Universidad de Buenos Aires

Facultades de Ciencias Económicas, Ciencias Exactas y Naturales e Ingeniería

Carrera de Especialización en Seguridad Informática

Trabajo final

Título:

Técnicas de inteligencia artificial usadas en seguridad informática

Autor: Ing. Alexander Pinilla Rodríguez

Tutor: Ing. Juan Alejandro Devincenzi

Noviembre de 2020

Cohorte 2019

Declaración jurada de origen de los contenidos

“Por medio de la presente, el autor manifiesta conocer y aceptar el Reglamento de Trabajos Finales vigente y se hace responsable que la totalidad de los contenidos del presente documento son originales y de su creación exclusiva, o bien pertenecen a terceros u otras fuentes, que han sido adecuadamente referenciados y cuya inclusión no infringe la legislación Nacional e Internacional de Propiedad Intelectual”.

_____

Alexander Pinilla Rodríguez
Pasaporte: AP907858

Resumen

Este trabajo busca recopilar y comparar algunas de las técnicas de inteligencia artificial (IA) que se vienen implementando en diferentes productos para seguridad en redes y sistemas de cómputo.

Se realizará un comparativo de las técnicas más comunes, así como de las que podrían tener una mayor relevancia en el mercado de la seguridad informática durante los próximos años.

Palabras Clave

Inteligencia artificial, aprendizaje automático, *machine learning, deep learning*, seguridad informática.

Tabla de Contenido

Resumen 3

Palabras Clave 3

Introducción 8

1. Definiciones y conceptos básicos 9

1.1. ¿Qué es e la IA? 9

1.2. Superinteligencia Artificial (ASI): 10

1.3. Inteligencia Artificial Fuerte (o general (AGI)): 10

1.4. Inteligencia artificial débil (o IA estrecha) 12

1.5. Agentes inteligentes 12

2. Machine Learning o aprendizaje automático 14

2.1. Aprendizaje automático supervisado, no supervisado y semisupervisado 15

2.2. Aprendizaje reforzado (Reinforcement learning) 17

2.3. Regresión: 18

2.4. Clasificación: 19

2.5. Arboles de decisión: 20

2.6. Agrupación (clustering): 22

2.7. Detección de anomalías: 22

2.8. Máquina de vectores de soporte (SVM): 23

2.9. Redes neuronales y Deep learning 24

2.10. Algoritmos genéticos 24

2.11. Aprendizaje automático de confrontación 25

3. Inteligencia artificial y seguridad informática 26

3.1. ¿Cómo estamos en Latinoamérica? 30

4. Requerimientos para la seguridad informática 32

5. Tendencias en seguridad informática y uso de IA 34

6. Futuro de inteligencia artificial 36

6.1. Modelo del DARPA 36

6.1.1. Primera ola: Handcrafted Knowledge o Conocimiento Artesanal 36

6.1.2. Segunda ola: aprendizaje estadístico..... 37

6.1.3. Tercera ola: adaptación contextual 37

6.2. Arquitecturas cognitivas 38

6.3. Principio de energía libre e inferencia activa 39

6.4. Inteligencia artificial y computación cuántica..... 40

7. Retos para el gobierno de la seguridad de la información 42

7.1. Retos para la gestión de riesgos:..... 43

8. Conclusión..... 45

9. Glosario:..... 46

10. Bibliografía 48

Tabla de Figuras

Figura 1. PANORAMA GENERAL DE LOS ALGORITMOS DE IA. NILS ACKERMANN. LICENCIADO BAJO CREATIVE COMMONS CC BY-ND 4.0 9

Figura 2 . AGENTES INTELIGENTES DE HARDWARE Y SOFTWARE 13

Figura 3. FLUJO DE TRABAJO DE APRENDIZAJE automático (SUPERVISADO)" DE NILS ACKERMANN. LICENCIADO BAJO CREATIVE Commons CC BY-ND 4.0 15

Figura 4. APRENDIZAJE REFORZADO 17

Figura 5. FUNCIONES DE REGRESIÓN 18

Figura 6. CLASIFICACIÓN 19

Figura 7. CLUSTERING 22

Figura 8. DETECCIÓN DE ANOMALIAS 23

Figura 9. SVM 23

Figura 10. REDES NEURONALES 24

Figura 11. NIVEL PERCIBIDO DE MADUREZ DE LA IA 28

Figura 12. PRINCIPALES APLICACIONES DE LA IA EN SI 29

Figura 13. PRINCIPALES CAPACIDADES DE LA IA PARA SI 33

Figura 14. INTERESES EN IA POR SECTOR 33

Figura 15. MODELO DE COGNICIÓN 39

Figura 16. DARPA- TERCERA OLA DE IA 38

Agradecimientos

A mi familia y amigos
A mis profesores de la maestría

Introducción

Existen diferentes productos de hardware y software que se pueden implementar en un sistema de cómputo para intentar prevenir, identificar e incluso bloquear ataques y, de esta forma, mantener su seguridad.

Sin embargo, los distintos productos de seguridad que se pueden utilizar en una red requieren de una cantidad significativa de capacitación y tiempo para que operen correctamente. Existe, además, un déficit en la cantidad de profesionales capacitados para administrar, operar y mantener estos sistemas de manera satisfactoria.

Por lo tanto, se han venido desarrollando productos que hacen uso de diferentes técnicas de inteligencia artificial (IA) para automatizar procesos de seguridad en las empresas, por ejemplo, detección de malware, detección de intrusiones en la red, detección de intrusiones en host, filtrado de correo, protección de datos, segmentación automatizada de redes, entre otros.

Este trabajo busca realizar un comparativo de las técnicas más comunes utilizadas en la actualidad, así como de las que podrían tener una mayor relevancia en el mercado de la seguridad informática durante los próximos años.

1. Definiciones y conceptos básicos

1.1. ¿Qué es e la IA?

Hay diferentes definiciones de inteligencia artificial (IA). Para entender mejor su relación con la seguridad informática consideramos la IA como una rama dentro de la informática que estudia cómo crear máquinas que posean capacidades similares a la inteligencia humana. Para esto se vale de diferentes técnicas matemáticas, estadísticas y algoritmos.

Para entender mejor el contexto de la IA varios autores clasifican la IA según su nivel de madurez y el propósito para el cual fue creada, así, en una primera clasificación tenemos los conceptos de superinteligencia, inteligencia artificial general e inteligencia artificial estrecha[1]:

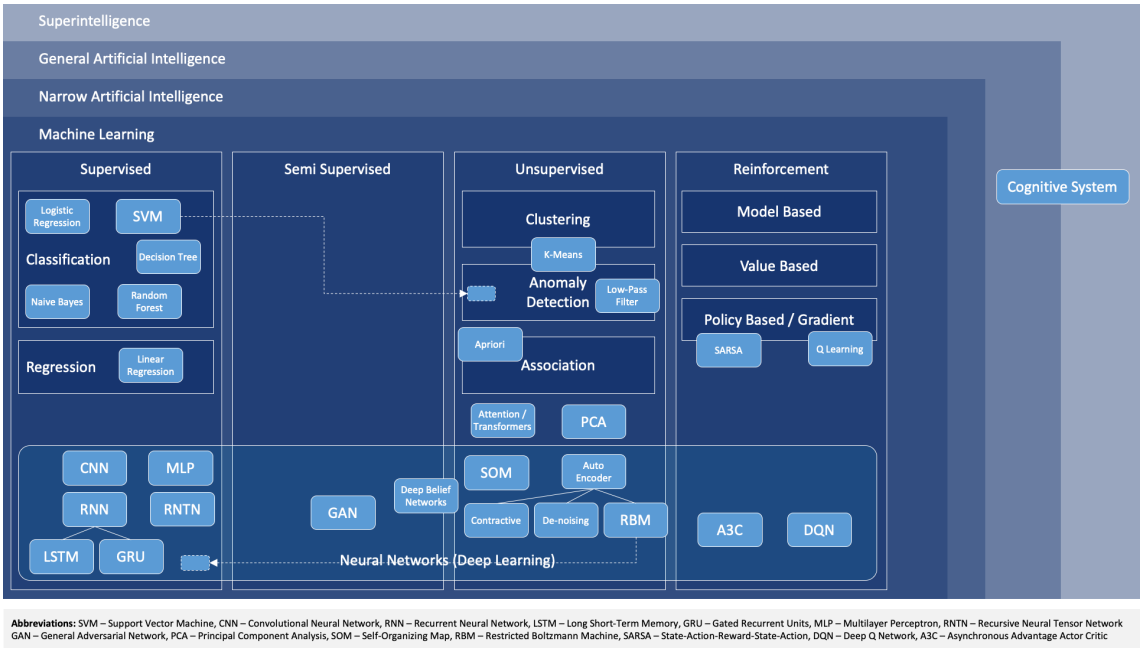

Figura 1. PANORAMA GENERAL DE LOS ALGORITMOS DE IA. NILS ACKERMANN. LICENCIADO BAJO CREATIVE COMMONS CC BY-ND 4.0

[1] https://towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87

1.2. Superinteligencia Artificial (ASI):

La superinteligencia artificial abarca desde una AI que sería un poco más inteligente que un ser humano hasta una que sería muchas veces más inteligente: "un intelecto que es mucho más inteligente que los mejores cerebros humanos en prácticamente todos los campos, incluida la creatividad científica, la sabiduría general y las habilidades sociales"[2].

En la actualidad no existe consenso científico sobre si la ASI tendrá un impacto positivo o negativo sobre los seres humanos, o incluso si algún día se podrá crear.

1.3. Inteligencia Artificial Fuerte (o general (AGI)):

La Inteligencia Artificial General se refiere a una IA que es tan inteligente como un ser humano promedio, es decir que puede realizar cualquier tarea intelectual que un ser humano puede hacer, esto implica, entre otras cosas: razonar, planificar, resolver problemas, pensar de manera abstracta, comprender ideas complejas, aprender rápidamente y aprender de la experiencia. Al menos en teoría, una AGI podría hacer todas esas cosas tan fácilmente como un ser humano.

