de IA Generativa, el **escepticismo profesional[16] debe intensificarse**, porque el riesgo no se limita a que la herramienta produzca errores, omisiones, distorsiones o razonamientos defectuosos, sino también a que el equipo auditor acepte sus resultados con una confianza excesiva por su aparente coherencia, fluidez o verosimilitud.

[16] Financial Reporting Council, Generative and Agentic AI Guidance, March 2026.

Como se ha señalado al inicio, en la NIA-ES 220 (Revisada) se advierte que el uso de herramientas tecnológicas puede generar **riesgos de sesgos inconscientes, como el sesgo de automatización**, que pueden llevar a confiar en exceso en la relevancia y fiabilidad de los resultados obtenidos por dichas herramientas por el mero hecho de haber sido generado por la tecnología, en lugar de valorar su alcance, sus limitaciones y su adecuación al objetivo del trabajo.

Reconocer la existencia de sesgos inconscientes ayuda al auditor a diseñar y ejecutar procedimientos de auditoría de manera objetiva. La susceptibilidad al sesgo de automatización puede aumentar cuando los procedimientos realizados con herramientas tecnológicas son complejos, como ocurre si hay múltiples entradas o relaciones entre ellas, o si existe menor transparencia sobre cómo se generan los resultados como sucede con la IA Generativa.

Entre las posibles medidas que el auditor puede considerar para **reducir el riesgo de sesgo de automatización** al usar herramientas tecnológicas como la IA Generativa se encuentran:

- Aplicar diligentemente el escepticismo y el juicio profesional para analizar si la evidencia obtenida de las herramientas de IA Generativa al diseñar y ejecutar procedimientos de auditoría es **pertinente y fiable**.
- Informar de forma clara al equipo de auditoría sobre las situaciones o casos en los que la susceptibilidad al sesgo por automatización pueda ser mayor.
- Proporcionar formación adecuada a los miembros del equipo de auditoría que emplean herramientas tecnológicas, incluyendo capacitación sobre el funcionamiento y las limitaciones de dichas herramientas y sobre la detección de posibles sesgos en los modelos.
- Resaltar la importancia de contar con la participación de miembros del equipo de auditoría con mayor experiencia, o con habilidades y conocimientos especializados, según corresponda, para:
    - Comprender las entradas de datos y los pasos de procesamiento, incluidos cálculos y modificaciones de datos, utilizados en las herramientas tecnológicas, identificando posibles incoherencias técnicas o conceptuales.
    - Interpretar los resultados obtenidos al aplicar las herramientas tecnológicas.

Cuanto más complejo sea el entorno informático mayor será el grado de escepticismo profesional requerido para evaluar la evidencia electrónica.

El escepticismo profesional reforzado implica no aceptar la salida de la IA por su sola apariencia técnica, sino someterla a contraste crítico, comprender sus limitaciones, corroborarla cuando proceda con la documentación fuente y mantener siempre el juicio del auditor como elemento decisivo en la formación de las conclusiones.

Se seguirán estos criterios en la evaluación por el auditor de la relevancia y fiabilidad de la información generada por la herramienta de IA Generativa destinada a ser utilizada como evidencia de auditoría:

- Cuanto más dependan los resultados de interpretaciones o juicios, más exhaustiva deberá ser la evaluación del auditor.
- Si resulta más complicado vincular los resultados a los datos de entrada, o replicar o verificar dichos resultados, la evaluación del auditor también deberá ser más amplia.
- A mayor grado de incoherencia con las expectativas del auditor o con otras evidencias de auditoría disponibles, más amplia deberá ser la evaluación del auditor y se requerirá un análisis adicional de las causas de dicha desviación.

En particular, el escepticismo abarcará, al menos las siguientes cuestiones:

- Evaluación de posibles sesgos del modelo.
- Detección de “alucinaciones” o respuestas no fundamentadas.
- Trazabilidad de las fuentes empleadas.
- Calidad y pertinencia de los datos de entrada.
- Capacidad de reproducir los resultados generados.
- Verificación de los controles de integridad aplicados a los ficheros y documentos procesados por la herramienta de IA.

