

# Forcepoint AI Security



## Qué gobierna:

- › Protección de datos en las interacciones con IA
- › Detección y control de Shadow AI
- › Controles de acceso a datos para IA autorizada
- › Monitoreo y gobernanza de agentes de IA personalizados
- › AI Agent Gateway: control y aplicación en tiempo de ejecución

## Beneficios clave:

- › **Panel de gobernanza de IA:** Vista única de cada plataforma de IA autorizada, herramienta de Shadow AI y agente.
- › **Cero reclasificaciones:** Las políticas de DLP existentes se extienden a todos los canales de IA desde el primer día. Sin reconstruir políticas ni crear nuevos clasificadores.
- › **Cobertura completa:** Una sola plataforma gobierna los prompts de los empleados, las herramientas de Shadow AI, los agentes personalizados y los agentes alojados en la nube.
- › **Cumplimiento integrado:** Cada interacción y acción de los agentes se registra con atribución completa para fines de seguridad, cumplimiento, auditoría e investigación forense.
- › **Valor inmediato:** Forcepoint Adaptive Risk Intelligence Assistant (ARIA) guía a los administradores desde el primer inicio de sesión hasta un entorno gobernado.

## Descripción general

La mayoría de los programas de seguridad empresarial se diseñaron a partir de una premisa sencilla: los datos confidenciales circulan por canales que el equipo de seguridad puede monitorear. Esa premisa ya no se cumple en el entorno actual de IA. Cuando un empleado pega un registro de cliente en ChatGPT o le pide a Copilot que resuma archivos en SharePoint, nada de esa actividad pasa por filtros de correo, proxies web ni controles de endpoints. Las herramientas de Shadow AI agravan aún más el efecto: los empleados usan aplicaciones de IA que el equipo de seguridad nunca ha revisado, y cada herramienta no autorizada es una exposición de datos sin visibilidad ni registro de auditoría.

Los agentes de IA aumentan este riesgo. Un agente personalizado puede encadenar decenas de llamadas a herramientas, consultar almacenes de datos confidenciales y escribir en sistemas externos durante una sola sesión, por completo fuera del alcance de visibilidad de cualquier herramienta de seguridad actual. No son riesgos teóricos. Así se usa la IA hoy en los entornos de producción.

Forcepoint AI Security cierra esa brecha. Gobierna cada interacción con IA en toda la empresa, desde el prompt que escribe un empleado hasta la llamada a una herramienta que realiza un agente autónomo a velocidad de máquina, con los mismos controles de aplicación que ya protegen todos los demás canales de datos: SaaS, web, correo electrónico y endpoints. Políticas unificadas aplicadas desde una plataforma centrada en los datos que clasifica los datos confidenciales, gobierna cada interacción y detiene todos los riesgos de forma automática.

## DetECCIÓN Y RESPUESTA DE IA

Monitoreo y respuesta basados en API en plataformas de IA autorizadas y agentes personalizados

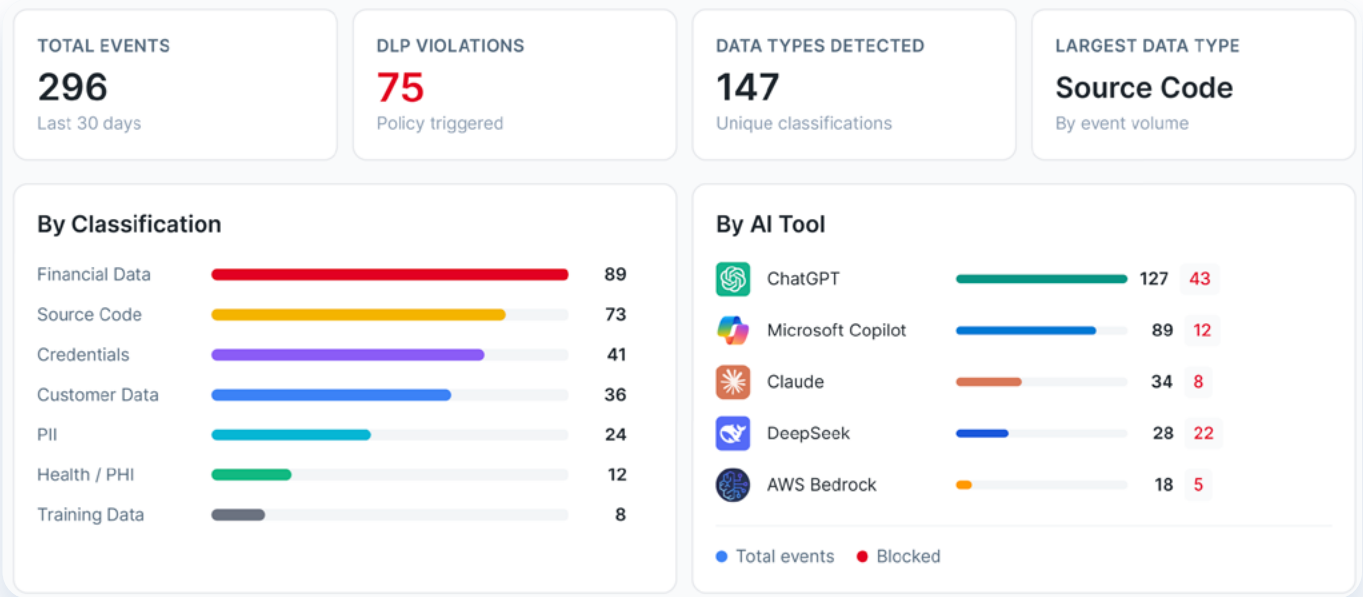
### Protección de datos en las interacciones con IA

