

Continuidad de la investigación bajo presión

Memoria externa y control de la autoridad · Anna Tchijova

Reportado ≠ verificado

Cuando alguien dice «ya está», queda registrada una afirmación. Incluso un archivo intacto requiere verificación antes de exportarse.

> **Conservá la evidencia. Mantené el hilo.**



ILUSTRACIÓN EDITORIAL

TaskResponsibility · demo/run_case_001.py

Vuelo 2574 de Continental Express · 1991

47 tornillos. 14 vidas.

El mantenimiento pasó de un turno a otro. Tras retirar 47 tornillos, una pieza del estabilizador quedó sin fijar. El avión se desintegró en vuelo.

> **El relevo debe transmitir el estado real del trabajo pendiente.**



ILUSTRACIÓN EDITORIAL

FAA · NTSB AAR-92/04

Mar del Norte · 1988

167 personas murieron en Piper Alpha

Se reinició una bomba con su válvula de seguridad fuera de servicio. El estado del trabajo sobre la válvula no llegó a quienes decidieron reiniciarla.

> **Los trabajos sobre un mismo recurso deben estar a la vista antes de actuar.**



ILUSTRACIÓN EDITORIAL

Investigación Cullen · HSE

Caso Liam Allan · 2017

Se dio por revisada una descarga de 57.000 líneas

La descarga del teléfono contenía más de 57.000 líneas de mensajes. El juicio se cayó al aparecer material que no se había revelado a la defensa.

> **Una revisión declarada necesita una verificación identificable.**

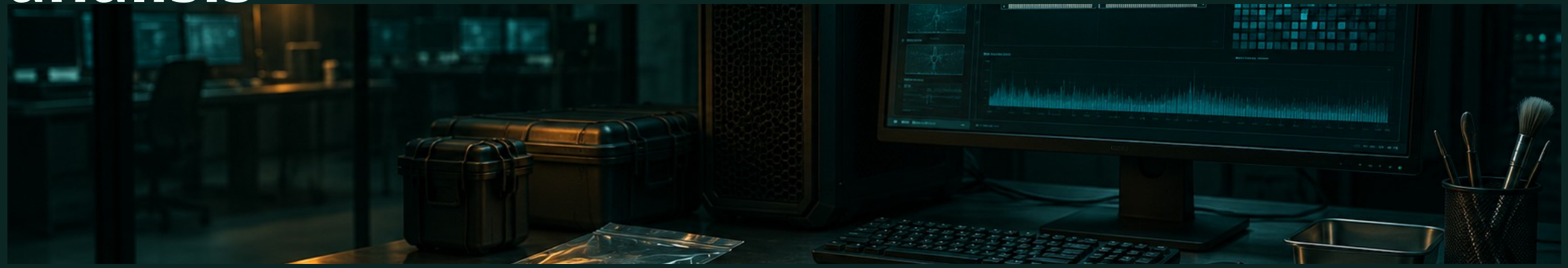


ILUSTRACIÓN EDITORIAL

Revisión conjunta de la fiscalía y la Policía Metropolitana · 2018

INVESTIGACIÓN SIMULADA

La evidencia puede desaparecer antes del análisis



01
A informa «ya está»

›

02
B da por hecha la captura

›

03
La fuente queda fuera de línea

›

04
El análisis se abstiene con razón

La última oportunidad de actuar llega antes del análisis.

Un mismo resultado puede ocultar dos historias

Inevitable

La evidencia nunca existió, era inaccesible o seguía siendo insuficiente tras una investigación bien realizada.

Potencialmente evitable

La evidencia era necesaria, se podía recuperar en un plazo conocido o estimable y había una vía autorizada en ese momento.

↓ ↓

MISMO RESULTADO FINAL / INCONCLUSO

El registro de lo que se sabía en ese momento permite distinguir ambos casos.

