Universidad de Buenos Aires

Facultades de Ciencias Económicas, Ciencias Exactas y Naturales e Ingeniería

Carrera de Especialización en Seguridad Informática

Trabajo Final

Aplicación de Inteligencia Artificial en el Proceso de Respuesta a Incidentes de Ciberseguridad: Avances, Desafíos y Perspectivas

Autor: Leonardo Perosio

Tutor: Mara Misto Macias

Año de presentación: 2024

Cohorte: 2023

Índice

1. Introducción

1.1. Contexto y Motivación

La ciberseguridad se ha convertido en una preocupación primordial para organizaciones de todos los tamaños y sectores, debido al aumento exponencial en la frecuencia, sofisticación y gravedad de los ataques cibernéticos. La protección de datos sensibles y la garantía de la integridad de los sistemas informáticos son cruciales en un entorno donde la interconexión digital y las nuevas tecnologías crecen a un ritmo acelerado.

1.1.1. La Evolución de las Amenazas Cibernéticas

En las últimas décadas, hemos sido testigos de una evolución significativa en el panorama de amenazas cibernéticas. Desde los primeros virus informáticos y gusanos, que eran en gran parte relativamente simples y se propagaban principalmente por medio de medios físicos como disquetes, hasta los sofisticados ataques dirigidos y ransomware actuales, que emplean técnicas avanzadas para explotar vulnerabilidades específicas. Estos ataques son realizados por una variedad de actores, desde delincuentes individuales hasta organizaciones criminales, y pueden tener consecuencias devastadoras, como la pérdida de datos, interrupciones operativas y daños a la reputación.

1.1.2. Importancia de una Respuesta Efectiva a Incidentes

La capacidad para detectar, analizar y responder rápidamente a incidentes de ciberseguridad es esencial para minimizar el impacto de los ataques. Un incidente mal gestionado puede llevar a consecuencias graves, como la exposición de datos personales, la pérdida de propiedad intelectual, daños económicos significativos y hasta la interrupción de servicios. En contraste, una respuesta efectiva puede limitar el alcan-

ce del daño, restaurar la normalidad operativa y proporcionar lecciones valiosas para mejorar la postura de seguridad de la organización.

1.1.3. Limitaciones de los Métodos Tradicionales

A pesar de la existencia de marcos robustos y herramientas de respuesta a incidentes, muchas organizaciones enfrentan desafíos significativos. Los métodos tradicionales, basados en análisis manuales y la experiencia de los analistas de seguridad, pueden ser insuficientes frente a la velocidad y la sofisticación de las amenazas modernas. Las limitaciones de los métodos tradicionales incluyen:

- **Escalabilidad:** La cantidad de datos y alertas generadas por sistemas modernos supera la capacidad de análisis manual.
- **Velocidad de Respuesta:** Los ataques cibernéticos pueden desarrollarse en minutos, requiriendo una respuesta rápida que los métodos tradicionales pueden no ser capaces de proporcionar.
- **Precisión:** La clasificación y el análisis manual de amenazas pueden llevar a errores humanos, como falsos positivos o negativos.

1.1.4. La Promesa de la Inteligencia Artificial

La Inteligencia Artificial (IA) ofrece un enfoque innovador para abordar estas limitaciones. La IA, particularmente a través del aprendizaje automático y el aprendizaje profundo, tiene el potencial de transformar el proceso de respuesta a incidentes de varias maneras:

- **Automatización:** La IA puede automatizar tareas repetitivas y analíticas, permitiendo que los equipos de seguridad se concentren en tareas más complejas y de alto nivel.

- **Detección y Análisis Avanzado:** Los modelos de IA pueden identificar patrones y anomalías que podrían ser pasados por alto por los humanos, mejorando la detección precoz de amenazas.
- **Respuesta Proactiva:** La IA puede prever y mitigar ataques potenciales antes de que se materialicen, basándose en datos históricos y patrones de comportamiento.

1.1.5. Motivación para el Trabajo

La creciente integración de la IA en la respuesta a incidentes de ciberseguridad destaca la necesidad de una investigación sobre cómo estas tecnologías están siendo aplicadas, cuáles son sus beneficios potenciales, y cuáles son los desafíos asociados. El trabajo en este campo es motivado por:

- **La Necesidad de Innovación:** Las amenazas cibernéticas evolucionan rápidamente, y la industria necesita soluciones innovadoras y avanzadas que puedan mantenerse al día.
- **La Demanda de Eficiencia:** Las organizaciones buscan mejorar la eficiencia y efectividad de sus operaciones de respuesta a incidentes para reducir costos y minimizar el impacto de los ataques.
- **La Evolución del Entorno Tecnológico:** A medida que las tecnologías emergentes, como la IA, continúan desarrollándose, es crucial entender cómo se pueden aplicar eficazmente para abordar problemas actuales y futuros en la ciberseguridad.

1.2. Objetivos del Trabajo

El propósito de este trabajo es examinar cómo la Inteligencia Artificial (IA) está revolucionando el proceso de respuesta a incidentes en ciberseguridad. Queremos

entender cómo la IA está siendo utilizada actualmente, identificar los avances más importantes, reconocer los desafíos que enfrentan las organizaciones y explorar qué nos depara el futuro en este ámbito.

1.2.1. Objetivos Específicos

Para alcanzar este objetivo principal, el trabajo se desglosa en varios objetivos específicos:

1. **Evaluar los Avances Recientes en la Aplicación de IA en Respuesta a Incidentes:**
    - Análisis de Tecnologías Actuales: Examinar las herramientas y sistemas basados en IA que se están utilizando actualmente en la respuesta a incidentes. Esto incluye la evaluación de tecnologías como sistemas de detección de anomalías, análisis predictivo, automatización de respuestas y correlación de eventos.
    - Estudio de Casos y Aplicaciones: Investigar cómo estas tecnologías se han aplicado en situaciones reales y qué resultados han producido. Los estudios de casos proporcionarán ejemplos concretos de la efectividad y el impacto de la IA en la respuesta a incidentes.
2. **Identificar y Analizar los Desafíos y Limitaciones:**
    - Desafíos Técnicos y Operativos: Investigar las dificultades técnicas asociadas con la implementación y el funcionamiento de soluciones de IA en ciberseguridad, como la calidad de los datos, la interpretación de resultados y la integración con sistemas existentes.
    - Problemas de Seguridad y Privacidad: Explorar los riesgos inherentes a la utilización de IA, incluyendo la posibilidad de nuevas vulnerabilidades y la gestión de la privacidad de los datos.

3. **Explorar las Perspectivas Futuras y Tendencias Emergentes:**

    - Tendencias en Innovación y Desarrollo: Analizar las innovaciones emergentes en el campo de la IA aplicada a la ciberseguridad y cómo estas podrían transformar la respuesta a incidentes en el futuro.
    - Proyecciones a Largo Plazo: Evaluar cómo la evolución de la IA podría influir en las estrategias y técnicas de respuesta a incidentes, y qué cambios se podrían anticipar en el panorama de la ciberseguridad.

4. **Proporcionar Recomendaciones para la Industria y la Investigación:**

    - Sugerencias Prácticas: Ofrecer recomendaciones basadas en los hallazgos del trabajo para la adopción y mejora de soluciones de IA en la respuesta a incidentes, dirigiéndose a profesionales y organizaciones del sector.
    - Áreas para Futuras Investigaciones: Identificar vacíos en el conocimiento y sugerir áreas para investigaciones adicionales que podrían beneficiar el desarrollo y la implementación de la IA en la ciberseguridad.

1.2.2. Impacto Esperado

El impacto esperado de este trabajo es proporcionar una visión clara y fundamentada sobre cómo la IA está transformando el proceso de respuesta a incidentes de ciberseguridad. Los hallazgos contribuirán a:

- Mejorar la Práctica Profesional: Ofrecer insights y recomendaciones prácticas para los profesionales de ciberseguridad, ayudando a optimizar sus estrategias de respuesta a incidentes.
- Informar la Toma de Decisiones: Facilitar la toma de decisiones informadas en la adopción e implementación de tecnologías de IA en el ámbito de la ciberseguridad.

- Avanzar en el Conocimiento Académico: Contribuir al cuerpo académico existente en ciberseguridad y IA, proporcionando una base para futuras investigaciones y desarrollos en el campo.