#### ¿Sabes qué información están ingresando tus empleados en ChatGPT?

Cada prompt de IA es una transferencia de datos. Los empleados pegan con frecuencia registros de clientes, datos financieros y código fuente en herramientas de IA, lo que saca información confidencial del control de la empresa sin interceptación ni registro. Los controles de DLP existentes no se diseñaron para inspeccionar las cargas de prompts y respuestas de IA.

Forcepoint AI Security integra las interacciones con IA en el mismo marco de políticas que gobierna todos los demás canales de datos. Las políticas de clasificación existentes se aplican a prompts, respuestas y cargas de archivos en ChatGPT Enterprise, Microsoft 365 Copilot, Claude for Enterprise y AWS Bedrock. Las infracciones de política activan una acción de remediación: auditar, poner en cuarentena, copiar o eliminar.

- **Cero reclasificaciones:** Para las organizaciones que ya usan Forcepoint DLP, las políticas existentes se extienden a los canales de IA desde el primer día. No se requiere una nueva taxonomía.
- **Inspección de prompts y respuestas:** La detección abarca texto escrito, contenido pegado, cargas estructuradas y respuestas generadas por IA, no solo archivos adjuntos.
- **Descubrimiento retroactivo:** El relleno histórico revela eventos de datos confidenciales que eran invisibles antes de implementar el producto.
- **Sin interrupciones en el flujo de trabajo:** El escaneo basado en API no requiere cambios en la forma de trabajar de los empleados y no agrega carga a las herramientas autorizadas que usan.



## Controles de acceso a datos para IA autorizada

### "Copilot está leyendo tus archivos de SharePoint ahora mismo. ¿Sabes a qué puede acceder?"

Microsoft Copilot tiene acceso directo a todos los archivos de OneDrive y SharePoint a los que un empleado puede llegar. Los permisos empresariales se diseñaron para personas, no para asistentes de IA capaces de escanear y resumir miles de archivos en segundos. Los archivos compartidos en exceso o mal configurados muestran al instante registros de recursos humanos, datos financieros, materiales de la junta directiva y propiedad intelectual a usuarios que nunca debieron verlos.

Forcepoint AI Security usa la inteligencia de clasificación existente para entender a qué pueden acceder los asistentes de IA autorizados antes de que se escriba cualquier prompt. El etiquetado MIP impide que Copilot resuma o muestre archivos confidenciales. La remediación basada en API revoca el acceso a los archivos compartidos en exceso antes de que puedan ingerirse. El mismo marco de conectores extiende la auditoría completa de sesiones a ChatGPT Enterprise y Claude for Enterprise.

- **Clasificación previa al acceso:** Los archivos confidenciales se identifican antes de que comience cualquier interacción con un asistente de IA. Los controles se aplican antes de la recuperación, no después de la exposición.
- **Etiquetado MIP para Copilot:** Los archivos confidenciales de OneDrive y SharePoint se etiquetan para evitar que Copilot acceda a ellos, los resuma o los muestre en las respuestas a los prompts.
- **Remediación automática del uso compartido excesivo:** La remediación basada en API revoca el acceso al contenido compartido en exceso sin necesidad de una revisión manual de permisos.
- **Auditoría completa de prompts y respuestas:** Cada sesión en una plataforma de IA conectada se registra y se atribuye, e incluye lo que se accedió, lo que se preguntó y lo que devolvió la plataforma.

