

DARKTRACE

Guía del CISO para la compra de IA

Investigar antes de invertir:
inicie su proceso de adopción de la IA



Secure
What AI
Can Do

Contenido

02	Introducción
03	¿Qué debe tener en cuenta al invertir en IA?
03	Lo que debe saber antes de evaluar un AISP
04	Comprender las limitaciones de la IA en ciberseguridad
05	Cómo evaluar proveedores de IA
06	Gobernanza, seguridad y controles de datos
07	Recopilación de datos y entrenamiento
07	Elección de modelos y técnicas
08	Validación del rendimiento y la precisión
08	Interpretabilidad, capacidad de ajuste y transparencia
09	Darktrace como proveedor de ciberseguridad de IA
10	Conclusión

Introducción

La IA se ha integrado rápidamente en los productos de seguridad y seguirá desempeñando un papel importante en la forma en que estos productos detectan, analizan y responden a las amenazas. Al ahorrar tiempo y recursos, extraer inteligencia y conclusiones de datos valiosos y llevar a cabo análisis más profundos y complejos que los de una solución de seguridad tradicional, la IA mejora las capacidades y el nivel de seguridad de una organización.

Con esta aceleración en la adopción de la IA, acompañada por el reciente auge de la IA agéntica y los agentes autónomos, los CISO deben mirar más allá de la superficie de estas capacidades para entender cómo se construyen, entrenan y validan los diferentes modelos, y cómo se comportan cuando se enfrentan a datos nuevos o desconocidos.

Los proveedores utilizan una amplia variedad de enfoques, y cada decisión de diseño influye en la precisión, la fiabilidad y la probabilidad de introducir puntos ciegos. Comprender cómo funciona una funcionalidad de IA y si es realmente adecuada para la tarea que afirma resolver permite a los responsables de seguridad plantear preguntas más precisas y garantizar que la tecnología se aplique de la forma correcta sin añadir riesgos innecesarios.

Igual de importante es comprender si la IA se despliega de forma segura y se gestiona de manera responsable. Una visibilidad deficiente, unos controles débiles y un mantenimiento irregular pueden provocar la degradación del rendimiento, la deriva del modelo, sesgos y vulnerabilidades explotables.

Las soluciones bien diseñadas incluyen salvaguardas sólidas, procesos de razonamiento transparentes y controles claros en torno al tratamiento y la protección de datos. Los CISO deben sentirse seguros a la hora de cuestionar a los proveedores sobre sus marcos de gobernanza, los flujos de trabajo de pruebas y validación, la resiliencia frente a entradas adversarias y las salvaguardas para evitar el uso indebido o la fuga de datos. Este nivel de diligencia debida es especialmente importante dado el reciente aumento de la inversión y el desarrollo en materia de IA, así como de las afirmaciones de los proveedores.

Si bien la IA puede aportar mejoras significativas en la productividad, la generación de información de valor y la eficacia de la seguridad, el mercado también está condicionado por las expectativas desmedidas, las ideas erróneas y la falta de transparencia.

Esta guía está diseñada para ayudar a los CISO a sortear ese ruido y distinguir las capacidades probadas de los términos de moda, para que puedan evaluar con confianza las herramientas de seguridad habilitadas por IA y garantizar que cualquier adopción sea fiable, responsable y esté alineada con los requisitos de riesgo y gobernanza de su organización.



¿Qué debe tener en cuenta al invertir en IA?

Un proveedor de servicios de IA (AISP) es una empresa que desarrolla y ofrece IA como servicio (AI-as-a-Service), gestionando todo el proceso: desde la elección del modelo, el diseño y el entrenamiento, hasta el perfeccionamiento, la validación y la gobernanza continua. Las soluciones de un AISP pueden basarse en un único modelo o en una combinación de modelos estructurados conjuntamente para generar un resultado más complejo y completo.

Casi todos los proveedores de seguridad empiezan a utilizar la IA como parte de sus soluciones. En el contexto de la ciberseguridad, la IA puede ser sumamente valiosa. Mientras que los sistemas de seguridad tradicionales se apoyan en la detección basada en firmas, los sistemas de IA son capaces de identificar patrones y tendencias, generalizando para identificar ataques inéditos.

