


## Facultad de Tecnología Informática

# Seguridad En Redes Sociales: Prevención De Ataques De Ingeniería Social.

**Profesora de Trabajo Final: Dra. Samela Marcela**
**Alumna: Dileo Daiana**

**Trabajo Final de Carrera presentado para obtener el título de**
**Lic. en Gestión de Tecnología informática.**

**2025**

## Resumen

En la actualidad, los usuarios de plataformas digitales están cada vez más expuestos a ataques de ingeniería social, una modalidad de ciberdelito que ha aumentado en complejidad y alcance. Estas amenazas se potencian por la gran cantidad de información personal disponible en línea y por la manipulación de aspectos humanos como la confianza, la urgencia o la falta de conocimiento. Esta combinación de vulnerabilidades convierte a los usuarios en blancos frecuentes y difíciles de proteger mediante métodos tradicionales. El presente trabajo de investigación ha abordado la aplicación de métodos de inteligencia artificial, en particular del aprendizaje profundo, como herramienta para detectar patrones de comportamiento inusuales y prevenir estas amenazas en entornos digitales.

Se realizó una revisión de los principales tipos de ataques, así como de los algoritmos más utilizados para la detección automatizada de anomalías, incluyendo redes neuronales convolucionales, recurrentes y autoencoders. Asimismo, se compararon estos enfoques con técnicas tradicionales de aprendizaje automático, con el objetivo de identificar sus ventajas y limitaciones en términos de adaptabilidad, precisión y capacidad de respuesta.

La investigación aporta al campo de la ciberseguridad no solo una visión integral sobre el potencial del aprendizaje profundo, sino también propuestas orientadas a mejorar su eficacia a partir de la incorporación de la perspectiva del usuario, considerando sus conocimientos, percepciones y niveles de confianza frente a sistemas automatizados de protección.


*Palabras clave*: aprendizaje automático, aprendizaje profundo, ciberseguridad, ingeniería social, inteligencia artificial

**Abstract**

Currently, social media users are increasingly exposed to the alarming risk of social engineering attacks a form of cybercrime that has evolved significantly in both sophistication and scope. These attacks are enabled by the massive exposure of personal information in digital environments and the manipulation of human factors such as trust, urgency, or lack of awareness. This combination of vulnerabilities makes users frequent targets and difficult to protect using traditional methods.This research explores the application of artificial intelligence methods particularly deep learning as a tool to detect unusual behavioral patterns and prevent threats in digital environments. The study includes a review of the main types of attacks, as well as the most widely used algorithms for automated anomaly detection, including convolutional neural networks, recurrent neural networks, and autoencoders. These approaches are also compared with traditional machine learning techniques, with the aim of identifying their advantages and limitations in terms of adaptability, accuracy, and responsiveness.

The investigation seeks to contribute to the field of cybersecurity not only by offering a comprehensive perspective on the potential of deep learning, but also by proposing ways to improve its effectiveness through the incorporation of the user's viewpoint taking into account their knowledge, perceptions, and levels of trust in automated protection systems.

*Keywords:* artificial intelligence, cybercriminals, cybersecurity, social engineering, social networks

## Dedicatoria

A mi familia, por estar siempre a mi lado y apoyarme en cada decisión tomada.
A mi compañero de vida por su paciencia infinita y por comprender mis momentos de ausencia.
A los amigos que me acompañaron en este proceso, brindándome ánimo en los momentos más necesarios.