Storm-0558 · El período sin registros

No se pudo establecer cómo se obtuvo la clave de firma. Los registros del período relevante ya no estaban disponibles.

> **Una ventana de recuperación ayuda a actuar sobre fuentes ya registradas en el caso.**

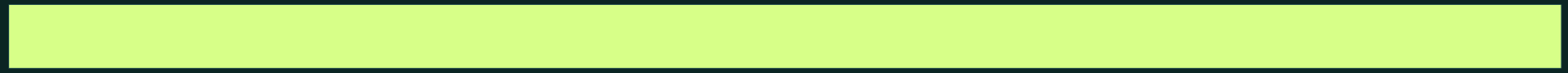


ILUSTRACIÓN EDITORIAL

CSRB · Revisión de la intrusión en Microsoft Exchange Online

393 días de intrusión. 90 días de registros.

Permanencia promedio



393 días

Retención estándar



90 días

Según Mandiant, la retención limitada impidió identificar el acceso inicial en los casos descritos.

Mandiant · M-Trends 2026

¿Quién necesita continuidad entre turnos?

Respuesta a incidentes

Una fuente volátil puede desaparecer durante un turno largo.

Periodismo de investigación

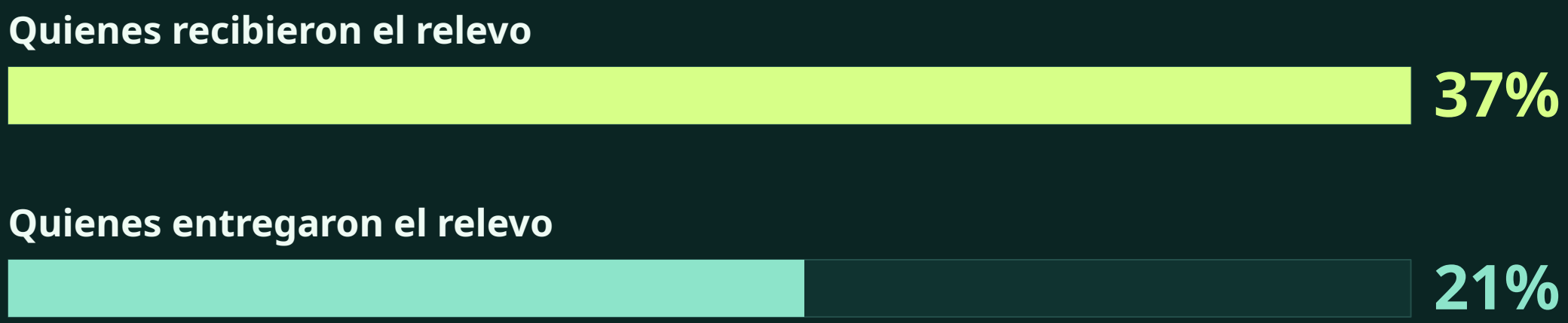
Una publicación o cuenta puede desaparecer mientras el equipo verifica su procedencia.

Investigación de seguridad

Al volver después de varios días, el equipo necesita saber qué venció, qué cambió y quién puede actuar.

Encuesta SANS 2025 · 79 % de los SOC encuestados operaban 24/7

Quien entrega y quien recibe ven relevos distintos



Porcentaje de relevos considerados fallidos en el proyecto citado. La alerta da pocos detalles sobre la muestra.

Joint Commission · Alerta 58

Los relevos estructurados redujeron el daño evitable



Reducciones relativas del 23 % y 30 %, respectivamente, en un programa hospitalario de relevos.

Starmer y cols. · NEJM · 2014

Una ventana puede cerrarse por plazo o por evento

Target: Stroke

El tratamiento dentro de los 60 minutos pasó del 26,5 % al 41,3 %. La mortalidad hospitalaria bajó del 9,93 % al 8,25 %.

Evidencia volátil

Un reinicio puede cerrar la ventana para capturar RAM en cualquier momento. La oportunidad y la autorización tienen que coincidir.