La revisión humana eficaz exige que quienes revisan los resultados obtenidos de la IA Generativa, tengan la **competencia adecuada** y evalúen si estos son apropiados para el uso previsto, completos, exactos, internamente consistentes y coherentes con su conocimiento de la entidad, aplicando expresamente escepticismo profesional y siendo conscientes del riesgo de sesgo de automatización.

11. Bibliografía

- **Artificial Intelligence Strategy for 2026-2030,** European Court of Auditors, Publications Office of the European Union, 2026.
- **Cyber AI Profile (oficialmente NIST IR 8596)** es una guía del Instituto Nacional de Estándares y Tecnología (NIST) diseñada para ayudar a las organizaciones a gestionar los riesgos de ciberseguridad relacionados con la Inteligencia Artificial. 16 de diciembre de 2025.
- **Directrices de uso de la Inteligencia Artificial en el Senado**. 16 de febrero de 2026.
- **Generative and Agentic AI Guidance.** Financial Reporting Council. Marzo de 2026.
- **Guía PR29 – Normas para el uso de los servicios de Inteligencia Artificial de la Sindicatura de Comptes de la Comunitat Valenciana**. Abril 2024.
- **Herramienta para el uso responsable de IA en las Administraciones Públicas.** Laboratorio de Privacidad de la Agencia Española de Protección de Datos (AEPD). Enero 2026.
- **ISO/IEC 42001:2023 – Sistemas de gestión de la Inteligencia Artificial.** International Organization for Standardization (ISO) y International Electrotechnical Commission (IEC). Organismos internacionales de normalización responsables de elaborar estándares técnicos.
- **ISO/IEC 23894:2023 – Gestión del riesgo de la Inteligencia Artificial.** International Organization for Standardization (ISO) y International Electrotechnical Commission (IEC). Organismos internacionales de normalización responsables de elaborar estándares técnicos.
- **Manual de Procedimientos de Fiscalización del Tribunal de Cuentas (MPFR TCu)**.
- **Proposed Roadmap for Non-Authoritative Material on Gen AI-Enabled Technological Tools Used in Engagements**, Technology Quality Management Workstream, International Auditing and Assurance Standards Board (IAASB), IAASB Main Agenda June 2026.
- **Reglamento (Unión Europea) 2024/1689 del Parlamento Europeo y del Consejo**, de 13 de junio de 2024 (Reglamento de Inteligencia Artificial o RIA).
- **Requisitos para Auditorías de Tratamientos que incluyan IA**. Agencia Española de Protección de Datos (AEPD). Enero 2021.
- **Securing Artificial Intelligence (SAI). Baseline Cyber Security Requirements for AI Models and Systems.** Instituto Europeo de Normas de Telecomunicaciones. Diciembre 2025.
- **Use of technology in audits: Innovation and audit quality 2026**, International Forum of Independent Audit Regulators (IFIAR). Abril 2026

Anexo 1. Recomendaciones para el diseño de los *prompt* en la ejecución de pruebas de auditoría con IA Generativa

Tabla 11 - Recomendaciones para el diseño de un *prompt* en pruebas de auditoría con IA Generativa