Índice General

Resumen ..... **2**
**Capítulo 1: Introducción** ..... **10**
1.1 Identificación y planteamiento del problema ..... 10
1.2 Justificación del tema ..... 11
1.3 Objetivo General ..... 13
*1.3.1 Objetivos Particulares* ..... 13
1.4 Hipótesis ..... 13
**Capítulo 2: Estado del Arte** ..... **14**
2.1 Introducción ..... 14
2.2 Modalidades de Ingeniería Social en redes sociales ..... 14
2.3 Aplicaciones de Machine Learning en ciberseguridad ..... 16
2.4 Deep Learning aplicado a ciberseguridad ..... 17
2.4 Estudios comparativos entre ML y DL ..... 19
2.5 Conclusión ..... 20
**Capítulo 3: Marco teórico** ..... **22**
3.1 Introducción ..... 22
3.2 Relevancia de la Ciberseguridad en el entorno digital ..... 22
3.3 La Ingeniería Social como Amenaza en la Ciberseguridad ..... 23
3.4 Inteligencia Artificial en Ciberseguridad ..... 24
*3.4.1 Introducción a la IA* ..... 24
*3.4.2 Historia de la IA en el ámbito de la ciberseguridad* ..... 25
*3.4.3 Detección de anomalías* ..... 26
3.5 Machine Learning y Deep Learning ..... 27
*3.5.1 Introducción al ML* ..... 27
*3.5.2 Introducción al Deep Learning* ..... 31
*3.5.3 Las redes neuronales artificiales* ..... 31
3.6 Tipos de Modelos de DL ..... 32
*3.6.1 Redes convolucionales (CNN)* ..... 32

*3.6.2 Redes Neuronales Recurrentes (RNN)* ..... 33
*3.6.3 Autocodificadores (Autoencoders)* ..... 34
3.7 Conclusión ..... 35
**Capítulo 4: Enfoque Metodológico ..... 37**
4.1 Metodología ..... 37
4.2 Recolección de Datos ..... 37
4.3 Preguntas de Investigación ..... 38
4.4 Selección de muestra ..... 38
4.5 Análisis de los Datos ..... 39
**Capítulo 5: Análisis de Resultados ..... 40**
5.2 Presentación de resultados ..... 40
*5.2.1 Edad del Entrevistado* ..... 40
*5.2.2 Nivel Educativo* ..... 41
*5.2.3 Área de Trabajo o Estudio* ..... 42
*5.2.4 Experiencia con Fraudes en Redes Sociales* ..... 43
*5.2.5 Red Social del Último Intento de Fraude* ..... 44
*5.2.6 Medidas de Seguridad Reconocidas* ..... 45
*5.2.7 Frecuencia de Mensajes Sospechoso* ..... 46
*5.2.8 Reacción ante Mensajes Sospechosos* ..... 47
*5.2.9 Percepción sobre Ciberseguridad e IA* ..... 49
*5.2.10 Preguntas relacionadas a ciberseguridad* ..... 50
*5.2.11 Familiaridad con IA en Ciberseguridad* ..... 50
*5.2.12 Confianza en la IA para proteger sus redes sociales.* ..... 51
*5.2.13 Conocimiento de Conceptos de Aprendizaje Profundo* ..... 52
*5.2.14 Comparación entre IA y Métodos Tradicionales* ..... 53
**Capítulo 6: Análisis relacional de la encuesta ..... 55**
6.1 Percepción de eficacia sobre medida de seguridad automatizadas ..... 55
6.2 Conocimiento de conceptos de IA vs. percepción de efectividad de la IA ..... 59
6.3 Ocupación en tecnología o ciberseguridad vs. percepción sobre la capacidad de la IA para detectar comportamientos anómalos ..... 63

6.4 Recepción de alertas automáticas, reacción ante mensajes sospechosos y frecuencia de exposición ..... 66
6.5 Edad y Reacción ante mensajes sospechosos ..... 69
**Capítulo 7: Conclusiones ..... 71**
Capítulo 8: Líneas futuras de investigación **..... 74**
Referencias **..... 75**
Anexo **..... 81**
Anexo 1: Encuesta distribuida ..... 81