Lo que debe saber antes de evaluar un AISP

-  Los AISP pueden utilizar varios modelos diferentes dentro de una única solución
-  La IA abarca una amplia gama de tecnologías (incluida, entre otras, la IA generativa)
-  La IA no siempre es correcta y necesita optimizarse
-  Todos los sistemas de IA introducen sus propios riesgos, lo que requiere gobernanza

La IA evoluciona constantemente, lo que dificulta mantener una definición clara de «qué es un sistema de IA». La Organización para la Cooperación y el Desarrollo Económicos (OCDE) define la IA como un sistema basado en máquinas que puede inferir resultados a partir de entradas sin la necesidad de una programación explícita.

El campo de la IA es amplio, con una gran variedad de técnicas que se engloban bajo el paraguas de la IA, entre las que se incluyen el clustering, la clasificación, la regresión, la visión artificial, los modelos de lenguaje de gran tamaño (LLM), los modelos de lenguaje específicos de dominio (DSLM) y la IA agéntica.

LLM y DSLM

Un campo de la IA que ha experimentado un reciente auge de popularidad son los modelos de lenguaje de gran tamaño (LLM). Los LLM son modelos de uso general diseñados para procesar entradas de texto libre y resultan idóneos para tareas de lenguaje natural.

Del mismo modo, los modelos de lenguaje específicos de dominio (DSLM) son LLM diseñados para una tarea específica o para operar en un dominio determinado, como el modelo de lenguaje DEMIST-2 de Darktrace, que se ha entrenado específicamente para tareas relacionadas con la seguridad.

IA generativa

La IA generativa se refiere a un sistema de IA capaz de producir resultados originales, incluidos texto en lenguaje natural, imágenes y audio. Estos sistemas se han vuelto cada vez más populares con el auge de los sistemas de chatbots de IA basados en la nube y respaldados por LLM. La IA generativa no se refiere exclusivamente a los LLM: existen muchos otros sistemas de IA generativa, como los modelos de generación de imágenes y vídeo (como Dalle-2 de OpenAI o Veo de Google), que se basan en la misma tecnología que los LLM, además de otras técnicas estadísticas generativas.

También es importante señalar que los LLM no siempre tienen que utilizarse de forma generativa. Existe una amplia gama de otros enfoques basados en LLM que pueden resultar más adecuados que la IA generativa para algunas tareas.

IA agéntica

Más recientemente, algunos AISP se han centrado en gran medida en desarrollar soluciones de IA agéntica. «Agéntico» es una palabra de moda que actualmente se utiliza mucho en el sector de la IA, con numerosas definiciones en circulación y muchos AISP que afirman que sus sistemas son agénticos.

La definición de IA agéntica varía enormemente de un proveedor a otro, abarcando desde cualquier sistema autónomo hasta cualquier sistema de IA generativa capaz de utilizar herramientas, o incluso un simple término de marketing que puede aplicarse a cualquier sistema de IA.

Con esta variación, es necesario preguntar a los posibles AISP qué definen como IA agéntica, cómo se implementa la IA agéntica en su solución y considerar qué riesgos puede introducir su implementación.

IA compuesta

Otra técnica clave que pueden utilizar los AISP es la IA compuesta: una combinación de múltiples técnicas de IA diferentes en una sola solución. Esta combinación permite que múltiples sistemas se apoyen mutuamente, lo que reduce los puntos ciegos y amplía el alcance de la solución.

Comprender las limitaciones de la IA en ciberseguridad

Si bien incorporar sistemas de IA a una solución puede ser beneficioso, al aportar conclusiones únicas y mejoras de la eficiencia mediante acciones autónomas, es importante recordar que todos los sistemas de IA conllevan sus propias desventajas y posibles riesgos.

Por ejemplo, aunque los sistemas de IA generativa son cada vez más populares por su flexibilidad y su capacidad para procesar y generar datos no estructurados en lenguaje natural, funcionan mejor en algunas tareas que en otras.

