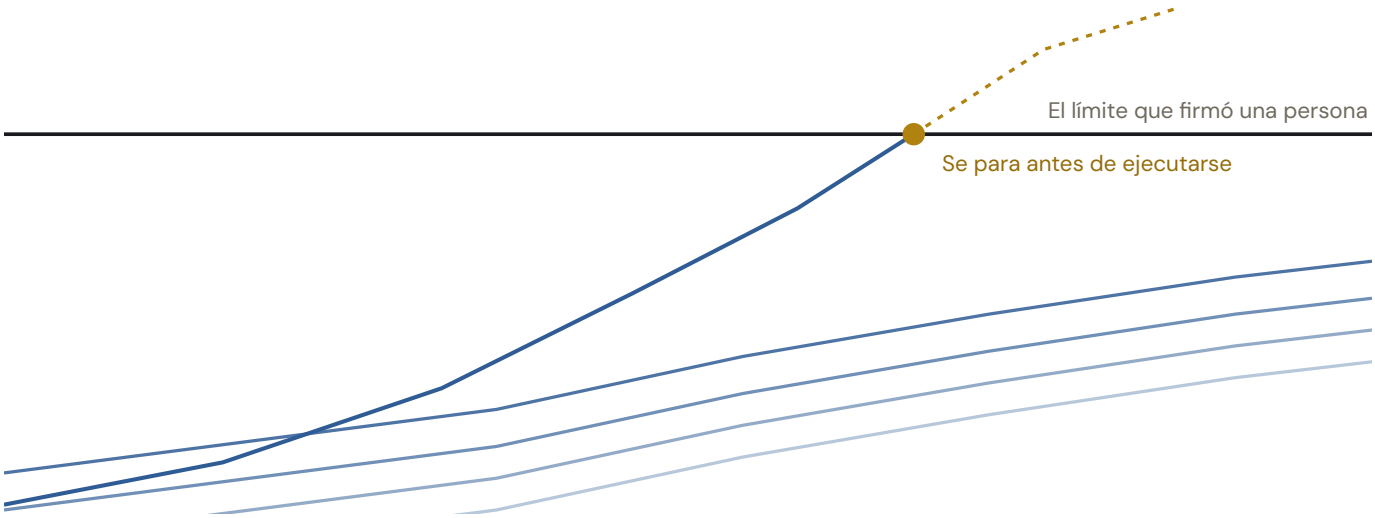
Todo lo que tocan *tus agentes,* en un solo lugar.

Cómo D4A reúne tus reglas, tus jueces automáticos y a las personas, antes de que tus agentes actúen. Para tu equipo de IT, de seguridad y de cumplimiento.



Contacto diacroma@veritglobal.com
Web www.diacroma.com
De Verit Global Labs

Pruébalo 30 días gratis en Microsoft Marketplace

marketplace.microsoft.com/en-us/product/veritglobal.d4a

Antes, la maraña. *Después, una sola puerta.*

Con D4A, cada acción de tus agentes pasa por una sola puerta, donde se comprueban tus reglas y se firma la decisión. El dinero que mueven se queda dentro de los límites que firmó una persona.

Antes: la maraña

Cada agente llega por su cuenta a tus herramientas. Tus políticas viven en un sistema, los filtros de contenido en otro, los evaluadores en un tercero, y las personas de tu empresa se enteran de un problema cuando se queja un cliente. Cada equipo monta sus propias comprobaciones, así que la misma pregunta recibe una respuesta distinta según el agente. ¿Qué regla manda? ¿A quién había que avisar? ¿Dónde está la prueba?

Después: el ovillo sencillo

D4A se coloca en el único camino que ya usa cada agente para llegar a una herramienta. Antes de que salga una llamada, consulta tus fuentes de reglas: los límites que una persona firmó para ese agente y las políticas de tu empresa, mientras the Watch revisa el patrón.† Decide, y anota la decisión en un registro firmado que vive en tu nube. Cuando algo necesita a una persona, abre un caso, que le llega a quien firmó el límite.† Tus agentes no cambian.

En nuestro propio entorno de pruebas, un reembolso de $150 consultó primero la regla OPA de la empresa. La regla dijo que no, la llamada se detuvo y el motivo quedó en el registro firmado.

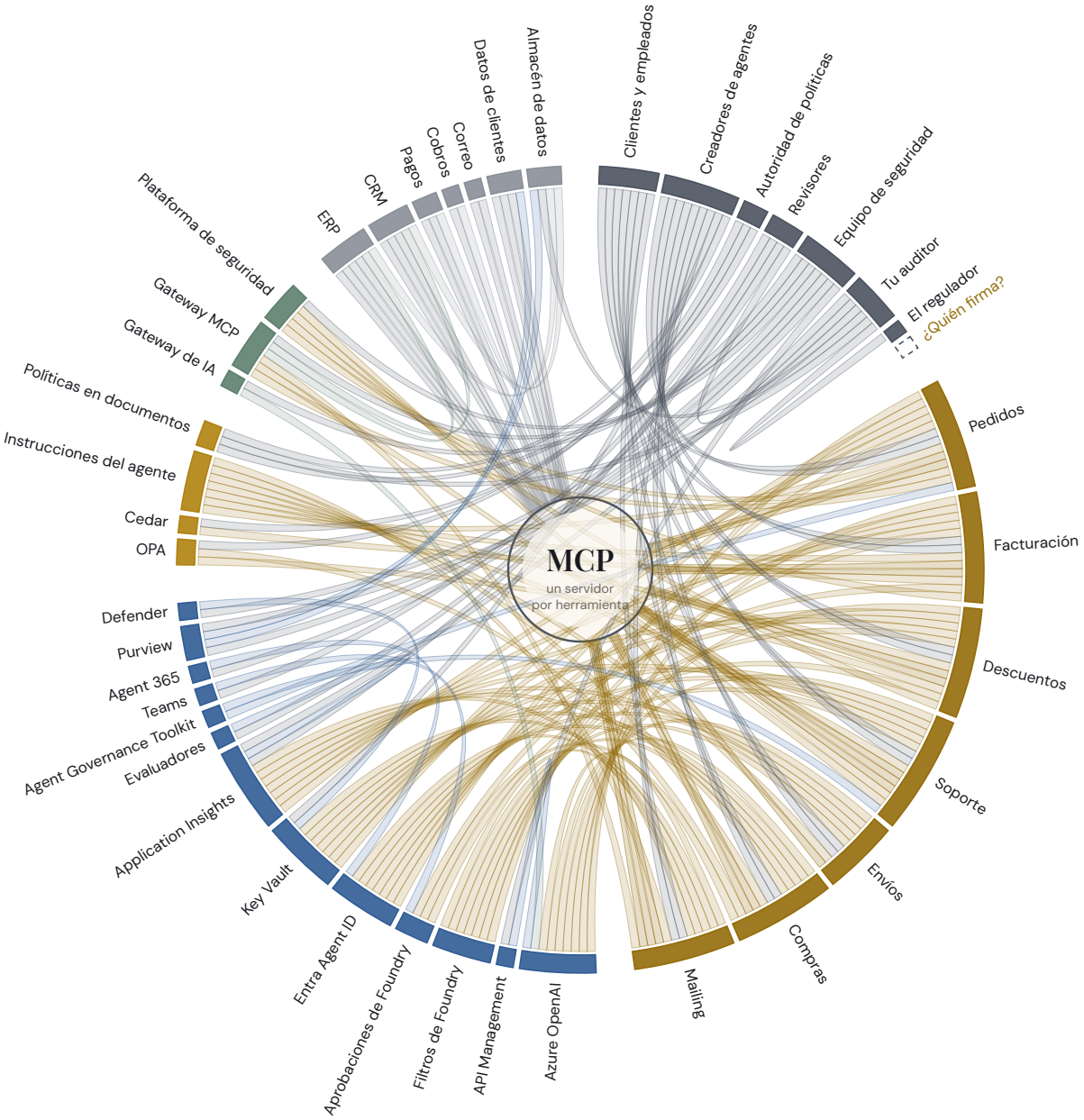
En este folleto

La maraña y el ovillo	3
Las dos facturas de un agente	5
La caja, y lo que pasa dentro	6
Los 24 problemas, con su código OWASP	7
Por qué MCP es el lugar donde decidir	10
Robusta y segura por construcción	11
Tu revisión de seguridad	12
Quién hace qué	13
Una persona decide	14
Tipos de agente	15
Cómo se instala y se opera	17
Anexo A · La comparativa a fondo	18
Anexo B · Todo lo que hace, en una lista	22
Anexo C · Lo que preguntará tu equipo de IT	23
Anexo D · Cómo se compra	25
Anexo D · A tu medida	26
Anexo E · Las herramientas que D4A reúne	27