Adicionalmente, la inteligencia humana está relacionada con la capacidad de un individuo para adaptarse efectivamente y hacer frente a un entorno cambiante. La adaptación significa que el individuo pueda realizar cambios en si mismo, por ejemplo, patrones de comportamiento, y que también pueda cambiar o modificar su entorno[3]. Estas características clave de la inteligencia humana son las que la Inteligencia Artificial General trata de recrear [1]:

Percepción: capacidad de entender el entorno, descomponer los objetos separados que lo componen y comprender las relaciones espaciales entre ellos.

[2] Definición del autor Nick Bostrom. https://towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87
[3] https://www.britannica.com/science/human-intelligence-psychology#ref13341

Aprendizaje: capacidad de alterar el comportamiento basado en experiencias pasadas.

Memoria: capacidad para codificar, almacenar y recuperar experiencias pasadas.

Razonamiento y abstracción: capacidad para llegar a conclusiones lógicas y de generalizarlas, es decir, entender las reglas a partir de las cuales se establece un patrón.

Resolución de problemas: capacidad de encontrar sistemáticamente posibles soluciones y obtener la mejor respuesta a un problema.

Pensamiento divergente/convergente: capacidad de generar múltiples soluciones a un problema dado y de reducir la lista de opciones hasta encontrar la mejor respuesta posible.

Inteligencia emocional: reconoce e interpreta las emociones humanas.

En la actualidad, el enfoque para explicar la inteligencia humana esta en entenderla como combinación de las características expuestas. La inteligencia, en total, ha llegado a ser considerada no como una habilidad única sino como un conjunto de muchas habilidades.

Para que un AGI sea efectiva, las características mencionadas anteriormente deben ocurrir en un período de tiempo razonable, casi en tiempo real, además, requisitos como precisión, capacidad de actualización, interacción con otros sistemas, requieren cantidades significativas de procesamiento. Por todo esto, el desarrollo de una AGI conlleva retos técnicos enormes que con las tecnologías actuales apenas empiezan a ser entendidos y superados, aún así, algunos autores estiman que se dispondrán de sistemas AGI para el año 2045[4] [2].

[4] https://towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87

1.4. Inteligencia artificial débil (o IA estrecha)

La IA estrecha (o débil) se concentra en casos de uso muy específicos. Esto significa que una IA de este tipo está entrenada para un propósito particular. Por lo tanto, la IA solo puede manejar eventos en los que ha sido entrenada. Algunos casos de uso típicos de IA estrecha son:

- Reconocimiento y procesamiento del lenguaje natural
- Conducción autónoma
- Reconocimiento e interpretación de imágenes visuales
- Detección de intrusos en ciberseguridad
- Reconocimiento de la actividad humana

Muchos sistemas actuales de IA estrecha son bastante potentes, pero casi siempre altamente especializados. La mayoría de las aplicaciones que se han hecho famosas en los últimos años corresponden a esta categoría, por ejemplo, sistemas como AlphaGo, que en octubre de 2015 se convirtió en la primera máquina de venció a un jugador profesional de Go[5] en un tablero de 19x19. Este sistema es particularmente bueno en el juego, pero si se le presentara una situación distinta o las reglas de juego cambian (incluso poco) no tendrá un buen desempeño.

1.5. Agentes inteligentes

Un concepto importante que viene tomando relevancia dentro del estudio de la inteligencia artificial es el de agente. Un agente es un sistema que procesa información y produce una salida a partir de una entrada.

En la informática clásica, se emplean principalmente agentes de software (Fig. 2). En este caso, el agente consta de un programa que calcula un resultado a partir de la entrada del usuario:

[5] El juego de Go es considerado extremadamente complejo por la enorme cantidad de combinaciones posibles de jugadas.

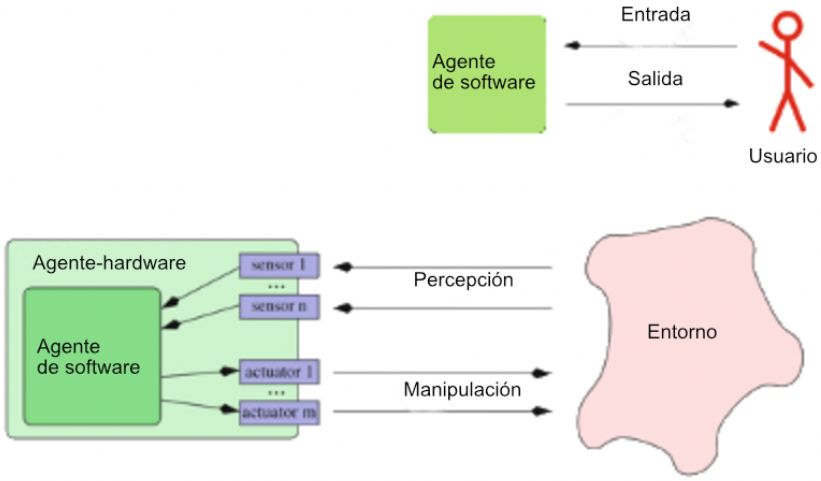

Figura 2 . AGENTES INTELIGENTES DE HARDWARE Y SOFTWARE

También existen los agentes de hardware (también llamados robots autónomos), que cuentan con sensores y actuadores para percibir su entorno, realizar acciones e interactuar con el ambiente que los rodea (Fig. 2).

En principio, es posible que una IA se pueda catalogar en cualquiera de las categorías vistas: como un ASI, AGI o ANI. Los agentes que se expondrán en este documento están dentro de la categoría de ANI.

2. Machine Learning o aprendizaje automático

Machine learning (o aprendizaje automático) hace referencia a un conjunto de técnicas de aprendizaje estadístico para identificar patrones en un conjunto de datos de manera automática. En este sentido, *machine learning* es "el estudio de algoritmos informáticos que mejoran automáticamente a través de la experiencia".[6]

Los conceptos y técnicas usados en *machine learning* integran elementos de las ciencias de la computación, la estadística, ingeniería, psicología, entre otras. El aprendizaje automático ha sido clave para el avance de la Inteligencia Artificial estrecha en los últimos años. En el campo de la seguridad informática se ha desarrollado muy bien porque muchos problemas se pueden estudiar y atacar con una gran precisión. Además, es posible disponer de grandes volúmenes de datos, una condición necesaria para que los algoritmos de *machine learning* operen de manera correcta.

En seguridad informática es posible clasificar los casos de uso de los algoritmos de *machine learning* en dos categorías: reconocimiento de patrones y detección de anomalías. En el reconocimiento de patrones, se busca identificar características explícitas u ocultas en un conjunto de datos. Estas características se pueden utilizar para enseñar a un algoritmo a reconocer estas características en otro conjunto de datos. Por otra parte, la detección de anomalías busca establecer una noción de normalidad que describa a la mayoría de los elementos de un conjunto de datos determinado. A partir de ahí cualquier tipo de desviación de esta normalidad se detectará como anomalía.

Como se mencionó previamente, el aprendizaje automático se basa en modelos estadísticos para realizar predicciones. Durante el proceso de aprendizaje, los parámetros del modelo se adaptan de manera óptima de acuerdo con los datos de entrenamiento proporcionados. Se puede decir que el sistema (o agente) aprende por

[6] Machine Learning is the study of computer algorithms that improve automatically through experience. T. Mitchell. Machine Learning. McGraw Hill, 1997. http://www-2.cs.cmu.edu/tom/mlbook.html

experiencia si mejora su rendimiento en datos nuevos y desconocidos a lo largo del tiempo, es decir, después de que haya visto muchos ejemplos de entrenamiento.

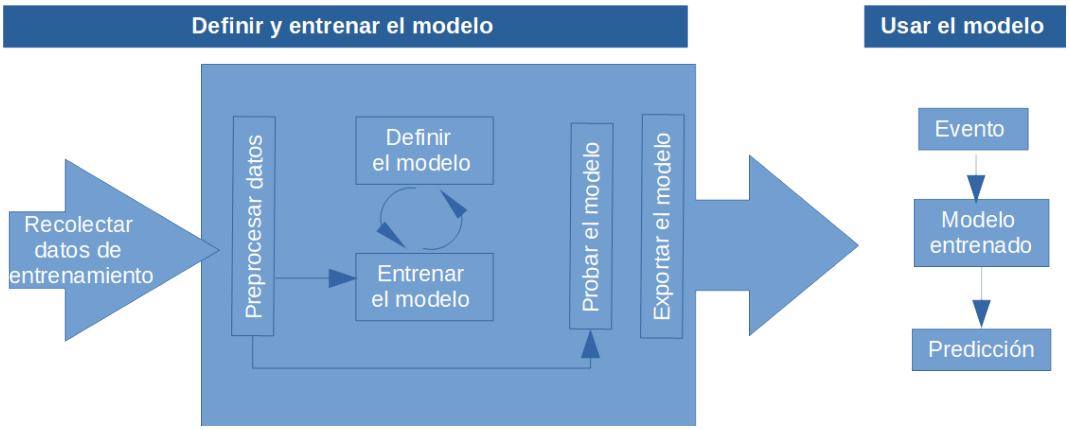

Figura 3. FLUJO DE TRABAJO DE APRENDIZAJE automático (SUPERVISADO)" DE NILS ACKERMANN. LICENCIADO BAJO CREATIVE Commons CC BY-ND 4.0

Para que un modelo se generalice bien para eventos no vistos es importante definir un algoritmo que se adapte bien al problema que debe solucionar, y tener datos de capacitación disponibles en la cantidad y calidad necesarias. Es por esto por lo que los sistemas de aprendizaje automático son extremadamente buenos solucionando los problemas específicos para los que fueron programados, pero inevitablemente fallarán si tratan de solucionar problemas nuevos o con situaciones que estén fuera de los rangos de operación con los que fueron diseñados.

En *machine learning* existen tres paradigmas: aprendizaje supervisado, aprendizaje no supervisado y el aprendizaje reforzado.

2.1. Aprendizaje automático supervisado, no supervisado y semisupervisado

De manera general los métodos de aprendizaje automático supervisado (*supervised machine learning)* adoptan un enfoque bayesiano para el descubrimiento de conocimiento, para esto utilizan las probabilidades de los eventos observados con anterioridad para inferir las probabilidades de los nuevos eventos, para esto es necesario alimentar al algoritmo datos de entrenamiento totalmente etiquetados, es

decir que estos conjuntos de datos ya han sido revisados y validados por un operador humano y se conocen completamente las características que los identifican.