2. Marco Teórico, Normativo y Legal

2.1. Fundamentos de Ciberseguridad

La ciberseguridad se refiere a las prácticas, tecnologías y procesos diseñados para proteger sistemas informáticos, redes y datos de ataques, accesos no autorizados y daños. Su objetivo principal es garantizar la confidencialidad, integridad y disponibilidad de la información.

- **Confidencialidad:** Asegura que la información solo esté disponible para personas autorizadas.
- **Integridad:** Garantiza que la información no sea alterada de manera no autorizada.
- **Disponibilidad:** Asegura que los sistemas y datos estén accesibles para los usuarios autorizados cuando se necesiten.

2.2. Modelo de Respuesta a Incidentes de Ciberseguridad

La respuesta a incidentes de ciberseguridad es un concepto, un marco, un conjunto de soluciones y mucho más. La respuesta a incidentes ha crecido junto con Internet y otras tecnologías digitales y emergieron como una disciplina estructurada en la década de 1980, cuando la Universidad Carnegie Mellon estableció los primeros Equipos de Respuesta a Emergencias Informáticas (Computer Emergency Response Team CERT) [21].

Actualmente, la respuesta a incidentes es un enfoque estructurado para manejar y mitigar los efectos de los incidentes de ciberseguridad. Para esto, existen varios modelos definidos y utilizados y entre los más comunes se encuentran:

- **Modelo NIST:** El marco del Instituto Nacional de Estándares y Tecnología (NIST) para la respuesta a incidentes se basa en la guía del NIST Special Publication

800-61r2, que describe un proceso en cuatro fases: Preparación, Detección y Análisis, Contención, Erradicación y Recuperación, y Post-Incidente.

- **Modelo SANS:** El modelo del SANS Institute enfatiza un ciclo de respuesta a incidentes que incluye Identificación, Contención, Erradicación, Recuperación y Revisión. Este modelo provee un enfoque práctico el cual se puede encontrar en [12]

- **Modelo ISO/IEC 27035:** Esta norma internacional proporciona directrices para la gestión de incidentes de seguridad de la información, incluyendo la planificación, la detección y la respuesta.

2.3. Proceso de Respuesta a Incidentes

El proceso de respuesta a incidentes según el National Institute of Standards and Technology (NIST) está bien definido en la Guía para la Respuesta a Incidentes de Seguridad de la Información, específicamente en el documento NIST [19].

Las etapas del proceso de respuesta a incidentes, según el NIST, son las siguientes:

1. **Preparación (Preparation)**:En esta etapa, se trata de establecer y mantener una infraestructura para manejar incidentes, incluyendo la capacitación del personal y la implementación de políticas y procedimientos. Esta etapa incluye la formación de un equipo de respuesta a incidentes, la implementación de herramientas y la preparación de planes de respuesta. Referencia [19] sección 3.2.

2. **Detección y Análisis (Detection and Analysis):** Consiste en identificar y confirmar la ocurrencia de un incidente. Incluye la recopilación y análisis de datos para determinar la naturaleza y el impacto del incidente. Aquí se valoran las alertas, se realiza la correlación de eventos y se evalúa la magnitud del incidente. Referencia [19] sección 3.3.

3. **Contención, Erradicación y Recuperación (Containment, Eradication, and Recovery):** Esta fase abarca las acciones para limitar el impacto del incidente (contención), eliminar las causas raíz y vulnerabilidades (erradicación), y restaurar y validar la funcionalidad del sistema afectado (recuperación). Referencia [19] sección 3.4.

4. **Post-Incidente (Post-Incident Activity):** Después de que el incidente ha sido manejado, se realiza una revisión post-incidente para aprender de la experiencia, mejorar las respuestas futuras y actualizar los planes y procedimientos según sea necesario. Esto incluye el análisis de lo que salió bien y lo que se podría mejorar. Referencia [19] sección 3.5.

2.4. Inteligencia Artificial en Ciberseguridad

La Inteligencia Artificial (IA) se define como la capacidad de las máquinas para replicar funciones cognitivas humanas, tales como el aprendizaje, la toma de decisiones y la resolución de problemas. En esencia, el objetivo de la IA es desarrollar sistemas computacionales capaces de llevar a cabo tareas que típicamente requieren inteligencia humana, como identificar patrones, razonar lógicamente y comprender el lenguaje natural.

Las áreas dentro de la Inteligencia Artificial se dividen en:

- **IA débil o estrecha (Weak AI):** Se refiere a sistemas creados para realizar tareas específicas dentro de un ámbito definido. Estos sistemas carecen de conciencia y no pueden aprender de manera autónoma fuera de su área de aplicación. Ejemplos de esto son los sistemas de recomendación, el reconocimiento de voz, los chatbots y los sistemas de visión por computadora.

- **IA general (AGI - Artificial General Intelligence):** También conocida como inteligencia artificial fuerte, esta categoría se refiere a sistemas con la capacidad de entender, aprender y ejecutar cualquier tarea que un ser humano pueda realizar.

La IA general aspiraría a alcanzar una inteligencia comparable o incluso superior a la humana en todos los aspectos. En la actualidad, la IA general sigue siendo un objetivo en fase de investigación y desarrollo sin haber sido completamente alcanzada.

- **IA superinteligencia (ASI - Artificial Superintelligence):** Describe una forma de inteligencia artificial que supera de manera significativa la inteligencia humana en todos los aspectos. Los sistemas de IA superinteligente tendrían la habilidad de resolver problemas complejos de forma rápida y eficaz, y podrían tener un impacto profundo en la sociedad. La IA superinteligente es un concepto teórico que ha suscitado debates y especulaciones acerca de sus posibles implicaciones éticas y sociales.

2.5. Aplicaciones Actuales de IA en Ciberseguridad

La Inteligencia Artificial se está utilizando en diferentes aspectos de la Ciberseguridad como ser:

- **Detección de Amenazas** Las aplicaciones de IA en la detección de amenazas incluyen Sistemas de Detección de Intrusiones (IDS) que utilizan IA para identificar comportamientos anómalos y Análisis de Malware con herramientas basadas en IA para la clasificación de malware.

- **Análisis Forense** La IA en el análisis forense incluye la recopilación de Evidencias con herramientas para organizar datos digitales y el Análisis de Evidencias con automatización del análisis de datos

- **Respuesta a Incidentes** IA en la respuesta a incidentes se aplica en Automatización de Respuestas con la implementación de respuestas automáticas y en la Coordinación de Incidentes con la gestión de respuesta en tiempo real

- **Gestión de Vulnerabilidades** La IA en la gestión de vulnerabilidades es utilizada para la Evaluación de Vulnerabilidades con herramientas para priorizar y gestionar vulnerabilidades así también como la Automatización del Parcheo.

2.6. Normativas y Regulaciones

2.6.1. Respuesta a Incidentes

A nivel internacional, en referencia a la gestión de respuesta a incidentes de ciberseguridad se pueden encontrar:

- **ISO/IEC 27001**: Proporciona un marco para un Sistema de Gestión de Seguridad de la Información (SGSI).
- **NIST Special Publication 800-61 [19]**: Ofrece directrices para la gestión de la respuesta a incidentes. Esta normativa será tomada como base en gran parte de la información de este documento.

A nivel local, en Argentina, no existe regulaciones propias de esta temática a nivel general. En forma particular se encuentran:

- **Disposición DI-2023-3-APN-SSTI#JGM [6]:** Tiene como objetivo administrar y gestionar toda la información sobre reportes de incidentes de ciberseguridad ocurridos en los organismos del Sector Público Nacional. También proporciona una taxonomía de incidentes cibernéticos para facilitar el proceso de reporte, el posterior análisis y las fases de contención y erradicación de los mismos.
- **BCRA Comunicación A 7266 [3]:** Destinada a entidades bancarias, sugiere de una serie de prácticas efectivas de respuesta y recuperación ante ciberincidentes con el fin de limitar los riesgos en la estabilidad financiera e impulsar la ciberresiliencia del ecosistema en su conjunto.

Cabe destacar que existen varias leyes tangenciales como la de Protección de Datos Personales (Ley 25.326) y la Ley de Delitos Informáticos (Ley 26.388) que pueden aplicar según la forma del uso de la información.