Requisito	Detalle
Asignación del rol	▪ Qué perfil o rol debe asumir la IA (auditor público, auditor informático, analista de datos, especialista en contratación pública, especialista en control interno, revisor jurídico …). "Experto en …." ▪ Marco mental o perspectiva de análisis ("Adopta una actitud de escepticismo profesional reforzado", "actúa con un enfoque estrictamente orientado al cumplimiento legal"…)
Objetivo exacto de la prueba	▪ Qué quieres lograr: control (diseño/implantación/eficacia operativa) o sustantiva (detalle/analítica). ▪ Qué afirmación aborda: existencia, completitud, exactitud/valoración, corte, derechos/obligaciones, presentación. ▪ Qué riesgo estás abordando (riesgo de incorrección material o de control).
Alcance y delimitación	▪ Entidad / proceso (contratos, subvenciones…) / ciclo (ingresos, compras, nóminas, tesorería…). ▪ Periodo (p. ej., ejercicio 2025; cierre a 31/12/2025). ▪ Unidades/ubicaciones incluidas o excluidas. ▪ Nivel de materialidad / tolerancia (si aplica).
Criterio / marco normativo o política	▪ Marco normativo. ▪ Planes. ▪ Políticas internas, manuales de procedimientos internos, guías técnicas…
Población, fuentes y datos disponibles	▪ Qué población se va a probar (p. ej., "todas las facturas emitidas", "altas de usuarios", "pagos > 10.000€"). ▪ Campos disponibles (fecha, importe, cliente, usuario, aprobador, etc.). ▪ Url, campos en las páginas webs…
Método esperado / técnica que se espera emplear	El prompt debe indicar la técnica de auditoría que se espera aplicar; observación, inspección documental, confirmaciones, procedimientos analíticos, de acuerdo con el objetivo de la prueba, el tipo de evidencia disponible y el riesgo identificado. ▪ Especificar si la prueba se realizará mediante muestreo o sobre el 100 % de la población, indicando el tamaño, método de selección, estratificación, umbrales o criterios de excepción aplicables.
Umbrales, criterios de selección y definición de excepción	▪ Umbral de selección (por importe, riesgo, aleatorio, dirigido). ▪ Qué cuenta como hallazgo o desviación. ▪ Qué hacer si aparecen excepciones (ampliar muestra, procedimientos alternativos, cuantificar error y evaluar el impacto en las conclusiones).
Entregable y formato de salida	Indicar expresamente el formato de la respuesta esperada, de modo que pueda incorporarse directamente a los papeles de trabajo, por ejemplo: ▪ Procedimiento paso a paso (numerado), para explicar la lógica del resultado. ▪ Evidencia a obtener. ▪ Criterio de evaluación para determinar si el resultado es satisfactorio. ▪ Documentación en papel de trabajo (capturas y referencias). ▪ Conclusión tipo y cómo redactarla. ▪ Riesgos/limitaciones y procedimientos alternativos.
Restricciones y supuestos	▪ Confidencialidad (usar datos ficticios). ▪ No inventar cifras, no asumir controles no mencionados. ▪ Idioma, nivel de detalle, longitud máxima.

Requisito	Detalle
Incorpora en el *prompt* la generación de un informe de validación	▪ Incluye en el *prompt* la generación de un Informe que confeccione la propia IA en la que se incluya la relación de instrucciones/*prompts*, los *outputs*/resultados obtenidos por cada una de ellas, la huella de tiempo (correspondiente al día y hora en el que se efectúa la prueba, la versión del sistema empleado IA, el parámetro/nivel de aleatoriedad o *temperatura*, la versión del sistema empleado IA y resto de evidencias anexas. ▪ Salida requerida: pasos numerados + evidencia esperada por paso + criterios de aceptación + tratamiento de excepciones + procedimientos alternativos + texto de conclusión para papel de trabajo.
Restricciones a la IA	▪ No inventes. ▪ No generes datos no fundamentados en los *prompts*. ▪ Si la información no está presente en los datos suministrados, responde expresamente “Dato no disponible”. ▪ No sustituyas ni incorpores datos que no se suministren para la obtención del resultado. ▪ Limita tus conclusiones a los datos aportados sin realizar inferencias fuera del alcance establecido.

Anexo 2. Ejemplos de informes de validación de *prompt*

El informe de validación del *prompt* deberá contar al menos con un título identificativo, ámbito/área de auditoría, programa, prueba, huella de tiempo (fecha-hora de exportación), modelo, versión, archivo generado y listado de descripción de *prompts* recopilados.

El informe de validación requerido tendrá que ser solicitado a la propia herramienta IA empleada identificando los campos requeridos por parte del equipo auditor

Tabla 12 - Propuesta de elementos que debe incluir el Informe de Validación de Pruebas obtenidas con el empleo de IA (*Log* de Auditoría)