† en desarrollo

Sin D4A

Más de 40 actores y más de cien conexiones que alguien configura y mantiene



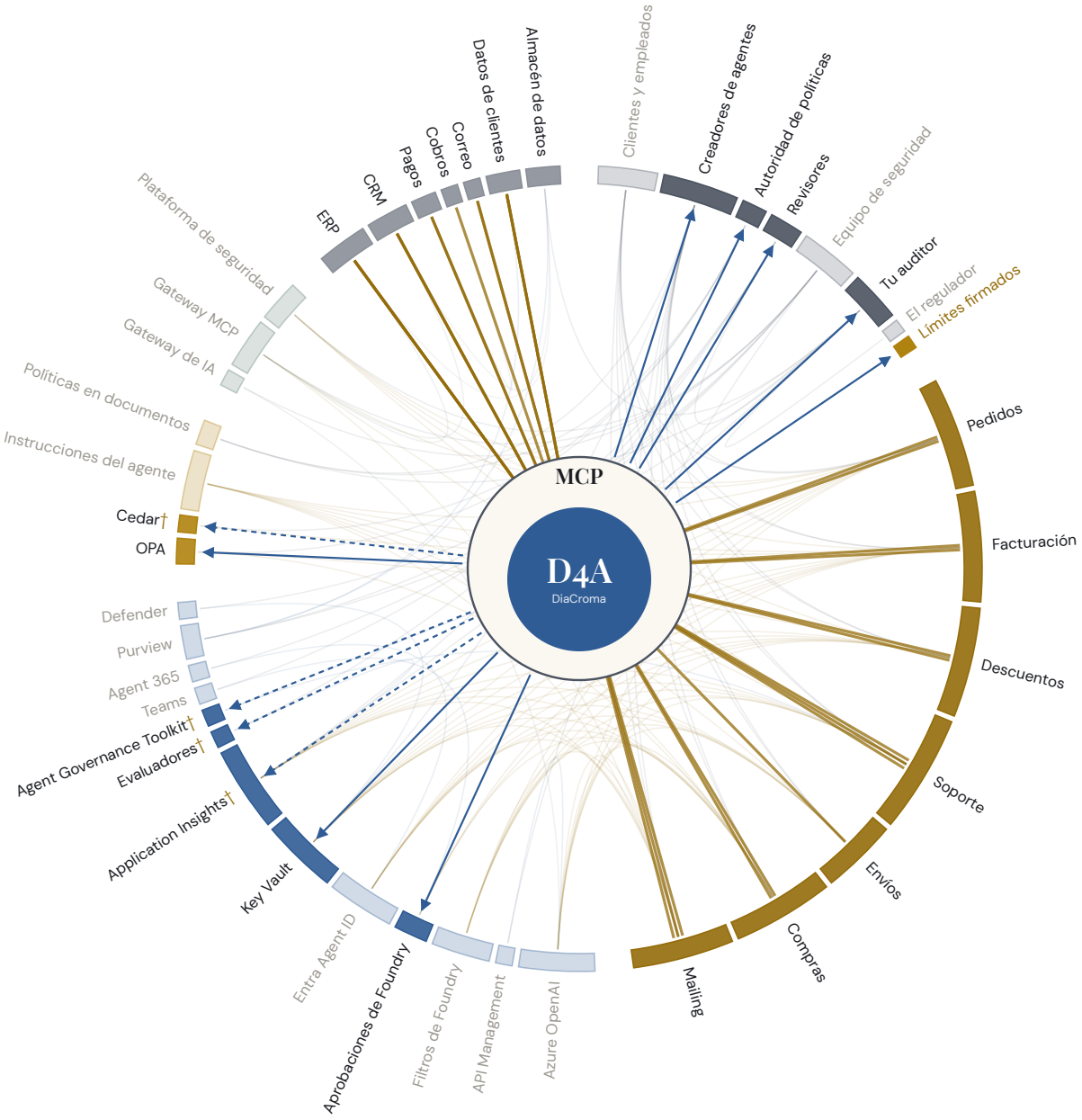
¿Qué regla manda? ¿A quién se avisa? ¿Dónde está la prueba?

Hoy, en una empresa de ejemplo. Cada arco es un actor: personas, agentes, piezas de Microsoft, reglas, otras guardas, sistemas y datos. Cada cuerda es una conexión que alguien configura y mantiene. Las llamadas de los agentes a tus sistemas pasan por MCP, con un servidor por herramienta, y cada agente llega con sus propias llaves. Solo algunas consultan un motor de reglas, los jueces automáticos puntúan después a partir de las trazas, y las personas lo revisan todo a mano. El arco de «¿Quién firma?» se queda vacío.

† en desarrollo

Con D4A

Tus agentes siguen llamando a sus herramientas por MCP; D4A está en el camino y da paso o para



Una sola puerta para cada acción y un registro firmado de cada decisión.

Los mismos actores, con D4A. Tus agentes no llaman a D4A: siguen llamando a sus herramientas por MCP, y D4A está dentro de ese camino, en tu red. Las flechas azules son lo que hace D4A: consulta tus reglas y las aprobaciones de Foundry, usa tus llaves de Key Vault, llama a los jueces automáticos y lleva los casos a las personas por la consola; después firma el resultado. Lo que sale en claro no pasa por D4A y sigue como estaba. Discontinua y con †: en desarrollo.

† en desarrollo

Lo que gastan tus agentes: *dos facturas.*

Un agente genera dos facturas. La de lo que hace con tu dinero llega de golpe: un error cuesta todo lo que el agente movió. La de los tokens, lo que gasta en pensar, llega cada mes y no deja de crecer. D4A detiene la primera antes de que ocurra, y te avisa de la segunda antes de que el mes se dispare.†

	Lo que hace con tu dinero	Lo que gasta en pensar
Qué es	Un reembolso, un descuento, un pago.	Los tokens que usa el modelo en cada paso.
Cómo lo trata D4A	Lo detiene antes de que llegue a tu sistema, y abre un caso.	Avisa, nunca detiene: D4A no se coloca delante del modelo.
En D4A, hoy	Ningún reembolso de más de $250 por llamada; no más de $2.000 en total, la cuenta acumulada del agente.	Nada todavía: el techo de tokens se fija en la plataforma que sirve el modelo (por ejemplo, Microsoft Foundry).
En desarrollo †	$600 por cliente al mes; un presupuesto de equipo, contado para todos los agentes del equipo a la vez; el caso le llega a quien firmó el límite.	Lo que cuesta cada acción en tokens y si sube semana a semana; bucles cortados: la misma llamada con los mismos datos no más de 5 veces, y no más de 50 llamadas por hora; un aviso a quien firmó el agente cuando sus tokens suben, o cuando una herramienta devuelve una respuesta tan grande que encarece el paso siguiente.

En producción, un reembolso de $10.000 contra un límite de $250 por llamada *nunca se ejecutó.*

Lo que D4A no hace con los tokens

D4A está entre el agente y sus herramientas, no entre el agente y su modelo. No ve las llamadas al modelo, no es un proxy del modelo y no pone el techo de tokens: ese techo se pone en la plataforma que sirve el modelo. Lo que sí corta son llamadas a herramientas: los bucles y las llamadas por hora.† Que se ahorren tokens es una consecuencia.

Por qué importa ahora

Google pasó de 9,7 billones de tokens al mes en sus servicios en mayo de 2024 a más de 3.200 billones en mayo de 2026, siete veces más que un año antes. Gartner prevé que más del 40 % de los proyectos de IA agéntica se cancelen antes de finales de 2027, por costes crecientes, un valor de negocio poco claro o controles de riesgo insuficientes.

Fuentes: Google I/O 2026 · Gartner, 25 de junio de 2025.

La caja, *y lo que pasa dentro.*

Una jaula no hace que un agente se porte bien. Tu plataforma y tu sandbox construyen la caja; D4A acota lo que el agente hace dentro. Conserva las dos: cada una cubre lo que la otra no ve.

La caja: tu plataforma y tu sandbox	Dentro de la caja: D4A
Qué ficheros puede leer y escribir el agente	Qué acciones puede hacer sobre tus sistemas
A qué destinos de red puede llegar	Si cada llamada está permitida, una a una y en total
Qué procesos y llamadas al sistema puede ejecutar, incluido el código que escribe	Si las reglas de tu empresa lo permiten
Adónde van sus peticiones al modelo y qué palabras puede producir el modelo	Si sigue en su tarea con el paso del tiempo†
Cuántos tokens puede gastar el modelo (su techo)	Cuánto cuesta en tokens cada acción, y a quién se avisa cuando sube†
Quién es el agente	Qué tenía permitido, quién lo firmó y la prueba

Quién construye la caja

Tu plataforma de agentes y tu sandbox. Por ejemplo, Microsoft Foundry ejecuta el agente, sus filtros de contenido controlan las palabras y Entra Agent ID le da una identidad; NVIDIA OpenShell ejecuta cada agente en un sandbox aislado por el sistema operativo, donde ficheros, red y procesos quedan confinados y el tráfico de salida se deniega salvo que una política lo permita. D4A no cambia ninguno de ellos.

