MONOGRÁFICO 5

CerbIA

CerbIA: aplicar ISO/IEC 42001 a la selección, operación y evaluación de un componente

Siete lecciones, ocho mapas visuales, matriz de correspondencia y veinte preguntas con sus respuestas.

José Manuel Hernández Agudo

Fundamentos del SGIA e ISO/IEC 42001
Itinerario avanzado · 1 de noviembre de 2026

Código y documentación fijados al commit **86d063106a8b**. Decisiones del SGIA, controles, evidencias y evaluación razonada.

Índice

Presentación 3

Lección 1 · Necesidad, recorrido del contenido y alcance del SGIA 5

Lección 1 · Mapa de aplicación al SGIA 8

Lección 2 · Hallazgos, puntuaciones y decisiones sobre el riesgo 9

Lección 2 · Mapa de aplicación al SGIA 12

Lección 3 · Errores, omisiones y actuación de la aplicación 13

Lección 3 · Mapa de aplicación al SGIA 16

Lección 4 · Integración, proveedores y datos 17

Lección 4 · Mapa de aplicación al SGIA 20

Lección 5 · Cobertura y métricas: qué puede concluirse 21

Lección 5 · Mapa de aplicación al SGIA 24

Lección 6 · Latencia, carga operativa y recalibración 25

Lección 6 · Mapa de aplicación al SGIA 28

Lección 7 · Dictamen del SGIA: responsabilidades, evidencias y decisión 29

Lección 7 · Mapa de aplicación al SGIA 32

Flujo de decisiones y responsabilidades 33

Matriz de correspondencia aplicada 34

Actividad del alumno y solución 36

Preguntas y respuestas razonadas 37

Cierre del monográfico 57

Síntesis para la dirección 59

Fuentes y revisión del estudio 60

Presentación

La pregunta del monográfico

¿Cómo debe gestionar una organización la incorporación de CerbIA a su sistema de IA? Para responder, primero entenderemos el componente y después decidiremos qué uso puede justificarse. El objeto técnico es un proyecto público; el objeto de gestión es el sistema de la organización que lo incorpora. Confundir ambos lleva a atribuir a una biblioteca decisiones que corresponden a personas con autoridad.

CerbIA es una biblioteca modular publicada por InditexTech. Compone mecanismos para evaluar contenido que atraviesa aplicaciones de IA. Sus señales dependen de reglas, modelos y configuración; la propia presentación advierte de sus límites como garantía de seguridad. La atribución pública no acredita una implantación interna en Inditex. Presentación y límites del proyecto

Un caso para aprender a decidir

Trabajaremos con un asistente hipotético que recibe consultas, utiliza información recuperada y puede llamar a herramientas. Una entrada puede contener datos personales, una credencial o una instrucción que intente desviar al modelo. Una respuesta también puede divulgar información. La organización necesita elegir qué contenido inspecciona, antes de qué efecto y con qué respuesta operativa.

El recorrido general combina carga, preprocesamiento, inspección y agregación. El resultado alimenta una decisión de la aplicación. Este diseño permite discutir selección de controles y riesgo residual, pero su existencia no prueba que una instalación concreta sea eficaz. La revisión del repositorio es estática: no hemos ejecutado el producto ni medido latencia, ahorro o eficacia.

Qué aprenderás

Las siete lecciones siguen una secuencia: necesidad y alcance; interpretación del resultado; errores y actuación; integración, proveedores y datos; cobertura y métricas; operación y recalibración; dictamen del SGIA. Cada una explica primero el funcionamiento y termina con una decisión razonada. Las referencias de código permiten comprobar afirmaciones; no se requiere programar ni instalar componentes.

El ejercicio de 60 entradas utiliza datos sintéticos. Sirve para distinguir detección, falsos bloqueos, exactitud y precisión. Sus porcentajes no describen el rendimiento de CerbIA. Los casos resueltos muestran cómo separar un hecho, una hipótesis y una evidencia pendiente, sin sustituir una evaluación de la organización.

Vocabulario básico

Término	Significado en este monográfico
SGIA	Sistema de gestión de la inteligencia artificial de una organización.
Entrada	Unidad de contenido que llega a una evaluación.
Escáner	Componente que busca determinadas señales.
Hallazgo	Resultado individual de un escáner, con contexto.
Puerta	Mecanismo que reúne resultados y calcula un veredicto.
Frontera	Punto donde el contenido pasa entre componentes.
Riesgo residual	Riesgo que permanece tras aplicar las medidas elegidas.
Evidencia	Información verificable que sostiene una conclusión.

Cómo utilizaremos ISO/IEC 42001

Relacionaremos el caso con alcance, responsabilidades, riesgos, impactos, operación, seguimiento y mejora. La numeración se apoya en el índice oficial de IEC. Para el Anexo A empleamos áreas de control, contrastadas en la presentación de BSI, página 6. Las correspondencias son propuestas docentes de aplicación; no declaran conformidad.

Los nombres de fichas, registros y roles son opciones prácticas del curso. Deben adaptarse al contexto. Identificar una versión facilita reconstruir una evaluación; una firma, un repositorio único o una revisión manual de cada texto no son exigencias universales que aquí atribuyamos a ISO. El criterio será siempre explicar qué se decidió, por qué, quién podía decidirlo y con qué evidencia.

Lección 1 · Necesidad, recorrido del contenido y alcance del SGIA

Sitúa el componente dentro de un sistema de IA y delimita qué contenido y operaciones quedan bajo control.

1. Qué necesidad se intenta cubrir

Un asistente puede recibir una petición legítima que incluya un documento con instrucciones maliciosas. También puede generar una respuesta que exponga una credencial. Son situaciones distintas: cambia la procedencia del contenido, su finalidad y el momento en que podría producirse el daño. Antes de elegir componentes, la organización necesita describir ese recorrido.

CerbIA ofrece piezas para inspeccionar contenido. No basta con afirmar «protegemos el asistente»: hay que señalar qué entrada se examina, qué señal se busca y qué operación depende del resultado. Una buena delimitación evita convertir una capacidad de la biblioteca en una cobertura supuesta del sistema entero.

2. Del contenido al resultado

Runner organiza el recorrido completo: reúne entradas de los cargadores, aplica los preprocesadores y envía cada entrada final a SecurityGate. La puerta evalúa el texto; el resultado se vincula después con su entrada. Llamar directamente a una puerta no incorpora automáticamente la carga y el preprocesamiento de Runner. Recorrido ejecutado por Runner

Pensemos en una consulta que incluye texto codificado. El cargador lo incorpora; un preprocesador puede transformar una representación; los escáneres examinan la entrada resultante. El preprocesador de decodificación especulativa explora candidatos acotados y acepta determinadas sustituciones. Eso no equivale a descifrar cualquier contenido oculto. Decodificación con condiciones de aceptación

3. Qué unidad estamos evaluando

Entry representa texto, origen, metadatos y tipo de contenido; puede conservar información de derivación. Es útil para preguntar de dónde procede una conclusión y qué transformación intervino. Entrada y derivación

Hay que distinguir documento original, fragmento extraído y entrada evaluada. FileLoader lee archivos como texto UTF-8; no aporta extracción nativa de PDF o DOCX ni OCR. Si el asistente recibe esos documentos, la extracción exige otro componente y una comprobación propia. Lectura real del archivo

La organización debe comprobar que lo evaluado representa el contenido que usará el modelo. Una extracción incompleta puede dejar fuera una instrucción o un dato sensible aunque el fragmento recibido por la puerta se examine correctamente.

4. Delimitar el alcance con ISO/IEC 42001

Aplicamos 4.3 y 4.4 al sistema de la organización. La decisión propuesta consiste en describir finalidad, usuarios, operaciones, fronteras y exclusiones. El componente se sitúa dentro de ese mapa; su repositorio no define el alcance del SGIA de quien lo utiliza.

Por ejemplo, «asistente interno de consulta documental» resulta demasiado amplio si puede enviar mensajes. Deben incluirse la consulta, la recuperación, la composición de respuestas y las operaciones externas pertinentes. Una exclusión necesita explicación: no se vuelve irrelevante un flujo porque el equipo todavía no lo haya integrado.

El responsable del sistema propone el alcance y la autoridad designada lo aprueba. En relación con A.6 y A.9, conviene situar la evaluación en el ciclo de vida y en el uso previsto, incluyendo las condiciones que obligarían a reconsiderarlo.

5. Aporte, brecha y evidencia

Aporte observado	Brecha que debe resolver la organización	Evidencia propuesta
Recorrido modular de contenido	Qué rutas reales pasan por él	Diagrama contrastado con la aplicación
Entradas asociadas a resultados	Correspondencia con el material original	Muestra de extracción y trazabilidad
Transformaciones configurables	Qué alteran y qué dejan fuera	Criterios y ejemplos revisados

Una ficha breve puede reunir finalidad, contenidos permitidos, tipos de datos, efectos externos, límites y dueño del sistema. No es un formulario oficial de ISO. Su valor está en permitir que desarrollo, operación y quien decide el riesgo compartan una descripción comprobable.

6. Ejercicio resuelto: el adjunto invisible

Supuesto docente. El equipo afirma que todos los adjuntos están protegidos porque configuró FileLoader para admitir la extensión PDF. El informe muestra únicamente texto de una ruta.

Solución. Admitir una extensión no demuestra extracción documental. Pediríamos el procedimiento de extracción, ejemplos comparados con el original y el punto en que el texto entra en la evaluación. Hasta disponer de esa evidencia, no afirmaríamos cobertura de los adjuntos. Tampoco concluiríamos que se ha producido una filtración: solo conocemos una brecha de justificación.

La medida inicial sería limitar el uso o completar y verificar el recorrido antes de autorizar contenido sensible.

7. La decisión que queda documentada

La organización autoriza un uso delimitado cuando conoce el recorrido y sus límites. Registra qué inspecciona, en qué momento y quién responde si cambia una ruta. Esa decisión se revisa al incorporar formatos, herramientas o usuarios nuevos. Lo aprendido aquí será necesario para interpretar cualquier resultado de las siguientes lecciones: primero debemos saber qué llegó realmente a la evaluación.

Amplía este punto en el curso

Mapa del curso SGIA e ISO/IEC 42001. Recupera la relación entre contexto, alcance y sistema de gestión antes de seleccionar controles.

Fuentes de esta lección

runner.py · líneas 76-114

speculative-decoding.adoc · líneas 8-36

entries.py · líneas 25-61

file.py · líneas 71-81

Lección 1 · Mapa de aplicación al SGIA

CERBIA · MONOGRÁFICO 5 · LECCIÓN 1

De la necesidad al alcance del control

Sitúa el componente dentro del sistema de IA de la organización.

NECESIDAD

Revisar contenido ante amenazas elegidas: datos, secretos o manipulación.

RUNNER · organiza el recorrido completo

1 · Cargar

TextLoader o FileLoader: texto UTF-8.

↓

2 · Preprocesar

Transformaciones configuradas, cuando se aplican.

↓

3 · SECURITYGATE · evalúa cada texto

Escáneres recogen hallazgos → **Agregador** combina riesgo BLOCK

↓

Veredicto, hallazgos y omisiones

ALCANCE DEL SGIA

QUÉ

Sistema, finalidad y amenazas.

DÓNDE

Entradas, salidas y herramientas incluidas.

LÍMITES

Contenido excluido, idioma y componentes activos.

4 · LA APLICACIÓN ACTÚA

Debe conectar el resultado con permitir, restringir o revisar.

ISO/IEC 42001 · 4.3 · Alcance / 8.1 · Operación

Delimitar la frontera y comprobar la actuación real del sistema.

Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática

Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001

José Manuel Hernández Agudo

Lección 2 · Hallazgos, puntuaciones y decisiones sobre el riesgo

Interpreta el veredicto como resultado de una configuración y justifica qué riesgo puede quedar después.

1. Qué devuelve la evaluación

Un escáner identifica determinadas señales. La puerta reúne resultados y devuelve GateVerdict con is_safe, score, rationale, findings y skipped_scanners. Es una estructura para interpretar una evaluación; no contiene por sí misma una autorización de negocio. Campos del veredicto

Conviene leer esos elementos juntos. Un booleano favorable sin contexto oculta preguntas esenciales: qué componente examinó el contenido, con qué acción, qué comprobaciones quedaron fuera y cómo se calculó la puntuación. La explicación ayuda a investigar, pero también necesita una lectura crítica acorde con la finalidad del sistema.

2. Acción, severidad y umbral

Solo los hallazgos positivos con acción BLOCK aportan puntuaciones a la agregación; cada puntuación se multiplica por el peso de su severidad. WARN y PASS permanecen en los resultados, pero no elevan ese agregado. La puerta es favorable si no hay hallazgos bloqueantes o si el agregado queda por debajo del umbral. Construcción del resultado

Por eso acción y resultado no son sinónimos. BLOCK identifica una clase de hallazgo que participa en el cálculo; el umbral y el agregador intervienen después. WARN tampoco significa «sin importancia»: una señal puede requerir atención organizativa aunque no cambie el booleano.

El PiiScanner de Core utiliza WARN por defecto. Si se pretende que su detección participe en el bloqueo, esa decisión debe quedar expresada en la configuración. Acción predeterminada de PII

3. Qué cambia con la agregación

MaxScoreAggregator utiliza la mayor puntuación ponderada; MeanScoreAggregator calcula la media. Elegirlos puede conducir a decisiones diferentes con los mismos hallazgos. Agregación por máximo Agregación por media

No existe una escala universal en la que cierta cifra signifique siempre «aceptable». Hay que considerar amenaza, efectos, datos de prueba y coste de los errores. Además, una puntuación de un detector no se interpreta automáticamente como probabilidad de que ocurra un daño.

La severidad HIGH tiene peso 0,75. Pesos de severidad Con un máximo y umbral 0,80, un único hallazgo HIGH, aun con puntuación original 1,00, aporta 0,75. Esta consecuencia debe entenderse antes de aprobar la regla.

4. Aplicar la evaluación y el tratamiento del riesgo

Usamos 6.1.2 y 8.2 para estructurar la evaluación del riesgo, y 6.1.3 y 8.3 para justificar el tratamiento elegido. La propuesta docente es formular escenarios concretos: una credencial pasa, una consulta legítima se bloquea o una señal relevante solo genera una advertencia.

