https://doi.org/10.61530/redtis.vol7.n1.2023.154.115-122

PERCEPCIÓN DE CIBERSEGURIDAD EN SISTEMAS DE INTELIGENCIA ARTIFICIAL EN LA EDUCACIÓN SUPERIOR

Héctor Luis López López[1], Josué Joaquín Aguilera Zatarain [1], Saraí Rojas Solís [1], María de los Ángeles Rendón Rendón[1]

[1]Universidad Autónoma de Sinaloa (MÉXICO)

Resumen

Este artículo presenta los resultados de un estudio que se centró en la percepción de los alumnos de la Facultad de Informática Mazatlán en relación con la ciberseguridad en sistemas de inteligencia artificial (IA). La creciente importancia de la IA en la educación y la industria ha llevado a una mayor necesidad de comprender la percepción de los alumnos en cuanto a la seguridad de estos sistemas. El estudio se basó en encuestas y entrevistas a alumnos de diferentes niveles académicos. Los resultados revelan una serie de hallazgos significativos. En primer lugar, se identificó que la mayoría de los alumnos tiene una conciencia básica sobre la seguridad en sistemas de IA, pero existe una falta de comprensión en profundidad. La mayoría de los encuestados expresaron preocupación por la seguridad de los datos y la integridad de los sistemas de IA, pero muchos carecían de conocimientos técnicos para abordar estas inquietudes.

Además, se encontró que la percepción de la importancia de la seguridad en IA estaba influenciada por factores como la educación previa en seguridad cibernética, la exposición a amenazas cibernéticas y la conciencia mediática. Los resultados también resaltan la necesidad de una mayor educación en ciberseguridad en el currículo de la Facultad, lo que puede ser valioso para desarrollar estrategias de enseñanza y concienciación de la seguridad cibernética siendo fundamental en un mundo donde la IA desempeñará un papel cada vez más central en la sociedad, y la educación es clave para empoderar a los alumnos en este aspecto.

Palabras clave: amenazas, ataques cibernéticos, ciberseguridad, inteligencia artificial, sistemas inteligentes.

Abstract

This article presents the results of a study that focused on the perception of students at the Mazatlán Faculty of Informatics in relation to cybersecurity in artificial intelligence (AI) systems. The growing importance of AI in education and industry has led to a greater need to understand student perceptions of the security of these systems. The study was based on surveys and interviews with students of different academic levels. The results reveal a series of significant findings. Firstly, it was identified that the majority of students have a basic awareness of security in AI systems, but there is a lack of in-depth understanding. Most respondents expressed concern about data security and the integrity of AI systems, but many lacked the technical knowledge to address these concerns.

Additionally, the perception of the importance of security in AI was found to be influenced by factors such as prior cybersecurity education, exposure to cyber threats, and media awareness. The results also highlight the need for greater cybersecurity education in the Faculty's curriculum, which can be valuable in developing teaching strategies and awareness of cybersecurity being fundamental in a world where AI will play an increasingly central role. In society, and education is key to empowering students in this regard.

Keywords: threats, cyber attacks, cybersecurity, artificial intelligence, intelligent systems.

1 INTRODUCCIÓN

La seguridad cibernética es sin duda, crucial en la era actual en la que vivimos, ya que permite la rápida detección de análisis de datos y distintos contextos que podrían dar lugar a ciberataques. En este contexto, la inteligencia artificial (IA), desempeña un papel fundamental al mejorar su comprensión de las amenazas

cibernéticas, utilizando millones de datos. La IA está transformando la forma en que se analizan diversas alternativas, soluciones y errores. Su aplicación en ciberseguridad abarca desde la detección de intrusiones hasta el análisis cibernético. Al utilizar tecnologías como el aprendizaje automático y el procesamiento del lenguaje natural, la IA proporciona datos rápidos que simplifican las alertas diarias, reduciendo los tiempos de respuesta [1].