Fonarow y cols. · JAMA · 2014

La verificación tiene que resistir la presión

Knight Capital

Uno de ocho servidores quedó sin actualizar. La SEC informó pérdidas superiores a USD 460 millones tras unos 45 minutos de operaciones erróneas.

Texas City

Murieron 15 trabajadores. La CSB identificó deficiencias en los relevos entre fallas organizacionales más amplias.

Mars Climate Orbiter

Una interfaz transmitió datos de impulso en unidades incompatibles. La NASA identificó la discrepancia como causa raíz.

SEC · CSB · NASA

Las interrupciones y los supuestos se acumulan

Interrupciones

Cada interrupción se asoció con un 12,1 % más de fallas de procedimiento al administrar medicamentos.

Lista quirúrgica

En un estudio antes y después realizado en ocho hospitales, la mortalidad hospitalaria bajó del 1,5 % al 0,8 %.

Criterio forense

El caso Mayfield ilustra el riesgo de anclaje. KRAKADYL registra el estado de la evidencia y su revisión. Detectar sesgos queda fuera de su alcance.

Westbrook y cols. · Haynes y cols. · DOJ OIG

Ocho fallas antes de empezar el análisis

FALLA	CONTROL EXISTENTE	MECANISMO DE KRAKADYL
Ventana perdida	Target: Stroke · 26,5 % → 41,3 % en 60 min	Registra ventanas con plazo o cierre por evento
Tarea declarada hecha	Inspección independiente en aviación	Exige verificación independiente del requisito
Relevo incompleto	I-PASS · 30 % menos eventos adversos evitables	Vincula la transferencia al estado registrado
Información sin vincular	Errores en el historial de medicación · 10–67 %	Muestra conflictos sobre un mismo recurso

Ocho fallas antes de empezar el análisis

FALLA	CONTROL EXISTENTE	MECANISMO DE KRAKADYL
Autoridad vencida	Pausas de seguridad y bloqueo	Revalida mandato y evidencia al ejecutar
Trabajo interrumpido	Chalecos «No interrumpir» · 48 % apoyó adoptarlos	Muestra cambios y pendientes al retomar
Supuesto sin verificar	Controles de sesgo forense	Reabre una hipótesis si cambia su evidencia decisiva
Ventana sin autoridad	RAM · equipo encendido, especialista y autorización	Expone la oportunidad y el permiso por separado

