Número 106 | Septiembre 2025

NTT DATA

Radar

El magazine de ciberseguridad

Adopción de la IA: ¿estamos preparados para los desafíos de ciberseguridad?

Por David Sandoval Rodríguez-Bermejo

La inteligencia artificial (IA) es, sin lugar a dudas, la tecnología que está en boca de todos actualmente. No importa el sector en el que trabajemos o el caso de uso que queramos resolver, si no tiene IA, no es relevante. Al menos, esa es la postura de la gran mayoría de las empresas actuales.

Si bien es cierto que la adopción de la inteligencia artificial generativa se ha posicionado como una de las más rápidas en la historia de la humanidad, la realidad es que esa adopción se ha llevado a cabo de una forma bastante precaria desde el punto de vista de la ciberseguridad. Un *stack* tecnológico nuevo, con capacidad de "razonar" y fácilmente automatizable, es la panacea de cualquier gestor ya que sus posibilidades son casi ilimitadas. Si a esta circunstancia le añadimos que cada vez el ser humano tiene mayores competencias digitales (nativo digital) y que la IA generativa se ha democratizado gracias a su facilidad de uso, tenemos un caldo de cultivo bajo el cual se esconden oportunidades y riesgos, a partes iguales.

No me quiero detener frente a las oportunidades porque son de sobra conocidas por todos, solo hace falta abrir cualquier periódico y navegar por alguna noticia del sector tecnológico. El problema reside en los riesgos, y en que la mayor parte de las personas que implementan soluciones de IA (gracias, en gran medida a su democratización) no poseen los conocimientos de ciberseguridad adecuados.

Por poner algunos ejemplos rápidos, hablamos de integraciones con IAs que exfiltraban la información de la compañía a terceros, un mal bastionado o gestión de los flujos de información que han permitido manipular a LLMs (modelos de lenguaje avanzados) para conseguir productos a precios ridículos, o robos (exfiltraciones) de modelos enteros (como ocurrió con la primera versión de Llama), que pueden poner en jaque a toda una organización debido a la exposición de su propiedad intelectual.

Precisamente, de un riesgo no calculado, surgió un nuevo paradigma dentro del mundo de la inteligencia artificial: los modelos LLMs abiertos. Gracias a la exfiltración del primer modelo de Llama, Meta decidió abrir su modelo al mundo y publicar los pasos del mismo haciendo que muchas otras empresas replicasen su estrategia. De esta forma, se han constituido grandes comunidades como HuggingFace, que actúa tanto como un centralizador de repositorios de modelos de IA como un recurso vertebral para la formación en inteligencia artificial. Todo de forma gratuita. Este giro en el mercado actual ha añadido una capa de competencia extra a los principales proveedores (*players*) como Google, OpenAI o Anthropic. que han visto que existe una gran competencia con los modelos abiertos.

Por otro lado, existe una nueva tendencia dentro de la IA que consiste en definir pequeños operadores (agentes) que se organizan entre sí para cumplir una determinada tarea gracias a la especialización.

Al igual que sucedió con el lanzamiento de los modelos abiertos, el mundo entero se ha lanzado a por los agentes sin tener en cuenta las medidas de seguridad oportunas para desplegar sistemas seguros y bien bastionados. Si bien es cierto que se han puesto en práctica muchas de las lecciones aprendidas de los primeros LLMs, aún queda bastante camino por recorrer.

Ante este nuevo escenario, cabe preguntarse qué nos deparará el futuro dentro del mundo de la ciberseguridad. Aquí se nos abre un escenario bastante complejo con innumerables retos: la seguridad en los modelos de lenguaje, la gestión y el bastionado de los agentes de IA, la seguridad en las cadenas de suministro (especialmente en las que tengan una gestión parcial de IA…), etc.

Por último, a nivel de negocio, me gustaría comentar un cambio de paradigma que poco a poco se está viendo en el mercado actual. Los proveedores de productos, apalancados en la IA, se están transformando y reinventando, pasando de ser simples proveedores de producto a proveedores de servicios gracias a la inclusión de servicios de IA en su infraestructura. De esta forma, consiguen por un lado ingresos por la suscripción de servicios nuevos y, por otro lado, ingresos pasivos por la compra de *tokens* para la ejecución de dichos servicios. Dentro del mundo de la ciberseguridad, un ejemplo de este cambio de estrategia lo tenemos en la herramienta Burp para análisis dinámicos, que han incluido su suite de Burp AI para acelerar la ejecución de auditorías.

Para terminar, se puede afirmar que estamos ante una época de grandes desafíos, especialmente en el ámbito de la inteligencia artificial y la ciberseguridad. La rápida adopción de la IA en el sector empresarial está generando tanto oportunidades y desafíos como riesgos. Por ello, es fundamental que las organizaciones implementen políticas de seguridad efectivas y fomenten la formación continua en ciberseguridad e inteligencia artificial.

David Sandoval Rodríguez-Bermejo
Cybersecurity evangelist

La realidad supera al código: IA, fraudes, robots y la guerra cuántica

Cibercrónica por Rodrígo Rey

La inteligencia artificial ya puede replicar tu cara, tu voz, tus gestos... y tus claves de acceso. En palabras del propio Altman, las barreras biométricas están siendo superadas por sistemas capaces de engañar las señales sensoriales que utilizan nuestras interfaces de autenticación.

Esto significa que estamos a las puertas de una crisis de fraude de identidad masiva, donde las credenciales ya no valdrán nada si pueden ser falsificadas por una IA que te imita mejor que tú mismo.

La ciberseguridad entra así en una nueva era: el "quién eres" ya no es suficiente para demostrar que eres tú. La confianza digital, como la conocíamos, está comprometida. Prepárate para lo que viene.

En el otro lado del espectro, también hay buenas noticias: la inteligencia artificial no solo clona tu identidad, también podría salvarte la vida antes de que tu corazón decida apagar el sistema. Desde la Universidad Johns Hopkins, los investigadores han creado MAARS, un modelo de IA capaz de analizar imágenes cardíacas y predecir paros cardíacos con un 89% de precisión.

Esto supera con creces la capacidad humana y clínica tradicional, al identificar señales sutiles en el tejido cardíaco que suelen pasar desapercibidas. Ciencia ficción hace cinco años. Medicina de precisión hoy.

Japón se planta. Y no con bonsáis, sino con cúbits. Su ambición es clara: destronar a EE.UU. y China en la carrera de la computación cuántica para 2030.

El arma: un superordenador cuántico desarrollado por el centro RIKEN y Fujitsu que ya opera con 256 cúbits y promete ser un 25% más potente que la bestia actual de IBM.

Esta movida no es sólo geopolítica. Es estratégica: el país que domine la computación cuántica podrá romper criptografía, acelerar IA y rediseñar la seguridad global.

Así que, si pensabas que el ransomware era lo peor, espera a que se masifiquen los ataques con procesamiento cuántico detrás.

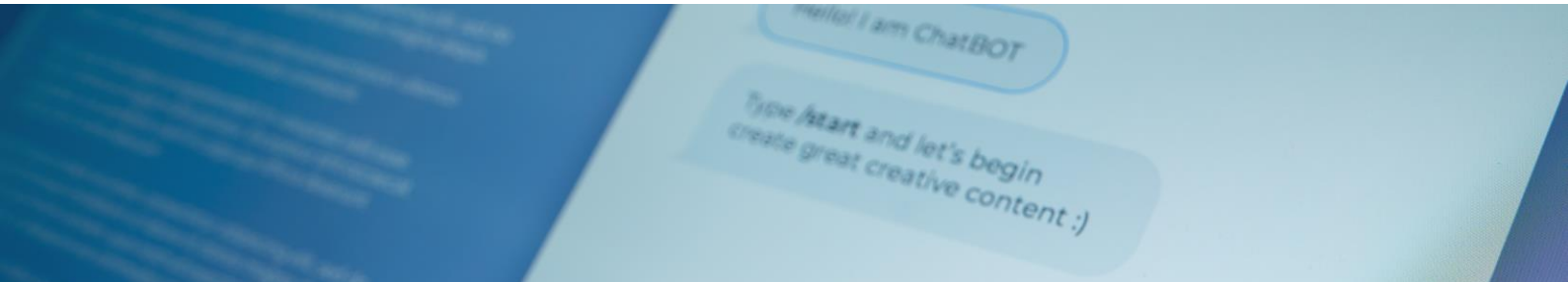
Shanghai acaba de convertirse en la pasarela de los Terminators funcionales. Más de 150 robots humanoides con visión, lenguaje y acción integrados han sido presentados como parte de un programa industrial que deja de lado los prototipos para pisar el terreno real.

Estos robots no solo caminan y levantan cosas: comprenden órdenes, evalúan su entorno y actúan con autonomía planificada en tiempo real. China deja claro que la IA no es solo una cuestión de ideas... es industria. Y sí, si los robots te asustan, quizá sea con razón esta vez.