2.6.2. Inteligencia Artificial

En el contexto de la ciberseguridad, La Unión Europea elaboró el Reglamento (UE) 2024/1689 [4] que representa un avance crucial en la regulación de la Inteligencia Artificial dentro de los países que la componen, estableciendo un marco que asegura el desarrollo y la implementación segura y éticas de tecnologías de IA. Este reglamento clasifica los sistemas de IA en cuatro niveles de riesgo: inaceptable, alto, moderado y bajo e impone requisitos específicos según el nivel de riesgo identificado. Por ejemplo, los sistemas de alto riesgo, que incluyen aplicaciones críticas en ciberseguridad, deben someterse a rigurosas evaluaciones de conformidad, mantener documentación exhaustiva y ser objeto de supervisión continua para garantizar su correcto funcionamiento y cumplimiento normativo. Además, el reglamento fomenta la creación de autoridades nacionales para la supervisión y establece sanciones significativas para garantizar la conformidad. A través de la regulación, el reglamento busca equilibrar la protección de los derechos fundamentales con el impulso a la innovación en el campo de la IA, apoyando a startups y promoviendo la cooperación internacional para una aplicación uniforme de las normas. Esta regulación es fundamental para la ciberseguridad, ya que establece directrices claras para el uso responsable de la IA en la protección contra amenazas y ataques cibernéticos.

A nivel local, en Argentina, no existen leyes de la temática de Inteligencia Artificial. La Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación emitió en el 2023 [5] una a convocatoria para busca impulsar el desarrollo de investigaciones científicas y tecnológicas en temas de IA y Ciencia de Datos y sus aplicaciones a los procesos de reconversión de la producción y la gestión.

3. Avances en la Aplicación de IA en la Respuesta a Incidentes

La Inteligencia Artificial (IA) está revolucionando la forma en que las organizaciones gestionan la respuesta a incidentes de ciberseguridad. La capacidad de los sistemas basados en IA para analizar grandes volúmenes de datos y detectar patrones ocultos permite una respuesta más rápida y precisa a amenazas emergentes. Los avances en este campo se centran en la automatización de tareas, la mejora de la precisión en la detección de amenazas y la optimización de la gestión de incidentes.

Las medidas tradicionales de ciberseguridad a menudo se basan en reglas y firmas predefinidas para identificar actividades maliciosas. Sin embargo, estos métodos son cada vez más ineficaces contra las ciberamenazas sofisticadas y en constante evolución. Las soluciones de ciberseguridad impulsadas por IA aprovechan los algoritmos de aprendizaje automático para analizar patrones, detectar anomalías e identificar amenazas previamente desconocidas en tiempo real. Al aprender continuamente de nuevos datos, los sistemas impulsados por IA pueden adaptar y mejorar sus capacidades de detección de amenazas, haciéndolos más resistentes contra las ciberamenazas emergentes, como los exploits de ”zero day”, ransomware y amenazas persistentes avanzadas (APT-Advanced Persistent Threat).

La IA no sólo mejora la detección de amenazas, sino que también permite adoptar un enfoque proactivo en materia de ciberseguridad. El análisis predictivo y los algoritmos de aprendizaje automático pueden analizar datos históricos para identificar vulnerabilidades potenciales y predecir vectores de ataque futuros. Al abordar de forma preventiva estas vulnerabilidades, se puede reducir significativamente su exposición al riesgo y fortalecer su postura general de ciberseguridad.

Una de las aplicaciones más prometedoras de la IA en ciberseguridad es la respuesta automatizada a incidentes. Los sistemas impulsados por IA pueden analizar y priorizar alertas, investigar incidentes de seguridad y orquestar una respuesta coordinada sin intervención humana. Esto no sólo acelera el tiempo de respuesta a inci-

dentes, sino que también minimiza el riesgo de error humano y garantiza un manejo coherente y eficaz de los incidentes de seguridad.

En esta sección estaremos enunciando los casos de uso más habituales en los que IA es aplicada en la actualidad. Los casos están organizados según las diferentes etapas del proceso de respuesta a incidentes indicado en la sección 2.3 según se ilustra en el siguiente gráfico:

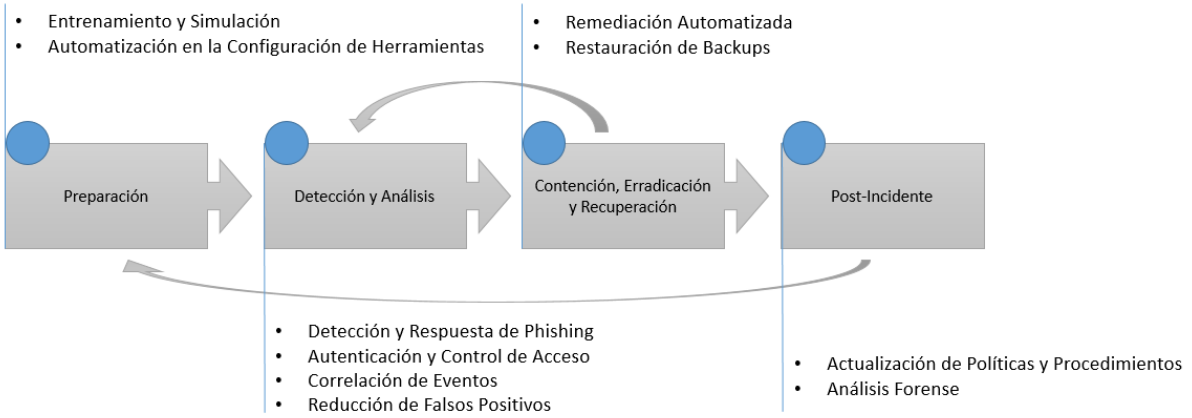

Figura 1: Uso de IA en Respuesta a Incidentes

3.1. Preparación

La inteligencia artificial en la etapa de preparación permite una configuración más efectiva de herramientas, una capacitación más robusta y adaptativa, y una evaluación continua y precisa de las políticas y capacidades de respuesta. Esto fortalece la capacidad de una organización para anticiparse a los incidentes y estar mejor preparada para manejarlos cuando ocurran.

3.1.1. Entrenamiento y Simulación

En las actividades de Entrenamiento y Simulación de la respuesta a incidentes, la inteligencia artificial (IA) proporciona herramientas avanzadas para mejorar la preparación del personal de seguridad y optimizar los ejercicios de respuesta a incidentes.

- **Escenarios Dinámicos:** Son aquellos que se ajusten en tiempo real a las nuevas amenazas y técnicas emergentes. La IA puede replicar técnicas y procedimientos utilizados por atacantes. Dos herramientas que podemos mencionar son **AttackIQ** y **Cymulate**.

- **Escenarios Personalizados:** La IA tiene la capacidad de crear programas de capacitación que se ajusten al nivel de conocimiento y habilidades de cada usuario ajustando el contenido de formación para abordar áreas de debilidad específicas. Un ejemplo de herramienta es **Cybrary**.

- **Análisis Post-Simulación:** Después de una simulación, la IA puede realizar un análisis del desempeño del equipo, identificando áreas de mejora y documentando lecciones aprendidas. Herramientas como **RangeForce** utilizan IA para analizar la respuesta durante los ejercicios y generar informes detallados.

Keepnet Labs [13] es un ejemplo completo de lo mencionado. Es una plataforma especialmente diseñada para el entrenamiento y la simulación de ataques, con un enfoque particular en el phishing. Utilizando IA, Keepnet Labs crea simulaciones altamente realistas que imitan ataques de phishing mediante la generación de correos electrónicos y sitios web falsos adaptados a los perfiles de los usuarios y a las técnicas de ataque más recientes. Esta capacidad de personalización asegura que las simulaciones reflejen con precisión las amenazas actuales y los escenarios de ataque más relevantes.

La plataforma proporciona una capacitación adaptativa que ajusta los módulos formativos en función del rendimiento individual de los empleados, ofreciendo retroalimentación inmediata y ajustando la dificultad de las simulaciones según las áreas de debilidad identificadas. Además, Keepnet Labs realiza un análisis exhaustivo de los resultados post-simulación, generando informes detallados que identifican riesgos, áreas de mejora y recomendaciones específicas para optimizar las políticas de seguridad y los procedimientos de respuesta.

La inteligencia artificial en Keepnet Labs también permite la actualización continua del contenido de formación para incorporar nuevas tácticas de ataque y vulnerabilidades emergentes, garantizando que el personal se mantenga al día con las amenazas más recientes. La automatización de la programación de entrenamientos regulares asegura una preparación constante y eficaz.