Cada escenario requiere consecuencias y criterios de aceptación. Un bloqueo injustificado puede impedir acceso al servicio; un escape puede divulgar información. La organización compara ambos efectos según contexto, sin resolverlos únicamente con una media global.

También conectamos 6.1.4 y 8.4 con las consecuencias para personas y colectivos. Una advertencia sobre datos personales necesita analizar finalidad y destinatarios; detectar una secuencia no resuelve por sí solo las obligaciones de privacidad aplicables.

5. Evidencia que falta para decidir

Aporte	Brecha	Evidencia propuesta
Hallazgos interpretables	Relación con escenarios del sistema	Riesgos y criterios definidos
Agregador y umbral configurables	Justificación de la tolerancia elegida	Comparación sobre muestras etiquetadas
Advertencias conservadas	Tratamiento posterior	Regla de revisión o escalado cuando proceda

El expediente debe mostrar configuración efectiva, versiones, ejemplos representativos y responsable de aceptar el riesgo residual. Una revisión humana puede ser apropiada en casos seleccionados, pero no se presume necesaria para cada texto. Debe tener propósito, capacidad y una actuación prevista.

6. Ejercicio resuelto: dos resultados favorables

Supuesto docente. En el caso A hay una detección de PII configurada como WARN. En el caso B hay un único hallazgo BLOCK, puntuación 1,00 y severidad HIGH, agregador máximo y umbral 0,80. No hay errores ni otros hallazgos.

Solución. A puede devolver resultado favorable porque WARN no alimenta el agregado. B también: 1,00 × 0,75 = 0,75, que no alcanza 0,80. Ninguno demuestra ausencia de datos personales o aceptación organizativa del riesgo.

Si la finalidad era impedir esas salidas, revisaríamos conjuntamente acción, severidad y umbral, y probaríamos el efecto sobre contenido legítimo y prohibido antes de autorizar una nueva configuración.

7. Decisión y límite de la conclusión

Quien tenga autoridad sobre el riesgo decide qué comportamiento necesita el sistema y por qué. El equipo técnico traduce esa decisión y aporta pruebas; operación verifica su aplicación continuada. La evidencia vincula el criterio de aceptación con resultados observables. No declaramos «seguro» todo contenido permitido: afirmamos qué controles se aplicaron y qué riesgo permanece bajo las condiciones evaluadas.

Amplía este punto en el curso

Módulo 6 · Tratamiento y controles. Ayuda a justificar la selección del control y la aceptación del riesgo que permanece.

Fuentes de esta lección

_models.py · líneas 6-22

_gate.py · líneas 212-228

pii.adoc · líneas 15-25

max.adoc · líneas 5-12

mean.adoc · líneas 5-12

types.py · líneas 30-35

Lección 2 · Mapa de aplicación al SGIA

CERBIA · MONOGRÁFICO 5 · LECCIÓN 2

Leer un veredicto sin exagerar su alcance

Cada campo aporta evidencia distinta para decidir sobre el riesgo.

CAMPO	QUÉ DICE	QUÉ FALTA COMPROBAR
findings	Hallazgos de los escáneres.	Cobertura y relevancia de cada señal.
score	Agregado de BLOCK positivos, ponderados por severidad.	No es una probabilidad universal de daño.
is_safe	Resultado binario de la puerta.	No autoriza por sí solo una herramienta.
skipped_scanners	Comprobaciones omitidas y motivo.	Un resultado favorable puede tener omisiones.
rationale	Explicación del resultado.	No equivale a una certificación de seguridad.

EJEMPLO DOCENTE · UN SOLO HALLAZGO BLOCK

1 × 0,75 = 0,75

riesgo × peso HIGH = ponderado

0,75 < 0,80

máximo < umbral de puerta

Sin errores, con MaxScoreAggregator: is_safe puede ser True.

WARN y PASS se conservan; no elevan el agregado de bloqueo.

ISO/IEC 42001 · 6.1.2 / 6.1.3 · Riesgo

Interpretar la señal, valorar sus límites y justificar el tratamiento.

Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática

Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001

José Manuel Hernández Agudo

Lección 3 · Errores, omisiones y actuación de la aplicación

Distingue un fallo del escáner, un resultado inseguro y una operación realmente impedida.

1. Cuando el control no termina como esperamos

Supongamos que un escáner necesita una dependencia que deja de responder. El contenido puede ser legítimo o peligroso; el error técnico no permite saberlo. Para interpretar lo ocurrido debemos separar tres planos: la comprobación que se realizó, el resultado que devolvió la biblioteca y la actuación posterior de la aplicación.

CerbIA contempla políticas para excepciones de escáneres. No se debe confundir esa gestión con un plan completo de continuidad del servicio. La organización tiene que definir cómo informa al usuario, qué operación se detiene y quién interviene si la situación persiste.

2. Tres políticas con consecuencias distintas

Con fail se propaga el error; block fuerza un resultado inseguro, continuando según fail_fast; skip registra el escáner omitido. Una omisión no se convierte en un análisis sin hallazgos. Políticas de errores del escáner

La elección expresa una tolerancia al riesgo. Detener toda consulta ante un fallo puede proteger información sensible y afectar a la disponibilidad. Continuar sin una comprobación mantiene servicio y deja una brecha. Ninguna política es correcta por su nombre: se justifica según finalidad, consecuencias y medidas complementarias.

La aplicación también debe gestionar una excepción propagada. Capturarla y continuar con la llamada al modelo puede anular la intención de quien eligió fail. Por eso la evaluación necesita observar el recorrido completo.

3. Omisiones por tipo y corte anticipado

La puerta puede omitir un escáner cuando el tipo de contenido conocido no coincide con los tipos aceptados. Un tipo unknown no lo excluye por esa razón. Enrutamiento y omisiones

Existe además una salida anticipada: con fail_fast, un hallazgo BLOCK de riesgo positivo corta cuando su puntuación ponderada individual alcanza el umbral; un error tratado como block también puede hacerlo. No se compara aquí el agregado acumulado. Condiciones del corte anticipado

Deben distinguirse incompatibilidad, error y trabajo no realizado tras el corte. Este último no equivale necesariamente a una entrada explícita en skipped_scanners. Reconstruir cobertura exige mirar también el orden y las condiciones de ejecución.

4. El resultado debe producir un efecto

Obtener is_safe=False e impedir una operación son hechos diferentes. En una integración hay que comprobar que la decisión se toma antes de la llamada o publicación que pretende controlar. La aplicación puede detener, devolver un mensaje o dirigir un caso a una vía autorizada, según su diseño.

El ejemplo LangChain revisado incorpora lógica antes de ejecutar una herramienta y también examina su resultado. Actuación alrededor de una herramienta No debe generalizarse ese comportamiento a cualquier aplicación por haber instalado el paquete.

Un control posterior puede limitar la circulación del resultado, pero no deshace un correo enviado o un registro ya modificado. Este orden temporal es parte del tratamiento del riesgo.

5. Decisión operacional y evidencia

Aplicamos 8.1 a la definición y comprobación de las reglas operativas. En conexión con 5.3, alguien debe tener autoridad para aprobar excepciones y reanudar un servicio detenido. Con 9.1 se seguirá la frecuencia y las consecuencias de errores y omisiones.

Aporte	Brecha organizativa	Evidencia propuesta
Política de excepción configurable	Cuándo se permite continuar	Criterio por operación y tipo de dato
Resultado y motivo de omisión	Qué comprobación falta	Registro y aviso pertinente
Posibilidad de bloquear el flujo	Efecto real en la aplicación	Prueba de integración antes de la operación

Las reglas deben indicar plazo, responsable y condición de retorno cuando se habilite una excepción. Es una propuesta proporcional de operación, no una autorización genérica para ignorar fallos.

6. Ejercicio resuelto: el mensaje ya se envió

Supuesto docente. Un asistente envía un mensaje y después evalúa su contenido. La puerta devuelve resultado inseguro. El equipo concluye que el envío estuvo bloqueado.

Solución. La evaluación posterior acredita una señal, no la prevención del envío. Solicitaríamos evidencia del orden de las llamadas y del mensaje efectivamente publicado. La medida propuesta es situar la evaluación pertinente antes de publicar y comprobar que el resultado inseguro impide la operación.

Si el escáner falla con skip, hay que registrar la ausencia de esa comprobación y aplicar la regla aprobada para esa operación. No podemos denominar ese caso «texto revisado y seguro».

7. Qué se autoriza y quién responde

La autorización operacional concreta cómo tratar resultado inseguro, excepción y omisión. Desarrollo demuestra el efecto, operación atiende fallos y la autoridad del riesgo decide las excepciones significativas. Si se incumple un criterio establecido, se investigará el caso y se valorará la acción correspondiente. Un fallo aislado no se etiqueta automáticamente como no conformidad del SGIA sin identificar el requisito o criterio incumplido.

Amplía este punto en el curso

Módulo 6 · Tratamiento y controles. Relaciona las políticas de fallo y la actuación del anfitrión con el tratamiento elegido.

Fuentes de esta lección

gate.adoc · líneas 45-58

_validation.py · líneas 6-36

_gate.py · líneas 183-197

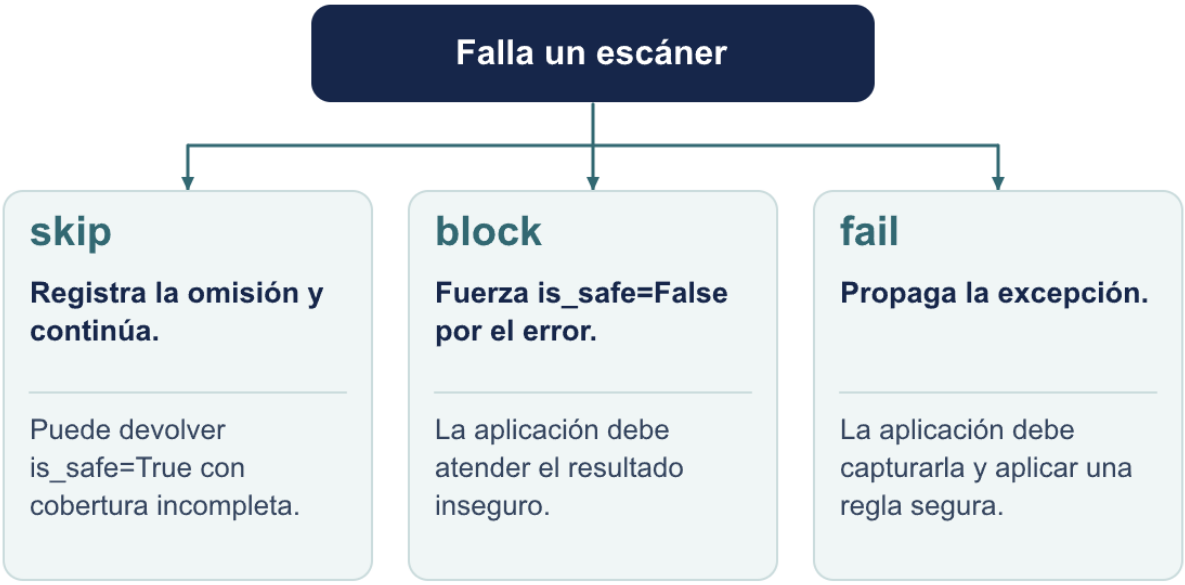
cerbia_middleware.py · líneas 231-267

Lección 3 · Mapa de aplicación al SGIA

CERBIA · MONOGRÁFICO 5 · LECCIÓN 3

Un error exige una actuación definida

Separa lo que hace la puerta de lo que debe hacer la aplicación.



DEL RESULTADO A LA PROTECCIÓN EFECTIVA

CERBIA — **Devuelve el veredicto o propaga el error.**
No sustituye la decisión operacional de la aplicación.

APLICACIÓN — **Permitir, restringir o derivar a revisión según política.**
Comprobar el efecto antes de llamar al modelo o herramienta.

ISO/IEC 42001 · 8.1 · Operación / 7.5 · Registros
Acordar la respuesta al fallo y conservar prueba de su efecto.

Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática
Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001
José Manuel Hernández Agudo

Lección 4 · Integración, proveedores y datos

Elige las fronteras de inspección y examina qué añade cada dependencia al sistema y a sus responsabilidades.

1. Dónde se coloca la evaluación

El ejemplo de LangChain permite configurar cuatro fronteras: input_gate para entradas del modelo; output_gate para sus respuestas; tool_gate para argumentos de herramientas; tool_output_gate para resultados de herramientas. Son puntos distintos del recorrido, con contenidos y consecuencias diferentes. Las cuatro puertas configurables

Una consulta aparentemente inocua puede provocar recuperación de contenido hostil. Una respuesta del modelo puede construir argumentos peligrosos. Por eso revisar solo la pregunta inicial deja otras superficies sin examinar. A la vez, no toda inspección tiene que emplear la misma combinación de escáneres: la selección debe responder a lo que sucede en esa frontera.

2. Qué aporta cada frontera

Frontera	Pregunta de control	Límite que debe considerarse
Entrada del modelo	¿Qué contenido se introduce?	No muestra toda respuesta futura
Salida del modelo	¿Qué se pretende entregar?	Debe actuar antes de la entrega
Argumentos de herramienta	¿Qué se va a solicitar fuera?	Examinar texto no concede permisos
Resultado de herramienta	¿Qué información externa vuelve?	El efecto de la herramienta puede haber ocurrido

El ejemplo es una demostración de integración, no evidencia de un despliegue interno. Su documentación recuerda que las herramientas reales necesitan restricciones de destino y manejo seguro de argumentos independientes. Alcance del ejemplo y responsabilidad de las herramientas

La aplicación determina qué campos se convierten en texto y qué contenido pasa por cada puerta. Esa selección merece revisión: un campo omitido puede contener precisamente la información que se quería controlar.

3. La configuración también es una decisión

El esquema real utiliza una raíz con nombre, listas de componentes, agregador, umbral y políticas. Cada componente se identifica por ruta de clase y argumentos de construcción. Campos de la configuración Selección de componentes

No necesitamos escribir YAML para entender su importancia: quien cambia una ruta o acción puede alterar qué código se ejecuta y qué señales cuentan. La organización debe tratar esa configuración como información controlada, con acceso y revisión acordes a su efecto.