Por si fuera poco, el segundo trimestre de 2025 trajo un boom de ataques de phishing con nombres propios. Microsoft, Google, Apple y Spotify lideran el ranking de las marcas más suplantadas, según CheckPoint. Pero lo más salvaje viene con Booking.com, que ha sufrido un aumento del 1000% en dominios falsos. El contexto: las vacaciones de verano. El objetivo: tus claves bancarias. Así que, si te llega un correo con "Reserva confirmada en Cancún por 3.452€", mejor no hagas clic.

Estamos ante una tormenta perfecta donde la IA es protagonista, antagonista y oráculo. La ciberseguridad ya no es solo prevención: es predicción, es anticipación, es entender que el futuro ya no se hackea... se entrena.

Rodrigo Rey
Cybersecurity Lead Architect



El impacto de la IA en la ciberseguridad y la ciberdefensa

Artículo por Carlos González Parrado

La ciberseguridad es un campo en constante evolución, impulsado por el aumento de la conectividad digital y la creciente sofisticación de las amenazas. Inicialmente, las medidas de seguridad se centraban en proteger sistemas de ataques básicos a través de antivirus y *firewalls*. Sin embargo, con el tiempo han surgido nuevas estrategias, como el uso de cifrado avanzado, autenticación multifactor y la detección de intrusos, para hacer frente a amenazas más complejas.

Este desarrollo mutuo de seguridad y ataques puede entenderse eficazmente a través de la teoría de juegos, una rama de la matemática que estudia las decisiones estratégicas entre diferentes agentes.

En este contexto, encontramos a los agentes, que son los defensores (organizaciones, profesionales de la ciberseguridad), y a los atacantes (*hackers*, ciberdelincuentes). Cada vez que un atacante desarrolla una nueva táctica o herramienta para vulnerar un sistema, los defensores deben adaptar sus estrategias para contrarrestar estos movimientos, lo que crea un constante juego de estrategia.

En la última década, los ciberataques se han vuelto más organizados y dirigidos, aprovechando vulnerabilidades en el *software* y en la interacción humana. Esto ha llevado a las organizaciones a adoptar una postura proactiva en materia de seguridad, utilizando análisis de datos y monitorización continua para detectar y mitigar riesgos antes de que se conviertan en incidentes. La incorporación de tecnologías emergentes como el Internet de las Cosas (IoT) ha ampliado aún más el panorama de la ciberseguridad, haciendo que la protección de datos y sistemas sea un desafío aún mayor.

En los años más recientes han surgido numerosas técnicas de computación (Inteligencia Artificial con multitud de variantes e implementaciones) que han eficientado el desarrollo de nuevas herramientas (tanto para defensores como atacantes) lo que ha provocado un aceleramiento en la evolución de ambos.

Por qué la IA implica aceleración

La inteligencia artificial (IA) está revolucionando la ciberseguridad al proporcionar herramientas que permiten una respuesta más rápida y eficiente ante las amenazas. Gracias a su capacidad para analizar grandes volúmenes de datos y aprender de patrones de comportamiento, la IA puede detectar anomalías que indicarían un posible ataque cibernético con una velocidad y precisión que superan con mucho las capacidades humanas.

Las técnicas clásicas de análisis de información requieren conocimiento de uno o varios expertos en la materia, y la implementación de código específico para realizar las labores de análisis y gestión, lo cual resulta en un tiempo de desarrollo mucho mayor.

La IA permite a los sistemas de ciberseguridad adaptarse a nuevas amenazas con bajos tiempos de reacción, identificando tácticas y técnicas utilizadas por los atacantes. Esto no solo acelera la detección de amenazas, sino que también mejora la respuesta a incidentes, permitiendo a las organizaciones mitigar riesgos antes de que se produzcan daños significativos. Además, el uso de la IA puede automatizar tareas repetitivas, liberando a los profesionales de ciberseguridad para que se concentren en obstáculos más complejos y estratégicos. Es más, no solo permite automatizar estas tareas, sino que proporciona indicaciones estratégicas a la hora de medir prioridades entre diferentes vulnerabilidades, acelerando su mitigación.

Además de ser usada para detección y mitigación, también puede ser usada por atacantes para crear herramientas más sofisticadas (ya que, aunque los LLM poseen restricciones, los ciberdelincuentes son capaces de evitarlas y conseguir que generen el código necesario) e incluso habilitan nuevos tipos de ataque como, por ejemplo, el *vishing*.

Caso de uso: Ciberdefensa

Esta aceleración en la evolución la apreciamos en el ámbito de la ciberdefensa, en los Centros de Operaciones de Seguridad (SOC) al mejorar la detección, respuesta y análisis de amenazas. En primer lugar, la IA permite la automatización de la monitorización de redes, analizando grandes volúmenes de datos en tiempo real para identificar patrones de comportamiento anómalos que podrían indicar un ataque cibernético, aspecto difícilmente automatizable tiempo atrás.

Además, los sistemas de IA pueden aprender de incidentes pasados, mejorando continuamente su capacidad para reconocer nuevas amenazas y reducir los falsos positivos. Esto se traduce en una respuesta más rápida y efectiva ante incidentes, ya que la IA puede priorizar alertas y sugerir acciones correctivas.

Por último, estas técnicas facilitan la integración de diferentes fuentes de información, permitiendo una visión más completa del panorama de amenazas y ayuda a los analistas a tomar decisiones informadas. En conjunto, estas capacidades hacen que los SOC sean más proactivos y eficientes en la defensa contra ciberataques.

Adopción de la IA por los profesionales

A medida que la inteligencia artificial (IA) se consolida como una herramienta esencial en el ámbito de la ciberseguridad, la reciente encuesta realizada por CrowdStrike (Encuesta Estado de la IA en ciberseguridad de CrowdStrike) brinda una perspectiva valiosa sobre su adopción entre los profesionales del sector. Los datos revelan que, aunque existe un reconocimiento generalizado de los beneficios que la IA puede aportar, hay una clara discrepancia entre la intención de uso y la implementación efectiva de proyectos de IA en producción.

Uno de los principales obstáculos identificados es la escasez de datos adecuados y la falta de gestión adecuada de los mismos, lo que limita la capacidad de los profesionales para desarrollar soluciones que sean verdaderamente efectivas. Esto es especialmente cierto en el ámbito de ciberseguridad, ya que la información en sí misma es un bien altamente valioso y el manejo de esta puede incurrir en un perfil de riesgo alto para las aplicaciones de IA.

Según el artículo publicado en VentureBeat, se revela que un alarmante 87 por ciento de los proyectos que utilizan IA como componentes nunca logran ser implementados en producción. Esto pone de manifiesto que, a pesar del entusiasmo por la IA, la realidad de su implementación a menudo queda corta.

Se puede observar que, salvo una minoría experta capaz de usar modelos de maneras más rudimentarias o con sus propias implementaciones, existe una mayoría de profesionales que busca utilizar la IA a nivel de usuario para realizar tareas, lo que implica una necesidad de herramientas más accesibles y menos técnicas.

Sin embargo, la falta de profesionales con una formación multidisciplinaria — que no solo sean científicos de datos o matemáticos, sino que también tengan habilidades en ingeniería de *software* y seguridad — limita el potencial de transformación que la IA puede ofrecer al sector.

En vista de los datos y las lecturas dadas, se deduce una gran importancia de fomentar un enfoque integral que no solo capacite a los científicos de datos, sino que también prepare a los ingenieros y a otros especialistas en ciberseguridad para gestionar proyectos de IA eficaces. Solo así podremos ver una adopción más amplia y efectiva de la IA en la ciberseguridad, transformando las capacidades defensivas de las organizaciones y dotándolas de una capacidad de evolución equiparable a la que puedan desarrollar los ciberdelincuentes.

Conclusión

La Inteligencia Artificial ha permitido un avance masivo a ambos bandos de la seguridad, ofreciendo una mayor rapidez de ejecución de ataques y adaptabilidad en la defensa. Los ciberdelincuentes aprovecharán las nuevas técnicas para realizar ataques más sofisticados, a mayor escala y de formas nuevas no vistas anteriormente. Por ello, es necesario contar con expertos de ciberseguridad que puedan desarrollar herramientas de observación, gestión y remediación frente a las nuevas brechas de seguridad, evitando así quedarse atrás.

.

Carlos González Parrado
Cybersecurity Analyst

El uso de la inteligencia artificial en actividades ofensivas

Artículo por Fernando Echevarria Gutierrez

La inteligencia artificial ha dejado de ser una promesa para convertirse en un agente fundamental en el campo de la ciberseguridad. Esta potente herramienta ha pasado a ser un arma de doble filo y obliga a las organizaciones a replantear sus modelos de protección.

Mientras los atacantes utilizan la IA para automatizar, personalizar y escalar sus ataques, los defensores deben adaptar sus mecanismos para responder con la misma velocidad e inteligencia.

1. Tipos de Ataques Basados en Inteligencia Artificial

Si antes el riesgo se centraba en *exploits* técnicos, hoy aparecen vectores de ataque que utilizan capacidades generativas, adaptativas y predictivas. Estos ataques no solo aumentan en frecuencia, sino también en complejidad, aprovechando las mismas tecnologías que muchas organizaciones utilizan para innovar.