La importancia de la seguridad de la información en sistemas de IA, radica en la capacidad de la tecnología artificial para imitar la aptitud de respuesta humana. La IA y el aprendizaje automático ayudan a los equipos de tecnologías de la información a gestionar amenazas de manera efectiva. La inteligencia artificial mejora significativamente los sistemas de seguridad, analizando grandes volúmenes de datos en tiempo real y detectando patrones y comportamientos anormales que podrían indicar un ataque en curso.

En definitiva, la ciberseguridad en la IA es crucial para anticipar amenazas, acelerar respuestas y reducir riesgos cibernéticos, a pesar de sus beneficios, la dependencia de aplicaciones críticas en la IA plantea desafíos éticos y legales. ¿Quién es responsable en caso de un accidente causado por un vehículo autónomo? ¿Cómo se protege la privacidad de los datos recopilados por sistemas de inteligencia? Estas preguntas son cruciales a medida que estas tecnologías evolucionan.

En los últimos años, las amenazas de ciberseguridad han aumentado, desde programas diseñados para robar información confidencial hasta ataques que colapsan servidores [2]. Comprender estos ataques, su funcionamiento y las medidas para prevenirlos es fundamental. Como ejemplo, el phishing, es un ataque común, se utiliza para obtener información confidencial de los usuarios en internet, como números de tarjetas de crédito, contraseñas, entre otros datos de información. Asimismo, el malware un software malicioso, detona las vulnerabilidades de un sistema para obtener permisos no autorizados, como ejecutar programas o acceder a funciones del equipo, como sería ejecutar o desinstalar programas, acceder a funciones del equipo o sistema en donde está instalado, permite la extracción de información confidencial de los equipos de trabajo.

La ciberseguridad en sistemas de inteligencia artificial (IA), enfrenta una serie de desafíos y tendencias emergentes, incluyendo ataques dirigidos a modelos de IA y técnicas de aprendizaje automático adversarial. Existen muchos desafíos por los cuales los modelos de Inteligencia pueden destacar, como son los ataques cibernéticos. En particular, los ataques dirigidos a modelos de IA, pueden ser subcategorizados como adversarial attacks y data poisoning [3].

Los adversarial attacks, son los atacantes que pueden introducir perturbaciones imperceptibles en los datos de entrada para engañar al modelo de IA, lo que lleva a predicciones incorrectas y arroja soluciones erróneas, por otro lado los data poisoning, son los atacantes que pueden manipular los datos de entrenamiento del modelo para alterar su comportamiento, llevando a predicciones erróneas o a decisiones incorrectas.

Existen técnicas de aprendizaje automático adversarial, son un conjunto de técnicas que los adversarios utilizan para atacar los sistemas de aprendizaje automático, estos ataques explotan las vulnerabilidades y especificidades de los modelos de aprendizaje automático. Entre las técnicas de aprendizaje adversarial, están:

1. La generación de ejemplos adversariales: los atacantes pueden generar ejemplos específicamente diseñados para engañar al modelo.
2. Ataques transferibles: los ataques desarrollados para un modelo pueden ser transferidos y ser efectivos en otros modelos similares.

Uno de los tipos más comunes de ataques dirigidos a modelos de IA, son los ataques adversarios [3]. Los ataques adversarios se pueden clasificar en términos generales en dos tipos: dirigidos y no dirigidos [4].

En los ataques dirigidos, el atacante pretende manipular la salida del modelo de IA para una entrada específica o un conjunto de entradas [4]. Por otro lado, los ataques no dirigidos tienen como objetivo hacer que el modelo haga predicciones incorrectas para cualquier entrada [5]. Estos ataques se pueden llevar a cabo introduciendo pequeñas perturbaciones o ruido en los datos de entrada, lo que puede hacer que el modelo clasifique erróneamente la entrada [6].

