

Anexo C

Objetivos y fuentes de riesgo de IA

Del valor esperado a los escenarios y los impactos

Guía práctica con once dimensiones de objetivos, seis familias de fuentes de riesgo y tres casos integradores de Tech Server IIT Spain.

José Manuel Hernández Agudo

Versión 0.2 · 13 de octubre de 2026

Lectura y práctica orientativas: 4-6 horas.

Conecta los módulos 5, 6 y 10 y las guías A y B.

Puedes leerlo en pantalla, imprimirlo y completar las actividades en un visor que admita formularios. Las orientaciones están al final para que respondas primero.

Anexo informativo. Explicaciones y ejercicios originales del curso. Casos y resultados de práctica ficticios.

Índice y recorrido

Método y alcance	4
Qué aporta esta guía	4
Cómo seguir la guía	5
Del valor esperado a una decisión justificable	6
Fuente, escenario, riesgo e impacto	7
Usar la experiencia de seguridad sin limitar la evaluación	8
Dimensiones para formular objetivos	10
C.2.1 · Responsabilidad y rendición de cuentas	11
C.2.2 · Competencia en IA	12
C.2.3 · Disponibilidad y calidad de datos de entrenamiento y prueba	13
C.2.4 · Impacto ambiental	14
C.2.5 · Equidad	15
C.2.6 · Mantenibilidad	16
C.2.7 · Privacidad	17
C.2.8 · Robustez	18
C.2.9 · Seguridad frente al daño	19
C.2.10 · Seguridad de la información y del sistema	20
C.2.11 · Transparencia y explicabilidad	21
Fuentes y escenarios	22
C.3.1 · Complejidad del entorno	23
C.3.2 · Falta de transparencia y explicabilidad	24
C.3.3 · Nivel de automatización	25

C.3.4 · Fuentes ligadas al aprendizaje automático	26
C.3.5 · Problemas del hardware del sistema	27
C.3.6 · Problemas a lo largo del ciclo de vida	28
Tres casos integradores	29
Caso 1. El chatbot: responder antes y conservar atención adecuada	29
Caso 2. El clasificador: el promedio y las personas afectadas	30
Caso 3. El copiloto: ahorro total y condiciones del uso	32
Actividad y Excel	33
Hojas de trabajo	33
Orientaciones	40
Glosario	52
Revisión del expediente	54
Clases de audio y precisiones docentes	55
Conexiones y referencias	59

Método y alcance

Qué aporta esta guía

La IA suele llegar con una promesa: atender antes, clasificar mejor, ahorrar trabajo o crear un servicio nuevo. El Anexo C ayuda a examinar qué condiciones necesita esa promesa y de dónde pueden venir los problemas. Lo utilizaremos como cantera de preguntas para concretar objetivos y construir escenarios de riesgo e impacto.

El Anexo C de ISO/IEC 42001:2023 es informativo. Ofrece posibilidades que cada organización debe considerar según su contexto. Esta guía presenta esas categorías con explicaciones y ejemplos propios para servicios TI. No reproduce el texto de la norma ni convierte todos sus ejemplos en requisitos universales.

La automatización, la personalización, la eficiencia y la innovación son metas de negocio que emplearemos en los ejercicios. Deben conectarse con dimensiones como la privacidad, la equidad, la robustez o la responsabilidad. El beneficio buscado y las condiciones que lo hacen aceptable se documentan juntos.

Al terminar tendrás propuestas de objetivos, escenarios descritos y preguntas de impacto para el chatbot, el clasificador y el copiloto. Podrás trasladarlas a los documentos que ya utiliza el curso. Una idea seleccionada sigue necesitando valoración, responsables y comprobaciones.

Cómo seguir la guía

Empieza por el método y el ejemplo cotidiano. Recorre las once dimensiones de objetivos y las seis familias de fuentes de riesgo. Responde a las preguntas antes de consultar las orientaciones. Después trabaja los tres casos integradores y completa el registro Excel.

Reserva una primera sesión para relacionar valor y objetivos. Dedica otra a formular escenarios y reconocer personas afectadas. En una tercera sesión completa el caso y revisa la trazabilidad con los módulos 5, 6 y 10. El tiempo orientativo de lectura y práctica es de 4-6 horas. Puedes consultar una familia concreta cuando reaparezca en una actividad.

Los audios son clases complementarias. El documento de estudio contiene el recorrido necesario para trabajar sin escucharlos. Las cuatro infografías sintetizan relaciones que después debes explicar con palabras y ejemplos.

Tech Server IIT Spain y los resultados de práctica son ficticios. Cuando copies una propuesta al expediente de tu organización, identifica los datos reales, los criterios aprobados y las verificaciones aún pendientes. Conserva la diferencia entre una hipótesis, un resultado sintético y una observación de producción.

Del valor esperado a una decisión justificable

Imagina un servicio de soporte que quiere responder antes. Contar respuestas emitidas por minuto resulta sencillo. Sin embargo, una respuesta rápida puede ser incorrecta, revelar un dato o impedir que una persona llegue al canal humano. El objetivo necesita expresar el beneficio y las condiciones del servicio que no deben deteriorarse.

Primero concreta la necesidad y el alcance: quién usa el servicio, para qué tarea, en qué idioma y con qué datos. Después identifica qué resultado sería útil. Formula una propuesta medible cuando sea practicable, define cómo observarla y señala qué condiciones acompañan su aceptación.

Un objetivo necesita un responsable, acciones, medios, una fecha o periodo de revisión y un método para evaluar el resultado. Si faltan una línea de base o datos suficientes, planifica obtenerlos. No sustituyas la ausencia de información por un ahorro ficticio ni por una tasa del cien por cien.

Al preparar un objetivo para el SGIA, comprueba su coherencia con la política y los requisitos aplicables. Define una medida cuando sea practicable, seguimiento, comunicación y actualización. Conserva la información documentada. Por ejemplo, la meta, las condiciones y el criterio de revisión deben llegar tanto a quien opera la herramienta como a quien evalúa sus resultados. El banco de ideas ayuda a formular esa propuesta; su aprobación sigue el proceso de la organización.

Por ejemplo: evaluar durante un piloto acotado si el chatbot reduce el tiempo de resolución de consultas elegibles, incluyendo la revisión y las derivaciones humanas, sin publicar respuestas fuera de su alcance. La meta numérica queda pendiente de línea de base y aprobación. Es una propuesta de trabajo, no una autorización de producción.

Una misma oportunidad puede generar varios objetivos. La eficiencia necesita una medida de tiempo total. La robustez necesita pruebas frente a variaciones de entrada. La equidad necesita examinar resultados por grupos relevantes sin crear una nueva recogida innecesaria de datos personales. La privacidad necesita conocer los flujos y condiciones de uso de datos.

Fuente, escenario, riesgo e impacto

Una fuente de riesgo es una condición de la que puede surgir un problema: documentos desactualizados, una tarea demasiado automatizada o un contexto diferente al de prueba. Un escenario cuenta qué podría suceder en ese contexto. La consecuencia explica a quién afectaría y de qué modo. La evaluación de riesgos valora el escenario con los criterios del SGIA.

Escribe el escenario con esta secuencia: en un uso y alcance concretos, una condición puede favorecer un evento que afecte a unas personas o al servicio y produzca unas consecuencias. Añade lo que ya sabes, la evidencia disponible y las preguntas que faltan. Una etiqueta como «sesgo» no proporciona por sí sola esa información.

Ejemplo: si cambian las formas de describir las incidencias y el clasificador mantiene sus reglas de operación sin una revisión adecuada, podría asignar prioridad normal a un ticket crítico. La atención se retrasaría. Deben examinarse los clientes afectados, la urgencia y las consecuencias posibles, junto con la detección y la intervención disponibles.

La evaluación de impacto amplía la mirada hacia individuos, grupos y sociedad. Considera uso previsto y usos indebidos previsibles, contexto y efectos, incluyendo quién soporta el coste de un error. Sus resultados alimentan la evaluación de riesgos. La columna de impacto del Excel prepara preguntas; no sustituye el informe IMP-001.

La evaluación de riesgos de 6.1.2 utiliza el proceso y los criterios definidos por la organización. La evaluación de impacto de 6.1.4 aporta consecuencias para las personas y colectivos. Los objetivos de 6.2 expresan qué resultados se buscan y cómo se planifica alcanzarlos. Son procesos relacionados; cada uno mantiene su finalidad y sus evidencias.

Usar la experiencia de seguridad sin limitar la evaluación

La gestión de accesos, los contratos, los cambios, la continuidad y la respuesta a incidentes proporcionan una base útil desde un SGSI. Algunos problemas de IA tienen antecedentes en TI tradicional. La dependencia de un proveedor y la caída de una API, por ejemplo, también existen en otros servicios digitales.

La IA añade mecanismos y condiciones que requieren pruebas propias: comportamiento aprendido, variación de entradas y poblaciones, respuestas plausibles sin fundamento, diferencias entre subgrupos o dificultad para explicar una recomendación. Un servicio puede estar disponible y protegido frente a accesos indebidos y aun así recomendar una prioridad inadecuada.

La contaminación deliberada de datos de entrenamiento y la inyección de instrucciones durante el uso afectan a momentos y activos diferentes. El primer escenario exige examinar datos y proceso de aprendizaje. El segundo puede entrar por mensajes, documentos recuperados o conectores y requiere limitar qué instrucciones y acciones se consideran confiables.

La clasificación o la predicción tampoco garantizan un comportamiento determinista ni correcto. Debemos estudiar la técnica concreta, las condiciones de ejecución y el efecto de sus errores. El análisis se adapta al papel de desarrollador, integrador o comprador y a aquello que puede controlarse, exigirse o verificarse.

Del valor al objetivo

FUNDAMENTOS DEL SGIA · ANEXO C

Del valor al objetivo

La mejora buscada necesita alcance, condiciones y una forma de comprobar el resultado.

01 · Necesidad concreta

Atender antes, reducir retrabajo o crear una tarea útil. Define para quién y en qué alcance.

02 · Resultado y condiciones

Combina beneficio con privacidad, calidad, revisión y consecuencias relevantes.

03 · Medición y línea de base

Población, periodo, fuente y método. Incluye el coste de comprobar y corregir.

04 · Plan para lograrlo

Responsable, acciones, medios y revisión. La meta ausente sigue pendiente.

05 · Decisión y seguimiento

Valora resultados, límites y cambios. Traslada indicadores al MON-001.

Ejemplo: tiempo total de resolución

Generar más respuestas no acredita resolver mejor. Una condición crítica no se compensa con un promedio.

José Manuel Hernández Agudo

Objetivos y fuentes de riesgo · ISO/IEC 42001:2023

Dimensiones para formular objetivos

Selecciona las dimensiones pertinentes para el uso. Las fichas siguientes ofrecen preguntas y ejemplos originales, no una lista de objetivos obligatorios para todas las organizaciones.

Once dimensiones para preguntar

FUNDAMENTOS DEL SGIA · ANEXO C

Once dimensiones para preguntar

Elige las pertinentes para el uso. Las metas de negocio se relacionan con estas condiciones.

Responsabilidad
Quién responde y puede intervenir

Competencia
Quién sabe revisar y contrastar

Datos
Calidad y cobertura de prueba

Medio ambiente
Actividad, recursos y alcance

Equidad
Diferencias y personas afectadas

Mantenibilidad
Cambios, versiones y alternativa

Privacidad
Finalidad, flujo y conservación

Robustez
Variación y límites de uso

Seguridad frente al daño
Consecuencias y recuperación

Seguridad del sistema
Datos, permisos e instrucciones

Transparencia
Información y vías de revisión

Anexo C informativo. Ejemplos didácticos del curso.