En los métodos no supervisados (*unsupervised machine learning)* se extraen los parámetros que se necesitan de datos sin etiquetar y se aplican a datos o eventos nuevos. Ambos tipos de métodos se pueden aplicar a problemas de clasificación (asignación de elementos a categorías) o regresión (predicción de propiedades numéricas de una observación) [3].

En los casos en que el etiquetado de cantidades masivas de datos de entrenamiento es demasiado costoso, se utiliza una mezcla entre aprendizaje automático supervisado y no supervisado. El proceso comienza con el etiquetado de una cantidad razonable de datos. A partir de ahí se pueden usar varias técnicas para extrapolar el comportamiento de datos nuevos. El ejemplo más popular de aprendizaje semisupervisado es un tipo especial de red neuronal llamada GAN (red general de confrontación), en la cual se utilizan datos de entrenamiento etiquetados para generar nuevos datos (a través de la red del generador) y luego se envían estos datos a otra red neuronal (discriminador). Las redes generadoras y discriminadoras mejoran conjuntamente en un ciclo de retroalimentación positiva. El generador mejorará al crear datos falsos convincentes y el discriminador está mejorando al separar los datos falsos de los datos de entrenamiento.

Para aprendizaje automático supervisado (*supervised machine learning)* se puede emplear cualquier algoritmo de aproximación de funciones:

- Clasificación
- Árboles de decisión
- Random forest (Bosques Aleatorios
- Vecino más cercano
- Regresión
- Aprendizaje basado en casos
- Aprendizaje de redes bayesianas (Naive Bayes)
- Redes neuronales
- Procesos gaussianos

- Máquinas de vectores de soporte (SVM - kernel)
- Wavelets, splines, funciones de base radial

Para aprendizaje automático no supervisado (*unsupervised machine learning:*

- Clustering (agrupamiento)
- Máquinas vectoriales de apoyo
- Algoritmo vecino más cercano
- Algoritmo vecino más lejano
- k-medias
- Redes neuronales

2.2. Aprendizaje reforzado (Reinforcement learning)

El aprendizaje reforzado se basa en la suposición de que un comportamiento o una acción óptimos pueden imponerse mediante recompensas positivas que se otorgan por una acción favorable. La configuración básica para el aprendizaje de refuerzo comprende un agente que interactuará con su entorno. En base a la retroalimentación positiva / negativa que el agente recibe de sus interacciones en una situación dada, se recompensa un determinado comportamiento y, por lo tanto, se refuerzan las acciones favorables. El aprendizaje profundo es otro término popular que comúnmente se confunde con el aprendizaje automático.

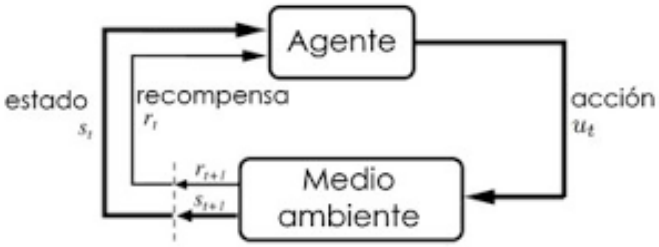

Figura 4. APRENDIZAJE REFORZADO[7]

Alguno de los algoritmos mas usados para aprendizaje reforzado (*reinforcement machine learning)* son*:*

[7] https://commons.wikimedia.org/wiki/File:Markov_diagram_v2.svg

- Iteración de valor
- Q aprendizaje
- aprendizaje de TD
- Métodos de gradiente
- Redes neuronales

Debido a que este trabajo se centra en las aplicaciones prácticas de la inteligencia artificial en la seguridad informática, no se hará énfasis en el detalle matemático de cada algoritmo. Si se desear consultar en mas detalle se recomienda el libro *Introduction to Artificial Intelligence Second Edition* de Wolfgang Ertel [4].

A continuación, se describen brevemente algunos de los algoritmos y técnicas mas usados en *machine learning.* En algunos casos estos algoritmos se pueden usar en mas de una tarea, por ejemplo, los arboles de decisión de pueden usar en tareas de clasificación o de agrupación de datos:

2.3. Regresión:

Algoritmo que predice un valor numérico de respuesta para una entrada dada de manera continua.

El algoritmo toma datos que consisten en variables de entrada (expresadas como vectores en un espacio vectorial) y una variable de respuesta (un número real) y produce un modelo lineal de "mejor ajuste", que asigna cada punto del espacio vectorial a su respuesta predicha.

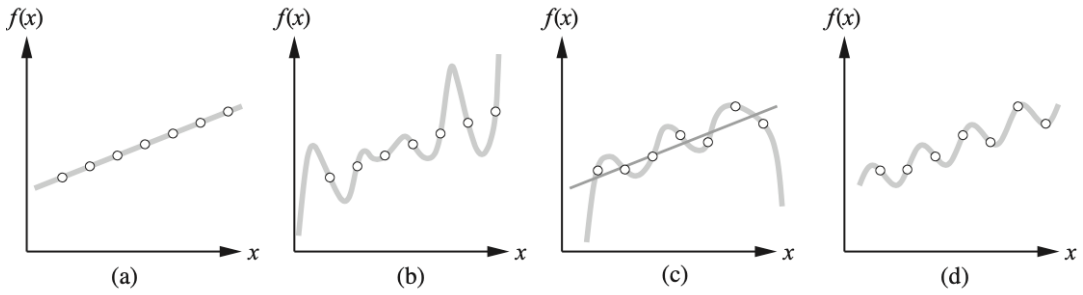

Figura 5. FUNCIONES DE REGRESIÓN[8]

[8] Artificial Intelligence A Modern Approach, 3rd Edition by Stuart J. Russell, Peter Norvig

Usualmente, en aplicaciones *machine learning* se utiliza la regresión logística

2.4. Clasificación:

Un algoritmo que predice el valor de respuesta mediante decisiones binarias sobre la naturaleza del evento observado. El algoritmo predecirá un valor discreto y segmentará los valores de respuesta en clases distintas. Estos algoritmos también se denominan clasificadores.

En seguridad informática muchos de los eventos que ocurren en una red o sistema deben ser clasificados como legítimos o maliciosos, por ejemplo: un archivo enviado a través de la red, ¿contiene malware? Por cada intento de inicio de sesión, ¿es un inicio de sesión válido? Por cada correo electrónico recibido, ¿es un intento de phishing? Para cada solicitud a un servidor, ¿es un ataque de denegación de servicio (DoS)? Para cada solicitud saliente de la red, ¿es un bot que llama a su servidor de comando y control? [3].

La clasificación puede ser binaria, por ejemplo, entre un evento malicioso o uno legitimo, o multiclase, por ejemplo, si desea determinar si una pieza de malware es *ransomware*, un *keylogger* o un troyano de acceso remoto (RAT)[9].

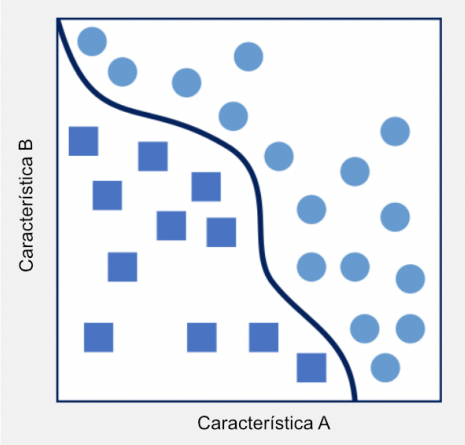

Figura 6. CLASIFICACIÓN[10]

[9] Remote Access Trojan, por su sigla en inglés
[10] "Supervised machine learning" by Nils Ackermann is licensed under Creative Commons CC BY-ND 4.0 [2]

2.5. Arboles de decisión:

Son modelos de aprendizaje supervisado muy versátiles que tienen la importante propiedad de ser fáciles de interpretar, por lo tanto, es fácil explicar las respuestas producidas por el algoritmo.

Los árboles toman una decisión de forma binaria. Empezando en la raíz del árbol, el conjunto de datos se divide en función de una condición binaria en dos subconjuntos secundarios. Por ejemplo, si la condición es "PRECIO ≥ 1000 USD", todos los datos para los que esta condición es verdadera van a la rama izquierda y todos datos para los que esta condición es falsa van a la rama derecha. Luego, los subconjuntos secundarios se dividen de forma recursiva en subconjuntos más pequeños en función de otras condiciones. Las condiciones de división se seleccionan automáticamente en cada paso en función de qué condición divide mejor el conjunto de elementos. Finalmente, el algoritmo genera una estructura de árbol en la que cada nodo representa una decisión binaria, las ramas de cada nodo representan los dos posibles resultados de esa decisión y cada decisión representa la clasificación de los puntos de datos que siguen el camino desde la raíz hasta ese punto.

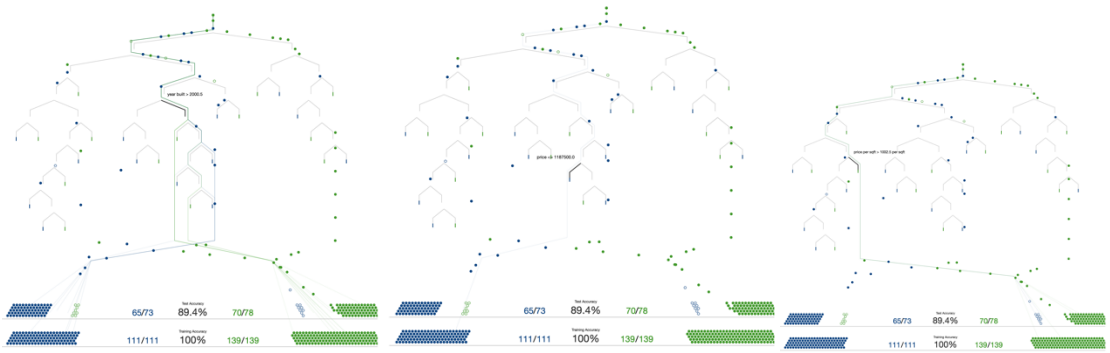

Figura 6. ÁRBOL DE DECISIÓN[11]

Los árboles son una forma muy intuitiva de mostrar y analizar datos. Son populares incluso fuera del campo de aprendizaje automático. Además, se pueden usar para

[11] Tomado de http://www.r2d3.us/visual-intro-to-machine-learning-part-1/. En este ejemplo el árbol clasifica bienes inmuebles en las ciudades de Nueva York y San Francisco y muestra diferentes rutas que se pueden tomar en el árbol para llegar a un resultado.

predecir tanto categorías (árboles de clasificación) como valores reales (árboles de regresión).