3.1.2. Automatización de la Configuración de Herramientas

La automatización de la configuración de herramientas es una actividad muy importante en la etapa de Preparación de la respuesta a incidentes, ya que establece las bases para una detección y respuesta eficiente. La inteligencia artificial desempeña un papel fundamental en esta área al facilitar la configuración y ajuste dinámico de herramientas de seguridad. A continuación, se detalla cómo la IA mejora esta actividad:

- **Adaptación en Tiempo Real:** La IA analiza grandes volúmenes de datos históricos y actuales para ajustar las configuraciones de las herramientas de seguridad incluyendo la actualización automática de reglas y políticas basadas en patrones emergentes de ataque y en la evolución de los comportamientos de los usuarios. Herramientas como **Splunk** e **IBM QRadar**, emplean algoritmos de aprendizaje automático para ajustar automáticamente las reglas de correlación y los umbrales de alerta en respuesta a cambios en el tráfico de red y patrones de datos.

- **Detección y Ajuste de Falsos Positivos:** Los algoritmos de IA ayudan a minimizar los falsos positivos ajustando las configuraciones de alerta en forma proactiva en función del análisis de incidentes previos. Esto se logra mediante la identificación de patrones que comúnmente causan falsos positivos y la adaptación de las reglas para reducir su incidencia. Por ejemplo, puede predecir cambios en los patrones de tráfico y ajustar las reglas de detección antes de que se conviertan en un problema significativo.

3.2. Detección y Análisis

Un campo en el que la inteligencia artificial destaca es en el análisis de información y detección de amenazas. La IA puede procesar grandes volúmenes de datos provenientes de diversas fuentes y detectar patrones anómalos en el comportamiento de los usuarios que podrían señalar un ciberataque. Por ejemplo, si un empleado, sin darse cuenta, hace clic en un correo electrónico de phishing, la IA puede detectar rápidamente el cambio en su comportamiento y generar una alerta sobre una posible violación de seguridad.

Cuando se identifica una amenaza potencial, los sistemas basados en IA pueden emitir alertas y notificaciones en tiempo real a los equipos de ciberseguridad, facilitando una respuesta rápida y eficaz. Al automatizar las acciones de respuesta a incidentes, como el aislamiento de los sistemas comprometidos o el bloqueo de actividades maliciosas, la IA reduce las oportunidades para que los atacantes lleven a cabo sus acciones y limita el impacto de una posible violación de seguridad.

3.2.1. Detección y Respuesta de Phishing

La IA utiliza técnicas de Procesamiento de Lenguaje Natural (NLP) para analizar el contenido de correos electrónicos y mensajes. Estos modelos pueden identificar patrones y frases comunes en ataques de phishing. También, los modelos de aprendizaje automático pueden identificar anomalías en los patrones de acceso a correos electrónicos o en la interacción con enlaces externos y examinar estos destinos validando si presentan características sospechosas. De esta manera, la IA puede automatizar el filtrado de mensajes sospechosos y bloquear automáticamente correos electrónicos o enlaces maliciosos antes de que lleguen al usuario y alertar a los administradores. Algunos ejemplos de herramientas que realizan estas acciones son:

- **Microsoft Defender for Office 365:** Integra capacidades de IA para proteger contra phishing y otras amenazas. Utiliza modelos de aprendizaje automático

para identificar correos electrónicos y enlaces sospechosos antes de que lleguen a la bandeja de entrada del usuario.

- **Barracuda Sentinel:** Usa IA para detectar y bloquear intentos de phishing y ataques de spear-phishing. Analiza patrones de comportamiento, contenido de mensajes y señales de amenazas para proporcionar protección en tiempo real.

3.2.2. Autenticación y Control de Acceso

La IA ofrece soluciones de autenticación más sólida que la tradicional considerando factores como la biometría y patrones de comportamiento del usuario a través de la utilización de algoritmos de aprendizaje automático. Algunos ejemplos de herramientas que realizan estas acciones son:

- **Apple y Google:** Tecnologías de reconocimiento facial y de voz, como las implementadas en aplicaciones de autenticación multifactor como FaceID de Apple y Google Authenticator, aprovechan la IA para proporcionar una capa adicional de seguridad que es difícil de eludir.

- **Darktrace:** Utilizan IA para detectar comportamientos anómalos en tiempo real, previniendo accesos no autorizados antes de que ocurran. Analiza patrones de datos para establecer un ”patrón de vida.específico para cada usuario e identificar anomalías emergentes que podrían indicar amenazas generando un enfoque de seguridad distinto y sin precedentes.

3.2.3. Correlación de Eventos

La correlación de eventos es una etapa crucial en la detección de amenazas y la respuesta a incidentes, ya que permite integrar y analizar datos provenientes de diversas fuentes para identificar patrones que podrían indicar un ataque. La inteligencia artificial ha revolucionado este proceso al mejorar significativamente la eficacia y la precisión en la correlación de eventos. Las herramientas de seguridad como **Splunk** e

IBM QRadar, emplean algoritmos de aprendizaje automático para consolidar y analizar datos de múltiples fuentes, incluyendo registros de servidores, tráfico de red y sistemas de gestión de eventos e información de seguridad (SIEM).

La IA facilita la correlación contextualizada de eventos al identificar relaciones entre eventos que podrían parecer aislados. Esto es particularmente importante para detectar ataques sofisticados que utilizan técnicas de múltiples etapas, como la explotación de vulnerabilidades combinada con movimientos laterales en la red. Los sistemas basados en IA pueden correlacionar eventos en tiempo real, proporcionando alertas más precisas y relevantes al correlacionar patrones de actividad sospechosa con indicadores conocidos de compromiso.

Además, la IA permite la correlación predictiva, donde los modelos de aprendizaje automático analizan datos históricos para prever posibles ataques futuros, mejorando la capacidad de respuesta proactiva. Al consolidar y analizar información de diversas fuentes, la IA optimiza la visibilidad y el contexto, permitiendo una identificación más eficaz de amenazas y una priorización adecuada de las respuestas.

3.2.4. Reducción de Falsos Positivos

La inteligencia artificial colabora fuertemente en la reducción de falsos positivos optimizando la precisión en la detección y el análisis de amenazas. Mediante el uso de algoritmos de aprendizaje automático, la IA puede ajustar dinámicamente las reglas de alerta y umbrales, correlacionar eventos de múltiples fuentes y proporcionar un contexto adicional para distinguir entre actividad legítima y maliciosa.

Herramientas como **Splunk** [14] implementan IA para adaptar las reglas de correlación y filtrar alertas, mientras que **Darktrace** utiliza modelos de comportamiento para detectar anomalías en tiempo real y reducir alertas erróneas. **IBM QRadar** integra datos y realiza un análisis contextualizado para mejorar la precisión en la detección de amenazas, ajustando las alertas y priorizando eventos críticos.

Estas tecnologías automatizan la respuesta inicial a incidentes y ajustan los um-

brales de alerta en función del aprendizaje continuo, optimizando la priorización de eventos para enfocarse en amenazas reales. Además, la IA incorpora retroalimentación para ajustar modelos y minimizar la incidencia de falsos positivos con el tiempo, mejorando la eficacia de la seguridad informática.

3.3. Contención, Erradicación y Recuperación

Automatización de Respuestas: La IA automatiza la respuesta a incidentes, como el aislamiento de sistemas afectados, el bloqueo de direcciones IP sospechosas o la aplicación de parches. Evaluación de Impacto: Los modelos de IA ayudan a evaluar el impacto del incidente y priorizar las acciones de contención y erradicación en función del riesgo.

3.3.1. Remediación Automatizada

La Remediación Automatizada de incidentes de seguridad informática se beneficia enormemente de la Inteligencia Artificial al mejorar la detección, respuesta y mitigación de amenazas de manera más eficiente y rápida.

La IA mejora la contención al automatizar el aislamiento de sistemas comprometidos para prevenir la propagación de ataques. Esto se logra mediante la identificación rápida de dispositivos afectados y la implementación de medidas automáticas, como la desconexión de la red o el bloqueo de accesos sospechosos. Herramientas como **Darktrace** y **CrowdStrike Falcon** ejemplifican esta capacidad. Darktrace utiliza algoritmos de IA para detectar anomalías en el tráfico de red y aislar sistemas comprometidos, mientras que CrowdStrike Falcon ofrece funcionalidades similares en sus endpoints, restringiendo procesos maliciosos de manera automática.