José Manuel Hernández Agudo
Objetivos y fuentes de riesgo · ISO/IEC 42001:2023

C.2.1 · Responsabilidad y rendición de cuentas

Qué pregunta abre

Una recomendación automática puede repartir la actuación entre varias personas. Conviene que la organización pueda reconstruir quién definió el uso, quién supervisó la salida y quién podía limitarla. Asignar un propietario no resuelve por sí solo la disponibilidad de sustitutos o la autoridad para detener una tarea.

Valor y condiciones

Una mesa de ayuda puede automatizar borradores para liberar tiempo. Para sostener esa mejora necesita identificar quién conserva la responsabilidad del servicio cuando el borrador falla y qué información recibe para actuar.

En Tech Server IIT Spain

En el chatbot, una propuesta es relacionar propietario del uso, supervisor, canal de derivación y sustituto en SUP-001. El registro de decisiones y el acceso efectivo a una persona deben comprobarse antes de dar por implantada la supervisión.

Cómo comprobarlo

Comprueba en ejercicios si el equipo identifica al responsable y logra ejecutar la intervención. Registra población de ejercicios, condiciones, resultados y pendientes. Evita contar únicamente nombres asignados en un organigrama.

Un ejemplo cotidiano

En una clínica de bicicletas, un programa recomienda cambiar una pieza. El mecánico sigue necesitando decidir si corresponde a esa bicicleta y responder al cliente. Si el encargado no está, otra persona debe conocer qué puede aceptar y dónde buscar la información. En soporte TI, identifica del mismo modo una decisión concreta y reconstruye qué persona podía modificarla, con qué acceso y en qué horario.

Ahora prueba tú

¿Qué falta si el chatbot tiene propietario, pero nadie sabe quién lo sustituye por la tarde?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 40](#)

C.2.2 · Competencia en IA

Qué pregunta abre

La competencia combina conocimiento del servicio, de los datos y de las limitaciones de la técnica. Las personas que aprueban cambios, revisan resultados o atienden reclamaciones necesitan comprender la tarea que realizan y reconocer cuándo solicitar ayuda.

Valor y condiciones

Un copiloto puede agilizar documentación técnica. El ahorro depende de que quien revisa detecte una referencia inventada, compruebe el código y conozca qué información puede introducir.

En Tech Server IIT Spain

La propuesta para el copiloto es formar al equipo con tareas habituales y comprobar su capacidad para contrastar una salida. La modalidad contratada y las condiciones del proveedor siguen pendientes hasta que PROV-001 disponga de evidencia.

Cómo comprobarlo

Observa la resolución de casos: identificar límites, rechazar una salida y justificar la comprobación. Separa asistencia a formación y capacidad demostrada. Registra las necesidades detectadas y un siguiente paso.

Un ejemplo cotidiano

Un traductor puede trabajar más deprisa con un borrador automático, pero necesita reconocer una negación mal traducida. La velocidad de lectura no acredita esa capacidad. Para soporte, prepara una salida que contenga un error relevante y observa si el revisor lo detecta y justifica la corrección con una fuente. El ejercicio debe tener criterio previo y no introducir datos reales sensibles.

Ahora prueba tú

¿Por qué contar cursos realizados no demuestra una revisión humana eficaz?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 40](#)

C.2.3 · Disponibilidad y calidad de datos de entrenamiento y prueba

Qué pregunta abre

Un modelo aprendido necesita datos adecuados para su propósito. El conjunto de prueba debe permitir observar comportamientos relevantes. La etiqueta, la representatividad, la procedencia y la separación respecto del entrenamiento condicionan la confianza en los resultados.

Valor y condiciones

Mejorar el enrutamiento de incidencias requiere ejemplos de categorías y situaciones críticas. Una muestra formada solo por tickets fáciles puede producir un porcentaje atractivo sin demostrar utilidad en la operación.

En Tech Server IIT Spain

Para SYS-003 / USO-004, documenta la procedencia y el alcance de la prueba, la definición de ticket crítico y la referencia humana. Los 100 tickets sintéticos del ejercicio sirven para practicar cálculos; no acreditan el comportamiento real.

Cómo comprobarlo

Registra cobertura de situaciones, defectos de etiquetas, separación de datos y límites de la muestra. Si se compra un modelo, distingue los datos bajo control propio de las evidencias que deben pedirse al proveedor.

Un ejemplo cotidiano

Una escuela prepara un examen a partir de los mismos problemas y soluciones que el alumnado ya memorizó. El resultado aporta menos información sobre una situación nueva. En un modelo, revisa si la prueba comparte ejemplos, etiquetas o transformaciones que contaminan la comparación. Si no puedes verificar el entrenamiento de un modelo comprado, declara esa limitación y diseña la comprobación de la tarea que sí puedes observar.

Ahora prueba tú

¿Qué pregunta harías antes de aceptar un 94 % de acierto?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 41](#)

C.2.4 · Impacto ambiental

Qué pregunta abre

El uso de IA puede consumir recursos y también ayudar a reducirlos en una actividad. Para examinar ese equilibrio es necesario definir el alcance: consultas, infraestructura, repetición de tareas y cambios inducidos por el nuevo servicio.

Valor y condiciones

La eficiencia operativa puede incluir reducir retrabajo y llamadas innecesarias. Generar más borradores o usar modelos más grandes para tareas simples puede aumentar consumo sin aportar una mejora útil.

En Tech Server IIT Spain

En el copiloto, una propuesta de exploración es comparar modalidades para tareas equivalentes, registrar número de solicitudes y retrabajo y solicitar información pertinente al proveedor. No se atribuyen emisiones concretas sin datos y método.

Cómo comprobarlo

Utiliza medidas disponibles con límites explícitos. Consultas y volumen procesado son aproximaciones de actividad, no una conversión automática a energía o emisiones. Conserva las fuentes, el ámbito y las estimaciones.

Un ejemplo cotidiano

Una oficina sustituye impresiones por resúmenes digitales, pero genera varias versiones que nadie utiliza. La actividad evitada y la actividad añadida deben entrar en el alcance del análisis. En IA, empieza por identificar tareas, repetición y retrabajo. Si solo dispones de número de consultas, úsalo como actividad observada y explica qué información ambiental falta para sostener una conclusión.

Ahora prueba tú

¿Puede un menor número de consultas probar por sí solo un menor impacto ambiental?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 41](#)

C.2.5 · Equidad

Qué pregunta abre

Un buen promedio puede ocultar diferencias perjudiciales para personas o grupos. Examina si la calidad del servicio, el acceso a corrección y las consecuencias de errores se distribuyen de manera injusta en el contexto del uso.

Valor y condiciones

Personalizar la atención resulta útil cuando mejora la adecuación del servicio. Debe evitar que el idioma, la forma de escribir o un canal de acceso impidan una atención comparable para necesidades equivalentes.

En Tech Server IIT Spain

Para el clasificador, plantea revisar incidencias con formas de expresión o idiomas relevantes para el servicio. Define qué diferencias resultan pertinentes y qué datos pueden utilizarse legítimamente. No inventes grupos sensibles ni recojas más información personal por defecto.

Cómo comprobarlo

Compara resultados por segmentos justificados con tamaños de muestra y condiciones. Una diferencia observada necesita investigación. La ausencia de datos de un grupo no equivale a demostrar igualdad de resultados.

Un ejemplo cotidiano

Un servicio de citas funciona bien por teléfono, pero quienes usan otro canal reciben menos información para corregir una solicitud. El promedio puede ocultar esa diferencia. En el caso TI, define qué situaciones equivalentes compararías, qué consecuencia importa y cómo acceder a revisión. Una desigualdad observada necesita investigación y contexto; una muestra insuficiente necesita quedar visible como límite.

Ahora prueba tú

Si un grupo tiene cinco ejemplos sin errores, ¿puedes afirmar que el sistema es equitativo?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 42](#)

C.2.6 · Mantenibilidad

Qué pregunta abre

Mantener una solución implica poder corregirla, adaptarla y conocer qué versión se utiliza. En servicios de terceros, la organización necesita saber cómo enterarse de cambios y qué parte puede probar o limitar por sí misma.

Valor y condiciones

Innovar sin interrumpir el servicio exige ciclos de cambio que permitan aprender y revertir. La velocidad de actualización debe acompañarse de criterios, pruebas y capacidad de alternativa.

En Tech Server IIT Spain

En CAM-001 del clasificador, separa versión evaluada, versión aprobada y versión operativa. La v2 del ejercicio se probó y no se aprobó para desplegarla. Ese estado no cambia por presentar un porcentaje global superior.

Cómo comprobarlo

Mide la capacidad de identificar versiones, reproducir una prueba y ejecutar una alternativa. El tiempo de corrección puede ayudar si su inicio, final y alcance están definidos. Un procedimiento escrito sigue necesitando ejecución.

Un ejemplo cotidiano

Una aplicación de facturación cambia el formato de una exportación. El equipo prepara una prueba y conserva una alternativa antes de adoptarlo. Con un componente de IA también necesitas contrastar el comportamiento de la tarea. Conserva configuraciones y versiones disponibles y verifica si la alternativa cumple su alcance. Una reversión escrita en el procedimiento no demuestra que pueda ejecutarse a tiempo.

Ahora prueba tú

¿Qué problema aparece cuando una API cambia sin que sepamos qué versión produjo cada salida?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 42](#)

C.2.7 · Privacidad

Qué pregunta abre

La privacidad requiere examinar información de personas, finalidad, flujo, accesos, conservación y condiciones del servicio. Que una herramienta no utilice datos para entrenamiento no resuelve todos esos aspectos.

Valor y condiciones

Automatizar resúmenes de soporte puede reducir trabajo. Para ello debe evitar introducir información innecesaria y comprobar qué queda en mensajes, adjuntos, registros y servicios conectados.

En Tech Server IIT Spain

En FD-001 del chatbot y PROV-001 del motor, documenta datos enviados, ámbito contractual, retención y evidencias. Si no conocemos la modalidad o ubicación, mantenemos la pregunta abierta y aplicamos los límites acordados.

Cómo comprobarlo

Revisa muestras autorizadas de flujos, permisos y minimización. Distingue una declaración comercial, una cláusula aplicable a la cuenta concreta y una configuración comprobada. El análisis legal se realiza en su procedimiento correspondiente.

Un ejemplo cotidiano

Para pedir presupuesto de una reparación no hace falta enviar la agenda completa de clientes. En un prompt, un adjunto puede contener más información de la que exige la tarea. Revisa también comentarios, hojas ocultas y metadatos cuando correspondan. La minimización y el canal autorizado reducen exposición, mientras el contrato y la configuración aclaran condiciones que no puedes inferir de la marca comercial.

Ahora prueba tú

El proveedor promete no entrenar con datos del cliente. ¿Qué queda por comprobar?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 43](#)

C.2.8 · Robustez

Qué pregunta abre

Un sistema útil debe soportar variaciones relevantes de su entrada y contexto dentro de su alcance. Conviene saber cómo responde ante ambigüedad, información incompleta o condiciones diferentes a las de la evaluación.

Valor y condiciones

Clasificar más tickets resulta beneficioso si la calidad permanece aceptable cuando cambian la redacción, los productos o la carga. Las pruebas deben cubrir esas variaciones y la alternativa disponible.

En Tech Server IIT Spain

Para SYS-003, propone pruebas con faltas de ortografía, casos ambiguos y cambios de vocabulario. Conserva el criterio humano y examina la infrapriorización de críticos. El resultado favorable de una muestra no elimina la necesidad de seguimiento.

