MONOGRÁFICO 5 | ACTIVIDAD DE CONOCIMIENTOS

Ficha de calibración de frontera de protección

SOLUCIÓN ORIENTATIVA

Fundamentos del SGIA e ISO/IEC 42001

Identificador didáctico C5-E01 | Edición 2026-10-26

Validar la interpretación de errores, cobertura, agregación, métricas y actuación del anfitrión.

Responde a las nueve preguntas y a los cuatro ejercicios de ampliación con los hechos aportados. Justifica la conclusión y señala la evidencia pendiente. No inventes datos ni ejecutes los productos.

Estas respuestas son una referencia de corrección. Admiten formulaciones equivalentes si conservan las distinciones y límites del caso.

Hechos aportados

1. Caso didáctico C5-E01. No son resultados de una ejecución de CerbIA. El uso propuesto protege argumentos de una herramienta con efectos externos.

2. La configuración hipotética utiliza política de error SKIP, umbral 0,5 y corte anticipado. Una entrada obtiene is_safe=True con un escáner omitido por error. No consta una excepción o degradación aprobada ni evidencia de actuación del anfitrión.

3. Un conjunto sintético independiente contiene 20 textos prohibidos y 40 legítimos. Se bloquean 16 prohibidos y 3 legítimos; pasan 4 prohibidos y 37 legítimos.

4. No se han aportado medidas de latencia, coste de revisión, versiones completas ni calendario justificado de reevaluación.

Referencias de estudio

CerbIA lecciones 2, 3, 5, 6 y 7; Módulos 6, 9, 10 y 11 del curso.

Modelo didáctico de la forma de evidencia. No es un formulario obligatorio de ISO/IEC 42001 ni una firma electrónica verificada.

Preguntas y respuestas

Solución orientativa del caso C5-E01

C01 Frontera y objetivo

Define qué se quiere proteger y qué queda fuera de esa frontera.

Se evalúan argumentos antes de llamar a una herramienta. El anfitrión debe hacer cumplir el resultado antes del efecto. La evaluación de texto no sustituye permisos ni autorización, y una inspección posterior no revierte una acción ya realizada.

C02 Política ante fallos

Justifica una política para este supuesto y cómo tratarías la alternativa.

No se acepta el paso inadvertido de contenido no evaluado. Para este uso puede justificarse BLOCK con actuación efectiva del anfitrión, o FAIL con la excepción capturada e impedimento de la llamada. SKIP solo sería defendible con cobertura degradada conocida, criterio explícito, justificación y decisión pertinente. No es una prohibición universal de SKIP.

C03 Umbral y agregador

Qué significa 0,5 y qué falta para justificarlo.

Es un parámetro de decisión técnica, no una probabilidad calibrada ni un umbral normativo. Se deben documentar agregador, escáneres, severidades, acciones, población de prueba y criterio de aceptación. Cambiar la estrategia o el orden con fail_fast puede cambiar recorrido y resultado.

Preguntas y respuestas

Solución orientativa del caso C5-E01

C04 Excepciones y cobertura

Puede considerarse plenamente evaluada la entrada favorable.

No. Un escáner fue omitido por error y falta la decisión que justifique operar con esa cobertura. Deben registrarse motivo, contenido afectado, controles ejecutados y omitidos, restricciones, responsable y vigencia de cualquier excepción. No se inventa una excepción aprobada.

C05 Detección de prohibidos

Calcula la tasa y declara el denominador.

16 / 20 = 80 %. El denominador son todos los textos prohibidos del conjunto sintético.

C06 Falsos bloqueos

Calcula la tasa sobre contenido legítimo.

3 / 40 = 7,5 %. El denominador son todos los textos legítimos, no los bloqueados.

Preguntas y respuestas

Solución orientativa del caso C5-E01

C07 Exactitud y precisión

Calcula ambas y explica su diferencia.

Exactitud global: (16 + 37) / 60 = 88,3 % aproximadamente. Precisión de los bloqueos: 16 / 19 = 84,2 % aproximadamente. Responden a preguntas diferentes; la exactitud global no borra los falsos negativos ni los falsos bloqueos.

C08 Matriz de pruebas periódicas

Qué categorías y disparadores registrarías sin inventar una periodicidad normativa.