Mejor juntos

Cuando la caja solo deja una salida hacia tus herramientas, y esa salida es the Boundary, cada acción del agente pasa por D4A. La caja cierra las demás puertas; D4A decide qué pasa por la que queda abierta.

Fuente: NVIDIA OpenShell overview.

† en desarrollo

Los 24 problemas de los agentes de IA *que resuelve D4A.*

Un agente puede fallar de 24 maneras, y cada una tiene su sitio en D4A. Léelas como la vida de una llamada: antes de salir, mientras se ejecuta, con el tiempo y después. Lo que no lleva † D4A lo resuelve hoy; lo que lleva † está en desarrollo.

La última columna da el código de cada problema en el OWASP Top 10 for Agentic Applications. OWASP (Open Worldwide Application Security Project) es la fundación sin ánimo de lucro detrás de las listas de riesgos de seguridad más conocidas; en diciembre de 2025 publicó esta, dedicada a los agentes de IA: diez riesgos, de ASI01 a ASI10. Una casilla vacía significa que ningún código concreto encaja.

#	El problema	En una tienda online	Qué lo resuelve	OWASP
Antes de la llamada				
1	Usa una herramienta que nunca debió usar	El agente de soporte intenta hacer un reembolso	Tu catálogo firmado: solo las herramientas que aprobó una persona	ASI02
2	Una llamada va demasiado lejos	Un reembolso de $10.000	Un límite firmado por llamada: $250 por reembolso	ASI02
3	Cada llamada está bien; el total no	Cuarenta reembolsos pequeños al mismo cliente	Un límite con memoria: $2.000 en total; $600 por cliente al mes†	ASI10
4	Varios agentes agotan juntos un mismo presupuesto, o se pasan el trabajo	Diez agentes de descuentos se reparten $500 al día	Un límite de equipo, contado para todos a la vez†	ASI07, ASI08
5	Incumple una regla de la empresa que su desarrollador nunca conoció	Un reembolso de $150 o más sin un responsable	Tus políticas de OPA y de Cedar†, consultadas antes de cada llamada; las reglas de tu kit de agentes, como el de Microsoft†	ASI02
6	Se definió más permisivo que tus reglas	Permitido $300 por reembolso; tu regla dice $150	La llamada se detiene por tu regla. D4A señala el agente y muestra dónde ajustarlo†	
7	Dos conjuntos de reglas no coinciden, o uno no dice nada	Tu OPA aún no tiene regla para una herramienta nueva	Tu selector decide cuánto cuentan tus reglas; Explicit† detiene lo que ninguna regla permite; el informe de desacuerdos muestra los dos conjuntos lado a lado†	
8	Lo convencen: un cliente, un correo o un documento	Un mensaje dice «ignora tus límites»	Los filtros de contenido de tu plataforma controlan las palabras; los límites se mantienen diga lo que diga el agente	ASI01
9	Se usan mal sus llaves: filtradas, o prestadas a otro agente	Una llave de servidor pegada en la configuración de un agente	Cada agente llama con su propia llave, comprobada en cada llamada; tus credenciales se quedan en tu Key Vault, nunca en nuestra consola	ASI03

† en desarrollo

#	El problema	En una tienda online	Qué lo resuelve	OWASP
Antes de la llamada (sigue)				
10	Un servidor de herramientas cambia sin avisar	Aparece una herramienta nueva, delete_customer, en tu servidor de CRM	Nada nuevo entra hasta que lo firme una persona. Si cambia una herramienta ya firmada, D4A la retiene cuando llega una política nueva; y la vigila siempre, también tras un reinicio†	ASI04
11	Alguien que no debe cambia las reglas	Instrucciones editadas después de aprobarlas; un revisor intenta aflojar un límite	Solo una firma cambia lo que se aplica, y solo la autoridad de políticas firma. Quien revisa no puede cambiar las reglas	ASI03
12	Funciona un agente que nadie aprobó	Un agente que alguien montó y nadie firmó	D4A lista todos los agentes; uno que nadie firmó enciende «te necesita» y no queda gobernado hasta que lo firme una persona	ASI10
13	Usa una herramienta que D4A no ve	Una herramienta propia de la plataforma que no pasa por MCP	D4A no la ve, y lo dice: esa ruta aparece como «No gobernada», nunca se oculta, para que la cubras con los controles de tu plataforma	ASI05
Mientras se ejecuta				
14	Falla algo de lo que depende	El servidor de pagos se cuelga; tu OPA no responde a tiempo	La llamada se corta y el agente siempre recibe respuesta; tu selector decide si seguir bajo los límites firmados o parar (Strict†)	ASI08
15	Da vueltas en círculo	La misma consulta de un pedido, con los mismos datos, cincuenta veces seguidas	La misma llamada con los mismos datos no más de 5 veces†; no más de 50 llamadas por hora†	ASI02, ASI08
Con el tiempo				
16	Herramienta correcta, uso equivocado	Un descuento permitido al cliente equivocado	Un límite: el agente actúa solo sobre lo del cliente que pide†	ASI02
17	Se aparta de su tarea	Deja de ofrecer alternativas y solo da descuentos	El motor de trayectoria de DiaCroma y los jueces de adherencia a la tarea, con un caso para quien firmó†	ASI10
18	Le envenenan la memoria o la conversación	Un mensaje temprano dirige todos los pasos siguientes	Cada paso sigue cumpliendo tus límites; the Watch ve el patrón a lo largo de la sesión†	ASI06

† en desarrollo

#	El problema	En una tienda online	Qué lo resuelve	OWASP
Con el tiempo (sigue)				
19	Salen datos sensibles	Una respuesta lleva el teléfono de otro cliente	Evaluadores de riesgo, como los de Microsoft†; un límite de cuánto puede exportar por llamada	
20	La factura de tokens que nadie ve venir	Los tokens que gasta cada reembolso se doblan en una semana	Lo que cuesta cada acción en tokens, y un aviso a quien firmó el agente†; un aviso cuando una herramienta devuelve una respuesta tan grande que encarece el paso siguiente†	
Después				
21	Nadie se entera	Un reembolso detenido a las 3:00 de la madrugada	El límite se mantiene sin que nadie haga clic, y se enciende «te necesita»; un caso que le llega a quien firmó†; Teams y correo†	
22	Nadie lo lee	La alerta número cuarenta de la semana	Casos agrupados por agente y motivo, para que las repeticiones no hagan ruido	ASI09
23	Nadie puede demostrar qué pasó	El regulador pregunta por marzo	El registro firmado; el paquete para el auditor†	
En cualquier momento				
24	Hay que pararlo ya	Hay que retirar un agente hoy	Retíralo: su plaza queda libre y the Boundary deja de atenderlo	ASI10

Tus reglas de negocio pueden ser tan complejas como tu negocio.
Nosotros las hacemos cumplir a tus agentes.

† en desarrollo

Por qué MCP es *el lugar donde decidir.*

D4A decide en MCP porque es el único lugar donde cada acción de un agente aparece con su nombre y sus argumentos antes de que ocurra.

Es donde está la acción

Las palabras de un agente van al modelo; sus acciones van a las herramientas. MCP lleva las acciones: qué herramienta y con qué argumentos. Decidir ahí es juzgar lo que el agente hace. Lo que dice es tarea de los filtros de contenido.

Es el estándar abierto

Anthropic presentó MCP en 2024 y en diciembre de 2025 lo donó a la Agentic AI Foundation, dentro de la Linux Foundation, con el apoyo de Google, Microsoft, AWS, Cloudflare y Bloomberg. Plataformas como Microsoft Foundry y Copilot Studio ya lo hablan, así que una sola Boundary cubre agentes construidos con herramientas distintas.

Nada que cambiar en el agente

Ninguna librería en cada agente, nada que un desarrollador pueda olvidarse de llamar. Lo único que se toca es la dirección de su herramienta MCP: la consola te enseña el cambio, no escribe nada hasta que confirmas, y se puede deshacer.