Una extensión de los arboles de decisión son los bosques de decisión o *Decision Forests*. En aprendizaje automático, la combinación de varios algoritmos y clasificadores independientes en un conjunto (también conocido como apilamiento) es una forma común de aprovechar las ventajas de cada método, básicamente lo que se hace es una combinación de múltiples clasificadores que crean un clasificador más complejo y, a menudo, de mejor rendimiento. La combinación de árboles de decisión para formar bosques de decisión es una técnica probada para crear clasificadores de alta calidad.

Los dos tipos más comunes de bosques que se utilizan en la práctica son los bosques de decisión aleatorios (*ramdom forest*) y los árboles de decisión potenciados por gradientes (*gradient-boosted decision trees*).

Los bosques aleatorios son agrupaciones de árboles de decisión que se agrupan colectivamente para hacer predicciones.

Los árboles de decisión potenciados por gradientes (GBDT) utilizan combinaciones más inteligentes de árboles de decisión para producir mejores predicciones. En el refuerzo de gradientes, se combinan de forma selectiva varios arboles débiles usando una técnica de optimización por descenso de gradientes en la función de pérdida [3]. La técnica básica consiste en agregar árboles individuales al bosque, de a uno a la vez, utilizando un procedimiento de descenso de gradiente para minimizar la pérdida al agregar árboles. La adición de más árboles al bosque se detiene cuando se alcanza un límite fijo, cuando la pérdida del conjunto de validación alcanza un nivel aceptable o cuando la adición de más árboles ya no mejora el desempeño del algoritmo.

2.6. Agrupación (clustering):

Algoritmo que intenta crear agrupaciones de puntos de datos similares. Las técnicas de agrupación utilizan una serie de parámetros de datos y pasan por varias iteraciones antes de poder agrupar los datos.

Una aplicación común es determinar que tipo de solicitudes tiene un sitio web. En este caso, el trafico se puede caracterizar en grupos, por ejemplo, las solicitudes que provienen de bots, las que provienen de dispositivos móviles y las de usuarios legítimos etc.

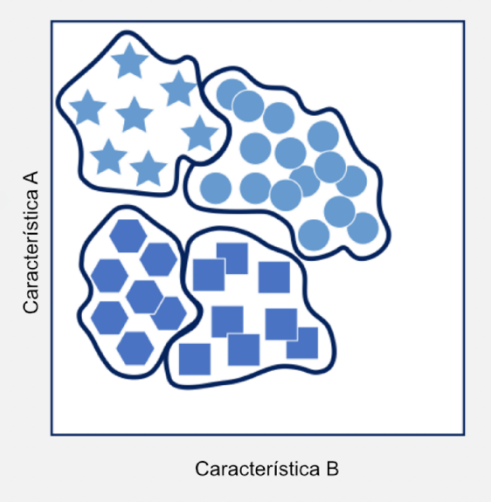

Figura 7. CLUSTERING[12]

2.7. Detección de anomalías:

Algoritmo que intenta encontrar valores atípicos (anomalías) dentro de un conjunto de datos. Un ejemplo típico es la identificación del comportamiento fraudulento en las transacciones bancarias. Si una persona realiza muchas transacciones mientras viaja entre Europa y los Estados Unidos, con cantidades relativamente pequeñas, una transacción de USD 10,000 desde Nigeria podría ser sospechosa.

[12] https://towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87

La detección de anomalías se puede considerar como un nivel adicional de un algoritmo de regresión. Matemáticamente se interpreta como el problema de determinar cuándo un valor observado es lo suficientemente diferente de un valor predicho para indicar que algo inusual está sucediendo:

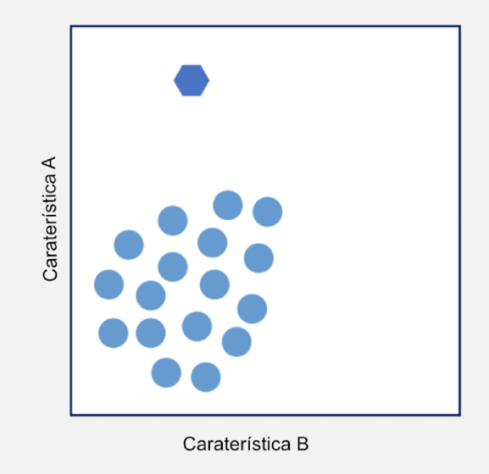

Figura 8. DETECCIÓN DE ANOMALIAS[13]

2.8. Máquina de vectores de soporte (SVM):

Es un algoritmo clasificador lineal que produce un hiperplano en un espacio vectorial para separar las clases en un conjunto de datos. El algoritmo opera implementando una función de decisión que intenta dividir el subconjunto de entrenamiento en dos niveles de clases o categorías. Al seleccionar la función de decisión, se asegura que se maximice la distancia a los puntos más cercanos en los datos de entrenamiento a cada lado del hiperplano.

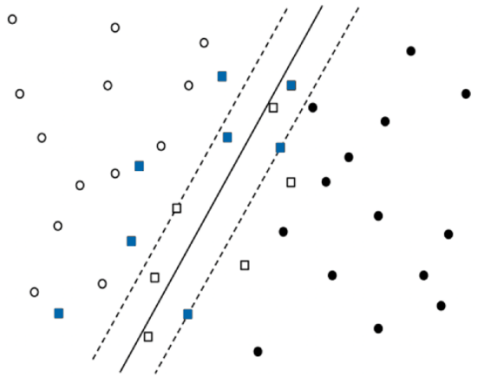

Figura 9. SVM

[13] https://towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87

2.9. Redes neuronales y Deep learning

Una red neuronal es un método de aprendizaje automático inspirado en la biología. Intenta modelar, simular, e incluso reconstruir en hardware el funcionamiento del cerebro humano mediante la aplicación de modelos multicapa que utilizan capas de componentes estadísticos más simples para aprender representaciones de datos.

Uno de los usos mas comunes de las redes neuronales, y por extensión de los algoritmos de *deep learning*, es la identificación de patrones muy complejos dentro de datos en los cuales otros algoritmos de clasificación mas simples como SVM o regresiones logísticas no son eficientes.

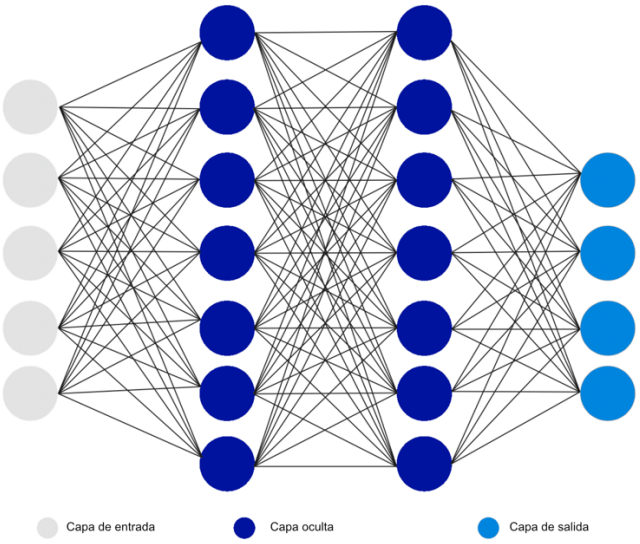

Figura 10. REDES NEURONALES

2.10. Algoritmos genéticos

Los algoritmos genéticos imitan el proceso de evolución fisiológica de los seres humanos y utilizan estos conocimientos para resolver problemas. Aplican métodos propios de la metaheurística para encontrar, generar o seleccionar algoritmos de búsqueda que pueden proporcionar una solución muy buena para un problema.

Aplicaciones recientes de los algoritmos genéticos permiten seleccionar diferentes alternativas de solución a un problema de manera automática, rápida y altamente

eficiente. Por ejemplo, Geneva [14] es una herramienta de IA desarrollada por investigadores de la Universidad de Maryland que utiliza un algoritmo genético para crear estrategias de evasión de censura. A la fecha ha encontrado varios errores previamente desconocidos en los sistemas de censura de países como China, India y Kazajstán. Las estrategias de Geneva manipulan el trafico de la red para confundir a los sistemas del censor sin afectar la comunicación entre el cliente servidor[15]. Esto hace que sea eficaz contra muchos tipos de censura [5].

Al igual que los sistemas biológicos, este sistema forma conjuntos de instrucciones a partir de componentes genéticos. Pero en lugar de utilizar cadenas de ADN como bloques de construcción, utiliza pequeños fragmentos de código que realizan instrucciones muy especificas, por ejemplo, SYN y ACK en el protocolo TCP[16]. Individualmente, las partes de código hacen muy poco, pero cuando se combinan en funciones, pueden realizar sofisticadas estrategias de evasión para dividir, ordenar o enviar paquetes de datos.

La herramienta evoluciona su código genético a través de sucesivos intentos (o generaciones). Con cada generación mantiene las instrucciones que funcionan mejor para evadir la censura y expulsa al resto. Luego muta y cruza sus estrategias eliminando instrucciones al azar, agregando nuevas instrucciones o combinando instrucciones exitosas y probando nuevamente. A través de este proceso evolutivo puede identificar múltiples estrategias de evasión novedosas y eficientes.

2.11. Aprendizaje automático de confrontación

Adversarial Machine Learning o aprendizaje automático de confrontación, es el estudio de las vulnerabilidades de los algoritmos de *machine learning* desde la perspectiva del atacante, con el fin de para hacerlos más robustos a los ataques o

[14] https://geneva.cs.umd.edu/
[15] Geneva's strategies manipulate the network stream to confuse the censor without impacting the client/server communication. This makes Geneva effective against many types of in-network censorship (though it cannot be used against IP-blocking censorship), traducido de https://github.com/Kkevsterrr/geneva
[16] El código fuente completo está disponible en https://github.com/Kkevsterrr/geneva

para reducir los errores que presentan cuando entran en producción. El aprendizaje automático de confrontación ha venido tomando mayor relevancia a medida que los sistemas que utilizan alguna forma de ma*chine learning* se han vuelto mas comunes en las empresas.

