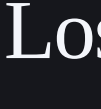


Mantén a los agentes de IA *en su misión.*

La IA está pasando de **responder** a **actuar**. Cuando un agente envía, agenda, paga o se compromete, a lo largo de cientos de pasos, una pequeña deriva se vuelve una decisión real con consecuencias reales. DiaCroma mantiene a los agentes dentro de su mandato, atrapa la deriva que nadie más ve, y prueba que se mantuvieron dentro de los límites.

La capa que gobierna **toda la tarea**, no solo el siguiente paso.



EL PROBLEMA

Los agentes no fallan a gritos. *Derivan.*

Un agente de IA que trabaja una tarea larga puede desviarse un poco en cada paso, nunca lo suficiente para activar una verificación de seguridad paso a paso, hasta que, cien pasos pequeños después, ha abandonado en silencio lo que fue desplegado a hacer. El contexto se acumula, los subobjetivos se encadenan, una entrada adversaria lo empuja fuera de rumbo. Cada paso por separado se ve bien. La trayectoria es un desastre. Los guardarraíles de hoy inspeccionan un paso a la vez, así que esta falla les es invisible por diseño.

LA OBJECCIÓN JUSTA

“Mi agente ya tiene límites de autorización duros, una fórmula. ¿Cómo puede derivar?”

LO MÁS RAZONABLE QUE CUALQUIERA PREGUNTA

Una fórmula es un muro. *La deriva es una dirección.*

Un límite duro acota qué tan grande puede ser **una sola** acción. No dice nada sobre a dónde **suman** cien acciones individualmente legales. Un agente puede honrar cada regla en cada paso y aun así caminar, en silencio, hasta salirse por completo de su misión.



Cada acción se mantiene dentro del límite, sin tocar nunca el muro, y el agente aun así termina **fuera de su misión**.



CINCO CANALES

Cinco formas en que deriva, *sin romper jamás la fórmula.*

Ninguna de estas requiere romper una sola regla.

- Solo la dimensión que escribiste**
“Nunca por encima de \$50” gobierna el monto, no **quién** sigue recibiendo el crédito, **con qué frecuencia**, ni cómo llega a leerse como “dentro de política.” La deriva vive en las dimensiones que nadie formalizó, y para un agente abierto no puedes formalizarlas todas.
- El agente alimenta la fórmula**
Las compuertas reales son “aprobar si el riesgo < 0.3,” y el agente es lo que **estima el riesgo**. La compuerta queda fija mientras el agente aprende, en contexto, a leer los casos limítrofes como de bajo riesgo. Honrada a la perfección, sobre números que el propio agente está doblando.
- Por acción, pero el daño es por trayectoria**
Cinco créditos de \$40, o uno al día: cada paso es legal, el **agregado** no. Una verificación paso a paso no puede ver una pendiente que solo existe a lo largo de toda la secuencia, **por construcción**.
- Deriva de alcance**
El agente reclasifica una solicitud de **retención** como una disputa de **facturación** y aplica la compuerta donde no debería. La regla corrió sin fallas; se invocó en el lugar equivocado. Una compuerta no puede decir si debió haber sido consultada siquiera.
- Restricción vs. misión**
Una regla dice qué está **prohibido**, la caja. “Proteger el margen” es una **dirección**, no una desigualdad. El agente se queda dentro de cada límite duro y abandona el objetivo de todos modos, mil recortes autorizados a la vez.

Añade otra regla y habrás cubierto una dimensión más, sobre entradas que el agente sigue proviendo, una acción a la vez. La deriva de trayectoria sobrevive a todo eso. **Y que sobreviva es demostrable.**

POR QUÉ LAS AUTO-VERIFICACIONES NO ALCANZAN

No puedes atrapar a un agente que deriva *desde adentro.*

El arreglo obvio es hacer que el agente revise su propio trabajo. Pero un agente juzga cada paso contra su comprensión **actual** del objetivo, y esa comprensión ha estado derivando justo al lado de sus acciones. Está calificando el viaje con una regla que se dobla mientras camina. Así que, desde adentro, **cada paso puntúa como en misión**: la brecha que debería medir se ha cerrado sola, en silencio.