Dónde más se podría decidir

Dónde	Qué ve	Qué se le escapa
Una librería dentro de cada agente	Lo que ve el propio agente	Cada agente tiene que incluirla, y se ejecuta donde el agente puede razonar para esquivarla
Un proxy de red	Bytes y destinos	Qué acción, con qué argumentos
Los filtros del modelo	Palabras que entran y salen	La acción en sí
Los logs, después	Todo	Ya es tarde
The Boundary, en MCP	La acción, por su nombre, con sus argumentos, antes de que ocurra, fuera del agente	

Fuente: MCP joins the Agentic AI Foundation (Linux Foundation, diciembre de 2025).

† en desarrollo

Robusta y segura *por construcción.*

Fuera del agente. The Boundary se ejecuta en tu red, no en el proceso del agente. Ningún prompt puede convencerla de saltarse un límite.

Una llave propia para cada agente. Cada llamada lleva la llave del propio agente, y the Boundary la comprueba. Se rota en caliente en un minuto.

El agente nunca tiene las llaves de tus servidores. Las tiene the Boundary, que las lee de tu almacén de claves (por ejemplo, Azure Key Vault) con la identidad de tu máquina.

Cifrado. Las llamadas llegan por HTTPS, con un certificado que the Boundary renueva sola.

Solo se aplican reglas firmadas. The Boundary solo acepta una política si su firma es válida. La clave que firma vive en Azure Key Vault: no se puede exportar y no sale de él.

Sin permiso, no se ejecuta. Cada acción necesita un permiso de un solo uso que no se puede repetir. Las repeticiones se rechazan.

Toda llamada recibe respuesta. Una herramienta que se cuelga se corta, y al agente se le avisa; nunca se queda esperando.

Toda decisión queda registrada. Las paradas, los pases y los «habría parado» entran en una cadena firmada en tu nube.

Que sea la única puerta. Configura tus servidores de herramientas para que solo acepten la credencial de the Boundary, y la caja para que solo encamine hacia ella. Así no queda otro camino hasta la herramienta.

Las partes aburridas son el producto

Cuenta el tipo, no la herramienta. En una ejecución grabada, un agente paró los reembolsos en $600 y luego pagó $395 como abonos. D4A cuenta todo el dinero que sale.

No hace falta lista de herramientas. Los agentes que nunca declararon sus herramientas se gobiernan por lo que realmente llaman.

Nada nuevo se cuela. Una herramienta añadida después de la aprobación, o una que nadie clasificó, no llega al agente hasta que una persona la aprueba.

Honesto con los resultados. Un tiempo agotado después de enviar se registra como desconocido, no como fallido, para que nadie pague un reembolso dos veces.

Nunca a medio montar. Si no puede sostener sus claves, sus límites firmados o su almacén, no abre.

† en desarrollo

Corre en tu nube. *Lo que permite nunca sale de ella.*

The Boundary es un contenedor en tu red. Nuestra consola ve lo justo para mostrarte qué pasa; el contenido de cada acción permitida, tus credenciales y tus políticas se quedan contigo.

Lo que ve nuestra consola

- Nombres de agentes y herramientas, nunca valores
- Los límites firmados y si se aplican
- Cada paro: la herramienta, sus argumentos, el motivo
- Cuántas habría parado, y un latido

Lo que nunca nos llega

- El contenido de cada acción permitida
- Tus credenciales y tus políticas
- Tus modelos y tus bases de datos

Sigue funcionando

Si no se nos puede alcanzar, D4A sigue haciendo cumplir lo que firmaste.

Tus credenciales se quedan contigo

Las declaras en tu propia instalación, nunca a través de nuestra consola. Donde tu nube lo permite, usa la identidad de tu host en su lugar.

Cuánto añade a cada llamada

En general, milisegundos. Con las políticas de tu empresa, el caso extremo es de dos segundos.†

Cuánto se guarda el registro

D4A no borra el registro firmado. Qué se purga, y cuándo, lo decides tú desde el producto, según tu región y tu regulación.†

Demuestra que el registro no se alteró, *no que el agente razonara bien.*

D4A no detecta alucinaciones, no garantiza que el agente razone bien y no promete que sea criptográficamente imposible saltárselo. Hace cumplir lo que firmó una persona en cada acción que pasa por él, y lo deja firmado.

† en desarrollo

Quién hace qué. *Cuatro papeles, separados a propósito.*

Quien revisa un caso nunca puede reescribir la regla que hay detrás. Y ningún agente cambia sus propios límites: viven fuera de él, firmados, y nada que se escriba en la conversación puede sobrescribirlos.

El propietario

Contrata D4A en el marketplace de tu nube (por ejemplo, Microsoft Marketplace), entra con la cuenta de la empresa, conecta tus agentes e invita a tus colegas.

- Creas tu organización con la cuenta de tu empresa.
- Invitas a un colega como revisor y a otro como autoridad de políticas.
- Cada cambio de papel queda escrito en el registro.

La autoridad de políticas

La persona que firma lo que un agente puede hacer.

- D4A propone límites a partir de las instrucciones del agente, cada uno con su frase exacta.
- La autoridad de políticas los revisa y los firma.
- Desde entonces, los casos de ese agente le llegan a ella.†

El revisor

Lo lee todo, pero no puede reescribir las reglas.

- Un reembolso detenido a las 3:00 de la madrugada llega a la bandeja de revisión.
- El revisor lo marca como «parada correcta» o lo libera, con su firma.†
- Quien revisa nunca cambia la política: los dos papeles siguen separados.

El auditor

Comprueba qué pasó sin preguntar a nadie.

- El auditor pregunta qué hicieron tus agentes en marzo.
- Exportas ese periodo desde tu propia Boundary, con una cabecera sellada.†
- Él comprueba cada decisión con un verificador abierto† y la clave pública de D4A: quién actuó, qué tenía permitido, qué hizo y por qué sí o por qué no.

† en desarrollo

The Watch te dice dónde mirar. *Una persona decide.*

The Watch† revisa el patrón entero, siempre activo y sin gastar tokens propios: la deriva que se acumula, agentes que se mueven a la vez y una misión que se deja de cumplir; tus jueces de Microsoft añaden su nota, en tu propio modelo de Azure OpenAI. Nunca para una acción ni cambia un límite: eso lo hace una persona.

Lunes	Martes	Miércoles	Miércoles, 11:02	Miércoles, 11:20
Descuento máximo al cliente 4471. Permitido.	Máximo otra vez, mismo cliente. Permitido.	Máximo otra vez. Cada uno dentro de su límite. Permitido.	The Watch abre un caso con las acciones que forman el patrón. Se enciende «te necesita».	Una persona aprieta el límite, firmado con su nombre.

Una semana ilustrativa: diez agentes de descuentos, cada uno dentro de sus límites, y uno de ellos ha dado el máximo al mismo cliente tres días seguidos.

Autoriza a mano

Una acción parada se aprueba una vez, con firma y atada a esa acción exacta, y el agente la reintenta. Ninguna llamada se queda retenida mientras tanto.†

El caso, a quien firmó

Cada caso le llega a quien firmó el límite,† en la consola y también por Teams o por correo.† Hoy, el caso se abre en la consola y se enciende «te necesita».

Sin permiso, no se ejecuta

Cada acción necesita además un permiso de un solo uso que no se puede repetir. Las repeticiones se rechazan.

El paquete para tu auditor

El regulador pregunta qué hicieron tus agentes en marzo. Le entregas un fichero, y lo verifica con una herramienta abierta, sin llamarte a ti ni a nosotros.† La cadena firmada está hoy en D4A y vive en tu nube.

d4a-2026-03.audit · cabeza sellada · ES256

quién actuó	billing-agent
permitido	reembolsos hasta 250 · firmado por una persona
hizo	issue_refund 320
decisión	parado · por encima del límite firmado
tus reglas	OPA: no
verificador	código abierto · válido

Así se lee una decisión en el paquete para tu auditor†. Ejemplo ilustrativo. La exportación por periodo, la cara legible por decisión y el verificador abierto están en desarrollo.

† en desarrollo

Tipos de agente. *Cinco tipos de efecto.*

Siete agentes de uso diario y un equipo cubren los cinco tipos de efecto: dinero que sale, dinero que entra, datos que salen, mensajes que salen y registros que cambian. Cada herramienta se asigna a uno, y D4A cuenta el tipo, no la herramienta.

Agente de facturación dinero que sale

- Hace reembolsos de hasta $250 y abonos de hasta $100, con $2.000 en total.
- La regla de tu empresa dice que los reembolsos de $150 o más necesitan a un responsable.
- Un reembolso de $150 consulta primero a OPA; OPA dice que no; la llamada se detiene y el motivo queda firmado.

