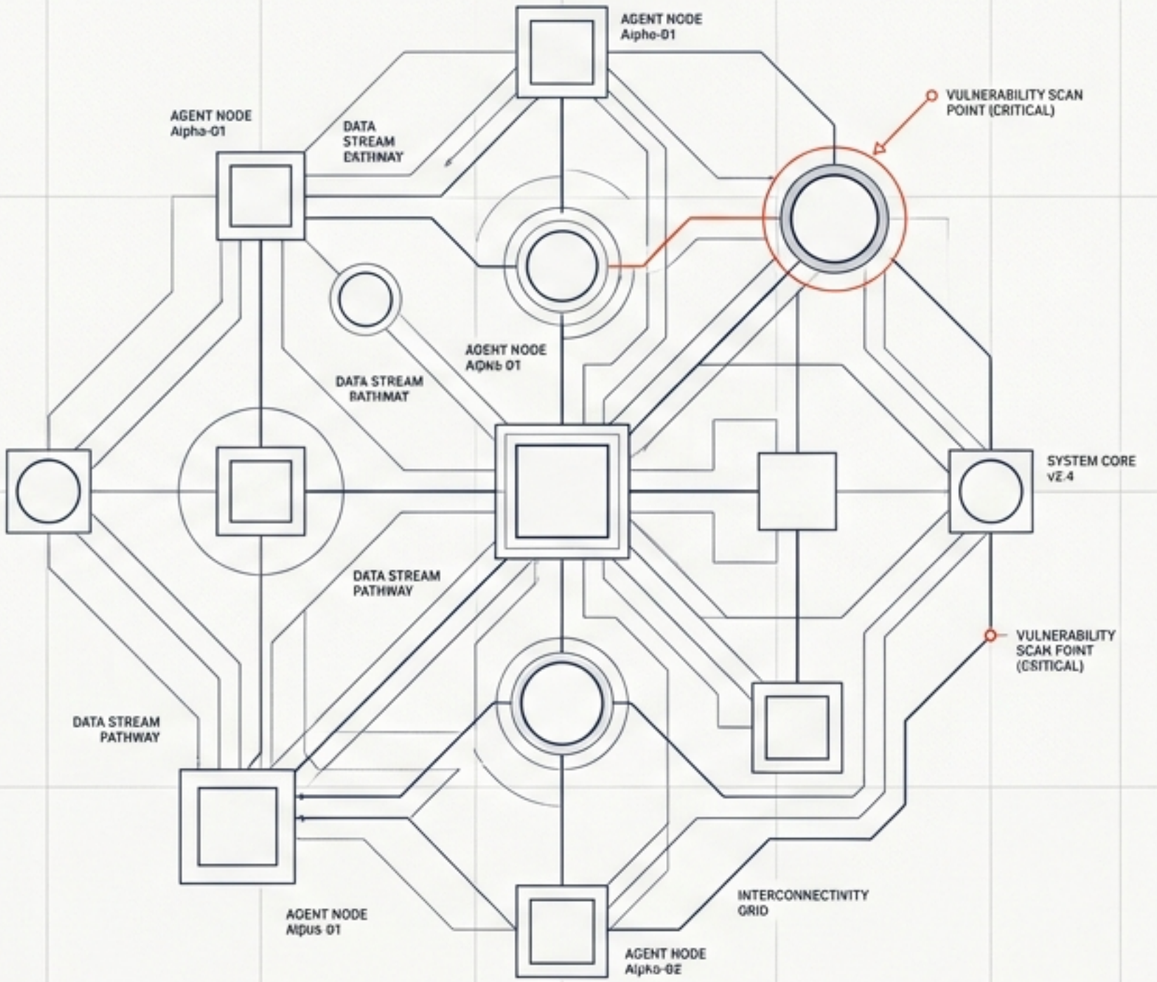
MDASH: Ciberdefensa Autónoma a Velocidad de IA

La arquitectura multi-agente de Microsoft que redefine el descubrimiento de vulnerabilidades a escala empresarial.



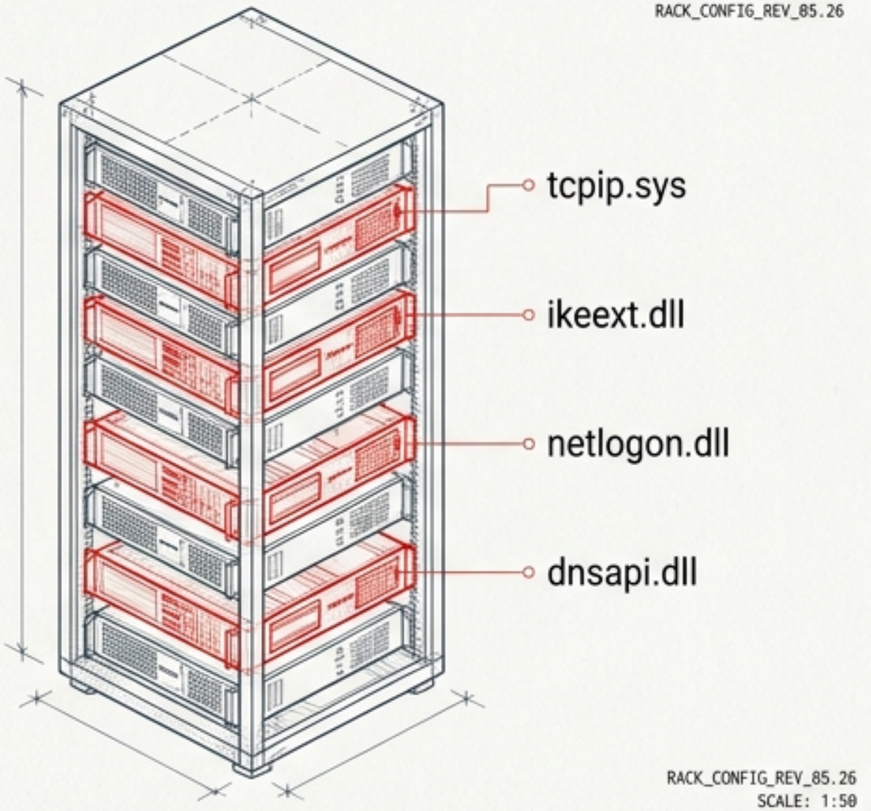
Investigación y Desarrollo: Autonomous Code Security Division (Microsoft)