ARQUITECTURA

Una propuesta de IA pasa por un control determinista

01
Muse propone una acción tipada

›

02
El entorno aporta hechos verificados

›

03
La política revisa alcance y evidencia

›

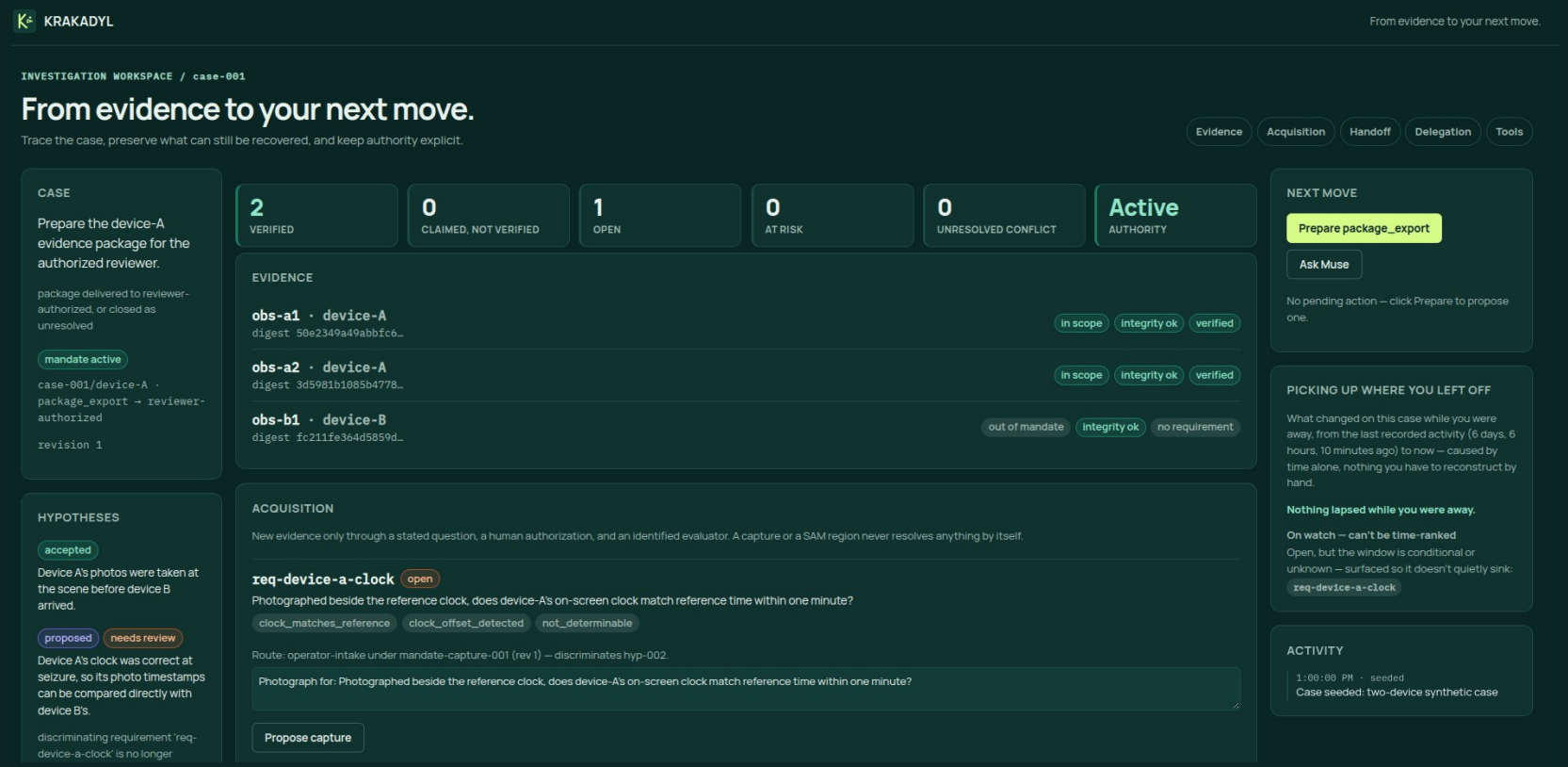
04
El ejecutor vuelve a verificar y registra

Laya estructura señales. SAM 3.1 localiza regiones. La autorización humana y la evaluación independiente rigen su uso.

PRODUCTO · CASO 001

El espacio de trabajo muestra el estado del caso

Estado de la evidencia, preguntas de adquisición y autoridad vigente en una misma vista.



01
Estado de verificación

02
Pregunta de adquisición abierta

03
Mandato vigente

La recuperación empieza con una pregunta

¿Qué se puede recuperar?

Registrá ventanas fijas, condicionales y desconocidas. Mantené visible la incertidumbre.

¿Qué separa las hipótesis?