Agente de cobros mensajes que salen, registros que cambian, dinero que sale

- Lista facturas vencidas, envía recordatorios, aplaza facturas y da deudas por perdidas.
- Un aplazamiento nunca puede pasar de 30 días.
- Dar una deuda por perdida borra la factura entera, así que el total se vigila por cuenta, no por llamada.

Agente de envíos registros que cambian

- Notifica unidades perdidas en un envío.
- No más de 25 unidades en una sola notificación.
- Diez notificaciones pequeñas sobre el mismo pedido en un día aparecen como un patrón para que lo vea una persona.†

Agente de pedidos registros que cambian

- Modifica y cancela pedidos en tu tienda.
- Cancelar un pedido de más de $500 lo decide tu regla de Cedar, no el agente.†
- El agente sigue trabajando en todo lo demás.

† en desarrollo

Agente de compras dinero que sale hacia proveedores

- Emite órdenes de compra a tus proveedores.
- $1.000 por orden, y $5.000 al mes por proveedor.†
- El total del mes se cuenta en the Boundary, sumando a todos los agentes que compran.

Agente de soporte mensajes que salen, datos que salen

- Responde a los clientes y consulta sus pedidos.
- Nunca exporta más de 50 filas de clientes de una vez. Nunca escribe fuera del dominio de tu empresa, con tu regla de Cedar.†
- Una respuesta con datos de otro cliente se marca para que la vea una persona.†

Agente de mailing mensajes que salen, datos que salen

- Envía avisos de pedidos y campañas a tus clientes, a partir de tus datos de clientes.
- Como mucho 50 correos por hora, y solo a clientes que lo aceptaron.†
- En el 51 se detiene, y la parada queda en el registro firmado.

Agentes de descuentos, como equipo dinero que entra

- Diez agentes dan descuentos: $100 por operación y $300 al día cada uno, $500 al día entre todos.†
- A las 11:40 van por $400 y tres piden $100 en el mismo segundo.
- Uno pasa y dos se detienen. Nadie agota un presupuesto compartido.

Ejemplos ilustrativos: los importes y las reglas son los que firmaría una persona en tu empresa.

Cómo se instala *y se opera.*

Desde Microsoft Marketplace, con 30 días gratis: un contenedor en tu red y tu sesión de Microsoft. Seis pantallas, sin código. Tus agentes, tus reglas y tu negocio no cambian.

1
Store
Consíguelo en Marketplace.

2
Discover
Encuentra tus agentes.

3
Scan
Propone límites, con su cita.

4
Approve
Tú firmas el mandato.

5
Gateway
Un contenedor en tu red.

6
Health
Gobernado o no, medido hoy.

Los límites no se escriben desde cero

D4A lee las instrucciones y las herramientas de cada agente, y propone límites, cada uno con la frase exacta de la que sale. Una persona los revisa y los firma. Solo se aplican los límites firmados, y no cambia nada en Azure hasta que lo aprueba una persona.

Empieza sin parar nada

En Observing, D4A cuenta lo que habría parado y no bloquea nada. Lo eliges al firmar, y después decides con qué dureza para:

Observing — Enforcing — Strict† — Explicit†

Health, medido en vivo

Cada agente, en una de tres luces:

Gobernado | Necesita atención | No gobernado

Un agente que nadie ha firmado no cuenta como gobernado. Su luz no se pinta en rojo: enciende «te necesita», porque espera tu firma; no es una avería.

Plazas que se ocupan y se liberan

Un agente ocupa una plaza cuando apruebas su política, no cuando lo escaneas. Retíralo y su plaza queda libre en ese mismo instante. Volver a aprobar un agente que cambia nunca ocupa una segunda plaza.

Lo único que cambia en tu agente de Foundry: la dirección de su herramienta MCP

```
tool = MCPTool(
    server_label="payments",
    server_url="https://<your-boundary>/mcp",
    require_approval="never",
)
```

† en desarrollo

Anexo A · *La comparativa a fondo.*

Hoy los agentes se protegen de tres maneras: un **gateway de IA** delante del modelo, un **gateway MCP** delante de las herramientas, o una **plataforma de seguridad de agentes**, que encuentra tus agentes, detecta amenazas y bloquea acciones de riesgo. Cada una hace bien su trabajo, y D4A no sustituye a ninguna. Añade una cosa: antes de que una acción se ejecute, comprueba que cabe en los límites que firmó una persona de tu negocio, contando todo lo que ya se hizo.

Lo básico: lo que esperas de cualquiera de las tres

Lo básico	En D4A
Qué herramientas puede usar cada agente	✓ Solo las que firmó una persona. Una herramienta nueva no entra sin firma.
Una llave propia para cada agente	✓ Se comprueba en cada llamada, y se rota en caliente en un minuto.
El agente no tiene las llaves de tus sistemas	✓ Las guarda the Boundary, leídas de tu Key Vault con la identidad de tu máquina.
Un registro de cada llamada	✓ Firmado y en tu nube: nadie puede reescribir lo que hizo un agente sin que se note.
Un aviso a tu gente cuando algo se para	✓ En la consola: el caso se abre y se enciende «te necesita». Por correo y por Teams.†
Límites de uso	✓ Por llamada y en total: ninguno de más de $250, y no más de $2.000 en total. Por periodo: llamadas por hora, y dinero por cliente y por mes.†
Una herramienta que cambia se detiene	✓ La retiene cuando llega una política nueva. La vigila siempre, también tras un reinicio, y una persona aprueba el contrato nuevo.†
Una persona aprueba antes de ejecutar	Se aprueba una vez, y el agente lo reintenta.†
Datos sensibles en lo que entra y sale	**Lo aportan las herramientas de tu nube** (p. ej., el evaluador «Sensitive Data Leakage» de Microsoft Foundry, en preview). D4A lo usa en tu nube, y the Watch avisa a una persona.†
Un techo a los tokens	**Con tu gateway.** Lo pone tu gateway de IA (Foundry o API Management), y D4A no lo duplica. D4A avisa a quien firmó el agente cuando sus tokens suben.†
Corre en tu entorno	✓ Un contenedor en tu red. Si no se nos puede alcanzar, sigue haciendo cumplir lo que firmaste.
Precio publicado y prueba gratis	✓ En Microsoft Marketplace, con 30 días gratis.

Hoy = en producción, en una muestra, con operador. Nadie externo lo ha comprado ni instalado todavía. † En desarrollo.

Frente a un gateway de IA

Se pone delante del modelo y de las herramientas: controla quién entra y cuántos tokens se gastan.

	Tu gateway de IA	Con D4A
Lo que el agente hace con tu dinero	Decide si el agente puede usar la herramienta de reembolsos	Decide si este reembolso, por este importe, cabe en lo que firmaste: ninguno de más de $250, no más de $2.000 en total, y por cliente y por mes†
Quién pone las reglas	Un administrador las configura	Una persona con autoridad las firma, y quien revisa los casos no puede cambiarlas
La prueba	Un registro de las peticiones	Una cadena firmada de cada decisión, en tu nube
La prueba para tu auditor	La exportas y la explicas tú	Tu auditor la comprueba sin preguntar a nadie†
Tus reglas de empresa	Las vuelves a escribir en el gateway	D4A le pregunta a tu OPA, o a Cedar†, antes de cada acción
El agente que se desvía poco a poco	Ve cada petición por separado	The Watch ve la tendencia: cada paso permitido, la suma no†
Dónde corre	Un clúster que montar	Un contenedor en tu red

Frente a un gateway MCP

Se pone delante de las herramientas: decide qué herramientas usa cada agente y protege lo que entra y sale.

	Tu gateway MCP	Con D4A
Lo que el agente hace con tu dinero	Decide si la herramienta está permitida para ese agente	Decide si esta acción, con estos datos, cabe en lo que firmaste, y lleva la cuenta del total; también por cliente y por mes†
Quién pone las reglas	Un administrador las configura	Las propone D4A a partir de las instrucciones del agente, las firma una persona, y quien revisa no puede cambiarlas
La prueba	Un registro de llamadas	Una cadena firmada de cada decisión, en tu nube
La prueba para tu auditor	La exportas y la explicas tú	Tu auditor la comprueba sin preguntar a nadie†
Tus reglas de empresa	Las vuelves a escribir en el gateway	D4A le pregunta a tu OPA, o a Cedar†, antes de cada acción
Lo que cambia con el tiempo	Vigila si cambia una herramienta	The Watch vigila si cambia el agente: tendencia, agentes que se mueven juntos, misión que se deja de cumplir†

† en desarrollo

Frente a una plataforma de seguridad de agentes

Descubre los agentes de tu empresa, detecta amenazas y bloquea acciones en el momento. D4A no la sustituye: se pone a su lado y añade lo que no cubre.