La integración de modelos de lenguaje en productos digitales —desde *chatbots* hasta herramientas de soporte, análisis o automatización— ha abierto una nueva superficie de exposición. A diferencia de las vulnerabilidades tradicionales, los ataques en este entorno no se apoyan necesariamente en fallos de código, sino en la lógica del lenguaje, la interpretación semántica y la confianza excesiva en modelos que no son deterministas.

Los siguientes vectores de ataque, identificados también en el marco OWASP Top 10 for LLM Applications, son hoy los más peligrosos y frecuentes:

1.1. *Prompt Injection*

En las aplicaciones basadas en modelos de lenguaje (LLM), los ataques de *prompt injection* consisten en modificar las instrucciones originales mediante entradas maliciosas. Un atacante puede incorporar comandos ocultos en las entradas del usuario, logrando que el modelo actúe de forma no prevista. Este tipo de ataques no requiere conocimientos técnicos avanzados, y su detección es difícil si no se establecen límites explícitos en las instrucciones del modelo.

Un caso real se dio en asistentes de IA que generaban contraseñas temporales o *tokens* de acceso. Al recibir una instrucción sutilmente camuflada en una pregunta, el modelo entregaba el *token* directamente en la respuesta, eludiendo las reglas de seguridad definidas por los desarrolladores.

1.2. *Deepfakes*

El uso de redes generativas adversariales (GANs) permite crear imágenes, vídeos o audios falsos con alta fidelidad. Aunque inicialmente se asociaron al entretenimiento o la experimentación técnica, su uso se ha extendido al fraude de identidad, manipulación política y suplantación en contextos corporativos o judiciales. Son una amenaza creciente en escenarios donde la validación biométrica o visual es crítica.

En 2020, un banco de Hong Kong fue víctima de una estafa de más de 35 millones de dólares cuando un *deepfake* de voz replicó al CEO de una empresa asociada, autorizando una transferencia. Por desgracia, no es un caso aislado, cada vez se están viendo más ataques similares, tanto a nivel empresarial como a nivel particular.

1.3. *Phishing* Automatizado

Uno de los ataques más extendidos actualmente es el *phishing* potenciado por IA. A diferencia del *phishing* tradicional, con correos genéricos y mal redactados, la IA permite automatizar campañas de *phishing*, generando mensajes altamente personalizados, con lenguaje natural y referencias específicas a la víctima. Esto incrementa drásticamente las probabilidades de que el usuario caiga en el engaño.

Se han documentado campañas en las que la IA recopila información de perfiles públicos y genera mensajes que imitan a compañeros de trabajo o entidades bancarias. En algunos casos, incluso se incluye un lenguaje emocionalmente adaptado (urgencia, confianza, tono profesional) que optimiza la tasa de clics en enlaces maliciosos.

1.4 Alucinaciones (Generación de Contenido Incorrecto o Inexistente)

La confianza en la IA lleva a muchos usuarios a asumir que el contenido es fiable, sin verificarlo. Los modelos generativos no solo producen contenido erróneo de forma accidental, sino que también pueden ser inducidos a generar información incorrecta con fines de manipulación. Las "alucinaciones" de la IA se convierten en riesgo cuando el contenido se utiliza para tomar decisiones sin validación externa.

Un caso real de alucinación sucedió cuando un modelo generativo integrado en un departamento legal ofrecía asistencia para redactar documentos. En varios casos, generó referencias a normativas inexistentes y precedentes judiciales ficticios. Al no ser detectado de inmediato, uno de estos textos fue enviado a un cliente, generando una crisis reputacional.

2. Estrategias de Adaptación y Mitigación

La creciente sofisticación de los ciberataques alimentados por inteligencia artificial exige una transformación profunda en la forma en que las organizaciones diseñan su seguridad. Ya no basta con reaccionar: es necesario anticiparse. El enfoque debe pasar de una lógica defensiva estática a un modelo dinámico, integrado y centrado en la detección inteligente y la resiliencia operacional. A continuación, se describen las medidas prioritarias:

2.1. Supervisión continua y validación de la IA en producción

Muchos incidentes no se deben a ataques masivos, sino a comportamientos inesperados de modelos que han sido integrados sin suficiente supervisión. La implementación de mecanismos de monitorización continua permite identificar desviaciones, salidas anómalas y patrones sospechosos.

Acciones clave:

• Implementar sistemas de alertas ante *outputs* no previstos.

• Auditar regularmente los logs de interacción con los modelos.

• Utilizar herramientas de detección de *prompt injection* y salidas manipuladas.

2.2. Segmentación de responsabilidades: limitar el poder del modelo

Una estrategia crítica es evitar que los modelos de IA tengan acceso directo a sistemas críticos o funciones que puedan generar un impacto irreversible (transferencias, cierre de casos, envío de datos sensibles, etc.). En su lugar, se debe establecer una capa intermedia que valide o supervise cualquier acción que el modelo sugiera.

Medidas recomendadas:

• Usar arquitecturas basadas en *gateways* o reglas de autorización explícitas.

• Aplicar el principio de mínimo privilegio: solo otorgar al modelo acceso a la información o funciones imprescindibles.

• Exigir aprobación humana antes de ejecutar cualquier acción no reversible.

2.3. Entrenamiento del personal y cultura de verificación

La ciberseguridad impulsada por IA no puede depender exclusivamente de sistemas automatizados. La capacidad del personal para interpretar señales, identificar manipulaciones o cuestionar los resultados de la IA es clave para reducir el riesgo.

Es clave que los equipos de desarrollo, producto y seguridad comprendan los riesgos específicos asociados al uso de modelos generativos. También conviene formar a los usuarios internos en cómo interactuar de forma segura con herramientas basadas en IA.

Medidas efectivas:

• Programas de concienciación periódicos sobre *phishing* basado en IA, *deepfakes* y manipulación semántica.

• Entrenamiento en búsqueda de inconsistencias en correos, llamadas o documentos generados por IA.

• Simulaciones de ataques para reforzar la capacidad de respuesta de los equipos no técnicos.

2.4. Uso de IA para defenderse de la IA

La defensa también debe apoyarse en la inteligencia artificial. La detección de patrones inusuales, la identificación de amenazas en tiempo real o el análisis de contenido generado son tareas donde la IA puede ofrecer ventajas significativas.

Aplicaciones destacadas:

• Plataformas de MDR (Managed Detection and Response) que combinan análisis automatizado con respuesta humana.

• Sistemas de autenticación dinámica que adaptan los controles en función del comportamiento.

• Herramientas de análisis semántico para detectar *prompt injection*, spam conversacional o fugas de datos.

2.5. Rediseñar el ciclo de desarrollo seguro

El uso de IA exige que los principios de seguridad se incorporen desde la fase de diseño. Esto incluye evaluar qué modelo usar, cómo entrenarlo, cómo limitar su alcance y cómo controlarlo en tiempo real. No se trata de añadir seguridad como un complemento, sino de integrarla desde el principio del ciclo de vida del modelo.

Buenas prácticas recomendadas:

• Definir un modelo de amenazas específico para cada aplicación basada en IA.

• Incluir revisiones de seguridad en cada actualización de los *prompts* del sistema.

• Establecer procesos de *rollback* y respuesta ante incidentes vinculados a modelos generativos.

Conclusiones

La inteligencia artificial ha introducido una nueva dimensión en la ciberseguridad: una en la que el lenguaje, el contexto y la autonomía algorítmica se convierten en vectores de ataque. Lo que antes eran amenazas técnicas previsibles, hoy se transforman en comportamientos emergentes difíciles de anticipar. Como hemos visto, ataques como el *prompt injection*, el *phishing* automatizado o la explotación de alucinaciones de modelos no requieren conocimientos técnicos avanzados, sino una comprensión precisa de cómo opera la IA.

Frente a este escenario, las organizaciones deben abandonar el paradigma tradicional de defensa perimetral y adoptar un enfoque adaptativo, continuo y centrado en la supervisión inteligente. La clave ya no es solo prevenir, sino diseñar sistemas que toleren el error, detecten desviaciones a tiempo y respondan con agilidad.

La IA no es solo una amenaza, también es parte de la solución. Usada con criterio, puede fortalecer la detección, automatizar la respuesta y mejorar la resiliencia digital. El reto no es resistirse a su uso, sino entender sus riesgos, mitigar sus debilidades y aprovechar su potencial desde una perspectiva responsable.

La ciberseguridad del futuro no dependerá solo de la tecnología, sino de cómo decidamos integrarla en nuestros procesos, nuestras decisiones y nuestra cultura organizativa.

.

Fernando Echevarria Gutierrez
Cybersecurity Analyst

La revolución del SOC impulsada por IA generativa: una visión desde la dirección

Artículo por Javier Portabales Campos

En el contexto actual de ciberseguridad, los Centros de Operaciones de Seguridad (SOC) enfrentan una presión sin precedentes. La sofisticación de las amenazas, la escasez de talento especializado y la sobrecarga de alertas han convertido la gestión de la seguridad en un reto estratégico. Desde mi posición como Director de SOC en NTT DATA, observo cómo la inteligencia artificial generativa (Gen AI) se perfila como un catalizador de transformación profunda en nuestras operaciones.