Respecto a la tarea de erradicación, la IA facilita la eliminación de amenazas al identificar y remover automáticamente componentes maliciosos, como malware o procesos dañinos, del sistema afectado. **IBM QRadar** automatiza la eliminación de malware detectado y aplica políticas para la remoción de amenazas, mientras que **Palo Alto**

Networks Cortex XSOAR integra capacidades para eliminar amenazas mediante la automatización de respuestas y la integración con diversas herramientas de seguridad.

La IA también juega un papel importante en la recuperación al automatizar la restauración de sistemas a su estado operativo normal. Esto incluye la aplicación de parches de seguridad, la restauración de configuraciones y la recuperación de datos desde copias de seguridad. **Cortex XSOAR** facilita esta recuperación al implementar procedimientos estandarizados para restaurar sistemas y validar la eliminación de amenazas, mientras que **CrowdStrike Falcon** ofrece capacidades para ayudar en la recuperación de sistemas comprometidos

3.3.2. Restauración de Backups

Luego de sufrir un ataque una de las formas de volver a estar operativos es mediante la restauración de backups. La Inteligencia Artificial puede facilitar una recuperación rápida y precisa colaborando en las siguientes tareas:

- **Detección de Datos Corruptos o Comprometidos:** La IA puede analizar la integridad de los backups para identificar datos corruptos o comprometidos antes de iniciar la restauración. Utiliza algoritmos de aprendizaje automático para detectar anomalías en los datos y asegurar que solo se restauren copias limpias y funcionales.

- **Automatización de la Restauración:** La IA facilita la automatización del proceso de restauración al seleccionar y aplicar automáticamente los backups más adecuados. Analiza las versiones de los datos y el estado de los sistemas para determinar la restauración más eficiente y efectiva.

- **Optimización del Tiempo de Recuperación:** La IA ayuda a optimizar los tiempos de recuperación al priorizar la restauración de datos críticos y garantizar que los sistemas esenciales vuelvan a estar operativos lo más rápido posible. Esto es crucial para minimizar el impacto en las operaciones empresariales.

- **Verificación Post-Restauración:** La IA verifica automáticamente la integridad y la funcionalidad de los datos restaurados, asegurando que todos los sistemas y aplicaciones funcionen correctamente después de la restauración y que no queden restos de vulnerabilidades. Ejemplos de Aplicaciones y Software

Veeam Backup & Replication es un software que utiliza IA para optimizar la restauración de backups mediante la automatización de la selección de copias de seguridad y la validación de la integridad de los datos. Su tecnología también permite la rápida recuperación de datos y sistemas críticos.

3.4. Post-Incidente

3.4.1. Actualización de Políticas y Procedimientos

Los modelos de IA pueden sugerir mejoras en las políticas de seguridad y los procedimientos basándose en patrones detectados en incidentes anteriores. También, la IA puede monitorear los cambios del sistema y actualizar automáticamente la documentación relevante [16]

3.4.2. Análisis Forense

El análisis forense en el contexto de seguridad informática es un proceso crítico que se realiza después de un incidente de seguridad para comprender cómo ocurrió el ataque, qué impacto tuvo y cómo se puede prevenir en el futuro. La Inteligencia Artificial permite automatizar y mejorar las capacidades de análisis forense de varias maneras [17]:

- **Recopilación y Procesamiento de Datos:** La IA ayuda a gestionar y procesar grandes volúmenes de datos generados durante y después del incidente. Utiliza técnicas avanzadas de procesamiento de lenguaje natural (PLN) y aprendizaje

automático para analizar registros de sistemas, archivos de eventos y comunicaciones de red, extrayendo información relevante y organizándola de manera coherente.

- **Identificación de Patrones y Correlación de Eventos:** Los algoritmos de IA pueden detectar patrones inusuales y correlacionar eventos aparentemente desconectados para identificar la secuencia de actividades que llevaron al incidente. Esta capacidad para ver patrones en grandes conjuntos de datos permite a los analistas entender cómo se ejecutó el ataque y qué vectores de entrada se utilizaron.

- **Análisis de Causa Raíz:** La IA facilita la identificación de la causa raíz del incidente mediante el análisis de datos históricos y el comportamiento de los sistemas afectados. Los modelos de IA pueden correlacionar las vulnerabilidades explotadas con las técnicas de ataque conocidas, ayudando a los investigadores a comprender cómo se comprometió el sistema.

- **Generación de Informes Automatizados:** La IA automatiza la creación de informes forenses detallados, que incluyen la cronología de eventos, la descripción de las vulnerabilidades explotadas y las acciones tomadas durante la respuesta al incidente. Estos informes son esenciales para la documentación del incidente, la auditoría y el cumplimiento normativo.

- **Análisis Predictivo y Recomendaciones:** Utilizando datos históricos de incidentes, la IA puede hacer predicciones sobre posibles futuros ataques y proporcionar recomendaciones sobre cómo fortalecer las defensas. Los sistemas de IA aprenden de cada incidente para mejorar continuamente las estrategias de prevención y respuesta.

IBM QRadar utiliza algoritmos de IA para correlacionar eventos de seguridad y detectar patrones en los datos de registro, proporcionando una visión integral del incidente. **Splunk**, por su parte, ofrece capacidades avanzadas de análisis de datos y

generación de informes automatizados, facilitando la investigación forense y la toma de decisiones basada en datos

3.5. Respuesta a Incidentes tradicional vs potenciada con IA

La siguiente tabla, tomada de [2], resume una comparación de diferentes actividades que se encuentran dentro del proceso de respuesta a incientes realizadas de la forma tradicional versus las realizadas con IA.

Aspecto	Tradicional	Potenciado con IA
Análisis de Datos	Depende en gran medida del análisis manual de registros de seguridad y alertas	Utiliza IA y aprendizaje automático para analizar y correlacionar datos de diversas fuentes en tiempo real
Clasificación de Incidentes	La clasificación suele ser manual y se basa en reglas y procedimientos predefinidos	Clasificación automatizada de incidentes utilizando algoritmos de IA para priorizar los incidentes según su gravedad, urgencia e impacto potencial
Ejecución de Tareas	Se basa en la ejecución manual de tareas por parte de analistas humanos	Automatiza tareas rutinarias como la generación de tickets, la validación de alertas y la asignación de recursos, reduciendo la intervención humana
Tiempo de Respuesta	El tiempo de respuesta puede ser más lento debido al análisis manual y la ejecución de tareas	Tiempos de respuesta más rápidos gracias al análisis en tiempo real y la ejecución automatizada de tareas, lo que reduce el tiempo para detectar y responder a incidentes
Escalabilidad	Escalabilidad limitada ya que las capacidades de respuesta dependen de la capacidad humana	Altamente escalable, capaz de manejar grandes volúmenes de incidentes con una mínima intervención humana
Adaptabilidad	Adaptabilidad limitada a los cambiantes panoramas de amenazas y la evolución de las técnicas de ataque	Se adapta a las amenazas en evolución aprendiendo continuamente de nuevos datos y actualizando las estrategias de respuesta en consecuencia
Toma de decisiones	Se basa en la toma de decisiones humana basada en la experiencia y procedimientos predefinidos	Respalda la toma de decisiones con conocimientos, recomendaciones y análisis predictivos basados en IA

Supervición de desempeño	El seguimiento del desempeño de la respuesta a incidentes puede ser manual y retrospectivo	Monitoreo e informes de desempeño automatizados, que brindan información en tiempo real sobre la efectividad de la respuesta y áreas de mejor
Mejora Contínua	La mejora depende del análisis de retroalimentación manual y del refinamiento del proceso	Facilita la mejora continua a través del análisis de datos de incidentes impulsado por IA, la identificación de tendencias y la optimización de estrategias de respuesta

Cuadro 1: Resolución de Incidentes Tradicional vs IA

4. Desafíos y Limitaciones

Dentro de las organizaciones, la dependencia excesiva de los sistemas con IA introduce una variedad de riesgos potenciales y el uso indebido de los mismos. A pesar de las pruebas y salvaguardas implementados por las organizaciones proveedoras, los sistemas de IA no son inmunes a errores o mal funcionamiento. El uso indebido de la tecnología de inteligencia artificial, ya sea intencional o no, puede tener consecuencias de gran alcance, desde violaciones de la privacidad hasta manipulación social y más.

En este capítulo abordaremos las problemáticas actuales y poteciales desde diversos aspectos organizados en cada una de las secciones del capítulo. Para cada aspecto ilustraremos con ejemplos reales y actuales las problemáticas con la que la población y las organizaciones se enfrentan al utilizar esta tecnología.