	Tu plataforma de seguridad	Con D4A
Lo que el agente hace con tu dinero	Bloquea por herramienta, tipo de acción o destino	Decide si este reembolso, por este importe, cabe en lo que firmaste, y lleva la cuenta del total; también por cliente y por mes†
Quién pone las reglas	El equipo de seguridad, o un asistente de IA que las redacta	Las propone D4A a partir de las instrucciones del agente, las firma una persona del negocio, y quien revisa no puede cambiarlas
Tus reglas de empresa	Las vuelves a escribir en su consola	D4A le pregunta a tu OPA, o a Cedar†, antes de cada acción
La prueba	Un registro de auditoría	Una cadena firmada de cada decisión, siempre, en tu nube
La prueba para tu auditor	La exportas y la explicas tú	Tu auditor la comprueba sin preguntar a nadie†
El agente que se desvía	Lo compara con su comportamiento habitual	The Watch lo compara con lo que se le encargó: un agente que se porta mal desde el primer día parece normal frente a su costumbre, no frente a su encargo†
Una persona aprueba antes de ejecutar	Algunas ya lo hacen	Se aprueba una vez, firmada, y el agente lo reintenta†

† en desarrollo

Conserva lo que tienes.
D4A le añade lo que no ve.

Tu gateway de IA controla quién usa el modelo y cuánto gasta. Tu gateway MCP controla qué herramientas usa cada agente. Tu plataforma de seguridad descubre tus agentes y detecta amenazas. Ninguno mira la acción en sí, contra lo que firmó tu negocio: este reembolso, por este importe, a este cliente, después de los de esta semana. D4A sí. Se pone encima de tu gateway y al lado de tu plataforma, sin rehacer tus agentes, y los datos de tus agentes no salen de tu nube.

Si tienes un gateway de IA

Él decide quién entra y cuántos tokens se gastan. D4A añade un límite en cada acción, que cuenta todo lo ya hecho, y avisa a quien firmó el agente cuando sus tokens suben.†

Si tienes un gateway MCP

Él decide qué herramientas usa cada agente. D4A decide si lo que va a hacer con ellas cabe en lo que firmaste. Si no cabe, lo para, abre el caso a una persona y guarda la prueba firmada.

Si tienes una plataforma de seguridad de agentes

Ella descubre tus agentes y detecta amenazas. D4A añade el límite que firma una persona del negocio, tus reglas de empresa y una prueba que tu auditor comprueba solo.†

Si tienes varios, o ninguno

D4A gobierna en un solo sitio tus límites: los de tu empresa y los de cada agente, y también los de cada equipo.† Y le da a tu auditor una prueba que nadie puede reescribir.

Otras guardas que conservas

Filtros de contenido de Foundry

Controlan las palabras que entran y salen del modelo. Déjalos al máximo.

Presupuestos de Foundry

El techo de tokens y de coste. D4A no lo duplica.

Entra Agent ID

Quién es cada agente. D4A añade qué tenía permitido.

Tu jaula

Dónde corre el agente: ficheros, red y procesos.

Si un agente está definido más flojo que las políticas de tu empresa, *mandan tus políticas.* Si no tienes políticas, mandan los límites del agente, firmados por una persona.

Un agente no responde ante la ley. Tú sí, y además lo pagas. Quédate con lo que tienes y añádele los límites.

† en desarrollo

Anexo B · *Todo lo que hace, en una lista.*

Para quien quiere comprobarlo punto por punto. Lo que no lleva † está hoy en D4A; lo que lleva † está en desarrollo.

Antes y en cada llamada

- Encuentra tus agentes de Foundry y sus herramientas desde tu sesión de Microsoft.
- Propone cada límite a partir de las instrucciones del propio agente, con su frase. Firma una persona.
- Comprueba tus políticas OPA y Cedar† antes de cada acción.
- Observing o Enforcing, elegido al firmar. Strict y Explicit con tus políticas.†
- Límites por llamada, y un total acumulado por agente a lo largo de su vida.
- Límites por hora, día y mes; por cliente; entre todos los agentes de una tarea.†
- Corta los bucles: la misma llamada con los mismos datos, más veces de las que firmas.†
- Cinco tipos de efecto; cada herramienta asignada a uno.
- Totales compartidos por los agentes que sirve una instalación; sobreviven a reinicios y a nuevas firmas.
- Las herramientas que nadie aprobó se rechazan; un agente cambiado se detecta.
- Los agentes sin lista de herramientas se gobiernan.
- Un permiso de un solo uso por acción; las repeticiones se rechazan.
- Respuesta en menos de dos segundos, contando tus políticas.†

Siempre

- Una acción parada nunca se ejecuta y abre un caso para revisión.
- Señala a un agente cuyas reglas son más flojas que tus políticas.†
- El informe de desacuerdos: tus políticas y las reglas del agente, lado a lado.†
- The Watch revisa el patrón entero y llama a una persona.†
- Cuánto cuesta cada acción en tokens. Avisa a quien firmó un agente cuando sus tokens suben, o cuando una herramienta devuelve una respuesta tan grande que encarece el siguiente paso.†
- Cada decisión en un registro firmado.
- La cabeza de la cadena firmada con ES256, comprobada con nuestra clave pública; una línea legible por decisión; cada límite marcado como garantía firme o mejor esfuerzo.†
- Health: Gobernado, Necesita atención, No gobernado, medido en vivo.
- Corre en tu nube; sigue haciendo cumplir si no se nos puede alcanzar.
- Autoservicio en Microsoft Marketplace.
- Esta versión corre en Microsoft. Otras nubes, plataformas de agentes y tiendas vienen después.

Anexo C · *Lo que preguntará tu equipo de IT.*

¿Quién decide que una acción se para?

Ningún modelo de lenguaje. Un modelo solo lee las instrucciones del agente y propone límites, cada uno con la frase de la que salió. Una persona los firma. La comprobación en sí es aritmética simple, y se puede recalcular.

¿De dónde salen los límites?

De las propias instrucciones del agente: «nunca reembolses más de $200 sin una persona» ya es un límite. D4A lo propone con la cita, y una persona lo firma. Tus reglas OPA y Cedar se suman.†

¿Qué significa «en total»?

Un total acumulado por cada tipo de efecto: dinero que sale, dinero que entra, datos que salen, mensajes que salen, registros cambiados. Vive fuera del agente y sobrevive a los reinicios. Firmar un límite nuevo no pone a cero lo ya gastado, y dos agentes que llegan en el mismo instante no pueden colarse los dos por debajo del techo. Los totales por hora, por día y por cliente funcionan igual.†

¿Qué pasa cuando una acción se para?

Nunca llega a tu sistema. El agente recibe el motivo («solo reembolsos hasta $250») y puede decírselo a su usuario, y se abre un caso para una persona. Cada acción necesita además un permiso de un solo uso que no se puede repetir: sin permiso, no se ejecuta.

¿Puede un agente cambiar sus propios límites?

No. Los límites viven fuera del agente, firmados. Nada que se escriba en la conversación puede sobrescribirlos.

¿Y lo que el agente gasta en tokens?

D4A no lo para: el techo de tokens se pone en la plataforma del modelo (p. ej. Microsoft). Lo que hace es decirte cuánto cuesta en tokens cada acción, avisar a quien firmó el agente cuando sus tokens suben, cortar los bucles de llamadas repetidas y avisar cuando una herramienta devuelve una respuesta tan grande que encarece el siguiente paso.†

¿Qué agentes y herramientas gobierna?

En esta versión, los agentes de Microsoft Foundry, y toda herramienta que les llegue por MCP, aunque el agente no declare su lista. Otras plataformas vienen después.

† en desarrollo

¿Cómo convive con nuestro gateway?

Se pone encima del que tengas. Tu gateway decide quién entra y qué herramientas usa cada agente; D4A decide si cada acción cabe en lo que firmaste.

¿Quién puede hacer qué?

Invitas a tu equipo con un rol: la autoridad firma los límites y el revisor lo lee todo. Autorizar a mano una acción parada y cerrar su caso: la autoridad y el revisor.† Tu auditor recibe el paquete firmado de cada periodo.†

¿Una persona puede autorizar a mano una acción parada?

Sí. La aprueba una vez, con su firma y atada a esa acción exacta, y el agente la reintenta. Ninguna llamada se queda retenida mientras tanto.†

¿Qué pasa si una herramienta cambia?

Si cambia lo que la herramienta dice que hace, D4A la retiene cuando llega una política nueva. Vigilarla siempre, también tras un reinicio, y que una persona apruebe el contrato nuevo.†

¿Cuánto añade a cada llamada?

En general, milisegundos. Con las políticas de tu empresa, el caso extremo es de dos segundos.†

¿Cuánto tiempo se guarda el registro?