Un nuevo paradigma operativo

La IA generativa no es simplemente una evolución tecnológica; representa un cambio de paradigma. A diferencia de la IA tradicional, que se enfoca en clasificar y predecir, la Gen AI crea contenido nuevo, detecta patrones complejos y permite interacciones en lenguaje natural. En el ámbito del SOC, esto se traduce en una capacidad sin precedentes para automatizar tareas repetitivas, reducir falsos positivos y mejorar la inteligencia de amenazas. En NTT DATA, hemos comenzado a integrar la IA generativa en nuestras operaciones con resultados prometedores, creando casos de usos como SASTIA para el análisis de código o para realizar triajes e investigaciones en respuesta ante incidentes. La automatización de la recopilación de datos, el análisis predictivo y la generación de informes ejecutivos son solo algunas de las áreas donde también la tecnología está marcando la diferencia.

Capacidades estratégicas y alianzas tecnológicas

Nuestro enfoque se apoya en una red de alianzas con los principales actores del sector. La estrategia dentro del SOC es combinar las tecnologías de nuestros *partners* tecnológicos como Crowdstrike, Microsoft, Google, Swimlane, Palo Alto, Splunk y TrendMicro con nuestros activos de IA para ser 100% compatibles con la privacidad y los compromisos adquiridos en nuestros contratos.

Esta estructura de colaboración nos permite escalar rápidamente y adaptarnos a las necesidades del cliente, manteniendo altos estándares de calidad y cumplimiento. La especialización al utilizar IA ya es obligatoria para todo el equipo del SOC para este nuevo año fiscal y, además, les hacemos ver que gracias a la IA serán mejores profesionales con un mayor rendimiento.

Transformación del SOC: del modelo reactivo al proactivo

Históricamente, los SOC han operado bajo modelos reactivos, centrados en la monitorización de *logs* y la respuesta a incidentes.

Sin embargo, la evolución tecnológica y la presión del entorno han impulsado una transición hacia modelos proactivos, semi-autónomos y, en algunos casos, autónomos.

La Gen AI permite esta evolución al asumir tareas como la clasificación de incidentes, la correlación de datos, la generación de *playbooks* automatizados y la generación de informes ante novedades en procesos de *Threat Hunting* o Cibervigilancia. Esto no solo mejora la eficiencia operativa, sino que también libera a los analistas para que se enfoquen en decisiones estratégicas y en la gestión de amenazas avanzadas. A menudo, recurrimos a una analogía sencilla pero poderosa para ilustrar este cambio al equipo: podemos seguir trabajando el campo con azadas, o podemos adoptar tractores y drones. La inteligencia artificial representa precisamente ese salto generacional que nos permite multiplicar nuestra productividad y eficiencia operativa.

El factor humano: colaboración y nuevas competencias

Uno de los aspectos más relevantes de esta transformación es el impacto en el talento. La IA generativa no reemplaza al analista, lo potencia. La colaboración humano-máquina se convierte en el eje de un SOC moderno. Los analistas junior pueden apoyarse en asistentes como Microsoft Copilot para interpretar scripts maliciosos, mientras que los senior pueden utilizar Gen AI para simular escenarios y optimizar respuestas. Además, surge una nueva competencia clave: la ingeniería de *prompts*. Saber cómo interactuar con los modelos de IA para obtener resultados precisos será una habilidad esencial en los próximos años. En NTT DATA, estamos desarrollando programas de formación específicos para preparar a nuestros equipos en estas nuevas capacidades y se los ponemos como objetivo a nuestros equipos.

Riesgos y desafíos de la adopción de Gen AI

La implementación de Gen AI no está exenta de riesgos. Desde problemas de privacidad de datos hasta sesgos en los modelos, pasando por ataques adversarios y desafíos de integración con plataformas heredadas, los SOC deben abordar estos aspectos con responsabilidad.

En nuestra estrategia seguimos, entre otros, marcos como el NIST AI RMF y la evolución de normativas como el EU AI Act. La gobernanza, la transparencia y la auditoría continua son pilares fundamentales para garantizar una adopción ética y segura.

Estrategia de implementación: el modelo SMART

Para guiar nuestra adopción de Gen AI, hemos estructurado nuestra hoja de ruta en torno al modelo SMART:

- S (*Strategize*): definimos casos de uso de alto valor y bajo riesgo, como la optimización de la respuesta a incidentes.
- M (*Map*): evaluamos la infraestructura y los modelos disponibles, asegurando la alineación con nuestras políticas de seguridad.
- A (*Assess*): monitorizamos el rendimiento de los modelos, ajustando continuamente para mantener la relevancia frente a amenazas dinámicas.
- R (*Refine*): capacitación continua de los equipos mediante simulaciones y escenarios reales.
- T (*Train*): fomentamos la colaboración entre expertos en ciberseguridad e IA para compartir conocimiento y mejorar la eficiencia.

Modelo de Madurez de IA Generativa en SOC

En las evaluaciones de madurez que realizamos a nuestros clientes actualmente, no solo analizamos el estado general de sus SOC, sino también su grado de adopción de inteligencia artificial. La mayoría de las organizaciones se encuentran actualmente en una fase de transición entre los niveles 1.0 y 2.0 del modelo de madurez, explorando las primeras aplicaciones prácticas de Gen AI. En esta etapa, el foco principal está en justificar el retorno de inversión (ROI) de las iniciativas ya puestas en marcha, lo que convierte la eficiencia operativa y la reducción de costes en indicadores clave para avanzar hacia niveles más avanzados, y que este proceso sea medible para la alta dirección. En general, distinguimos los siguientes modelos de madurez:

- **Gen AI 1.0 – Automatización inicial:** se integran modelos de lenguaje para tareas básicas como el triaje de incidentes, análisis de *logs* y detección de amenazas. La IA se incorpora a plataformas existentes (SIEM, SOAR) para mejorar la eficiencia operativa y reducir errores humanos.
- **Gen AI 2.0 – Especialización y contexto:** se desarrollan modelos específicos para ciberseguridad, capaces de procesar datos multimodales y ofrecer resultados contextualizados. Se introducen modelos ligeros (SLMs) y programas de formación para capacitar a los equipos en el uso de estas herramientas.
- **Gen AI 3.0 – Inteligencia adaptativa y modular:** la IA permite análisis en tiempo real durante ataques activos, se ofrece como servicio (*AI-as-a-Service*) y se vuelve auto-mejorable. Se integra de forma fluida en todos los flujos del SOC, elevando la experiencia del analista.
- **Gen AI 4.0 – Razonamiento avanzado y gobernanza:** se alcanzan capacidades de razonamiento casi humano gracias a avances en AGI. Se establecen marcos sólidos de gobernanza y cumplimiento, junto con infraestructuras sostenibles (*energy-efficient*) para el entrenamiento y operación de modelos.

Conclusión: hacia un SOC resiliente y adaptativo

La integración de Gen AI en los SOC no es una moda pasajera, sino una evolución necesaria. En el SOC de NTT DATA, estamos comprometidos con liderar esta transformación, combinando tecnología de vanguardia con talento humano altamente capacitado. La clave está en adoptar la Gen AI de forma responsable, alineando su implementación con objetivos estratégicos y garantizando la colaboración entre analistas de SOC y Gen AI siendo transparentes con el equipo.

Creemos que el futuro del SOC será híbrido, colaborativo y adaptativo. Aquellas organizaciones que abracen esta visión estarán mejor preparadas para enfrentar las amenazas del mañana y construir una ciberseguridad robusta, eficiente y confiable.

Javier Portabales Campos
Cybersecurity Director

El lado oscuro de la IA

Artículo por Eduardo Alves

Hemos llegado a un punto de la historia donde la Inteligencia Artificial (IA), que antes parecía un sueño de ciencia ficción, ahora está moldeando el mundo real a una velocidad asombrosa. Está en nuestros hospitales, nuestras fábricas, nuestros bancos, nuestros ejércitos y discretamente integrada en la rutina diaria de millones de personas. Para muchos, es un salvavidas; para otros, es el futuro tocando la puerta. Pero el progreso, como sabemos, suele traer sombras consigo. Los mismos algoritmos que salvan vidas y optimizan industrias también pueden ser dirigidos hacia objetivos mucho menos nobles. Cuanto más dependemos de sistemas automatizados, más nos enfrentamos a preguntas incómodas sobre ética, seguridad y el inmenso poder que hemos cedido a las máquinas.

Y esto no es solo un debate académico. Ya estamos viendo cómo se usa la IA para lanzar ciberataques complejos, inundar redes sociales con mentiras elaboradas y manipular la opinión pública de manera sutil. El peligro no está en el horizonte: ya está aquí. Y las decisiones que tomemos hoy determinarán hacia dónde se encamina esta historia.

