CAPÍTULO V

Herramientas de ciberdefensa: los sistemas de inteligencia artificial aplicados a la ciberseguridad

Icía Masid Urbina
Ingeniera del Área de Ciberdefensa (ISDEFE).
Desarrolla su labor en el Mando Conjunto del Ciberespacio (MCCE)

SUMARIO. 1. Introducción. 2. Descripción de los instrumentos y técnicas de ciberataque más empleados. 2.1. Principales ciberataques. *2.1.1. Malware. 2.1.2. Ransomware. 2.1.3. Phishing. 2.1.4. Ingeniería social. 2.1.5. Explotación de vulnerabilidades. 2.1.6. Denegación de servicio. 2.1.7. Acceso no autorizado a la información. 2.1.8. Suplantación. 2.1.9. Hacktivismo. 2.1.10. Amenaza interna.* 2.2. Impacto de los ciberataques en las entidades locales. **3. Revisión de los mecanismos técnicos adecuados para hacer frente de forma proactiva a ciberataques.** 3.1. Detección de patrones. 3.2. Clasificación. 3.3. Automatización. 3.4. Procesamiento del lenguaje natural. 3.5. Caso de estudio. 3.6. Beneficios de la implementación de sistemas de IA guardianes en ciberseguridad para las entidades locales. **4. Directrices para la implementación de medidas de seguridad en el uso de medios electrónicos que empleen IA. 5. Conclusiones. 6. Bibliografía.**

1. Introducción

La inteligencia artificial (IA) está transformando rápidamente el panorama de la ciberseguridad, marcando una revolución que afecta tanto a los atacantes como a los defensores. En este entorno dinámico, las entida-

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

des locales se encuentran en una posición única. Estas Administraciones manejan una vasta cantidad de datos históricos y sensibles diariamente en casi todos los ámbitos. Todos estos datos son imposibles de gestionar por los humanos. Los algoritmos permiten estudiarlos, extraer patrones y exprimir todo su potencial. Su adecuada explotación ofrece innumerables ventajas para ellas y para los ciudadanos a través de su aplicación en la gestión de los servicios públicos, en la toma de decisiones y en la ciberseguridad.

Desde una perspectiva de ciberseguridad, la IA está jugando un papel crucial en dos direcciones. Por un lado, está siendo utilizada por los cibercriminales para lanzar ataques más sofisticados, y por otro, los profesionales de la seguridad la utilizan para desarrollar defensas más robustas y proactivas. Es decir, la IA juega un papel importante en ambos bandos, tanto del lado del atacante como del lado del defensor.

En el primer caso, los atacantes están empleando técnicas de IA para automatizar y mejorar sus métodos de infiltración. Los algoritmos de aprendizaje automático les permiten identificar vulnerabilidades en los sistemas de manera más eficiente, lanzar ataques de *phishing* altamente personalizados, y evadir las detecciones tradicionales. Este uso malicioso de la IA aumenta la frecuencia y la sofisticación de los ciberataques dirigidos a las entidades locales, poniendo en riesgo la integridad de los datos y la continuidad de los servicios públicos.

Por su parte, los defensores están aprovechando la IA para fortalecer sus estrategias de ciberseguridad. Los sistemas de IA guardianes, por ejemplo, permiten a los ayuntamientos monitorizar continuamente sus redes, detectar comportamientos anómalos en tiempo real, y responder de manera automatizada a las amenazas. Estas soluciones avanzadas pueden identificar patrones de ataque previamente desconocidos, predecir posibles vulnerabilidades, y proporcionar informes detallados que facilitan la toma de decisiones informadas.

En este escenario de constante evolución, es imperativo que las entidades locales comprendan y adopten las tecnologías de IA tanto para protegerse como para anticiparse a los ciberataques. Al hacerlo, no solo salvaguardan la información y los servicios esenciales de sus ciudadanos, sino que también refuerzan la confianza pública en su capacidad para gestionar la seguridad en un mundo digital cada vez más complejo.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

2. Descripción de los instrumentos y técnicas de ciberataque más empleados

Nuestras Administraciones reciben ataques informáticos de diverso tipo y gravedad cada día. Un factor determinante que contribuye al auge de los ciberataques contra las Administraciones públicas es el desarrollo cada vez mayor de las herramientas de ataque gracias a la IA. Y es que la IA ha revolucionado la forma en que se ejecutan los ataques cibernéticos, ya que ofrece a los ciberdelincuentes herramientas más sofisticadas y difíciles de detectar por el *software* y expertos en ciberseguridad, permitiendo lanzar ataques más complejos, precisos, personalizados y a gran escala, y, por tanto, mucho más efectivos.

El Prontuario de ciberseguridad para entidades locales, elaborado por el Centro Criptológico Nacional (CCN) y la Federación Española de Municipios y Provincias (abril, 2021), muestra la realidad de los riesgos y amenazas que emanan del ciberespacio, y que pueden amenazar el normal desenvolvimiento de los procedimientos administrativos, las funciones involucradas en el desarrollo institucional provincial o municipal, y la gestión y administración de las entidades locales. Además, establece que, aunque el nivel de amenaza varía según los ayuntamientos, todos ellos poseen información o infraestructura de interés para los ciberatacantes.

A continuación, se describen las principales amenazas presentadas en dicho prontuario, y se presenta un análisis de cómo pueden ser potenciadas por la IA cuando es utilizada de forma malintencionada por los ciberatacantes, así como el riesgo que suponen para las entidades locales.

2.1. Principales ciberataques

2.1.1. *Malware*

El prontuario lo define como un *software* malicioso, como puede ser un virus, troyano, gusano, o cualquier código o contenido que pueda tener un impacto adverso en organizaciones o individuos.

Una de las aplicaciones más importantes de la IA en el mundo del cibercrimen es la generación automatizada de *malware*. Los algoritmos de aprendizaje automático pueden analizar grandes conjuntos de datos de *malware* existente y aprender a crear variantes nuevas y únicas. Esto significa que los ciberdelincuentes pueden crear *malware* adaptado a

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

objetivos específicos, aumentando la eficacia y reduciendo la probabilidad de detección.