Elemento del informe de validación	Contenido a consignar	Contenido / Descripción
Título identificativo	Denominación única del informe o *log* de auditoría generado para la prueba realizada mediante IA.	[Ej.: Validación_Análisis_Contratos_Q1_2026]
Ámbito / Área de auditoría	Identificación del ámbito funcional o material en el que se encuadra la prueba.	[Ej.: Auditoría Interna / Cumplimiento Normativo / Gestión de Riesgos]
Programa	Referencia al programa de trabajo o actuación de auditoría en el que se integra la prueba.	[Ej.: Programa de Validación de Modelos Generativos - Proyecto C-2026]
Prueba	Descripción concreta de la verificación o contraste efectuado mediante la herramienta de IA.	[Ej.: Verificación de la no existencia de sesgos en la selección de proveedores]
Huella de tiempo	Fecha y hora exactas de exportación o generación del informe, a efectos de trazabilidad.	[Insertar marca de tiempo: YYYY-MM-DD HH:MM:SS UTC]
Modelo y versión	Identificación precisa de la herramienta de IA utilizada, incluyendo versión concreta del modelo.	[Ej.: GPT-4o-2024-05-13 / Claude 3.5 Sonnet / Llama-3-70b]
Archivo generado	Denominación del fichero o ficheros resultantes de la ejecución de la prueba.	[Ej.: Informe_Validacion_Final.pdf / Resultados.csv]
Listado descriptivo de *prompts*	Relación ordenada de las instrucciones suministradas a la IA durante la prueba, distinguiendo, en su caso, entre *prompt* inicial, refinamientos sucesivos y *prompt* final de validación o contraste.	*Prompt* 1 (Inicio): [Descripción detallada de la instrucción inicial de contexto]

Propuesta de Instrucciones para solicitar el informe de validación generado por la IA

Para obtener este tipo de informe automáticamente, se proponen orientaciones como la siguiente en cuanto a la configuración del *prompt* al finalizar la tarea de ejecución de la prueba de auditoría:

"Basado en esta sesión, genera un informe de validación formal en formato estructurado que incluya: Título, Área, Programa, Prueba, fecha-hora exacta, modelo y versión, archivo generado y una lista con las descripciones de los prompts utilizados para asegurar la trazabilidad."

Ejemplos de informes de validación de *prompts* generados por la IA

Huella de tiempo (UTC): **2026-03-12T15:31:37+00:00**

Prompts recopilados

1 **1. Acceder a Entidades Ejecutoras (página principal)**
¿Puedes acceder a esta url: https://www.juntadeandalucia.es/organismos/consejerias.html?

2 **2. Extraer listado de Entidades Ejecutoras y exportar a Excel**
¿Puedes generar una tabla con las diferentes Entidades Ejecutoras y pasarlas a un excel?

3 **3. Verificación PRTR dentro de "Información económica y presupuestaria" por Entidad Ejecutora (sí/no)**
Accede a la url principal de Entidades Ejecutoras. Entra en cada URL secundaria de cada Entidad Ejecutora. Localiza en la parte inferior "INFORMACIÓN ECONÓMICA Y PRESUPUESTARIA" y accede. Dentro, entra en "Plan de Recuperación, Transformación y Resiliencia - PRTR". En la parte inferior verifica si aparecen campos/enlaces (o similares): componentes/inversiones, buscador de licitaciones, buscador de ayudas y convocatorias, y "Infórmate de la ejecución del MRR". Genera un Excel que marque por Entidad Ejecutora si esos elementos están presentes (sí/no).

4 **4. Extraer fechas de "Ejecución del MRR" por Entidad Ejecutora**
Dentro de PRTR, entrar en el enlace "INFÓRMATE DE LA EJECUCIÓN DEL MRR". En la página "Ejecución del MRR", extraer por Entidad Ejecutora la fecha asociada a: "Resumen de Ejecución del Mecanismo", "Ejecución Acumulada" y "Boletín del Mecanismo de Recuperación y Resiliencia". Volcarlo en una tabla/Excel.

5 **5. Rehacer el proceso uno a uno, sin inventar, y con informe de incidencias**
Repetir el proceso secuencial (principal → Entidad Ejecutora → información económica y presupuestaria → PRTR → "Infórmate de la ejecución del MRR") para cada Entidad Ejecutora. Generar una tabla con 3 columnas (Resumen de ejecución del mecanismo, Ejecución Acumulada, Boletín del Mecanismo de Recuperación y Resiliencia) incluyendo la fecha asociada. No inventar. Si algo falla, generar un informe de incidencias para verificación.

Notas de integridad

La huella de tiempo se muestra en formato ISO 8601 (UTC). El hash SHA-256 permite verificar que el contenido del listado no ha sido alterado: si se recalcula sobre el mismo texto canónico de los prompts, debe coincidir.