Cómo comprobarlo

Define escenarios de variación y observa errores por categoría y criticidad. Registra cuándo el sistema debe abstenerse o derivar. Una señal de cambio en entradas no demuestra por sí sola una degradación de rendimiento.

Un ejemplo cotidiano

Una persona reconoce una dirección aunque falte una tilde; un servicio automático puede interpretar de otro modo una expresión. Prepara variaciones relevantes de la misma necesidad, con una referencia consistente. Examina si cambian las decisiones y quién puede corregirlas. Si la tarea no está cubierta, la respuesta adecuada puede ser abstenerse o derivar, siempre que esa alternativa esté disponible y comprobada.

Ahora prueba tú

¿Qué distingue detectar drift de confirmar que el servicio ha empeorado?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 43](#)

C.2.9 · Seguridad frente al daño

Qué pregunta abre

En esta dimensión examinamos la posibilidad de daño en el uso y sus consecuencias. Conviene distinguirla de la protección frente a ataques, aunque un ataque también pueda acabar produciendo daño.

Valor y condiciones

Un soporte más rápido debe preservar la atención a situaciones cuyo retraso puede tener consecuencias graves. La urgencia, la reversibilidad y las personas expuestas importan además de la tasa de acierto.

En Tech Server IIT Spain

En el clasificador, pregunta qué ocurre si una incidencia crítica llega tarde al equipo adecuado. La consecuencia depende del servicio y del cliente. No se afirma un daño real ni una ausencia de daño sin investigación.

Cómo comprobarlo

Relaciona tipos de error con consecuencias y posibilidad de recuperación. Define señales de intervención y comprueba su funcionamiento. Los objetivos pueden incluir condiciones críticas justificadas, no solo promedios.

Un ejemplo cotidiano

Confundir dos documentos de una consulta común puede generar retrabajo. Confundir un procedimiento de emergencia puede provocar consecuencias mucho más graves. Antes de definir un indicador, relaciona las clases de error con el contexto y las personas expuestas. La gravedad no se deduce de que haya una sola equivocación ni del número de mensajes emitidos por el sistema.

Ahora prueba tú

¿Por qué dos errores de clasificación pueden tener gravedad diferente?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 44](#)

C.2.10 · Seguridad de la información y del sistema

Qué pregunta abre

La protección de datos, modelos, componentes y accesos necesita medidas adecuadas al uso. Además de amenazas conocidas, considera instrucciones maliciosas, manipulación de datos y acciones que una integración permita ejecutar.

Valor y condiciones

Una integración puede ahorrar pasos en soporte. Si concede permisos excesivos, una salida manipulada puede llegar a modificar registros o revelar información con más facilidad.

En Tech Server IIT Spain

Para el chatbot, examina documentos recuperados y conectores como entradas que pueden contener instrucciones no confiables. Propón límites de permisos y pruebas acotadas. La arquitectura concreta y su eficacia quedan por comprobar.

Cómo comprobarlo

Combina verificación de accesos y registros con pruebas de comportamiento pertinentes. No introduces datos reales sensibles en ejercicios. Define qué resultado constituye exposición o acción indebida antes de ejecutar.

Un ejemplo cotidiano

Una hoja recibida de fuera puede contener instrucciones que el equipo nunca quiso ejecutar. En una integración de IA, distingue información que ayuda a responder e instrucciones que pretenden cambiar permisos o acciones. Prueba límites de confianza en un entorno acotado y documenta qué comportamiento sería indebido. No atribuyas eficacia a una protección por su nombre o por activarla sin verificar.

Ahora prueba tú

¿Basta con autenticar al usuario para impedir una inyección de instrucciones?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 44](#)

C.2.11 · Transparencia y explicabilidad

Qué pregunta abre

La información debe permitir que las personas entiendan el uso, sus límites y las vías de revisión. La explicación necesaria depende de la decisión y del destinatario. Una salida redactada con seguridad no demuestra que sus razones sean correctas.

Valor y condiciones

La personalización puede mejorar una recomendación si el usuario conoce qué está recibiendo y puede corregir información pertinente. Un técnico necesita detalles diferentes a los de un cliente que pide atención.

En Tech Server IIT Spain

En el chatbot, SUP-001 puede recoger aviso de interacción con IA, alcance y canal humano. Para el clasificador, registra entradas relevantes, salida y revisión. No atribuyas una explicación fiel del mecanismo a una justificación generada sin contrastar.

Cómo comprobarlo

Comprueba si un usuario entiende el límite y encuentra la revisión. Evalúa explicaciones con la evidencia de la tarea. No mides transparencia únicamente por la presencia de un aviso o el número de palabras.

Un ejemplo cotidiano

Un usuario que recibe una recomendación necesita saber cómo corregir un dato y pedir revisión. Un técnico necesita reconstruir entradas, versiones y transformaciones pertinentes. Diseña información distinta para cada destinatario. Una explicación extensa puede resultar inútil si omite el límite o la vía de intervención. Observa comprensión y capacidad de actuar en lugar de contar únicamente avisos publicados.

Ahora prueba tú

Una IA explica por qué asignó una prioridad. ¿Podemos aceptar esa explicación como prueba causal?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 45](#)

Fuentes y escenarios

Las fuentes ayudan a explorar condiciones que pueden generar un problema. Formula el evento, las personas afectadas y las consecuencias. Después valora el escenario mediante el proceso del SGIA.

Seis familias de fuentes de riesgo

FUNDAMENTOS DEL SGIA · ANEXO C

Seis familias de fuentes de riesgo

Una familia abre preguntas. El escenario concreta qué podría pasar, a quién y con qué consecuencias.

Entorno Población, canales y dependencias	Opacidad Información y trazabilidad insuficientes
Automatización Acciones permitidas e intervención	Aprendizaje automático Datos, patrones, errores y cambios
Hardware Capacidad, fallos y ejecución	Ciclo de vida Diseño, cambios, mantenimiento y retirada

Fuente → evento → afectados → consecuencias

Después: evidencia, incertidumbre y valoración con los criterios del SGIA. El impacto prepara preguntas para IMP-001.

Drift, alucinaciones, sesgos o APIs externas son ejemplos que se analizan dentro del contexto. No constituyen una lista exhaustiva.

Un servicio protegido y disponible puede dar una salida inadecuada.

José Manuel Hernández Agudo
Objetivos y fuentes de riesgo · ISO/IEC 42001:2023

C.3.1 · Complejidad del entorno

Qué pregunta abre

El entorno reúne personas, sistemas, proveedores, normas de trabajo y situaciones cambiantes. Una evaluación limitada a un conjunto de entradas puede omitir dependencias que alteran el resultado durante la operación.

Escenario didáctico

El chatbot responde bien en un piloto con documentos estables. Después se integra con varias bases, crece la demanda y aparecen consultas en otro idioma. Una respuesta puede utilizar una fuente inadecuada o retrasar el acceso a un operador.

Señales que investigar

Cambios en mezcla de consultas, carga, horarios, documentación o canales. Reclamaciones que se concentran en un segmento y resultados que difieren entre prueba y operación.

Cómo empezar a comprobar

Dibuja el recorrido de una solicitud y sus dependencias. Examina población, contextos no probados y capacidad humana. Relaciona FD-001, SUP-001 y seguimiento MON-001. No atribuyas el problema al modelo sin descartar otras capas.

Un ejemplo cotidiano

Un restaurante prueba un sistema de reservas solo con mesas pequeñas. Al llegar grupos grandes, aparecen condiciones que la prueba no cubría. En el chatbot, los cambios de idioma, horario y tipo de consulta exigen revisar la cobertura. Una señal favorable en una población reducida puede justificar un siguiente ensayo, pero no basta para autorizar cualquier extensión del servicio.

Ahora prueba tú

¿Qué cambia si el piloto solo incluyó consultas sencillas en horario de oficina?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 45](#)

C.3.2 · Falta de transparencia y explicabilidad

Qué pregunta abre

La organización puede desconocer por qué se produjo una salida, qué datos influyeron o qué limitaciones tiene el componente adquirido. Esa opacidad dificulta revisión, reclamación y diagnóstico.

Escenario didáctico

Un clasificador asigna prioridad normal a una incidencia crítica. El equipo solo conserva la etiqueta final y ha perdido el texto de entrada y la versión. No puede reconstruir si influyeron los datos, el modelo o la transformación aplicada después.

Señales que investigar

Salidas sin contexto registrable, documentación insuficiente, explicaciones no contrastables y cambios que no pueden compararse. La confidencialidad y la minimización también deben guiar qué registro se conserva.

Cómo empezar a comprobar

Identifica la información necesaria para decidir y las limitaciones aceptables. Pide evidencia pertinente al proveedor, examina trazabilidad de la integración y registra preguntas abiertas. Una explicación convincente necesita validación en la tarea.

Un ejemplo cotidiano

Una persona reclama un cargo y el equipo solo conserva el importe final. Sin referencias pertinentes le cuesta revisar lo ocurrido. Para el clasificador, define el registro mínimo que permite contrastar una prioridad sin conservar datos indiscriminadamente. La explicación debe apoyar la decisión que el destinatario necesita tomar. Un texto generado por el sistema puede parecer razonable y seguir pendiente de validación.

Ahora prueba tú

¿Cómo mejorar la investigación sin conservar indiscriminadamente todas las conversaciones?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 46](#)

C.3.3 · Nivel de automatización

Qué pregunta abre

Cuanto mayor sea la capacidad de una salida para provocar una actuación, más relevantes serán los límites, la reversibilidad y la intervención. El riesgo depende del uso efectivo y de la autoridad concedida.

Escenario didáctico

El copiloto comienza proponiendo borradores revisados. Una integración posterior permite enviar la respuesta o modificar un ticket automáticamente. Una salida incorrecta puede afectar al cliente antes de que alguien la compruebe.

Señales que investigar

Permisos nuevos, eliminación de revisiones, intervención demasiado tardía, sobrecarga de supervisores y aceptación habitual de propuestas sin contraste.

Cómo empezar a comprobar

Compara alcance anterior y nuevo, acciones permitidas y capacidad de parada. Evalúa usos indebidos previsibles y costes de recuperación. Relaciona CAM-001, SUP-001 e IMP-001 cuando corresponda.

Un ejemplo cotidiano

Una calculadora sugiere una cantidad y una persona decide pagarla. Si la herramienta puede hacer el pago directamente, cambia la exposición. En un agente, revisa herramientas conectadas, alcance de permisos y puntos de confirmación. El sistema no necesita un modelo nuevo para que ese cambio de autoridad requiera otra evaluación y una prueba de intervención.

Ahora prueba tú

¿Es equivalente sugerir una prioridad y aplicarla sin revisión?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 46](#)

C.3.4 · Fuentes ligadas al aprendizaje automático

Qué pregunta abre

El comportamiento aprendido depende de datos, etiquetas, supuestos y técnicas de desarrollo y evaluación. Puede fallar al encontrar entradas diferentes, reproducir patrones perjudiciales o producir una salida plausible sin fundamento suficiente.

Escenario didáctico

Si cambian las expresiones de clientes, el clasificador puede dejar de detectar situaciones críticas. Si las etiquetas históricas reflejan errores, puede aprenderlos. En un uso generativo, una respuesta puede inventar un procedimiento aunque la infraestructura funcione.

Señales que investigar

Variaciones de entrada o de relación entre entrada y resultado, diferencias por segmentos, referencias inexistentes y desacuerdo con revisión humana. Una señal de drift necesita contraste antes de afirmar una causa.

Cómo empezar a comprobar

Revisa datos, procedencia, etiquetas, separación de prueba y versiones. Distingue contaminación accidental o deliberada del entrenamiento de una inyección de instrucciones durante el uso. Si el modelo es de terceros, registra evidencia disponible y preguntas al proveedor.