Otro tipo de ataque dirigido a modelos de IA son los ataques de envenenamiento [7]. Estos ataques implican la introducción de datos maliciosos en el conjunto de entrenamiento utilizado para desarrollar el modelo de IA, con el objetivo de hacer que el modelo haga predicciones incorrectas para entradas específicas [6]. Los ataques de envenenamiento pueden ser particularmente dañinos porque pueden ser difíciles de detectar y los datos maliciosos pueden diseñarse cuidadosamente para mezclarse con datos legítimos [7]. Estos ataques pueden ser llevados a cabo por atacantes con acceso a los datos de capacitación, como personas internas maliciosas o proveedores externos [6].

Por último, los ataques de puerta trasera son otro tipo de ataque dirigido a modelos de IA [8]. Estos ataques implican insertar un disparador oculto o una puerta trasera en el modelo durante la fase de entrenamiento [9]. Luego, la puerta trasera puede activarse mediante una entrada específica o un conjunto de entradas, lo que hace que el modelo se comporte de maneras inesperadas [10]. Los ataques de puerta trasera pueden ser difíciles de detectar, ya que el desencadenante puede diseñarse cuidadosamente para combinarse con entradas legítimas. Estos ataques pueden ser llevados a cabo por atacantes con acceso a los datos de entrenamiento o por personas internas maliciosas [6].

Existen tecnologías que se encargan de mantener la seguridad en sistemas de IA, es por eso que daremos a conocer en este artículo algunas de ellas e implementaremos medidas para brindarles conocimiento acerca de cifrado, la autenticación de usuarios y detección de intrusiones. “La inteligencia artificial es muy útil para combatir con los usuarios que practican el robo de identidad o de información, esto es mediante un estudio de los comportamientos de patrones para detectar esas actividades sospechosas. Brinda ayuda en la protección contra el robo y manipulación de datos” [11].

El proyecto diseño de ambiente CTF para Entrenamiento en ciberseguridad busca diseñar un ambiente de aprendizaje de la seguridad de la tecnología en la información está centrado en competencias tipo CTF (Capture the Flag), mediante la identificación, desarrollo e implementación de un conjunto de herramientas, recursos tecnológicos y material de apoyo que permitan contar con un espacio que propicie el estudio y la investigación en el área [12].

La ciberseguridad es un aspecto fundamental en el mundo digital actual, y en especial en la Facultad de Informática de Mazatlán, se están implementando sistemas de inteligencia artificial, específicamente sistemas de prevención de amenazas, con el propósito de salvar la integridad, confidencialidad y disponibilidad de los recursos informáticos de la institución. “La información es vital, ya que las amenazas y ataques son constantes, y estar informado es crucial para protegerse contra ellos” [13].

Estos sistemas de prevención de amenazas basados en inteligencia artificial son un componente esencial en la defensa de la red de la facultad contra ataques cibernéticos y vulnerabilidades de seguridad. Utilizan algoritmos de aprendizaje automático y análisis de comportamiento para identificar patrones anómalos y actividades maliciosas en tiempo real.

2 METODOLOGÍA

El enfoque utilizado en nuestra investigación es mixta, y mediante la combinación del enfoque cualitativo y cuantitativo, se obtiene una comprensión más completa de la investigación, mediante el método cualitativo corresponde a la revisión general de la literatura existente y en segundo lugar se utilizó el método cuantitativo, con la aplicación de un instrumento tipo encuesta. El objetivo de este enfoque es utilizar la revisión de literatura para contextualizar y comprender el estado actual del conocimiento sobre de investigación, identificar distintas lagunas en la investigación que ya existe y establecer una base teórica sólida. Y a su vez la encuesta se utilizó para recopilar datos cuantitativos que pueden ser analizados de manera estadística para responder a preguntas de investigación específicas y obtener información empírica sobre el tema.