Si bien es cierto que un porcentaje considerable de empresas planean implementar sistemas de IA en los próximos años, entre los que profesionales que se dedican a la investigación en seguridad informática hay un porcentaje muy alto de escepticismo sobre la madurez de los sistemas de IA. Un estudio elaborado en 2017 por *VMware Carbon Black* entrevistó[17] a 410 investigadores expertos en ciberseguridad y encontró que cerca de tres cuartas partes (74%) dijeron que las soluciones de seguridad impulsadas por la inteligencia artificial todavía tienen fallas. El 70% de los investigadores de seguridad dijeron que los atacantes pueden eludir las tecnologías de seguridad impulsadas por ML; y casi un tercio (30%) dijo que las soluciones de seguridad impulsadas por ML son fáciles de eludir [6].

3. Inteligencia artificial y seguridad informática

Las principales aplicaciones de los algoritmos y técnicas expuestas previamente en la seguridad informática se dan en los siguientes campos:

- Manejo de identidades:
  - Autenticación y control de acceso
  - Detección de identidades comprometidas
- Aplicaciones en ciberdefensa:
  - Detección de malware
  - Detección de intrusiones basada en red
  - Detección de intrusiones basada en host

[17] Los datos se tomaron de https://www.carbonblack.com/blog/beyond-hype-security-experts-weigh-artificial-intelligence-machine-learning-non-malware-attacks/

  - Análisis de seguridad (correlación de eventos)
  - Respuesta a incidentes

- Seguridad de correos electrónicos:

  - Filtrado de spam
  - Detección de *spearphishing*

- Evaluación de riesgos:

  - Evaluación de riesgo cibernético
  - Inteligencia de amenazas

- Protección de datos:

  - Insider Threat Detection (Data loss/disclosure prevention)
  - Clasificación de datos
  - Verificación de integridad de datos

- Automatización:

  - Segmentación de red automatizada
  - Seguridad y seguridad definidas por SW
  - Orquestación

- Detección de fraude y abuso

En el estudio *Security Gets Smart with AI*, realizado por el *SANS Institute* en 2019 entre profesionales de la Seguridad Informática [7], se encontró que el 57% de los encuestados está utilizando o planeando utilizar soluciones de seguridad que utilizan IA, aunque solo el 35% tiene experiencia directa con dichas plataformas. Por otra

parte[18], el 46% de los entrevistados considera que las soluciones de seguridad basadas en inteligencia artificial están "madurando", pero que aún no cuentan con la suficiente madurez tecnológica para ser implementadas en sus negocios. Esto concuerda con el estudio de *Wmware Carbon black* citado en la sección anterior sobre aprendizaje automático de confrontación:

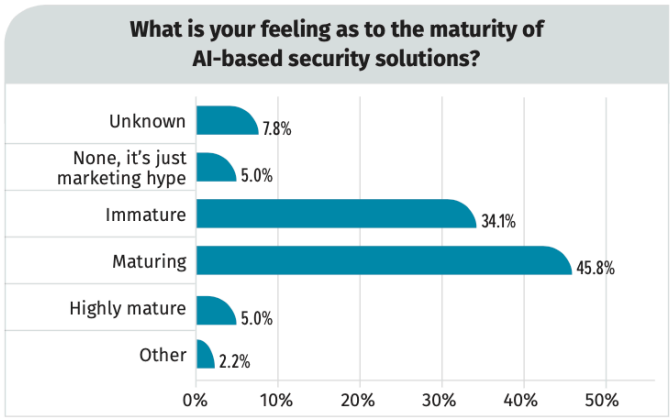

Figura 11. NIVEL PERCIBIDO DE MADUREZ DE LA IA

Esto es bastante interesante si se considera que inteligencia artificial lleva mas de 50 años siendo un campo de estudio dentro las ciencias de la computación. Sin embargo, no fue sino hasta alrededor del año 2010, después de varias décadas de investigación sobre aprendizaje automático y redes neuronales, que se empezaron a comercializar de manera masiva aplicaciones que hacen usos de diferentes formas de IA, de ahí que la percepción de falta de madurez de estos productos esté relacionada con su poco tiempo de ingreso al mercado.

El estudio del *SANS Institute* también indica las áreas en las que las empresas prevén aplicar IA para defender sus sistemas:

- Defensa cibernética (75,2%)
- Prevención de malware (70,5%)
- Detección/prevención avanzada de amenazas (68,6%)
- Detección de amenazas distintas de malware (44,8%)
- Respuesta automática de incidentes (40%)

[18] El documento original está en inglés y disponible en: https://www.sans.org/press/announcement/2019/03/14/1

- Prevención de phising (32,4%)
- Protección de redes IoT (31,4%)
- Analítica de comportamiento de usuario (31,4%)
- Respuesta a ransomware (29,5%)
- Automatización de procesos (26,7%)

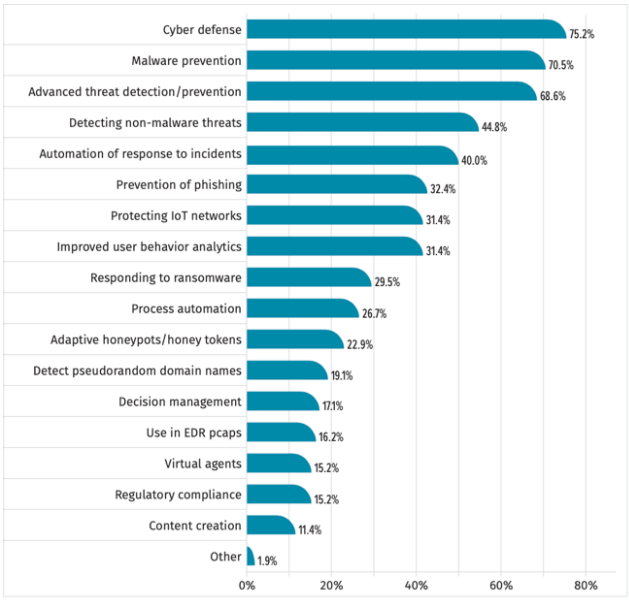

Figura 12. PRINCIPALES APLICACIONES DE LA IA EN SI

Otro estudio, elaborado en 2019 por la consultora Capgemini[19], entrevistó a 850 ejecutivos de diferentes industrias, incluyendo productos de consumo, comercio minorista, banca, seguros, automoción, servicios públicos y telecomunicaciones. Este reporte encontró que el 69% de las empresas creen que la IA será necesaria para responder a los ciberataques, sin embargo, tan solo una de cada cinco organizaciones utilizó sistemas de IA antes de 2019 [8].

De acuerdo con este estudio, el 63% de las organizaciones encuestadas planea implementar IA en 2020 para mejorar la ciberseguridad y cerca de la mitad (48%) dijo que los presupuestos para IA en ciberseguridad aumentarían en promedio el 29% para el año 2020. El aumento promedio en los presupuestos para año fiscal 2020 en una de cada diez organizaciones estaría cerca al 40% con respecto presupuestos del año fiscal 2019.

[19] El documento original está disponible en https://www.capgemini.com/research/reinventing-cybersecurity-with-artificial-intelligence/

Si bien ambos estudios prevén que la adopción de IA está lista para dispararse, con casi dos de cada tres organizaciones planeando emplear IA para 2020, aún no es claro como la pandemia desatada por el virus del COVID-19 haya afectado los planes de las empresas para la implementación de estas tecnologías.

3.1. ¿Cómo estamos en Latinoamérica?

Las tendencias presentadas en los estudios del SANS Institute y la consultora Capgemini se basan en entrevistas efectuadas entre ejecutivos de empresas de Alemania, Australia, España, Estados Unidos, Francia, Italia, Países Bajos, Reino Unido, Suecia e India. Nada raro si consideramos que en la mayoría de estos países existe una cultura de desarrollo y adopción tecnológica bastante establecida. Por otra parte, los países de América Latina se encuentran en el grupo de aquellos que se atreven a incorporar nuevos desarrollos, pero a un ritmo mas lento, por lo tanto, algunas de las tecnologías y aplicaciones que se han presentado en las secciones previas no tienen la misma velocidad de adopción que en Europa, Asia o Norteamérica.

Si bien aún es difícil encontrar en la región un gran número de empresas que hagan uso de la IA en sus sistemas, según el estudio *La inteligencia artificial en las compañías latinoamericanas*[20]*,* realizado por la consultora Everis y la publicación *MIT Tech Review,* alrededor de un 47%[21] de las empresas han aplicado la IA en algunos de sus procesos o departamentos, especialmente en proyectos de eficiencia operativa y atención al cliente en sectores como el financiero o telecomunicaciones.

El estudio resalta que la adopción en Latinoamérica es incipiente. Al respecto *MIT tech review* dice que: “Al igual que en otras regiones del mundo, la brecha entre los sectores más avanzados y los más rezagados amenaza con ser cada vez más grande. Esto puede causar que los sectores más lentos en la implementación de IA

[20] https://www.technologyreview.es/s/12427/el-uso-de-la-ia-en-las-grandes-organizaciones-en-latinoamerica

[21] MIT tech review entrevisto más de 40 directivos y a más de 100 personas responsables de liderar la adopción de IA en sus empresas. Comparado con el estudio de Capgemini (850 personas), la muestra es muy baja.

tarden más en adquirir datos, talento y competencias clave, ralentizando así el propio desarrollo de la tecnología en la empresa"[22].

Brasil y Argentina lideran la adopción de tecnologías basadas en IA. Según Microsoft[23], en Argentina a 5 de cada 10 Pymes les interesa invertir en Inteligencia Artificial. En Colombia, las empresas ven que el uso de la IA es más una necesidad que una opción en el medio plazo.

Sin embargo, los estudios citados coindicen en que la seguridad informática no es una prioridad para la inversión de IA en la región. Aunque la mayoría de las empresas reconocen el potencial de estas tecnologías, proyectos de eficiencia de operaciones, atención al cliente, reducción de costes, predicción de ventas y el cálculo avanzado de riesgos son priorizados.