NotebookLM

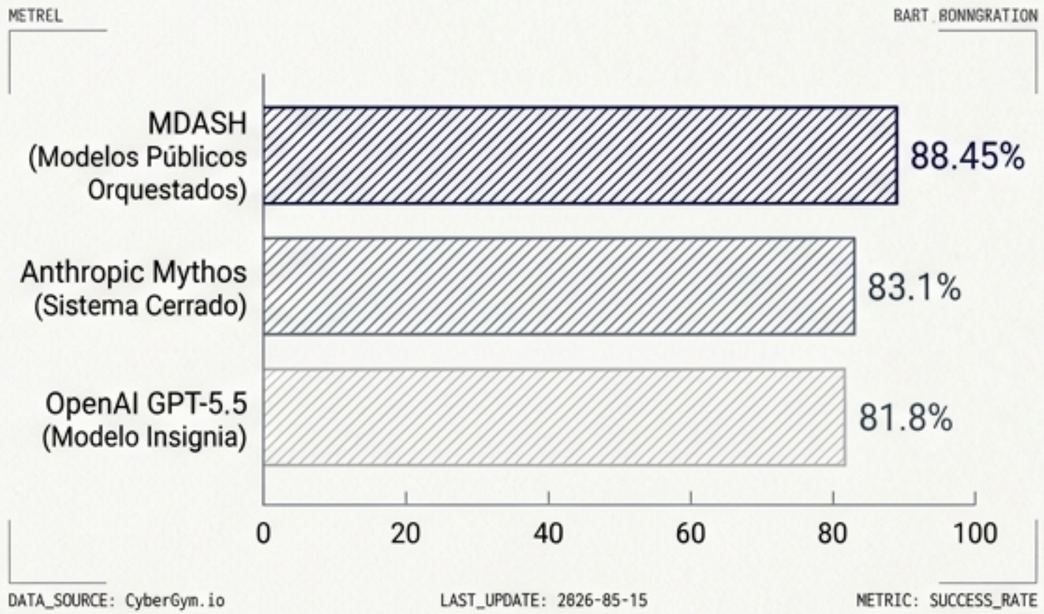
El Impacto: Resultados Tangibles y Dominio de Benchmarks

16 Nuevas Vulnerabilidades (Patch Tuesday, Mayo 2026)

4 Fallas Críticas RCE (Remote Code Execution)



Benchmark CyberGym (1,507 tareas del mundo real)



MDASH encabeza la tabla de clasificación utilizando modelos públicos, superando a las IAs de seguridad más potentes y cerradas del mundo.

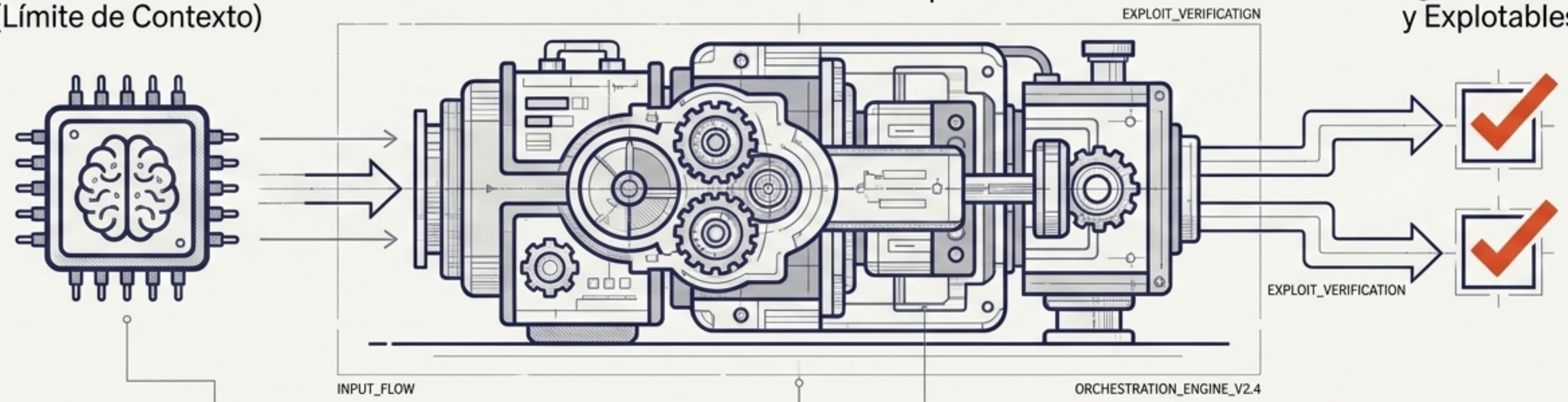
NotebookLM

La Tesis de Microsoft: El Valor Reside en el Sistema, No en el Modelo

El arnés hace el trabajo, el modelo es solo un insumo.