4.1. Precisión y Confiabilidad

Uno de los desafíos significativos en el campo de la ciberseguridad es la complejidad de los datos y el procesamiento de los mismos. Generalmente, los datos provienen de múltiples fuentes / sistemas entre los cuales se pueden encontrar registros de dispositivos (EndPoints), eventos de red, diversos sistemas y comportamientos de los usuarios (como ser lugares de conexión, horarios de trabajo, etc.). Para homogeneizar, estandarizar y relacionar datos, es fundamental la implementación de técnicas avanzadas de preprocesamiento de datos los cuales son requeridos para que los sistemas de IA utilicen información consistente. Esta preparación es esencial para mejorar la precisión y la eficacia del análisis automatizado en la detección y respuesta a incidentes.

Una restricción adicional en la automatización de tareas de seguridad mediante IA es la presencia de errores y sesgos en los algoritmos. Los modelos de IA pueden ser entrenados con datos que contienen sesgos o imprecisiones, lo que puede afectar tanto la exactitud como la equidad de los resultados y las decisiones del sistema. Es

crucial aplicar técnicas para mitigar y detectar sesgos en los algoritmos de IA revisando y validando los datos de entrenamiento para asegurar su calidad y representatividad, así como realizar un seguimiento continuo de los resultados y decisiones del sistema para identificar y corregir posibles sesgos o errores.

Por otro lado, muchos algoritmos de IA operan como “cajas negras”, lo que significa que su proceso de toma de decisiones no es conocido y es difícil de entender para los usuarios. La falta de visibilidad sobre cómo se toman las decisiones puede generar desconfianza, especialmente cuando las decisiones afectan áreas sensibles como la ciberseguridad.

4.2. Privacidad y Ética

La recopilación masiva de datos personales y/o sensibles para entrenar a los modelos de IA, sumada a la falta de transparencia de los algoritmos, genera preocupaciones sobre la protección de la privacidad y la posibilidad de sesgos algorítmicos que podrían llevar a decisiones injustas o discriminatorias. Además, la autonomía creciente de los sistemas de IA plantea interrogantes sobre la responsabilidad en caso de errores y sobre los límites éticos de la toma de decisiones automatizada. La falta de transparencia y la dificultad para explicar cómo los algoritmos llegan a sus conclusiones erosionan la confianza en los resultados. Por otro lado, la implementación de la IA en la ciberseguridad puede generar desigualdades al amplificar las brechas existentes en el acceso a la tecnología y a los datos.

Los siguientes son ejemplos en donde la Privacidad y Ética de las IA entraron en conflicto:

- **Recopilación y Uso de Datos Sensibles:** La IA a menudo requiere grandes volúmenes de datos para entrenar modelos y tomar decisiones informadas. Estos datos pueden incluir información personal sensible, lo que plantea riesgos para la privacidad si no se gestionan adecuadamente.

Por ejemplo, **Samsung Electronics** en el 2023 prohibió a los empleados usar asistentes de inteligencia artificial como ChatGPT debido a una filtración de códigos internos sensibles. Ver artículo completo en [8].

- **Consentimiento y Control del Usuario:** Obtener el consentimiento informado para el uso de datos personales puede ser complicado, especialmente cuando los datos se utilizan para entrenar modelos de IA. Además, los usuarios pueden tener un control limitado, o hasta no tenerlo, sobre cómo se utilizan sus datos.

  Por ejemplo, **Whastapp** desplegó Meta, su IA en diversos países. En Europa rige el Reglamento General de Protección de Datos donde se tienen en cuenta el “Derecho al olvido”, “Portabilidad de datos”, “Transparencia” y “Concentimiento”. Este último debe ser explícito para el procesamiento de los datos de cada personas. Baja este reglamento, los ciudadanos Europeos cuentan con la posibilidad de desactivar Meta de sus dispositivos. Por el contrario, en Argentina no se cuenta con la posibilidad de desactivar la IA. Teniendo en cuenta el uso que Meta hace de los datos personales, se efectuó una denuncia con la empresa [15] por el incumplimiento de la Ley 25.326.

4.3. Complejidad de Adopción

La adopción e implemnentación efectiva de la IA en ciberseguridad presenta una serie de desafíos y dificultades a afrontar tanto desde el punto de vista técnico como económico. Ya vimos en la sección anterior temás de éticas y de privacidad que forman parte de los desafíos de adopción. Desde el punto de vista técnico podemos enumerar los siguientes temas:

- **Integración con sistemas heredados:** La mayoría de las organizaciones cuentan con sistemas de seguridad legados que pueden ser difíciles de integrar con las nuevas tecnologías de IA.

- **Gestión de datos:** Requiere la recopilación, limpieza y estructuración de grandes volúmenes de datos para entrenar modelos de IA efectivos.
- **Desarrollo de modelos:** Crear modelos de IA precisos y robustos es un proceso complejo que requiere experiencia y recursos.
- **Mantenimiento y actualización:** Los modelos de IA requieren actualizaciones constantes para adaptarse a las nuevas amenazas y tecnologías emergentes.
- **Interpretabilidad:** Muchos modelos de IA, especialmente los basados en redes neuronales profundas, son cajas negras, lo que dificulta entender cómo llegan a sus conclusiones.

En caso que una organización desee o requiera implementar su propia IA, debe considerar un proceso de adopción más extenso y costoso. Entre las actividades a tener en cuenta se agregan:

- Investigación y desarrollo de algoritmos y modelos de aprendizaje automático
- Adquisición de equipos de alto rendimiento
- Reclutamiento y capacitación de profesionales capacitados en IA y ciberseguridad
- Configuración de la infraestructura de red segura y sistemas de almacenamiento de datos

4.4. Ataques de IA

Llamaremos ataques de IA a las actividades maliciosas en las que los atacantes utilizan técnicas de inteligencia artificial para explotar las vulnerabilidades de los sistemas. Estos ataques pueden implicar la manipulación de algoritmos de IA, manipular conjuntos de datos o crear entradas adversas que engañen a los modelos de IA. Existen varios métodos que los atacantes pueden utilizar, entre los cuales se encuentran:

- **Ataques de IA basados en aprendizaje automático:** El aprendizaje automático puede automatizar la identificación de puntos débiles en una red, haciendo que los ataques sean más eficientes y difíciles de detectar.
- **Ataques adversarios de IA:** Los ataques adversarios implican manipular las entradas de un sistema de inteligencia artificial para provocar que tome decisiones incorrectas.
- **Ataques de evasión de IA:** Estos ataques ocurren cuando los atacantes manipulan las entradas de datos para evitar su detección. Por ejemplo, alterar firmas de malware para evitar el software antivirus basado en aprendizaje automático o modificar los patrones de tráfico de la red para evadir los sistemas de detección de intrusiones.
- **Explotación de las vulnerabilidades de la inteligencia artificial:** Los sistemas de inteligencia artificial tienen vulnerabilidades inherentes que los atacantes pueden aprovechar. Estos incluyen sesgos en los datos de entrenamiento, sobreajuste y debilidades en las arquitecturas de modelos.

Detectar ataques de IA es un desafío debido a su naturaleza en constante evolución. Actualmente, las organizaciones basan sus defensas de ciberseguridad en métodos tradicionales que a menudo no logran identificar y mitigar los ataques de IA. A medida que los ataques de IA se vuelven más sofisticados, las limitaciones de estas defensas tradicionales se volverán cada vez más evidentes.

Para mantenerse al día con las sofisticadas técnicas de ataque de IA, las organizaciones deben adoptar estrategias proactivas. Estas incluyen actualizaciones periódicas de los modelos de IA, la incorporación de inteligencia sobre amenazas en los marcos de seguridad y la inversión en investigación y desarrollo para anticipar futuros vectores de ataque. Otro factor muy importante es la formación y la educación continua de sus equipos de ciberseguridad para mantener un alto nivel de preparación. En otras palabras, las organizaciones deben pasar de una defensa reactiva a una defensa proactiva.

5. Perspectivas Futuras

La incorporación de la Inteligencia Artificial (IA) en la gestión de incidentes de ciberseguridad ha marcado un progreso considerable en el área. Sin embargo, este campo sigue desarrollándose rápidamente, ofreciendo un panorama lleno de oportunidades y retos que influirán en el futuro de la ciberseguridad. En este capítulo, se examinarán las perspectivas futuras de la IA en esta área, resaltando las tendencias emergentes y las oportunidades que se presentan.