Monitoreo y gobernanza de agentes de IA personalizados

“¿Puedes mostrarme todos los agentes de IA personalizados que se ejecutan en tu entorno en este momento?”

Las empresas están implementando agentes de IA personalizados en AWS Bedrock y Copilot Studio con acceso a fuentes de datos internas. La mayoría de las organizaciones no tiene un inventario de qué agentes se ejecutan, a qué datos pueden acceder ni si alguno opera fuera de los límites previstos. Cuando algo sale mal, no hay registro de auditoría ni atribución.

Forcepoint AI Security ingiere y analiza los registros de invocación de AWS Bedrock a través de CloudWatch y S3, reconstruye los rastros de sesión y atribuye la actividad a identidades mediante IAM, Okta o Entra ID. Cada agente recibe una puntuación de riesgo según la sensibilidad de los datos. Los equipos de seguridad pueden correlacionar las acciones de los agentes con los eventos de DLP y actuar antes de que ocurra un incidente.

- **Atribución completa:** Cada acción de un agente se vincula tanto con la identidad del agente como con la persona que la activó. Ese vínculo es lo que hace posible la investigación de incidentes de IA.
- **Alerta temprana ante la proliferación de agentes:** Se activan alertas automáticas cuando se publican agentes nuevos o cuando los agentes existentes agregan nuevas conexiones externas.
- **Controles de gobernanza accionables:** Pausa una ejecución, exige aprobación humana, revoca el acceso o pon en cuarentena los archivos escritos por agentes.
- **Registros de auditoría listos para el cumplimiento:** Los registros de actividad de los agentes respaldan los requisitos de EU AI Act (artículo 12), GDPR (artículo 30), DORA, NIS2, LFPDPPP y de divulgación de IA de la SEC.

| Agent                                                  | Owner / Team                    | Profile   | Scope   | Token      | Last Active           |
|--------------------------------------------------------|---------------------------------|-----------|---------|------------|-----------------------|
| <b>Sale-Copilot-Prod</b><br>cl_748651dfbd...sdgd       | Mary Rodriguez<br>Revenue Ops   | Read-Only | 6 tools | ● Active   | 18-May-26<br>11:11:11 |
| <b>Market-Analyzer-Prod</b><br>cl_853462dfbf...ghij    | John Smith<br>Marketing Lead    | Editable  | 4 tools | ● Expiring | 20-May-26<br>12:15:30 |
| <b>Customer-Insights-Prod</b><br>cl_948273dfbe...ijkl  | Emma Johnson<br>Data Analyst    | Read-Only | 5 tools | ● Active   | 22-May-26<br>14:30:45 |
| <b>SalesTracker-Prod</b><br>cl_748652dcba...mnop       | Liam Brown<br>Sales Manager     | Editable  | 3 tools | ● Inactive | 23-May-26<br>09:00:00 |
| <b>Product-Feedback-Prod</b><br>cl_652348dfgh...qrst   | Olivia Taylor<br>UX Designer    | Editable  | 2 tools | ● Active   | 24-May-26<br>10:05:12 |
| <b>Competitor-Tracker-Prod</b><br>cl_763291dfgh...vwvx | Noah Wilson<br>Research Analyst | Read-Only | 6 tools | ● Active   | 25-May-26<br>13:45:20 |
| <b>Campaign-Optimizer-Prod</b><br>cl_253471dfgh...yzab | Ava Lee<br>Marketing Specialist | Editable  | 5 tools | ● Inactive | 26-May-26<br>15:30:00 |
| <b>Lead-Management-Prod</b><br>cl_847162dfgh...cdef    | Sophia White<br>Lead Strategist | Read-Only | 7 tools | ● Active   | 27-May-26<br>11:20:33 |

## DetECCIÓN Y CONTROL DE SHADOW AI

DetECCIÓN, puntuación de riesgo y control de acceso para cada herramienta de IA no autorizada de la empresa

### “¿Cuántas herramientas de IA usan tus empleados sin que las hayas aprobado?”

Las herramientas de Shadow AI proliferan a la velocidad de una extensión de navegador. Los empleados adoptan cuentas personales de IA, asistentes de redacción, herramientas de programación y procesadores de documentos sin pasar por una revisión de seguridad, y las organizaciones ahora gestionan en promedio 37 agentes de IA mientras que muchos equipos de seguridad carecen de visibilidad sobre a qué acceden esos agentes. Cada herramienta no autorizada es una posible exposición de datos que los controles existentes no pueden ver.