D4A no borra el registro firmado. Qué se purga, y cuándo, lo decides tú desde el producto, según tu región y tu regulación.†

¿Cómo se instala?

Desde Microsoft Marketplace, con 30 días gratis: un contenedor en tu red y tu sesión de Microsoft. Tus agentes no cambian; D4A cambia la dirección de su herramienta MCP. Para gobiernos, grandes empresas y partners, también desde tu propia instalación, bajo acuerdo.†

¿Otra pregunta? *Escríbenos.*

diacroma@veritglobal.com · www.diacroma.com

† en desarrollo

Anexo D · *Cómo se compra.*

Pagas por los agentes que gobiernas. Una cuota mensual fija por un número de agentes gobernados. Las llamadas gobernadas son ilimitadas y nunca se miden, y leer un agente no cuesta nada extra. Cada plan empieza con 30 días gratis: cancela antes de que terminen y no se cobra nada.

Te suscribes en Microsoft Marketplace

Se factura con tu factura de Azure, bajo el acuerdo comercial de Microsoft que ya tienes. Inicias sesión con tu cuenta de Microsoft y tu organización se crea sola. Los planes y precios actuales solo se publican allí.

Cuándo se ocupa y se libera una plaza

Al aprobar su política, no al escanearlo. Retíralo y queda libre al instante. Un agente que no se puede gobernar aparece en la lista y nunca se factura. El recuento sale de lo que está firmado de verdad.

Sube cuando se te quede pequeño

Nada se rompe al llegar al techo: lo ya gobernado sigue aplicándose y solo espera el siguiente agente nuevo. Cambias de plan desde la consola. Más allá del plan más grande, una oferta privada.

Para clientes y partners con necesidades únicas

Opera D4A desde tu propia instalación, en lugar de Microsoft Marketplace, bajo acuerdo.† D4A instalado en tu propio Azure, operado por tu equipo o por un partner de confianza; los casos, lo que D4A sabe de tus agentes y las claves que firman cada decisión se quedan contigo. Para gobiernos, grandes empresas y partners estratégicos.

El kit para partners†

Quien opera D4A puede añadirle lo que solo él sabe y ajustar cómo se comporta. Cómo decide D4A no cambia, y cada límite lo sigue aprobando una persona.

Añade tu propio «no»

Tu herramienta de seguridad puede parar una acción que D4A permitiría. Nunca puede dejar pasar una que D4A para.

Tus paquetes, por sector

Escribe una vez los límites y los ajustes de un sector, para clínicas o para bancos, como un preset firmado. Los aprueba la gente de cada cliente.

Cada caso, donde ya trabajas

Cada parada y cada caso llegan a los sistemas que tu equipo ya vigila.

Una prueba que cualquiera comprueba

Exporta un periodo y compruébalo con una herramienta abierta. Cada decisión dice con qué configuración se tomó.

Si construyes agentes · integrado u OEM

Tú construyes el agente. Juntos, lo entregamos gobernado: integramos D4A con tu plataforma de agentes, con la gobernanza incorporada, no añadida.

Habla con nosotros

diacroma@veritglobal.com
www.diacroma.com

† en desarrollo

Anexo D · *A tu medida.*

D4A es muy configurable. Funciona solo desde el primer minuto: cientos de decisiones vienen con un valor por defecto, cada una con su porqué. Te quedas con ellos o los ajustas en tres sitios, según tu licencia.†

Settings

Para todo cliente, con cualquier plan. Decenas de decisiones sobre cómo gobierna tu empresa. Por ejemplo:

- El modo: viene marcado Observing y te recomendamos Enforcing.
- Si un agente cambiado se para o solo avisa, y si un borrado que la nube a veces puede deshacer cuenta como irreversible.
- Adónde van los avisos, por correo o en Teams, tu zona horaria y cuánto se guarda el registro.

SDK de tu organización

Licencias enterprise y de gobierno, y partners. Más de cien ajustes finos del motor y de la Boundary. Por ejemplo:

- Cuánto espera D4A a tus reglas.
- Cuánta deriva avisa o para a un agente.
- Cuántas llamadas por minuto puede hacer tu organización.

SDK de operador

Para quien opera su propia consola: una empresa, un gobierno o un partner. Decenas de ajustes de cómo funciona esa consola. Por ejemplo:

- Qué modelo usa el escaneo.
- Cada cuánto comprueba la consola todas las fronteras.
- Cuántas operaciones pesadas lleva a la vez y desde qué buzón salen los avisos.

Se aplican cuatro capas: el valor por defecto de D4A, el preset de un partner, tu empresa y cada instalación, y gana la más concreta. Cada cambio se firma con quién, cuándo y por qué, y toda la configuración se exporta en un fichero JSON legible.

Lo que nunca se configura

Lo que protege una garantía no está en ninguna pestaña: toda firma se comprueba, lo que no se puede evaluar se para y la cadena firmada no se borra para que algo parezca verde. Los planes y los precios tampoco son ajustes.

† en desarrollo

Anexo E · *Las herramientas que D4A reúne.*

Cada herramienta del diagrama: qué es, un ejemplo de tres líneas y, cuando existe, el código real.

Dónde viven tus agentes

MCP (Model Context Protocol). El estándar abierto que usan los agentes para llamar a herramientas.

- Tu agente de facturación llama a issue_refund en tu servidor de pagos por MCP.
- D4A se coloca en ese camino. El agente y el servidor no cambian; solo cambia la dirección.
- Desde entonces, cada llamada pasa por un punto que puede decir sí o no, y abrir un caso para una persona.

Lo que envía el agente, y lo que D4A lee antes de que se mueva nada:

```
{"jsonrpc": "2.0", "id": 7, "method": "tools/call",
  "params": {"name": "issue_refund", "arguments": {"order_id": "A-1042", "amount": 150}}}
```

Inicio de sesión de Microsoft. Tu cuenta de trabajo de Microsoft: con ella D4A encuentra tus agentes.

- Inicias sesión una vez con tu cuenta de trabajo.
- D4A lista todos los agentes de tus proyectos y las herramientas a las que llega cada uno.
- No cambia nada en Azure hasta que lo aprueba una persona.

Microsoft Foundry. Donde se construyen y ejecutan los agentes de esta versión. Después llegarán otras plataformas.

- D4A lee las instrucciones y las herramientas de cada agente.
- Un modelo propone límites a partir de las instrucciones, cada uno con la frase exacta de la que sale.
- Una persona los revisa y los firma. Solo se aplican los límites firmados.

La herramienta MCP del agente en Foundry. La única línea que cambia D4A es la dirección:

```
tool = MCPTool(
    server_label="payments",
    server_url="https://<your-boundary>/mcp",
    require_approval="never",
)
```

Copilot Studio†. El constructor de agentes low-code de Microsoft.

- Un agente de ventas hecho en Copilot Studio llega a tu CRM por MCP.
- D4A se coloca en ese mismo camino, con los mismos límites firmados.
- Una sola política cubre los agentes que hicieron tus desarrolladores y los que hicieron tus equipos de negocio.

† en desarrollo

Filtros de contenido de Foundry (Azure AI Content Safety). Controlan lo que entra en el modelo y lo que sale de él.

- En nuestra prueba, el filtro detuvo un jailbreak cinco veces de cinco.
- El agente aun así superó su tope de gasto cinco noches de cinco.
- Los filtros vigilan las palabras; D4A acota las acciones. Mantén los dos al máximo.

Entra Agent ID. Una identidad de Microsoft para cada agente.

- Tu equipo de seguridad ve al agente de reembolsos como una identidad propia en Entra, no como una cuenta compartida.
- Una revisión de incidentes pregunta: ¿qué agente hizo esto, y tenía permiso?
- Entra responde quién; el registro firmado de D4A responde qué tenía permitido y por qué se detuvo.

Azure Key Vault. Tus credenciales se quedan en tu almacén, en tu nube.

- La llave de tu servidor de pagos vive en tu Key Vault.
- The Boundary la lee con la identidad de tu propia máquina.
- Ninguna credencial pasa nunca por nuestra consola.

La única llamada que hace the Boundary, con la identidad de tu máquina:

```
GET https://<your-vault>.vault.azure.net/secrets/payments-key?api-version=2025-07-01
```

Tus reglas

OPA (Open Policy Agent). El motor de código abierto más usado para políticas de empresa. Sus reglas se escriben en un lenguaje llamado Rego.

- D4A consulta a tu OPA antes de cada llamada, en la ruta que declaras.
- Si tu regla dice que no y tu selector la hace cumplir, la llamada se detiene, con el motivo de tu regla.
- El «no» de tu regla queda en el registro firmado.