La validación de la CLI comprueba configuración y construcción de componentes; no carga entradas ni ejecuta escáneres. Alcance de validate Una validación favorable sirve como evidencia técnica parcial, pero no demuestra cobertura ni eficacia.

4. Datos y dependencias que hay que conocer

Los componentes opcionales añaden condiciones propias. La integración Presidio revisada utiliza un analizador en inglés y el modelo en_core_web_sm. Esa configuración no acredita reconocimiento multilingüe universal. Idioma y modelo utilizados

También distinguimos detectar y transformar. Los resultados de escáner pueden incluir coincidencias; no constituyen un flujo general de anonimización. Contrato del resultado individual En KeywordScanner, redact oculta fragmentos en la explicación; no cambia el texto enviado al modelo. Redacción de la explicación

Para cada instalación se necesita conocer qué datos salen de cada componente, dónde se procesan y conservan, quién accede y qué dependencias efectivas intervienen. No se presume uso de un servicio remoto ni ausencia de transferencias por el nombre de un paquete.

5. Decisión ISO y evidencia pendiente

Aplicamos 7.5 y 8.1 a la configuración y operación. En las áreas A.7 y A.10 relacionamos datos y terceros; A.8 se refiere a información para partes interesadas. Estas relaciones orientan preguntas de gestión y no atribuyen títulos inventados a subcontroles.

Aporte	Brecha	Evidencia propuesta
Puertas configurables	Rutas realmente cubiertas	Diagrama y prueba por frontera
Integraciones opcionales	Condiciones heredadas	Inventario, versiones y evaluación del tercero
Señales sobre contenido	Tratamiento efectivo de datos	Flujo, accesos y política de conservación

La organización debe elegir qué información ofrece a usuarios y responsables sobre límites y fallos. La explicación será proporcionada al contexto y compatible con la protección de información sensible.

6. Ejercicio resuelto: una herramienta de mensajería

Supuesto docente. El asistente consulta documentos y envía mensajes. Solo se inspeccionan la pregunta inicial y la respuesta final. El equipo declara que también controla las herramientas.

Solución. Falta demostrar la evaluación de argumentos antes del envío y del contenido recuperado antes de introducirlo en el contexto. Propondríamos las fronteras pertinentes, permisos específicos y prueba de efecto. Revisar el mensaje final no demuestra que un envío anterior se haya impedido.

Solicitaríamos además qué campos se serializan y qué datos se registran. La existencia del ejemplo LangChain no acredita una conexión ya implementada con otra plataforma.

7. La autorización de integración

El responsable técnico acredita el recorrido; quienes gestionan datos y terceros aportan sus condiciones; la autoridad designada acepta el uso y sus límites. Una nueva dependencia, idioma o herramienta puede modificar el riesgo y requiere revisión proporcional. El resultado es una integración delimitada y evaluable, con evidencia propia de la organización.

Amplía este punto en el curso

Módulo 9 · Proveedores y ciclo de vida. Amplía la evaluación de dependencias y responsabilidades más allá del modelo principal.

Fuentes de esta lección

cerbia_middleware.py · líneas 82-105

README.md · líneas 9-14

configuration.adoc · líneas 33-45,97-115

_command.py · líneas 16-50

pii.py · líneas 47-80

scans.py · líneas 20-31

_scanner.py · líneas 110-120

Lección 4 · Mapa de aplicación al SGIA

CERBIA · MONOGRÁFICO 5 · LECCIÓN 4

Cuatro fronteras que hay que integrar

El ejemplo LangChain permite configurar estas puertas.

INTERCAMBIO CON LA PERSONA



Decidir qué mensajes y fragmentos se incluyen en cada revisión.

INTERCAMBIO CON HERRAMIENTAS



Inspeccionar texto no equivale a autorizar la ejecución de una herramienta.

DOS RESPONSABILIDADES COMPLEMENTARIAS

Integrador

Selecciona contenido, coloca puertas y aplica permisos y respuestas.

Configura clases e init_args; conserva versión y configuración activa.

Dependencias

Cada componente tiene cobertura, idioma y requisitos propios.

Si usa terceros, verificar qué datos reciben y bajo qué condiciones.

ISO/IEC 42001 · 4.3 · Alcance / 5.3 · Responsabilidades

Asignar las fronteras y las obligaciones de cada dependencia.

Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática

Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001

José Manuel Hernández Agudo

Lección 5 · Cobertura y métricas: qué puede concluirse

Comprueba primero que se evaluó el contenido esperado y después mide aciertos, escapes y falsos bloqueos con denominadores claros.

1. Cobertura antes de eficacia

La cobertura pregunta qué contenido y comprobaciones llegaron a evaluarse. La eficacia pregunta cómo se comportó el control frente a casos cuyo resultado esperado conocemos. Son dimensiones relacionadas, pero diferentes. Un excelente comportamiento sobre un fragmento no compensa que otras rutas relevantes no se hayan inspeccionado.

La arquitectura admite extensiones de preprocesamiento que pueden alterar la cantidad de entradas. Los resultados corresponden a las entradas finales. Cardinalidad de entradas y resultados Por ello no siempre basta comparar cuántos archivos había con cuántos veredictos se obtuvieron: hay que reconstruir el recorrido.

2. El resultado favorable vacío

ScanResult calcula is_safe comprobando que todas sus entradas sean favorables. Una lista vacía también devuelve True. Cálculo del resumen La documentación explica que un cargador configurado puede devolver cero entradas y entonces no se llama a la puerta. Resultado vacío sin evaluación

Esto enseña una precaución práctica: registrar resultado favorable y registrar evaluación realizada son cosas distintas. Si se esperaba contenido y no aparece ninguna entrada, el proceso debe reconocer esa diferencia conforme a una regla organizativa.

Para investigar, pediremos población esperada, entradas cargadas, entradas finales, escáneres aplicables y ejecutados, omisiones, cortes y errores. Las medidas deben indicar qué cuentan y cómo evitan duplicar contenido derivado.

3. Un conjunto de referencia útil

Para evaluar eficacia necesitamos ejemplos etiquetados con un criterio conocido. Deben reflejar amenazas y usos legítimos pertinentes, con idiomas, longitudes y formatos representativos. Incluir usos benignos de palabras sospechosas permite estudiar falsos bloqueos.

Un escáner de prompt injection trabaja con patrones; la alternativa de clasificación es otro componente opcional. Patrones y contexto de detección MaliciousUrlScanner utiliza heurísticas locales sobre enlaces. Naturaleza heurística de la señal Estas capacidades orientan casos de prueba; no prueban detección exhaustiva ni reputación actualizada de todo dominio.

La clasificación de referencia debe justificar qué se considera prohibido. Si dos revisores discrepan sobre un ejemplo, conviene resolver el criterio antes de usarlo como medida del detector.

4. Ejercicio resuelto con datos sintéticos

Supuesto docente, sin ejecución de CerbIA. Hay 20 entradas prohibidas y 40 legítimas. Se bloquean 16 prohibidas y 3 legítimas; se permiten 4 prohibidas y 37 legítimas.

Contenido de referencia	Bloqueado	Permitido	Total
Prohibido	16	4	20
Legítimo	3	37	40
Total	19	41	60

Solución. Detección: 16/20 = 80 %. Falsos bloqueos: 3/40 = 7,5 %. Exactitud: (16 + 37)/60 = 53/60 = 88,3 %. Precisión de los bloqueos: 16/19 = 84,2 %.

Las cuatro prohibidas permitidas son escapes respecto al criterio del ejercicio. Las tres legítimas bloqueadas representan una carga o un perjuicio potencial para usuarios. La exactitud resume ambas clases, mientras que la precisión responde qué proporción de lo bloqueado era prohibido. Sus denominadores no se intercambian.

5. Lo que esas cifras no prueban

El titular «88,3 % de precisión y seguridad garantizada» sería incorrecto. El 88,3 % es exactitud; además, una muestra sintética no acredita comportamiento real ni cumplimiento. Tampoco permite predecir qué ocurrirá si elevamos el umbral.

Antes de comparar configuraciones debemos mantener criterios y condiciones comparables, documentar cambios y separar segmentos relevantes. Si las entradas prohibidas son escasas en uso real, cambia la interpretación operacional de la cola de bloqueos.

También hay que presentar omisiones y errores junto a los aciertos. Si solo calculamos eficacia sobre lo que pudo completarse, debemos declarar ese denominador y mostrar aparte lo que quedó sin evaluar.

6. Aplicación de ISO/IEC 42001

Relacionamos 9.1 con seguimiento y evaluación, y 8.2 y 8.3 con revisar si el tratamiento sigue siendo adecuado. En A.5 consideramos los impactos de escapes y bloqueos; A.6 ayuda a situar la evaluación en el ciclo de vida.

Aporte	Brecha	Evidencia propuesta
Resultados por entrada	Población cubierta	Recuento y justificación de exclusiones
Hallazgos y omisiones	Eficacia frente a criterio	Muestra etiquetada y resultados completos
Métricas comparables	Riesgo aceptable en el uso	Umbrales de aceptación y decisión responsable

Estas propuestas conectan medición con decisiones; no convierten un porcentaje aislado en certificado.

7. Decisión que permite el informe

La organización decide si la cobertura y eficacia observadas son suficientes para un uso definido. Identifica limitaciones, escapes relevantes y medidas complementarias. Si faltan datos, limita la conclusión y planifica su obtención. El informe conserva los denominadores y distingue el ejercicio didáctico de cualquier evaluación futura de una instalación real.

Amplía este punto en el curso

Módulo 6 · Tratamiento y controles. Permite relacionar criterios de eficacia con selección y revisión del tratamiento del riesgo.

Fuentes de esta lección

architecture.adoc · líneas 87-99,151-160

results.py · líneas 30-50

prompt-injection.adoc · líneas 5-30

url-malicious.adoc · líneas 5-36

Lección 5 · Mapa de aplicación al SGIA

CERBIA · MONOGRÁFICO 5 · LECCIÓN 5

Mide el resultado y conserva el denominador

CASO SINTÉTICO · 60 entradas · no es un ensayo de CerbIA.

20 prohibidas + 40 legítimas = 60 entradas

REALIDAD CONOCIDA	BLOQUEADAS	PERMITIDAS
Prohibidas 20 entradas	**16** correctas	**4** escapes
Legítimas 40 entradas	**3** falsos bloqueos	**37** correctas

Detección	**16 / 20 = 80 %**	sobre las prohibidas
Falsos bloqueos	**3 / 40 = 7,5 %**	sobre las legítimas
Exactitud	**53 / 60 = 88,3 %**	correctas sobre el total
Precisión del bloqueo	**16 / 19 = 84,2 %**	prohibidas sobre bloqueadas

Cobertura: registrar también errores, omisiones y evaluaciones vacías.

ISO/IEC 42001 · 9.1 · Seguimiento y evaluación

El 88,3 % global no elimina los 4 escapes ni los 3 falsos bloqueos.

Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática

Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001

José Manuel Hernández Agudo

Lección 6 · Latencia, carga operativa y recalibración

Evalúa el coste completo del control y decide cambios sin perder de vista cobertura, riesgo y efectos sobre las personas.

1. Qué tiempos hay que distinguir

Un control puede ser eficaz y resultar difícil de sostener si incrementa excesivamente el tiempo de respuesta o genera una cola de revisiones. Para comprender ese coste debemos separar tiempo de la puerta, carga y preprocesamiento, llamada al modelo, herramientas y entrega final.

En la implementación revisada los escáneres compatibles se ejecutan secuencialmente. Orden de evaluación Esa característica orienta qué medir; no permite calcular una latencia universal. Algunos componentes opcionales incorporan carga de modelos y opciones de ejecución propias. Componente con modelo y condiciones propias

No hay en este monográfico un ensayo ejecutado ni cifras acreditadas de milisegundos. Las mediciones futuras deben describir sus condiciones para que una comparación tenga sentido.

2. Diseñar una medición que sirva

Proponemos fijar hardware, versiones, configuración, componentes, revisión del modelo cuando corresponda, corpus, idiomas, longitudes y carga simultánea. Separaremos arranque y carga de recursos del escaneo con el servicio preparado.

Conviene presentar distribución y percentiles, no solo promedio. El percentil 95 indica un valor bajo el que queda aproximadamente el 95 % de las observaciones de esa muestra. No describe por sí mismo el peor caso ni sustituye los datos de volumen.

La latencia debe segmentarse por frontera y tipo de contenido cuando eso explique diferencias. También se anotan errores, omisiones y cortes: un tiempo menor debido a comprobaciones no realizadas necesita una interpretación distinta de una optimización que conserva el control.

3. El coste incluye personas y servicio

Además de recursos de cómputo, hay tiempo dedicado a revisar alertas, atender reclamaciones y recuperar operaciones. Una política de WARN sin destinatario ni capacidad de respuesta puede producir señales que nadie utiliza.

La revisión humana debe tener un objetivo definido. Puede aplicarse a excepciones, muestras o casos de impacto elevado, según la evaluación organizativa. No se deduce que una persona deba revisar cada texto para que exista supervisión responsable.

También valoramos efectos sobre usuarios: demoras, bloqueos repetidos o pérdida de acceso. Una reducción de costes técnicos que traslade una carga desproporcionada a personas afectadas no es automáticamente una mejora del sistema.

4. Qué puede cambiar y qué debe permanecer visible

Ordenar escáneres puede cambiar cuánto trabajo se realiza cuando se permite el corte anticipado; con fail_fast desactivado no se obtiene ese corte. Efecto de fail_fast La decisión necesita comparar evidencia y finalidad: un corte útil para bloquear puede dejar menos información para investigar.

Elevar el umbral o cambiar una acción puede reducir bloqueos, pero también modificar escapes. Cambiar la política de error a skip no acelera una comprobación: permite continuar sin ella si falla. Una propuesta debe expresar qué mecanismo altera y qué riesgo acepta.

Aplicamos 8.1 al ajuste operacional y 9.1 a su evaluación. La cláusula 6.3 se examina cuando cambia el propio SGIA; no etiquetamos cada ajuste técnico como un cambio completo del sistema de gestión.