En la recopilación de la literatura para la investigación fue a través del acceso a fuentes confiables de artículos científicos en Internet, es fundamental para la comunidad académica y científica, mediante Google Académico, ya que es una herramienta de búsqueda que indexa una gran variedad de fuentes académicas, incluyendo revistas, conferencias y libros. Muchos de los artículos listados en Google Académico son revisados por pares y se pueden acceder de forma gratuita. Por otra parte, también se recopilo información en Web of Science, que es una plataforma que proporciona acceso a revistas científicas y ofrece herramientas de análisis de citas.

El cuestionario utilizado, se aplicó utilizando la plataforma de Google Forms. Los participantes completaron el cuestionario en línea a través de un enlace proporcionado, lo que permitió la recopilación eficiente de datos de manera electrónica. Este enfoque facilitó la organización de las respuestas, la administración de las preguntas y la posterior exportación de los datos para su análisis. El uso de Google Forms también ofreció la ventaja de contar con herramientas para la creación de formularios y la generación de informes que contribuyeron a la gestión efectiva de la encuesta.

La encuesta busca evaluar el conocimiento de los encuestados sobre los conceptos básicos, preguntas como: ¿Estás familiarizado con los conceptos básicos de ciberseguridad en sistemas de inteligencia artificial (IA)?, están destinadas a determinar el nivel de comprensión inicial. Se intenta recopilar información sobre experiencias prácticas de los encuestados en proyectos que involucren la implementación de medidas de seguridad en sistemas de IA. Las preguntas sobre proyectos previos y testimonios de incidentes de seguridad pretenden proporcionar ejemplos concretos y aplicables. Se busca entender la opinión de los encuestados sobre si consideran que la seguridad de datos, es una preocupación importante en el desarrollo y despliegue de sistemas de tecnologías artificiales.

A su vez el instrumento indaga sobre las medidas de ciberseguridad que los expertos consideran más efectivas para proteger sistemas de IA. Esto proporciona información sobre las prácticas recomendadas y las estrategias que los profesionales encuentran útiles. En conjunto, estas preguntas buscan obtener información detallada sobre diversos aspectos relacionados con la seguridad en sistemas de inteligencia artificial, proporcionando una comprensión completa de las percepciones, experiencias y opiniones de los expertos en el campo.

3 RESULTADOS

Dentro de la investigación de este artículo se llevó a cabo una encuesta entre los estudiantes de la Facultad de Informática Mazatlán (FIMAZ), y personas que están fuera del ámbito estudiantil para poder profundizar más en el tema y saber cuál es su conocimiento en este tema, o si es que lo tienen.

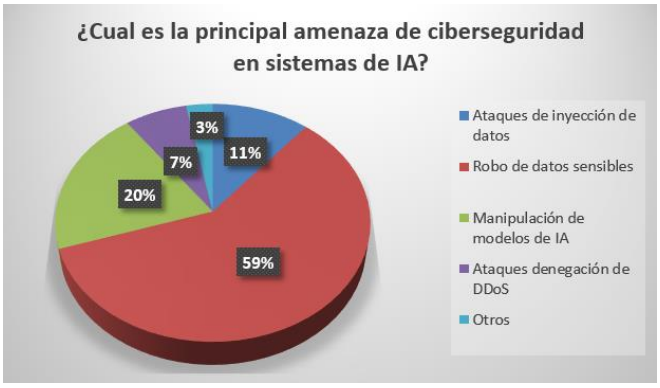
En este estudio, se llevó a cabo una evaluación exhaustiva de la ciberseguridad en los sistemas de inteligencia artificial (IA), utilizados en la Facultad de Informática Mazatlán. Los resultados de esta evaluación revelan hallazgos importantes en relación con la protección y vulnerabilidades de estos sistemas. A continuación, se presentan los resultados más relevantes de la investigación.

Gráfica 1.Conceptos básicos de ciberseguridad.