Texto canónico usado para el hash: concatenación de cada prompt como `ID|TITULO|TEXTO` (una línea por prompt), en UTF-8.

Los Informes de validación que presenten carencias y/o errores deberán ser descartados y vueltos a solicitar.

Registro acumulativo - PRTR (Hoja 'Resultado Prueba Verificacion')

Registro acumulativo (prompts + resultados) - PRTR

Ambito: Hoja 'Resultado Prueba Verificacion' del archivo Excel aportado por el usuario.

Fecha/hora de exportacion (Europe/Madrid): **09/02/2026 11:40:23**

Modelo/Version: **GPT-5.2 Thinking**

Archivo de trabajo: **001. PRTR_consejerias.xlsx**

Nota sobre tiempos por operacion: En esta conversacion no se registraron/cronometraron los tiempos de respuesta de las operaciones 0 a 9 en el momento de producirse; por tanto, se consignan como N/D para evitar inventar datos. En adelante, se pueden incluir tiempos medidos de forma consistente.

Resumen de operaciones

Operacion	Contenido	Marca de fecha/hora	Tiempo empleado
0	Aprobacion del marco: trabajar solo con la tabla; no inventar; resultados	09/02/2026 (no registrado en el momento)	N/D
1	Aceptacion de registro acumulativo con fecha/hora y tiempo de respuesta	09/02/2026 (no registrado en el momento)	N/D
2	Recepcion del Excel y descripcion de estructura y universo (13 consejerías)	09/02/2026 (no registrado en el momento)	N/D
3	Analisis 1-7 (consejerias, URLs, fechas, componentes).	09/02/2026 (no registrado en el momento)	N/D
4	Analisis 8-9 (futuras convocatorias).	09/02/2026 (no registrado en el momento)	N/D
5	Analisis 10-12 (licitaciones y expedientes).	09/02/2026 (no registrado en el momento)	N/D
6	Analisis 13-14 (ayudas/subvenciones).	09/02/2026 (no registrado en el momento)	N/D
7	Analisis 15-16 (convenios).	09/02/2026 (no registrado en el momento)	N/D
8	Analisis 17 (ejecucion MRR: Q-R-S-T + notas 3-5).	09/02/2026 (no registrado en el momento)	N/D
9	Reejecucion 17 por R15/S15/T15 (sin cambios en el archivo aportado).	09/02/2026 (no registrado en el momento)	N/D
10	Consolidacion del registro en chat.	09/02/2026 11:35:27 - 11:36:20	53 s (estimado)

Anexo 3. Propuestas de procedimientos de corroboración del resultado de la IA Generativa

A continuación, la guía propone procedimientos de corroboración que pueden emplearse para contrastar los resultados ofrecidos por la inteligencia artificial y determinar, con base suficiente, su idoneidad como soporte de las conclusiones de auditoría.

Los resultados obtenidos mediante herramientas de IA deben ser objeto de corroboración mediante procedimientos de auditoría y actuaciones complementarias de validación técnica, de forma que el hallazgo no descanse exclusivamente en la salida generada por el sistema, sino en evidencia adicional suficiente, adecuada, trazable y susceptible de revisión crítica por el auditor. El uso de otra herramienta de IA Generativa para corroborar el resultado de una primera no permite obtener información corroborativa sólida y concluyente ni es aconsejable a fin de evitar el riesgo de sesgos compartidos o error circular.

La siguiente tabla ofrece de manera separada una propuesta de **procedimientos para la obtención de evidencia de auditoría** y **procedimientos específicos de validación del componente de IA**. Los procedimientos identificados descansan en los contenidos de las NIA-ES-SP 1500, 1505, 1520 y 1530, así como por el contenido de la guía de la AEPD (Agencia Española de Protección de Datos) sobre auditorías de tratamientos que incluyan IA.