A la hora de considerar una inversión en una solución de IA, los compradores deben sopesar qué ventajas ofrece la IA frente a otros enfoques y evaluar la variedad de metodologías de IA disponibles, junto con las ventajas únicas que aporta cada una. En este documento, abordaremos algunas de las preguntas clave que podría plantear a su AISP para comprender su enfoque respecto al desarrollo, la implementación y la gobernanza de la IA.

Por último, aunque la adopción de sistemas de IA se acelera rápidamente, gran parte de la gobernanza y la gestión de riesgos en torno a ellos aún se encuentra en una fase inicial. Además, los requisitos en materia de gobernanza y gestión de riesgos no hacen más que aumentar a medida que los sistemas se vuelven más complejos. Comprender los controles que los proveedores han implementado en este ámbito puede ser un componente crucial para diferenciar entre demostraciones técnicas impresionantes y sistemas capaces de un uso empresarial fiable y sólido.

En el caso particular de la IA generativa, Gartner sugiere que es:

-  Muy útil para generación de contenido, interfaces de usuario conversacionales, descubrimiento de conocimiento
-  Moderadamente útil para segmentación/clasificación, sistemas de recomendación, percepción, automatización inteligente, detección de anomalías/monitorización, sistemas autónomos
-  Apenas útil para predicción/elaboración de previsiones, planificación, inteligencia de decisiones

Además, una investigación de Gartner ha llegado a la conclusión de que:

Para lograr soluciones de IA más precisas y robustas, los líderes de IA deben ir más allá del uso de un único modelo o técnica, adoptar prácticas de IA compuesta y asumir una perspectiva holística del sistema de IA.

Cómo evaluar proveedores de IA

En la investigación inicial sobre los AISP, hay varios lugares a los que los compradores pueden recurrir para reducir su lista de candidatos. Para ello, busque pruebas concretas y testimonios de clientes.

Estos proporcionarán pruebas de que la herramienta de IA funciona y de que también puede funcionar para su organización. Busque especialmente información de clientes que se encuentren en situaciones similares a la de su organización, ya sea por sector, tamaño de la organización, región u otros atributos.

Asimismo, al evaluar un AISP, sería muy útil valorar si el proveedor cuenta con una cultura de buena gobernanza. Ello implica comprender cómo se integra la gobernanza en el desarrollo, las pruebas y la implementación de la IA dentro de la organización.

La investigación de nuestro informe Estado de la ciberseguridad de la IA de Darktrace reveló que solo el 37 % de las organizaciones cuenta con una política de gobernanza de la IA.