Por ejemplo, los ciberatacantes podrían utilizar IA para crear un *malware* de espionaje (*spyware*) altamente sofisticado, que se instalara en los sistemas de una entidad local sin ser detectado. Este *spyware* podría tener capacidades avanzadas para analizar grandes volúmenes de datos y extraer información valiosa de forma automática, como correos electrónicos, documentos internos, grabaciones de audio, y cualquier otra información sensible almacenada en los sistemas de la entidad. Gracias a la IA, el *malware* puede aprender y adaptarse para evitar ser detectado por los sistemas de seguridad, modificando su comportamiento dinámicamente en respuesta a las defensas cibernéticas, y haciendo que sea extremadamente difícil de identificar y eliminar. Esto supondría una amenaza significativa para las entidades locales, ya que puede llevar a la filtración de información sensible, comprometer la privacidad de los ciudadanos, y causar un daño duradero a la reputación y funcionalidad de la entidad.

2.1.2. *Ransomware*

Se trata de un tipo de *malware* que bloquea los sistemas o los datos de los ordenadores de sus víctimas, permitiéndoles el acceso una vez que se satisface un pago (extorsión).

La IA puede ser utilizada por ciberatacantes para crear un *ransomware* altamente sofisticado y específico para una entidad local mediante la automatización de diversas etapas del ataque.

Utilizando algoritmos de aprendizaje automático, los atacantes pueden analizar el tráfico de red y los patrones de comportamiento de los usuarios para identificar las vulnerabilidades más críticas en la infraestructura de la entidad. Una vez infiltrado, el *ransomware* potenciado por la IA puede evadir la detección mediante la modificación dinámica de su código y comportamiento. Además, puede emplear técnicas avanzadas de cifrado para asegurar que los datos sean irrecuperables sin el pago del rescate, y moverse lateralmente dentro de la red para maximizar su impacto, comprometiendo sistemas críticos y servicios esenciales.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

2.1.3. *Phishing*

El *phishing* consiste en enviar correos electrónicos que simulan proceder de un organismo público o de una persona, persiguiendo extraer información sensible de los ciudadanos, de la propia entidad, o de sus responsables o empleados. Con la ayuda de la IA, los ciberdelincuentes suplantan la identidad de empresas en correos electrónicos muy persuasivos, animando al usuario a facilitar información personal, a clicar en enlaces o a descargar archivos adjuntos que pueden contener *software* malicioso. Es, sin duda, uno de los usos malintencionados más frecuentes de la IA por parte de los atacantes.

Los riesgos derivados de una campaña de *phishing* dirigida contra una entidad local son numerosos y graves. Entre los principales se encuentra el robo de credenciales, que puede dar a los atacantes acceso no autorizado a sistemas internos y datos sensibles, incluyendo información personal de ciudadanos y documentos críticos. Esta brecha de seguridad puede llevar a violaciones de privacidad, pérdidas económicas significativas debido a transacciones fraudulentas, y la interrupción de servicios esenciales como agua, electricidad y transporte público. Además, un ataque exitoso puede dañar gravemente la reputación de la entidad, erosionando la confianza de los ciudadanos y otras partes interesadas.

2.1.4. Ingeniería social

Consiste en la recopilación de información personal sin el uso de la tecnología, como, por ejemplo, a través de mentiras, trucos, sobornos, etc.

La IA mejora significativamente la efectividad de los ataques de ingeniería social al permitir una recopilación de información más exhaustiva, ya que puede analizar grandes volúmenes de datos de redes sociales, foros, correos electrónicos y otras fuentes públicas, para obtener información detallada sobre los objetivos. También puede crear mensajes altamente personalizados y convincentes que se dirigen a las vulnerabilidades específicas del objetivo, utilizando técnicas como el procesamiento del lenguaje natural (NLP). Estos mensajes pueden parecer provenir de colegas, amigos o familiares, aumentando su credibilidad. Además, puede generar *deepfakes*, que son simulaciones de vídeo y audio realistas, para imitar a personas de confianza. Esto puede ser utilizado para hacer llamadas telefónicas fraudulentas o enviar mensajes de vídeo falsos que persuadan al objetivo de realizar acciones específicas, como transferir dinero o revelar información confidencial.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

Por ejemplo, un ciberatacante podría utilizar algoritmos de aprendizaje automático para analizar las redes sociales y los correos electrónicos de los empleados de una entidad local. La IA podría identificar a un empleado clave del departamento de finanzas, y recopilar información sobre sus interacciones y horarios. Usando esta información, el atacante podría crear un *deepfake* convincente de la voz del alcalde solicitando urgentemente una transferencia de fondos para un proyecto municipal crítico. El *deepfake* se enviaría como un mensaje de voz a través del sistema de comunicación interno del ayuntamiento. Debido a la personalización y el alto nivel de realismo, el empleado de finanzas, confiando en la autenticidad del mensaje, podría realizar la transferencia de fondos a una cuenta controlada por el atacante, resultando en una pérdida financiera significativa para la entidad local.

2.1.5. Explotación de vulnerabilidades

Consiste en un intento de comprometer un sistema o interrumpir un servicio mediante la explotación de las vulnerabilidades organizativas o técnicas del sistema atacado.

Los ciberdelincuentes utilizan la IA para identificar vulnerabilidades en sistemas y aplicaciones. Los algoritmos pueden analizar miles de líneas de código para encontrar debilidades que puedan explotarse. Esto acelera el proceso de encontrar vulnerabilidades y, en última instancia, facilita la creación de *exploits* que aprovechen estas debilidades.

Un ataque de explotación de vulnerabilidades con IA a una entidad local podría involucrar a un ciberatacante que utiliza una herramienta de IA avanzada para escanear y analizar continuamente la infraestructura de TI de una entidad local en busca de debilidades. La IA identifica una vulnerabilidad no parcheada en el servidor web que gestiona los servicios ciudadanos en línea. Aprovechando esta vulnerabilidad, el atacante despliega un *exploit* que le permite obtener acceso no autorizado al servidor. Una vez dentro, la IA ayuda al atacante a moverse lateralmente dentro de la red, identificando y explotando otras debilidades en sistemas interconectados. Esto permite al atacante extraer datos sensibles, como información personal de los ciudadanos y registros financieros, y potencialmente instalar *ransomware* para cifrar los sistemas críticos, dejando a la entidad local incapacitada para operar hasta que se pague un rescate.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

2.1.6. Denegación de servicio

Se trata de la interrupción o ralentización de un servicio por múltiples peticiones, normalmente aplicaciones web.

La IA, al analizar patrones de tráfico en tiempo real, podría organizar ataques para sobrecargar los servidores de una empresa específica, adaptándose continuamente para superar sin esfuerzo las medidas de ciberseguridad, y causando interrupciones prolongadas en los servicios *online*.