# Índice de figuras

Figura 1. Reporte de cibercrimen en Argentina....12
Figura 2. Cuestiones abiertas en la detección de ataques de ingeniería social. ....15
Figura 3. Aprendizaje supervisado....28
Figura 4. Aprendizaje No supervisado....29
Figura 5. Aprendizaje por refuerzo....30
Figura 6. Ejemplo de arquitectura de una red profunda prealimentada o feedforward....32
Figura 7. Ejemplo de arquitectura de red CNN....33
Figura 8. Arquitectura de Red Neuronal Recurrente....34
Figura 9. Arquitectura de Autoencoders....35
Figura 10. Resultados pregunta N° 1:¿Cuál es la edad del entrevistado?....40
Figura 11. Resultados pregunta N°2: Nivel educativo del entrevistado....41
Figura 12. Resultados pregunta N° 3: Área de trabajo o estudio del entrevistado. ....42
Figura 13. Resultados pregunta N°4: Experiencia del entrevistado con fraudes en redes sociales. ....43
Figura 14. Resultados pregunta N° 5: Red social del último intento de fraude el entrevistado. ..44
Figura 15. Resultados de pregunta N° 7: Frecuencia en que los encuestados reciben mensajes sospechosos. ....47
Figura 16. Resultados pregunta N°8: Reacción de los encuestados ante mensajes sospechosos. 48
Figura 17. Resultados de pregunta N° 9: Percepción que tiene los encuestados sobre la Ciberseguridad e Inteligencia Artificial ....49
Figura 18. Resultados pregunta N°12: Familiaridad con el uso de la IA en ciberseguridad.....50
Figura 19. Resultados pregunta N°14: Confianza en Inteligencia Artificial para proteger redes sociales. ....51
Figura 20. Resultados pregunta N°15: Conocimiento de los Conceptos de Inteligencia Artificial. ....52
Figura 21. Resultados pregunta N°17: Comparación de los métodos de Inteligencia Artificial con los tradicionales....54
Figura 22. Representación gráfica del NT de las preguntas analizadas....57
Figura 23.
Confianza en un sistema basado en IA para proteger las redes sociales.....60
Figura 24. Mapa de calor sobre la confianza en los sistemas de IA. ....61
Figura 25. Resultado prueba chi-cuadrado....63
Figura 26. Percepción de la efectividad de la IA por área profesional. ....64
Figura 27. Acuerdo del uso de la IA en detección de anomalías por área de desempeño ....65

## Índice de tablas

Tabla 1. Resultados pregunta N° 6: Medidas de seguridad reconocidas por los encuestados .....46
Tabla 2: Preguntas N° 11, 12, 14, 16 del cuestionario. .....56
Tabla 3. Resultados de respuestas de preguntas N°11, 12, 14 y 17. .....57
Tabla 4. Confianza y detección de alertas según conocimientos en Deep Learning. .....58
Tabla 5. Frecuencias observadas para variables perfectamente independientes .....62
Tabla 6. Frecuencias esperadas para variables perfectamente independientes .....62
Tabla 7. Comportamiento ante mensajes sospechosos. Usuarios que recibieron alertas automáticas. .....66
Tabla 8. Comportamiento ante mensajes sospechosos. Usuarios que no recibieron alertas automáticas. .....67
Tabla 9. Comportamiento ante mensajes sospechosos. Usuarios que no saben si recibieron alertas automáticas. .....67
Tabla 10. Red Social y recepción de alerta automática. .....68
Tabla 11. Cruce de edad con reacción ante mensajes sospechosos. .....69

# Capítulo 1: Introducción

## 1.1 Identificación y planteamiento del problema

La creciente proliferación de las redes sociales aumentó la exposición de los usuarios a diversos delitos informáticos. Coronel Ayala y López Sevilla (2023) mencionan que el uso masivo de sistemas digitales a nivel global ha expuesto tanto a personas como a organizaciones a distintos tipos de ciberataques. Esto sucede debido a que, en ciertas ocasiones, los datos no cuentan con la protección adecuada.

A pesar de la existencia de medidas preventivas, como la capacitación y concientización sobre ciberseguridad, muchos usuarios continúan siendo víctimas de este tipo ataques debido a la falta de conocimiento o a la implementación deficiente de políticas de seguridad (Sancho-López, 2023).