Un ejemplo cotidiano

Un filtro aprende de correos etiquetados con errores. Puede repetir un patrón incorrecto aunque el programa ejecute lo previsto. Con datos deliberadamente manipulados aparece otra hipótesis, que necesita evidencia sobre procedencia y alteración. Un comprador de API debe pedir información pertinente y comprobar su tarea, sin afirmar que controla el entrenamiento del tercero. En generativa, además, conviene revisar fundamento y fuentes de las salidas.

Ahora prueba tú

¿Qué capas revisarías si aumentan los tickets críticos mal clasificados?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 47](#)

C.3.5 · Problemas del hardware del sistema

Qué pregunta abre

La ejecución depende de infraestructura física y de sus condiciones. Capacidad, fallos de componentes o cambios de ejecución pueden afectar disponibilidad, tiempos y comportamiento. Un comprador cloud mantiene preguntas sobre esas dependencias aunque no gestione el hardware.

Escenario didáctico

Una demanda superior a la capacidad genera espera o respuestas incompletas. Si una integración trata una respuesta incompleta como válida, puede aumentar el error operativo. Una modificación técnica de ejecución exige comprobar qué efectos produce en el servicio.

Señales que investigar

Saturación, timeouts, respuestas incompletas y diferencias entre entornos. Una latencia alta no demuestra por sí sola un fallo físico: también pueden intervenir red, cuotas o software.

Cómo empezar a comprobar

Relaciona observabilidad, límites de capacidad, evidencias del proveedor y manejo de errores. Prueba la alternativa y la conservación de solicitudes abiertas. Separa hipótesis de hardware y resultado comprobado.

Un ejemplo cotidiano

Un modelo se traslada a una plataforma de ejecución distinta. Errores de componente, incompatibilidades o condiciones técnicas pueden alterar el resultado o la continuidad. Compara el mismo alcance antes y después y registra configuración observada. Cuotas de servicio, problemas de red y fallos físicos son hipótesis diferentes: una demora no permite seleccionar una de ellas sin contraste.

Ahora prueba tú

¿Qué puede hacer una organización que no administra los servidores del modelo?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 47](#)

C.3.6 · Problemas a lo largo del ciclo de vida

Qué pregunta abre

El uso cambia desde la concepción hasta la retirada. Una finalidad que se amplía, un proveedor que cambia de versión o un componente que pierde soporte pueden invalidar supuestos y evidencias anteriores.

Escenario didáctico

El motor del chatbot modifica sus condiciones o retira una API. El equipo migra con prisa y conserva la antigua prueba como justificación. Pueden cambiar comportamiento, flujo de datos y disponibilidad de la alternativa.

Señales que investigar

Versiones no identificadas, avisos de deprecación, pruebas desactualizadas, ampliación de tareas y cuentas no autorizadas. La dependencia contractual es conocida en TI; la IA añade variaciones de comportamiento a revisar.

Cómo empezar a comprobar

Conecta SOL-PROV-001, PROV-001, CAM-001 y RET-001 con los escenarios e indicadores. Define qué cambio obliga a revisar y qué datos, accesos y obligaciones quedan al retirar el uso.

Un ejemplo cotidiano

Un departamento abre una cuenta gratuita para resolver una necesidad urgente. Ese uso fuera del canal autorizado puede quedar sin revisión de datos, permisos y condiciones. La Shadow AI entra aquí como ejemplo del curso, no como un epígrafe literal de C.3. Usa SOL-PROV-001 para detectar la necesidad, ofrecer alternativas y activar evaluación antes de transmitir información o ampliar la tarea.

Ahora prueba tú

¿La misma marca comercial garantiza que siga siendo válido el análisis anterior?

Tu respuesta antes de consultar la orientación

[Consultar orientación · página 48](#)

Tres casos integradores

Los escenarios son didácticos. Las cifras del clasificador son sintéticas. Los datos y condiciones ausentes permanecen pendientes de comprobar.

Caso 1. El chatbot: responder antes y conservar atención adecuada

La oportunidad es reducir esperas en consultas de soporte elegibles. El uso de referencia es USO-003. Antes de fijar una meta de ahorro, deben conocerse tiempo actual, mezcla de consultas, derivaciones y retrabajo. Contar mensajes producidos no representa la resolución de solicitudes.

La propuesta de objetivo incluye medir el tiempo total de resolución de consultas elegibles durante un piloto acotado. Las condiciones acompañantes son respetar alcance de respuesta, minimizar datos y facilitar una derivación efectiva. El responsable propuesto es el propietario del servicio con apoyo de TI y supervisión. Recursos, periodo y meta quedan por acordar.

Escenario: una base de procedimientos contiene versiones antiguas y el sistema recupera una instrucción no vigente. El chatbot podría ofrecer una solución plausible que agrave una incidencia o haga perder tiempo al cliente. «Alucinación» y «documentación obsoleta» describen mecanismos posibles diferentes; hay que investigar cuál interviene.

Preguntas de impacto: ¿quién recibe la respuesta y puede reconocer el error?, ¿qué ocurre si el cliente actúa antes de una revisión?, ¿puede acceder al canal humano?, ¿se ven afectados usuarios que necesitan atención urgente o un canal accesible? La gravedad depende del servicio y del contexto, no solo de que la respuesta sea falsa.

Otros escenarios incluyen instrucciones maliciosas dentro de documentos recuperados, información personal innecesaria en mensajes y una interrupción de API que desborde el canal humano. Son hipótesis didácticas para formular preguntas. No se declara que esta arquitectura concreta exista ni que esos incidentes hayan ocurrido.

Evidencias a preparar: FD-001 con fuentes y flujos, SUP-001 con límites y derivación, PROV-001 con modalidad y garantías aplicables, RET-001 con capacidad de alternativa y MON-001 con población, ventana e indicadores. La disponibilidad de una API y la calidad de la respuesta se observan por separado.

Actividad: escribe un objetivo con beneficio y condiciones, dos escenarios y tres preguntas de impacto. Señala una comprobación previa al piloto y una señal de revisión durante la operación. No completes los datos ausentes con garantías supuestas.

Caso 2. El clasificador: el promedio y las personas afectadas

SYS-003 / USO-004 clasifica incidencias. La oportunidad es mejorar enrutamiento y reducir trabajo repetido. R-001 e IMP-001 conservan la relación con riesgos e impactos. El objetivo debe incluir lo que importa al servicio cuando un ticket es crítico, además de la tasa global.

En el ejercicio hay 100 tickets sintéticos, de los que 10 son críticos. v1 acierta 90 de 100 en total y detecta 9 de 10 críticos. v2 acierta 94 de 100 y detecta 7 de 10 críticos. El acierto global sube del 90 % al 94 %. La sensibilidad a críticos baja del 90 % al 70 %. Los falsos negativos críticos pasan de 1 a 3.

La sensibilidad utiliza verdaderos positivos críticos divididos entre verdaderos positivos más falsos negativos críticos. El denominador de diez es pequeño y debe mantenerse visible. No se calcula precisión positiva ni tasa de falsos positivos porque faltan los datos necesarios para esas métricas.

El criterio didáctico de sensibilidad mínima del 90 % fue fijado antes de comprobar resultados. No es una exigencia numérica de ISO. Con ese criterio, v2 no se aprueba para desplegarla. Sí fue probada en el ejercicio. No se afirma que esté en producción ni que la versión anterior esté validada para cualquier contexto.

Escenario: un cambio de vocabulario o de mezcla de incidencias puede favorecer la infrapriorización de críticos y retrasar atención. La fuente puede relacionarse con aprendizaje, contexto o integración. La consecuencia requiere preguntar por urgencia, servicio y personas expuestas. El promedio por sí solo no responde a esas preguntas.

En T10-014, el registro didáctico señala prioridad normal a las 09:00, discrepancia reconocida a las 09:12 y recuperación/derivación a atención prioritaria a las 09:20. Son 12 minutos hasta detección, 8 desde detección hasta derivación y 20 en total. No se confirma causa raíz, daño efectivo ni incumplimiento de un SLA sin evidencia adicional.

El objetivo propuesto es mejorar el enrutamiento de la población elegible conservando el criterio crítico acordado y una revisión efectiva de discrepancias. CAM-001 documenta cambio y decisión, MON-001 define indicadores e INC-001 conserva el incidente y sus preguntas pendientes.

Actividad: explica por qué v2 no se aprueba, formula un escenario sin afirmar la causa y plantea dos preguntas para IMP-001. Indica qué evidencia adicional permitiría revisar la decisión y qué cambiaría si se ampliase la población.

El promedio y las situaciones críticas

FUNDAMENTOS DEL SGIA · ANEXO C

El promedio y los casos críticos

SYS-003 / USO-004 · 100 tickets sintéticos, con 10 críticos. Prueba didáctica; v2 sin aprobar.

v1

Acierto global

90 / 100 · 90 %

Detección de críticos

9 / 10 · 90 %

v2

Acierto global

94 / 100 · 94 %

Detección de críticos

7 / 10 · 70 %

Mejora global, peor detección crítica

Los falsos negativos críticos pasan de 1 a 3. Falta información para calcular precisión positiva o falsos positivos.

Decisión del ejercicio: no aprobar v2

Criterio previo de sensibilidad mínima: 90 %. Es un criterio didáctico del caso, no un umbral impuesto por ISO.

R-001 + IMP-001 + CAM-001 + MON-001

Relaciona riesgo, consecuencias, decisión de cambio y seguimiento. Conserva la incertidumbre de una muestra pequeña.

José Manuel Hernández Agudo

Objetivos y fuentes de riesgo · ISO/IEC 42001:2023

Caso 3. El copiloto: ahorro total y condiciones del uso

La oportunidad es reducir tiempo en borradores técnicos o apoyo al código. El identificador del uso debe tomarse del inventario. No asignamos uno nuevo. La modalidad del servicio y las condiciones aplicables se confirman mediante el canal de solicitud y evaluación de proveedor.

Una salida rápida puede requerir más revisión y corrección. La propuesta de objetivo compara el tiempo total de tareas equivalentes, incluyendo contraste de fuentes, pruebas y edición. Una reducción de tiempo de generación por sí sola no demuestra ahorro operativo.

Escenario de exactitud: el copiloto inventa una función o una referencia y quien revisa la incorpora sin comprobar. Puede aumentar retrabajo o introducir un defecto. Escenario de privacidad: se adjunta un documento con información personal que excede la finalidad. Escenario de propiedad intelectual: una salida destinada a un activo comercial exige verificar condiciones y derechos pertinentes.

Preguntas de impacto: ¿quién usará el código o documento?, ¿puede el error alcanzar a clientes?, ¿quién puede corregirlo y con qué demora?, ¿la revisión perjudica a empleados por carga o presión para aceptar propuestas?, ¿las personas saben cómo tratar información confidencial?

Las condiciones acompañantes incluyen revisión competente, pruebas para código de producción, datos autorizados y verificación de modalidad y condiciones. Se trata de propuestas de alcance. La guía no afirma una garantía contractual sobre ninguna marca ni una licencia automáticamente válida.

Relaciona SOL-PROV-001 para detectar la necesidad antes de adquirir, PROV-001 para diligencia debida, SUP-001 para revisión, CAM-001 para cambios pertinentes y MON-001 para seguimiento del uso. Registra una alternativa autorizada cuando exista y mantén abierta la comprobación cuando se desconozca.

Actividad: define una comparación justa de ahorro, un escenario de salida incorrecta y uno de flujo de datos. Señala qué necesitas conocer del proveedor y qué puede comprobar el equipo sin acceso a los datos de entrenamiento del modelo.