Por ejemplo, un ciberatacante podría emplear una red de bots controlada por IA para coordinar un ataque masivo. La IA analiza el tráfico de red normal de los servidores de la entidad local para identificar patrones y horarios de menor resistencia. Aprovechando esta información, el atacante lanza un ataque de denegación de servicio sofisticado durante un momento crítico, como el periodo de inscripción para servicios municipales, o el pago de impuestos. La IA ajusta dinámicamente la intensidad y los vectores de ataque en tiempo real para evadir las defensas y maximizar el impacto, inundando los servidores con un volumen abrumador de solicitudes falsas. Esto provoca que los sistemas se sobrecarguen y dejen de funcionar, interrumpiendo los servicios esenciales para los ciudadanos y causando caos y frustración, además de posibles daños a la reputación de la entidad local.

2.1.7. Acceso no autorizado a la información

El prontuario define este ataque como la sustracción de credenciales de acceso mediante interceptación de tráfico o mediante el acceso a documentos físicos.

Los algoritmos de IA pueden aprender de los patrones de detección de sistemas de seguridad como *firewalls*, sistemas de detección de intrusos (IDS) y sistemas de prevención de intrusos (IPS). Utilizando esta información, la IA puede adaptar el comportamiento del ataque para evadir estas defensas, por ejemplo, modificando el tráfico de red para que parezca legítimo, o fragmentando el ataque en paquetes más pequeños y menos detectables.

También puede optimizar ataques de fuerza bruta para adivinar contraseñas de manera más eficiente. Utilizando algoritmos de aprendizaje automático, la IA puede analizar patrones comunes en la creación de contraseñas y priorizar intentos que tienen más probabilidades de éxito. Esto puede reducir significativamente el tiempo necesario para romper

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

una contraseña. Además, puede mejorar el funcionamiento de *keyloggers*, que son programas maliciosos diseñados para registrar las pulsaciones de teclas en un dispositivo infectado. La IA puede filtrar y analizar los datos recogidos para identificar automáticamente credenciales de acceso, como nombres de usuario y contraseñas, y otros datos sensibles, como números de tarjetas de crédito. O puede ser utilizada para analizar grandes volúmenes de datos robados con el fin de identificar rápidamente información valiosa, como credenciales de acceso, números de tarjetas de crédito y otros datos sensibles, facilitando así el robo de información crítica de manera eficiente y efectiva.

El riesgo de un ataque de este tipo a una entidad local es diverso y puede tener consecuencias graves y duraderas. Primero, la exposición de credenciales y documentos confidenciales puede llevar a un acceso no autorizado continuo, permitiendo a los atacantes explotar la información robada para cometer fraudes financieros, como el desvío de fondos municipales o el robo de identidades de empleados y ciudadanos. Segundo, la pérdida de datos sensibles puede resultar en la interrupción de servicios esenciales, afectando la capacidad de la entidad local para operar eficazmente y brindar servicios críticos a la comunidad. Además, la filtración de información podría dañar gravemente la reputación de la entidad local, minando la confianza de los ciudadanos y otras partes interesadas. La entidad también podría enfrentar consecuencias legales y regulatorias, incluyendo multas y sanciones por no proteger adecuadamente la información. Por último, la reparación de los sistemas comprometidos y la recuperación de la información robada pueden ser costosas y llevar mucho tiempo, impidiendo el funcionamiento normal de la entidad, y poniendo en riesgo la seguridad y el bienestar de la comunidad que sirve.

2.1.8. Suplantación

Consiste en un tipo de ataque en el que una entidad suplanta a otra para obtener beneficios ilegítimos.

La IA puede analizar grandes volúmenes de datos públicos y privados para recopilar información detallada sobre los funcionarios clave de la entidad que se pretende suplantar, como sus nombres, cargos y patrones de comunicación. Esta información se utiliza para crear perfiles detallados que facilitan la personalización del ataque.

Luego, puede generar *deepfakes* muy realistas, imitando a los funcionarios de la entidad suplantada. Estos *deepfakes* pueden ser utilizados en

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

llamadas telefónicas, videoconferencias o mensajes de voz para convencer a la entidad objetivo de que está interactuando con representantes legítimos.

También puede generar documentos falsificados pero realistas, como cartas oficiales, formularios de solicitud de fondos y contratos, que pueden ser presentados a la entidad objetivo como parte de una solicitud de fondos o recursos. Estos documentos falsos se respaldan con la información y los perfiles previamente recopilados, haciendo que el engaño sea más difícil de detectar. En conjunto, estas técnicas permiten a los atacantes obtener beneficios ilegítimos, como la transferencia de fondos o el acceso a recursos, al hacer creer a la entidad objetivo que está interactuando con otra entidad local legítima.

2.1.9. *Hacktivismo*

Se trata de ataques a sitios web o cuentas de redes sociales para publicitar una causa concreta.

Por ejemplo, un grupo activista podría querer publicitar una causa concreta, como la oposición a un proyecto de construcción en una entidad local. Para lograr una mayor visibilidad y apoyo, podría utilizar técnicas avanzadas de IA para comprometer sitios web y cuentas de redes sociales. Por ejemplo, podrían generar bots impulsados por IA que pueden crear y gestionar cuentas en redes sociales para difundir desinformación a gran escala. Estos bots pueden interactuar con usuarios reales, compartir contenido falso y amplificar narrativas engañosas. También podrían generar comentarios automáticos en la página web de la entidad local, en sus foros y redes sociales, para influir en discusiones y sembrar discordia. Estos comentarios pueden parecer genuinos y provenientes de usuarios reales.

2.1.10. Amenaza interna

Las amenazas mencionadas hasta ahora surgían por el uso malintencionado de la IA por parte de los ciberatacantes. Pero hay que tener en cuenta que no todas las amenazas provienen de fuentes externas.

Los errores humanos en el uso de la IA dentro de una entidad local representan una amenaza interna que puede tener repercusiones graves en términos de seguridad de datos, eficiencia operativa y equidad social.

Uno de los principales riesgos radica en la posibilidad de que datos sensibles o confidenciales sean manejados inadecuadamente, lo que po-

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

dría llevar a filtraciones de información o a la exposición de datos personales de los ciudadanos. Esto no solo compromete la privacidad de los individuos, sino que también puede erosionar la confianza pública en la entidad local y sus capacidades para gestionar de manera segura y efectiva la información.

Además, los errores cometidos por los empleados en el uso de la IA pueden tener consecuencias operativas significativas. Por ejemplo, si un empleado introduce datos incorrectos en un sistema de IA encargado de optimizar la asignación de recursos municipales, esto podría resultar en una distribución ineficiente de esos recursos, afectando negativamente la prestación de servicios públicos. En un contexto más crítico, la interpretación inadecuada de los análisis proporcionados por la IA podría llevar a decisiones mal fundamentadas que afecten negativamente a la planificación urbana, la seguridad pública o la gestión de emergencias.