5. Un ciclo de recalibración razonado

La secuencia propuesta empieza con una línea base y un problema definido. Después reúne muestras de bloqueados, WARN y permitidos, con etiquetas y segmentos pertinentes. Compara la alternativa con la referencia, revisa cobertura, errores, escapes y falsos bloqueos, y añade latencia y carga humana.

Quien tenga autoridad decide con esa comparación. Antes del despliegue se acuerdan seguimiento, límites y condición de retorno. Si existe una no conformidad acreditada, se conecta con 10.2 y se comprueba la eficacia de la acción; otras mejoras pueden encuadrarse en 10.1.

Aporte	Brecha	Evidencia propuesta
Parámetros y componentes ajustables	Beneficio real del cambio	Comparación bajo condiciones equivalentes
Resultados para investigar	Carga de atención	Volumen, tiempos y capacidad responsable
Posible retorno de configuración	Recuperación practicable	Versión anterior y criterio de reversión

6. Ejercicio resuelto: menos cola, más incertidumbre

Supuesto docente. Suben la latencia p95 y la cola de advertencias. Se propone elevar el umbral y pasar de block a skip en errores. No se entrega una muestra comparable.

Solución. No aprobaríamos la propuesta con esa evidencia. Sabemos que existe una preocupación operativa, pero desconocemos el efecto sobre escapes y cobertura. Pediríamos línea base, segmentación, muestras etiquetadas y comparación de alternativas, incluyendo cambios en la capacidad de revisión.

La autoridad del riesgo decidiría después, con una condición de retorno y seguimiento definidos. Una cola menor, por sí sola, no demuestra mayor eficacia ni resuelve la causa del incremento.

7. La decisión queda unida a su propósito

Una recalibración se autoriza para resolver un problema concreto conservando un riesgo aceptable. Operación observa el comportamiento y escala desviaciones; desarrollo mantiene la comparación reproducible; la autoridad competente acepta las consecuencias. Así se evita que los ajustes sucesivos conviertan una medida de protección en una configuración cuyo propósito ya nadie puede explicar.

Amplía este punto en el curso

Módulo 9 · Proveedores y ciclo de vida. Sitúa actualizaciones y dependencias dentro de la operación y mantenimiento del sistema.

Módulo 6 · Tratamiento y controles. Permite valorar si el cambio mantiene el propósito y la eficacia del tratamiento.

Fuentes de esta lección

architecture.adoc · líneas 113-117

_scanner.py · líneas 149-175

gate.adoc · líneas 39-43

Lección 6 · Mapa de aplicación al SGIA

CERBIA · MONOGRÁFICO 5 · LECCIÓN 6

Recalibrar con pruebas y responsables

Un ajuste de umbral cambia decisiones: compara antes y después.

QUIÉN	QUÉ HACE Y QUÉ DEJA COMO EVIDENCIA
1 · Operación	**Detecta la señal** Escapes, falsos bloqueos, omisiones o aumento de la carga de revisión. **Registro del caso y configuración activa.**
2 · Equipo técnico	**Compara alternativas** Fija corpus, versiones y condiciones; prueba acción, severidad y umbral. **Resultados antes/después y plan de reversión.**
3 · Responsable del riesgo	**Autoriza el cambio** Contrasta eficacia, riesgo residual, latencia y recursos con criterios acordados. **Decisión con atribuciones y condiciones.**
4 · Operación	**Verifica tras el cambio** Confirma el efecto real; vigila desviaciones y revierte cuando se cumpla el criterio. **Seguimiento de eficacia y cierre del cambio.**

MEDIR LA LATENCIA EN SU CONTEXTO

Separar arranque y escaneo; medir p50, p95 y p99 con carga definida.

ISO/IEC 42001 · 5.3 · Roles / 9.1 · Medición / 10.1 · Mejora

Cambiar con autoridad, conservar la comparación y verificar la mejora.

Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática

Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001

José Manuel Hernández Agudo

Lección 7 · Dictamen del SGIA: responsabilidades, evidencias y decisión

Construye una conclusión proporcionada que distinga capacidad técnica, evidencia de la instalación y decisión organizativa.

1. Qué puede aportar el componente al expediente

Después del recorrido, resultados, errores e integración, ya podemos responder con precisión. CerbIA aporta mecanismos configurables de evaluación y resultados interpretables. Para que contribuyan a una decisión del SGIA deben conectarse con el sistema concreto, su riesgo y la actuación de la aplicación.

GateVerdict no incorpora firma, fecha ni identidad de firmante. Contenido del modelo de resultado El registro utiliza logging de Python y la aplicación anfitriona elige destinos y configuración. Registro dependiente del anfitrión Conservar evidencia recuperable requiere por tanto un diseño de la organización; no se presume un expediente de auditoría completo por obtener un veredicto.

2. De la responsabilidad técnica a la autoridad

Aplicamos 5.3 a distribuir responsabilidades y autoridades. El desarrollador puede explicar y probar la integración; no se deduce que pueda aceptar por sí solo cualquier riesgo de negocio. Operación necesita saber cuándo detener, escalar y reanudar. Quien gestiona datos y terceros aporta criterios de su ámbito.

La dirección o autoridad designada decide sobre límites, recursos y riesgo según la estructura real. La revisión independiente evalúa evidencia y criterios; no sustituye a quien opera el servicio. En A.3 relacionamos estas funciones con la organización interna.

Con 7.2 se comprueba competencia para interpretar omisiones, porcentajes y consecuencias de cambios. Conocer un comando no basta para justificar la autorización de uso.

3. Preparar evidencia que responda preguntas

El expediente propuesto reúne alcance y flujos; evaluación de riesgos e impactos; configuración y versiones; cobertura y pruebas; reglas operativas; resultados de seguimiento y decisiones de cambio. Cada pieza debe responder una pregunta y tener una persona responsable de mantenerla.

Aporte comprobado	Brecha de la instalación	Evidencia que solicitaríamos
Evaluación configurable	Adecuación a la finalidad	Riesgos, criterios y tratamiento elegido
Resultado estructurado	Efecto preventivo real	Pruebas de integración y excepciones
Extensibilidad	Mantenimiento de componentes propios	Responsable, validación y seguimiento

Aporte comprobado	Brecha de la instalación	Evidencia que solicitaríamos
Señales y registros posibles	Conservación y uso posterior	Política aplicada y muestra recuperable

El contrato de Scanner permite componentes que satisfacen un protocolo con identidad, acción, severidad, tipos y método de análisis. Contrato de extensibilidad Ampliar controles también amplía trabajo de evaluación y mantenimiento: no convierte una política específica en una función nativa ya acreditada.

4. Auditoría y revisión por la dirección

En 9.2, la revisión debe contrastar criterios aplicables con evidencia suficiente. En 9.3, la dirección utiliza resultados para decidir sobre recursos, prioridades, límites y continuidad. Son propósitos relacionados, pero una captura de un resultado favorable no satisface por sí sola ninguno de ellos.

Una conclusión útil dice qué se examinó, bajo qué versión y condiciones, qué resultados se obtuvieron y qué queda pendiente. Si falta evidencia, se expresa esa limitación. Para declarar una no conformidad se identifica el requisito o criterio incumplido y la evidencia que lo demuestra.

La selección de referencias al Anexo A se justifica desde el riesgo. No se asigna «cumplido» a un área porque exista un escáner con un nombre parecido.

5. Impactos e información a partes interesadas

El dictamen también debe explicar efectos sobre personas: exposición de información, errores de bloqueo, demoras o imposibilidad de acceder al servicio. Relacionamos esta evaluación con 6.1.4 y 8.4, y con A.5.

En A.8 planteamos qué información necesitan usuarios y otras partes interesadas sobre finalidad, límites, tratamiento de incidencias y vías de respuesta. En A.10 preguntamos por compromisos y condiciones de terceros; en A.7, por los datos utilizados.

Las medidas concretas dependen del contexto. No imponemos firma universal, revisión manual de cada mensaje ni una arquitectura única. Buscamos evidencia suficiente para la decisión que se pretende sostener.

6. Ejercicio resuelto: redactar una conclusión admisible

Supuesto docente. El equipo aporta el repositorio, una configuración validada y un ejemplo con is_safe=True. Afirma que el asistente cumple ISO/IEC 42001 y que no puede filtrar datos.

Solución. La evidencia acredita, como máximo, capacidad documentada, una comprobación técnica parcial de configuración y un resultado bajo condiciones concretas. Faltan alcance, riesgos, impactos, cobertura, efecto de la aplicación, eficacia y gestión continuada. No permite declarar conformidad ni ausencia de filtraciones.

El dictamen propuesto es: «La incorporación del componente puede contribuir al tratamiento definido. La autorización del uso sensible queda pendiente de las evidencias identificadas y de una decisión responsable sobre el riesgo residual». No afirmamos una infracción o una fuga sin evidencia.

7. Cerrar la decisión y revisarla

El responsable reúne el expediente; las funciones competentes revisan sus condiciones; la autoridad designada autoriza, limita o aplaza. Se documentan pendientes, plazo, seguimiento y desencadenantes de revisión. Si se acredita un incumplimiento, se aplica el proceso correspondiente y se verifica su eficacia. Así, el componente pasa de ser una promesa técnica a formar parte de una decisión del SGIA que otra persona puede comprender y contrastar.

Amplía este punto en el curso

Módulo 6 · Tratamiento y controles. Completa la justificación del tratamiento y de la aplicabilidad de las medidas elegidas.

Módulo 9 · Proveedores y ciclo de vida. Ayuda a sostener la decisión durante la operación, mantenimiento y evolución de las dependencias.

Fuentes de esta lección

_models.py · líneas 18-22

logging.adoc · líneas 9-12

base.py · líneas 7-37

Lección 7 · Mapa de aplicación al SGIA

CERBIA · MONOGRÁFICO 5 · LECCIÓN 7

De la evidencia a una decisión del SGIA

La conclusión debe poder reconstruirse desde hechos y criterios.

EXPEDIENTE PARA DECIDIR

1 · Alcance
Sistema, finalidad, fronteras y exclusiones.

2 · Riesgo
Consecuencias, criterios y tratamiento previsto.

3 · Configuración
Commit, clases, acciones, pesos y umbral.

4 · Pruebas
Corpus, cobertura, errores y métricas con base.

5 · Trazabilidad
Resultado, actuación real, cambios y revisión.

RAZONAMIENTO

Hecho
¿Qué se ha comprobado realmente?

Criterio
¿Qué exigía la organización para este uso?

Conclusión
¿Qué riesgo queda y qué decisión admite?

DECISIÓN DE QUIEN TIENE ATRIBUCIONES

Autorizar
Si satisface los criterios acordados.

Condicionar
Con límites y medidas verificables.

No autorizar
Si falta evidencia de un riesgo aceptable.

ISO/IEC 42001 · 7.5 · Evidencia / 10.1 · Mejora
Registrar quién decide, por qué, bajo qué límites y cuándo revisar.

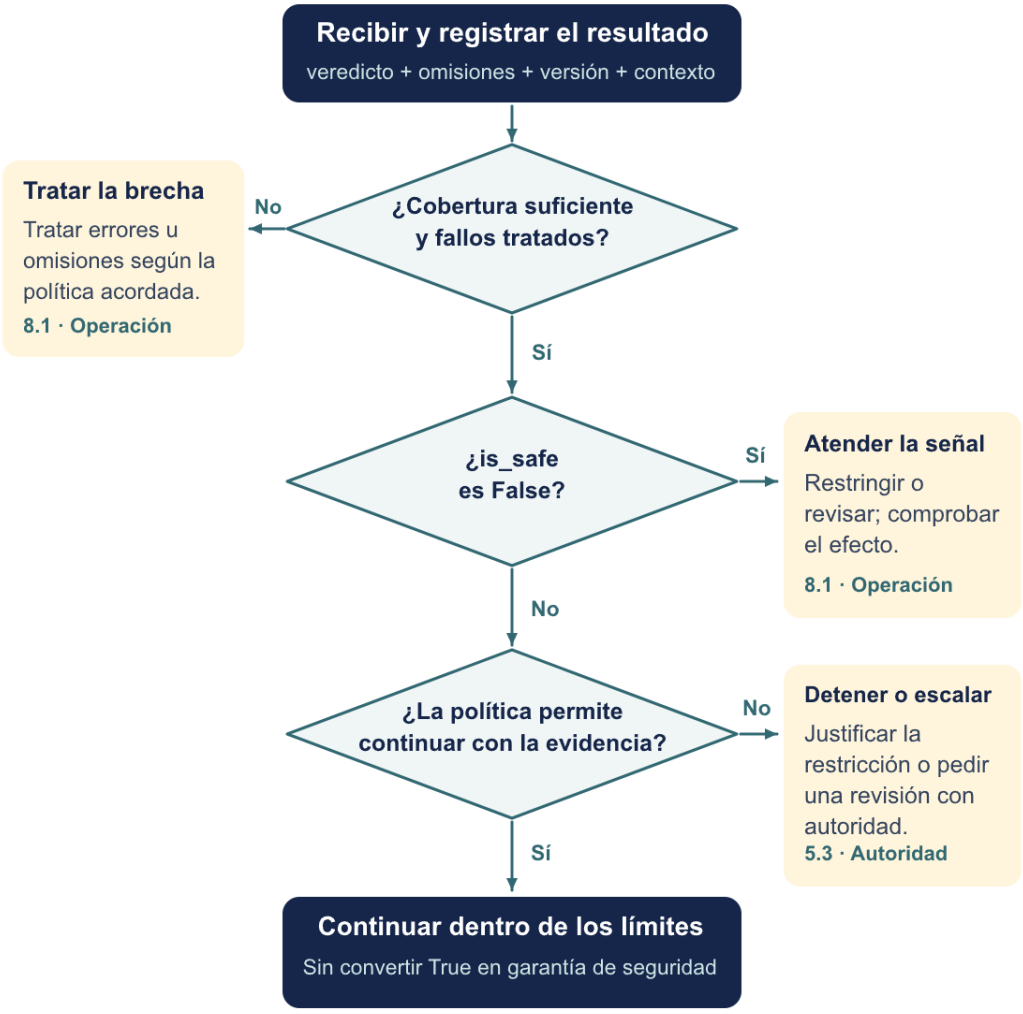
Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática
Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001
José Manuel Hernández Agudo

Flujo de decisiones y responsabilidades

CERBIA · MONOGRÁFICO 5 · DECISIONES

Del resultado técnico a la decisión del SGIA

Propuesta docente: la organización diseña esta política y asigna autoridad.