Tu regla, y la consulta que hace D4A:

```
deny contains "refunds of 150 or more need a manager" if {
    input.tool == "issue_refund"; input.args.amount >= 150
}

POST /v1/data/<your-policy-path>
```

† en desarrollo

Cedar†. Un lenguaje de políticas de código abierto creado por AWS, pensado para ser rápido y fácil de analizar. Lo que no está permitido se deniega.

- Tu regla dice que ningún agente escribe fuera del dominio de tu empresa.
- D4A la consulta antes de cada envío, junto a tus límites firmados.
- Un correo a otro dominio no sale, y el motivo queda firmado.

```
// No agent sends email outside the company
forbid (principal, action == Action::"send_email", resource)
unless { context.to like "*@yourstore.com" };
```

Microsoft Agent Governance Toolkit†. El motor de políticas de código abierto de Microsoft para agentes. Sus reglas se escriben en YAML.

- D4A también las consulta, junto a OPA y Cedar, si responden en milisegundos.
- Una regla bloquea las acciones destructivas, como borrar o truncar.
- Su «no» se suma al de OPA, al de Cedar y al de tus límites firmados.

```
rules:
  - name: block-destructive
    condition: "action.type in ['drop', 'delete', 'truncate']"
    action: deny
```

Aprobaciones de Foundry. Foundry puede pedir a una persona que apruebe cada llamada a una herramienta.

- Tu herramienta de ERP está configurada como «preguntar siempre».
- D4A lo muestra en la fila del agente, para que nadie crea que funciona solo.
- La decisión sigue siendo tuya, en Foundry.

El ajuste que D4A lee en cada herramienta MCP:

```
tool = MCPTool(server_label="erp", server_url="https://<your-erp-server>/mcp",
               require_approval="always")
```

Límites que firmó una persona. Las reglas propias del agente, en números.

- Agente de facturación, issue_refund: como mucho $250 por llamada y $2.000 en total, y la misma llamada con los mismos datos no más de 5 veces.†
- D4A lo propuso a partir de una frase de las instrucciones del agente: «los reembolsos tienen un tope de $250».
- Lo firmó tu responsable de finanzas. Sin firma, no se aplica nada.

† en desarrollo

Los jueces automáticos (the Watch†)

Motor de trayectoria de DiaCroma†. Nuestro propio motor. Ve la deriva que ninguna llamada sola muestra, sin gastar tokens.

- Un agente de exportación puede sacar 50 filas por llamada. Al tercer día ha sacado 1.900 filas en 40 llamadas.
- Cada llamada estaba permitida. El total no era lo que nadie quería.
- The Watch abre un caso con las 40 llamadas que forman el patrón.

Lo que cuesta cada acción†. Cuántos tokens gasta cada acción, leídos de tu propia plataforma, y un aviso cuando una herramienta devuelve una respuesta tan grande que encarece el siguiente paso.

- Los tokens por reembolso de un agente de facturación se doblan en una semana.
- The Watch los suma por acción, sin gastar tokens propios.
- Quien firmó el agente recibe el aviso antes de que el mes se dispare; el techo de tokens sigue en tu plataforma.

Aquí todavía no hay código: qué datos de tokens envía cada plataforma está aún por medir.

Evaluadores de llamadas a herramientas de Microsoft†. ¿Eligió el agente la herramienta correcta, con los datos correctos?

- El agente consulta una factura que el cliente nunca nombró.
- El evaluador, que se ejecuta en tu propio despliegue de modelo en tu Azure, señala el dato equivocado.
- El hallazgo entra en el registro firmado y en la ficha del agente.

El evaluador de Microsoft, tal como lo ejecuta su SDK, y el tipo de respuesta que da:

```
from azure.ai.evaluation import ToolCallAccuracyEvaluator

tool_call_accuracy = ToolCallAccuracyEvaluator(model_config)
result = tool_call_accuracy(query=query, tool_calls=tool_calls, tool_definitions=tool_definitions)
# {"tool_call_accuracy": 3, "tool_call_accuracy_result": "fail", "tool_call_accuracy_threshold": 4, ...}
```

Task Adherence de Microsoft†. ¿Sigue el agente haciendo lo que se le pidió?

- A un agente de soporte se le pide seguir un pedido y empieza a cambiar la dirección del cliente.
- Task Adherence puntúa cuánto se aparta de su tarea.
- Una persona recibe el caso antes de que se convierta en costumbre.

```
from azure.ai.evaluation import TaskAdherenceEvaluator

task_adherence = TaskAdherenceEvaluator(model_config=model_config)
result = task_adherence(query=query, response=response)
# {"task_adherence": 2, "task_adherence_result": "fail", "task_adherence_threshold": 3,
"task_adherence_reason": "..."}
```

Evaluadores de riesgo de Microsoft†. Acciones prohibidas y salida de datos sensibles.

- Una respuesta incluye el teléfono de otro cliente.
- El evaluador señala la fuga.
- El caso llega a la persona que firmó los límites de ese agente.

Los dos evaluadores para agentes, tal como los declara Foundry (en preview):

```
{"type": "azure_ai_evaluator", "name": "Prohibited Actions",
  "evaluator_name": "builtin.prohibited_actions",
  "data_mapping": {"query": "{{item.query}}", "response": "{{item.response}}", "tool_calls": "
{{item.tool_calls}}"}}
{"type": "azure_ai_evaluator", "name": "Sensitive Data Leakage",
  "evaluator_name": "builtin.sensitive_data_leakage", ...}
```

Application Insights†. Donde ya llegan las trazas y las puntuaciones de tus agentes.

- La evaluación continua de Foundry marca las conversaciones de soporte del martes como fuera de tarea.
- The Watch pone esa puntuación junto a su propia tendencia para el mismo agente.
- Quien firmó recibe un solo caso con las dos, no dos paneles que cruzar.

Dónde deja sus puntuaciones la evaluación continua de Foundry:

```
traces
| where message == "gen_ai.evaluation.result"
| where timestamp > ago(7d)
```

Azure OpenAI†. Tu propio despliegue de modelo, en tu propio Azure.

- Para juzgar si un descuento fue al cliente correcto, un modelo tiene que leer la llamada.
- Ese modelo es tu propio despliegue, en tu propia suscripción.
- La llamada se juzga sin salir de tu tenant.

El modelo del juez es un despliegue que tú nombras:

```
model_config = AzureOpenAIModelConfiguration(
    azure_endpoint=os.environ["AZURE_OPENAI_ENDPOINT"],
    api_version=os.environ["AZURE_OPENAI_API_VERSION"],
    azure_deployment=os.environ["MODEL_DEPLOYMENT_NAME"],
)
```

Las personas y la prueba

Teams y correo†. Los avisos llegan a las personas donde ya están.

- Una parada a las 3:00 de la madrugada llega a la persona de guardia en Teams a las 3:01, con el motivo.
- La cierra como «parada correcta» desde el caso.
- Su decisión queda firmada con su nombre.

El registro firmado. Cada decisión se convierte en un hecho dentro de una cadena que vive en tu nube.

- Cada parada, cada pase y cada «habría parado» se escribe en el momento.
- El «no» de tu regla también está, con el motivo.
- Nadie puede reescribir en silencio lo que hicieron tus agentes.

Los marcos a los que respondes

D4A te da los registros y la supervisión humana que piden ISO/IEC 42001, NIST AI RMF, el Reglamento Europeo de IA (EU AI Act) y el OWASP Agentic Top 10.

ISO/IEC 42001

Sistema de gestión de IA.

NIST AI RMF

Gobernar, mapear, medir, gestionar.

EU AI Act

Registros y supervisión humana.

OWASP Agentic Top 10

Mal uso de herramientas, agentes descontrolados.

Qué significa la †

Construimos D4A a la vista. Un elemento marcado con † está en desarrollo a noviembre de 2026; todo lo que no la lleva está hoy en D4A. Hoy quiere decir en producción, en una muestra, con operador: nadie externo lo ha comprado ni instalado todavía. Acme es nuestra empresa de muestra, no un cliente: el negocio es inventado, el despliegue es real.

Afiliación

DiaCroma la construye Verit Global Labs Inc., una empresa de Delaware con sede en Parkland, Florida. DiaCroma no está afiliada a Microsoft, NVIDIA, AWS ni a ninguna de las empresas citadas; sus productos se describen a partir de su documentación pública. OPA y Cedar son marcas de The Linux Foundation. Patente en trámite.

Míralo en tus propios agentes.

D4A está hoy en Microsoft Marketplace, para agentes en Microsoft Foundry. Otras plataformas vienen después.

diacroma@veritglobal.com
www.diacroma.com