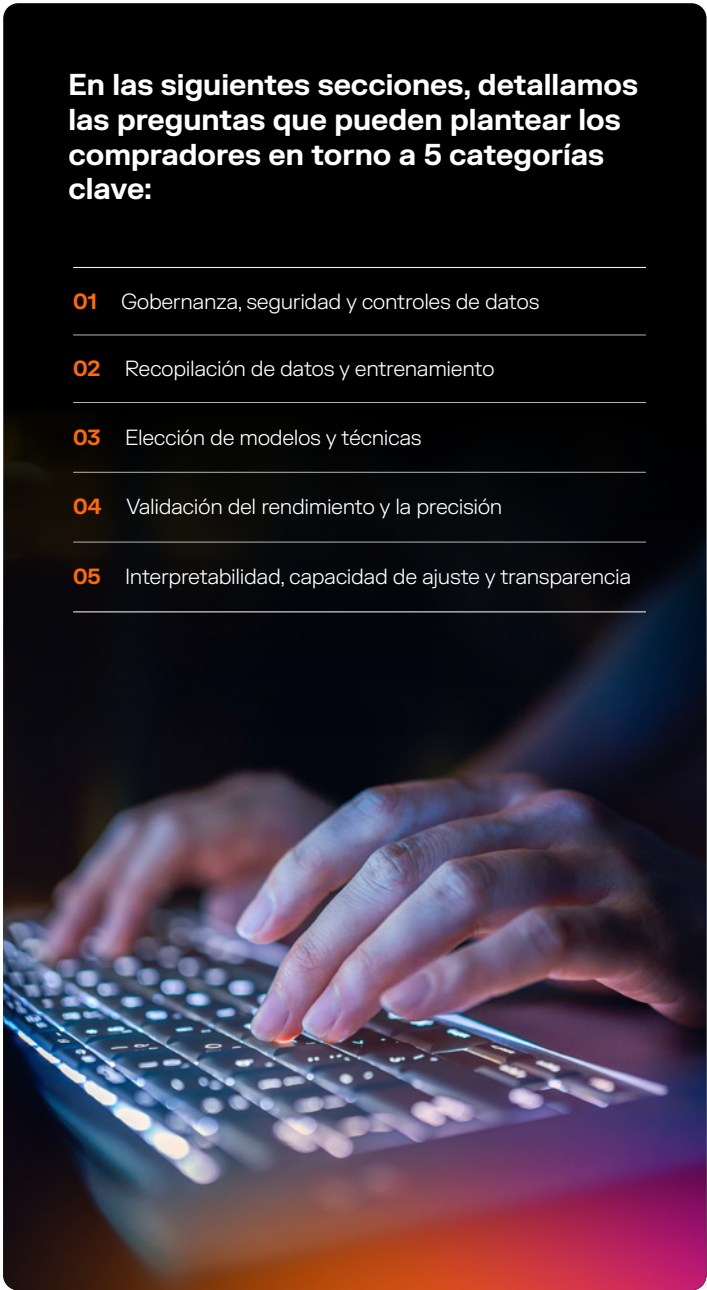


Es importante reconocer que, aunque una empresa cuente con una política de gobernanza de la IA o afirme alinearse con un estándar conocido de gobernanza de la IA, las preguntas importantes que deben plantearse son cómo implementa dicha empresa esas políticas y procesos. Cuanto más sólidas sean la cultura y la adopción de la gobernanza de la IA en una empresa, mayor será la probabilidad de que esto se refleje en sus productos.

Los compradores deben formular preguntas a los posibles AISP sobre los procesos de gobernanza y seguridad de datos que implementan, las tecnologías que utilizan y los flujos de trabajo que siguen para las pruebas, la evaluación, la validación y la verificación. Al formular estas preguntas, un comprador puede recopilar más información para confirmar que una solución cumple con sus requisitos y políticas empresariales.

En las siguientes secciones, detallamos las preguntas que pueden plantear los compradores en torno a 5 categorías clave:

- 01 Gobernanza, seguridad y controles de datos
- 02 Recopilación de datos y entrenamiento
- 03 Elección de modelos y técnicas
- 04 Validación del rendimiento y la precisión
- 05 Interpretabilidad, capacidad de ajuste y transparencia



Gobernanza, seguridad y controles de datos

La gobernanza de la IA es esencial para garantizar que todos los sistemas y modelos de IA de una solución sean seguros, precisos y reciban un mantenimiento adecuado, así como para verificar que todos los datos introducidos en un sistema se utilicen de forma responsable, ética y legal. Un sólido sistema de gobernanza proporciona suficiente visibilidad interna sobre todos los sistemas, lo que permite a una organización garantizar la seguridad, la protección y la integridad de los datos.

Cuando se trata de sistemas de IA compuestos más complejos que incorporan un gran número de modelos, la gobernanza es fundamental. En las soluciones que incorporan múltiples sistemas de IA, sería difícil garantizar que todos los modelos se mantengan actualizados y seguros, y se perfeccionen de forma constante para mejorar la precisión, sin una solución de gobernanza suficientemente sólida que permita rastrear y documentar todos los sistemas.

Un mantenimiento irregular y la falta de visibilidad sobre los sistemas a nivel interno podrían dar lugar a la introducción de vulnerabilidades, puntos débiles y sesgos, lo que afectaría negativamente al rendimiento y a la precisión de la solución.

A continuación, se incluyen algunos ejemplos de preguntas que un comprador debe plantear a los posibles proveedores de servicios de IA al recopilar información sobre sus políticas y procesos de gobernanza.

PREGUNTAS PARA PROVEEDORES

¿Qué políticas y marcos de gobernanza de la IA sigue, y/o qué certificaciones mantiene actualmente?

¿Cómo garantiza la seguridad y la privacidad de los datos, así como la compartimentación de la información de clientes (IP, PII, etc.), dentro de sus soluciones de IA?