[22] https://www.technologyreview.es/s/12427/el-uso-de-la-ia-en-las-grandes-organizaciones-en-latinoamerica

[23] https://www.argencon.org/microsoft-en-argentina-a-5-de-cada-10-pymes-les-interesa-invertir-en-inteligencia-artificial/

4. Requerimientos para la seguridad informática

Las características que deben tener los sistemas de IA que se utilicen para seguridad informática se pueden dividir en dos grupos: por un lado, las relacionadas con el funcionamiento del sistema, es decir, la función para la cual fue creado, y por otro, las relacionadas con el diseño y la operación segura del sistema en si mismo.

Para analizar las características o requerimientos de funcionamiento del un sistema IA, se pueden retomar los estudios de SANS, Capgemini, la revista Forbes[24] y Bloomberg[25], los cuales concuerdan en que los sistemas de seguridad informática que incorporen inteligencia artificial deben contar con algunas características funcionales avanzadas como, propias de una Inteligencia Artificial Fuerte (o general (AGI)) [ver sección 1.3]. Algunas de estas características deseables son [6]:

• Aprender: que el sistema detecte y reaccione a medida que cambian los datos.

• Descubrir: que el sistema aprenda de los datos sin intervención humana.

• Predecir: que el sistema proporcione alternativas sobre lo que sucederá en el futuro, basándose en la comprensión autónoma de los conjuntos de datos que ha adquirido.
• Justificar: que el sistema permita explicar las acciones tomadas, haciendo que los resultados sean reconocibles y creíbles. Como veremos mas adelante, esto es particularmente importante cuando hablamos de regulación y cumplimiento (*compliance)*.

• Responder: ante un evento de seguridad el sistema produzca una salida acorde con el tipo de ataque.

[24] El artículo original esta disponible en: https://www.bloomberg.com/professional/blog/five-components-artificial-intelligence-must-succeed/

[25] https://www.forbes.com/sites/louiscolumbus/2019/11/24/10-predictions-how-ai-will-improve-cybersecurity-in-2020/#32e3963e6dd7

Los requerimientos de un sistema IA también varían de acuerdo con el nivel (técnico o gerencial) dentro de una organización, sin embargo, algunas características como descubrir, actuar y predecir son consideradas fundamentales [9]:

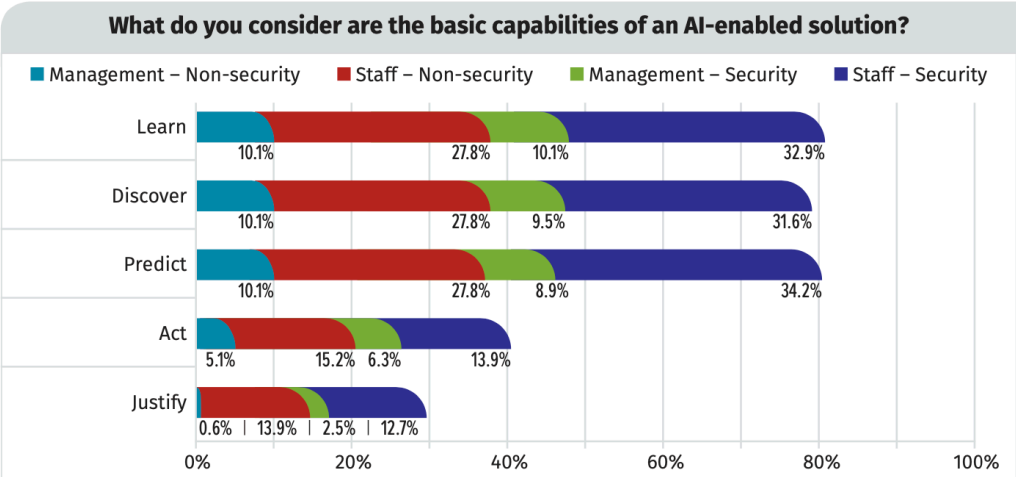

Figura 13. PRINCIPALES CAPACIDADES DE LA IA PARA SI

Los intereses sobre las capacidades que deben tener los sistemas de IA también varían de acuerdo con la industria [7]:

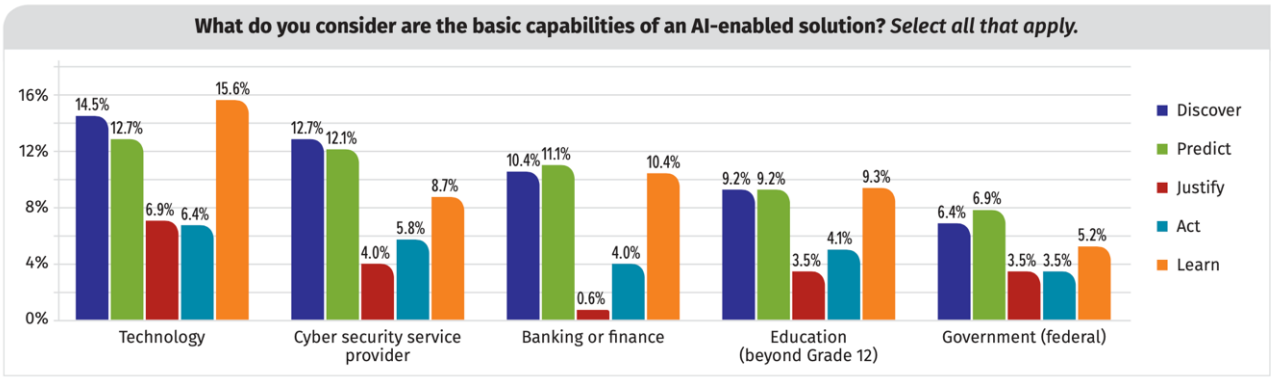

Figura 14. INTERESES EN IA POR SECTOR[26]

En los sectores de tecnología, proveedores de servicios de ciberseguridad y banca, las capacidades para descubrir y predecir son especialmente relevantes. Es decir que "'la forma en que una organización percibe el mejor uso de la IA para mejorar su seguridad inevitablemente esta determinada por la misión y las capacidades existentes de esa organización"[27]

[26] https://www.sans.org/press/announcement/2019/03/14/1

[27] Security Gets Smart with AI página 8: Bottom line: How an organization perceives the best use of AI in improving its security posture will inevitably be driven by the mission and existing capabilities of that organization.

5. Tendencias en seguridad informática y uso de IA

Los estudios citados de SANS y Capgemini, así como la revista Forbes[28], concuerdan en que la inteligencia artificial continuará integrándose en las organizaciones, no solo en para la seguridad informática, sino también en la parte de gestión de TI. Algunas de las tendencias que darán en el mercado de las soluciones de IA son [10]:

- Se incrementarán la cantidad de productos para administración de TI que hagan uso de IA. Esto impactará directamente en los resultados de seguridad, permitiendo la automatización de tareas como: identificación de dispositivos en una red, identificación de procesos que se están ejecutando, medición y control del consumo de ancho de banda, actualización de equipos y protección de terminales de usuario, entre otras.

- Las soluciones de inteligencia artificial y técnicas de minería de datos (*data mining),* permitirán el análisis y correlación de datos diversos y complejos, mejorando la predicción y respuesta a eventos de ciberseguridad.

- Los actores maliciosos incrementarán el uso de IA para analizar los mecanismos de defensa y evadir controles de seguridad. Es así como se prevé que aparezcan ataques nuevos y sofisticados, en especial por parte de atacantes que disponen de recursos y financiamiento, como son gobiernos y organizaciones criminales altamente profesionales, que aprovecharán la inteligencia artificial para explotar vulnerabilidades y brechas de seguridad para obtener acceso a sistemas y datos valiosos.

- Dada la escasez de recursos humanos con experiencia y el gran volumen de datos con el que la mayoría de las organizaciones están tratando de trabajar, muchas organizaciones buscarán capacidades de IA para automatizar sus procesos de operaciones de seguridad informática.

[28] El artículo completo está disponible online en: https://www.forbes.com/sites/louiscolumbus/2019/11/24/10-predictions-how-ai-will-improve-cybersecurity-in-2020/#210c81a36dd7

- Debido a que el teletrabajo se incrementó durante todo el año 2020, el uso de inteligencia artificial se volverá más frecuente en los controles de fraude. Es previsible un incremento en las tecnologías de IA que analicen patrones de comportamiento de los usuarios, desde sistemas de control de acceso hasta patrones de comportamiento en la web.

6. Futuro de inteligencia artificial

Los avances en los productos para seguridad informática que usan IA se han dado principalmente mediante la aplicación de técnicas de *machine learning* y *deep learning*. Sin embargo, estas técnicas presentan limitaciones inherentes que dificultan el desarrollo de muchas características deseables expuestas en la sección anterior y que corresponden con capacidades avanzadas propias de una Inteligencia Artificial General (AGI) [ver sección 1.3].

Una perspectiva interesante para referenciar las tendencias y avances en IA es la que usa el DARPA (*Defense Advanced Research Projects Agency*)[29] de los Estados Unidos. En este modelo se plantean tres olas o tendencias así[30]:

6.1. Modelo del DARPA

6.1.1. Primera ola: Handcrafted Knowledge o Conocimiento Artesanal

Los sistemas de esta generación están en capacidad de razonar sobre problemas muy definidos, mas no cuentan con capacidad de aprendizaje y, si bien tienen alguna capacidad de predicción, tienen un muy mal manejo de la incertidumbre.

En esta primera ola, se crearon conjuntos de reglas para presentar el conocimiento en dominios bien definido y son quienes programan los algoritmos quienes definen la estructura del conocimiento o la inteligencia que tiene el sistema. Ejemplo de esta ola serían los sistemas basados en lógica difusa.

[29] La Agencia de Proyectos de Investigación Avanzados de Defensa, más conocida por su acrónimo DARPA, por su sigla en inglés Defense Advanced Research Projects Agency, es una agencia del Departamento de Defensa de Estados Unidos responsable del desarrollo de nuevas tecnologías para uso militar.

[30] El documento original se encuentra en: https://www.darpa.mil/about-us/darpa-perspective-on-ai

6.1.2. Segunda ola: aprendizaje estadístico.

Los sistemas de esta ola están en capacidad de resolver problemas muy específicos, para lo cual se crean modelos estadísticos y se entrenan con grandes cantidades de datos (*big data*). Las aplicaciones de esta ola cuentan con capacidades de clasificación y predicción, mas no cuentan con capacidad de entendimiento contextual y, aunque podrían mostrar algún nivel de razonamiento, este es muy limitado.