Es importante destacar que en muchos de los ciberataques se emplean técnicas de ingeniería social. Al respecto, Borbón Sanabria (2012) explica que la recopilación de información constituye la base de este tipo de delitos; para ello, es crucial explorar todos los medios posibles que puedan contener datos sobre el objetivo, destacando en este aspecto el papel de Internet y de las redes sociales. De acuerdo con lo anterior, Berenguer Serrato (2018) menciona que la forma más efectiva de prevenir los ataques de ingeniería social es aprender a reconocerlos cuando ocurren. Por lo tanto, comprender e identificar comportamientos sospechosos se convierte en un aspecto clave para la prevención de los delitos.

En este contexto la Inteligencia Artificial (IA) emerge como una herramienta crucial, como menciona Moreno Barea “uno de los campos más prósperos de las ciencias de la computación es la inteligencia artificial (IA). Los problemas a resolver mediante la IA han ido cambiando con el tiempo, conforme ha avanzado esta tecnología” (Moreno Barea, 2023, p.4). Los avances en este campo tienen un impacto relevante en el ámbito de la ciberseguridad.

Considerando la creciente sofisticación de estos delitos y la importancia de reconocer patrones sospechosos, esta investigación plantea la pregunta: ¿Cómo mejoran los métodos de IA la eficacia en la prevención de ataques de ingeniería social a partir de la detección de anomalías de comportamiento de usuarios en redes sociales?

Para poder responder a ese interrogante, es necesario evaluar las técnicas más recientes empleadas en ciberseguridad, Ruiz Martínez explica que “en los últimos años, se ha escuchado hablar mucho sobre Inteligencia Artificial y, en especial, de una de sus ramas más destacadas, como lo es el machine learning” (Ruiz Martínez, 2021, p.44). Por otro lado, otros autores añaden que:

> La concesión a las máquinas de la capacidad de aprender se conoce como aprendizaje automático o machine learning (ML), y ha sido el principal contribuyente al resurgimiento de la IA. Especialmente en la última década, el enfoque con mayor progresión y éxito del ML moderno se centra en el modelado mediante redes neuronales artificiales con múltiples etapas de procesamiento, actualmente conocido como aprendizaje profundo, representado comúnmente mediante su anglicismo deep learning. (Moreno Barea, 2023, p. 4)

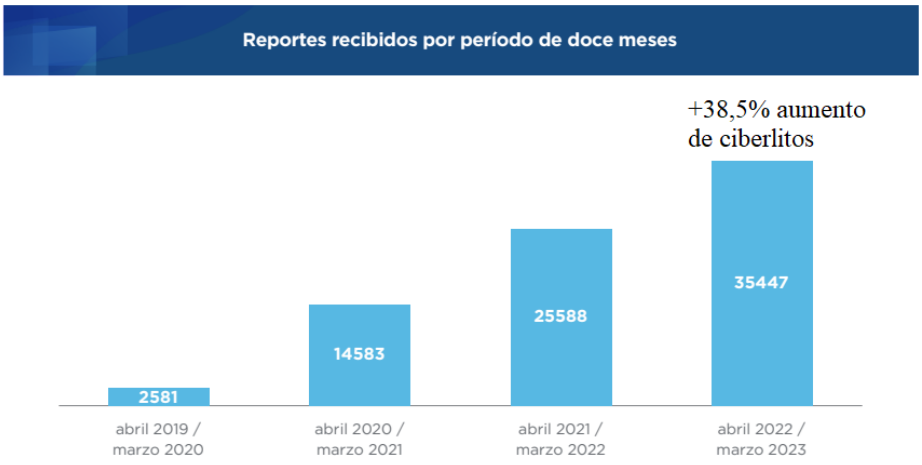
Con el avance de esta investigación se pretende generar aportes concretos para el diseño de estrategias de ciberseguridad más eficaces. Para ello, se explora el potencial de la inteligencia artificial, y particularmente de las técnicas de aprendizaje profundo (deep learning), como herramientas clave para mejorar la detección de anomalías y prevenir ataques de ingeniería social en entornos digitales cada vez más complejos y dinámicos.