Actividad y Excel

El registro de trabajo relaciona propuestas de objetivos y escenarios. Utiliza el libro en blanco para tu actividad y el ejemplo para examinar cómo se mantienen evidencias y pendientes. Después desarrolla los documentos formales del curso.

Hoja Objetivos

Identifica uso y alcance, beneficio buscado y resultado. Añade condiciones acompañantes, método del indicador, línea de base, meta propuesta, acciones, medios, responsable, estado y evidencia. La ausencia de una línea de base debe aparecer como pendiente.

Hoja Escenarios

Relaciona la fila con un ID objetivo. Selecciona una familia de fuente y explica condición y evento posible, personas y consecuencias. Añade preguntas de impacto, referencias, comprobación y responsable. Conserva la diferencia entre una hipótesis y una observación.

Documentos de trabajo

Plantilla: Anexo_C_Objeticivos_Riesgos_Vacio.xlsx

Ejemplo: Anexo_C_Objeticivos_Riesgos_Ejemplo.xlsx

Cómo trasladar el resultado

R-001 mantiene la valoración con los criterios acordados. IMP-001 desarrolla consecuencias para individuos y colectivos. MON-001 define indicadores y respuesta. CAM-001 documenta los cambios pertinentes. El registro de esta guía prepara esas relaciones y no crea una autorización automática.

Hojas de trabajo

Responde antes de consultar las orientaciones. Cada hoja conserva espacio para tu propuesta y sus comprobaciones. Identifica si trabajas con el caso ficticio o con información de tu organización.

1. Una meta de negocio bien delimitada

Elige automatización, personalización, eficiencia o innovación. Describe la tarea, los beneficiarios, el alcance y la línea de base que necesitarías.

Propuesta y justificación

Comprobaciones, responsable y siguiente paso

[Consultar orientación](#)

2. Objetivo con condiciones acompañantes

Formula una propuesta, un indicador de beneficio y dos condiciones de aceptación. Añade responsable, medios, periodo y forma de revisar.

Propuesta y justificación

Comprobaciones, responsable y siguiente paso

[Consultar orientación](#)

3. Escenario que pueda investigarse

Elige una fuente y redacta condición, evento, afectados y consecuencias para uno de los tres usos. Señala evidencia y pendiente.

Propuesta y justificación

Comprobaciones, responsable y siguiente paso

[Consultar orientación](#)

4. Preguntas de impacto

Identifica individuos o colectivos que puedan soportar el error y las vías de revisión. Considera uso previsto y uso indebido previsible.

Propuesta y justificación

Comprobaciones, responsable y siguiente paso

[Consultar orientación](#)

5. Trazabilidad y seguimiento

Relaciona objetivo, escenario, valoración prevista, tratamiento e indicador. Asigna una comprobación y un responsable.

Propuesta y justificación

Comprobaciones, responsable y siguiente paso

[Consultar orientación](#)

6. Revisión cuando cambia el uso

Elige una modificación de modelo, una nueva tarea o una nueva población. Explica qué objetivos, riesgos e impactos revisarías.

Propuesta y justificación

Comprobaciones, responsable y siguiente paso

[Consultar orientación](#)

Orientaciones para revisar respuestas

Una respuesta adecuada concreta el alcance, justifica la relación y conserva aquello que aún necesita comprobación. Estas orientaciones no sustituyen la valoración del caso real.

C.2.1 · Responsabilidad y rendición de cuentas

Qué: Hay un vacío de cobertura: conocemos al propietario del chatbot, pero no quién puede intervenir cuando no está disponible. La responsabilidad asignada y la capacidad de actuar son aspectos que deben comprobarse juntos.

Por qué: Si llega una consulta que excede el alcance, alguien debe poder revisarla, detener una actuación o derivarla. Tener un nombre en la ficha no demuestra que esa respuesta exista por la tarde.

Cómo: Define un sustituto, sus facultades, cómo recibe la alerta y cómo accede a las solicitudes. Después ensaya una derivación en ese horario y observa si llega a una persona capaz de actuar.

Cuándo: Antes de ampliar el horario del servicio y cuando cambien turnos, responsables o canales. Si la cobertura sigue pendiente, identifica los límites de uso que deben mantenerse mientras se decide.

Dónde: En la operación del chatbot USO-003 y su canal humano; documenta la cobertura y las responsabilidades en SUP-001 y enlaza el escenario con R-001 cuando corresponda.

Con qué: Con la planificación de turnos, permisos de intervención, contactos y una solicitud de prueba. Conserva resultado, incidencias y responsable del siguiente paso; no marques el ensayo como realizado si solo está previsto.

[Volver a la actividad](#)

C.2.2 · Competencia en IA

Qué: Contar cursos demuestra participación en formación. Para afirmar competencia necesitamos observar que la persona aplica lo aprendido en una tarea de revisión relevante.

Por qué: Un certificado no indica si el revisor detecta una referencia inventada, comprueba un fragmento de código o reconoce datos que no puede introducir. Además, puede conocer la técnica y carecer de tiempo o autoridad para intervenir.

Cómo: Prepara un ejercicio controlado del copiloto con una salida que necesite contraste. Pide al revisor que localice el problema, justifique su decisión y muestre cómo impediría utilizar una propuesta inadecuada.

Cuándo: Antes de asignarle la función y cuando cambien tareas, herramientas o condiciones de revisión. Repite las comprobaciones pertinentes si la práctica revela dificultades; no midas competencia únicamente al finalizar un curso.

Dónde: En un entorno de ensayo o con casos autorizados de su actividad, relacionando formación y práctica con SUP-001 y la gestión de competencia de la organización.

Con qué: Con criterios previos, fuentes válidas, pruebas pertinentes, tiempo y permisos. Registra qué capacidad se observó, qué necesita mejorar y cómo se comprobará esa mejora.

[Volver a la actividad](#)

C.2.3 · Disponibilidad y calidad de datos de entrenamiento y prueba

Qué: Antes de aceptar el 94 %, pregunta qué se evaluó y qué errores importan al servicio. El acierto global necesita población, denominador, referencia y resultados de las situaciones críticas.

Por qué: En el ejercicio, v2 acierta 94 de 100 tickets, pero detecta solo 7 de 10 críticos: 70 % de sensibilidad. Un promedio mejor puede ocultar el incumplimiento del criterio específico necesario para la tarea.

Cómo: Revisa procedencia y cobertura de la muestra, etiquetas, separación respecto del entrenamiento cuando sea pertinente, versión y condiciones. Compara acierto y sensibilidad con los criterios fijados antes de ver resultados.

Cuándo: Al diseñar la prueba y antes de decidir sobre el despliegue. En este ejercicio, v2 no se aprueba porque su 70 % incumple el mínimo didáctico previo del 90 %; no cambies el mínimo para acomodar el resultado.

Dónde: En el expediente de prueba de SYS-003 / USO-004, con referencias a CAM-001, R-001, IMP-001 y MON-001 según su finalidad.

Con qué: Con los cien tickets sintéticos, diez críticos de referencia, resultados y criterios. Conserva los límites: noventa no críticos no significa noventa de prioridad normal; faltan datos para precisión positiva y falsos positivos.

[Volver a la actividad](#)

C.2.4 · Impacto ambiental

Qué: Un menor número de consultas puede indicar menor actividad, pero no demuestra por sí solo menor consumo energético o impacto ambiental.

Por qué: Una consulta puede utilizar más recursos que otra y las tareas pueden necesitar varias generaciones o revisión. El tamaño del modelo, infraestructura, localización y alcance de la comparación influyen en la conclusión.

Cómo: Compara tareas equivalentes con calidad comparable. Define qué operaciones incluyes, registra actividad y retrabajo y pide datos ambientales pertinentes al proveedor. Explica el método y la incertidumbre de las estimaciones.

Cuándo: Cuando explores un objetivo ambiental, compares modalidades o revises una ampliación del uso. Si faltan datos, registra la obtención de una línea de base como siguiente paso, en lugar de anunciar un ahorro.

Dónde: En el uso del copiloto y su cadena de servicio, dentro del alcance que hayas definido. El registro de objetivos debe señalar qué parte de la actividad examina la comparación.

Con qué: Con registros de solicitudes, tareas completadas, información de infraestructura y fuentes del proveedor. Los recuentos son medidas de actividad; su conversión a energía o emisiones necesita datos y método compatibles.

[Volver a la actividad](#)

C.2.5 · Equidad

Qué: Puedes afirmar que en esos cinco ejemplos no se observaron errores, pero no que el sistema sea equitativo para todo el grupo o cualquier situación.

Por qué: La muestra puede no cubrir necesidades, formas de expresión o dificultades de acceso relevantes. Además, la equidad incluye consecuencias y posibilidad de revisión; no se reduce a una tasa de acierto.

Cómo: Define qué diferencia merece estudiarse y por qué. Examina cobertura y tamaño de las muestras, compara condiciones pertinentes y comprueba si los afectados pueden reconocer el problema y solicitar corrección.

Cuándo: Al preparar pruebas y evaluación de impacto, y ante señales o reclamaciones que sugieran diferencias. Revisa también la pertinencia del análisis cuando cambien usuarios, idiomas o canales.

Dónde: En la prueba y el recorrido del servicio, con preguntas desarrolladas en IMP-001 y resultados considerados en R-001. La ausencia de información no se registra como igualdad demostrada.

Con qué: Con referencias de evaluación, muestras pertinentes y datos cuyo tratamiento esté autorizado. Utiliza solo información necesaria y justifica cualquier segmentación; no recojas atributos personales por rutina para completar una tabla.

[Volver a la actividad](#)

C.2.6 · Mantenibilidad

Qué: Sin identificar la versión o configuración de una salida, cuesta comparar resultados y reconstruir qué condiciones existían cuando apareció una discrepancia.

Por qué: Dos respuestas diferentes pueden relacionarse con entradas, documentos, integración o un cambio del componente. Si mezclamos condiciones desconocidas, podemos atribuir una mejora o un fallo a una causa que no hemos contrastado.

Cómo: Identifica qué referencias ofrece el servicio, conserva los datos pertinentes de ejecución y solicita al proveedor información de cambios. Si no expone una versión, documenta qué puedes observar y el límite que permanece.

Cuándo: Al contratar o integrar la API, al recibir avisos de cambio y ante señales de comportamiento distinto. Define previamente qué cambios o incertidumbres desencadenan pruebas adicionales.

Dónde: En el registro de operación y el expediente de cambios CAM-001, enlazados con PROV-001 y el seguimiento MON-001. Distingue versión evaluada, aprobada y operativa.

Con qué: Con avisos del proveedor, referencias de solicitudes, configuración registrada y pruebas comparables. Las evidencias previas conservan su alcance; decide y justifica qué reutilizar y qué volver a comprobar.

[Volver a la actividad](#)

C.2.7 · Privacidad

Qué: La garantía de no entrenamiento cubre una finalidad concreta. Sigue pendiente comprobar qué modalidad y datos cubre y cómo se gestionan acceso, conservación, localización, eliminación y otros tratamientos pertinentes.

Por qué: Evitar entrenamiento no impide necesariamente conservar entradas o permitir accesos bajo otras condiciones del servicio. Tampoco convierte cualquier envío de información en autorizado ni prueba por sí solo una brecha.

Cómo: Dibuja el flujo, identifica categorías de datos y destinatarios y contrasta finalidad, contrato, configuración y condiciones de la modalidad exacta. Distingue lo confirmado por evidencia de una promesa comercial general.