Modelo Único
(Límite de Contexto)

Sistema MDASH: Arnés Orquestador



Bugs Verificados
y Explotables

El Límite del Modelo Único

Los modelos por sí solos carecen de contexto cruzado en repositorios masivos privados.

La Solución MDASH

Enmarca modelos públicos dentro de una tubería orquestada con herramientas y estado.

Pedigrí de Ingeniería

Desarrollado por veteranos del Team Atlanta (Ganadores del DARPA AI Cyber Challenge de $29.5M).

SYSTEM_ARCH_REF_MS

NotebookLM

Contraste Estratégico: Fuerza Bruta vs. Orquestación Agéntica

Dimensión	Enfoque de Modelo Único	**Enfoque Multi-Agente (MDASH)**
Visión de Código	○ Aislada / Un solo archivo	● Holística / Rastreo de 6+ archivos
Método de Validación	○ Aceptación ciega del primer output	● Debate adversarial entre agentes
Manejo de Falsos Positivos	○ Alto volumen de ruido	● 0% en pruebas de caja blanca privada
Prueba de Concepto (PoC)	○ Generación de hipótesis estáticas	● Ejecución dinámica y validación real
Dependencia Tecnológica	○ Atado a una versión específica de IA	● Totalmente agnóstico al modelo

NotebookLM

Arquitectura Base: Una Red de Más de 100 Agentes Especializados

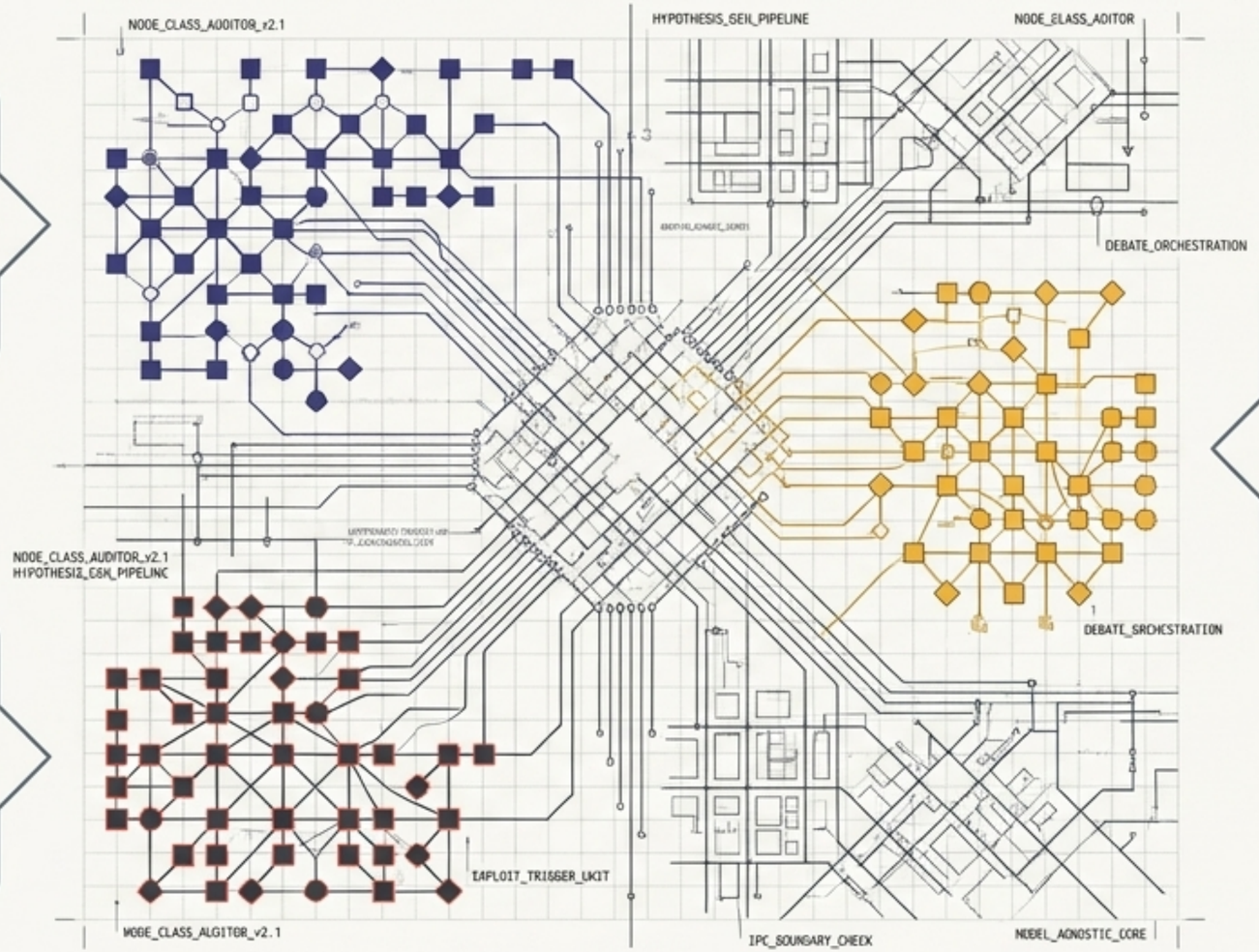
MDASH no ejecuta un único prompt monolítico. Orquesta más de 100 agentes independientes con roles y criterios de detención específicos.