La falta de formación y entendimiento sobre cómo funcionan los sistemas de IA también contribuye a la amenaza interna. Los empleados que no están adecuadamente capacitados pueden no ser conscientes de los sesgos inherentes en los algoritmos de IA, o de cómo sus propias acciones pueden influir en los resultados generados por estos sistemas. Esto puede perpetuar desigualdades y dar lugar a decisiones basadas en datos sesgados, exacerbando problemas sociales y económicos dentro de la comunidad.

2.2. Impacto de los ciberataques en las entidades locales

Los ciberataques a entidades locales pueden tener impactos profundos y diversos, afectando tanto a sus operaciones como a la confianza de los ciudadanos.

En primer lugar, los ataques pueden causar una interrupción significativa de los servicios esenciales. Por ejemplo, los sistemas informáticos que gestionan el suministro de agua, la recolección de basura, el transporte público y los servicios de emergencia pueden quedar paralizados, lo que resulta en la interrupción de la vida cotidiana de los ciudadanos y en retrasos y cancelaciones de servicios.

Además, los ciberataques pueden comprometer datos sensibles. La exposición de información personal, como números de seguridad social, direcciones, información de salud y datos financieros, puede tener graves consecuencias para la privacidad de los ciudadanos. La pérdida o alteración

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

de datos críticos también puede afectar la capacidad de la entidad local para operar de manera efectiva y tomar decisiones informadas.

El impacto económico de un ciberataque puede ser significativo. Los costos asociados con la recuperación de sistemas afectados, la restauración de datos y la mejora de las medidas de seguridad pueden ser elevados. En algunos casos, las entidades locales pueden verse obligadas a pagar rescates para recuperar el acceso a sus sistemas y datos.

La reputación de la entidad local también puede sufrir. La confianza del público en la capacidad de la entidad para gestionar y proteger sus intereses puede erosionarse, afectando su imagen y su relación con los ciudadanos y otras partes interesadas.

Desde una perspectiva legal y regulatoria, los ciberataques pueden tener consecuencias severas. Las entidades locales están sujetas a regulaciones de protección de datos y pueden enfrentar sanciones legales y multas si se violan estas normas. Además, los ciudadanos afectados por la exposición de sus datos personales pueden emprender acciones legales, resultando en litigios costosos.

La seguridad pública puede estar en riesgo debido a los ciberataques. Los ataques a infraestructuras críticas, como la red eléctrica o los sistemas de agua, pueden tener consecuencias graves para la seguridad de los ciudadanos. Asimismo, los ataques a los sistemas de comunicación y operación de servicios de emergencia pueden dificultar la respuesta a incidentes y emergencias.

El impacto en los empleados de la entidad local también es notable. El estrés y la ansiedad pueden aumentar debido a la presión adicional para gestionar las consecuencias del ataque, lo que puede afectar su moral y desviar su atención de otras tareas importantes.

Finalmente, los ciberataques pueden afectar la planificación y los presupuestos futuros de la entidad local. Los recursos financieros destinados a proyectos y mejoras pueden tener que ser redirigidos para abordar las consecuencias del ataque y fortalecer las defensas cibernéticas. Esto puede resultar en ajustes en las prioridades y en la planificación estratégica a largo plazo.

En definitiva, los ciberataques a entidades locales pueden tener impactos diversos que afectan todos los aspectos de su funcionamiento y relación con el público. La recuperación de un ataque requiere tiempo, re-

cursos y un esfuerzo concertado para restaurar la confianza y la seguridad. Por eso, es necesario establecer los mecanismos técnicos adecuados para hacer frente de forma proactiva a estos ataques.

3. Revisión de los mecanismos técnicos adecuados para hacer frente de forma proactiva a ciberataques

Para contrarrestar estas amenazas, se han desarrollado mecanismos técnicos avanzados, como los sistemas de IA guardianes, que permiten una defensa proactiva contra los ciberataques. En este sentido, la IA puede potenciar la ciberseguridad, y ser muy beneficiosa integrándose en estos mecanismos, mejorándolos frente a enfoques más tradicionales.

Los sistemas de IA guardianes son soluciones avanzadas diseñadas para proteger redes, sistemas y datos frente a amenazas cibernéticas. Estos sistemas utilizan tecnologías de IA, como el aprendizaje automático (*Machine Learning*) y el procesamiento del lenguaje natural (NLP), para identificar, analizar y responder a actividades sospechosas en tiempo real. A diferencia de los sistemas de seguridad tradicionales que dependen de reglas predefinidas y firmas conocidas de *malware*, los sistemas de IA guardianes tienen la capacidad de aprender y adaptarse continuamente, mejorando su efectividad a medida que recopilan y procesan más datos.

Las principales características de los sistemas de IA guardianes incluyen la detección de patrones, la clasificación, la automatización, y el procesamiento de lenguaje natural. A continuación, se detalla cómo pueden dichas características ayudar a las entidades locales a hacer frente de forma proactiva a los ciberataques.

3.1. Detección de patrones

Los sistemas de IA utilizan técnicas de aprendizaje automático para detectar patrones de comportamiento en tiempo real, y esto permite identificar comportamientos anómalos en los sistemas, y predecir posibles ataques antes de que ocurran.

Esto va a ser muy útil en la detección de vulnerabilidades en un sistema. Los sistemas de detección de intrusión (IDS) tradicionales se basan en detectar firmas o patrones conocidos. Esto significa que un tipo de ataque completamente nuevo puede no ser detectado en absoluto, porque la firma no existe en la base de datos. Por el contrario, la IA revisa el tráfico de

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

red, la actividad de los usuarios, los registros de eventos y de auditoría del sistema, en busca de actividad inusual o comportamientos sospechosos, que puedan indicar una actividad maliciosa que implique una vulnerabilidad.

Lo mismo ocurre con la detección de *malware*. Los antivirus tradicionales se basan en la detección de firmas. La IA analiza los ficheros para decir si son buenos o malos con un cierto grado de probabilidad. El veredicto no está basado en una característica, sino en múltiples características que van a dar una clasificación benigna o maligna del fichero. Esto permite extraer "el ADN" del *malware*, las características fundamentales por las que podemos decir que esto es un *malware* que se comporta como otro *malware* conocido. Por lo tanto, debe ser de la misma familia. Es decir, la IA no utiliza firmas, sino ciertas características de comportamiento o estructura de los ficheros, y determina si ese fichero está infectado por un *malware* conocido, o por un *malware* relacionado con esa familia. De esa manera, la defensa contra campañas es más efectiva, porque durante una campaña la misma familia de *malware* va mutando, y esto nos puede ayudar a ser más efectivos a la hora de detectar esa mutación de las campañas. Esto es especialmente interesante en las amenazas de tipo *zero day*, que no se pueden detectar por medios convencionales, porque aún no se han creado las firmas.