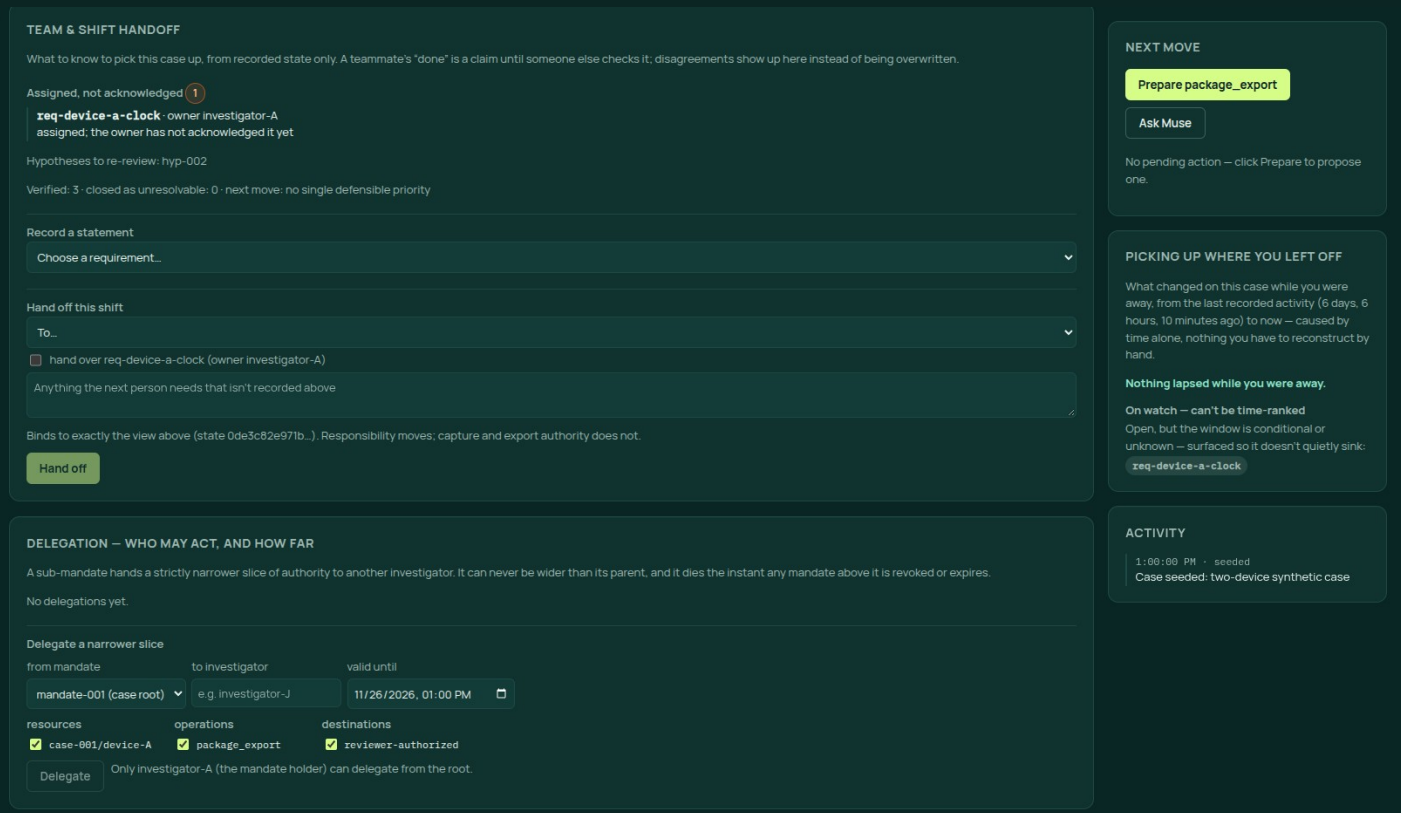
Definí la pregunta y los resultados posibles antes de recolectar material.

¿Quién evalúa?

Una persona autoriza la captura. Otra puede resolver el requisito mediante una evaluación independiente. SAM marca regiones para revisar.

El relevo transfiere responsabilidad

La vista de salida queda ligada al hash del estado. Quien recibe ve los pendientes. La autoridad conserva su propio alcance.