EN TODAS LAS RUTAS **Guardar decisión, actuación y criterio de revisión.**

ISO/IEC 42001 · 6.1.2 / 6.1.3 · Riesgo / 7.5 · Evidencia / 9.1 · Seguimiento
Revisar límites, recurrencia de errores y eficacia del tratamiento.

Fuentes: InditexTech/cerbia · revisión 86d0631 · lectura estática
Aplicación docente al SGIA · Fundamentos del SGIA e ISO/IEC 42001
José Manuel Hernández Agudo

Matriz de correspondencia aplicada

Matriz de aplicación de ISO/IEC 42001

La correspondencia relaciona mecanismos del componente con decisiones y evidencias de la organización. Se concreta según alcance, riesgos y controles seleccionados. Los registros propuestos son modelos docentes.

Mecanismo o decisión	Cláusulas y área	Decisión organizativa	Evidencia a valorar
Recorrido de carga, preprocesamiento y puerta	4.3, 4.4; A.6 · Ciclo de vida; A.9 · Uso	Delimitar sistema, contenidos, fronteras y exclusiones.	Ficha del sistema, diagrama contrastado y muestra del recorrido.
Resultados y escenarios de escape o bloqueo	6.1.2, 8.2; A.5 · Impactos; A.9 · Uso	Evaluar riesgos según consecuencias y contexto.	Registro de riesgos, criterios y evaluación por escenario.
Acción, severidad, agregador y umbral	6.1.3, 8.3; A.6 · Ciclo de vida; A.9 · Uso	Justificar selección y combinación de controles y riesgo residual.	Plan de tratamiento, aplicabilidad razonada y pruebas comparativas.
Consecuencias de escapes y falsos bloqueos	6.1.4, 8.4; A.5 · Impactos	Valorar efectos sobre personas y colectivos y medidas apropiadas.	Evaluación de impactos y seguimiento de consecuencias.
Interpretación y mantenimiento por personas competentes	5.3, 7.2; A.3 · Organización interna	Asignar autoridad, tareas y competencia suficiente.	Responsabilidades y comprobación de competencia aplicada.
Configuración y versiones identificadas	7.5, 8.1; A.6 · Ciclo de vida	Controlar información relevante para reconstruir la evaluación.	Configuración efectiva e informe recuperable, con versiones y condiciones.
Políticas fail, block y skip; actuación del anfitrión	8.1; A.9 · Uso	Definir respuesta ante fallo, omisión y resultado inseguro.	Reglas operativas, excepciones y comprobación antes del efecto.
Datos, extracción y componentes opcionales	6.1.2, 8.1; A.7 · Datos; A.10 · Relaciones con terceros	Conocer flujos, condiciones y responsabilidades heredadas.	Inventario efectivo, evaluación de terceros y flujo de datos.
Información sobre límites y vías de respuesta	8.1; A.8 · Información para partes interesadas	Determinar qué información necesita cada parte interesada.	Información proporcionada y canales de respuesta definidos.

Mecanismo o decisión	Cláusulas y área	Decisión organizativa	Evidencia a valorar
Cobertura, eficacia, omisiones, latencia y carga	9.1; A.5 · Impactos; A.6 · Ciclo de vida	Determinar si el control mantiene su propósito en el uso.	Métricas con denominadores, condiciones y segmentos.
Revisión del proceso y decisiones de dirección	9.2, 9.3; A.3 · Organización interna	Contrastar criterios y decidir límites, continuidad y recursos.	Hallazgos sustentados, resultados y decisiones responsables.
Recalibración y tratamiento de incumplimientos acreditados	8.1, 10.1, 10.2; 6.3 cuando cambia el propio SGIA; A.6 · Ciclo de vida	Comparar cambios, autorizar seguimiento y verificar eficacia.	Línea base, comparación, retorno y análisis de causa cuando proceda.

A.8 trata información para partes interesadas; A.10, relaciones con terceros y clientes. La cláusula 6.3 se refiere a cambios del propio SGIA. Las comprobaciones y la autorización de una configuración se concretan en el proceso aplicable.

ISO/IEC 42001 · Editor · Índice de la norma · IEC

Actividad del alumno y solución

El paquete incluye una ficha de calibración de frontera de protección en Word y PDF rellenable: versión del alumno y solución orientativa. Sus trece cuestiones cubren políticas de error, cobertura, métricas y recalibración.

Trabaja primero la ficha vacía. El campo de PDF puede desplazar texto largo; utiliza la página de ampliación e inspecciona la vista de impresión antes de imprimir. La solución ofrece conclusiones razonadas y criterios para contrastarlas.

Acceder a la actividad en el curso. Los archivos también están en Actividades/01_Alumno y Actividades/02_Soluciones del ZIP.

Preguntas y respuestas razonadas

Dieciséis preguntas de respuesta única y cuatro casos. Responde antes de consultar la solución y fundamenta qué puede concluirse, qué falta y qué decisión corresponde. Son preguntas originales del curso, no un banco oficial de acreditación.

Pregunta 1 · Runner, preprocesamiento y unidad de cobertura

Un servicio carga 100 documentos mediante Runner. Un preprocesador externo divide algunos documentos y elimina otros; el resultado contiene 140 EntryResult favorables. La política interna exige analizar todos los documentos recibidos. ¿Qué conclusión y siguiente comprobación están mejor fundamentadas?

A. Runner garantiza una salida por documento porque los preprocesadores incorporados conservan la cardinalidad.

B. Los resultados corresponden a entradas finales; debe reconstruirse la relación entre documentos recibidos, entradas cargadas, derivaciones y exclusiones.

C. El resultado favorable acredita el criterio interno, siempre que las 140 puntuaciones sean cero.

D. La cobertura es del 140 %: cada EntryResult representa necesariamente un documento original distinto.

Respuesta correcta: B

La arquitectura documentada sitúa el preprocesamiento entre la combinación de salidas de los cargadores y la llamada a SecurityGate por entrada final. Aunque los preprocesadores incorporados conservan cardinalidad, el contrato de extensión permite eliminar, expandir y reordenar entradas. Por ello, 140 resultados no acreditan que los 100 documentos hayan atravesado el control. Para 8.1 interesa contrastar el recorrido con el criterio operativo establecido; para 9.1 debe definirse la unidad de medida antes de calcular cobertura. La trazabilidad y las exclusiones justificadas permiten valorar el control del ciclo de vida asociado al área A.6 sin confundir fragmentos con documentos. La evidencia necesaria es una conciliación de identificadores, fuentes, derivaciones y exclusiones, vinculada a la versión de configuración. El responsable del control debe decidir si las exclusiones cumplen el alcance y documentar cualquier cobertura pendiente; un mayor número de fragmentos no compensa expedientes nunca inspeccionados.

Por qué se descartan las otras opciones

A. Traslada el comportamiento de componentes incorporados a un preprocesador externo cuyo contrato permite cambiar la cantidad de entradas.

C. La puntuación describe entradas evaluadas; no identifica documentos eliminados antes de la puerta.

D. Divide unidades diferentes y presupone una correspondencia que las extensiones no garantizan.

ISO/IEC 42001: 8.1; 9.1; A.6. Lección 1.

runner.py · líneas 76-105

architecture.adoc · líneas 87-99

Pregunta 2 · Enrutamiento por tipo y omisiones

Un escáner declara content_types=[code]. En una entrada tipada text aparece omitido; otra entrada unknown sí se examina. El responsable considera que existe un fallo porque unknown debería quedar excluido. ¿Cuál es la mejor interpretación para auditar la cobertura?

A. El comportamiento documentado acepta unknown en todos los escáneres; hay que verificar el tipado y separar omisiones de análisis completados.

B. La restricción code garantiza que también se rechaza unknown, por lo que ambas entradas debieron omitirse.

C. Toda restricción content_types se aplica durante la carga e impide crear las entradas incompatibles.

D. La omisión de text equivale a una ejecución sin hallazgos y puede sumarse como detección negativa completada.

Respuesta correcta: A

La documentación de configuración define el enrutamiento como elegibilidad del escáner, no selección del cargador ni transformación de entradas. Un tipo conocido incompatible genera una omisión, mientras unknown es aceptado por todos los escáneres; además, TextLoader y FileLoader producen actualmente unknown. Esta regla concreta obliga a examinar cómo se asigna el tipo en la integración. En 8.1, el auditor contrasta ese comportamiento con el diseño operativo aprobado; en 9.1, los recuentos deben distinguir ejecuciones y omisiones. El área A.7 resulta pertinente para la gestión de los datos y metadatos utilizados, sin convertir una omisión en evidencia de ausencia de riesgo. Solicitaría la configuración activa, ejemplos de tipos conocidos y desconocidos y contadores separados de elegibilidad, ejecución y omisión. Si el proceso exige un detector para todo el contenido, la persona responsable deberá justificar ese enrutamiento o corregirlo, conservando la comprobación posterior y sin reinterpretar unknown como análisis especializado.

Por qué se descartan las otras opciones

B. Contradice expresamente la regla que acepta las entradas unknown.

C. El enrutamiento decide qué escáner se invoca después de cargar y preprocesar; no filtra la creación de entradas.

D. El escáner omitido no analizó el contenido y no aporta un resultado negativo de detección.

ISO/IEC 42001: 8.1; 9.1; A.7. Lección 1.

_validation.py · líneas 6-37

_gate.py · líneas 84-114

configuration.adoc · líneas 194-210

Pregunta 3 · Contrato GateVerdict y presentación de resultados

Un panel del servicio traduce el resultado de CerbIA a las etiquetas Pass e Insecure, y su procedimiento comprueba solo esas cadenas. El responsable quiere que el panel sirva como evidencia del control. ¿Qué corrección ofrece una base más sólida para verificar la integración y comunicar sus límites?

A. Usar Action.PASS como sustituto del veredicto completo y descartar el resto de los campos.

B. Aceptar Pass e Insecure como valores oficiales porque son comprensibles para la persona usuaria.

C. Decidir exclusivamente con score, pues un valor cero equivale siempre a una evaluación completa y favorable.

D. Comprobar is_safe como booleano y preservar score, rationale, findings y skipped_scanners; las etiquetas visuales deben mapearse explícitamente.

Respuesta correcta: D

GateVerdict declara is_safe como booleano y conserva puntuación, explicación, hallazgos y escáneres omitidos. Action.PASS pertenece al enum de acciones de escáner, por lo que no sustituye ese contrato. Una interfaz puede mostrar etiquetas propias si documenta su correspondencia y no pierde información necesaria para interpretar el resultado. La relación con 7.5 está en mantener una especificación controlada y coherente con la versión utilizada; 8.1 permite contrastar que la aplicación consume correctamente el contrato. El área A.8 ayuda a revisar la información comunicada a las partes interesadas, evitando que una etiqueta favorable sugiera cobertura absoluta o ausencia de omisiones. Conservaría una prueba de extremo a extremo con advertencias, omisiones y errores, además de la correspondencia de etiquetas y el responsable que la aprueba. El registro debe ser suficiente para reconstruir la decisión, con acceso y conservación definidos: una explicación puede contener datos que no deban copiarse indiscriminadamente a paneles o registros.

Por qué se descartan las otras opciones

A. Confunde la acción de un escáner con el resultado agregado y elimina información de cobertura y errores.

B. La utilidad de una etiqueta visual no la convierte en un valor definido por el modelo de la API.

C. Un fallo tratado con BLOCK puede forzar is_safe=False sin un agregado positivo; score no reemplaza el veredicto.

ISO/IEC 42001: 7.5; 8.1; A.8. Lección 2.

_models.py · líneas 6-22

types.py · líneas 38-49

logging.adoc · líneas 9-22,25-42

_scanner.py · líneas 110-120

Pregunta 4 · Agregación de hallazgos bloqueantes ponderados

Sin errores y con fail_fast=False, una entrada produce tres hallazgos: BLOCK con riesgo 0,8 y severidad HIGH; WARN con riesgo 1,0 y severidad CRITICAL; BLOCK con riesgo 0,0 y severidad CRITICAL. Se usa MaxScoreAggregator y umbral 0,7. Al justificar el tratamiento, ¿qué resultado debe reconocerse y contrastarse con la intención de protección?

A. El agregado es 0,3 y la entrada es favorable porque siempre se promedian los dos hallazgos BLOCK.

B. El agregado es 0,6 y la entrada es favorable; solo el BLOCK positivo aporta puntuación y HIGH pesa 0,75.

C. El agregado es 0,8 y la entrada es insegura porque la severidad es una etiqueta sin peso numérico.

D. El agregado es 1,0 y la entrada es insegura porque el máximo siempre incluye todos los hallazgos.

Respuesta correcta: B

_build_verdict selecciona primero riesgos positivos y después acciones BLOCK; multiplica cada riesgo elegible por el peso de su severidad antes de llamar al agregador. Así, 0,8 por 0,75 produce 0,6; WARN y el BLOCK de riesgo cero no entran en ese cálculo. El máximo devuelve 0,6 y, al quedar por debajo de 0,7, is_safe resulta verdadero. Para 6.1.3 y 8.3, esta semántica debe ser coherente con el tratamiento elegido y su justificación de aplicabilidad. En 9.1, conocer la fórmula evita presentar la puntuación como probabilidad calibrada o confundir la intensidad de una advertencia con un bloqueo efectivo. La evidencia sería una tabla de hallazgos y un cálculo reproducible asociado a la configuración aprobada. Si se pretendía impedir ese contenido, el resultado revela que la combinación seleccionada no implementa esa intención; corresponde revisar el tratamiento y aceptar o reducir el riesgo residual mediante una decisión autorizada.

Por qué se descartan las otras opciones

A. Sustituye el agregador máximo por una media e incorpora un BLOCK que no tiene riesgo positivo.

C. Ignora el peso 0,75 definido para HIGH.

D. Incluye WARN pese a que no forma parte del conjunto enviado al agregador bloqueante.

ISO/IEC 42001: 6.1.3; 8.3; 9.1. Lección 2.