En definitiva, gracias a la capacidad de detección de patrones de la IA, una entidad local puede identificar y responder rápidamente a actividades sospechosas antes de que se conviertan en incidentes de seguridad graves. Esto no solo protege la infraestructura de red y los datos sensibles de los ciudadanos, sino que también asegura la continuidad de los servicios municipales y mantiene la confianza pública en la capacidad del ayuntamiento para salvaguardar la información y los recursos críticos.

3.2. Clasificación

Esta característica de la IA se refiere a su capacidad para categorizar datos en diferentes grupos o clases basándose en sus características y patrones. Este proceso implica que un algoritmo de IA aprende a identificar y diferenciar entre distintos tipos de datos a partir de ejemplos previamente etiquetados durante una fase de entrenamiento. Una vez entrenado, el modelo puede asignar nuevas entradas a las categorías adecuadas con un alto grado de precisión.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

En el ámbito de la ciberseguridad, la capacidad de clasificación de la IA puede ser extremadamente útil en la caza de amenazas. Esta se basa en el descubrimiento de concatenaciones de sucesos que puedan responder a un patrón de ataque. La IA puede clasificar la información que le llega y predecir qué tipo de ataque estamos sufriendo, incluso aunque sea un ataque desconocido. O verificar si un conjunto de esos ataques que estamos viendo en los millones de ficheros que estamos analizando al día se puede atribuir a una campaña. En definitiva, enriquecer toda la inteligencia de la caza de amenazas y así mejorar las medidas preventivas.

Asimismo, la IA puede ser útil en el análisis de comportamiento. Puede clasificar comportamientos de usuarios y dispositivos para detectar actividades anómalas que podrían indicar una brecha de seguridad. Por ejemplo, si un usuario que normalmente accede a ciertos sistemas durante el horario laboral comienza a descargar grandes cantidades de datos fuera de horario, un sistema de IA podría clasificar este comportamiento como sospechoso y alertar a los administradores de seguridad.

O también en la respuesta a incidentes. Durante un incidente de seguridad, la clasificación de la IA puede ayudar a priorizar las alertas y gestionar los recursos de respuesta. Por ejemplo, la IA puede clasificar las alertas de seguridad en función de su gravedad y el potencial impacto, permitiendo a los equipos de respuesta centrarse primero en las amenazas más críticas.

Finalmente, puede ayudar en el filtrado de correo electrónico. Los sistemas de IA pueden clasificar correos electrónicos entrantes en categorías como “spam”, “phishing” o “legítimos”. Esta capacidad ayuda a prevenir que los correos electrónicos maliciosos lleguen a los usuarios finales, reduciendo el riesgo de ataques de ingeniería social y robo de información.

En resumen, la clasificación por IA puede mejorar significativamente la detección de amenazas, la protección de datos y la eficiencia operativa en la ciberseguridad de una entidad local. La IA puede analizar y clasificar correos electrónicos entrantes para detectar y bloquear mensajes maliciosos, monitorizar el comportamiento de los empleados para identificar actividades anómalas que podrían indicar una brecha de seguridad, y gestionar vulnerabilidades en los sistemas municipales, priorizando las más críticas. Además, puede proteger la infraestructura de red, al clasificar el tráfico en tiempo real y asegurar los datos personales en las solicitudes de servicios ciudadanos.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

3.3. Automatización

La automatización es una de las características más poderosas y transformadoras de la IA. Se refiere a la capacidad de los sistemas de IA para realizar tareas y procesos de manera autónoma, sin intervención humana constante, basándose en algoritmos y modelos entrenados. En el ámbito de la ciberseguridad, la automatización puede desempeñar un papel crucial, especialmente en entidades locales como ayuntamientos y Administraciones municipales, al mejorar la eficiencia, la precisión y la rapidez de las respuestas ante amenazas.

La automatización a través de la IA mejora significativamente la ciberseguridad de una entidad local, al reducir la dependencia de la intervención humana, lo que permite una respuesta más rápida y efectiva ante amenazas. La automatización de las tareas mencionadas anteriormente, como la detección de vulnerabilidades, de *malware*, la caza de amenazas, o la respuesta a incidentes, libera a los equipos de TI de tareas repetitivas y laboriosas, y les permite centrarse en actividades estratégicas y de mayor valor añadido, como proteger los datos y servicios críticos de los ciudadanos.

Además, la IA puede automatizar la generación de informes de seguridad y cumplimiento normativo, recopilando y organizando datos relevantes de manera precisa y eficiente. Esto es especialmente útil para entidades locales que deben cumplir con regulaciones específicas sobre protección de datos y seguridad de la información. Esto reduce la carga de trabajo manual y garantiza que los informes estén siempre actualizados y listos para auditorías.

3.4. Procesamiento del lenguaje natural

El procesamiento del lenguaje natural (PLN) es una rama de la IA que se enfoca en la interacción entre los ordenadores y los seres humanos a través del lenguaje natural. El objetivo del PLN es permitir que las máquinas comprendan, interpreten y generen el lenguaje humano de una manera que sea valiosa. Estamos hablando de la IA Generativa, que está suponiendo una revolución en estos momentos, y también promete revolucionar la ciberseguridad, y ser una herramienta poderosa en este ámbito, especialmente en las entidades locales.

En primer lugar, puede ayudar a la detección de amenazas. Utilizando el PLN para analizar grandes volúmenes de datos, como correos electrónicos, mensajes y registros de actividad, se pueden identificar patrones sos-

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

pechosos y posibles amenazas. Por ejemplo, detectar correos electrónicos de *phishing* mediante el análisis del contenido del mensaje.

También puede ayudar en el análisis de vulnerabilidades, analizando informes de vulnerabilidades y alertas de seguridad, resumiendo información crucial, y sugiriendo acciones prioritarias para mitigarlas. Además, puede monitorizar estas plataformas en busca de menciones de posibles amenazas o actividades maliciosas que puedan afectar a la entidad.

Los chatbots y asistentes virtuales que utilizan PLN pueden dar soporte para responder automáticamente a consultas relacionadas con la ciberseguridad, proporcionando información y recomendaciones a los usuarios en tiempo real.