En la pregunta: ¿Estas familiarizado con los conceptos básicos de ciberseguridad en sistemas de inteligencia artificial (IA)? Los resultados arrojaron, que el 55% de los encuestados informaron si estar familiarizados con los conceptos básicos de seguridad en sistemas de IA, el 20% de los participantes comentaron no y el 25% de los encuestados respondió no estar seguro. Lo que indica un alto grado a no estar familiarizado con los conceptos.

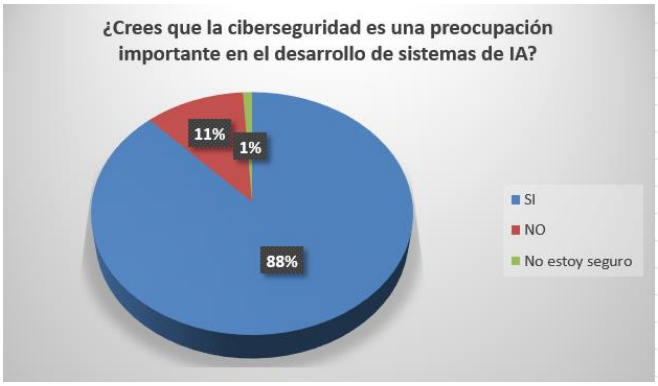
Gráfica 2. Proyectos que implementan medidas de ciberseguridad.

A los participantes se les preguntó si han trabajado en proyectos que involucran la implementación de medida de ciberseguridad en sistemas de IA. Los datos mostraron que el 73% de los encuestados declararon no haber trabajado en proyectos que implementen medidas de seguridad de los datos, el 18% manifestó no estar seguros y el 9% de los encuestados indico si haber trabajado en proyectos con medidas de ciberseguridad. Esta cifra sugiere un fuerte compromiso para la optar por la utilización de medidas de seguridad para futuros proyectos.



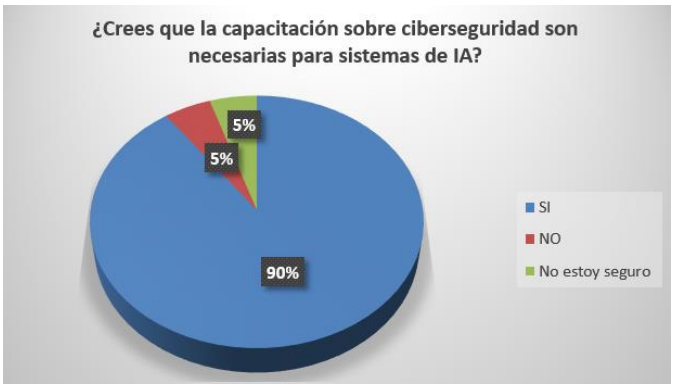
Gráfica 3. Principales amenazas de ciberseguridad.

Otra de las preguntas fundamentales y más relevantes utilizadas en la encuesta, se les solicitó que identificaran las principales amenazas de ciberseguridad en sistemas de IA. Entre las respuestas más frecuentes, fue el robo de datos sensibles con un 59%, en segundo lugar la manipulación de modelos de IA con un 20%, a continuación ataques de inyección de datos con 11%, en cuanto a ataques de denegación de servicio (DDoS), el 7% y finalmente el 3% con otros tipos de amenazas. Estos hallazgos reflejan un alto porcentaje en amenazas de seguridad de los datos en los alumnos de la Facultad de Informática Mazatlán.



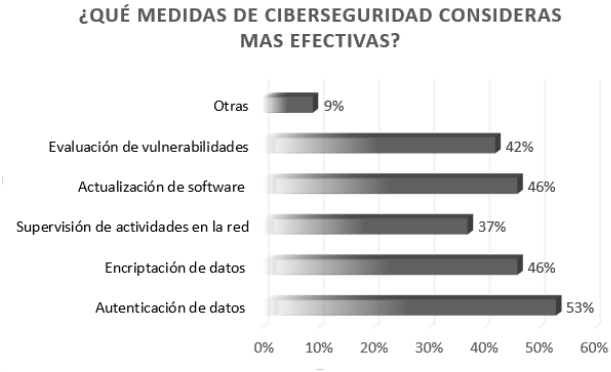
Gráfica 4. La ciberseguridad es una preocupación importante.