Cuándo: Antes de utilizar datos en la herramienta y cuando cambien modalidad, finalidad, proveedor, condiciones o configuración. Si falta una condición necesaria, concreta qué uso queda pendiente de decisión.

Dónde: En FD-001 para el flujo y PROV-001 para el tercero, relacionados con la evaluación de privacidad y riesgos. Una cuenta de equipo y una individual de la misma marca pueden tener condiciones diferentes.

Con qué: Con contrato y condiciones vigentes, configuración observada, clasificación de datos e información pertinente sobre terceros y ubicaciones. Solicita revisión competente cuando corresponda y conserva el alcance de cada conclusión.

[Volver a la actividad](#)

C.2.8 · Robustez

Qué: Detectar drift y confirmar deterioro responden a preguntas distintas: qué ha cambiado y si ese cambio afecta a la tarea. Ninguna de las dos conclusiones sustituye automáticamente a la otra.

Por qué: Puede variar el vocabulario sin empeorar la clasificación. También puede aumentar el error porque una integración transforma una etiqueta, aunque las entradas y el comportamiento del modelo permanezcan iguales.

Cómo: Compara distribuciones o relaciones de referencia para examinar la deriva y evalúa resultados contra una referencia pertinente para examinar desempeño. Conserva población, ventana, versión y condiciones para interpretar ambas señales.

Cuándo: Durante el seguimiento y cuando cambien entradas, referencias o resultados. Una alerta inicia contraste y medidas proporcionadas; no demuestra por sí sola que haya que reentrenar.

Dónde: En las pruebas y MON-001 del uso concreto, llevando los resultados pertinentes a R-001, IMP-001 y CAM-001 si la respuesta implica un cambio.

Con qué: Con una línea de base, entradas y etiquetas autorizadas, resultados de referencia y registros de integración. Define el responsable y la comparación siguiente si los datos todavía no permiten distinguir las hipótesis.

[Volver a la actividad](#)

C.2.9 · Seguridad frente al daño

Qué: Dos errores de la misma clase pueden tener gravedad diferente porque cambian las personas afectadas, el servicio, el momento y la posibilidad de corregir las consecuencias.

Por qué: Una infrapriorización puede generar una demora leve en una consulta rutinaria o prolongar una interrupción relevante. La etiqueta técnica del error no expresa por sí sola ese alcance.

Cómo: Describe quién puede verse afectado y examina urgencia, exposición, duración, alcance y reversibilidad. Aplica los criterios del SGIA y considera las preguntas de impacto, separando posibles consecuencias de daño observado.

Cuándo: Al identificar y valorar escenarios y cuando un incidente o nueva evidencia cambia la comprensión de sus efectos. No esperes a demostrar una causa técnica para plantear un escenario prospectivo.

Dónde: En R-001 e IMP-001; cuando haya un evento registrado, utiliza INC-001 para los hechos y actuaciones. Las marcas temporales de T10-014 no prueban por sí solas daño ni SLA incumplido.

Con qué: Con contexto del servicio, cronología, requisitos pertinentes, evidencia del efecto y criterios acordados. Si falta información, registra su comprobación y evita asignar una gravedad solo por intuición.

[Volver a la actividad](#)

C.2.10 · Seguridad de la información y del sistema

Qué: Autenticar al usuario es útil para controlar acceso, pero no basta para impedir que contenido no confiable desvíe una interacción o actuación.

Por qué: Una instrucción puede venir de un usuario autenticado o de documentos y herramientas que el sistema procesa. La autenticación no convierte automáticamente todo ese contenido en instrucciones autorizadas.

Cómo: Identifica entradas y límites de confianza; separa instrucciones de material de consulta y limita herramientas, permisos y acciones. Prueba el recorrido completo con casos controlados y comprueba cómo se trata la salida.

Cuándo: Durante diseño e integración, antes de ampliar permisos o conectores y al revisar señales de uso indebido. Las medidas se vuelven a examinar cuando cambian esas condiciones.

Dónde: En el servicio, sus documentos recuperados, conectores y herramientas autorizadas. Relaciona datos, integración, supervisión y cambios con los registros pertinentes del SGIA.

Con qué: Con inventario de conexiones, permisos, fuentes y pruebas autorizadas. Registra resultados y límites; alterar una respuesta y ejecutar una acción son efectos distintos que dependen también de la integración.

[Volver a la actividad](#)

C.2.11 · Transparencia y explicabilidad

Qué: La explicación del sistema es información para contrastar, no prueba causal suficiente por el hecho de parecer razonable.

Por qué: Una salida puede ofrecer una justificación plausible sin describir fielmente qué produjo la prioridad. Aceptarla sin contraste puede cerrar prematuramente una investigación o impedir una revisión adecuada.

Cómo: Compara explicación, entrada pertinente, referencia, respuesta del componente y transformaciones. Formula pruebas que distingan hipótesis y adapta la información al operador que actúa o a la persona que solicita revisión.

Cuándo: Al diseñar explicaciones y trazabilidad, al validar su utilidad y durante discrepancias o reclamaciones. Conserva explícita la incertidumbre si la evidencia no permite establecer el mecanismo.

Dónde: En la interfaz, el registro de operación y el expediente de investigación o evaluación. Cada destinatario necesita información acorde con su tarea; no necesariamente todos los detalles internos.

Con qué: Con documentos pertinentes, referencias de configuración, registros protegidos y pruebas comparables. Establece acceso y conservación adecuados: investigar requiere evidencia suficiente, no acumular datos indiscriminadamente.

[Volver a la actividad](#)

C.3.1 · Complejidad del entorno

Qué: El piloto aporta evidencia sobre consultas sencillas y horario de oficina. Antes de extenderlo, hay que examinar las condiciones que la nueva propuesta añade.

Por qué: La dificultad de las preguntas, idioma, fuentes, carga o cobertura humana pueden cambiar el resultado y sus consecuencias. Que el código sea el mismo no demuestra que todo el recorrido siga siendo adecuado.

Cómo: Compara alcance original y propuesto, identifica diferencias relevantes y prepara pruebas de entradas, fuentes y derivación. Ensaya qué ocurre cuando el chatbot no puede responder dentro de sus límites.

Cuándo: Antes de ampliar horario, población, idiomas o tareas y cuando el seguimiento indique un contexto no cubierto. Las pruebas anteriores se conservan; se completa lo pertinente para la ampliación.

Dónde: En el recorrido de USO-003, desde la consulta hasta la respuesta y su alternativa. El ejemplo nocturno exige comprobar la cobertura real, no suponer que por la noche siempre desaparece.

Con qué: Con resultados del piloto, descripción del alcance, fuentes vigentes, turnos y casos representativos. Documenta qué cubre la evidencia, qué sigue pendiente y quién decide sobre la ampliación.

[Volver a la actividad](#)

C.3.2 · Falta de transparencia y explicabilidad

Qué: Mejorar la investigación significa conservar evidencia pertinente para reconstruir hechos e investigar hipótesis, con protección y conservación definidas.

Por qué: Guardar todas las conversaciones puede añadir exposición innecesaria; conservar solo la etiqueta final puede impedir comprobar qué ocurrió. La selección debe responder a la finalidad de la investigación.

Cómo: Define qué entrada, salida, transformación y configuración necesitas contrastar. Utiliza referencias o muestras controladas cuando sean suficientes y conserva originales autorizados cuando resulten necesarios, cuidando integridad y acceso.

Cuándo: Al diseñar registros y conservación, y al preparar una investigación concreta. Verifica que la evidencia se pueda obtener antes de que se elimine conforme a sus condiciones de conservación.

Dónde: En los registros pertinentes de datos, integración e interfaz y en un expediente de acceso limitado. El informe general puede contener referencias sin reproducir toda la información sensible.

Con qué: Con un propósito definido, clasificación de datos, permisos, referencias de ejecución y método de custodia. Ensaya la reconstrucción de un caso y registra qué límites impiden todavía contrastar alguna hipótesis.

[Volver a la actividad](#)

C.3.3 · Nivel de automatización

Qué: Sugerir una prioridad y aplicarla automáticamente no son usos equivalentes: cambia la capacidad de la salida para producir una actuación antes de la intervención.

Por qué: En la sugerencia, una persona puede contrastar antes de actuar. En la aplicación automática, un error puede modificar el proceso de atención de inmediato; la revisión posterior no deshace necesariamente todas las consecuencias.

Cómo: Dibuja propuesta, aprobación y ejecución. Identifica permisos, consecuencias y reversibilidad; prueba que la persona puede intervenir con información, tiempo y autoridad suficientes o que funciona la alternativa autorizada.

Cuándo: Antes de ampliar automatización y cuando cambien permisos, acciones, destinatarios o capacidad de supervisión. Un mismo modelo con nuevas herramientas puede requerir una revisión aunque su versión sea idéntica.

Dónde: En la integración que asigna la prioridad y el proceso de atención afectado, con cambios en CAM-001 y condiciones de supervisión en SUP-001.

Con qué: Con mapas del recorrido, permisos, criterios previos, casos de prueba y registros de actuación. Revisa R-001 e IMP-001 y documenta decisión, límites y seguimiento del alcance evaluado.

[Volver a la actividad](#)

C.3.4 · Fuentes ligadas al aprendizaje automático

Qué: Hay que revisar las capas que pueden explicar las discrepancias: entrada y referencia, modelo, integración, interfaz y operación humana, además del contexto.

Por qué: Un aumento de críticos mal detectados describe un resultado, pero no identifica automáticamente drift, un ataque o un defecto del modelo. Las mismas etiquetas finales pueden surgir de mecanismos diferentes.

Cómo: Contrasta primero hechos pertinentes y formula hipótesis verificables. Por ejemplo, compara la prioridad devuelta con la registrada para examinar una transformación, o revisa entradas y etiquetas para estudiar un cambio de referencia.

Cuándo: Ante una señal de deterioro, discrepancia o incidente, priorizando según evidencia y posibles consecuencias. Se pueden aplicar límites proporcionados mientras se investiga; no existe un orden universal para todas las capas.

Dónde: En las pruebas y el recorrido operativo de SYS-003 / USO-004. MON-001 conserva la señal y los resultados; INC-001 documenta hechos y actuaciones cuando haya un evento que investigar.

Con qué: Con referencias válidas, muestras autorizadas, ventanas comparables y registros de configuración e integración. Conserva supuestos y pendientes; la causa del resultado v2 no está demostrada en el ejercicio.

[Volver a la actividad](#)

C.3.5 · Problemas del hardware del sistema

Qué: Aunque no gestione los servidores, la organización puede definir requisitos, evaluar evidencia del tercero y comprobar el comportamiento de su integración y alternativa.

Por qué: Comprar cloud cambia la capacidad de control directo, pero mantiene la dependencia. Una respuesta incompleta puede producir un problema si el sistema receptor la trata como válida.

Cómo: Observa tiempos, errores y condiciones; pide evidencia pertinente al proveedor y ensaya cómo se reconocen fallos, se conservan solicitudes y se ejecuta la alternativa. Contrasta hipótesis antes de atribuir latencia a hardware.

Cuándo: Al seleccionar e integrar el servicio, en pruebas de continuidad y ante fallos o cambios de ejecución. La frecuencia de seguimiento se define según el uso y sus riesgos.

Dónde: En la interfaz con la API y el proceso que depende de ella, relacionando PROV-001, MON-001 y el plan de continuidad o retirada pertinente.

Con qué: Con requisitos de servicio, condiciones del proveedor, registros de respuesta y un ensayo autorizado. Comprueba la capacidad de la ruta elegida; una derivación humana es una posibilidad, no la solución universal a cualquier fallo.