Por último, pueden ayudar a generar y analizar contenido de ciberseguridad, como políticas de seguridad, definiéndolas, revisándolas, y asegurando que estén actualizadas y cumplan con las normativas vigentes.

3.5. Caso de estudio

A continuación, se plantea un caso de uso detallado de cómo una IA puede mejorar la ciberseguridad de un ayuntamiento.

Imaginemos que un ayuntamiento ha experimentado un aumento en intentos de ciberataques dirigidos a sus sistemas administrativos, incluyendo el sistema de gestión de ciudadanos, el portal de servicios *online* y las bases de datos de empleados. Estos ataques han generado preocupación entre los funcionarios debido a la posibilidad de robo de datos sensibles y la interrupción de servicios municipales esenciales.

Para abordar estas preocupaciones, el ayuntamiento decide implementar una solución de ciberseguridad basada en IA, diseñada para detectar, responder y mitigar amenazas cibernéticas en tiempo real.

El primer paso es realizar una evaluación exhaustiva de riesgos. Esto implica identificar las áreas más vulnerables de la infraestructura tecnológica del ayuntamiento, tales como bases de datos que contienen información personal de los ciudadanos, sistemas de pago de impuestos, y redes internas que soportan la operación diaria del ayuntamiento. Basado en la evaluación inicial, el ayuntamiento selecciona una solución de IA guardián que mejor se adapta a sus necesidades específicas. Las características clave a considerar incluyen la capacidad de integración con los sistemas exis-

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

tentes, la robustez de los algoritmos de detección de anomalías, y la facilidad de uso y gestión.

Una vez seleccionada la solución, el siguiente paso es la integración con la infraestructura existente del ayuntamiento. El equipo de TI del ayuntamiento integra la IA con sus sistemas existentes. La IA se conecta a la red del ayuntamiento y a todos los dispositivos críticos, comenzando a recopilar datos sobre el tráfico y las actividades de la red.

El sistema de IA guardián necesita un período de aprendizaje para establecer una línea base de comportamiento normal dentro del entorno del ayuntamiento. Durante los primeros meses, la IA se entrena utilizando datos históricos y tráfico en tiempo real, y aprende los patrones normales de comportamiento en la red del ayuntamiento, así como las actividades típicas de los empleados y ciudadanos.

Una vez entrenada, la IA comienza a monitorizar activamente la red en busca de patrones anómalos que puedan indicar un ciberataque. Por ejemplo, detecta intentos inusuales de acceso a la base de datos de ciudadanos desde ubicaciones no autorizadas. Utilizando algoritmos de clasificación, la IA puede determinar si estos intentos son legítimos o sospechosos.

La IA detecta una actividad sospechosa que sugiere un posible ataque de *phishing* dirigido a los empleados del ayuntamiento. Utilizando su capacidad de PLN, analiza los correos electrónicos sospechosos para identificar patrones de *phishing*. Inmediatamente, la IA alerta al equipo de ciberseguridad y comienza a tomar medidas automatizadas para contener la amenaza, como bloquear los correos electrónicos sospechosos y aislar las cuentas comprometidas.

Con base en el análisis del ataque de *phishing*, la IA actualiza sus modelos de detección y mejora su capacidad para identificar correos electrónicos similares en el futuro. Además, el equipo de ciberseguridad implementa nuevas políticas de formación para empleados sobre cómo reconocer y reportar correos electrónicos sospechosos.

Es fundamental que el personal del ayuntamiento esté bien capacitado en el uso del sistema de IA guardián. Esto incluye la interpretación de alertas y reportes generados por el sistema, la comprensión de las acciones automáticas que el sistema puede tomar, y el manejo de incidentes de seguridad. La capacitación también debe abarcar la concienciación sobre las mejores prácticas de seguridad para minimizar errores humanos.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

Después de mitigar la amenaza, la IA realiza un análisis forense para entender cómo se originó el ataque y qué vulnerabilidades fueron explotadas. Utilizando sus capacidades de detección de patrones y clasificación, la IA identifica áreas de mejora en las políticas de seguridad. Los hallazgos se utilizan para actualizar las políticas de seguridad y mejorar la respuesta a futuros incidentes.

La IA continúa aprendiendo y adaptándose a nuevas amenazas. A través de actualizaciones regulares y del aprendizaje automático, se vuelve más eficiente en la detección y mitigación de ciberataques. Su capacidad de procesamiento del lenguaje natural también se mejora continuamente, permitiendo una mejor identificación y respuesta a amenazas basadas en texto.

3.6. Beneficios de la implementación de sistemas de IA guardianes en ciberseguridad para las entidades locales

Como se ha visto en el caso de estudio anterior, al aprovechar las capacidades avanzadas de detección, análisis y respuesta de la IA, las entidades locales pueden proteger mejor sus infraestructuras críticas y datos sensibles, responder de manera eficaz a las amenazas, y optimizar el uso de sus recursos. En última instancia, esto no solo fortalece la seguridad de la información, sino que también contribuye a la confianza y satisfacción de los ciudadanos, cumpliendo con el mandato de proporcionar servicios seguros y eficientes.

A continuación, se resumen los principales beneficios para una entidad local, después de la implementación exitosa de un sistema de IA guardián.

1. Reducción de incidentes de seguridad: Con la monitorización continua y la respuesta automatizada, el número de incidentes de seguridad puede reducirse significativamente.

2. Protección mejorada de datos: Los datos sensibles de los ciudadanos y las operaciones críticas del ayuntamiento están mejor protegidos contra accesos no autorizados y filtraciones.

3. Eficiencia operativa: La automatización de tareas de monitorización y respuesta libera al personal de TI para enfocarse en proyectos estratégicos y mejorar la infraestructura tecnológica del ayuntamiento.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

4. Confianza de los ciudadanos: Una postura de seguridad sólida y proactiva genera confianza entre los ciudadanos, quienes se sienten más seguros al saber que sus datos personales y la información crítica de la ciudad están protegidos. Esto puede mejorar la percepción pública del ayuntamiento y su gestión.

5. Cumplimiento normativo: Los sistemas de IA guardianes ayudan al ayuntamiento a cumplir con regulaciones y normativas de protección de datos y ciberseguridad. Esto es especialmente importante para evitar sanciones legales y para demostrar una gestión responsable de la información.

6. Detección temprana de amenazas: La capacidad de detectar anomalías y patrones de comportamiento sospechosos en tiempo real permite al ayuntamiento identificar y neutralizar amenazas antes de que puedan causar daños significativos.