01
Tarea sin acuse de recibo

02
Relevo ligado al estado

03
Delegación acotada

NIVEL 6 · IMPLEMENTADO

La delegación termina cuando cae su autoridad

RAÍZ

Autoridad sobre el caso

›

DELEGACIÓN 1

Equipo A · operaciones limitadas

›

DELEGACIÓN 2

Equipo A · solo exportación

RAÍZ REVOCADA → SE BLOQUEAN AMBAS DELEGACIONES

Si se revoca la raíz después de preparar la acción, el ejecutor vuelve a recorrer la cadena y rechaza la acción. El alcance puede acotarse por recurso, operación, destino y vigencia.

¿Qué cambió mientras no estabas?

4 días y 3 horas. El informe muestra qué mandatos vencieron, qué ventanas se cerraron y qué sigue abierto o bajo vigilancia condicional.

> **La ausencia se cuenta desde el último evento registrado del caso.**



ILUSTRACIÓN EDITORIAL

demo/run_case_004_gap.py

Cuatro tipos de ventanas de recuperación

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
Conocida	Plazo fijo	Prioriza por vencimiento explícito
Estimada	Plazo con fundamento	Muestra la estimación y su fundamento
Condicional	Se cierra por un evento	Vigila un equipo encendido hasta que se reinicie o apague
Desconocida	Sin información temporal	Mantiene el requisito visible sin inventar urgencia

La revocación alcanza a cada descendiente

La persona a cargo delega en otra, que a su vez delega un alcance menor. Revocar el mandato raíz bloquea a quien recibió la última delegación.

> **El submandato puede quedar intacto. La decisión usa la cadena vigente en un único almacén.**



ILUSTRACIÓN EDITORIAL

demo/run_case_003_level6.py · tests/test_delegation.py

El revisor puede comprobar el encadenado

01 **Registro canónico** › 02 **Sello SHA-256** › 03 **Hash del recibo anterior** › 04 **Verificación independiente**

El sello permite comprobar integridad y encadenado. Para revisar la autoridad y la base de verificación hacen falta los registros vinculados.

Qué entregamos durante el hackathon

1

01–02 · Autorizar y recuperar

Decisiones deterministas, recibos, ventanas de recuperación y revisión de hipótesis.

2

03–04 · Adquirir y relevar

Captura autorizada por personas, evaluación independiente, conflictos y relevos ligados al estado.

3

05–06 · Comparar y delegar

Escenarios hipotéticos aislados y submandatos acotados con revocación en cascada.

Qué reprodujo la investigación

451 pruebas pasaron

Python 3.11 y 3.13 con dependencias fijadas. Se excluyó el módulo opcional de pruebas de Laya.

4 demos deterministas

Evidencia intacta sin verificar, alteración y recuperación, revocación en cascada y retorno tras una ausencia.

HTTP e interfaz

Preparación y ejecución sobre HTTP real. Compilación y lint correctos. Git registra 205 commits en siete días calendario.

Informe global §3.1 · comprobación del verificador realizada por separado

Cómo mostrar el rechazo y la recuperación

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
01 · Empezá	Espacio de trabajo	Mostrá evidencia verificada y la pregunta abierta sobre el reloj del equipo
02 · Ponelo a prueba	Custodia compartida	Registrá una tarea como hecha y mostrá qué acción exige verificación
03 · Revocá	Delegación	Prepará bajo un submandato, revocá el ancestro y mostrá el rechazo al ejecutar
04 · Verificá	Recibo	Mostrá una cadena preparada intacta y otra con un veredicto alterado. Falta el exportador para casos reales.

Alcance de los controles

Grafo de evidencia

La procedencia heredada puede omitir una invalidación semántica. Quedan dependencias por migrar a transiciones tipadas.

Despliegue

El entorno documentado usa tokens locales y SQLite en un solo proceso. Falta robustecer la identidad y el despliegue multiproceso.

Evaluación

Los incidentes muestran patrones de falla. Las demos muestran el comportamiento en escenarios acotados. El impacto en campo está pendiente.