_gate.py · líneas 212-228

types.py · líneas 23-35

max.py · líneas 6-22

Pregunta 5 · Sentido del umbral y aceptación del riesgo

Una configuración produce, sin fallos, un único hallazgo BLOCK ponderado de 0,40. El responsable propone subir el umbral de 0,40 a 0,50 para reducir interrupciones. ¿Qué explicación describe correctamente el cambio y su evaluación en el SGIA?

A. Subir el umbral vuelve más restrictiva la puerta, por lo que esa entrada seguirá bloqueada y se reducirán los escapes.

B. Cualquier hallazgo BLOCK impide pasar con independencia del agregado y del umbral configurado.

C. El nuevo 0,50 demuestra un 50 % de riesgo y permite aceptar automáticamente los casos de menor probabilidad.

D. Con 0,40 la igualdad bloquea; con 0,50 la entrada será favorable, y deben evaluarse detección, falsos bloqueos y riesgo residual antes de aceptar el ajuste.

Respuesta correcta: D

Con hallazgos BLOCK elegibles, el código permite pasar únicamente cuando el agregado es estrictamente menor que threshold. La igualdad 0,40 frente a 0,40 queda del lado inseguro; al elevar el umbral a 0,50, el mismo agregado pasa. Reducir interrupciones puede disminuir falsos bloqueos, pero también permitir contenido prohibido que antes se detenía. Las cláusulas 6.1.2 y 8.2 orientan la reevaluación del riesgo en el uso concreto; 8.3 vincula la configuración con el tratamiento aplicado. Si el cambio requiere modificar el propio SGIA, debe examinarse esa planificación; no se identifica automáticamente todo ajuste técnico con un cambio del sistema de gestión. La evidencia mínima es una comparación etiquetada entre ambas configuraciones, con criterios de aceptación previos y una decisión del responsable. Deben examinarse también las entradas permitidas, porque observar menos interrupciones sin buscar escapes dejaría sin evaluar una consecuencia central del ajuste.

Por qué se descartan las otras opciones

A. Invierte el sentido de la comparación implementada.

B. Omite que un agregado BLOCK puede quedar por debajo de un umbral distinto de cero y producir resultado favorable.

C. Interpreta un parámetro de decisión como una probabilidad estadística de daño sin calibración que lo sustente.

ISO/IEC 42001: 6.1.2; 8.2; 8.3. Lección 2.

_gate.py · líneas 212-220

Pregunta 6 · WARN y eficacia del proceso de atención

Una entrada genera únicamente un hallazgo WARN de riesgo positivo, no hay fallos y la puerta devuelve is_safe=True y score=0. El registro interno exige revisar las advertencias de datos sensibles en dos horas. ¿Qué debería contrastar prioritariamente el auditor?

A. Que el equipo borre las advertencias favorables para evitar que distorsionen la tasa de bloqueos.

B. Que el hallazgo se conserve y active la atención prevista, midiendo plazo, carga y resolución sin confundir WARN con BLOCK.

C. Que WARN haya elevado el agregado hasta el umbral, pues esa es su función obligatoria.

D. Que cada WARN sea aprobado manualmente antes de continuar, aunque el proceso interno permita la operación.

Respuesta correcta: B

_build_verdict conserva los hallazgos recibidos, pero incorpora al agregado únicamente los riesgos positivos cuya acción es BLOCK. Por tanto, una WARN aislada puede coexistir con puntuación cero y resultado favorable sin contradecir la implementación. El criterio auditable del supuesto es el plazo interno de atención, no una obligación universal de bloquear advertencias. La cláusula 8.1 permite revisar la ejecución de ese proceso; 5.3 sitúa responsabilidades y 9.1 exige medidas útiles para conocer su funcionamiento. El área A.3 es pertinente para la organización de la atención. Volumen, antigüedad y resolución muestran si la advertencia conserva utilidad o se transforma en una cola inatendida. Revisaría asignación, fechas de recepción y cierre, pendientes y capacidad de revisión. PiiScanner usa WARN por defecto; seleccionar ese componente no crea un circuito humano. Si se pretendía bloquear, hay que comprobar conjuntamente acción, severidad, umbral y actuación del anfitrión, y autorizar la política resultante.

Por qué se descartan las otras opciones

A. Eliminar señales impide contrastar el criterio interno de atención y oculta carga operativa.

C. Atribuye a WARN una participación en la agregación que el código reserva a BLOCK positivo.

D. Introduce una revisión previa universal que no se deduce de la acción WARN ni del criterio establecido.

ISO/IEC 42001: 5.3; 8.1; 9.1; A.3. Lección 6.

_gate.py · líneas 212-228

_scanner.py · líneas 38-51

Pregunta 7 · Política FAIL y excepciones del anfitrión

Con on_scanner_error=FAIL, un escáner lanza una excepción. El anfitrión captura cualquier excepción y continúa llamando al modelo. La política aprobada del servicio exige no continuar si la evaluación obligatoria no termina. ¿Qué evaluación es la más fundada?

A. La captura del anfitrión es aceptable por definición porque mantiene la disponibilidad del modelo.

B. FAIL convierte automáticamente la excepción en GateVerdict(is_safe=False), así que la aplicación ya habrá bloqueado.

C. FAIL propaga la excepción original; debe probarse que el tratamiento del anfitrión cumple la política de no continuar ante esa evaluación incompleta.

D. FAIL equivale a una omisión registrada; basta revisar skipped_scanners para considerar completado el control.

Respuesta correcta: C

El caso FAIL de SecurityGate.scan ejecuta raise dentro del manejo de la excepción, sin construir por esa vía un veredicto ordinario ni registrar una omisión SKIP. El efecto protector depende entonces de cómo responda el anfitrión a la excepción. Aquí existe un criterio expreso que prohíbe continuar cuando falla la evaluación obligatoria; la captura que permite llamar al modelo puede contradecirlo si se verifica con evidencia suficiente. La cláusula 8.1 orienta esa comprobación operativa y 8.3 su conexión con el tratamiento aplicado. Si se acredita el incumplimiento, 10.2 exige abordar su causa y comprobar la eficacia de la corrección, sin atribuir el permiso al escáner. La prueba decisiva inyectaría un error controlado y observaría si se invoca el modelo, junto con la configuración y el procedimiento vigentes. El responsable debe asegurar que todas las rutas relevantes atienden el fallo; documentar únicamente que la biblioteca lanzó una excepción no demuestra el efecto operativo exigido.

Por qué se descartan las otras opciones

A. La disponibilidad no elimina el criterio operativo expresamente aprobado en el supuesto.

B. Confunde FAIL con BLOCK y además atribuye al componente el bloqueo efectivo de la aplicación.

D. La política que registra el escáner como omitido es SKIP; FAIL relanza la excepción.

ISO/IEC 42001: 8.1; 8.3; 10.2. Lección 3.

_gate.py · líneas 126-143

Pregunta 8 · BLOCK por fallo y cumplimiento del veredicto

Un escáner falla con on_scanner_error=BLOCK. Sin otros hallazgos bloqueantes, la puerta devuelve score=0 e is_safe=False. El desarrollador propone ignorar el booleano cuando la puntuación sea cero. ¿Cuál es la respuesta más sólida?

A. BLOCK impide físicamente todas las operaciones del proceso anfitrión, por lo que no hace falta integración adicional.

B. Sustituir BLOCK por SKIP y conservar score=0 como única regla del anfitrión para simplificar la integración.

C. El cero demuestra ausencia de riesgo, por lo que el booleano es un error de presentación que puede ignorarse.

D. El fallo fuerza el veredicto inseguro aunque no exista agregado positivo; el anfitrión debe aplicar la actuación aprobada usando el contrato completo.

Respuesta correcta: D

En _build_verdict, scanner_failures fuerza is_safe=False y sustituye la explicación, después de calcular el agregado de los hallazgos ordinarios. Por eso score=0 puede coexistir legítimamente con un veredicto inseguro. Ignorar el booleano eliminaría la vía de tratamiento del fallo prevista por la configuración. Para 8.1 y 8.3, el auditor necesita comprobar la respuesta real de la aplicación y su correspondencia con el tratamiento aprobado, no solo la presencia de BLOCK en un fichero. El área A.9 ayuda a situar la actuación en el uso del sistema. La selección de política debe justificarse según el contexto; la norma no impone universalmente este enum concreto. Solicitaría pruebas de fallo con trazas de las operaciones posteriores y evidencia de quién aprobó la regla de decisión. Solo si se demuestra que la aplicación incumple el criterio vigente procede formular la no conformidad y aplicar 10.2; la coexistencia de cero e inseguro no es, por sí misma, un defecto.

Por qué se descartan las otras opciones

A. SecurityGate devuelve una evaluación; detener o desviar una operación corresponde a la integración.

B. Cambiar a SKIP puede omitir la evaluación fallida y no corrige una integración que ignora el veredicto.

C. El agregado no representa todos los motivos que pueden forzar un resultado inseguro.

ISO/IEC 42001: 8.1; 8.3; A.9. Lección 3.

_gate.py · líneas 126-142,212-228

Pregunta 9 · SKIP y afirmaciones de cobertura

Para mantener el servicio durante una incidencia, se autoriza temporalmente SKIP en un detector. Tras un fallo, los demás escáneres no encuentran señales y el resumen indica is_safe=True. ¿Qué comunicación y seguimiento son adecuados?

A. Identificar el resultado favorable con cobertura reducida, conservar el motivo de la omisión y revisar la excepción, su vigencia y el riesgo residual.

B. Informar de evaluación completa sin amenazas porque el booleano es favorable.

C. Presentar el fallo como un falso negativo confirmado aunque no se conozca la clasificación del contenido.

D. Suprimir skipped_scanners para que la excepción autorizada no afecte a las métricas del servicio.

Respuesta correcta: A

El caso SKIP añade un SkippedScanner con motivo de fallo y continúa; no incorpora por esa vía un hallazgo bloqueante ni fuerza por sí mismo un veredicto inseguro. Un resultado favorable bajo esa política describe una evaluación con una parte omitida. La autorización temporal no convierte la omisión en detección completada. Las cláusulas 6.1.2 y 8.2 sitúan la valoración del riesgo residual; 7.5 permite conservar la decisión y su vigencia. El área A.8 resulta pertinente para comunicar con exactitud los límites del resultado. Revisar el motivo y la cobertura también evita mezclar fallos técnicos con falsos negativos, que requieren una referencia de clasificación externa. La excepción debería identificar responsable, periodo, contenido afectado, medidas compensatorias y condición de retorno. Contaría sus ejecuciones omitidas por separado y comprobaría si el nivel de exposición sigue siendo aceptable. Si vence la autorización o se incumple su condición, la conclusión debe sustentarse en ese criterio concreto.

Por qué se descartan las otras opciones

B. Atribuye al booleano una cobertura que la omisión documentada contradice.

C. Un fallo de ejecución no acredita que el contenido fuese prohibido ni que una decisión fuese incorrecta.

D. Oculta la pérdida de cobertura y elimina información necesaria para controlar la excepción.

ISO/IEC 42001: 6.1.2; 7.5; 8.2; A.8. Lección 3.

_gate.py · líneas 143-160,212-228

Pregunta 10 · fail_fast, orden y comparabilidad

Se comparan dos configuraciones con los mismos escáneres y fail_fast=True, pero distinto orden. La primera tarda menos y conserva menos hallazgos por entrada bloqueada. ¿Qué interpretación es la más defendible antes de declarar una mejora?

A. El menor número de hallazgos acredita necesariamente que la primera configuración tiene menos amenazas.

B. Las dos configuraciones son equivalentes para cualquier medida porque conservan la misma lista de componentes.

C. El orden puede anticipar una condición de corte; deben compararse etapas ejecutadas, finalidad de la medida y eficacia además de la latencia.

D. fail_fast detiene únicamente cuando el agregado final ya se calculó con todos los escáneres.

Respuesta correcta: C

SecurityGate recorre los escáneres secuencialmente y consulta _should_fail_fast tras cada ejecución. Con fail_fast activo, un fallo tratado como bloqueo o un hallazgo BLOCK de riesgo no aproximadamente nulo cuyo peso alcance el umbral puede interrumpir el recorrido. Reordenar componentes cambia qué señales posteriores llegan a recogerse y cuánto trabajo se realiza. La cláusula 9.1 exige que una comparación de rendimiento identifique ese recorrido y la finalidad de los recuentos; 8.1 sitúa su coherencia con el proceso aprobado. El área A.6 permite relacionar la configuración con su validación durante el ciclo de vida. Menor latencia con menos análisis no demuestra por sí sola mayor eficacia. Pediría corpus equivalente, versiones, orden, trazas de etapas ejecutadas y percentiles bajo la misma carga. El corte compara cada hallazgo individual con el umbral, no un agregado acumulado. La decisión debe distinguir ahorro operacional de pérdida de información para investigación y seguimiento, manteniendo los objetivos de protección acordados.

Por qué se descartan las otras opciones

A. Menos hallazgos recogidos pueden deberse a que los escáneres posteriores no llegaron a ejecutarse.

B. Ignora que el orden y el corte anticipado forman parte del comportamiento de la configuración.

D. El corte se evalúa durante el bucle y usa el hallazgo individual ponderado o el fallo bloqueante.

ISO/IEC 42001: 8.1; 9.1; A.6. Lección 6.

_gate.py · líneas 84-114,164-197

Pregunta 11 · Resultado vacío y suficiencia de evidencia

El panel muestra ScanResult.is_safe=True, cero hallazgos y cero entradas inseguras. Durante la ventana se recibieron 300 documentos y no se dispone de recuentos de carga. El responsable propone cerrar la comprobación de cobertura. ¿Cuál es la decisión de auditoría más adecuada?

A. Mantener pendiente la conclusión de cobertura y solicitar conciliación de entradas y criterio aplicable; una carga vacía también puede dar ese resumen.

B. Aceptar que los 300 documentos se analizaron porque el resultado global es favorable.

C. Calcular detección del 100 % a partir de la ausencia de hallazgos inseguros.

D. Cerrar la comprobación si se registraron 300 llamadas al servicio, sin exigir el recuento de entradas efectivamente cargadas.