Agentes Auditores

Especializados en reconocer patrones, convenciones del kernel y límites de confianza IPC. Generan las hipótesis iniciales.

Agentes Probadores

Ingenieros de explotación dinámicos que escriben y ejecutan código para disparar vulnerabilidades reales.



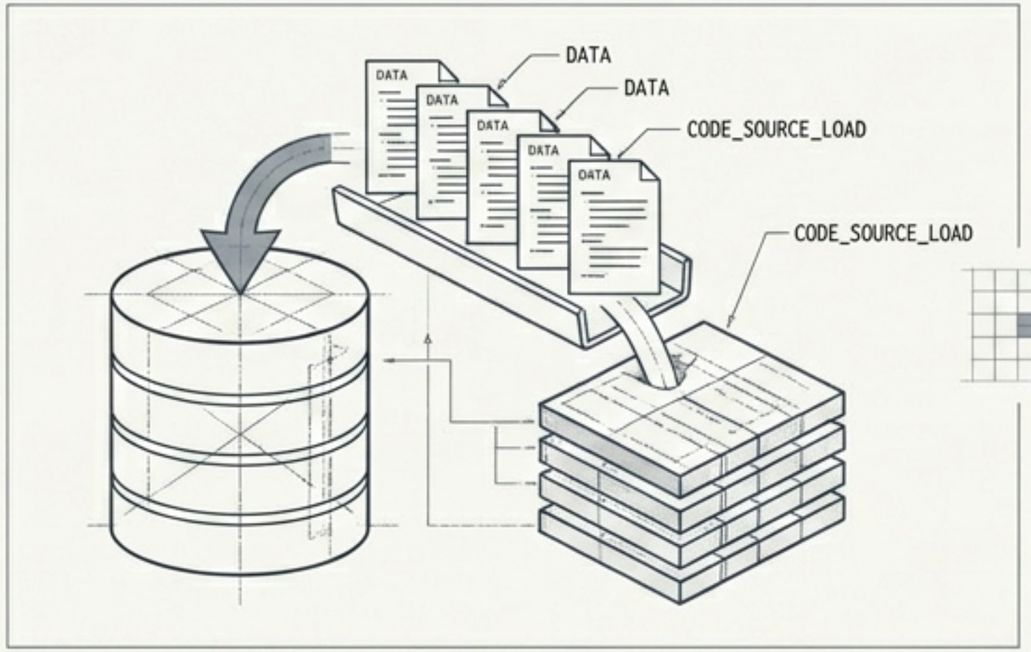
Agentes Debatientes

Modelos destilados y de vanguardia que actúan como fiscales y defensores, debatiendo l; debatiendo la viabilidad de cada hallazgo.

NotebookLM

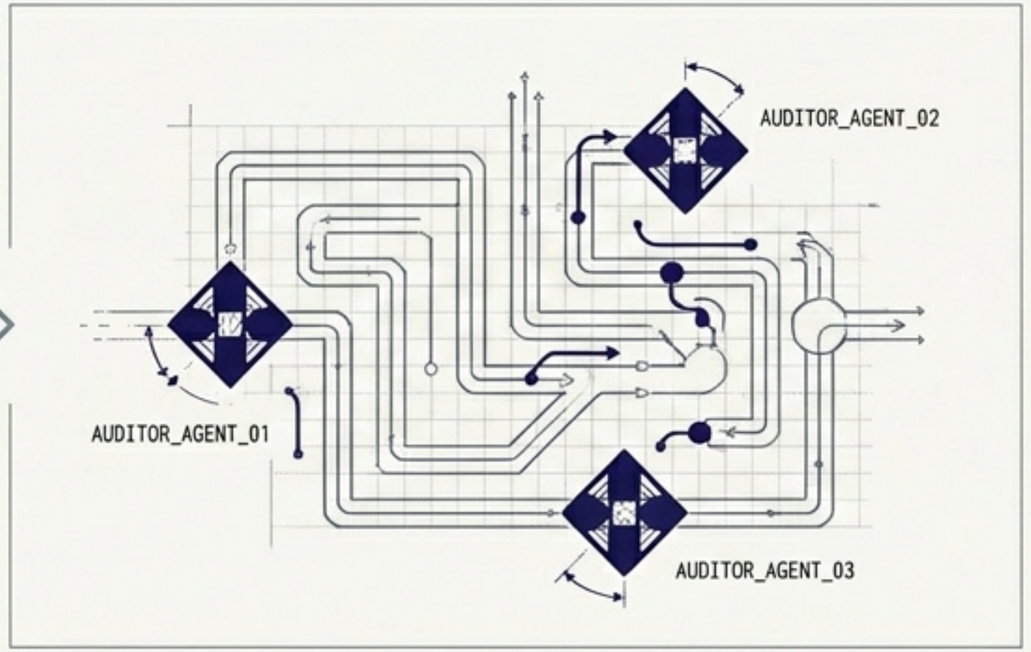
El Pipeline MDASH (Fases 1 y 2): Ingesta y Escaneo

Fase 1: Preparación (Ingesta de Contexto)



El sistema ingiere el código fuente, construye índices sensibles al lenguaje y dibuja superficies de ataque analizando commits históricos.