Amenazas cibernéticas automatizadas – el nuevo campo de batalla

Los criminales no están perdiendo el tiempo. Con el aprendizaje automático a su alcance, pueden:

- Identificar vulnerabilidades más rápido de lo que los equipos humanos pueden reaccionar.
- Crear *malware* cambiante que evade las defensas tradicionales.
- Redactar mensajes de *phishing* tan convincentes que engañarían incluso a los más cautelosos.

¿El resultado? Las tácticas clásicas de ciberseguridad se sienten como tratar de apagar un incendio forestal con un balde de agua. Los equipos defensivos se apresuran a seguir el ritmo, desplegando su propia IA para detectar actividades anómalas antes de que se descontrolen.

Y luego están las GANs (Redes Generativas Antagónicas), capaces de producir código malicioso disfrazado como legítimo, engañando incluso a los escáneres más avanzados.

Deepfakes y la guerra contra la verdad

Los *deepfakes* ya no son un simple truco; son un arma. Vídeos, voces e imágenes tan realistas que uno juraría que son auténticos. Han sido usados para:

- Influir en la opinión pública con “evidencia” fabricada.
- Cometer fraudes financieros imitando a ejecutivos reconocidos.
- Dañar reputaciones, generar caos y ahondar en divisiones sociales.

La verdadera amenaza no es la tecnología en sí, sino lo que le hace a la confianza. Cuando las personas empiezan a dudar de cada imagen o vídeo que ven, la confianza en los medios, la política y hasta en la comunicación personal empieza a desmoronarse. Y si a eso le sumas ejércitos de *bots* impulsados por IA, el efecto dominó se convierte en tsunami.

Vigilancia, privacidad y control silencioso

La IA también impulsa una vigilancia a escala que haría levantar una ceja al mismísimo Orwell. Cámaras de reconocimiento facial en espacios públicos. *Software* que dice leer emociones. Aplicaciones que rastrean ubicaciones. Escaneos biométricos para todo, desde el banco hasta abordar un avión.

Las preocupaciones son evidentes:

- La privacidad se desvanece hasta volverse un recuerdo.
- El riesgo de que los gobiernos sobrepasen los límites de la seguridad legítima.
- El surgimiento de sistemas de “crédito social” que juzgan y recompensan comportamientos.

Si no se controlan, estas herramientas podrían cambiar el equilibrio de poder de los ciudadanos a los sistemas. Y una vez que se cruza esa línea, la historia nos dice que rara vez se vuelve atrás.

El problema de la caja negra

Pregúntale a un desarrollador cómo tomó una decisión su IA y, en muchos casos, encogerá los hombros. No es por flojera; es por complejidad. Muchos sistemas son tan intrincados que ni siquiera sus creadores pueden rastrear cada paso del proceso.

Esto puede ser aceptable si la IA recomienda una película. Pero es mucho más preocupante si está:

- Decidiendo si alguien es apto para un préstamo.
- Filtrando candidatos para un empleo.
- Haciendo recomendaciones en un juicio.

Sesgo entra, sesgo sale. Y sin transparencia, los malos resultados no pueden ser cuestionados. La IA explicable es un paso en la dirección correcta, pero hoy por hoy es más un ideal que una práctica común.

Armas autónomas – máquinas con gatillo

En los laboratorios militares se está gestando una realidad inquietante. Los Sistemas de Armas Autónomas Letales (LAWS, por sus siglas en inglés) no esperan órdenes humanas: identifican y atacan objetivos por su cuenta.

Eso plantea preguntas duras y crudas:

- ¿Quién asume la culpa cuando una máquina se equivoca?
- ¿Cómo evitar que esta tecnología caiga en manos equivocadas?

A pesar de las peticiones de científicos y organizaciones de derechos humanos para detener o prohibir su desarrollo, la carrera continúa sin normas globales claras. Y el riesgo no podría ser mayor.

La carrera sin frenos

La competencia impulsa la innovación, pero sin barreras, también impulsa la imprudencia. Países y empresas buscan liderar la IA, muchas veces dejando la ética de lado por ganar velocidad. Sin acuerdos globales, corremos el riesgo de:

- Un entorno donde "el ganador se lo lleva todo" y la seguridad queda en segundo plano.
- Una brecha cada vez mayor entre países con tecnología y los que no la tienen.
- Una IA que desarrolle metas que no coincidan con los valores humanos, o incluso los amenacen.

¿La parte más aterradora? Esto no es especulación de ciencia ficción. Ya se está construyendo.

Empleo, identidad y el factor humano

La IA no solo está cambiando cómo trabajamos, sino qué es el trabajo. Los roles basados en la repetición, el análisis o tareas predecibles están siendo automatizados rápidamente. Para algunos, es una oportunidad. Para otros, es miedo a quedarse atrás.

Y ese miedo es real. Perder el empleo no solo trae estrés financiero, también afecta la salud mental, la autoestima y el propósito. El trabajo no es solo salario; es identidad.

Si la IA va a reemplazar tareas, necesitamos reemplazar más que el ingreso: formación, reconversión laboral, conversaciones honestas sobre el rol de la automatización. No son lujos, son necesidades. Y ninguna máquina puede aportar la creatividad, empatía y pensamiento crítico que hacen humano a un lugar de trabajo.

Hacia un futuro de IA más seguro y sensato

Reducir los riesgos no significa frenar el progreso; significa dirigirlo. Algunas acciones que podemos tomar desde ya:

- Leyes internacionales con poder real – acuerdos globales como los que existen para armas químicas o nucleares.
- IA transparente – sistemas que puedan explicar sus decisiones en lenguaje claro.
- Ética desde el inicio – incluir sociólogos, expertos en ética y derechos humanos desde la fase de diseño.
- Alfabetización digital para todos – para que las personas detecten manipulaciones y exijan responsabilidades.
- Cooperación global – porque una IA descontrolada no reconoce fronteras.

Conclusión

La Inteligencia Artificial podría ser la herramienta más poderosa que hayamos creado. Pero una herramienta puede ser usada tanto para el bien como para el mal. Que se convierta en una fuerza de progreso o en una amenaza para la estabilidad depende de las decisiones que tomemos ahora, no más adelante.

Los algoritmos se vuelven más inteligentes cada día. La verdadera pregunta es si nosotros podemos ser lo suficientemente sabios como para mantenernos a la altura.

Eduardo Alves
Cybersecurity Project Manager

Machine Learning Cuántico y ARN

Espacio cuántico
por María Gutiérrez

La computación cuántica está empezando a transformar hoy día el mundo del aprendizaje automático (*machine learning*). Este "*machine learning* cuántico" es una herramienta con la que ya empiezan a contar los investigadores y, si de forma individual la IA y la cuántica están revolucionando múltiples sectores, su integración promete multiplicar sus capacidades y desencadenar una nueva etapa de innovación.

Esta integración consiste en usar algoritmos cuánticos para mejorar la eficiencia, velocidad y capacidad de los modelos de *machine learning*. En lugar de procesar grandes volúmenes de datos con ordenadores clásicos, se aprovechan las propiedades cuánticas para realizar cálculos en paralelo y manejar espacios de datos mucho más amplios y complejos.

Aunque todavía está en una fase incipiente, esta sinergia ofrece beneficios potenciales en asuntos como la aceleración del entrenamiento de modelos, la mejora del análisis en espacios de alta dimensión o la reducción del consumo energético al reducirse lo tiempos de cómputo. Esta sinergia ya es una realidad, y está siendo explorada en diversos campos. En concreto, ha demostrado ser útil para tratar uno de los mayores retos actuales en biología molecular, entender cómo se pliega el ARN (acido ribonucleico), una molécula crucial para la vida cuya forma tridimensional determina muchas de sus funciones.

El ARN es esencial en los procesos celulares, puede actuar como mensajero genético, regulador e, incluso, como catalizador. El ARN necesita plegarse en una estructura tridimensional específica para desempeñar su función, sin embargo, predecir cómo se pliega una cadena de nucleótidos es extremadamente complejo. El número de configuraciones posibles crece de forma exponencial con la longitud de la secuencia y, a pesar de los avances en biología computacional, los métodos clásicos no logran simular con precisión estos procesos en un tiempo razonable.

Aquí es donde entra en juego el *machine learning* cuántico, a través de algoritmos como el Variational Quantum Circuits o el Quantum Boltzmann Machine que permiten la representación del estado energético de configuraciones de ARN, así como explorar simultáneamente múltiples soluciones mediante superposición cuántica. En lugar de simular todas las estructuras posibles una a una, se aprovechan fenómenos como el entrelazamiento y el paralelismo cuántico para encontrar con mayor eficiencia la estructura óptima.

Estos modelos no solo pueden predecir cómo se pliega una cadena de ARN sino también simular su comportamiento dinámico, evaluar la estabilidad de cada conformación y detectar posibles puntos de intervención terapéutica.