Respuesta correcta: A

La arquitectura describe expresamente que un cargador puede devolver cero entradas: la puerta no se invoca y el ScanResult vacío es favorable con cero hallazgos. El panel observado es compatible tanto con contenido evaluado favorable como con ausencia de evaluación, por lo que no resuelve la cobertura de los 300 documentos. La cláusula 9.1 requiere medidas interpretables y 9.2 una conclusión sustentada en evidencia y criterios de auditoría. La cláusula 4.3 ayuda a delimitar qué flujo está comprendido en el alcance. No hay base para acreditar cobertura ni para inventar una prohibición normativa universal del resultado vacío; debe completarse la comprobación antes de calificar el hallazgo. Solicitaría inventario recibido, salidas de los cargadores, exclusiones del preprocesamiento y resultados vinculados por identificador. El responsable debe explicar y aprobar el tratamiento de una evaluación vacía. Tampoco leer un PDF como texto UTF-8 acredita extracción documental: ese recorrido requiere evidencia de contenido útil y extracción cuando corresponda.

Por qué se descartan las otras opciones

B. Confunde una propiedad del conjunto de resultados con evidencia de que existieron entradas evaluadas.

C. No existe referencia etiquetada ni denominador de prohibidos para calcular detección.

D. El número de llamadas no demuestra que se cargaran y evaluaran todos los documentos recibidos.

ISO/IEC 42001: 4.3; 9.1; 9.2. Lección 5.

results.py · líneas 40-70

architecture.adoc · líneas 151-160

file.py · líneas 77-81

Pregunta 12 · Fronteras de herramientas y autorización

En el middleware de ejemplo, una llamada a una herramienta de publicación obtiene veredicto favorable para sus argumentos. El usuario carece de permiso para publicar y se propone confiar en el filtro como autorización suficiente. ¿Qué respuesta describe mejor la frontera real?

A. is_safe=True concede al usuario el permiso necesario para esa herramienta porque los argumentos no contienen señales detectadas.

B. El análisis de argumentos no acredita atribuciones; debe comprobarse autorización antes de ejecutar, y el análisis posterior de la salida no revierte la publicación.

C. El middleware solo examina mensajes al modelo y carece de una frontera previa a herramientas.

D. Basta usar tool_output_gate: si la publicación fue indebida, el veredicto posterior la deshará.

Respuesta correcta: B

wrap_tool_call evalúa primero los argumentos y, si el resultado es inseguro, evita llegar al handler por esa vía. Si continúa, ejecuta handler(request) antes de evaluar la salida de la herramienta. Este orden puede impedir una llamada detectada, pero no constituye una comprobación de permisos ni una reversión de efectos externos. La cláusula 8.1 permite contrastar la operación con la política de autorización; 5.3 ayuda a situar atribuciones y responsabilidades. El área A.9 conecta el control con el uso del sistema. La evidencia debe cubrir tanto evaluación de contenido como autorización efectiva del usuario, evitando inferir de una señal favorable una facultad que el contrato no concede. El diseño contempla cuatro fronteras configurables: entrada y salida del modelo, argumentos y resultados de herramientas. Debe documentarse cuáles se han activado, su finalidad y quién responde por cada una. Una prueba de usuario sin permiso con argumentos inocuos permite verificar la autorización sin confundirla con detección de contenido.

Por qué se descartan las otras opciones

A. Confunde ausencia de señales de contenido con autorización de la acción.

C. El ejemplo implementa explícitamente tool_input y tool_output en wrap_tool_call.

D. La salida se examina después de handler(request), por lo que el filtro no deshace el efecto externo.

ISO/IEC 42001: 5.3; 8.1; A.9. Lección 4.

cerbia_middleware.py · líneas 82-105,231-267

Pregunta 13 · Paquetes opcionales y cadena de suministro

El inventario de un servicio solo registra cerbia[cli]. Su memoria afirma que todos los datos pasan por Presidio y ProtectAI porque ambos figuran en el README. ¿Qué comprobación aporta más valor para evaluar esa afirmación?

A. Aceptar la cobertura porque la licencia raíz del proyecto incluye automáticamente todas las integraciones.

B. Concluir que existe una vulnerabilidad confirmada en Presidio por no aparecer en el inventario.

C. Verificar paquetes y versiones instalados, escáneres realmente construidos, modelos y rutas ejecutadas; la disponibilidad opcional no acredita su uso.

D. Suponer que ProtectAI opera mediante un servicio remoto en todas las configuraciones y solicitar solo facturas del proveedor.

Respuesta correcta: C

El README distingue los extras cli, ml, presidio, protectai y all, y presenta ProtectAI como escáner local opcional. Ese catálogo describe posibilidades de composición, no el inventario de una instalación ni su cobertura efectiva. La documentación de construcción añade que Runner importa las clases configuradas y comprueba sus contratos. Para el área A.10 interesa identificar las dependencias elegidas y sus condiciones; A.6 conecta versiones y mantenimiento con el ciclo de vida. Las cláusulas 7.5 y 8.1 sustentan la necesidad de información controlada y comprobación operativa. Sin esas evidencias no procede confirmar la afirmación de la memoria ni inventar una vulnerabilidad o un flujo remoto. Solicitaría el inventario real, configuración construida, artefactos de modelos y revisiones, sus condiciones de uso y pruebas de las rutas en servicio. La persona responsable de dependencias debe gestionar actualizaciones y retorno; validar el esquema o importar una clase tampoco acredita que sus detecciones sean eficaces en el corpus previsto.

Por qué se descartan las otras opciones

A. La licencia del repositorio no instala paquetes ni demuestra que un detector se ejecute.

B. La falta de inventario es una brecha de evidencia; no identifica por sí misma una vulnerabilidad concreta.

D. El README describe una integración local opcional y no acredita el flujo remoto supuesto.

ISO/IEC 42001: 7.5; 8.1; A.6; A.10. Lección 7.

README.md · líneas 30-38,83-89

configuration.adoc · líneas 22-31,99-115

Pregunta 14 · Impactos de falsos bloqueos

Un detector bloquea con frecuencia solicitudes legítimas en un idioma minoritario. La persona responsable propone excluir los falsos bloqueos de la evaluación de impacto porque no son ataques exitosos. ¿Qué enfoque resulta más adecuado?

A. Dar por aceptable cualquier diferencia entre idiomas si el agregado procede de una clase importada correctamente.

B. Elevar automáticamente el umbral hasta eliminar los falsos bloqueos de esa muestra, sin volver a contrastar la detección de prohibidos.

C. Excluirlos de impactos y conservarlos solo como un coste informático del escáner.

D. Evaluar quién queda perjudicado, la interrupción y vías de revisión por segmento, relacionando estos impactos con el tratamiento y la operación.

Respuesta correcta: D

El veredicto depende de acciones, severidades, agregador y umbral; una señal heurística puede detener contenido legítimo en la integración. Ese efecto sobre personas y servicio merece evaluación aunque ningún ataque haya prosperado. Las cláusulas 6.1.4 y 8.4 distinguen la evaluación de impactos y su realización operativa; el área A.5 resulta pertinente para los efectos identificados. Comparar idiomas con referencia revisable, interrupciones y vías de recurso permite valorar distribución y gravedad, además de la tasa global. La construcción válida de una clase no demuestra ausencia de efectos adversos. Tampoco se deduce una obligación universal de eliminar bloqueos: el ajuste requiere evidencia y una decisión coherente con los riesgos del uso. La evidencia incluiría muestras legítimas revisadas por idioma, frecuencia y gravedad de interrupciones, personas afectadas y resolución de reclamaciones. Una persona con atribuciones debe decidir medidas, revisión y seguimiento; comparar solo el promedio global ocultaría grupos con pérdida de acceso aunque la cifra conjunta parezca aceptable.

Por qué se descartan las otras opciones

A. La validez del componente como objeto no demuestra validez o equidad de sus decisiones.

B. Optimiza solo los falsos bloqueos de una muestra y puede aumentar escapes sin evaluar impactos ni eficacia conjunta.

C. Reduce los efectos sobre personas y acceso al servicio a consumo técnico y omite su distribución.

ISO/IEC 42001: 6.1.4; 8.4; A.5. Lección 7.

_gate.py · líneas 212-228

README.md · líneas 24-32

pii.py · líneas 47-53,76-80

Pregunta 15 · Matriz sintética y denominadores de medición

En un ejercicio original y exclusivamente sintético se etiquetan 20 entradas prohibidas y 40 legítimas. Hay 16 prohibidas bloqueadas, 4 prohibidas permitidas, 3 legítimas bloqueadas y 37 legítimas permitidas. ¿Qué informe diferencia correctamente las métricas sin atribuir rendimiento empírico a CerbIA?

A. Detección 16/20=80 %; falsos bloqueos 3/40=7,5 %; exactitud 53/60≈88,3 %; precisión de bloqueos 16/19≈84,2 %, solo para este supuesto sintético.

B. Detección 88,3 % y precisión 80 %; las dos usan las 60 entradas como población equivalente.

C. Detección 84,2 %, falsos bloqueos 5 % y exactitud 80 %; son resultados de producción del proyecto.

D. Detección 80 %, falsos bloqueos 15 % y precisión 88,3 %; basta informar estas tasas sin sus poblaciones.

Respuesta correcta: A

La detección pregunta por los prohibidos y usa 20; los falsos bloqueos preguntan por los legítimos y usan 40. La exactitud reúne las 53 decisiones correctas entre 60, mientras la precisión examina los 19 bloqueos y cuántos eran prohibidos. La cláusula 9.1 requiere mantener visibles estas poblaciones y el método de etiquetado. El contrato GateVerdict aporta decisiones y hallazgos, pero no una verdad de referencia ni estos porcentajes por sí mismo. La cláusula 7.5 ayuda a conservar el origen y límite del ejercicio. Las cifras son didácticas: ni el código revisado ni este supuesto equivalen a una evaluación de producción de CerbIA. La lectura de SGIA añade los cuatro prohibidos que escaparon y los tres usuarios legítimos detenidos a la valoración del riesgo y del impacto. Conservaría etiquetas revisables y matriz de confusión asociadas a la configuración. El responsable decidirá si los errores de esta muestra satisfacen sus criterios, sin convertir una tasa favorable en garantía.

Por qué se descartan las otras opciones

B. Intercambia exactitud y detección y asigna denominadores que no responden a las preguntas de cada métrica.

C. Confunde precisión con detección, calcula falsos bloqueos sobre el total y atribuye datos sintéticos a producción.

D. Usa 20 como denominador de legítimos y confunde exactitud con precisión; omitir poblaciones impide interpretar las tasas.

ISO/IEC 42001: 7.5; 9.1. Lección 5.

_models.py · líneas 6-22

Pregunta 16 · Fatiga, latencia y recalibración continua

La cola de WARN aumenta y la latencia p95 empeora. El equipo propone elevar el umbral y cambiar BLOCK a SKIP ante errores sin conservar una línea base. ¿Qué actuación permite una mejora verificable del control?

A. Activar SKIP silenciosamente hasta que mejore p95 y reconstruir la justificación si aparece un incidente.

B. Muestrear únicamente bloqueos, pues los textos permitidos no aportan información sobre errores del control.

C. Muestrear alertas y permitidos, medir cobertura, escapes, falsos bloqueos y revisión, probar una versión candidata con regresión y decidir su liberación y retorno con responsable.

D. Aprobar el ajuste cuando baje la cola, aunque disminuyan las ejecuciones de escáneres y no se mida detección.

Respuesta correcta: C

La ejecución secuencial y fail_fast hacen que el tiempo dependa del trabajo realmente realizado. Además, WARN no participa en el agregado BLOCK, de modo que elevar el umbral no es una solución demostrada para su cola. SKIP puede permitir resultados favorables con análisis omitido. Las cláusulas 9.1 y 9.3 conectan mediciones de eficacia, carga y recursos con decisiones responsables; 10.2 se aplica cuando hay una no conformidad acreditada que exige corregir causas. Una recalibración defendible incluye permitidos para buscar escapes, versiones comparables, criterios previos y retorno. El área A.6 encuadra la validación del cambio técnico; menos alertas o menor latencia aisladas no prueban mejor protección. Conservaría hardware, longitud del corpus, calentamiento, concurrencia y tiempos de puerta y servicio, además de espera y esfuerzo humano. La mejora de 10.1 se demuestra comparando resultados completos; un ajuste necesita autorización, vigilancia y retorno. Sin incumplimiento acreditado, no corresponde presentar automáticamente la recalibración como acción correctiva.

Por qué se descartan las otras opciones

A. La omisión silenciosa elimina trazabilidad y pospone la decisión sobre el riesgo hasta después del daño.

B. Muestrear solo bloqueos impide descubrir prohibidos que el control permitió y limita la estimación de escapes.

D. La cola puede bajar por pérdida de cobertura; falta comprobar la eficacia y el riesgo residual.

ISO/IEC 42001: 9.1; 9.3; 10.2; A.6. Lección 6.

_gate.py · líneas 84-114,143-165,183-228

Pregunta 17 · Auditoría de cobertura con evidencia incompleta

Caso original de preparación. Una organización afirma que ha examinado los 500 expedientes de una ventana porque su panel CerbIA muestra resultado favorable y cero hallazgos. Solo entrega el README y una captura. Desconoce cuántas entradas devolvieron los cargadores y un preprocesador externo puede excluir expedientes. No se aporta aún el alcance ni el criterio interno de cobertura. Formula una conclusión provisional de auditoría y un plan de evidencia que permita resolverla sin inventar incumplimientos.

Respuesta razonada

La conclusión provisional es que no hay evidencia suficiente de cobertura de los 500 expedientes. El README describe posibilidades del componente, y la captura no identifica población ni configuración. Runner carga y preprocesa antes de evaluar cada entrada final; las extensiones pueden eliminar o expandir entradas y ScanResult vacío también resulta favorable. No puede deducirse ni evaluación completa ni una fuga consumada. Primero obtendría el alcance de 4.3, la finalidad, los intercambios incluidos y el criterio interno aplicable. Solicitaría inventario recibido, salidas por cargador y preprocesador, motivos de exclusión, versiones, configuración y resultados vinculados por identificador. Conciliaría esas unidades y probaría rutas vacías o excluidas. Conforme a 8.1 y 9.1, el responsable debe definir qué ocurre si falta contenido y cómo se mide cobertura. Para 9.2, la calificación depende de evidencia frente a un criterio concreto: si se demuestra su incumplimiento se documentará el hallazgo; mientras tanto procede mantener abierta la comprobación. La decisión y sus límites se conservan conforme a 7.5.