Tabla 13 Procedimientos de corroboración de la evidencia obtenida por herramientas IA

Procedimientos de corroboración	Finalidad corroborativa frente al resultado de la IA	Aplicación práctica en la prueba de auditoría	Soporte principal
Recálculo	Verificar la exactitud aritmética de cálculos, importes, ratios o recálculos realizadas por la herramienta de IA.	Recalcular manualmente o con herramientas convencionales una muestra de resultados generados por la IA.	La NIA-ES-SP 1500 define el recálculo como comprobación de la exactitud de los cálculos matemáticos incluidos en documentos o registros.
Reejecución independiente	Contrastar si el hallazgo de la IA se reproduce cuando el auditor ejecuta de nuevo el procedimiento por una vía independiente.	Repetir el proceso con reglas tradicionales, scripts de auditoría, consultas sobre la base original o pruebas manuales sobre una muestra.	La NIA-ES-SP 1500 contempla la reejecución como ejecución independiente por el auditor de procedimientos o controles previamente realizados.
Muestreo de auditoría sobre resultados de IA *(NIA-ES-SP 1530)*	Obtener una base razonable para concluir sobre el conjunto de resultados cuando no se revisa el 100% de los resultados.	Seleccionar una muestra de la población relevante de resultados generados por la IA, especialmente allí donde el riesgo valorado sea mayor.	La NIA-ES-SP 1530 regula el muestreo como aplicación de procedimientos a menos del 100% de la población relevante y exige evaluar si la muestra permite alcanzar conclusiones razonables.
Conciliación de integridad entre fuente y población procesada	Corroborar que la IA ha trabajado con el universo correcto, sin omisiones, duplicidades o transformaciones no controladas.	Cotejar número de registros, importes totales, diccionario de datos, transformaciones previas y formato de entrada frente a la base fuente o registros contables.	La NIA-ES-SP 1500 exige valorar la relevancia y fiabilidad de la información utilizada como evidencia; la AEPD exige documentación del origen de los datos, diccionario de datos, preprocesamiento y adecuación del formato de entrada.
Confirmación externa	Obtener evidencia independiente que valide o refute el resultado generado por la IA.	Circularizar saldos, condiciones contractuales, transacciones o hechos detectados por la IA a clientes, proveedores, bancos u otros terceros.	La NIA-ES-SP 1505 tiene por finalidad facilitar confirmaciones externas para obtener evidencia fiable y relevante; además, destaca que la evidencia procedente de terceros puede ser más fiable que la generada internamente.
Inspección documental o física	Confirmar la existencia, validez o soporte material/jurídico del hallazgo identificado por la IA.	Revisar facturas, contratos, expedientes, actas, documentos electrónicos o, cuando proceda, activos o evidencias físicas relacionadas con el resultado.	La NIA-ES-SP 1500 define la inspección como examen de registros, documentos o activos y la reconoce como procedimiento apto para obtener evidencia.

Procedimientos de corroboración	Finalidad corroborativa frente al resultado de la IA	Aplicación práctica en la prueba de auditoría	Soporte principal
Procedimientos analíticos de contraste	Valorar si el resultado de la IA es coherente con relaciones plausibles entre datos financieros y no financieros.	Comparar con series históricas, ratios esperadas, tendencias, presupuestos, variables operativas y conocimiento del negocio.	La NIA-ES-SP 1520 regula los procedimientos analíticos sustantivos y exige investigar variaciones o relaciones incongruentes con otra información o con los valores esperados.
Indagación corroborativa y revisión humana experta	Aportar contexto y someter el resultado de la IA a juicio profesional, sin aceptar la salida de la herramienta como conclusión automática.	Contrastar hallazgos con personal conocedor de la operación, responsables del proceso y revisores cualificados; documentar discrepancias y decisiones.	La NIA-ES-SP 1500 admite la indagación como procedimiento, aunque no basta por sí sola; la AEPD exige personal cualificado, mecanismos de supervisión y procedimientos de intervención humana ante resultados discrepantes.
Validación técnica del componente IA	Corroborar que el resultado no deriva de fallos de diseño, implementación o funcionamiento del modelo.	Revisar plan de pruebas, métricas, criterios de validación, pruebas de caja blanca y caja negra, casos límite y proceso de depuración de errores.	La AEPD exige documentación del proceso de verificación y validación, métricas y criterios, estrategia y plan de pruebas, pruebas de caja blanca y negra, comprobación de valores límite y depuración documentada.
Verificación de datos de entrada y control del sesgo	Corroborar que el resultado de la IA no está distorsionado por datos incompletos, mal formateados o sesgados.	Revisar calidad, completitud, homogeneidad, preprocesamiento, imputaciones, representatividad y medidas para identificar o limitar sesgos.	La AEPD exige controles sobre preprocesamiento, formato de datos, documentación de imputaciones y procedimientos para identificar, eliminar o al menos limitar sesgos, con supervisión humana sobre los resultados.
Trazabilidad, logs y control de versiones	Hacer reproducible el resultado y permitir su revisión posterior, incluida la investigación de contradicciones o anomalías.	Conservar versión del modelo, *datasets* usados, código, librerías, *prompts*/instrucciones, fecha de ejecución, logs, incidencias y correcciones aplicadas.	La AEPD exige trazabilidad, control de versiones, *logs*, registros de resultados, incidencias y disponibilidad de la monitorización para operadores humanos; además, la NIA-ES-SP 1500 subraya la importancia de la fiabilidad de la información empleada como evidencia.