Fase 2: Escaneo (Generación de Hipótesis)

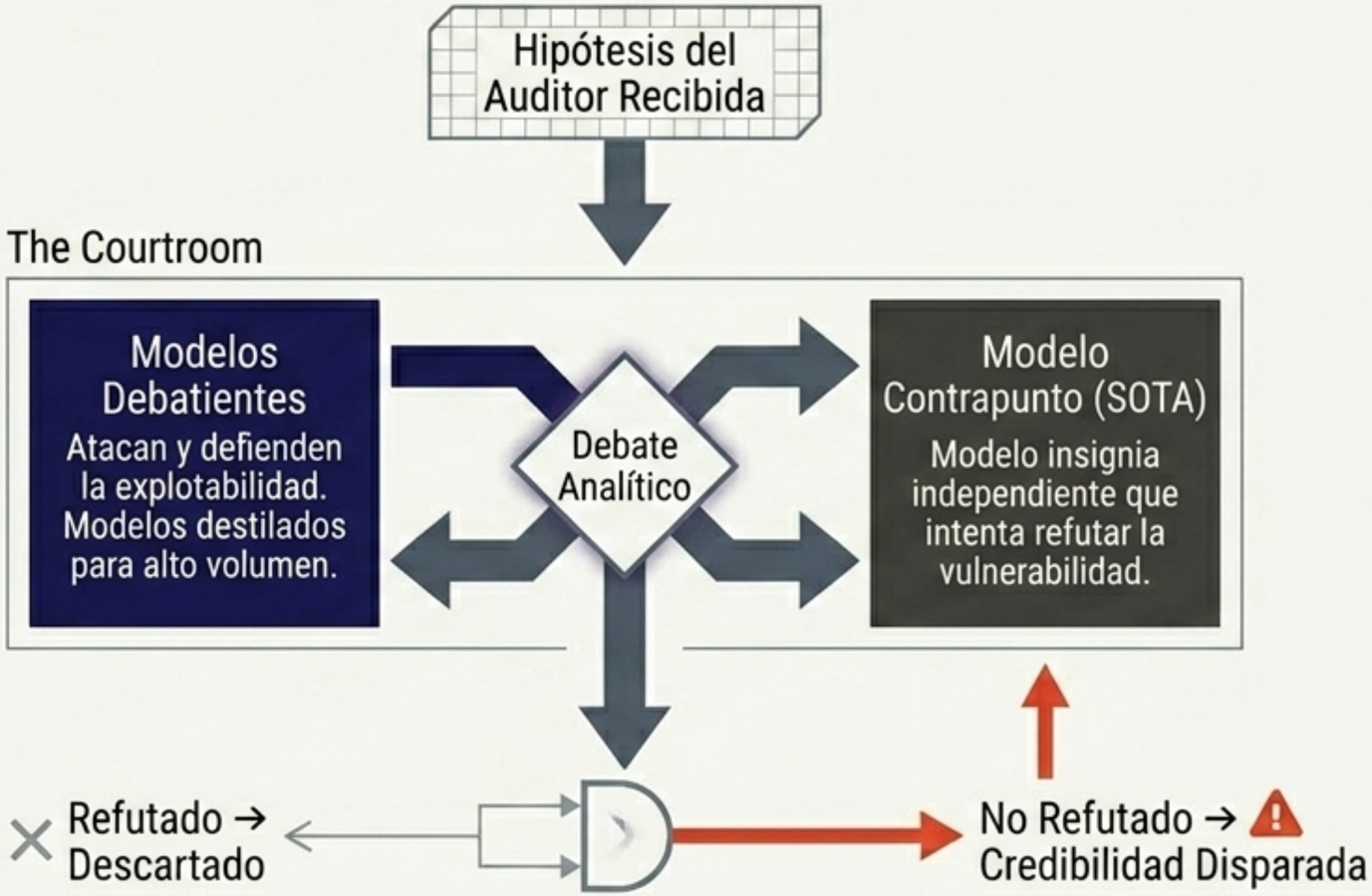


Agentes auditores recorren rutas de código candidatas. Emiten hipótesis formales respaldadas con evidencia de código específica, no solo búsquedas de patrones.

