

NeodyAI AI Gateway

Agentes de IA con datos confidenciales — de forma segura y bajo sus controles de riesgo de modelos

Resumen técnico · Octubre de 2026 · Patente en trámite

Este resumen describe el NeodyAI AI Gateway y las garantías para las que está diseñado. Se entrega por fases; contáctenos para conocer la disponibilidad para su implementación. Hay una presentación detallada de la arquitectura disponible bajo acuerdo de confidencialidad (NDA).

El problema

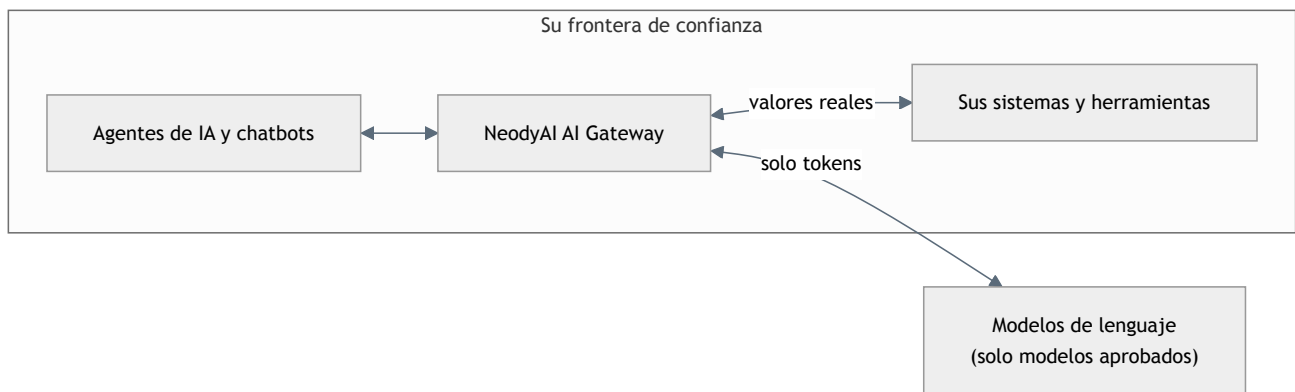
Los agentes de IA son útiles porque combinan dos cosas:

- **Un modelo de lenguaje** — normalmente operado por un tercero, fuera de su organización — que razona sobre el trabajo.
- **Sus sistemas** — CRM, pagos, siniestros, core bancario, documentos — donde el trabajo realmente ocurre.

El modelo no necesita los números de tarjeta, números de cuenta, información de salud ni nombres reales de sus clientes para razonar sobre una tarea. Sus sistemas no pueden actuar sin ellos. Y si usted es una institución regulada, cada modelo que influye en una decisión debe estar inventariado, validado, monitoreado y cambiarse bajo control — incluido un modelo de un tercero que su proveedor puede actualizar en cualquier momento.

