MONOGRÁFICO 5

CerbIA

Preguntas y respuestas razonadas

Dieciséis preguntas de respuesta única y cuatro casos de aplicación de ISO/IEC 42001 al proyecto CerbIA.

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.

Preguntas y respuestas

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