En esta gráfica se observa una tendencia constante hacia la preocupación el no utilizar la seguridad en los sistemas de IA con un 88%, el 11% de los encuestados respondieron no estar seguros que la ciberseguridad no es una preocupación y únicamente el 1% eligió no ser una preocupación. Esta tendencia refleja la importante preocupación en el área de la seguridad de datos en el ámbito informático.



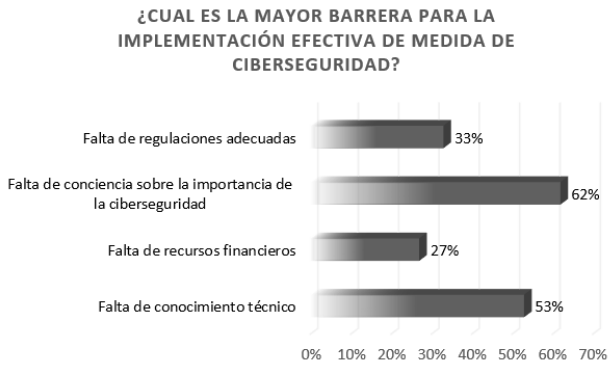
Gráfica 5. Capacitación sobre ciberseguridad es necesaria.

Observamos en esta gráfica que el 90% de los encuestados mencionaron ser necesaria la capacitación sobre ciberseguridad para los profesionales que laboran con sistemas de IA, y el alrededor de 10% respondió no o no estar seguros. Esta predicción nos demanda la necesita de mayor capacitación en temas de relacionados a la seguridad en los sistemas de inteligencia artificial.



Gráfica 6. Medidas de ciberseguridad más efectivas.

Se observa una tendencia constante hacia las medidas de ciberseguridad más efectivas, en la protección de los sistemas de IA, en primer lugar con el 53% de los encuestados afirmaron que la autenticación de usuarios ser la más segura, en segundo lugar empatados la encriptación de datos y la actualizaciones regulares de software con un 46%, seguido de la evaluación de vulnerabilidades con un 42%, le sigue la supervisión constante de actividades en la red con un 37% y por ultimo otras medidas de seguridad de sistemas con solamente el 9%.



Gráfica 7. Mayor barrera en la implementación de medidas de ciberseguridad.

Por último, en esta gráfica se observamos un claro ejemplo de los mayores obstáculos en la implementación de medidas de ciberseguridad, el 62% de los encuestados manifestó la falta de conciencia sobre la importancia del uso de la seguridad, el 53% la falta de conocimiento técnico, el 33 % la falta de regulaciones y únicamente el 27% contesto que se debe a la falta de recursos financieros.

Una limitación de nuestra investigación fue que esperábamos que más personas de las que teníamos contempladas hubieran participado en la encuesta que realizamos para que puedan estar enteradas de este tema y por medio de esta saber si cuentan con el conocimiento, aun así, la encuesta ha sido una herramienta que nos ha servido mucho para el tema.

4 CONCLUSIONES

En conclusión, la ciberseguridad en sistemas de inteligencia artificial (IA), se ha convertido en un tema crítico en la era digital actual. La rápida evolución de la tecnología ha llevado a un aumento en las amenazas cibernéticas, desde ataques de phishing hasta técnicas de aprendizaje automático adversarial. La IA desempeña un papel crucial al mejorar la comprensión de estas amenazas, utilizando millones de datos para detectar patrones y comportamientos anormales que podrían indicar un ataque en curso. Los ataques dirigidos a modelos de IA, como los ataques adversarios y de envenenamiento, representan desafíos significativos.