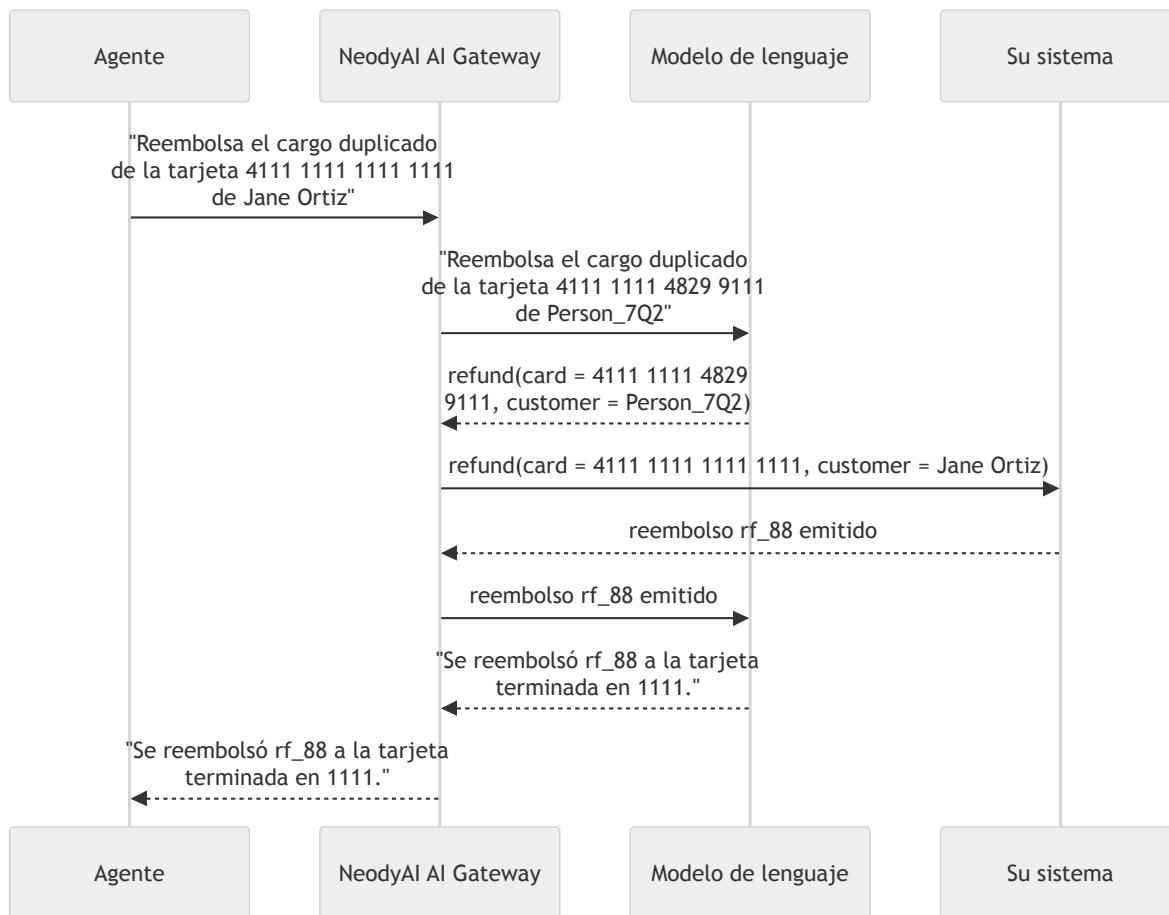
Qué hace el gateway

El NeodyAI AI Gateway se ubica en cada camino entre sus agentes, los modelos de lenguaje que usan y sus sistemas.



1. **Los datos protegidos nunca llegan al modelo.** Antes de enviar algo a un modelo de lenguaje, el gateway encuentra los valores protegidos — datos de tarjeta, información de salud, datos personales, números de cuenta, información confidencial del negocio, credenciales — y los reemplaza por **tokens**.
2. **Sus sistemas reciben los valores reales.** Cuando el modelo pide una acción, el gateway restaura los valores reales y los entrega solo a sus sistemas aprobados, actuando con los permisos de la persona para la que trabaja el agente.
3. **Los resultados se protegen en el camino de regreso.** Los datos que devuelven sus sistemas se convierten en tokens antes de que el modelo los vea.
4. **Cada llamada se gobierna y se registra.** Solo se puede llegar a modelos aprobados, cada uno aprobado para tipos de datos específicos, y cada llamada deja un registro de evidencia — sin sus datos.

Un recorrido completo



El modelo trabajó con la forma de los datos — un número de tarjeta, un cliente — y nunca con los valores reales.

Tokens con los que el modelo puede trabajar

- **Conservan su forma.** Una tarjeta convertida en token sigue pareciendo y validando como un número de tarjeta, así que el modelo puede razonar sobre ella y las validaciones existentes siguen funcionando.
- **Son consistentes dentro de una conversación**, así que el modelo sabe que dos menciones se refieren al mismo cliente o a la misma cuenta.
- **No significan nada fuera de ella.** Los tokens no se pueden revertir fuera del gateway.