## 1.2 Justificación del tema

EL incremento de ciberdelitos se debe, entre otras causas, al mayor uso de las redes sociales. Borbón Sanabria (2012) explica que la expansión de las redes de datos impulsada por el uso de dispositivos móviles, como teléfonos celulares, smartphones y tabletas, facilitó el acceso y uso de diversas aplicaciones, portales y herramientas, integrándolas en la rutina diaria. Sin embargo, también son objeto de numerosas críticas debido a la gestión de la privacidad de la información y a los controles de seguridad implementados en sus plataformas.

Un claro ejemplo del creciente aumento del cibercrimen puede observarse en la Figura 1 donde evidenciamos que, en Argentina, este tipo de delitos aumentaron un 38.5% entre abril de 2022 y marzo de 2023 (Unidad Fiscal Especializada en Ciberdelincuencia, [UFECI] 2023).

**Figura 1.**
Reporte de cibercrimen en Argentina




*Nota:* Adaptado de Reporte recibidos por periodo de doce meses por UFECI, 2022-2023. https://www.fiscales.gob.ar/wp-content/uploads/2023/12/UFECI_informe-de-Gestion_23_15-12.pdf

La ingeniería social ocupa un lugar destacado entre los distintos tipos de ciberataques. Los delincuentes utilizan esta práctica para robar información personal y financiera con el objetivo de ejecutar ataques más sofisticados. Esto se debe tanto a la constante mejora de las técnicas de los atacantes como a la falta de concientización y capacitación de los usuarios, quienes en muchos casos no están familiarizados con prácticas como el phishing. Para mitigar estos ataques, es esencial contar con soluciones de seguridad robustas, acompañadas de la educación y formación tanto a nivel individual como corporativo (Lubeck, 2021).

En esta misma línea Coronel Ayala y López Sevilla (2023) exponen que el conflicto entre los piratas informáticos y la ciberseguridad se intensificará en los próximos años, ya que los atacantes trabajan constantemente en desarrollar y explotar técnicas maliciosas mediante sistemas cada vez más avanzados, lo que obliga a la seguridad a evolucionar de forma acelerada.

## 1.3 Objetivo General

El objetivo de esta investigación es evaluar cómo los métodos de IA mejoran la eficacia de prevención de ataques de ingeniería social a partir de la detección de anomalías de comportamiento de usuarios en redes sociales.

### *1.3.1 Objetivos Particulares*

- Identificar diferentes tipos de ataques de ingeniería social dentro del ámbito de las redes sociales.
- Evaluar los métodos de aprendizaje profundo utilizados para la detección de anomalías de comportamiento.
- Comparar las técnicas de detección de anomalías de aprendizaje automático con las de aprendizaje profundo para determinar alcances y limitaciones de cada uno.

## 1.4 Hipótesis

Los métodos de aprendizaje profundo, como las redes convolucionales, las redes neuronales recurrentes y los autocodificadores, aumentan la eficacia en la prevención de ataques de ingeniería social mediante una detección más precisa de anomalías en el comportamiento de los usuarios en redes sociales.

# Capítulo 2: Estado del Arte

## 2.1 Introducción

El presente Estado del Arte se centra en el análisis de investigaciones publicadas entre los años 2022 y 2025 que abordan la aplicación del Deep Learning (DL) en la detección de anomalías dentro de redes sociales. En primer lugar, se presentan antecedentes relacionados con las distintas modalidades de ingeniería social, a fin de contextualizar el problema y comprender las técnicas de manipulación utilizadas por los atacantes. Luego, se analizan estudios que describen el rol de la tecnología en la mitigación de estas amenazas, con especial énfasis en el uso de algoritmos de Machine Learning para la detección automatizada de patrones sospechosos.

A continuación, se revisa la evolución desde el Machine Learning hacia el Deep Learning, incorporando investigaciones que evalúan empíricamente el rendimiento de distintos modelos en entornos reales. Asimismo, se presenta una discusión teórica sobre los enfoques basados en redes neuronales para reconocer comportamientos atípicos y detectar posibles amenazas emergentes. Finalmente, se incluyen estudios comparativos que contrastan el desempeño del Deep Learning frente a métodos tradicionales de aprendizaje automático, destacando sus ventajas, limitaciones y áreas de oportunidad.