¿Cómo protege sus sistemas de IA frente a entradas maliciosas, envenenamiento de datos y otros ataques adversarios?

¿Utiliza plataformas o modelos de terceros alojados externamente para procesar o gestionar datos de clientes o de la empresa? Y, de ser así, ¿cómo garantiza que estos sistemas sean seguros, protegidos y confidenciales?

Al revisar las respuestas a las preguntas sobre gobernanza y seguridad de los datos, los compradores deben buscar referencias a certificaciones, flujos de trabajo o marcos específicos que utilice su AISP para garantizar una gobernanza sólida.



Darktrace cuenta con la certificación de la norma ISO/IEC 42001. La ISO/IEC 42001 es la primera norma mundial para sistemas de gestión de la IA (AIMS) y proporciona orientación valiosa sobre los desafíos únicos que plantea la IA. Para las organizaciones, establece una forma estructurada de gestionar los riesgos y las oportunidades asociados con la IA, lo que favorece una gobernanza sólida y duradera de todos los sistemas de IA.

Es importante tener en cuenta que la norma ISO/IEC 42001 no es el único marco de gobernanza de la IA disponible, ya que otros marcos como el Marco de Gestión de Riesgos de la IA del NIST también se utilizan ampliamente. Sin embargo, la ISO/IEC 42001 proporciona un nivel adicional de garantía mediante auditorías externas, en lugar de basarse en una adhesión autodeclarada.

Además de las políticas de gobernanza de la IA, la seguridad y la privacidad de los datos son fundamentales para todos los sistemas, por lo que identificar marcos de seguridad de los datos y certificaciones también puede aportar a los compradores la confianza de que un sistema es lo suficientemente seguro y capaz de gestionar sus datos de manera segura.

Recopilación de datos y entrenamiento

Uno de los pasos más importantes a la hora de crear un sistema de IA es la recopilación inicial de datos de entrenamiento. Si un sistema de IA se entrena con datos de baja calidad, incoherentes, inexactos o sesgados, el sistema resultante nunca ofrecerá un rendimiento óptimo.

La norma ISO/IEC 42001 y otros marcos de gobernanza proporcionan controles que se deben seguir al recopilar datos para sistemas de IA. Estos controles incluyen la identificación del origen y los registros de auditoría para garantizar que los datos de entrenamiento sean trazables y no se hayan manipulado, la aplicación de una desidentificación o abstracción suficiente en los datos de usuario recopilados en función del contexto de riesgo, y la comprobación exhaustiva de posibles sesgos cuando proceda.

A continuación, se incluye una lista de preguntas que los compradores pueden tomar como punto de partida para las consultas sobre la recopilación de datos.

PREGUNTAS PARA PROVEEDORES

- ¿Qué medidas toma para evitar los sesgos en sus modelos de IA y en los datos de entrenamiento?
- ¿Cómo minimiza los riesgos asociados a los datos de entrenamiento (p. ej., regurgitación, procedencia de los datos)?

Al revisar las respuestas, los compradores deben evaluar si el proveedor cuenta con un sistema de gestión de IA que regule la obtención y el tratamiento de los datos de entrenamiento, y si existen procesos claros para identificar y mitigar sesgos o problemas sistémicos.

En los casos en que haya información de identificación personal (PII) en los datos de entrenamiento de la IA (fuera de los sistemas de autoaprendizaje específicos de la organización), los AISP deben poder hacer referencia a controles y procesos específicos utilizados para desidentificar los datos (eliminar la PII) cuando proceda, o a cualquier otro proceso de anonimización.

Elección de modelos y técnicas

Los distintos modelos y técnicas de IA se adaptan a diferentes tareas. Seleccionar el modelo adecuado para una tarea concreta es un paso importante para garantizar que una solución sea precisa y eficiente.

Los modelos y las técnicas también requieren una gobernanza sólida, y no se trata solo del cumplimiento de normas y certificaciones; también se trata de adoptar los modelos y procesos más eficaces y eficientes para lograr el objetivo previsto. Ello implica mantener registros detallados de los modelos y técnicas que se utilizan y garantizar que se analicen continuamente para detectar riesgos y vulnerabilidades, incluidos los riesgos de la cadena de suministro de los modelos de terceros y los modelos internos.

