Engineer

Seguridad para sus agentes de IA

Revisiones y controles de acceso para agentes de IA: inventario acordado, hallazgos priorizados, pruebas de permisos y registro de verificación.

Comprenda a qué pueden acceder sus agentes de IA y establezca controles sobre ese acceso. Un agente conectado al correo, a documentos o a una API introduce permisos y vías de fallo que deben revisarse. Inventariamos los sistemas acordados, evaluamos los riesgos, implementamos los controles aprobados y probamos las acciones restringidas. Usted recibe el mapa de accesos, los hallazgos y el registro de verificación.

Convierta el servicio en un resultado

Haga que «Seguridad para sus agentes de IA» contribuya a un resultado empresarial. Podemos conectar el trabajo técnico con los usuarios, las operaciones y las decisiones a los que debe dar soporte. El acceso de cada agente definido y limitado a lo necesario, cada acción registrada donde el agente no puede alterarla y cada protección comprobada en lugar de darse por segura.

Explore nuestros servicios Póngase en contacto →

Examine los trabajos relevantes

Utilice los trabajos relacionados de esta página para examinar nuestro enfoque. Pregunte por los compromisos entre objetivos y las condiciones detrás de cada resultado.

Casos prácticos Póngase en contacto →

Identificados: cada agente y aquello a lo que puede acceder

Inventariamos los agentes dentro del alcance acordado, incluidas las integraciones aprobadas y otras conexiones identificadas con su personal. Para cada uno, el mapa registra credenciales, herramientas y datos accesibles. Recibe el mapa junto con los hallazgos, y la responsabilidad de mantenerlo actualizado se acuerda como parte del traspaso. Es un registro del entorno revisado, con su alcance y sus lagunas expresados de forma explícita.

Protegidos: cada agente con el mínimo privilegio

Cada agente recibe su propia identidad y el mínimo privilegio que requiere su tarea: roles en lugar de cuentas compartidas (RBAC), solo lectura de forma predeterminada y escrituras únicamente mediante herramientas acotadas, con límites de frecuencia y volumen. Los secretos se separan por agente y por servidor, nunca se pegan en una instrucción y se rotan sin cambiar el código. El texto externo que lee un agente —un correo electrónico, una página web o un documento— se trata como datos no fiables. Los permisos de las herramientas y las comprobaciones de validación aplican los límites acordados; un mensaje que diga «ignora tus reglas» no autoriza ninguna acción. Los controles deniegan el acceso ante un fallo: si una comprobación no puede ejecutarse, se rechaza la acción.

Registrados: qué hizo y por cuenta de quién

Cada llamada a una herramienta se registra junto con la persona por cuya cuenta se realizó, qué se pidió, a qué accedió y qué devolvió, y el registro se conserva donde el propio agente no pueda modificarlo. Las reglas de alerta se acuerdan para el sistema. Pueden abarcar una ráfaga de lecturas, el uso fuera del horario previsto o el acceso a datos fuera del alcance acordado para el agente. Cuando algo falla, el registro convierte un incidente en una cronología, en lugar de una conjetura.

Un mapa, un cierre y un registro **son tres afirmaciones** — hasta que alguien intenta sortearlos.

Los controles acordados se comprueban con casos representativos de acceso, uso indebido y recuperación antes del traspaso.

Los controles se prueban, no se dan por buenos

Probamos los controles de acceso dentro del alcance del encargo y registramos qué se intentó, qué se rechazó y qué sigue requiriendo atención. Las pruebas pueden abarcar llamadas a herramientas denegadas, límites de frecuencia, aislamiento de credenciales e intentos de inyección de instrucciones. El resultado es un registro de verificación revisable y un conjunto de comprobaciones

repetibles, con sus límites indicados. La respuesta a incidentes y la rotación de credenciales se definen por separado cuando hay que contener una exposición existente.

El trabajo que lo respalda

Hemos realizado este trabajo en servidores, sistemas de identidad y plataformas. Escribimos 1.000 reglas RBAC para servidores Linux y de bases de datos, dirigimos la gestión de accesos en M365 y el inicio de sesión único y gestionamos un programa continuo de análisis de seguridad con 21 auditorías escritas. Separamos todos los secretos operativos por servidor, añadimos una comprobación previa de secretos de un segundo, encontramos tres cerraduras en la puerta, ninguna cerrada y demostramos en cada ejecución que funciona el bloqueo de intrusiones. Construimos límites de frecuencia que deniegan el acceso ante un fallo y protección contra abusos y escribimos una regla de alcance tras una filtración. Nuestra propia práctica se explica en cómo trabajamos de forma segura.

Cómo se desarrolla un encargo

Empieza con acceso de solo lectura. Identificamos los agentes dentro del alcance acordado, sus credenciales y los recursos a los que permiten acceder, y le entregamos los hallazgos por orden de riesgo, cada uno con su motivo y su solución. Después se aplican las correcciones en su configuración y su repositorio, pasando por su revisión, con comprobaciones que fallan cuando las vías de acceso cubiertas superan los permisos acordados. Usted conserva el mapa, el registro y las pruebas: el archivo que entrega a un auditor o a una aseguradora. Los datos que leen los agentes pueden recibir permisos del mismo modo.
Se contrata como el resto de nuestro trabajo, según las condiciones de cuánto cuesta.

Si los agentes ya tienen llaves de sus sistemas, cuéntenos cuáles.

Cada agente que ejecuta **tiene una llave para algo** — sepa cuál y para qué.

Comienza en modo de solo lectura: un mapa de los agentes y las credenciales incluidos en el alcance, con los hallazgos ordenados por riesgo.

Solicite un mapa de agentes Cuánto cuesta →

Defina el alcance antes de encargar el trabajo

Describa la ruta de despliegue, los riesgos operativos y quiénes se encargan de las guardias. Podemos definir el alcance de un cambio con comprobaciones, una vía de recuperación y responsabilidades claras.

Vea cuánto cuesta Póngase en contacto →

Conversemos sobre sus requisitos

Cuéntenos qué necesita de «Seguridad para sus agentes de IA», incluidos el plazo y el equipo que lo mantendrá. Podemos hablar de una forma práctica de avanzar.

Explore nuestros servicios

welcome@engineer.company +45 50 10 61 84 Copenhague, Dinamarca

Llévese esta página

Tarjeta de contacto

Creado con esmero y cariño por D.

CVR 45969703 · Tagensvej 105C, 3. th, 2200 København N

Derechos de autor © 2025 – actualidad · Engineer ApS · Todos los derechos reservados.

Impreso desde https://ditte.engineer.company/es/ai-security/