[Volver a la actividad](#)

C.3.6 · Problemas a lo largo del ciclo de vida

Qué: La misma marca no demuestra que se mantengan modalidad, finalidad, configuración y condiciones que sustentaban el análisis anterior.

Por qué: Una cuenta individual, un plan de equipo y una API pueden tener funciones y condiciones distintas. También puede cambiar el uso de la organización aunque el proveedor conserve el nombre comercial.

Cómo: Compara alcance anterior y nuevo: datos, tareas, permisos, versión o referencias disponibles, condiciones y resultados relevantes. Identifica qué evidencia se mantiene y qué pruebas o evaluaciones necesitan completarse.

Cuándo: Al contratar o cambiar modalidad, ampliar tareas o población, recibir actualizaciones o preparar retirada. Define disparadores para volver a revisar si cambian los supuestos del análisis.

Dónde: En el ciclo de vida del uso y sus relaciones con terceros, enlazando SOL-PROV-001, PROV-001, CAM-001 y RET-001 cuando sean pertinentes.

Con qué: Con inventario, condiciones vigentes, historial de decisiones, pruebas y plan de alternativa. La revisión documenta responsables y límites; no convierte cada cambio en obligación de rehacerlo todo desde cero.

[Volver a la actividad](#)

Una meta de negocio bien delimitada

Qué: Delimita una necesidad operativa, no una promesa genérica. Ejemplo didáctico: reducir el tiempo total para resolver consultas elegibles del chatbot, manteniendo calidad y acceso a revisión.

Por qué: Si solo medimos rapidez de generación, podemos trasladar esfuerzo al usuario o al operador y presentar una mejora que no existe en el recorrido completo.

Cómo: Describe tarea, beneficiarios y consultas incluidas. Compara un recorrido manual equivalente con el asistido e incluye generación, revisión, corrección, derivaciones y retrabajo pertinentes.

Cuándo: Antes de proponer una meta numérica de ahorro y al revisar una ampliación. Si falta la referencia, planifica obtenerla y deja la cifra de mejora pendiente de aprobación.

Dónde: En la hoja Objetivos del Excel y en el alcance de USO-003 que se quiere evaluar. El ejemplo es una propuesta, no una autorización operativa.

Con qué: Con tiempos de referencia, criterios de calidad, registros autorizados y participación del equipo de atención. Anota qué necesitas medir, quién lo hará y qué condiciones permitirán comparar los resultados.

[Volver a la hoja de trabajo](#)

Objetivo con condiciones acompañantes

Qué: Formula beneficio y dos condiciones que no deban deteriorarse. Por ejemplo, evaluar reducción del tiempo total en consultas elegibles, con revisión de respuestas contra fuentes pertinentes y derivación efectiva cuando excedan el alcance.

Por qué: Una mejora de promedio puede coexistir con errores importantes. Las condiciones aclaran qué resultado sería aceptable y qué no puede compensarse solo con eficiencia.

Cómo: Define población, periodo, método y línea de base. Propón acciones, recursos, responsable y criterios de revisión; fija cifras cuando sean practicables y estén justificadas, y distingue propuesta, aprobación y resultado observado.

Cuándo: Durante planificación y antes del piloto o cambio que se evalúa. Revisa el objetivo cuando la evidencia o las condiciones de uso alteren la comparación.

Dónde: En la hoja Objetivos y en la planificación del SGIA, con indicadores conectados a MON-001. Mantén coherencia con política y requisitos pertinentes.

Con qué: Con una referencia útil, fuente del indicador, criterios de calidad y pruebas de la alternativa. Si falta información, concreta su obtención; una condición cualitativa también necesita un criterio y una evidencia evaluables.

[Volver a la hoja de trabajo](#)

Escenario que pueda investigarse

Qué: Redacta un escenario investigable. Ejemplo: si se incorporan documentos cuya vigencia no se ha verificado, el chatbot podría recomendar un paso obsoleto y retrasar o agravar una intervención.

Por qué: La etiqueta «alucinación» no identifica por sí sola qué ocurrió ni quién puede sufrir consecuencias. El problema puede relacionarse con documentación, recuperación, generación o integración.

Cómo: Separa condición, evento posible, afectados y consecuencias. Identifica evidencia disponible y un pendiente; si investigas un evento ocurrido, conserva varias hipótesis mientras no haya base para elegir una.

Cuándo: Al identificar riesgos y antes de ampliar fuentes o tareas; también cuando una discrepancia actualice el escenario. La identificación prospectiva no requiere daño ni causa técnica ya demostrados.

Dónde: En la hoja Escenarios, enlazada con el objetivo pertinente, R-001 e IMP-001. Usa el caso ficticio o tu organización, dejando claro a cuál pertenecen los datos.

Con qué: Con alcance, fuentes documentales, entradas y respuestas pertinentes, criterios de referencia y registros protegidos. Propón una prueba que distinga hipótesis y un responsable del siguiente paso, sin inventar una probabilidad.

[Volver a la hoja de trabajo](#)

Preguntas de impacto

Qué: Identifica quién puede soportar el error y qué posibilidad tiene de reconocerlo, pedir revisión y recuperarse. Incluye individuos y grupos pertinentes al contexto.

Por qué: Un resultado técnico no expresa todas las consecuencias. Una respuesta obsoleta puede afectar de modo distinto según urgencia, dependencia y acceso a ayuda.

Cómo: Plantea una pregunta individual, como qué puede ocurrir antes de recibir revisión, y una de grupo, como si determinados idiomas, canales u horarios tienen menor acceso efectivo. Examina también uso indebido previsible y efectos más amplios cuando sean pertinentes.

Cuándo: Al evaluar impactos y al revisar uso, población, permisos o evidencia de incidentes. Los resultados aportan información a la valoración de riesgos y pueden requerir nuevas comprobaciones.

Dónde: En IMP-001 y en el recorrido real o propuesto del servicio, con referencia al escenario de R-001. En el laboratorio del clasificador se estudian consecuencias posibles, no daño demostrado en clientes reales.

Con qué: Con contexto, vías de revisión, muestras pertinentes y evidencia autorizada sobre el servicio. Justifica los grupos que estudias; evita recopilar atributos personales innecesarios o declarar discriminación por un promedio aislado.

[Volver a la hoja de trabajo](#)

Trazabilidad y seguimiento

Qué: Mantén una relación verificable entre objetivo, escenario, valoración, medida y señal de seguimiento. Cada referencia debe permitir localizar qué evidencia sostiene la decisión.

Por qué: Un expediente puede estar completo en apariencia y seguir sin pruebas ejecutadas. La trazabilidad muestra qué se sabe, qué falta y quién debe comprobarlo.

Cómo: Relaciona las filas del Excel con los registros pertinentes. Ejemplo: una restricción de fuentes propuesta para el chatbot se vincula con su objetivo, el escenario, la prueba de vigencia y una comprobación de derivación. Registra estado y siguiente decisión.

Cuándo: Al preparar el expediente y después de pruebas, decisiones o cambios. Actualiza resultados cuando se obtengan; no conviertas una fecha prevista en una actuación realizada.

Dónde: En Objetivos y Escenarios, con referencias a R-001, IMP-001, MON-001 y otros documentos pertinentes. Los códigos identifican documentación, no conceden autorización.

Con qué: Con identificadores, resultados de prueba, fuente del indicador, responsable y revisión prevista. Para SYS-003, enlaza el criterio previo y el 70 % de v2 con CAM-001; la negativa de aprobación se basa en el incumplimiento medido.

[Volver a la hoja de trabajo](#)

Revisión cuando cambia el uso

Qué: Identifica qué parte del análisis cambia. Ejemplo didáctico: ampliar el chatbot a un nuevo idioma y horario puede afectar entradas, cobertura de fuentes y acceso a intervención.

Por qué: Una decisión conserva el alcance que fue evaluado. El mismo nombre de herramienta no extiende automáticamente la prueba a otra finalidad, población o autoridad.

Cómo: Compara antes y después: tareas, datos, permisos, referencias, resultados relevantes y capacidad humana. Relaciona cada diferencia con objetivos, escenarios e impactos y justifica qué pruebas reutilizas o completas.

Cuándo: Antes de aprobar la ampliación y durante el seguimiento posterior. Una señal relevante o un incidente puede abrir otra revisión aunque no cambie la versión del modelo.

Dónde: En CAM-001 y el expediente del uso, conectado con Objetivos, R-001, IMP-001 y MON-001. En el ejemplo de v2 se conserva la decisión de no aprobar su despliegue evaluado.

Con qué: Con alcance anterior, propuesta, evidencias históricas, criterios previos, pruebas pertinentes y alternativa. Asigna responsables y revisión; las condiciones necesarias que sigan pendientes no se presentan como satisfechas.

[Volver a la hoja de trabajo](#)

Glosario

Oportunidad

Posibilidad de obtener un resultado útil que debe examinarse junto con condiciones y riesgos.

Objetivo de IA

Resultado que la organización busca con respecto a la IA y cuya consecución debe planificarse y evaluarse.

Condición acompañante

Límite o resultado que debe mantenerse al perseguir un beneficio. En esta guía es un recurso didáctico para expresar compromisos.

Fuente de riesgo

Condición o elemento del que puede surgir un riesgo; necesita concretarse en el contexto del uso.

Escenario

Descripción de cómo una condición podría favorecer un evento y producir consecuencias para personas o servicio.

Evaluación de impacto

Proceso que examina consecuencias potenciales para individuos, grupos y sociedad en el contexto del sistema.

Drift de datos

Cambio en características de las entradas o su distribución que requiere investigar su relevancia.

Drift de concepto

Cambio en la relación entre entradas y resultados de referencia. No equivale simplemente a cambiar de versión.

Alucinación o confabulación

Salida plausible que contiene información falsa o sin fundamento suficiente para la tarea.

Envenenamiento de datos

Manipulación deliberada de datos que puede alterar el aprendizaje o componentes basados en datos.

Inyección de instrucciones

Entrada que intenta redirigir el comportamiento o las acciones del sistema mediante instrucciones no confiables.

Sensibilidad

Proporción de positivos de referencia detectados: $VP / (VP + FN)$. Su denominador y población deben ser visibles.

Explicabilidad

Capacidad de proporcionar información útil sobre una salida o funcionamiento, adecuada al destinatario y contrastable.

Hipótesis

Explicación posible aún pendiente de contraste. No debe presentarse como causa demostrada.

Línea de base

Referencia documentada con la que se compara una mejora, bajo condiciones y población definidas.

Trazabilidad

Relación reconstruible entre alcance, objetivos, decisiones, pruebas, resultados y documentos.

Revisión del expediente

Marca los elementos que has revisado y conserva las observaciones necesarias. Completar el ejercicio no acredita la eficacia de una solución ni el cierre de los pendientes reales.

☐ Objetivos delimitados

Uso, beneficiarios, resultado y condiciones pertinentes.

Observación / pendiente

☐ Método de evaluación

Población, periodo, línea de base, fuente e indicador.

Observación / pendiente

☐ Escenarios concretos

Condición, evento, afectados y consecuencias.

Observación / pendiente

☐ Impacto preparado

Preguntas para individuos, colectivos y vías de revisión.

Observación / pendiente

☐ Trazabilidad

Referencias R-001, IMP-001, MON-001 y otras pertinentes.

Observación / pendiente

☐ Seguimiento y pendientes

Responsables, recursos, comprobación y siguiente decisión.

Observación / pendiente

Clases de audio y precisiones docentes

Las clases publicadas complementan el estudio. Consulta las precisiones que acompañan cada grabación y utiliza esta guía para desarrollar las actividades y sus comprobaciones.