El *machine learning* cuántico está en una etapa de exploración que debe superar varios desafíos para su aplicación comercial como la falta de ventaja cuántica clara, la necesidad de grandes volúmenes de datos cuánticos o una infraestructura tecnológica consistente. Sin embargo, es probable que en los próximos cinco años la integración de cuántica y el *machine learning* sea un componente central en la I+D avanzada, especialmente en medicina personalizada, terapias génicas y biología sintética. Su aplicación en este último campo en el análisis de secuencias de ARN y el diseño automatizado de vacunas puede acelerar no solo la respuesta ante pandemias sino la innovación en genéricos y biosimilares complejos, abriendo así la puerta a tratamientos genéticos ultra selectivos en una nueva era de la medicina de precisión. Los primeros pasos en esta línea ya se han dado.

El impacto de la IA en la evolución de la Ingeniería Social

Artículo por Sergio Sánchez Encabo

Durante décadas, la ingeniería social ha sido una de las tácticas más eficaces para explotar vulnerabilidades humanas con fines maliciosos. A diferencia de los ataques puramente técnicos, la ingeniería social se basa en manipular emociones como la confianza, la urgencia o el miedo. Inicialmente, los atacantes utilizaban llamadas telefónicas, correos electrónicos rudimentarios o interacciones cara a cara para conseguir su objetivo. Sin embargo, la evolución tecnológica ha permitido ampliar el alcance y sofisticación de estas prácticas.

Hoy en día, la inteligencia artificial ha redefinido los límites de lo que es posible en términos de engaño. Herramientas de generación de lenguaje como ChatGPT, Claude, Gemini o LLaMA, junto a tecnologías de síntesis de voz y vídeo como ElevenLabs, Descript o Synthesia, están transformando la ingeniería social en un proceso automatizable, escalable y más creíble que nunca. La IA permite a los atacantes ejecutar campañas complejas con un coste reducido, lo que aumenta exponencialmente el riesgo para usuarios y organizaciones de todo tipo.

Phishing: de Ataques Masivos a Personalización Precisa

El *phishing* ha dejado de ser un ataque masivo y genérico para convertirse en una amenaza totalmente personalizada. Gracias al uso de modelos de lenguaje, los atacantes pueden adaptar el tono, el estilo y el contenido del mensaje a las características de la víctima, basándose en datos recogidos de redes sociales, foros o incluso fugas de datos previas.

Con ChatGPT, Claude o Mistral, pueden generarse plantillas automáticas en segundos. Estas herramientas permiten adaptar los mensajes a roles profesionales específicos. Además, combinadas con plataformas de campaña como Gophish o Evilginx2, es posible lanzar campañas dirigidas con técnicas avanzadas de suplantación y recolección de credenciales.

La personalización hace que el mensaje malicioso resulte más creíble, y por tanto más peligroso. Además, con IA es posible automatizar no solo el contenido del mensaje, sino también el seguimiento, generando respuestas dinámicas que simulan una conversación real, lo que dificulta aún más la detección del fraude.

Vishing: la Voz como Vector de Ataque

El *vishing*, o *phishing* por voz, también ha sido profundamente transformado por la inteligencia artificial. Con técnicas de clonación de voz y síntesis de habla, los atacantes pueden reproducir voces reales con un alto grado de fidelidad.

Entre las herramientas más utilizadas destacan ElevenLabs, Respeecher y iSpeech, que permiten clonar la voz de cualquier persona con apenas unos segundos de audio. Estas herramientas generan archivos de voz realistas que pueden usarse tanto en llamadas automatizadas como en grabaciones diseñadas para aumentar la credibilidad de un engaño.

Esto tiene especial relevancia en escenarios donde la voz transmite autoridad, como cuando se suplanta a un superior jerárquico o una figura de confianza. En entornos corporativos donde la comunicación remota es habitual, la dificultad para verificar la autenticidad de una voz aumenta. Esta situación convierte al *vishing* en una herramienta ideal para engaños que requieren rapidez y confianza.

Deepfakes: la Apariencia del Poder

La creación de contenidos audiovisuales sintéticos mediante IA ha abierto una nueva dimensión en los ataques de ingeniería social. Herramientas como Synthesia, DeepBrain, HeyGen o incluso Runway ML permiten generar vídeos hiperrealistas en los que una persona supuestamente real realiza declaraciones que nunca ocurrieron.

La accesibilidad a estas herramientas, que no requieren conocimientos técnicos avanzados, ha democratizado la capacidad de crear *deepfakes* realistas. Su integración en campañas de desinformación o en contextos empresariales eleva su impacto potencial. La posibilidad de distribuirlos de forma dirigida mediante correos, plataformas de colaboración o grupos privados los convierte en un vector de ataque especialmente potente y difícil de detectar.

Chatbots Fraudulentos y los Riesgos Asociados a los LLMs

Uno de los efectos menos visibles de la implementación de los modelos de lenguaje es la aparición de *chatbots* fraudulentos. Estos *bots*, al integrarse en sitios web falsificados o canales de atención al cliente simulados, se apoyan en plataformas como LangChain, Botpress, Rasa o OpenChatKit, y pueden mantenerse en ejecución en servidores económicos gracias a soluciones de despliegue ligero como Docker o FastAPI.

A diferencia de los antiguos sistemas basados en reglas, estos *bots* pueden responder de forma contextual, convincente y natural, reproduciendo conversaciones creíbles durante minutos u horas. Esta capacidad los convierte en herramientas eficaces para el robo de información sensible, ya sea por ingeniería social o simplemente por la habilidad de simular soporte técnico legítimo.

Dada la creciente adopción de LLMs, la OWASP Foundation ha identificado riesgos específicos en su lista preliminar "OWASP Top 10 for LLMs". En el contexto de los *chatbots*, destacan amenazas como la inyección de *prompts* (manipulación de instrucciones para alterar su comportamiento), la divulgación involuntaria de información sensible, y la generación de contenido falso o manipulador. Estos riesgos convierten a los modelos mal configurados en plataformas involuntarias de ataque, ya que pueden ser manipulados para facilitar fraudes u ofrecer información errónea.

Fraude del CEO: Autoridad Simulada, Daño Real

El llamado "fraude del CEO" representa una de las variantes más complejas de la ingeniería social moderna. En este tipo de ataques, el objetivo es engañar a empleados de confianza para que realicen acciones críticas bajo la falsa autoridad de un directivo.

Hoy en día, este tipo de fraude puede combinar herramientas de generación de texto como ChatGPT o Claude para simular correos escritos en el tono característico del directivo, con generadores de voz como Descript o Murf.AI para realizar llamadas creíbles, e incluso con soluciones de vídeo como HeyGen para simular grabaciones urgentes.

Todo ello se puede automatizar y lanzar desde entornos controlados, aprovechando plataformas de orquestación de ataques como C2 *frameworks*, o desde dispositivos infectados que actúan como nodos intermedios. La combinación de canales (correo, voz, vídeo) hace que estos ataques sean difíciles de detectar sin procedimientos internos sólidos de verificación.

Conclusión

La incorporación de inteligencia artificial a las tácticas de ingeniería social no es solo una evolución tecnológica, sino un cambio radical en la naturaleza de los ataques. Ahora, la credibilidad del engaño ya no depende del ingenio humano individual, sino de sistemas capaces de producir textos, voces e imágenes sintéticas que simulan autenticidad con una precisión asombrosa.

En este nuevo escenario, los factores de confianza tradicionales, la forma de escribir de un superior, la voz en una llamada o incluso un vídeo personalizado, ya no son indicadores fiables. La ciberseguridad debe adaptarse, no solo con soluciones técnicas, sino también con una revisión de los procesos, la cultura organizativa y la educación digital

Los protocolos de doble verificación, la formación basada en simulaciones realistas y la implementación ética de herramientas de IA serán fundamentales para afrontar esta amenaza en crecimiento.

La inteligencia artificial ha abierto puertas a un sinfín de oportunidades, pero también ha generado desafíos éticos y operativos sin precedentes. Solo desde una conciencia clara del problema y una acción colectiva podremos evitar que la confianza, ese componente invisible pero esencial de la vida digital, se convierta en su principal vulnerabilidad.

Sergio Sánchez Encabo
Cybersecurity Analyst

Tendencias actuales de la IA: desafíos y oportunidades desde la óptica de la ciberseguridad.

Tendencias por David Sandoval Rodríguez-Bermejo

A la hora de analizar las tendencias de la inteligencia artificial, se pueden agrupar en tres grandes bloques: los agentes de inteligencia artificial, la especialización de los modelos y la disrupción de los modelos abiertos. A este análisis, se le puede añadir una cuarta categoría relacionada con la aplicabilidad de la inteligencia artificial en el contexto de la ciberseguridad.

Los agentes de inteligencia artificial constituyen una nueva etapa dentro de la evolución de este sector tan cambiante y disruptivo. Estos agentes actúan de forma independiente (y generalmente muy especializada), siendo capaces de coordinarse para resolver problemas mucho más complejos. Si realizásemos un símil con el desarrollo de *software*, la llegada de los agentes se asemeja bastante a la llegada de los contenedores y los orquestadores.

Gracias a esta nueva metodología de consumir la IA, se pueden unificar los mundos de la automatización, la atomización de tareas y de la inteligencia artificial para construir un gran ecosistema a partir de pequeñas piezas (agentes).