Forcepoint AI Security ofrece a los equipos de seguridad un inventario actualizado de forma continua y con puntuación de riesgo de cada aplicación de IA no autorizada en uso. Los equipos de seguridad ven el detalle completo de la actividad por usuario y por herramienta: visitas, prompts, respuestas, cargas de archivos y descargas de archivos. Cuando los empleados usan cuentas personales en lugar de las cuentas

corporativas autorizadas, AI Security detecta la diferencia y aplica la política correspondiente.

- **Inventario de IA con puntuación de riesgo:** Cada aplicación de IA no autorizada se categoriza y puntúa, de modo que las decisiones de acceso se basan en el riesgo real.
- **Análisis detallado de la actividad:** Visibilidad total de lo que hacen los empleados en cada herramienta de Shadow AI, incluida la actividad de prompts y las transferencias de archivos.
- **Detección de cuentas corporativas frente a personales:** Identifica si los empleados usan las herramientas de IA con cuentas corporativas autorizadas o con cuentas personales, y aplica la política correspondiente.
- **Control de acceso por herramienta:** Permite, bloquea o restringe las cargas y descargas por aplicación de IA, según el usuario o el departamento.

| Unsanctioned AI Apps               |                               |        |       |       |
|------------------------------------|-------------------------------|--------|-------|-------|
| App Name                           | Category                      | Risk   | Users | Trend |
| Google Gemini<br>gemini.google.com | Generative AI - Conversation  | Low    | 89    |       |
| Perplexity AI<br>perplexity.ai     | Generative AI - Conversation  | Low    | 87    |       |
| Midjourney<br>midjourney.com       | Generative AI - Multimodal    | Medium | 43    |       |
| DeepSeek<br>deepseek.com           | Generative AI - Conversation  | High   | 31    |       |
| Hugging Face<br>huggingface.co     | Other - AI App                | Medium | 29    |       |
| Poe<br>poe.com                     | Generative AI - Conversation  | Low    | 22    |       |
| Jasper AI<br>jasper.ai             | Generative AI - Text and Code | Low    | 18    |       |
| Cursor<br>cursor.so                | Generative AI - Text and Code | High   | 15    |       |

# AI Agent Gateway

Aplicación en línea nativa de MCP entre los agentes de IA alojados en la nube y las aplicaciones SaaS a las que llaman

## “¿Puedes impedir que un agente de IA acceda a datos que no debería tocar?”

Los agentes alojados en la nube se conectan directamente a aplicaciones SaaS, como Microsoft 365, Salesforce y Jira, y operan con todos los permisos de la cuenta de servicio que se les asignó al implementarlos. No existe ningún mecanismo para limitar qué datos lee un agente, bloquear los campos que nunca debería ver ni detener una operación a mitad de la ejecución. Un agente comprometido es, en la práctica, una credencial comprometida con acceso irrestricto a las aplicaciones y sin responsabilidad.

Forcepoint AI Agent Gateway es un gateway MCP diseñado específicamente que actúa como intermediario obligatorio entre los agentes de IA alojados en la nube y las aplicaciones SaaS a las que llaman. Ningún agente posee directamente credenciales de ninguna aplicación empresarial. Cada agente opera dentro de un alcance definido de herramientas aprobadas, con tokens de corta duración sujetos a revocación inmediata. Cada respuesta se inspecciona a nivel de campo antes de que el agente pueda leerla: la PII, el código fuente, las credenciales y los secretos se bloquean o se ocultan antes de la transmisión. Las operaciones de escritura y eliminación pueden retenerse hasta recibir aprobación humana a través de Teams, Slack o correo electrónico. Cada transacción se registra de forma inmutable.