Estos ataques manipulan los datos de entrada para engañar al modelo, lo que puede tener consecuencias graves. Además, la ciberseguridad en la IA se ve afectada por preguntas éticas y legales, como la responsabilidad en caso de accidentes causados por vehículos autónomos. A pesar de estos desafíos, la IA también ofrece soluciones. Utilizando tecnologías como el aprendizaje automático y el procesamiento del lenguaje natural, la IA proporciona datos rápidos que simplifican las alertas diarias, reduciendo los tiempos de respuesta y mejorando la capacidad de los equipos de Tecnologías de la Información para gestionar amenazas de manera efectiva.

Este estudio sobre la percepción de los alumnos de la Facultad de Informática Mazatlán en relación con la ciberseguridad en sistemas de inteligencia artificial subraya la importancia de abordar la concienciación y formación en este nuevo campo de investigación. Los hallazgos destacan la necesidad de una educación más completa en seguridad en los datos en el plan de estudios de la Facultad y resaltan la relevancia de capacitar a los alumnos para enfrentar los desafíos de seguridad en un entorno cada vez más digitalizado y orientado hacia la inteligencia artificial. Estos resultados ofrecen una base sólida para la implementación de estrategias de enseñanza y formación que empoderen a los estudiantes para abordar eficazmente los aspectos de seguridad en sistemas de IA a lo largo de sus carreras académicas y profesionales.

REFERENCIAS

[1] IBM. (s.f.). IBM. Obtenido de https://www.ibm.com/mx-es/security/artificial-intelligence.Modelado de amenazas en inteligencia artificial. (s.f.). Obtenido de https://learn.microsoft.com/es-es/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2

[2] J. J. Chavez. Delta protect. (8 de Abril de 2023). Obtenido de https://www.deltaprotect.com/blog/amenazas-de-ciberseguridad

[3] Adversarial Machine Learning: ataques a modelos de ML. (28 de 10 de 2023). Obtenido de https://www.welivesecurity.com/la-es/2022/05/30/adversarial-machine-learning-introduccion-ataques-modelos-ml/

[4] IA y ataques adversarios. (s.f.). Obtenido de https://ts2.space/es/ia-y-ataques-adversarios/

[5] Adversarial Machine Learning (parte V): ataques de evasión. (s.f.). Obtenido de www.bbvanexttechnologies.com

[6] Modelado de amenazas en inteligencia artificial. (s.f.). Obtenido de https://learn.microsoft.com/es-es/azure/machine-learning/concept-deep-learning-vs-machine-learning?view=azureml-api-2

[7] Como la inteligencia artificial está siendo envenenada. (s.f.). Obtenido de blog.f-secure.com

[8] ¿Qué es un ataque de puerta trasera? Ejemplos y cómo? (s.f.). Obtenido de nordvpn.com/es/blog/ataque-de-puerta-trasera/

[9] 25 Tipos de ataques informáticos y cómo prevenirlos. (s.f.). Obtenido de ciberseguridad.blog

[10] Puerta trasera/Backdoor: pc infectado. (s.f.). Obtenido de www.pandasecurity.com/es/security-info/back-door/

[11] L. S. Garmendia. (28 de 8 de 2023). Fortalecer las defensas: cómo abordar la ciberseguridad en la era de la IA. Obtenido de Forbes México: https://www.forbes.com.mx/mexico-puede-liderar-la-revolucion-del-hidrogeno-verde-en-latam-h2v2/

[12] F, B., I, R., & A, V. (s.f.). Plataformas de gestión de escenarios de ciberseguridad para aprendizaje y entretenimiento. En M. Unibertsitatea, Actas de las Cuartas Jornadas Nacionales de Investigación em Ciberseguridad (págs. 55-62).

[13] L. E. Casallas Rodríguez. Estado actual de la ciberseguridad aplicada a sistemas defensivos y ofensivos a partir de inteligencia artificial. Monografia, Repositorio institucional UNAD, 2020. Obtenido de https://repository.unad.edu.co/handle/10596/34627