Casos positivos y negativos por amenaza, formato e idioma pertinente; carga y omisiones, errores FAIL/BLOCK/SKIP, corte anticipado, actuación real del anfitrión y regresión tras cambios. Para cada prueba: entrada, resultado esperado, observado, versión y evidencia. La frecuencia y los disparadores se justifican por uso, riesgos, cambios e incidentes, no por un mínimo universal inventado. La ampliación desarrolla muestreo, desvío de reglas, cambios, regresión y retorno.

C09 Latencia y revisión humana

Qué datos faltan y dónde interviene el responsable.

Faltan latencia del control y del servicio, recursos y trabajo de atender WARN y falsos positivos. El responsable aprueba uso, política y excepciones y revisa eficacia. No se requiere revisión manual universal de cada texto favorable cuando el proceso aprobado permite automatizarlo. La carga sostenida debe disparar revisión y recalibración controlada, no desactivación silenciosa.

Recalibración continua y desvío de reglas

El equipo plantea reducir alertas relajando reglas. No hay nueva muestra etiquetada, mediciones operativas ni candidato validado. Los datos sintéticos iniciales se mantienen; no son una línea base de producción.

C10 Muestreo y fatiga de alertas

Cómo comprobarías si hay fatiga y desvío de reglas.

Definir ventana y población; muestrear bloqueos, WARN y contenido permitido con etiquetado revisable y minimización de datos. Separar falsos positivos, falsos negativos, omisiones y errores. Medir volumen, cola, tiempo y coste de revisión, latencia p50/p95 y uso de bypass. Comparar versiones y segmentos. Frecuencia, límites y responsables deben justificarse; no inventar cifras.

C11 Cambio controlado

Qué debe contener una propuesta de recalibración.

Ticket con causa, versión de configuración y diferencias en escáneres, preprocesadores, acciones, severidades, orden, fail_fast, on_error, umbral y agregador. Mantener la versión anterior y un conjunto de regresión etiquetado. Modificar un parámetro aislado no demuestra seguridad; SKIP o desactivar escáneres cambia cobertura.

Recalibración continua y desvío de reglas

El equipo plantea reducir alertas relajando reglas. No hay nueva muestra etiquetada, mediciones operativas ni candidato validado. Los datos sintéticos iniciales se mantienen; no son una línea base de producción.

C12 Aceptación y liberación

Cómo decidirías si una configuración candidata puede usarse.

Comparar detección, falsos bloqueos, escapes y cobertura con denominadores separados, además de latencia y carga de revisión. Fijar criterios antes de evaluar y probar actuación del anfitrión ante fallos. El responsable con atribuciones decide; liberar de forma acotada y vigilada con versión, evidencia y reversión definida. No relajar controles solo para vaciar la cola.

C13 Umbral, excepciones y retorno

Qué riesgo tiene subir el umbral y cómo cerrarías el ciclo.

Puede reducir bloqueos, pero permitir hallazgos BLOCK con puntuación positiva que antes se detenían; no representa una mejora automática. Mantener excepciones acotadas y con caducidad. Ante deterioro, volver a configuración conocida, impedir la acción protegida o degradar explícitamente según política, y registrar incidente/cambio. Aquí no hay datos para aceptar una nueva configuración.

Ampliación de las respuestas

Las respuestas desarrolladas figuran bajo cada pregunta. No se completan datos de una instalación ni se emite una autorización real.

Criterios de evaluación

C01 Identifica la frontera previa a la acción y sus límites.

C02 Justifica el tratamiento según riesgo; no exige siempre un único enum.

C03 Relaciona umbral, agregación y evidencia.

C04 Lee is_safe junto con omisiones y errores.

C05 Calcula detección y población correcta.

C06 No solapa denominadores.

C07 Distingue las dos métricas y sus límites.

C08 Relaciona pruebas, criterios y revisión basada en contexto.

C09 No atribuye rendimiento ni una barrera humana por texto que no están demostrados.

C10 Evita sesgo de observar solo alertas y mide carga operativa.

C11 Abarca la configuración efectiva y sus dependencias.

C12 Usa criterios previos, regresión y aprobación de cambios.

C13 Conserva seguridad, reversibilidad y evidencia pendiente.

Rúbrica común

Correcta: conclusión proporcionada, razonamiento y evidencia bien identificados.

Parcial: conclusión razonable con una omisión relevante.

A revisar: confunde capacidad y autorización, inventa datos o usa denominadores incorrectos.

Los criterios valoran la respuesta del alumno. No acreditan cumplimiento de una instalación ni completan evidencia ausente.