Cómo se encuentran los datos protegidos

La detección es **modular**. El gateway combina varios métodos, los más certeros primero:

- las definiciones de campos de sus propios sistemas,

- patrones validados (por ejemplo, números de tarjeta con dígito verificador, números de cuenta y de ruta bancaria),
- la coincidencia con registros que usted ya tiene,
- sus propias listas de términos confidenciales,
- modelos de reconocimiento para nombres y direcciones en texto libre,
- y, si así lo decide, **el sistema de prevención de fuga de datos (DLP) que ya utiliza.**

Cuando la detección no es segura, el gateway protege el valor. **Ningún servicio de detección ni de tokenización fuera de su frontera de confianza recibe jamás sus datos protegidos.**

La tokenización también es modular: use la de NeodyAI o **el producto de tokenización que ya tiene.**

Claves

- Cada agente tiene su propia identidad y sus propias claves; la clave privada del agente nunca sale del agente.
- Cada conversación tiene sus propias claves de sesión, y el tráfico entre un agente y el gateway se cifra en cada llamada, además de TLS.
- Las claves se guardan en un servicio administrado de claves; usted puede rotarlas o revocarlas, y la revocación termina las sesiones de inmediato.

Gestión del riesgo de modelos

Por el gateway pasa cada llamada a un modelo, así que es ahí donde se aplican los controles de riesgo de modelos:

Expectativa (p. ej., SR 11-7, OCC 2011-12)	Lo que ofrece el gateway
Inventario de modelos	Ningún agente puede llegar a un modelo que no esté registrado y aprobado
Validación antes de usar	Los modelos y agentes nuevos trabajan en paralelo con su equipo antes de que se les permita actuar
Cuestionamiento efectivo	Aprobación por alguien independiente de quien propone, aplicada por el sistema
Monitoreo continuo	Calidad, costo y hallazgos de datos protegidos, por modelo y por agente
Control de cambios	Cuando un modelo cambia, los agentes que dependen de él pierden autonomía automáticamente hasta volver a validarse
Límites documentados	Qué datos puede manejar cada modelo y cada agente se aplica en cada llamada
Evidencia	Registros a prueba de manipulación de cada llamada, exportables para los supervisores, sin datos de clientes

Los métodos de detección se gestionan igual: inventariados, validados contra tasas mínimas de detección y monitoreados.

Lo que afirmamos y lo que no

Afirmamos	No afirmamos
Los datos protegidos se convierten en tokens antes de llegar a un modelo de lenguaje	"Cifrado de extremo a extremo" — el gateway restaura los valores reales para que sus sistemas puedan actuar
El proveedor del modelo nunca recibe sus valores protegidos	"NeodyAI nunca ve sus datos" — en la nube, el gateway procesa los valores dentro de la frontera de confianza; una opción en sus instalaciones protege los datos más sensibles antes de que lleguen a NeodyAI
Las tasas de detección se miden y se reportan	Que siempre se pueda detectar cada valor protegido

Los agentes y la red

NeodyAI está construido sobre **AgentMesh** y **Agent Messaging**, el protocolo abierto para que los agentes trabajen entre sí. Los agentes de NeodyAI tienen su propia identidad y pueden trabajar con otros agentes — los de sus proveedores, socios y clientes — a través de Agent Messaging, con cada mensaje firmado.

En la red rige la misma regla que con los modelos: **los datos protegidos no salen de su frontera de confianza sin convertirse en tokens**. Hoy NeodyAI no envía valores protegidos por la red en absoluto; cada solicitud se revisa antes de salir y queda registrada como evidencia. Para implementaciones reguladas, el servidor de código abierto Agent Messaging Server puede operar en su propio entorno.

Implementación

- **Nube de NeodyAI**, aislada por cliente.
- **Protección en sus instalaciones** para datos de tarjeta y las clases más sensibles.
- **Sus propios agentes** pueden usar el gateway: registre el agente y dirija sus llamadas a modelos y herramientas al gateway.

NeodyAI · neody.ai · Patente en trámite. Hay una presentación detallada de la arquitectura disponible bajo NDA — contáctenos en neody.ai.