Para garantizar que los proveedores de servicios de IA sigan los procedimientos de gobernanza adecuados al seleccionar modelos y servicios alojados tanto interna como externamente, estas son algunas preguntas que los compradores deben formular:

PREGUNTAS PARA PROVEEDORES

- ¿Qué tipo(s) de modelo de IA utiliza en su solución?
- ¿Utiliza alguna IA generativa (alojada de forma interna o externa) y, en caso afirmativo, qué medidas de protección y barreras de seguridad tiene implementadas para ellas?
- ¿Cómo garantiza que su elección de modelos sea óptima para cada tarea requerida?
- ¿Utiliza IA agéntica? Si es así, ¿cómo y dónde se utiliza?
- ¿Utiliza modelos fundacionales o de propósito general? En caso afirmativo, ¿cómo se alojan?

Si bien es probable que la información detallada y específica sobre los sistemas personalizados que utilizan los AISP sea confidencial, los compradores deben esperar que los proveedores puedan ofrecer una visión general de las principales técnicas utilizadas. La transparencia en torno al uso de modelos fundacionales de terceros o alojados en la nube, las API asociadas y la residencia de datos pertinente sigue siendo un factor importante para los compradores que buscan garantías sobre el tratamiento y la integridad de los datos.

Los AISP deben hacer referencia a la diligencia debida y a una investigación y un análisis exhaustivos de terceros antes de implementar sus sistemas externos para garantizar la protección y la seguridad.

Validación del rendimiento y la precisión

Los sistemas de IA deben someterse a exhaustivos procesos de pruebas, evaluación, verificación y validación (TEVV) antes de su uso general. La evaluación de un sistema, seguida de optimizaciones, ajustes y mejoras, resulta fundamental para garantizar que los sistemas desarrollados estén calibrados de forma adecuada y segura, con el fin de ofrecer información útil para la tarea prevista.

El TEVV va más allá de la mera validación de la precisión. Es necesario aplicar numerosas técnicas para confirmar que un sistema está libre de sesgos, no sobreajusta ni infraajusta los datos, y funciona de forma eficiente y precisa. La evaluación continua tras el lanzamiento también es fundamental para evitar la desviación no deseada de los datos y del modelo. Las pruebas también deben abarcar el rendimiento físico del modelo, lo que garantiza que pueda generar resultados en un plazo adecuado y que funcione con un consumo de recursos razonable.

La norma ISO/IEC 42001 exige la creación de un flujo de trabajo para evaluar la relevancia y la precisión de un sistema, el cual debe completarse para cada sistema de IA antes de que pueda utilizarse en producción.

A continuación se incluye una lista de preguntas que los compradores podrían plantear a su proveedor de servicios de IA para asegurarse de que lleva a cabo las pruebas, la evaluación, la verificación y la validación (TEVV) adecuadas y de que existe una gobernanza suficiente en la validación del rendimiento y la precisión.

PREGUNTAS PARA PROVEEDORES

- ¿Cómo audita, prueba, evalúa, verifica y valida los resultados de sus modelos de IA?
- ¿Cómo se adaptan sus modelos a un panorama de amenazas externas en constante cambio y cómo reduce el riesgo de desviación de datos?
- ¿Incluye su flujo de trabajo de TEVV pruebas de rendimiento físico como la latencia, el consumo de memoria y los requisitos de energía y recursos?

Las respuestas a estas preguntas deben incluir referencias directas a un flujo de trabajo de TEVV, ya sea como parte de un AIMS existente (u otro sistema de gobernanza) o como un marco independiente.

Interpretabilidad, capacidad de ajuste y transparencia

Una parte importante de cualquier sistema de IA consiste en garantizar que los resultados de los datos sean interpretables, que se comprendan los pasos seguidos para obtenerlos y que los usuarios finales puedan confiar en que los resultados de la IA son correctos. Aunque un sistema de IA suele ser una caja negra, los usuarios deben poder comprender por qué ha llegado a una conclusión determinada, incluso si los pasos exactos permanecen ocultos.