029. Cómo definir objetivos auditables de IA: Anexo C

Escuchar la clase en iVoox: <https://go.ivoox.com/rf/182059453>

Anexo C y metas de negocio (05:57-06:07). En este pasaje, el audio atribuye al apartado C.2 unas metas que no son sus epígrafes. La afirmación necesita esta corrección: la eficiencia, la automatización, la personalización y la innovación son puntos de partida de los ejercicios del curso. No son los epígrafes de C.2. Sus dimensiones se refieren a responsabilidad, competencia en IA, disponibilidad y calidad de datos de entrenamiento y prueba, impacto ambiental, equidad, mantenibilidad, privacidad, robustez, seguridad frente al daño, seguridad del sistema y transparencia y explicabilidad. El Anexo C es informativo y cada organización selecciona lo pertinente para su contexto.

Objetivos y evaluación. La propuesta debe ser coherente con la política y los requisitos aplicables y concretar acciones, medios, responsable, plazo y evaluación. La medición se plantea cuando es practicable; una evaluación cualitativa necesita criterio y evidencia definidos. El tiempo total es una medida central para valorar el ahorro en la tarea del copiloto. También deben examinarse calidad, privacidad, revisión humana y otras condiciones pertinentes. No existe una única métrica suficiente para evaluar todo el uso.

Pendientes y decisiones. Un pendiente documenta una comprobación que falta y necesita responsable, plazo y límites. No autoriza automáticamente el uso. Cuando falta una condición necesaria para una actividad, esa actividad no debe darse por autorizada. Otros trabajos previstos y autorizados pueden continuar dentro de su alcance. No todos los pendientes equivalen por sí solos a prohibir cualquier avance del proyecto.

Riesgos e impactos. La evaluación de riesgos también considera consecuencias pertinentes para personas, grupos y sociedad. La evaluación de impacto desarrolla esas consecuencias, tanto potenciales como observadas cuando corresponda, y aporta resultados al proceso de riesgos. Ambos procesos pueden actualizarse e informarse mutuamente. No es necesario suponer un informe completo y definitivo de impacto antes de empezar a identificar escenarios de riesgo. La seguridad del sistema y la seguridad frente al daño pueden relacionarse en errores, usos indebidos y ataques.

Clasificador. La muestra contiene cien tickets sintéticos, con diez críticos de referencia. v1: 90/100 de acierto global y 9/10 de detección crítica. v2: 94/100 y 7/10, respectivamente. La sensibilidad pasa del 90 % al 70 % y los falsos negativos críticos de uno a tres. El mínimo del 90 % es un criterio previo del ejercicio, no un umbral universal de ISO. v2 se probó y no se aprobó para desplegarse. «Críticos reales», dentro de la explicación de la muestra, debe entenderse como críticos según la referencia sintética, no como incidentes reales de producción.

Otras métricas. La precisión positiva y la tasa de falsos positivos se calculan cuando se dispone de sus componentes pertinentes y aportan información útil para el objetivo. El caso no indica cuántos no críticos se predijeron como críticos, y no debe suponerse que el clasificador distingue solo dos categorías. El acierto global no permite inventar ese desglose.

Documentación y competencia. Conserva la información requerida y pertinente, con trazabilidad, accesos y conservación adecuados. Una celda rellena no acredita aprobación ni una prueba ejecutada. La pérdida de

práctica humana del cierre es una hipótesis para revisar competencia, sustituciones y capacidad de intervención; no una consecuencia inevitable de automatizar ni un resultado observado en Tech Server IIT Spain.

Referencias del caso. Tech Server IIT Spain es ficticia. Chatbot USO-003; clasificador SYS-003 / USO-004. La evaluación de proveedores se documenta en PROV-001. El identificador y la modalidad del copiloto se confirman en el inventario. R-001, IMP-001 y MON-001 mantienen la valoración, las consecuencias y el seguimiento.

Referencias de consulta ISO/IEC 42001:2023, cláusulas 6.1.2, 6.1.4 y 6.2 y Anexo C, en una copia autorizada. <https://www.iso.org/standard/42001> NIST AI RMF 1.0, apartados 1.1, 3.2 y 3.3, como recurso complementario para riesgos, consecuencias y seguridad.
<https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>

030. Evaluación práctica de riesgos en IA: Anexo C

Escuchar la clase en iVoox: <https://go.ivoox.com/rf/182076164>

Esta grabación utiliza ejemplos ficticios de Tech Server IIT Spain para relacionar fuentes de riesgo, escenarios, impactos y decisiones. Acompaña su escucha con estas precisiones. El Anexo C es informativo; las secuencias y preguntas de la clase son recursos pedagógicos, no una metodología única ni requisitos idénticos para todos los sistemas.

Drift y respuesta al cambio (09:49-10:13). Un modelo que obtiene peores resultados no demuestra por sí solo deriva. La deriva de datos se refiere a cambios en las entradas o su distribución; la de concepto, a cambios en la relación entre entradas y referencia. Deben contrastarse población, ventanas, versiones, referencia e integración. Un cambio de datos no implica siempre degradación. Reentrenar puede ser una medida pertinente, pero no es una respuesta obligatoria o suficiente para cualquier cambio: puede ser necesario revisar reglas, integración, fuentes, límites de uso o componente, según el mecanismo y la capacidad de la organización.

Decisión sobre v2 (08:05-09:13). En cien tickets sintéticos, diez críticos, v1 obtiene 90 aciertos globales y detecta 9 críticos; v2 obtiene 94 y detecta 7. El acierto global mejora cuatro puntos porcentuales; la sensibilidad pasa del 90 % al 70 % y los falsos negativos críticos de uno a tres. v2 se probó y no se aprobó para el despliegue evaluado porque incumple el mínimo previo de sensibilidad del 90 % del ejercicio. Esa es la razón de la decisión, no una ausencia genérica de datos. El mínimo es didáctico, no impuesto por ISO.

Alcance de la muestra (07:57-09:40). Los otros noventa tickets son no críticos; el caso no afirma que todos sean de prioridad normal. Los resultados no indican a qué categoría se asignó cada error, cuántos falsos positivos hubo ni la precisión positiva. Tampoco prueban daño en tres clientes reales ni discriminación contra un grupo de personas. Las diferencias observadas permiten plantear consecuencias posibles y preguntas por idioma, canal, acceso a revisión u otros factores pertinentes, sin inventar hechos o garantías.

Hardware, defensas y acciones (10:25-11:52). Los problemas de infraestructura o del ciclo de vida pueden afectar continuidad y controles, pero no implican que todas las defensas de software colapsen automáticamente. Se comprueba el alcance y la eficacia de cada medida y de la alternativa. La inyección puede ser directa o indirecta, también por documentos o herramientas; sus posibles acciones dependen de integración y permisos. La manipulación del aprendizaje puede afectar entrenamiento, ajuste o reentrenamiento. Son mecanismos que requieren evidencia, no causas demostradas por cualquier salida incorrecta.

Pruebas, idiomas y turnos (02:46-03:48). El piloto y su ampliación son hipótesis didácticas. La ausencia de operadores nocturnos es una condición del ejemplo, no una consecuencia inevitable de trabajar de noche. Añadir idiomas exige comprobar entradas y cobertura; no prueba un cambio automático de los parámetros internos del modelo. Las pruebas anteriores conservan evidencia de su alcance. Una ampliación exige revisar y completar lo pertinente, no empezar siempre desde cero.

Privacidad y supervisión (05:22-07:21). El tratamiento con un proveedor externo puede estar previsto y autorizado. Esa circunstancia no demuestra una fuga ni acredita conformidad total de cualquier uso. Una cuenta fuera del canal autorizado requiere examinar finalidad, datos, modalidad, condiciones y acceso. Rechazar la salida no revierte una transmisión previa. Datos confidenciales no equivalen siempre a datos personales, y las conclusiones sobre una brecha se contrastan según el caso. FD-001 y SUP-001 diferencian controles de flujo y de intervención, pero se relacionan con proveedores, permisos, cambios y riesgos.

Cronología y límites de la investigación (04:06-05:08). T10-014 tiene prioridad normal a las 09:00, discrepancia detectada a las 09:12 y recuperación/derivación a atención prioritaria a las 09:20. Doce minutos hasta detección, ocho hasta derivación y veinte en total. Esto no prueba resolución final, causa raíz, versión operativa, daño efectivo ni incumplimiento de un SLA. La trazabilidad ayuda a contrastar hipótesis por capas; cuando la causa sigue incierta, se conserva el pendiente y su siguiente comprobación, sin inventar una certeza.

Documentación y decisiones (12:05-13:18). R-001, IMP-001 y MON-001 vinculan valoración, consecuencias y seguimiento; sus códigos no autorizan por sí solos una medida. La evaluación de riesgos considera consecuencias pertinentes sobre personas, grupos, sociedad y organización, y la evaluación de impacto aporta resultados a ese proceso. Pueden emplearse métodos cualitativos, cuantitativos o mixtos adecuados. Las consecuencias del escenario del chatbot son posibles, no daño medido. Una prueba de derivación debe comprobar el funcionamiento y la cobertura de la alternativa; no exige siempre un ensayo estadístico. Las propuestas quedan pendientes de la comprobación y aprobación que correspondan.

Identificadores del caso. Chatbot USO-003; clasificador SYS-003 / USO-004. Son referencias de uso y sistema, no versiones de modelo. El ID y modalidad del copiloto se consultan en el inventario. T10-014 no se atribuye a v2. El panel de disponibilidad aporta información técnica; tampoco en TI tradicional basta por sí solo para demostrar calidad, seguridad o cumplimiento de todas las condiciones del servicio.

Referencias de consulta ISO/IEC 42001:2023, cláusulas 6.1.2, 6.1.4 y 6.2 y Anexo C, en copia autorizada. <https://www.iso.org/standard/42001> Google for Developers · glosario de aprendizaje automático, complemento técnico: <https://developers.google.com/machine-learning/glossary#concept-drift> Podkopaev y Ramdas · cambios de distribución y riesgo del modelo, complemento técnico: <https://arxiv.org/abs/2110.06177> NIST AI RMF 1.0 · riesgos, consecuencias y seguimiento: <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>

Conexiones y referencias

La guía acompaña el trabajo de los módulos 5 y 6 y la revisión del módulo 10. Consulta A para selección de controles y B para aplicación y comprobación. Los ejemplos amplían las preguntas del Anexo C con escenarios de servicios TI; no constituyen una reproducción del estándar.

Módulo 5

<https://jmhernandez.tech/sgia-modulo-5/>

Módulo 6

<https://jmhernandez.tech/sgia-modulo-6/>

Módulo 10

<https://jmhernandez.tech/sgia-modulo-10/>

ISO/IEC 42001:2023

Norma de referencia. Consulta las cláusulas 6.1.2, 6.1.4 y 6.2 y el Anexo C en una copia autorizada.

<https://www.iso.org/standard/42001>

Vista previa de ISO/IEC 42001:2023

Índice y presentación. No contiene el texto completo del Anexo C.

https://webstore.ansi.org/preview-pages/ISO/preview_ISO%2BIEC%2B42001-2023.pdf

NIST AI 600-1. Perfil de IA generativa

Recurso complementario para ampliar ejemplos de IA generativa. No sustituye los requisitos de ISO/IEC 42001.

<https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence>

Mapa del curso

Recorrido y módulos disponibles.

<https://jmhernandez.tech/curso-sgia-iso-42001/>

Guía del Anexo A

Consulta y selección de controles.

<https://jmhernandez.tech/sgia-anexo-a-controles/>

Guía del Anexo B

Aplicación y comprobación de medidas.

<https://jmhernandez.tech/sgia-anexo-b-implementacion-controles/>