NotebookLM

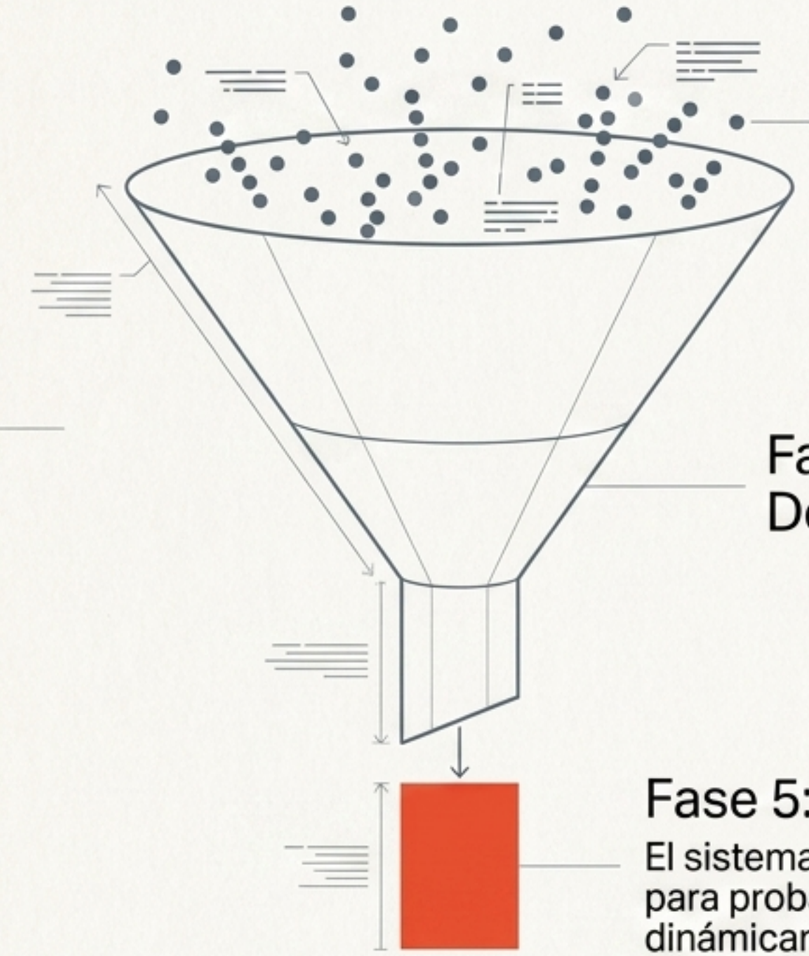
El Tribunal de IA: Validación por Debate (Fase 3)

El desacuerdo es la señal de precisión. Los falsos positivos mueren aquí.



NotebookLM

De la Teoría a la Evidencia Irrefutable (Fases 4 y 5)



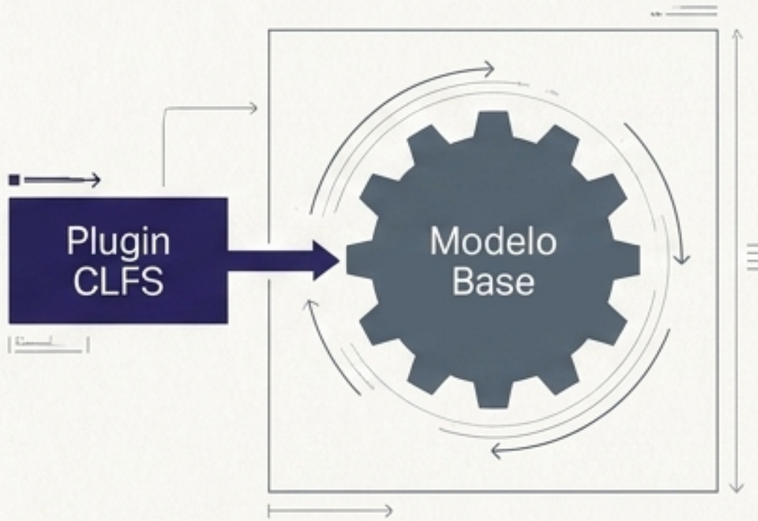
Cientos de rutas y hallazgos semánticamente equivalentes

Fase 4: Deduplicación

Fase 5: Prueba Dinámica

El sistema escribe scripts reales (PoC) para probar la vulnerabilidad dinámicamente en el entorno.

El Rol Crítico de los Plugins de Dominio



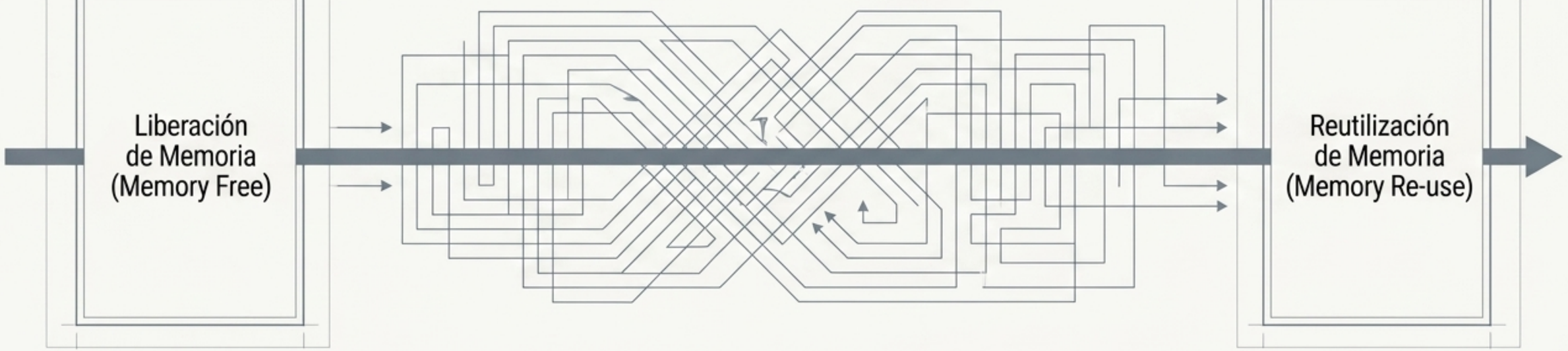
Un modelo base desconoce los invariantes del sistema de archivos de Microsoft. El plugin inyecta ese contexto, permitiendo al sistema construir un archivo de registro malicioso para confirmar el error.