En el mundo de la ciberseguridad, la confianza en un sistema de IA es esencial. Estos sistemas deben ejecutar acciones en tiempo real, actuando de forma autónoma para bloquear conexiones, rechazar mensajes y restringir el acceso a la red. Si no se puede confiar en que el sistema realice estas acciones de forma precisa y correcta, es posible que los usuarios finales desactiven las funciones o se inclinen a revertir las acciones, lo que reduce de manera crítica las capacidades defensivas del sistema. Los elevados requisitos de confianza en la ciberseguridad hacen que muchos de sus casos de uso no resulten adecuados para la automatización mediante IA generativa. La capacidad de ajuste de los sistemas de IA es una función importante.

Cada organización cuenta con un conjunto diferente de comportamientos, estructuras y culturas. La capacidad de ajustar los modelos de IA ayudará a alcanzar los objetivos globales y permitirá que una solución se adapte a una gama más amplia de casos de uso.

Las siguientes preguntas constituyen un punto de partida para las consultas que los compradores pueden plantear a su AISP con el fin de confirmar que su sistema será lo suficientemente interpretable y ajustable, y que fomentará una sólida relación de confianza entre los analistas humanos y las herramientas de IA.

PREGUNTAS PARA PROVEEDORES

- ¿Cómo garantiza que sus sistemas de IA sean interpretables?
- ¿Cómo fomenta una relación de confianza entre los analistas humanos y los resultados de la IA?
- ¿Cómo puede un equipo del SOC ajustar sus modelos para adaptarlos a sus políticas y requisitos empresariales específicos?

Un AISP debe ser capaz de describir cómo se puede adaptar su solución para ajustarse a la estructura, las políticas y las necesidades exclusivas de una organización, ya sea mediante el autoaprendizaje y la adaptación autónoma o a través de ajustes manuales disponibles para los usuarios finales.

También resulta valiosa la confirmación de que se tienen en cuenta la transparencia y la interpretabilidad. Un AISP debe proporcionar información sobre las políticas, los flujos de trabajo y los métodos con los que garantiza una total transparencia y fomenta la confianza en su solución.

Darktrace como proveedor de ciberseguridad de IA

Darktrace lleva más de una década creando y aplicando la IA a la ciberseguridad, desarrollando sus capacidades a la par de un panorama de amenazas cada vez más complejo y dinámico.

Esta experiencia se ha traducido en un enfoque maduro y multicapa de la IA, en el que la IA de autoaprendizaje de Darktrace aprende continuamente los patrones habituales de cada organización para comprender el comportamiento, interpretar el contexto e identificar desviaciones significativas, sin depender de reglas predefinidas ni firmas de ataque conocidas. Con el tiempo, esto ha facilitado una contrastada comprensión del comportamiento que ayuda a descubrir señales sutiles de riesgo que, de otro modo, podrían pasar desapercibidas.

Este enfoque se ha desarrollado y validado en entornos dinámicos y multidominio, donde la actividad abarca redes, identidades, nube, correo electrónico y aplicaciones. Al comprender el comportamiento a medida que se desplaza por los sistemas, Darktrace proporciona una visión continua de la actividad en toda la organización, lo que permite una detección, investigación y respuesta que se mantienen adaptativas a medida que evolucionan los entornos y las amenazas. Todo ello refleja un compromiso duradero con la innovación, basado en la aplicación y el perfeccionamiento en el mundo real, junto con un enfoque centrado en un desarrollo responsable y transparente de la IA.

Como pioneros en este ámbito, también nos hemos asegurado de que la responsabilidad sea un principio fundamental de nuestro proceso. Este compromiso se refleja en nuestra certificación ISO/IEC 42001, lo que demuestra la capacidad de Darktrace para ofrecer una IA de autoaprendizaje líder en el sector en aras de la resiliencia de la ciberseguridad.

Nuestros grupos de interés, clientes y partners pueden tener la seguridad de que Darktrace desarrolla sus sistemas de IA de manera responsable, ética y segura, y gestiona el uso de la IA en nuestras operaciones diarias de forma conforme a la normativa, segura y ética.