Anexo 4. Ejemplos de casos de uso en los que se puede aprovechar las capacidades de la IA Generativa

Tabla 14 Ejemplos

<table>
<tr><th>Uso descrito</th><th>Tipo de uso de IA / automatización</th><th>Clasificación de la información</th><th>Resultado matriz de evaluación (tabla 9)</th></tr>
<tr><td>Tratamiento de informes no publicados del ente auditado</td><td>Extracción y tratamiento</td><td>SPI: Datos sensibles</td><td rowspan="2">No permitido con sistemas de IA no corporativa o sistemas externos si no está contemplado en la política de IA del OCEX</td></tr>
<tr><td>Tratamiento de expedientes de modificación presupuestaria</td><td>Extracción, tratamiento y elaboración de cuadros</td><td>SPI: Datos sensibles</td></tr>
<tr><td>RPT, anexo de personal, efectivos reales, nóminas y conciliación</td><td>Contraste de resultados, conciliación y análisis de datos de personal</td><td>SPI si incluye nóminas, efectivos reales o datos personales no anonimizados.
Podría ser ANOM si se trabaja con datos anonimizados</td><td>Si son SPI no permitido con sistemas de IA no corporativa o sistemas externos si no está contemplado en la política de IA del OCEX
En caso de datos anonimizados no permitido en sistemas no corporativos</td></tr>
<tr><td>Elaboración de cuadros asociados a la cuenta general sobre estados financieros de agencias y consorcios</td><td>Extracción, tratamiento y elaboración de estados financieros y cuadros</td><td>ODS/OGD según origen.
Si son estados entregados en fiscalización: ODS.
Si son estados ya publicados: OGD</td><td>Si es ODS, no permitido con sistema de IA no corporativo.
Si es OGD, permitido en todos los sistemas, siempre que lo contemple las políticas de uso del OCEX para sistemas no corporativos</td></tr>
<tr><td>PRTR. Planes estratégicos de contratación depositados en los perfiles de contratante y/o de subvenciones publicados en portal de transparencia</td><td>Obtención, análisis y contraste de información publicada</td><td>OGD: datos abiertos o información pública accesible en perfiles de contratante</td><td rowspan="2">Permitido en todos los sistemas, siempre que lo contemple las políticas de uso del OCEX para sistemas no corporativos</td></tr>
<tr><td>Descarga estructurada de estados financieros contables publicados, asociados a la cuenta general, en una url pública</td><td>Descarga, estructuración y consolidación de información publicada</td><td>OGD: estados financieros publicados por ejercicio</td></tr>
<tr><td>Generación de scripts o programas informáticos mediante IA para tratar, posteriormente, con otra herramienta que no es IA Generativa, información financiera, por ejemplo:
Extracción de estados contables de Entidades incluidas en la Cuenta General.
Descarga de información pública en plataforma de contratación</td><td>Generar con IA un script o programa informático de Python, por ejemplo, para que ejecute tarea de descarga, extracción y análisis de documentación o datos económicos</td><td colspan="2" rowspan="2">La información utilizada deberá ser coherente con la matriz de la Tabla 9</td></tr>
<tr><td>Generación de scripts o programas informáticos mediante IA para realizar pruebas de datos, por ejemplo:
Revisión de los permisos de usuarios y objetos de autorización en SAP.</td><td>Generar con IA un script o programa informático de Python para pruebas de controles en SAP que son muy complejas de realizar con ACL.</td></tr>
</table>