NotebookLM

Detectando lo Invisible: Condición de Carrera en el Kernel

CVE-2026-33827 | Componente: tcpip.sys | Impacto: RCE Crítica sin Autenticación

El Punto Ciego: Flujo de control no trivial y verificaciones concurrentes



IA Tradicional

Busca patrones estáticos unidos. Las operaciones parecen independientes debido a la separación temporal. Pasa por alto el error.

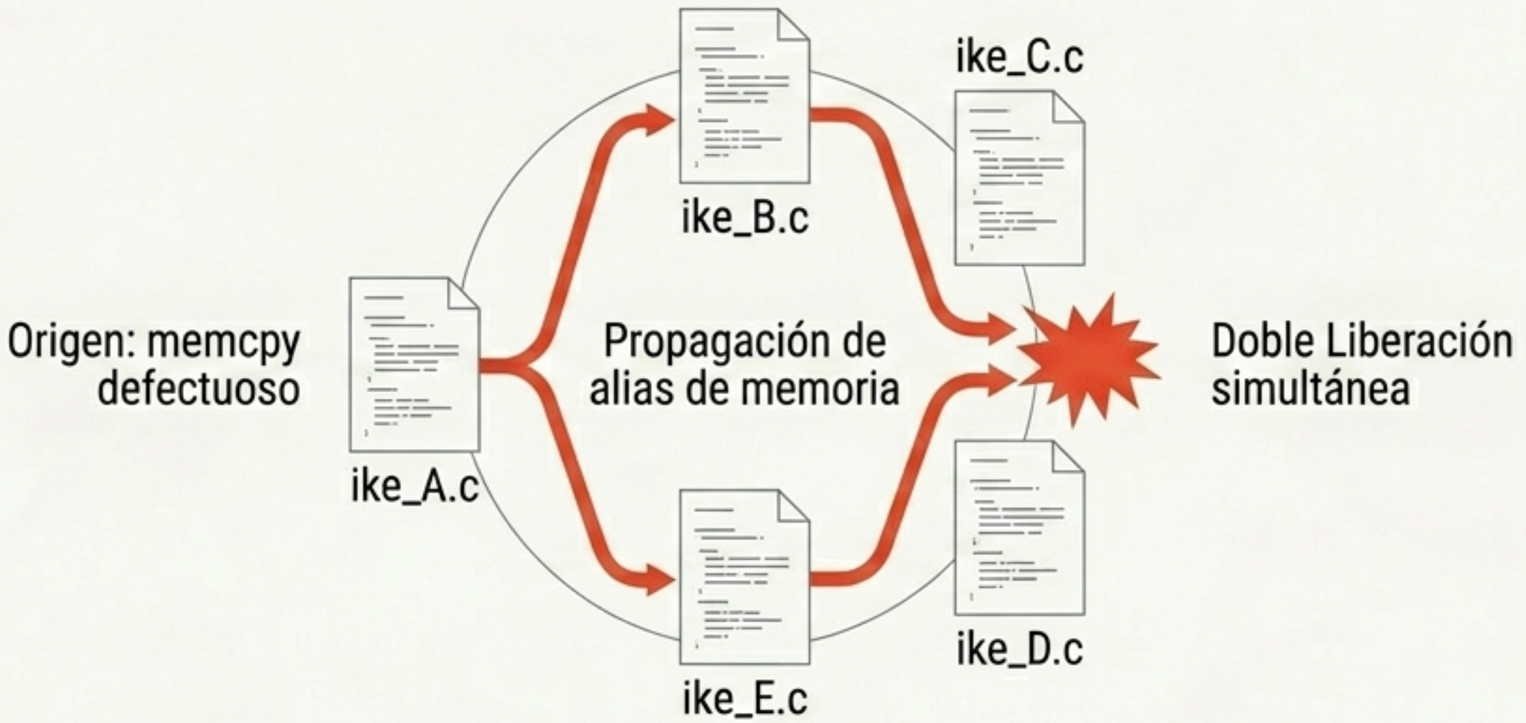
Ventaja MDASH

Rastrea la propiedad del objeto Path a través de estados intermedios. Identifica la recolección de basura guiada por estado concurrente.

NotebookLM

Rastreando Vulnerabilidades a Través de Múltiples Archivos

CVE-2026-33824 | Servicio IKEEXT | LocalSystem RCE (Double-Free)



El ‘Momento Aha’ Agéntico:

Ningún análisis de un solo archivo detectaría esto. MDASH encontró el error comparando el patrón defectuoso con un manejo correcto escondido en otra sección de ike_D.c.