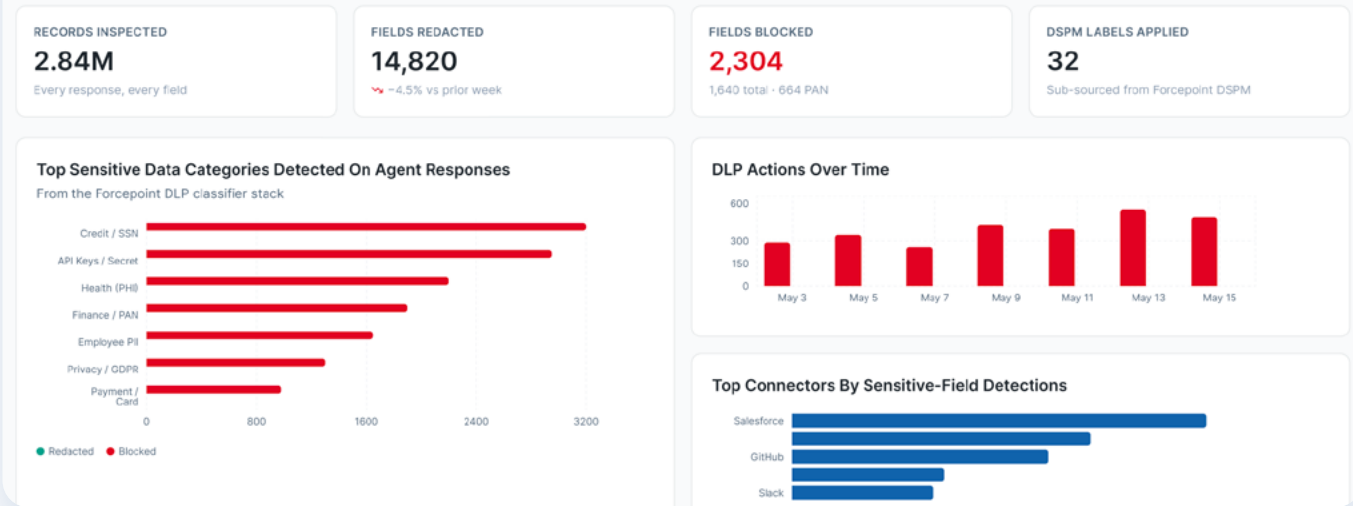
→ **Los agentes nunca poseen credenciales directas:**  
El Gateway intermedia todas las credenciales de las

aplicaciones. Un agente comprometido no puede autenticarse directamente en ninguna aplicación conectada.

- **Privilegios mínimos por diseño:** Cada agente se registra con un alcance definido de herramientas aprobadas y recibe tokens de corta duración. Los agentes no pueden acceder a datos ni operaciones fuera de su límite aprobado.
- **Protección de datos a nivel de campo:** La PII, el código fuente, las credenciales y los secretos se bloquean o se ocultan antes de que el agente pueda leerlos, agregarlos o transmitirlos.
- **Controles con intervención humana:** Las operaciones de escritura y eliminación pueden retenerse hasta recibir aprobación humana. El aprobador ve el nombre del agente, la identidad del usuario, la herramienta, los parámetros y la clasificación de los datos antes de decidir.
- **Detección de comportamiento anómalo:** Las consultas con un volumen inusual de registros se detectan y se detienen en la capa de ejecución antes de que se complete una extracción masiva de datos.
- **Rastro de auditoría inmutable:** Cada transacción genera un registro completo: identidad del agente, identidad del usuario, herramienta invocada, clasificación de los datos, decisión de la política, campos bloqueados u ocultos, cantidad de registros y marca de tiempo.

## Data protection — what no other MCP gateway shows

Sensitive fields detected, redacted, or blocked on agent responses — by category, connection, and over time



## Valor desde el primer día

Forcepoint AI Security está diseñado para generar valor desde el primer día. ARIA, el asistente de incorporación de Forcepoint basado en IA, se inicia automáticamente en el primer inicio de sesión y guía al administrador en la configuración de conectores y políticas, y le recomienda un conjunto inicial de políticas de DLP según la región, el sector y el perfil de sensibilidad de los datos de la organización. No se requieren servicios profesionales. Los datos históricos de interacciones se rellenan automáticamente en la primera conexión, de modo que el panel de gobernanza de IA tenga contenido real desde el momento en que se completa la configuración.

### AI Security Setup

✕ Exit

Complete setup to start monitoring AI usage across your organisation.

1

2

3

Connect AppsData ProtectionConfirm Setup

## Connect your AI Applications

Connect **at least one app** to begin monitoring. You can add more at any time.  
We will monitor user prompts sent to AI apps and any backend data access across services.