Rúbrica de razonamiento

Nivel 1 · Técnico: Identifica que resultado favorable y cero hallazgos no demuestran cobertura; solicita algún recuento, pero no delimita alcance ni unidades.

Nivel 2 · Cumplimiento: Añade alcance y criterio, concilia recibidos, cargados, derivados, excluidos y evaluados, y explica la posibilidad de resultado vacío.

Nivel 3 · Auditoría: Integra trazabilidad versionada, responsable y respuesta ante cobertura incompleta; formula una conclusión provisional y condiciona cualquier no conformidad a un criterio y evidencia verificados.

ISO/IEC 42001: 4.3; 7.5; 8.1; 9.1; 9.2. Lección 7.

runner.py · líneas 76-105

architecture.adoc · líneas 87-99,151-160

results.py · líneas 40-70

Pregunta 18 · Interpretación y límites de un ensayo sintético

Caso sintético original, sin mediciones del runtime. En 60 entradas etiquetadas hay 20 prohibidas y 40 legítimas. El filtro bloquea 16 prohibidas y 3 legítimas; permite las otras 4 prohibidas y 37 legítimas. La dirección recibe el titular «88,3 % de precisión y seguridad garantizada» y propone subir el umbral para eliminar los tres falsos bloqueos. Corrige el informe y define qué comprobar antes de aceptar ese cambio.

Respuesta razonada

Corregiría el titular: detección 16/20=80 %, falsos bloqueos 3/40=7,5 %, exactitud 53/60≈88,3 % y precisión de los bloqueos 16/19≈84,2 %. Son datos sintéticos del ejercicio; no miden el rendimiento observado de CerbIA ni garantizan seguridad. Los cuatro prohibidos permitidos representan escapes y los tres legítimos bloqueados, interrupciones que también deben valorarse. GateVerdict proporciona resultados, sin etiquetas externas de referencia. El código permite pasar si el agregado BLOCK queda por debajo del umbral; elevarlo puede reducir falsos bloqueos y aumentar escapes, sin que esta matriz permita predecir cuántos. Para 9.1 conservaría corpus, etiquetas revisadas, denominadores, configuración y distribuciones de puntuación. Conforme a 6.1.2 y 8.2 reevaluaría el riesgo residual; 6.1.4 y 8.4 permiten considerar los efectos de las interrupciones. Antes de aplicar el ajuste, compararía base y candidato sobre muestras equivalentes y segmentos pertinentes, con criterios previos, medición de omisiones y decisión autorizada sobre el tratamiento de 8.3. Prevería vigilancia y retorno.

Rúbrica de razonamiento

Nivel 1 · Técnico: Reconoce que el ejemplo es sintético e identifica algún error del titular, pero confunde tasas o no expresa sus denominadores.

Nivel 2 · Cumplimiento: Calcula las cuatro tasas correctamente, identifica escapes e interrupciones y explica el sentido del umbral sin prometer eliminar errores.

Nivel 3 · Auditoría: Además propone comparación etiquetada, criterios previos por riesgo e impacto, configuración trazable, responsable, vigilancia y retorno; rechaza la garantía y la atribución de rendimiento de producción.

ISO/IEC 42001: 6.1.2; 6.1.4; 8.2; 8.3; 8.4; 9.1. Lección 5.

_gate.py · líneas 212-228

_models.py · líneas 6-22

Pregunta 19 · Integración efectiva, permisos y efectos externos

Caso original de auditoría. Una aplicación utiliza CerbiaAgentMiddleware con raise_error=False, tool_gate y tool_output_gate. El procedimiento aprobado exige comprobar permisos antes de publicar y detener llamadas con veredicto inseguro. Una prueba muestra que una publicación de un usuario sin atribuciones se ejecuta con argumentos favorables; después tool_output_gate devuelve inseguro y se presenta un mensaje de error. Diseña la evaluación del hallazgo y las pruebas necesarias para distinguir autorización, contenido y efecto protector.

Respuesta razonada

El mensaje final de error no demuestra que se evitara la publicación. El middleware evalúa argumentos antes del handler; si son inseguros no lo ejecuta, incluso con raise_error=False. Cuando son favorables, ejecuta la herramienta antes de inspeccionar su salida, de modo que esta frontera posterior no revierte el efecto externo. Tampoco una señal favorable concede permisos. Las cuatro fronteras configurables cubren entrada y salida del modelo y entrada y salida de herramientas, con finalidades diferentes. Contrastaría atribuciones del usuario, procedimiento aprobado, configuración y trazas de la publicación. Si se acredita la ausencia o elusión de la autorización previa exigida, documentaría el incumplimiento respecto a 8.1 y las responsabilidades de 5.3, sin atribuirlo simplemente a raise_error=False. Probaría por separado usuario sin permiso con contenido inocuo, usuario autorizado con contenido inseguro, fallo de escáner y rutas síncronas y asíncronas relevantes. El área A.9 sitúa las condiciones de uso del sistema. Si procede 10.2, la corrección debe tratar causas y demostrar después que ninguna publicación protegida elude los controles requeridos.

Rúbrica de razonamiento

Nivel 1 · Técnico: Distingue contenido favorable de permiso, pero no explica el orden de ejecución ni solicita evidencia del efecto externo.

Nivel 2 · Cumplimiento: Explica entrada-handler-salida, reconoce que la inspección posterior no revierte la publicación y contrasta el procedimiento con las trazas.

Nivel 3 · Auditoría: Incluye las cuatro fronteras y responsabilidades, interpreta correctamente raise_error=False y propone pruebas separadas de permisos, contenido, errores y rutas; condiciona el hallazgo y verifica la eficacia correctiva.

ISO/IEC 42001: 5.3; 8.1; 10.2; A.9. Lección 4.

cerbia_middleware.py · líneas 82-105,109-138,231-267,285-335

Pregunta 20 · Recalibración responsable y capacidad operativa

Caso original de mejora. Tras actualizar un detector opcional local, crecen las WARN y la revisión se retrasa. Algunos operadores quieren cambiar a SKIP ante errores y elevar el umbral sin conservar una línea base; además proponen enviar prompts completos y explicaciones a registros sin plazo de conservación. No se ha acreditado incumplimiento de un plazo aprobado. Diseña una recalibración que considere eficacia, impactos, p95, carga humana, privacidad de registros y responsabilidades; precisa cuándo procede una acción correctiva o revisar el propio SGIA.

Respuesta razonada

Abriría una investigación con responsable, versiones y configuración de referencia, sin presuponer una no conformidad. El código excluye WARN del agregado y SKIP registra evaluaciones omitidas; subir el umbral no resuelve por sí mismo la cola y omitir controles puede ocultar exposición. Muestrearía alertas y permitidos para medir escapes, falsos bloqueos y omisiones por idioma y uso. Compararía p50/p95 de puerta y servicio bajo hardware, carga, longitud y calentamiento identificados, separando arranque, y mediría espera, volumen pendiente y esfuerzo humano. Conforme a 7.5 y 9.1 conservaría evidencia mínima recuperable: el anfitrión elige registros y destinos, y las explicaciones pueden contener fragmentos sensibles. Definiría acceso, minimización y conservación; redact=True en KeywordScanner oculta fragmentos en la explicación, no anonimiza el prompt. Evaluaría impactos según 8.4 y probaría un candidato con criterios previos, autorización, vigilancia y retorno. 9.3 permite decidir recursos; 10.1 orienta la mejora. Aplicaría 10.2 si se acredita incumplimiento y trataría causas y eficacia. Revisaría además si el cambio técnico requiere modificar el propio SGIA, sin equipararlos automáticamente.

Rúbrica de razonamiento

Nivel 1 · Técnico: Propone medir latencia o atender alertas, pero omite permitidos, cobertura, versiones, privacidad o responsabilidad de aceptación.

Nivel 2 · Cumplimiento: Incluye comparación base-candidato, eficacia y omisiones, p95 con condiciones, carga humana y controles de acceso y conservación de los registros.

Nivel 3 · Auditoría: Integra impactos, dependencias, minimización y límites de redact=True, criterios previos, autoridad, recursos, vigilancia y retorno; distingue mejora, cambio del SGIA y acción correctiva fundada.

ISO/IEC 42001: 5.3; 7.5; 8.1; 8.4; 9.1; 9.3; 10.1; 10.2; A.6; A.7; A.10. Lección 6.

_gate.py · líneas 143-165,212-228

README.md · líneas 83-89

logging.adoc · líneas 9-22,25-42

_scanner.py · líneas 110-120

Cierre del monográfico

Una decisión que puede explicarse

El aprendizaje del caso consiste en unir funcionamiento, riesgo y evidencia. Sabemos distinguir el recorrido completo de Runner de la evaluación por texto de SecurityGate; una señal del cálculo agregado; una omisión de un análisis terminado; y un resultado inseguro de una operación efectivamente impedida. Esas distinciones evitan que una decisión de gestión descanse únicamente en un booleano.

La incorporación de CerbIA debe responder a una finalidad definida y a escenarios de riesgo del sistema que lo utiliza. El expediente propuesto conecta alcance, responsables, configuración, cobertura, eficacia, impactos, operación y cambios. La revisión documental del proyecto aporta conocimiento sobre mecanismos; las conclusiones sobre una instalación necesitan además evidencia de esa instalación.

Repartir el trabajo y la autoridad

La siguiente RACI es una propuesta adaptable. R realiza el trabajo; A tiene autoridad y responde por la decisión; C aporta criterio; I recibe información. Los roles pueden combinarse si se preservan las responsabilidades y la independencia pertinente.

Decisión o tarea	R	A	C	I
Definir finalidad y alcance	Responsable del sistema	Autoridad del SGIA	Usuarios, datos y seguridad	Operación
Integrar y comprobar el efecto	Equipo técnico	Responsable del sistema	Seguridad y operación	Autoridad del riesgo
Evaluar riesgos e impactos	Equipo designado	Autoridad del riesgo	Datos, usuarios y negocio	Dirección
Aprobar uso y excepciones significativas	Responsable del sistema	Autoridad designada	Seguridad, datos y operación	Personas afectadas cuando proceda
Seguir resultados y proponer ajustes	Operación	Responsable del sistema	Equipo técnico y riesgo	Dirección
Revisar evidencia del proceso	Revisión independiente	Responsable de la función de revisión	Áreas examinadas	Dirección

No se exige que estos sean los nombres de la organización. Deben quedar claras las atribuciones, la competencia y la vía de escalado. La supervisión humana se diseña alrededor de decisiones significativas y capacidad de intervención.

Anexo breve: arquitectura conjunta hipotética

Como conexión con el curso, imaginemos una aplicación que consulta información mediante un componente de búsqueda estudiado en Kumoss y podría incorporar MCP Teams Server, previsto en el itinerario avanzado del curso. Situaríamos evaluaciones de contenido en las fronteras pertinentes y permisos propios en las herramientas.

El recorrido hipotético sería: consulta del usuario → evaluación de entrada → recuperación y revisión del contenido externo → modelo → comprobación de argumentos y autorización de herramienta → ejecución → revisión del resultado → evaluación antes de entregar la respuesta. Cada flecha exige decidir qué texto se observa y qué efecto todavía puede impedirse.

Esta es una propuesta docente para comparar responsabilidades entre proyectos. No describe una conexión implementada entre ellos ni una implantación de Inditex. La búsqueda, la evaluación de contenido y la autorización de operaciones cumplen funciones distintas.

Evidencia para mantener la decisión

Antes de autorizar, preguntamos si se conoce el recorrido, si la aplicación actúa y si las pruebas sostienen el criterio de aceptación. Después observamos omisiones, escapes, falsos bloqueos, tiempos y efectos sobre usuarios. Revisamos al cambiar finalidad, contenidos, dependencias o condiciones de uso.

La matriz que acompaña el módulo ayuda a relacionar aportes y brechas con las referencias de gestión. Sus correspondencias son razonamientos de aplicación, no equivalencias automáticas de cumplimiento. Una decisión responsable puede autorizar un uso limitado, exigir medidas adicionales o aplazarlo hasta disponer de evidencia suficiente.

Síntesis para la dirección

Ficha para la dirección: qué decisión permite sostener la evidencia

CerbIA aporta evaluaciones configurables de contenido. Su incorporación debe vincularse con riesgos concretos y con la actuación de la aplicación. La dirección necesita conocer la cobertura, las limitaciones, los recursos necesarios y las condiciones para mantener o restringir el uso.

Estado de la evidencia	Lectura para la decisión	Actuación propuesta
Verde	Se aportan pruebas suficientes frente a criterios definidos para un uso y versión acotados.	Mantener el uso autorizado y vigilar sus condiciones.
Ámbar	Hay incertidumbre, cobertura degradada o resultados pendientes relevantes.	Acotar el uso y resolver las evidencias pendientes con responsable y plazo justificados.
Rojo	Se demuestra un incumplimiento de un criterio crítico o el control deja pasar una acción que debía impedir.	Contener o suspender la operación afectada y tratar la causa antes de reanudarla.

Los colores describen estados de evidencia en este modelo docente; no son un riesgo medido del producto. Una media de resultados favorables no compensa una brecha crítica. La decisión debe identificar el uso, la versión, la autoridad, las restricciones y la condición de revisión.

Fuentes y revisión del estudio

Proyecto: InditexTech/cerbia.

Commit analizado: 86d063106a8b5b4ffc0e1244371a948a01b75204. El paquete conserva los archivos de referencia con sus avisos y licencias. Se han leído código y documentación; no se ha ejecutado el producto ni medido una implantación. Las correspondencias con el SGIA son elaboración docente.

ISO/IEC 42001 · Editor

Índice de ISO/IEC 42001 · IEC

Áreas de control · BSI, página 6