Las aplicaciones que cumplen con el requisito de inteligencia se basan en redes neuronales y aprendizaje automático (*machine learning*). Si bien esta tecnología es extremadamente poderosa en muchas aplicaciones actuales, es bastante inadecuada para resolver situaciones interactivas en la que el agente de IA no ha sido entrenado. Los sistemas de esta ola no se adecuan bien a problemas cambiantes o entornos nuevos y, para adecuarse de manera apropiada, requieren cantidades importantes de procesamiento de datos, por ejemplo, los vehículos autónomos requieren procesar enormes cantidades de datos provenientes de todo tipo de sensores con el fin conducir de manera segura.

6.1.3. Tercera ola: adaptación contextual

Los sistemas de esta ola están basados en modelos contextuales explicativos para fenómenos del mundo real. Es decir que estarán en capacidad de percibir, aprender, abstraer y razonar a partir de conjunto de datos que se les presenten en situaciones cotidianas. Esto es un paso mas en el desarrollo de la IA, sin llegar a ser completamente un AGI.

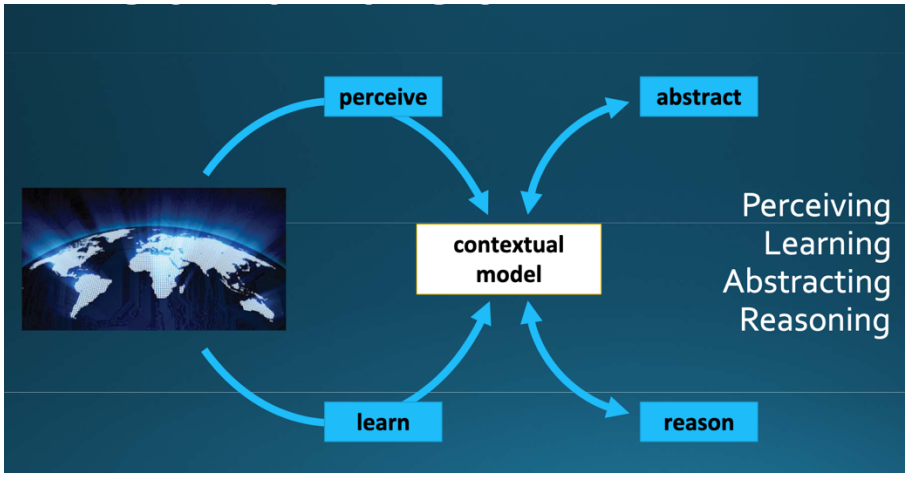

Figura 15. DARPA- TERCERA OLA DE IA

Las aplicaciones que estarían dentro de esta ola aún no llegan masivamente al mercado, aunque comercialmente ya hay algunos productos, en especial de interacción humana y lenguaje natural, que se comercializan con algoritmos que responderían a esta clasificación.[31]

De acuerdo con esta perspectiva, los próximos avances en IA se darán de la mano de las llamadas arquitecturas cognitivas

6.2. Arquitecturas cognitivas

Una arquitectura cognitiva se refiere tanto a una teoría sobre la estructura de la mente humana como a una instanciación computacional de dicha teoría utilizada en los campos de la inteligencia artificial (IA) y la ciencia cognitiva computacional.

La arquitectura cognitiva busca entender y modelar la forma como los individuos regulan y logran el equilibrio de sus estados internos, es decir la homeostasis, así como la información relativa a los sistemas sensoriales y de percepción de que disponen para interactuar con su entorno [11].

[31]Por ejemplo, la empresa Aigo ofrece productos basados en arquitecturas cognitivas, que al menos en teoría, serian mas potentes que las redes neuronales actuales: https://www.aigo.ai/technology

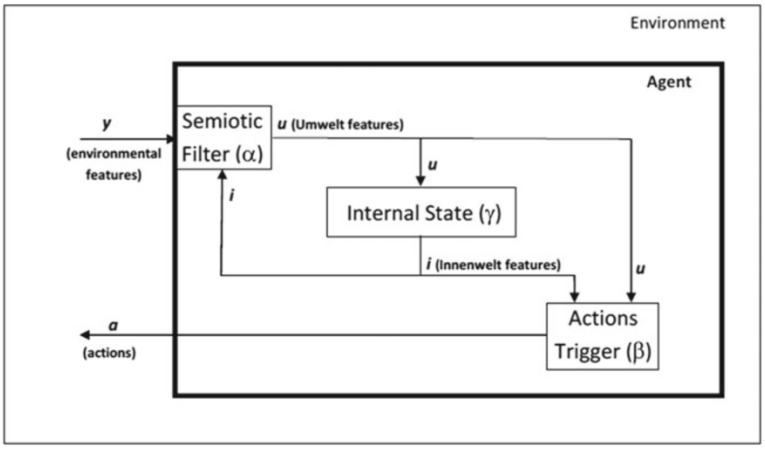

Figura 16. MODELO DE COGNICIÓN[32]

El Instituto de Tecnologías Creativas de la Universidad de California[33] define la arquitectura cognitiva como: "hipótesis sobre las estructuras fijas que proporcionan una mente, ya sea en sistemas naturales o artificiales, y cómo funcionan juntas, junto con el conocimiento y las habilidades incorporadas dentro de la arquitectura, para producir un comportamiento inteligente en una diversidad de entornos complejos " [12].

Como se mencionó previamente, ya hay algunos sistemas comerciales que dicen utilizar sistemas de este tipo en sus agentes de IA. Sin embargo, su estado actual de desarrollo no permite determinar claramente las ventajas que tendría en una aplicación real sobre sistemas basados en *machine learning* o *deep learning.*

6.3. Principio de energía libre e inferencia activa

El principio de energía libre es un modelo propuesto por el neurocientífico Karl Friston y se basa en la forma como los sistemas vivos e inertes permanecen en estados estacionarios de equilibrio al limitarse a un número reducidos de estados.

El principio de energía libre modela los sistemas través de cadenas de Markov que intentan minimizar la diferencia entre su modelo interno y lo que percibe a través de

[32]De [11] página 23.
[33] https://cogarch.ict.usc.edu/

diferentes sensores. Esta diferencia se reduce mediante la corrección continua del modelo creado en el sistema. El principio se basa en la idea bayesiana del cerebro como un motor de inferencia que minimiza el error entre el modelo interno y lo que percibe del mundo exterior a través de acciones constantes de interacción con el entorno. Al cambiar activamente el mundo al estado esperado, los sistemas también pueden minimizar la energía libre del sistema. El aporte de Friston está en asumir que este es el principio aplica a toda relación biológica, y por extensión, seria aplicable a la inteligencia artificial.

Al menos en el laboratorio, las implementaciones de IA basadas en el principio de inferencia activa han mostrado ventajas sobre otros métodos[34] y prometen sobrellevar algunas de las limitaciones de los sistemas actuales. Sin embargo, no es claro que tan viable sea su implementación en aplicaciones comerciales [13].

6.4. Inteligencia artificial y computación cuántica

En años recientes la investigación en computación cuántica se ha visto impulsada por empresas como google, IBM y Microsoft, así como varios gobiernos como Estados Unidos y China, que buscan ser los primeros en llegar al mercado con productos comerciales que apliquen sus desarrollos en campos como la computación integrada de alto rendimiento (HPEC), la criptografía cuántica, incluso sistemas de computo domésticos. Los investigadores argumentan que el procesamiento de información cuántica es más eficiente y seguro porque se basa en mecánica cuántica, lo cual promete abrir posibilidades casi ilimitadas de procesamiento. Es decir que estos sistemas, al menos en principio, podrán superar ampliamente a todos los sistemas de computo actuales.

Si bien la computación cuántica aún permanece en investigación y desarrollo, ya existen demostraciones prácticas en todo el mundo. Se prevé que en los próximos años el rendimiento de estas máquinas llegue al punto de demostrar una clara ventaja por primera vez, y quizás se alcance la supremacía quántica[35]. Por ahora, la

[34] https://www.wired.com/story/karl-friston-free-energy-principle-artificial-intelligence/
[35] Debido a que el significado original del término "supremacía cuántica", propuesto por John Preskill en 2012, era describir el punto en el que las computadoras cuánticas pueden hacer cosas que las

computación cuántica no tiene ventajas prácticas sobre la tecnología de computación estándar. Sin embargo, al ritmo actual de desarrollo de la computación cuántica, es previsible que sistemas comerciales estarán disponibles en las próximas décadas, lo cual abrirá la puerta a nuevos desarrollos que, posiblemente, impulsen una nueva generación de sistemas de IA.

Por otra parte[36], muchos investigadores advierten sobre la militarización de la IA y su combinación con sistemas de computo cuánticos. De momento se desconoce si será posible incorporar sistemas de IA en computadores cuánticos. Pero si se tiene certeza de que sistemas que utilizan aprendizaje automático y *deep learning* ya están en uso por parte de las fuerzas armadas de varios países. Es previsible que la combinación de sistemas cuánticos con IA será implementada en ámbitos militares por varias más naciones a medida que estén disponibles.

computadoras clásicas no pueden, este umbral no se ha alcanzado y es objeto de debate si los desarrollos actuales de empresas como google ya lo alcanzaron.

[36] https://www.militaryaerospace.com/computers/article/14182330/future-of-artificial-intelligence-and-quantum-computing

7. Retos para el gobierno de la seguridad de la información

La manera en que una empresa percibe el uso de la IA para mejorar su nivel de ciberseguridad está determinada por la misión y las capacidades existentes de esa organización. Una compañía que decide implementar sistemas con inteligencia artificial debe destinar recursos considerables y la alta dirección debe estar comprometida en apoyar con recursos humanos, técnicos y financieros.

El principal desafío del uso de inteligencia artificial está en que la organización tenga claras sus ventajas y limitaciones. Para implementar y mantener sistemas de IA se requieren cantidades considerables de recursos de procesamiento y almacenamiento de datos. Además, debido a que los sistemas de IA están entrenados a través de conjuntos de datos de aprendizaje, la empresa debe tener muchos conjuntos de datos diferentes con códigos de malware, códigos no maliciosos, anomalías y firmas de tráfico. Obtener todos estos conjuntos de datos lleva mucho tiempo y recursos que algunas empresas no pueden permitirse.