AWS Bedrock



ChatGPT



Claude



Copilot

## Capacidades clave

| FUNCIÓN                                         | BENEFICIO                                                                                                                                                                                                                                                                                                                                                             |
|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Conectores basados en API</b>                | Visibilidad y control profundos sobre ChatGPT Enterprise, Microsoft 365 Copilot, Claude for Enterprise y AWS Bedrock.                                                                                                                                                                                                                                                 |
| <b>Herencia de políticas de DLP</b>             | Las clasificaciones existentes de Forcepoint DLP se extienden automáticamente a prompts, respuestas y cargas de archivos de IA sin requerir una nueva taxonomía.                                                                                                                                                                                                      |
| <b>Registro de prompts y respuestas</b>         | Texto completo de cada prompt y respuesta de IA capturado, indexado y con búsqueda por usuario, rango de fechas, etiqueta de clasificación y herramienta de IA.                                                                                                                                                                                                       |
| <b>Visibilidad de operaciones con archivos</b>  | Cada archivo cargado en una herramienta de IA autorizada se detecta, se clasifica según la taxonomía de DLP de la organización y se registra con contexto de auditoría completo.                                                                                                                                                                                      |
| <b>Panel de gobernanza de IA</b>                | Vista única en tiempo real de todas las plataformas de IA autorizadas, las herramientas de Shadow AI y la actividad de los agentes. Indicadores clave (KPI), datos de tendencias, hallazgos clasificados por riesgo y próximas acciones guiadas en una sola vista operativa.                                                                                          |
| <b>Inventario y gobernanza de agentes</b>       | Inventario en vivo de los agentes de IA en Copilot Studio, ChatGPT Enterprise y AWS Bedrock, con la identidad del agente, el modo de autenticación, las fuentes de datos conectadas y la puntuación de riesgo por agente.                                                                                                                                             |
| <b>Detección y control de Shadow AI</b>         | Inventario actualizado de forma continua de herramientas de IA no autorizadas con puntuaciones de riesgo, detalle completo de la actividad por usuario y herramienta, detección de cuentas corporativas frente a personales y políticas de control de acceso.                                                                                                         |
| <b>Barreras de acceso a datos</b>               | Identificación basada en la clasificación de archivos compartidos en exceso o mal etiquetados antes de que los asistentes de IA puedan acceder a ellos. Etiquetado MIP y remediación basada en API disponibles para Microsoft 365, SharePoint, OneDrive y Microsoft Copilot.                                                                                          |
| <b>Acciones de remediación</b>                  | Permitir, permitir y monitorear, restringir la acción, eliminar el historial de conversaciones, revocar el acceso a archivos, poner en cuarentena archivos escritos por agentes, pausar la ejecución de un agente o exigir aprobación humana. Las acciones disponibles varían según el tipo de interacción y la superficie de aplicación.                             |
| <b>Configuración guiada por ARIA</b>            | Asistente de incorporación basado en IA que guía a los administradores en la configuración de conectores y políticas, y recomienda un conjunto inicial de políticas según la región, el sector y la sensibilidad de los datos. No se requieren servicios profesionales.                                                                                               |
| <b>AI Agent Gateway</b>                         | Intermediario obligatorio entre los agentes de IA alojados en la nube y las aplicaciones empresariales SaaS conectadas. Los agentes nunca poseen credenciales directas de las aplicaciones. Protección de datos a nivel de campo, aprobaciones configurables con intervención humana y registro de auditoría inmutable de cada transacción entre agente y aplicación. |
| <b>Exportación de auditoría de cumplimiento</b> | Registro de auditoría exportable que abarca todas las interacciones de IA, las acciones de los agentes y los eventos de políticas. Admite los requisitos de EU AI Act, NIST AI RMF, GDPR (artículo 30), DORA, NIS2, LFPDPPP y de divulgación de IA de la SEC. Exportable en CSV, JSON y PDF.                                                                          |

Forcepoint AI Security se basa en la plataforma Forcepoint AI Data Security, que clasifica los datos confidenciales antes de que la IA los toque y aplica políticas en cada interacción en toda la empresa. Un único marco de políticas gobierna el correo electrónico, la web, los endpoints, las aplicaciones en la nube y la IA desde un único plano de control.

La plataforma aborda los dos lados del mismo problema: proteger los datos empresariales frente a las herramientas de IA y gobernar los datos para que la IA pueda operar con seguridad. La detección, la aplicación y la información sobre riesgos se ejecutan en el mismo motor, lo que ofrece a los equipos de seguridad una vista operativa única de todas las superficies que toca la IA.