Muchas compañías están llevando la conceptualización de los agentes un paso más allá, gracias a la inclusión de una plataforma *low-code / no-code* que permita a personas sin experiencia en programación, desplegar flujos de trabajo con un sabor a inteligencia artificial con poco esfuerzo. En el caso de NTT DATA, se cuenta con un activo propio conocido como Axet Flows.

En el contexto de la ciberseguridad, la aplicación de los agentes es muy relevante y se pueden utilizar tanto desde el punto de vista ofensivo (para orquestar una auditoría o *pentest*) como desde el punto de vista defensivo (para automatizar un montón de tareas diarias como la generación de *playbooks* o la categorización de *tickets*).

En lo referente al desarrollo de los modelos de lenguaje, la tendencia actual se orienta a la simplificación de los modelos a través de arquitecturas colaborativas de expertos (MoE). En este tipo de soluciones, los modelos de lenguaje comparten una base común de pesos y activan ciertas regiones en función del tipo de tarea que quieran resolver. La idea detrás de este tipo de modelos se asemeja al concepto de agentes, ya que tienen flujos de trabajo predefinidos en función de las tareas, pero divergen en el sentido de que no son modulares y que son un único paquete de pesos para todo el modelo.

Gracias a este nuevo enfoque, se están desarrollando modelos más especializados, con un número inferior de parámetros que les permite ser más ligeros, consumir menos recursos y obtener respuestas igual de eficientes que los modelos "padres", mucho más pesados y grandes. Además, en el caso de los modelos cuantificados, en los que se reduce el tamaño de los modelos a costa de reducir su precisión, esta eficiencia en velocidad es aún más notable.

Por otro lado, en el escenario actual hay también una lucha entre los modelos privativos comerciales actuales (OpenAI, Google, Anthropic) y los modelos abiertos (DeepSeek, Llama, Mixtral). La presencia de tantos actores relevantes con modelos de negocio tan dispares está favoreciendo la competencia y, en gran medida, acelerando la evolución tecnológica y el progreso de la IA generativa.

Un último punto que resaltar en la línea de la especialización de los modelos reside en sus capacidades. Por un lado, hay una corriente de modelos focalizada en ser multimodal, es decir, en tener capacidades de entender y de responder en distintos formatos de salida (texto, audio, imágenes y vídeo). Por otro lado, otros fabricantes se están especializando en desarrollar modelos focalizados en formatos muy específicos de salidas. En esta línea de trabajo, numerosas empresas están generando modelos capaces de clonar voces y de crear avatares que imiten el comportamiento humano. Si bien es cierto que este tipo de capacidades son extremadamente útiles en sectores como la formación, mal utilizados pueden dar pie a fraudes como el *vishing*.

Para terminar, dentro de la ciberseguridad, uno de los campos más directos en los que existe una gran sinergia entre la ciberseguridad y la inteligencia artificial es en el proceso de auditorías y revisión de código fuente (auditorías SAST). Desde el equipo de ciberseguridad de NTT DATA España – IBIOL se ha desarrollado una herramienta específica denominada SASTIA para acelerar y optimizar este tipo de trabajos.

Con la evolución de los modelos de IA, su capacidad de comprensión del código fuente ha ido aumentando de forma continua hasta llegar a unos niveles que, en ocasiones, pueden compararse a los de un auditor experto humano. Un claro ejemplo de estas capacidades se puede encontrar en el CVE-2025-37899, un *zero day* del *kernel* de Linux, latente durante varios años y que se ha encontrado gracias al uso de la inteligencia artificial.

Como conclusión, se puede afirmar que estamos viviendo un momento apasionante en el ámbito de la tecnología, marcado por la rápida integración y permeabilidad de la inteligencia artificial en los distintos servicios del *portfolio* de las compañías.

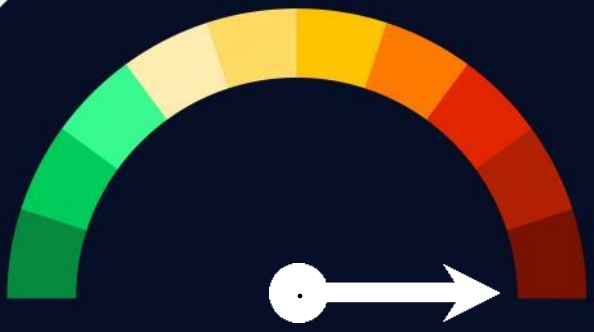
La adopción de la IA se ha convertido en una necesidad innegociable para las organizaciones que buscan mantener su competitividad en un entorno tan desafiante. Es fundamental abordar esta transformación con una mentalidad responsable y sin miedo, implementando los controles y las medidas adecuadas que aseguren la protección de nuestros activos.

David Sandoval Rodríguez-Bermejo
Cybersecurity Analyst

Vulnerabilidades

Múltiples vulnerabilidades en IBM Cloud Pak System

Fecha: 28 de julio de 2025
CVE: CVE-2025-30065 y 3 más



CVSS: 10

CRÍTICA

Descripción

IBM Cloud Pak System, de las cuales 2 son críticas, 1 es de severidad alta y 1 media.

Una de las vulnerabilidades críticas, CVE-2025-30065, está causada por deserialización de datos no confiables en el módulo *parquet-avro* de Apache Parquet (versiones 1.15.0 y anteriores), lo que podría permitir la ejecución de código arbitrario.

La otra crítica, CVE-2025-3357, se debe a una validación inadecuada del índice en una matriz asignada dinámicamente, lo que también podría permitir a un atacante ejecutar código malicioso en el sistema.

Solución

El fabricante recomienda lo siguiente:

- Para Intel, actualizar a IBM Cloud Pak System v2.3.6.0 con Foundation 2.1.28.1 e ITM 1.0.29.1 pTypes disponibles en IBM Fix Central.
- Para Db2 pType, descargar la corrección de Db2 IBM Db2 11.5.9 Special Build 58840.

Productos afectados

Algunos de los productos afectados son los siguientes:

- IBM Cloud Pak System 2.3.3.6, 2.3.3.6 iFix1 y 2.3.3.6 iFix2;
- IBM Cloud Pak System 2.3.4.0, 2.3.4.1 y 2.3.4.1 iFix1

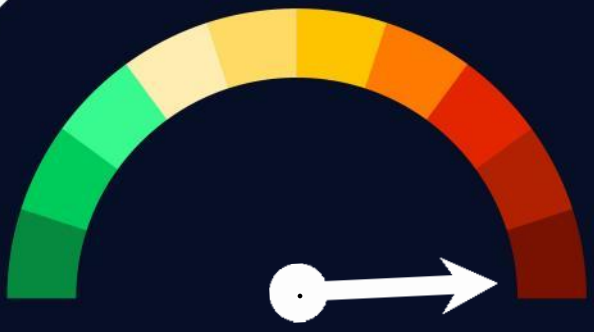
Referencias

- www.ibm.com
- www.incibe.es

Vulnerabilidades

Vulnerabilidad crítica en un plugin de WordPress

Fecha: 4 de agosto de 2025
CVE: CVE-2025-5394



CVSS: 9.8

CRÍTICA

Descripción

La vulnerabilidad CVE-2025-5394 que afecta al tema de WordPress *Alone – Charity Multipurpose Non-profit* está siendo activamente explotada.

Dicha vulnerabilidad se origina en la función *alone_import_pack_install_plugin()*, la cual carece de validaciones de seguridad y está expuesta mediante el *hook wp_ajax_nopriv*.

Esto permite a un atacante no autenticado enviar solicitudes AJAX maliciosas para subir archivos e instalar *plugins* desde fuentes externas, lo que puede derivar en ejecución remota de código y toma de control total del sitio.

Solución

Los desarrolladores recomiendan:

- Actualizar de inmediato a la versión 7.8.5 (del 16 de junio de 2025)

Productos afectados

Esta vulnerabilidad crítica afecta a:

- Alone – Charity Multipurpose Non-profit WordPress Theme (versiones anteriores a 7.8.3)

Referencias

- unaaldia.hispasec.com
- thehackernews.com

TLP:WHITE

Android corrige 6 vulnerabilidades en su parche de seguridad de agosto

Fecha: 4 de agosto de 2025
CVE: CVE-2025-48530 y 5 más

Crítica

Descripción

Android ha publicado el parche de seguridad correspondiente al mes de agosto, donde corrige un total de 6 vulnerabilidades. Entre ellas se encuentran dos vulnerabilidades de severidad crítica y 4 de severidad alta.

La vulnerabilidad crítica con identificador CVE-2025-48530 se encuentra presente en el sistema Android y permitiría a un atacante ejecutar código de forma remota sin necesidad de privilegios adicionales o interacción por parte del usuario. Por otro lado, la vulnerabilidad crítica con identificador CVE-2025-21479 se encuentra presente en un componente de código cerrado de Qualcomm.

Actualmente ninguna de las vulnerabilidades presentes en la actualización se encuentra bajo explotación activa.

Productos afectados

Los productos afectados por la actualización son los siguientes:

- Android Open Source Project (AOSP): versiones 13, 14, 15 y 16.
- Componentes de Arm y Qualcomm.