5.1. Camino a un IA confiable

La IA confiable intenta solucionar los desafíos y limitaciones indicados en el capítulo anterior. Este concepto, abarca varios aspectos incluyendo la transparencia, privacidad de datos, explicabilidad de la información, la responsabilidad del uso y la robustez. La transparencia en la IA busca hacer que los procesos y decisiones de estos sistemas sean comprensibles tanto para los usuarios como para las partes interesadas. La rendición de cuentas implica establecer protocolos adecuados para gestionar resultados adversos o sesgos, asegurando que haya una supervisión y una remediación efectivas.

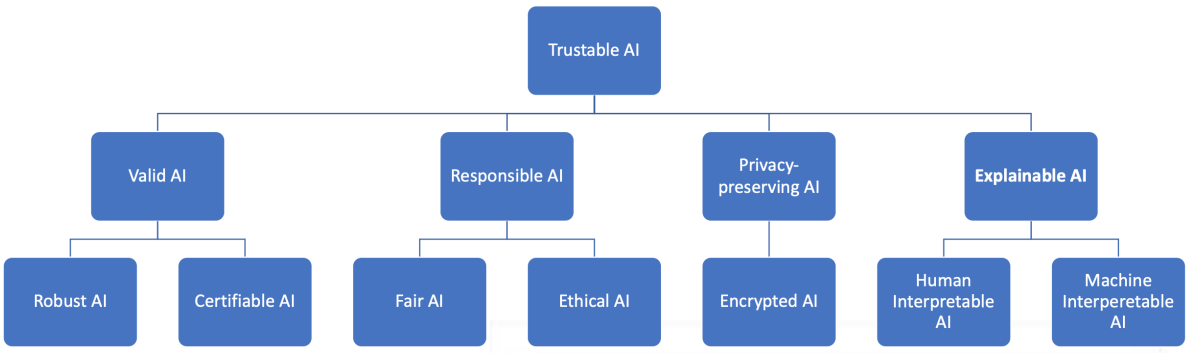

Figura 2: Conceptos de IA[1]

En el camino a una IA confiable se está trabajando en diferentes conceptos de IA.

Explicaremos a continuación los conceptos de la segunda línea:

- **Valid AI:** La IA válida o también conocida como IA confiable, es aquella que no solo funciona correctamente, sino que también es segura, justa, transparente y responsable. Esta IA cumple con las siguientes características:
  - **Robustez:** Los sistemas deben ser resistentes a ataques y errores, y funcionar de manera confiable incluso en condiciones adversas.
  - **Seguridad:** Deben proteger los datos de los usuarios y garantizar la privacidad.
  - **Equidad:** Los algoritmos deben evitar sesgos y discriminaciones, tratando a todos los usuarios de manera justa. Transparencia: Los procesos de toma de decisiones de la IA deben ser comprensibles y explicables.
  - **Responsabilidad:** Los desarrolladores y usuarios de la IA deben ser responsables de sus acciones y de las consecuencias de su uso.
- **Responsible AI:** La inteligencia artificial (IA) responsable es un conjunto de principios que ayudan a guiar el diseño, el desarrollo, la implementación y el uso de la IA, generando confianza en las soluciones de IA que tienen el potencial de empoderar a las organizaciones y a sus partes interesadas. La IA responsable implica la consideración del impacto social más amplio de los sistemas de IA y las medidas necesarias para alinear estas tecnologías con los valores de las partes interesadas, las normas legales y los principios éticos. La IA responsable tiene como objetivo incorporar dichos principios éticos en las aplicaciones y flujos de trabajo de la IA para mitigar los riesgos y los resultados negativos asociados con el uso de la IA, al tiempo que se maximizan los resultados positivos [10].
- **Privacy-preventing AI:** La IA que preserva la privacidad permite a las organizaciones aprovechar el poder de la IA mientras protegen los datos confidenciales y respetan los derechos de privacidad individuales [22].

En el marco de respuesta a incidentes, los algoritmos de IA requieren cantidades significativas de datos para identificar amenazas potenciales y aprender de patrones. Estos datos suelen contener información privada y sensible que debe ser protegida y tratada correctamente. Garantizar que las IA utilicen los datos de forma ética y cumplan con las normas de protección de datos es una preocupación constante en las organizaciones.

- **IA Explicable:** La inteligencia artificial explicable (XAI) es una clasificación de la IA en la cual permite a los usuarios comprender las razones exitentes detrás de las decisiones tomadas por los modelos de IA, facilitando la confianza y la depuración [1] [9]. Con la IA explicable, una empresa puede solucionar problemas y mejorar el rendimiento de los modelos al tiempo que ayuda a las partes interesadas a comprender los comportamientos de los modelos de IA entendiendo y eliminando los sesgos. En el proceso de gestión de incidentes, XAI es muy importante en las actividades de: [18]

  - **Detección y análisis de amenazas:** Permite comprender cómo evalúa las alertas de un IDS o SIEM y esto ayuda a los analistas a priorizar y responder a una amenaza de manera más eficiente.
  - **Automatización y respuesta de seguridad:** Permite comprender las reglas permite a las organizaciones puedan perfeccionarlas, ser más efectivas y minimizar falsos positivos.

  Según el análisis de mercado realizado en [11], el tamaño del mercado de IA explicable se estima en 8,63 mil millones de dólares en 2024, y se espera que alcance los 21,19 mil millones de dólares en 2029, creciendo a una tasa compuesta anual del 19,69 % durante el período previsto (2024-2029).

5.2. Tendencias Emergentes

Se enumeran a continuación algunas tendencias emergentes que podrían generar un gran impacta al momento de hacerse realidad.

- **Blockchain y IA:** Blockchain e Inteligencia artificial (IA) son dos tecnologías que están en están en constante evolución y desarrollo y, aunque son distintas en sus aplicaciones y funcionalidades, pueden complementarse y beneficiarse mutuamente en ciertos contextos. Su convergencia tiene el potencial de transformar diversas actividades como la gestión y análisis de datos.

  Una de las posibles convergencia de ambas tecnologías es en la privacidad y manejo de datos. Blockchain generalmente se asocia con la protección de datos descentralizada mientras que la IA puede requerir regulaciones específicas para el manejo ético de datos y la privacidad.

- **IA Auto-Adaptativa:** Representa una evolución significativa en el campo de la ciberseguridad. A diferencia de los sistemas tradicionales basados en reglas estáticas, la IA auto-adaptativa emplea algoritmos de aprendizaje automático para analizar de manera continua varios conjuntos de datos, identificando patrones y correlaciones que pueden indicar un comportamiento anómalo. Esta capacidad de auto-aprendizaje continuo permite a los sistemas de IA adaptarse de forma proactiva (sin la intervención humana ni la necesidad de recursos adicionales) a las tácticas en constante evolución de los ciberatacantes, lo que resulta en una mayor resiliencia ante amenazas emergentes. En la reciente publicación [20] se muestra un nuevo modelo desarrollado por especialistas del laboratorio de investigación de IA de T-Bank y el Instituto AIRI de Moscú, denominado 'Headless-AD', el cual permite crear sistemas de IA que pueden adaptarse a su entorno.

- **IA Cuántica:** La integración de la computación cuántica, la inteligencia artificial y la ciberseguridad podría anunciar una nueva era de avance tecnológico [7]. Si bien la computación cuántica plantea desafíos importantes a los sistemas de

seguridad actuales, también ofrece oportunidades para crear soluciones impulsadas por IA más seguras y eficientes. La computación cuántica tiene un inmenso potencial para transformar la seguridad cibernética al mejorar la prevención de amenazas, acelerar la recuperación de amenazas y reforzar la resiliencia cibernética. Al aprovechar los principios cuánticos, las organizaciones pueden desarrollar métodos de cifrado más seguros, detectar y responder rápidamente a las amenazas y mantener la continuidad del negocio frente a los ciberataques. A medida que la tecnología cuántica siga evolucionando, su integración con la IA y la ciberseguridad será crucial para crear sistemas de seguridad robustos, adaptables y preparados para el futuro. Particularmente para el caso de respuesta a incidentes, esta combinación podría ser muy efectiva para:

- **Tener una postura de fedensa sólida y adaptable:** la computación podría predicir y mitigar las amenazas antes de que puedan causar daños significativos adaptandose en tiempo real gracias a su capacidad de cómputo.
- **Continuidad del Negocio:** la velocidad y la eficiencia de la computación cuántica a la hora de analizar y responder a las amenazas garantizan un tiempo de inactividad mínimo.
- **Cifrado resiliente:** La computación cuántica podría ayudar a los sistemas de defensa impulsados por IA a contar con métodos dinámicos de encriptación que podrían cambiar en tiempo real ante un ataque lo cual dificultaría a los atacantes romper las diferentes encriptaciones utilizadas en la organización.