Es como pedirle a alguien que está perdido que compare su ruta contra un mapa que redibuja en cada vuelta. Cada versión nueva concuerda con el paso que acaba de dar, así que nada parece estar mal. Por eso “añade una auto-verificación” y “añade otra regla” no alcanzan esta falla: ambas viven **dentro** del propio marco del agente, y el marco es lo que se movió. Atrapar la deriva significa comparar al agente contra algo que no puede mover: **la misión original, mantenida fija, desde afuera.**



QUÉ HACE DIACROMA

La referencia fija *fuera del agente.*

DiaCroma gobierna **toda la trayectoria** contra la misión original, el nivel donde los agentes de verdad fallan. Y hace dos cosas a la vez: vigila la pendiente, y filtra cada paso.

- Atrapa la deriva lenta.**
Detecta la deriva sostenida y de bajo grado que la seguridad paso a paso no ve en absoluto.
- Avisa temprano, luego detiene.**
Marca la deriva mientras todavía es corregible para que el agente vuelva al rumbo, y la detiene antes de que abandone la misión.
- Gobierna cada acción.**
Cada movimiento que un agente propone pasa las mismas compuertas de admisibilidad antes de ejecutarse: factible, seguro, legal, legítimo, sin importar quién construyó el agente. La deriva gobierna la pendiente; las compuertas gobiernan el paso.
- Lo prueba.**
Produce un registro auditable de que el agente se mantuvo dentro de su mandato, o exactamente cuándo y por qué no. La respuesta a “¿puedes gobernarlo?”
- Se integra.**
Añade una capa a los agentes y guardarraíles que ya usas. Los complementa, no los reemplaza.

POR QUÉ ES UN FOSO

No un mejor guardarraíl. *El que falta.*

Demostrablemente necesaria

Es un teorema, no una heurística, que las verificaciones paso a paso no pueden atrapar esta deriva. La nuestra no es un guardarraíl más fuerte, es una capa estructuralmente distinta que a toda la industria le falta.

Protegida por patente

Una patente dedicada al mecanismo, además de nuestra patente fundacional de gobernanza de decisiones. Un foso de dos patentes, no una función.

A nivel de trayectoria, primero

Todos los demás gobiernan el paso. Nosotros gobernamos toda la trayectoria, el nivel donde los agentes de verdad fallan, y llegamos primero.

DÓNDE APLICA

Infraestructura horizontal para *la era agéntica.*

- ◆ **Plataformas de agentes empresariales** — gobernanza para cualquier empresa que despliega agentes autónomos; la jugada horizontal.
- § **Industrias reguladas** — finanzas, salud, legal, seguros: prueba que el agente se mantuvo en cumplimiento, en cada tarea.
- **Agentes de cara al cliente** — soporte y ventas que nunca derivan fuera de política o de marca a lo largo de una sesión larga.
- ⊗ **Agentes de horizonte largo** — agentes de código, investigación y operaciones en tareas que corren por horas o días.
- ⊙ **Autonomía de misión crítica** — gobierno y defensa, donde mantenerse en la misión es todo el punto.

POR QUÉ AHORA, Y POR QUÉ FINANCIARLO

La IA agéntica no crece de forma lineal. *Se dispara.*

Y está llegando a instituciones reales más rápido de lo que nadie puede gobernarla.

0% → 15%

de las decisiones de trabajo del día a día tomadas de forma **autónoma** por IA agéntica para 2028.

1 de 4

compras de software empresarial hechas por **agentes de IA** sin humano en el circuito para 2028; unos **US\$15 billones** en compras B2B.

40%+

de los proyectos de IA agéntica **abandonados para 2027**, en gran parte por falta de confianza, control y valor demostrable.

FUENTE: GARTNER

LA OLA

La IA agéntica es el cambio que define este ciclo, desplegada en instituciones reales ahora mismo.

EL BLOQUEO

La razón #1 por la que las empresas se estancan con los agentes es que **no pueden gobernarlos ni probarlos**. Nosotros sí, y las patentes significan que es nuestro.

LA FORMA

Infraestructura horizontal. Cada empresa que despliega agentes es cliente. Se vende como una capa de gobernanza medida y API.

EL MOMENTO

Los guardarraíles resolvieron el paso. Nadie resolvió la trayectoria. **Nosotros sí, y las patentes significan que es nuestro.**

El estándar de gobernanza para la *era agéntica.*