7. Respuesta rápida a incidentes: La automatización de la respuesta a incidentes asegura que las amenazas se gestionen de manera inmediata y eficiente, minimizando el tiempo de exposición y el impacto potencial de los ciberataques.

8. Análisis y reportes detallados: Los sistemas de IA pueden generar reportes detallados sobre las actividades de seguridad, proporcionando a los administradores una visión clara y completa del estado de la ciberseguridad. Esto facilita la toma de decisiones informadas y la planificación estratégica.

9. Mejora continua: La IA permite una mejora continua de las defensas cibernéticas. A medida que el sistema recopila y analiza más datos, se vuelve más efectivo para predecir y responder a nuevas amenazas, adaptándose a un entorno de amenazas en constante cambio.

10. Optimización de recursos: La automatización y la inteligencia avanzada permiten al ayuntamiento optimizar el uso de sus recursos de ciberseguridad. Esto es particularmente beneficioso para entidades locales que a menudo enfrentan limitaciones de presupuesto y personal.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

4. Directrices para la implementación de medidas de seguridad en el uso de medios electrónicos que empleen IA

Mejorar la resiliencia de las entidades locales frente a ciberataques implica adoptar un enfoque sistemático y bien planificado, que incluya la evaluación de riesgos, la selección de herramientas adecuadas, la capacitación del personal y la colaboración con entidades externas, entre otros aspectos.

A continuación, se describen una serie de directrices para implementar las medidas de seguridad en el uso de medios electrónicos que hagan uso de la IA, lo que puede transformar significativamente la capacidad de una entidad local para protegerse contra ciberamenazas.

1. Primero, se recomienda realizar una evaluación exhaustiva de las necesidades de ciberseguridad del ayuntamiento, identificando las áreas críticas que pueden beneficiarse de la implementación de la IA. Esto incluye definir objetivos claros, como la detección temprana de amenazas, la automatización de respuestas y la mejora de la capacidad de análisis forense.

2. Una vez identificadas las necesidades, se debe seleccionar una plataforma de IA adecuada que pueda integrarse sin problemas con los sistemas de TI y las infraestructuras de seguridad existentes. Es crucial entrenar los modelos de IA utilizando datos históricos y en tiempo real sobre el tráfico de red, los comportamientos del sistema y los incidentes de seguridad para detectar patrones anómalos y clasificar amenazas.

3. La implementación debería incluir agentes de detección de anomalías que monitoricen continuamente el tráfico de red y el comportamiento del sistema. Estos agentes deben estar actualizados con las últimas amenazas y técnicas de ataque conocidas. Además, se recomienda configurar agentes de respuesta a incidentes para automatizar la contención de amenazas, como el aislamiento de sistemas comprometidos y el bloqueo de direcciones IP sospechosas.

4. Para asegurar una respuesta efectiva, es importante realizar simulaciones de ciberataques y ejercicios de respuesta a incidentes para evaluar la efectividad de los agentes de IA y preparar al equipo de ciberseguridad. Los resultados de estas simulaciones deben utilizarse para ajustar y mejorar los modelos de IA.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

5. Para garantizar la seguridad de la infraestructura, es necesario mantener todos los sistemas y *software* de IA actualizados con los últimos parches de seguridad para proteger contra vulnerabilidades conocidas. Además, se debe asegurar que existan sistemas de respaldo y redundancia para mantener la operatividad en caso de fallos o ataques.

6. La protección de datos sensibles es fundamental. Se recomienda implementar medidas de anonimización y cifrado de datos, asegurando el cumplimiento con las regulaciones y normativas relevantes, como el Reglamento General de Protección de Datos (RGPD). También es esencial establecer canales de comunicación seguros para la coordinación entre los agentes de IA y el equipo de ciberseguridad, utilizando cifrado de extremo a extremo.

7. La capacitación continua del personal del ayuntamiento sobre las mejores prácticas de ciberseguridad y el uso de herramientas de IA es crucial. Esto incluye campañas de concienciación sobre *phishing* y otros ataques basados en ingeniería social, utilizando ejemplos reales y simulaciones para mejorar la preparación del personal. Es esencial invertir en la capacitación continua de los empleados y desarrollar protocolos robustos de manejo y supervisión de los sistemas de IA. Solo así se podrá garantizar que la integración de la IA en las operaciones de la entidad local se realice de manera segura y beneficiosa para todos los ciudadanos.

8. La colaboración con otras entidades locales y organismos de ciberseguridad es importante para compartir información sobre amenazas y mejores prácticas. Participar en redes y foros de ciberseguridad ayudará a mantenerse al día con las últimas tendencias y desarrollos en el campo.

9. Se recomienda realizar auditorías periódicas para evaluar la efectividad de la implementación de IA en ciberseguridad y ajustar los protocolos de respuesta según sea necesario. Mantener los sistemas de IA actualizados con las últimas técnicas de ciberseguridad y amenazas emergentes es esencial para una mejora continua.

10. Es importante implementar técnicas para la detección de anomalías de contenido, con el fin de asegurar que los modelos de IA no produzcan resultados que sean maliciosos, inexactos o ilegales. Los sistemas de IA generativa, como los modelos de lenguaje na-

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

tural, pueden generar respuestas o contenido que, sin una adecuada supervisión, podrían incluir información falsa, desviaciones no intencionadas (conocidas como "alucinaciones"), o incluso contenido que infrinja derechos de autor. Para mitigar estos riesgos, se deben emplear técnicas avanzadas de monitorización y validación, que permiten identificar y corregir anomalías en tiempo real, asegurando que el contenido generado sea preciso y conforme a las normativas legales.

11. Además, es conveniente asegurar que la infraestructura de IA sea escalable para manejar un aumento en el volumen de datos y la complejidad de las amenazas a medida que la entidad local crece. Los sistemas de IA deben ser flexibles para adaptarse a cambios en las políticas de seguridad y las necesidades operativas del ayuntamiento.

12. Finalmente, mantener una documentación completa y actualizada de todos los procesos, políticas y herramientas de ciberseguridad implementadas es crucial. Generar informes periódicos sobre el estado de la ciberseguridad ayudará a evaluar el rendimiento de los agentes de IA y tomar decisiones informadas sobre futuras mejoras.

Por su parte, los propios usuarios de la IA en una entidad local también son responsables de hacer un buen uso de esta tecnología. En este sentido, deben contemplar las siguientes directrices para garantizar una operación segura, responsable y efectiva:

1. Es crucial que los usuarios reciban una capacitación continua sobre el uso seguro y efectivo de los sistemas de IA, así como sobre las mejores prácticas en ciberseguridad. Deben mantenerse actualizados sobre las últimas amenazas y técnicas de ciberataque que podrían explotar vulnerabilidades en los sistemas de IA.