5.3. Recomendaciones para Futuras Investigaciones

Además de los temas indicados en este capítulo, en esta sección agregaremos otras temáticas que no fueron aboradadas durante el trabajo y que podrían ser un campo de investigación.

- **Personalización de la IA:** Investigar cómo desarrollar modelos de IA que puedan adaptarse a las características de seguridad específicas de cada organización, considerando su tamaño, sector, infraestructura, tecnología y tipo de amenazas.
- **Impacto de Redes 5G y 6G:** Analizar cómo la implementación de redes 5G y 6G afectará la seguridad de las organizaciones y cómo la IA puede ayudar a proteger estas redes ante ataques que aprovechen esta tecnología.

6. Conclusiones

La incorporación de la Inteligencia Artificial (IA) en la respuesta a incidentes de ciberseguridad empieza a ser crucial para las organizaciones. Las capacidades de IA, como el aprendizaje automático y el análisis de grandes volúmenes de datos en tiempo real, han mejorado significativamente la velocidad y precisión en la detección de amenazas y la respuesta a incidentes. Los sistemas de IA pueden identificar patrones anómalos y comportamientos sospechosos con una eficiencia que supera a los métodos tradicionales, permitiendo una reacción más ágil ante posibles ataques.

La implementación de soluciones basadas en IA están contribuyendo a reducir la carga operativa de los equipos de ciberseguridad. Las herramientas automatizadas para la clasificación de incidentes y la generación de informes están optimizando el proceso de gestión, permitiendo a los analistas enfocarse en tareas más estratégicas y menos repetitivas.

A pesar de los grandes avances que podemos ver y utilizar actualmente, la integración de la IA en la respuesta a incidentes presenta varios desafíos. La adversidad de los atacantes que también utilizan IA para evadir la detección exige una constante actualización y adaptación de los sistemas de seguridad. Los riesgos asociados con la privacidad y la integridad de los datos son también una preocupación significativa que debe ser gestionada con cuidado. Otro factor de riesgo a destacar es la posible dependencia excesiva en sistemas automatizados creyendo que no es requerida la intervención humana en la resolución de problemas.

La aplicación de IA en ciberseguridad plantea importantes cuestiones éticas y regulatorias. Es fundamental establecer políticas claras sobre el uso de algoritmos y la gestión de datos sensibles para evitar abusos y garantizar la protección de la privacidad. La conformidad con regulaciones y normativas debe ser una prioridad para asegurar que las tecnologías de IA se implementen de manera responsable y en alineación con las leyes vigentes (locales e internacionales).

Las perspectivas futuras prometen avances aún más significativos que los actua-

les. La investigación y el desarrollo continuo en el ámbito de IA y ciberseguridad son esenciales para abordar los desafíos emergentes y mejorar las capacidades de respuesta. La colaboración entre la academia, la industria y las entidades reguladoras será crucial para fomentar la innovación y asegurar la eficacia y ética en el uso de la IA en ciberseguridad.

Para maximizar los beneficios de la IA en la respuesta a incidentes de ciberseguridad, es recomendable realizar una implementación consciente, gradual y bien planificada. Tener definido porqué se utilizará una herramienta con IA y cuál va a ser el alcance de la misma así también como las posibilidades de adaptación a la organización, es crucial al momento de su adquisición e implementación.

“Si piensas que la tecnología puede solucionar tus problemas de seguridad, está claro que no entiendes los problemas ni entiendes la tecnología” Bruce Schneier

Referencias

[1] Explainable ai. https://xaitutorial2019.github.io/, 2019. (consultada el 21/09/2024).

[2] LeewayHertz AI Development Company. Ai in incident response: Exploring use cases, solutions and benefits. https://www.leewayhertz.com/ai-in-incident-response/, 2024. (consultada el 25/08/2024).

[3] Banco Central de la República Argentina. Lineamientos para la respuesta y recuperación ante ciberincidentes (rrci). https://www.bcra.gob.ar/pdfs/comytexord/A7266.pdf, 2021. (consultada el 12/08/2024).

[4] Parlamento de la Unión Europea y Consejo de la Unión Europea. Reglamento (ue) 2024/1689 del parlamento europeo y del consejo. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689, 2024. (consultada el 20/09/2024).

[5] Agencia Nacional de Promoción de la Investigación el Desarrollo Tecnológico y la Innovación. Picto ia. https://www.argentina.gob.ar/servicio/picto-ia, 2023. (consultada el 19/08/2024).

[6] Centro Nacional de Respuesta a Incidentes Informáticos (CERT.ar.). Disposición di-2023-3-apn-sstijgm. https://www.boletinoficial.gob.ar/detalleAviso/primera/289746/20230706, 2023. (consultada el 11/08/2024).

[7] DEFIANCE. In-context reinforcement learning for variable action spaces. https://www.defianceetfs.com/quantum-computing-and-ai-the-new-frontiers-of-cybersecurity/, 2024. (consultada el 01/10/2024).

[8] Mark Gurman. Samsung bans staff's ai use after spotting chatgpt data leak. https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-

and-other-generative-ai-use-by-staff-after-leak?embedded-checkout=true, 2023. (consultada el 15/09/2024).

[9] IBM. ¿qué es la ia explicable? https://www.ibm.com/es-es/topics/explainable-ai. (consultada el 21/09/2024).

[10] IBM. What is responsible ai? https://www.ibm.com/topics/responsible-ai, 2024. (consultada el 22/09/2024).

[11] Mordor Intelligence. Explainable ai market size & share analysis - growth trends & forecasts (2024 - 2029). https://www.mordorintelligence.com/industry-reports/explainable-ai-market/market-size, 2024. (consultada el 18/09/2024).

[12] Patrick Kral. The incident handler book. https://sansorg.egnyte.com/dl/6Btqoa63at, 2011. (consultada el 10/08/2024).

[13] Keepnet Labs. Keepnet labs - home page. https://keepnetlabs.com/, 2024. (consultada el 21/09/2024).

[14] Xiao Lin. Reduce false alerts – automatically! `https://www.splunk.com/en_us/blog/security/reduce-false-alerts-automatically.html`, 2024. (consultada el 17/09/2024).

[15] Daniel Monastersky. Argentina: Denuncian a meta por usar datos privados de whatsapp y otras redes para entrenar su ia. https://www.linkedin.com/pulse/argentina-denuncian-meta-por-usar-datos-privados-de-y-monastersky-riz2f/, 2024. (consultada el 28/09/2024).

[16] Magnus Oxenwaldt. Ai fighting ai: The future of cybersecurity - are you ready? https://www.columbusglobal.com/en/blog/ai-fighting-ai-the-future-of-cybersecurity, 2024. (consultada el 02/09/2024).

[17] Satyendra Pandey. Ai forensics: A new era in digital investigation techniques. https://www.linkedin.com/pulse/ai-forensics-new-era-digital-investigation-techniques-pandey-lzukc/, 2024. (consultada el 18/09/2024).

[18] Ace HR Partners. Xai for information security: Seeing through the fog. https://www.linkedin.com/pulse/xai-information-security-seeing-through-fog-acehrpartners-jzkse/, 2024. (consultada el 18/09/2024).

[19] Paul Cichonski, Tom Millar, Tim Grance and Karen Scarfone. NIST SP 800-61 Rev. 2: Computer Security Incident Handling Guide. `https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf`, 2012. (consultada el 10/08/2024).

[20] Viacheslav Sinii, Alexander Nikulin, Vladislav Kurenkov, Ilya Zisman, and Sergey Kolesnikov. In-context reinforcement learning for variable action spaces. https://arxiv.org/pdf/2312.13327v6, 2024. (consultada el 22/09/2024).

[21] Teresa Thomas and Carnegie Mellon Today Kelly Kimberkand. U.s. department of homeland security announces partnership with carnegie mellon's cert coordination center. `https://www.cmu.edu/cmnews/extra/030915_cyberpartner.html`, 1988. (consultada el 10/09/2024).

[22] Dial Zara. Privacy-preserving ai: Techniques frameworks. https://dialzara.com/blog/privacy-preserving-ai-techniques-and-frameworks/, 2024. (consultada el 04/10/2024).