Cómo protege Darktrace los sistemas de IA

Darktrace ahora aporta estas capacidades para supervisar y responder ante el riesgo generado por los sistemas de IA en todas las organizaciones con Darktrace / SECURE AI. Esta solución analiza cómo se utilizan los prompts, los agentes y los sistemas en el contexto de cada organización, e integra cada interacción de IA en una única vista. Este enfoque exclusivo ayuda a los equipos a comprender la intención, evaluar el riesgo, proteger los datos confidenciales y aplicar políticas tanto en la actividad humana como en la de los agentes de IA. Proporciona visibilidad y control en tiempo real sobre la IA generativa, los agentes de IA, los entornos de desarrollo y la IA en la sombra, lo que permite a los equipos identificar el uso indebido, la configuración incorrecta y las desviaciones que los enfoques basados en reglas a menudo pasan por alto.

Al comprender el contexto y la intención entre usuarios y sistemas, ayuda a las organizaciones a mantener la supervisión a medida que la adopción de la IA se expande en toda la empresa.

Inscríbese en el Secure AI Readiness Program aquí: Esto le brinda acceso exclusivo a las últimas noticias sobre las amenazas de IA más recientes, actualizaciones sobre los enfoques emergentes que transforman la seguridad de la IA y análisis de las últimas innovaciones, incluido el trabajo continuo de Darktrace en este ámbito.

¿Está preparado para hablar con un experto de Darktrace sobre cómo proteger la IA? Regístrese aquí para recibir orientación práctica sobre los riesgos de la IA que más importan a su empresa, además de claridad sobre dónde centrarse primero en gobernanza, visibilidad, reducción de riesgos y preparación a largo plazo.



El enfoque de Darktrace hacia una IA responsable en ciberseguridad

Explore los principios que sustentan el enfoque de IA responsable de Darktrace, basado en la colaboración con expertos internacionales del ámbito académico y gubernamental, y descubra cómo integra la rendición de cuentas, la explicabilidad y la validación continua en su tecnología de ciberseguridad.

Lea más [aquí](#)

Arsenal de IA

Profundice en cómo el aprendizaje automático supervisado y no supervisado y los LLM pueden aplicarse a la ciberseguridad, y cómo Darktrace combina técnicas específicas en una arquitectura multicapa para potenciar a los equipos humanos de seguridad en el informe técnico «Arsenal de IA»

Lea más [aquí](#)



Conclusión

A medida que la IA se integra cada vez más en la ciberseguridad, el factor más importante para los líderes de seguridad no es si un producto utiliza IA, sino cómo se diseña, gobierna, valida y controla esa IA. Los compradores deben esperar respuestas claras sobre el origen y la protección de los datos, la selección de modelos, las prácticas de prueba y validación, la resiliencia frente a entradas adversarias y la transparencia de las decisiones basadas en IA.

Las soluciones de IA sólidas se basan en combinaciones de técnicas cuidadosamente seleccionadas, una evaluación continua y marcos de gobernanza que reducen el riesgo a lo largo del tiempo, lo que garantiza que los sistemas sigan siendo precisos, adaptables y alineados con las políticas organizativas a medida que evolucionan las amenazas.

Cuando se implementa de forma responsable, la IA ofrece ventajas significativas: una mejor detección de amenazas inéditas, investigaciones más rápidas y precisas, una menor carga operativa para los equipos de seguridad y una mayor confianza en las respuestas autónomas. El enfoque multicapa y de autoaprendizaje de Darktrace demuestra cómo diversas técnicas de IA, una sólida gobernanza y resultados explicables pueden combinarse para ofrecer una ciberseguridad fiable a escala.

Póngase en contacto con Darktrace para una demostración personalizada

[más información ↗](#)

■ **Acerca de Darktrace**

Darktrace es líder mundial en ciberseguridad basada en IA que permite a las organizaciones adelantarse cada día al panorama de amenazas en constante evolución. Fundada en 2013 en Cambridge (Reino Unido), Darktrace ofrece la plataforma de ciberseguridad esencial para proteger a las organizaciones frente a amenazas desconocidas mediante una IA que aprende de cada empresa en tiempo real. La plataforma y los servicios de Darktrace cuentan con el respaldo de más de 2.700 empleados que protegen a cerca de 10.000 clientes en todo el mundo. Para más información, visite www.darktrace.com.