La verificación de casos reales necesita un exportador

Hoy

En la comprobación manual, el verificador aceptó una cadena preparada de dos recibos y rechazó un veredicto alterado.

Primero

Exportar chain.json desde un caso real y agregar pruebas de aceptación y manipulación.

Después

Definir un formato versionado de evidencia y autoridad que permita reconstruir la decisión fuera de línea.

Cómo abordan el problema otros sistemas

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
TheHive 5	Tareas obligatorias y registros	La documentación revisada no describe una segunda verificación como condición para actuar
Zhu y Wang	Controles distribuidos y efectos en curso	KRAKADYL combina revocación en un almacén con evidencia y recuperación
LangGraph	Interrupción persistente del flujo	KRAKADYL reorienta al investigador y revalida el mandato al retomar
AgentBound	Diseño de recibos con reconstrucción de la decisión	KRAKADYL ya comprueba el sello y el orden de los recibos

Cada mecanismo tiene antecedentes

Capacidades

Macaroons, UCAN y Biscuit describen cómo acotar permisos. La revocación depende del formato.

Control de acciones

Bounded Agents y Veto abordan combinaciones de llamadas permitidas individualmente. KRAKADYL controla el alcance de las acciones sobre evidencia.

Transparencia

Rekor aporta un registro externo de transparencia. Anclar las cabeceras de la cadena es trabajo futuro.

ECOSISTEMA

Las herramientas que acompañan la investigación

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
Adquisición	Velociraptor · SIFT	Recolectan y examinan evidencia
Gestión de casos	TheHive · DFIR-IRIS	Gestionan casos, tareas y observables
Evidencia y análisis	Guardian · Belkasoft · Timesketch · i2	Preservan custodia, examinan artefactos y conectan cronologías o entidades
Agentes y flujos	Muse · AgentScope · DeerFlow · GigaCowork · Temporal	Aportan ejecución, aprobación, memoria o reanudación persistente

EVIDENCIA EXTERNA DE ALCANCE ACOTADO

Los sistemas de agentes también fallan al autorizar

Éxito declarado

Tallam describe una delegación en producción que figuraba como exitosa y nunca ocurrió.

55 llamadas con un límite de 5

BROODS lo observó en un framework al delegar trabajo. El resultado del taller corresponde a esa configuración.

19,2 % de trayectorias

La cifra de AuthorizationBench proviene de cobertura secundaria y describe una muestra de evaluación.

Requisitos de custodia en distintas jurisdicciones

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
Argentina · Brasil	Registro de traspaso	Argentina pide una hoja de ruta. Brasil documenta etapas e identidades de custodia.
Japón · Corea del Sur	Registros de preservación	Japón registra detalles de la entrega. Para Corea se consultaron fuentes secundarias.
India · China	Autenticidad e integridad	Los certificados, hashes y reglas de datos electrónicos se revisaron en fuentes secundarias.
Alcance	Cobertura y aplicación	La admisibilidad y el mapeo por jurisdicción requieren evaluación específica.

Qué vamos a medir

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
Preservación	Resultado principal	Proporción de artefactos recuperables verificados a tiempo y tasa de cierres inconclusos evitables
Continuidad	Resultado operativo	Tiempo para reconstruir el caso y llegar a la próxima acción defendible
Seguridad	Límites de seguridad	Acciones sin autorización, verificaciones falsas y abstenciones justificadas preservadas
Diseño	Comparación	Comparar casos equivalentes con interrupciones, con y sin el espacio de trabajo

La próxima etapa conecta flujos reales

1

Adaptadores

Definir contratos acotados para IRIS, TheHive, Velociraptor, Timesketch, SIFT, i2 y VIGÍA.

2

Evaluación reproducible

Publicar escenarios de interrupciones, cambios de alcance y artefactos no disponibles.