## 2.2 Modalidades de Ingeniería Social en redes sociales

Un estudio reciente desarrollado por Rathod et al. (2025) aporta elementos relevantes para comprender los mecanismos detrás de los ataques de ingeniería social en redes. Entre sus principales hallazgos, los autores señalan que los atacantes se aprovechan de la confianza que las personas depositan en los medios digitales, utilizando enlaces falsos para obtener datos personales o financieros. Este tipo de ataques suele tener como origen el error humano, impulsado por emociones como el miedo, la codicia, la curiosidad o la empatía. En tales situaciones, es la propia víctima quien, sin saberlo, permite el acceso a su información.

Otro de los hallazgos importantes identificado por los autores es la existencia de una serie de desafíos que dificultan la detección oportuna de los ataques de Ingeniería Social. En la Figura 2 se ilustran dichos obstáculos.

**Figura 2.**
Cuestiones abiertas en la detección de ataques de ingeniería social.




*Nota:* Obtenido de Rathod et al. (2025*).”Un estudio exhaustivo sobre ataques de ingeniería social, contramedidas, estudios de casos y desafíos de investigación.”* https://doi.org/10.1016/j.ipm.2024.103928

Entre los principales se encuentran, la recurrencia del error humano, la escasa cultura de reporte ante incidentes, y el uso de técnicas de manipulación psicológica que explotan emociones como la urgencia o el miedo. A esto se suman la exposición de datos personales en redes sociales, la falta de formación específica en ciberseguridad, y el uso dual de la inteligencia artificial, tanto para ataques como para la defensa (Rathod et al. ,2025).

En relación con los ataques que manipulan emociones, la clasificación propuesta por Kaspersky (s.f.) permite identificar cinco modalidades: Phishing, Spear Phishing, Pretexting, Baiting y Quid Pro Quo.

El Phishing sigue siendo uno de crímenes cibernéticos más comunes. Según el informe anual del FBI (2023) encabezó el ranking de delitos informáticos denunciados entre los años 2019 a 2023, alcanzando 298.878 quejas.

El Spear Phishing, como evolución del anterior, se caracteriza por su alto nivel de personalización. Se advierte que incluso se utilizan técnicas de Machine Learning para perfeccionar los mensajes dirigidos a usuarios específicos, basándose en los datos que estos comparten públicamente (Banco Santander, s.f.).

Respecto al Pretexting, Burdova (2023) indica que este tipo de estafas se basa en la creación de contextos falsos para generar confianza en la víctima. Trevino (2023), añade que los delincuentes analizan los perfiles de las víctimas para obtener un conocimiento más profundo sobre ellas.

En cuanto al Baiting, Rudra (2023) explica que consisten en enviar enlaces hacia archivos maliciosos que pueden compartirse en redes sociales como Facebook e Instagram. Burdova (2023) agrega que estos ataques apelan a falsas recompensas para inducir una acción.

Por último, en los ataques Quid Pro Quo, Salman Nadeem (2022) señala que se ofrecen supuestos beneficios a cambio de información, con el objetivo de cometer fraudes o apropiarse de cuentas.

El rasgo central de estos delitos es la manipulación psicológica. En este sentido, resulta ilustrativo el estudio realizado por Benavidez Astudillo et al. (2020), quienes llevaron a cabo un experimento que simuló acciones engañosas que apelaban a la curiosidad de sus contactos, atrayéndolos hacia un blog con contenido falso accesible desde dos redes sociales: Whastapp y Facebook. Durante esta interacción, se instaló de forma encubierta un complemento malicioso que recopilaba información personal del usuario, como su dirección IP, país, sistema operativo y navegador. Los resultados mostraron que la mayoría de las víctimas accedieron desde sus teléfonos móviles y que, aunque muchos usuarios conocían conceptos como inteligencia artificial o Phishing, no sabían cómo protegerse frente a estos ataques. Esto evidenció la falta de formación y concientización en ciberseguridad.

Como proyección futura, los autores sugieren la integración de algoritmos de Aprendizaje Automático con enfoques de seguridad cognitiva orientados a optimizar la detección, el control y la mitigación de ataques de ingeniería social.