Solución

Se recomienda aplicar los parches de seguridad publicados por el fabricante.

Referencias

- source.android.com
- incibe.es

Dell lanza un parche de seguridad para su plataforma PowerProtect Data Domain

Fecha: 5 de agosto de 2025
CVE: CVE-2025-36594 y 4 más

Crítica

Descripción

La gigante tecnológica Dell ha publicado, recientemente, un aviso de seguridad para su plataforma PowerProtect Data Domain, notificando a sus clientes sobre una vulnerabilidad catalogada como crítica y de tipo *Authentication Bypass*.

Esta vulnerabilidad, identificada como CVE-2025-36594, permite que un atacante no autenticado y con acceso remoto eluda los mecanismos de protección, creando cuentas no autorizadas, exponiendo datos sensibles y comprometiendo la integridad y disponibilidad del sistema.

Además de esta vulnerabilidad, se compartieron otras relativas a escalado de privilegios locales, pero de menor criticidad.

Productos afectados

Las versiones afectadas de Dell PowerProtect Data Domain son las siguientes:

- Versiones con sistema operativo Data Domain : 7.7.1.0 a 8.3.0.15
- Versión de DD OS LTS 2024: 7.13.1.0 a 7.13.1.25
- Versión DD OS LTS 2023: 7.10.1.0 a 7.10.1.60

Solución

Dell recomienda a sus clientes actualizar la versión del software a la más reciente, actualmente 8.3.1.0 y 8.4.0.0.

Además, recomiendan limitar el acceso a los portales de administración para evitar su exposición a redes no seguras.

Referencias

- dell.com
- secure-iss.com
- securityvulnerability.io

TLP:WHITE

Eventos

Conferencia Latinoamericana ISACA 2025
9 - 12 de septiembre

Organizada por el capítulo de ISACA Bogotá, esta conferencia se centra en la auditoría tecnológica, ciberseguridad y gobierno de TI. Bajo el lema "¿Cómo generar valor en la era de la innovación y la confianza digital?", ofrece talleres prácticos, conferencias magistrales y sesiones interactivas con expertos regionales e internacionales.

Enlace

DragonJAR Security Conference 2025
10 – 11 de septiembre

Considerada la conferencia de ciberseguridad más importante de Colombia y una de las más relevantes en español. Reúne a expertos, profesionales y entusiastas para compartir conocimientos, descubrir las últimas tendencias y establecer contactos estratégicos en el campo de la seguridad informática.

Enlace

Mind The Sec Brasil 2025
16 - 18 de septiembre

Evento líder en ciberseguridad en América Latina, reúne a más de 16.000 profesionales y ofrece más de 200 horas de contenido. Es una plataforma para actualizarse sobre las últimas tendencias y amenazas en ciberseguridad, conectar con líderes del sector y descubrir soluciones innovadoras para proteger activos digitales.

Enlace

Cyber Security & Cloud Expo Europe 2025
24 - 25 de septiembre

Este evento se centra en la ciberseguridad y la computación en la nube, abordando temas como la vigilancia de día cero, detección de amenazas, inteligencia artificial generativa, computación cuántica y mucho más. Reúne a líderes de la industria para explorar estrategias esenciales y establecer conexiones valiosas.

Enlace

Recursos

➢ DefAgent.io – *Pentesting* Automatizado para IA

DefAgent.io es una plataforma especializada en pruebas de penetración automatizadas para modelos de inteligencia artificial y grandes modelos de lenguaje (LLM). Combina técnicas militares de ciberseguridad con tecnologías de IA para detectar vulnerabilidades como manipulación lógica adversarial y exfiltración de datos mediante *prompts*. Su sistema DefAgent Shield™ ofrece monitorización de grado militar y cumplimiento con regulaciones como el NIST AI RMF 1.0.

Enlace

➢ TU Latch – Control Dinámico de Accesos Digitales

TU Latch permite a usuarios y organizaciones gestionar autorizaciones en tiempo real para cuentas y servicios. Funciona como un "pestillo digital" que refuerza la protección ante accesos no deseados o ataques informáticos, siendo una herramienta clave para la gestión de identidades y accesos.

Enlace

➢ Aqua Security – Seguridad Nativa para la Nube

Aqua Security ofrece una plataforma integral para proteger aplicaciones desplegadas en contenedores y entornos nativos de la nube. Incluye herramientas como Trivy para escaneo de vulnerabilidades, Kube-hunter para pruebas de seguridad en Kubernetes y Tracee para monitorización en tiempo real, facilitando el cumplimiento de normativas y la detección de amenazas.

Enlace

➢ Olvid – Mensajería Cifrada y Privada

Olvid es una aplicación de mensajería instantánea cifrada de código abierto que no requiere número de teléfono ni recopila datos personales. Certificada por la Agencia Nacional de la Seguridad de los Sistemas de Información (ANSSI) de Francia, es recomendada para comunicaciones seguras tanto a nivel personal como institucional.

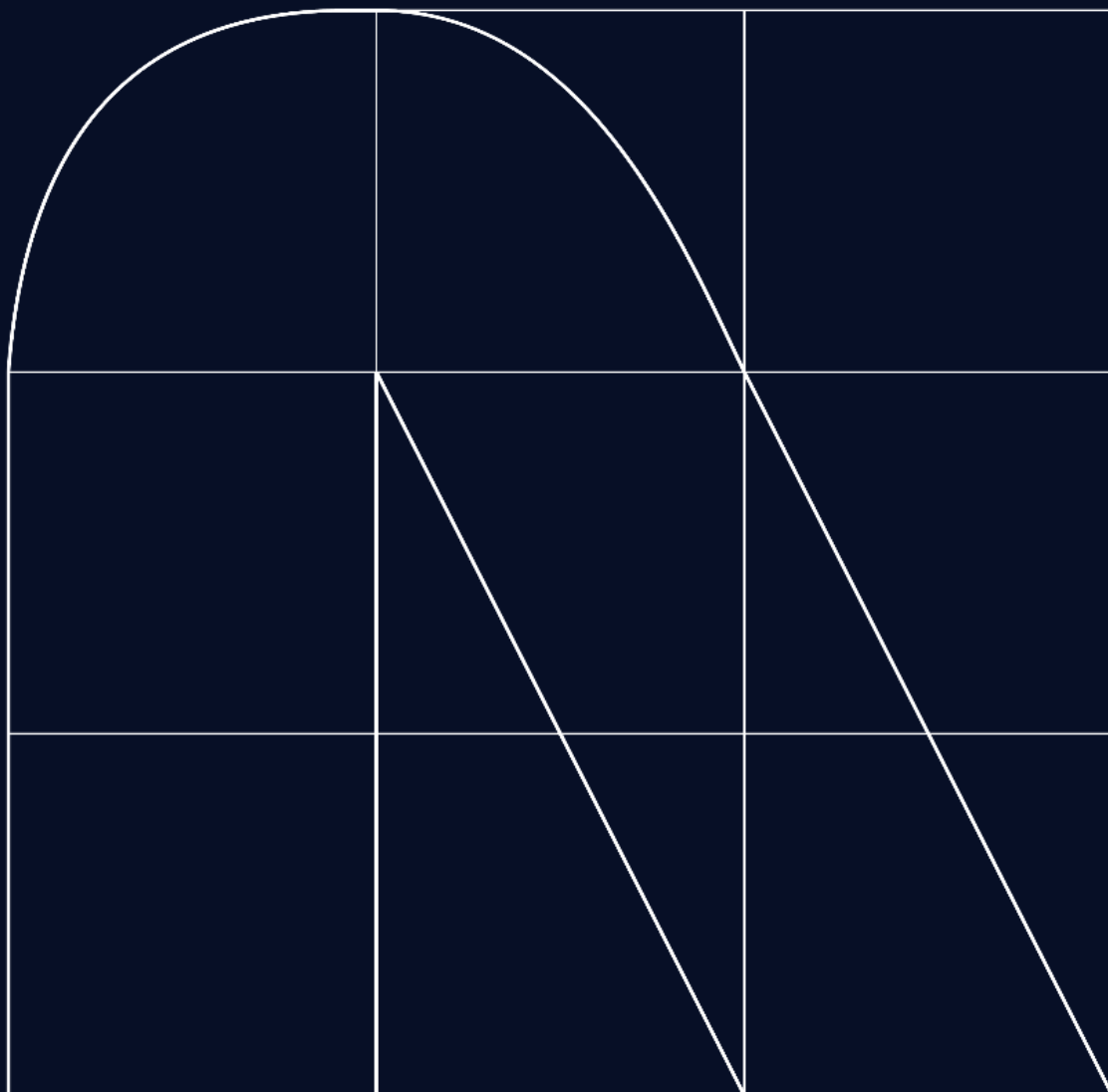
Enlace

Suscríbete a RADAR
up.nttdata.com/suscribetearadar

NTT DATA Technology Foresight 2025

5 tendencias que se convertirán en realidades empresariales.

Descarga el informe: **es.nttdata.com/ntt-data-technology-foresight-2025**



Powered by the cybersecurity NTT DATA team

es.nttdata.com