NotebookLM

Telemetría de Precisión y Recuperación Histórica



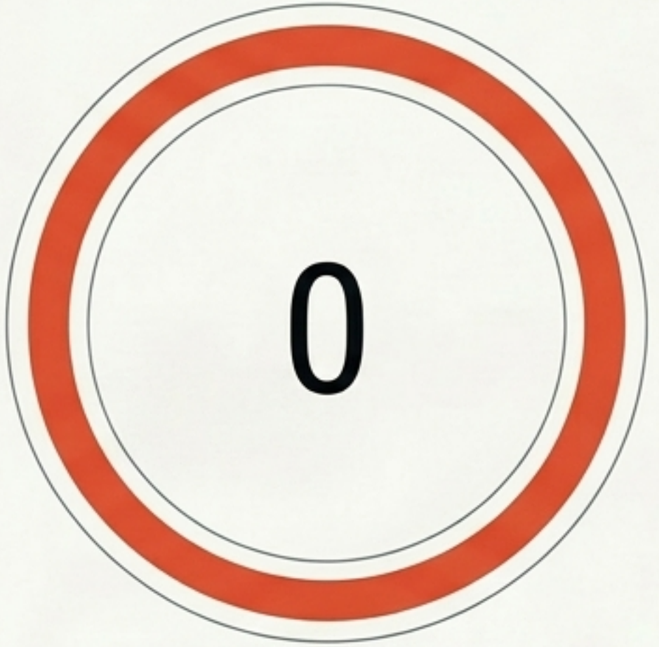
Recall en tcpip.sys

El sistema redescubrió el 100% de los casos confirmados por el MSRC en el componente central tcpip.sys abarcando los últimos 5 años.



Recall en clfs.sys

Recuperación exitosa de 28 casos históricos del MSRC en un periodo de 5 años. Errores reales que requirieron parches.



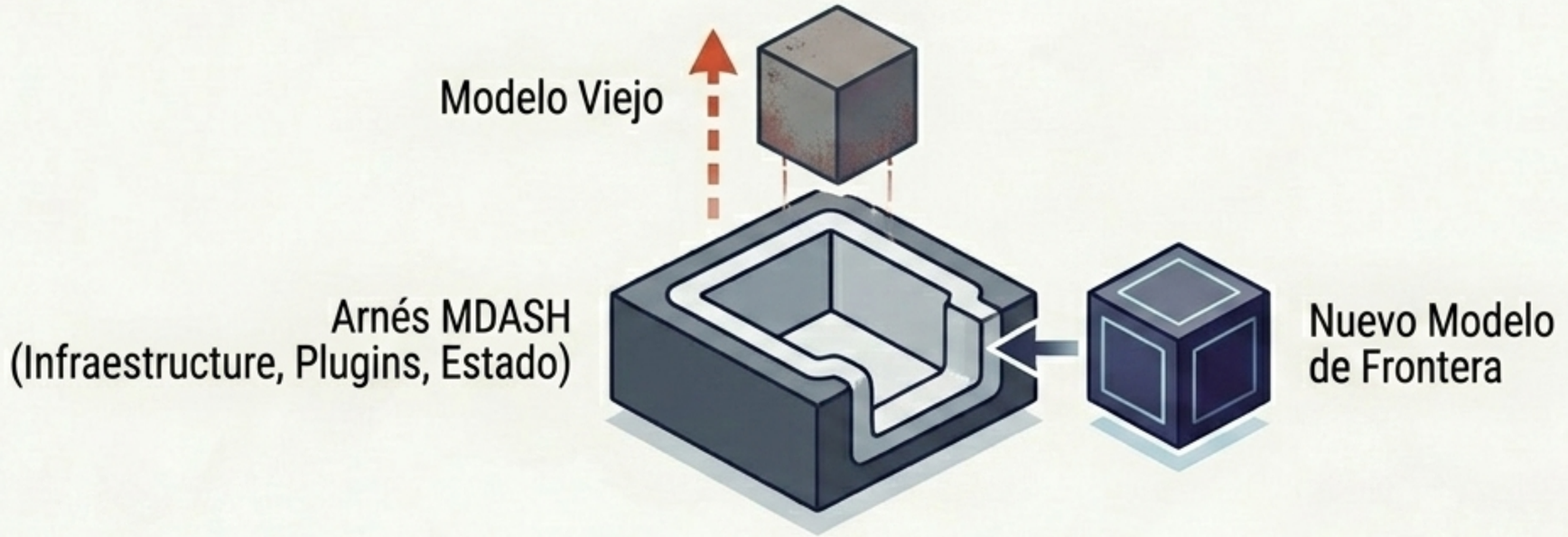
Falsos Positivos

Prueba de caja blanca escaneando el controlador privado StorageDrive. Encontró 21 vulnerabilidades inyectadas sin generar una sola alerta falsa.

NotebookLM

Arquitectura Agnóstica: Inmunidad a la Obsolescencia de Modelos

La carrera de la IA continuará, pero la infraestructura de seguridad no debe reconstruirse cada seis meses.



Sustitución Inmediata

La integración de un nuevo modelo se realiza mediante un simple cambio de configuración y una prueba A/B.

Preservación del Valor

Todo el trabajo de ingeniería profundo (plugins CLFS, calibraciones, herramientas) se mantiene intacto. El sistema simplemente se vuelve más inteligente.

Independencia

Supera ecosistemas cerrados al permitir a los defensores usar las mejores herramientas de OpenAI, Anthropic y otros simultáneamente.

NotebookLM

La Carrera Ofensiva-Defensiva Exige Sistemas Orquestados



- Los atacantes ya tienen acceso a los mismos modelos de IA públicos masivos.
- La ventaja táctica ya no pertenece a quien tiene el modelo más grande, sino a quien posee la mejor arquitectura de orquestación.
- MDASH transforma la IA de un generador de ruido especulativo a un motor de validación de grado de producción.

PRIVATE PREVIEW ACTIVA

MDASH está siendo utilizado actualmente por Microsoft. Evalúe el futuro de la **seguridad de código** autónoma en su propia infraestructura.

NotebookLM