3

Piloto

Convocar equipos de DFIR para medir preservación, tiempo de retorno y rechazos en turnos realistas.

PRÓXIMO HITO / VALIDAR CON INVESTIGADORES

Conservá la evidencia. Mantené el hilo.

Pedile dos pruebas a la demo. Una afirmación sigue siendo una afirmación hasta que se verifica. Una delegación termina cuando cae su autoridad.

github.com/annatchijova/krakadyl



Cómo vamos a evaluar la hoja de ruta

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
Interoperabilidad	Adaptadores de lectura y exportación	Una tarea de IRIS o TheHive queda declarada hasta su verificación. Velociraptor aporta procedencia.
Evaluación reproducible	Prerregistro	Comparar cierres inconclusos evitables y tiempo de retorno con un gestor de tareas como base.
Usabilidad	Interrupciones controladas	Medir identificación de ítems verificados, retorno y carga con investigadores, incluidos participantes neurodivergentes.
Federación y auditoría	Revocación entre almacenes	Bloquear efectos antes de confirmarlos, anclar la cadena y publicar una auditoría independiente.

Fuentes · investigaciones de incidentes

FAA / NTSB	Vuelo 2574 de Continental Express · AAR-92/04
Investigación Cullen / HSE	Investigación pública sobre Piper Alpha · 1990
Fiscalía / Policía Metropolitana	Revisión conjunta del proceso de revelación en R v Allan · 2018
SEC / CSB / NASA	Orden Knight Capital 34-70694 · BP Texas City · Informe Mars Climate Orbiter
CSRB / Mandiant	Revisión de Storm-0558 · Análisis de BRICKSTORM en M-Trends 2026

Métodos y límites de la revisión

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
Repositorio	Commit público 1d56d66	Pruebas con dependencias fijadas, HTTP, compilación y lint, cuatro demos deterministas
Investigación externa	Unas 75 búsquedas y lecturas	Primero código y normas. Luego estudios, investigaciones, documentación y prensa especializada.
Cobertura	Ocho regiones	Sin búsquedas en hindi o hebreo. SOAR y patentes fuera del alcance.
Límites	Cambios locales y servicios en vivo	La verificación excluyó la interfaz ES/EN local, el navegador, los modelos y las sondas en vivo.

Preguntas de un jurado especialista

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
¿Dos personas pueden ponerse de acuerdo?	Identidad verificadora	Se exige otra identidad autenticada. La garantía no cubre colusión.
¿Y si no registran una fuente?	Alcance del registro	La ventana aparece para las fuentes registradas en el caso.
¿Es admisible en juicio?	Admisibilidad	Sin evaluar. Los registros permiten mapear requisitos por jurisdicción.
¿Qué prueba un rechazo?	Alcance del recibo	Con el contenido canónico se comprueban sello y cadena. La autoridad requiere el estado de respaldo.

El ejecutor comprueba el estado vigente

PASO / CONDICIÓN	CONTROL	RESULTADO OBSERVABLE
Cambian los bytes	Integridad de la evidencia	Aborta la publicación y registra el motivo
Cambia la autoridad	Cadena de mandatos	Rechaza una acción vencida o revocada
Se disputa una tarea	Verificación independiente	Trata el requisito correspondiente como conflictivo
Se repite una solicitud	Nonce de un solo uso	Rechaza la repetición

Fuentes · estudios y evidencia del proyecto

Starmer y cols. · NEJM 2014	Cambios en errores médicos tras implementar un programa de relevos
Fonarow y cols. · JAMA 2014	Tiempo hasta el tratamiento y resultados antes y después de Target: Stroke
Westbrook y cols. · Arch Intern Med 2010	Interrupciones, frecuencia y gravedad de errores de medicación
Haynes y cols. · NEJM 2009	Lista de verificación quirúrgica para reducir morbilidad y mortalidad
KRAKADYL · 1d56d66	Repositorio público, cuatro demos reproducibles e informe global