### 2.3 Aplicaciones de Machine Learning en ciberseguridad

Comprendida la forma en que los atacantes explotan el factor humano, el siguiente paso es abordar cómo la tecnología busca contrarrestar estas amenazas. González y Quevedo-Sacoto (2025) realizaron un estudio de caso titulado “Tendencias actuales en ataques de ingeniería social” en el que la aplicación de inteligencia artificial para detectar patrones inusuales, junto con la implementación de autenticación multifactor y la capacitación de los usuarios, constituyen elementos fundamentales en la defensa, dado que el factor humano continúa siendo el eslabón más débil.

En la misma línea, Wazid et al. (2022), en su estudio “Uniendo la ciberseguridad y el aprendizaje automático: ventajas, desafíos e investigación futura”, destacan que la combinación

entre la ciberseguridad y el aprendizaje automático (ML) abarca dos dimensiones principales: por un lado, incorporar la ciberseguridad en el desarrollo y uso de modelos, y por otro, aplicar el ML como herramienta para fortalecer la ciberseguridad. Esta integración ofrece múltiples beneficios, entre ellos, aumentar la protección de los modelos, optimizar las técnicas de defensa cibernética y mejorar la detección de ataques de día cero con una menor necesidad de intervención humana.

En el mismo estudio, Wazid et al. (2022) explican que los modelos de aprendizaje automático (ML) son capaces de adquirir conocimiento sobre distintos tipos de ciberataques, tanto en entornos fuera como dentro de línea, utilizando conjuntos de datos previamente procesados. Estos algoritmos pueden identificar en tiempo real indicios de intrusiones o posibles ataques cibernéticos. Una de las ventajas es que reduce los falsos positivos. Dependiendo del entorno y los sistemas, pueden aplicarse distintos enfoques de ML, como el aprendizaje supervisado, no supervisado, por refuerzo o profundo.

Sin embargo, Wazid et al. (2022) también mencionan algunos desafíos respecto de esta integración, como la compatibilidad entre las técnicas de análisis, dado que la combinación de ciberseguridad y ML implica la utilización de diversos métodos de seguridad (por ejemplo, algoritmos de cifrado, firma o hash) en combinación con técnicas de agrupamiento, clasificación o redes neuronales convolucionales (CNN).

### 2.4 Deep Learning aplicado a ciberseguridad

La evolución de las técnicas de aprendizaje automático ha dado lugar al desarrollo del aprendizaje profundo diversos estudios recientes han explorado su incorporación en el ámbito de la ciberseguridad, demostrando su eficacia para detectar amenazas emergentes. Uno de los más relevantes en este aspecto es el realizado por la compañía operadora de internet Sarenet (2024) que destaca que el Deep Learning se ha convertido en un aliado estratégico gracias a su capacidad de adaptarse a nuevas amenazas, permitiendo una respuesta rápida y efectiva. En la detección de patrones anómalos, el DL ofrece una ventaja competitiva, ayudando a los usuarios a adelantarse en la protección de sus activos digitales. Otra afirmación relevante del informe es que la incorporación del aprendizaje profundo en soluciones de ciberseguridad proporciona una defensa más sólida y adaptativa representando una transformación profunda en la forma de entender la ciberseguridad, posicionándose como un componente esencial.

Otros autos como Quirumbay Yagual et al. (2022) sostienen que la efectividad de una solución de seguridad basada en redes neuronales y aprendizaje profundo está determinada tanto por la calidad de los datos como por el rendimiento de los algoritmos utilizados. Sin embargo, los autores advierten que obtener datos confiables en ciberseguridad es difícil, debido a registros ambiguos o incompletos.

Una línea de investigación actualmente en desarrollo es la evaluación empírica del rendimiento de distintos modelos de DL en entornos reales. Barcenas Medina y Alcaraz Cancio (2025) aplicaron una combinación de redes neuronales convolucionales (CNN) y recurrentes (RNN) para identificar amenazas del tipo zero-day en entornos empresariales. Aunque el estudio se centra en sistemas corporativos, aporta elementos valiosos para el presente trabajo que se explican a continuación.