2. El manejo seguro de los datos es fundamental. Los usuarios deben asegurarse de que todos los datos utilizados y procesados por los sistemas de IA estén protegidos mediante técnicas de cifrado y anonimización siempre que sea posible. Además, deben cumplir con todas las leyes y regulaciones de protección de datos relevantes, como el Reglamento General de Protección de Datos (RGPD), para evitar violaciones de privacidad. Un reto importante es la concienciación de los usuarios, para recordarles que siempre validen

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

la salida de la IAG para comprobar su precisión antes de incorporarlos a sus trabajos, que no introduzcan información confidencial, etc.

3. Los usuarios deben monitorizar continuamente el comportamiento de los sistemas de IA, y reportar cualquier actividad sospechosa o anomalía al equipo de ciberseguridad de inmediato. Para ello, es esencial establecer canales claros y seguros para que los usuarios puedan informar sobre posibles incidentes de seguridad o vulnerabilidades.

4. Es fundamental que los usuarios comprendan cómo funcionan los sistemas de IA y las decisiones que toman. Los modelos de IA deben ser transparentes y sus decisiones explicables. Además, se deben realizar evaluaciones de impacto en la privacidad y la seguridad, para identificar y mitigar los riesgos asociados al uso de la IA.

5. El uso responsable y ético de los sistemas de IA es otra directriz importante. Los usuarios deben evitar utilizar estos sistemas para actividades que puedan ser consideradas abusivas o poco éticas, como la vigilancia excesiva o la toma de decisiones discriminatorias. Deben ser conscientes de su responsabilidad en el uso de la IA, y actuar en consecuencia para minimizar los riesgos.

6. Es esencial que los usuarios conozcan y sigan los protocolos establecidos para la gestión de incidentes de ciberseguridad, incluyendo la contención, mitigación y recuperación de incidentes. Participar en simulacros y ejercicios de respuesta a incidentes ayudará a mejorar la preparación y la capacidad de respuesta ante ciberataques.

7. Finalmente, fomentar la colaboración y la comunicación entre los usuarios, así como con el equipo de ciberseguridad, es crucial para una respuesta efectiva y coordinada ante cualquier amenaza. Los usuarios deben trabajar en equipo y compartir información relevante para mejorar la seguridad general de los sistemas de IA.

5. Conclusiones

La IA está transformando rápidamente el panorama de la ciberseguridad, marcando una revolución que afecta tanto a los atacantes como a los defensores. En este entorno dinámico, las entidades locales se encuentran en

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

una posición única. Estas Administraciones manejan una vasta cantidad de datos históricos y sensibles diariamente en casi todos los ámbitos. Todos estos datos son imposibles de gestionar por los humanos. Los algoritmos permiten estudiarlos, extraer patrones y exprimir todo su potencial. Su adecuada explotación ofrece innumerables ventajas para ellas y para los ciudadanos, a través de su aplicación en la gestión de los servicios públicos, en la toma de decisiones y en la ciberseguridad.

Desde una perspectiva de ciberseguridad, la IA está jugando un papel crucial en dos direcciones. Por un lado, está siendo utilizada por los atacantes para desarrollar ciberataques más sofisticados y difíciles de detectar. Por otro lado, los defensores están empleando la IA para fortalecer sus mecanismos de defensa y proteger sus sistemas de información.

Entre los ciberataques más comunes se encuentran el *malware*, el *ransomware*, el *phishing*, la ingeniería social, la explotación de vulnerabilidades, la denegación de servicio, el acceso no autorizado a la información, la suplantación, el *hacktivismo* y las amenazas internas. Estos ataques representan una amenaza constante para las entidades locales, y pueden tener un impacto devastador en la continuidad de los servicios públicos, la confianza de los ciudadanos y la integridad de los datos.

Para hacer frente a estos desafíos, se han desarrollado diversos mecanismos técnicos que utilizan IA, tales como la detección de patrones, la clasificación, la automatización y el procesamiento del lenguaje natural. Estos mecanismos permiten una respuesta más rápida y precisa ante las amenazas, mejorando la capacidad de detección y respuesta de las entidades locales.

La implementación de sistemas de IA en ciberseguridad ofrece numerosos beneficios, incluyendo una mayor capacidad de detección de amenazas, una respuesta más rápida y eficiente, y una mejor gestión de los recursos. Además, la IA puede ayudar a identificar vulnerabilidades y prevenir ataques antes de que ocurran, proporcionando una capa adicional de protección para los datos y servicios públicos.

Finalmente, es esencial seguir directrices claras para la implementación de medidas de seguridad en el uso de medios electrónicos que emplean IA. Estas directrices son fundamentales para garantizar la protección de los datos y la continuidad de los servicios públicos, asegurando así la seguridad y confianza de los ciudadanos.

Fundación Democracia y Gobierno Local
Serie: Claves del Gobierno Local, 48
ISBN: 978-84-128852-8-6

En conclusión, la IA se presenta como una herramienta poderosa y esencial en la lucha contra los ciberataques. Las entidades locales deben adoptar y adaptar estas tecnologías para proteger sus datos y servicios, garantizando así la seguridad y confianza de los ciudadanos.

6. Bibliografía

Centro Criptológico Nacional y Federación Española de Municipios y Provincias. (2021). *Prontuario de ciberseguridad para entidades locales. Abril 2021*. Disponible en https://transparencia.gob.es/transparencia/es/transparencia_Home/index/MasInformacion/Informes-de-interes/Seguridad/ProntuarioCiberseguridad-Abr2021.html.

ENISA (European Union Agency for Cybersecurity). (2020). *AI Cybersecurity Challenges*. Disponible en https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges.

European Commission. (2020). *White Paper on Artificial Intelligence - A European approach to excellence and trust*. Disponible en https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_en?filename=commission-white-paper-artificial-intelligence-feb2020_en.pdf.

Ministerio para la Transformación Digital y de la Función Pública. (2024). *Estrategia de Inteligencia Artificial 2024*. Disponible en https://portal.mineco.gob.es/es-es/digitalizacionIA/Documents/Estrategia_IA_2024.pdf.

Presidencia del Gobierno. (2019). *Estrategia Nacional de Ciberseguridad*. Disponible en https://www.dsn.gob.es/sites/default/files/documents/Estrategia%20Nacional%20de%20Ciberseguridad%202019.pdf.