Además, existe un déficit en el número de profesionales capacitados en áreas relacionadas con seguridad de la información. La promesa de los sistemas de inteligencia artificial de operar de manera autónoma aún no se cumple plenamente porque también requieren de personal altamente capacitado que pueda administrarlos, operarlos y mantenerlos de manera satisfactoria. En este sentido, las organizaciones que buscan destinar menos personal a las áreas de seguridad deben contemplar que el nivel de especialización requerido es bastante alto, lo cual también es un reto importante para la gestión de los recursos humanos, que ya de por si enfrenta dificultades para conseguir y mantener el personal en el largo plazo.

Otro desafío importante para la gobernanza de la seguridad de la información es que quizás sea necesario implementar un programa de gobierno de datos para las plataformas de inteligencia artificial. No se debe confiar ciegamente en que al implementar estos sistemas la integridad, confiablidad y disponibilidad de los datos se mantenga de manera automática.

7.1. Retos para la gestión de riesgos:

Otro desafío que la organización debe considerar es que la inteligencia artificial es una superficie de ataque adicional, con nuevos riesgos que deben ser considerados y evaluados.

Los sistemas de IA están recibiendo datos de fuentes internas o externas a la organización, por lo tanto, es necesario filtrar y desinfectar estos datos o podrían corromper el sistema y llevarlo a presentar comportamientos inesperados. Ya se mencionó en la sección 2.11 que la se debe considerar la existencia del *adversarial machine learning,* o aprendizaje automático de confrontación.

Además, es necesario que la organización esté en capacidad de explicar el funcionamiento de los modelos de IA que utilice, incluso si el modelo es complejo, de forma que los resultados sean estables y predecibles. Para esto, las organizaciones deben ajustar sus Ciclos de Vida de Desarrollo de Software (SDLC) o DevSecOps para probar y asegurar los modelos de manera constante.

Si a esto le sumamos los riesgos que se presentan al implementar servicios e infraestructura en la nube, o la potencial incorporación de sistemas cuánticos, las empresas enfrentan un panorama incierto en donde surgen amenazas completamente nuevas e inesperadas. Esto conlleva a que la implementación de sistemas de inteligencia artificial implique la revisión de la estrategia de gestión del riesgo corporativa, para que estos nuevos riesgos se mantengan dentro de los niveles que la organización estime y esté dispuesta a asumir.

7.2. Retos de cumplimiento

Respecto al cumplimiento que deben observar las empresas en algunos sectores específicos, es necesario valorar que los sistemas de inteligencia artificial estén en capacidad de cumplir y exceder las expectativas que tiene la compañía sin afectar el cumplimiento de alguna regulación. Puede darse alguna circunstancia en la cual un sistema de inteligencia artificial tome una decisión que afecte la operación del negocio, y a su vez, genere un incumplimiento de alguna normativa, ¿de quién es la

responsabilidad? La alta dirección debe tener claras estas implicaciones. Como se mencionó previamente, la aplicabilidad de los resultados y las decisiones que toman los sistemas de IA son fundamentales para algunas organizaciones.

Un ejemplo de la relación entre cumplimento y IA lo podemos encontrar en GDPR: “Aunque la inteligencia artificial (IA) no se menciona explícitamente en el Reglamento General de Protección de Datos de la UE (GPDR), muchas de sus disposiciones son relevantes para el uso de la IA, y algunas de hecho enfrentan desafíos planteados por las nuevas formas de procesar datos personales que están habilitados por IA” [14]. Por ejemplo, si un usuario autoriza el tratamiento de datos para un propósito, y a partir de este, una IA extrapola información útil, ¿el consentimiento original que se dio cubre esto? La política de tratamiento de datos personales de la empresa debe contemplar estos escenarios.

Hay que recordar que el costo del “no cumplimiento” tanto de la normativa externa a la organización como de la normativa interna puede tener implicaciones civiles, penales, financieras, de imagen o reputación, y que con la inteligencia artificial el impacto varía sustancialmente, ya que se estará entregando la toma de algunas decisiones a sistemas automatizados, sin embargo, la responsabilidad, al menos a la luz de las regulaciones actuales, sigue siendo de la alta dirección.

Por lo tanto, es de esperar que regulaciones como PCI-DSS, NIST y similares, se actualicen para incluir aspectos relacionados con IA. La organización debe estar preparada para entender y asumir estos ajustes.

8. Conclusión

La inteligencia artificial promete un mundo de posibilidades técnicas que, en muchas situaciones, facilitarán el desarrollo de tareas complejas o tediosas para el personal de seguridad informática. Sin embargo, como toda tecnología novedosa, también implica retos que, en este caso, son bastante complejos y requieren compromisos importantes por parte de las organizaciones.

Aunque la mayoría de las empresas latinoamericanas reconocen el potencial de las tecnologías basadas en inteligencia artificial, la región aún está lejos de alcanzar el nivel de implementación y desarrollo que tienen países de Europa, Norteamérica y Asia. En este sentido, sino se implementan políticas que fortalezcan la adopción de estas tecnologías, la brecha tecnológica entre la región y otras zonas se seguirá ampliando.

9. Glosario:

Aprendizaje automático (ML): El aprendizaje automático es la ciencia de conseguir que las computadoras actúen sin ser programadas explícitamente.

Aprendizaje profundo (DL): Algoritmos inspirados en la estructura y función del cerebro, creando una red neuronal artificial.

Ciberseguridad: Conjunto de capacidades que permite a las organizaciones detectar, predecir y responder a las ciberamenazas.

Data mining: La minería de datos, también llamada descubrimiento de conocimiento en bases de datos en informática, es el proceso de descubrir patrones y relaciones interesantes y útiles en grandes volúmenes de datos. Combina herramientas de estadística e inteligencia artificial (como redes neuronales y aprendizaje automático) con la gestión de bases de datos para analizar grandes colecciones digitales, conocidas como conjuntos de datos. La minería de datos se usa ampliamente en negocios (seguros, banca, comercio minorista), investigación científica (astronomía, medicina) y seguridad gubernamental (detección de delincuentes y terroristas)[37].

Inteligencia artificial (IA): La inteligencia artificial (IA) es un término colectivo para las capacidades mostradas por los sistemas de aprendizaje que son percibidos por los seres humanos como representantes de la inteligencia. Hoy en día, las capacidades típicas de IA incluyen reconocimiento de voz, imagen y vídeo, objetos autónomos, procesamiento de lenguaje natural, agentes conversacionales, modelado prescriptivo, creatividad aumentada, automatización inteligente, simulación avanzada, así como análisis y predicciones complejas.

Keylogger: Puede ser un hardware o software destinado a automatizar el proceso de registro de pulsaciones de teclas. Estas herramientas registran los datos enviados por cada pulsación de tecla en un archivo de texto para ser recuperados en un momento

[37] https://www.britannica.com/technology/data-mining

posterior. Algunas herramientas pueden grabar todo en su portapapeles de copiar, cortar y pegar, llamadas, datos de GPS e incluso imágenes de micrófono o cámara[38].

Ransomware: es un tipo de software malicioso que cifra los datos de una víctima hasta que el atacante recibe un rescate predeterminado. Por lo general, el atacante exige el pago en forma de criptomoneda como bitcoin. Solo entonces el atacante enviará una clave de descifrado para liberar los datos de la víctima[39].

RAT: Troyano de acceso remoto

[38] https://usa.kaspersky.com/resource-center/definitions/keylogger
[39] https://www.cisco.com/c/en/us/solutions/security/ransomware-defense/what-is-ransomware.html

10. Bibliografía

[1] R. J. Sternberg, «Human intelligence,» Encyclopædia Britannica, Inc., 2020. [En línea]. Available: https://www.britannica.com/science/human-intelligence-psychology#ref13341.

[2] N. Ackermann, «Artificial Intelligence Framework: A Visual Introduction to Machine Learning and AI,» diciembre 2018. [En línea]. Available: https://towardsdatascience.com/artificial-intelligence-framework-a-visual-introduction-to-machine-learning-and-ai-d7e36b304f87.

[3] D. F. Clarence Chio, Machine Learning and Security, Sebastopol, CA : O'Reilly Media, Inc., , 2018.

[4] W. Ertel, Introduction to Artificial Intelligence, Weingarten Germany: Springer, 2017.

[5] G. H. X. Q. a. D. L. K. Bock, «Geneva: Evolving Censorship Evasion,» de *ACM Conference on Computer and Communications Security (CCS)*, 2019.

[6] Vmware carbon black, «Beyond the Hype: Security Experts Weigh in on Artificial Intelligence, Machine Learning and Non-Malware Attacks,» logo , 2017.

[7] B. F. G.W. Ray Davidson PhD, «Security Gets Smart with AI,» SANS Institute, 2019.

[8] Capgemini research institute, «Reinventing Cybersecurity with Artificial Intelligence,» Capgemini, 2019.

[9] Bloomberg, «Five components that artificial intelligence must have to succeed,» Bloomberg, 25 09 2017. [En línea]. Available: https://www.bloomberg.com/professional/blog/five-components-artificial-intelligence-must-succeed/. [Último acceso: 08 2020].

[10] L. Columbus, «10 Predictions How AI Will Improve Cybersecurity In 2020,» Robes, 24 11 2019. [En línea]. Available: https://www.forbes.com/sites/louiscolumbus/2019/11/24/10-predictions-how-ai-will-improve-cybersecurity-in-2020/#53e520d76dd7. [Último acceso: 07 2020].

[11] M. I. A. Ferreira y J. S. S. Rodrigo, Cognitive Architectures, Switzerland: Springer, 2019.

[12] Institute for Creative Technologies , «Cognitive Architecture,» ICT, 2020. [En línea]. Available: https://cogarch.ict.usc.edu/. [Último acceso: 08 2020].

[13] Wired Magazine, «The Genius Neuroscientist Who Might Hold the Key to True AI,» Wired, 13 11 2018. [En línea]. Available: https://www.wired.com/story/karl-friston-free-energy-principle-artificial-intelligence/. [Último acceso: 08 2020].

[14] J. M. Kizza, Guide to Computer Network Security, Chattanooga, TN, USA: Springer, 2017.