El modelo diseñado por los autores logró una alta precisión en la detección de anomalías y fue capaz de diferenciar correctamente entre actividades normales y maliciosas, incluso frente a ataques no conocidos previamente. Además, se destaca que el modelo presentó una tasa muy baja de falsos positivos y tiempos de respuesta ágiles, lo que permite su aplicación en entornos que requieren análisis en tiempo real. Sin embargo, los autores también identificaron algunas limitaciones, como la necesidad de incluir conjuntos de datos más amplios y representativos para mejorar su desempeño, el uso intensivo de recursos computacionales y la posibilidad, aunque baja, de que existan falsos negativos.

En relación con las limitaciones de las redes neuronales convolucionales (CNN), IBM (2024) ha publicado recientemente un artículo en el cual destaca que, si bien son capaces de detectar patrones dentro de imágenes y videos, también presentan ciertas desventajas: demandan considerable tiempo y recursos económicos, requieren múltiples unidades de procesamiento gráfico, profesionales con alta especialización, así como pruebas exhaustivas de configuraciones y ajustes específicos.  En cuanto a las redes recurrentes, IBM (2024) explica que pueden necesitar un tiempo de entrenamiento extenso y resultar complicadas de manejar con conjuntos de datos grandes. La optimización de estas se vuelve más compleja cuando cuentan con numerosas capas y parámetros.

Otro aporte relevante en esta línea es el estudio de González López et al.(2023) desarrollado en el marco de la Jornadas Nacionales de Investigación en Ciberseguridad. En este

trabajo, se ha analizado cómo la elección de hiperparámetros[1] influye en el rendimiento de los autoencoders utilizados para detectar ataques en redes. Los autores señalan que este tipo de modelos se emplea con frecuencia porque permite diferenciar patrones normales del tráfico de red frente a comportamientos anómalos. En su investigación, observaron que el rendimiento del modelo puede mejorar notablemente cuando estos parámetros se configuran correctamente.

En particular, advierten que cuando se trabaja con datos reales de tráfico de red en los que no siempre se sabe con certeza cuáles registros corresponden a ataques, sumar información no necesariamente mejora el desempeño y, en algunos casos, puede incluso empeorarlo. IBM (2024) añade que los autoencoders pueden no captar correctamente las conexiones compuestas en datos estructurados, lo que les impide identificar adecuadamente relaciones complejas.

En resumen, si bien las redes neuronales convolucionales (CNN), las recurrentes (RNN) y los autoencoders presentan una alta precisión en la detección de patrones anómalos, incluso en imágenes o videos, y muestran un alto grado de éxito en la detección en tiempo real, también presentan ciertas limitaciones. Entre ellas, se destacan la necesidad de contar con conjuntos de datos representativos, la sensibilidad a la configuración de los modelos, que puede afectar su rendimiento, y los tiempos de entrenamiento prolongados, lo cual puede reducir su eficacia en contextos dinámicos como el de las redes sociales.

### 2.4 Estudios comparativos entre ML y DL

En el apartado anterior se presentaron los antecedentes sobre la identificación de comportamiento anómalos a partir de la aplicación de los algoritmos de Deep Learning. En este, se analizan estudios que lo contrastan con el funcionamiento de los métodos tradicionales de Machine Learning. Estos antecedentes permiten comparar el desempeño de ambos enfoques.

Según Crabtree (2024), una diferencia relevante entre ambos enfoques radica en el volumen de datos requerido: mientras que los modelos de aprendizaje profundo demandan volúmenes significativos de información para alcanzar su máximo desempeño, el aprendizaje automático puede ofrecer buenos resultados con volúmenes menores. Esto se vuelve importante cuando la disponibilidad de datos es limitada, como ocurre con frecuencia en entornos reales.

[1] Los hiperparámetros son variables configurables que determinan cómo se entrena un modelo y afectan directamente su rendimiento (Microsoft, 2024).

Respecto a los tiempos de procesamiento, entrenar una red de aprendizaje profundo puede llevar un tiempo considerable, extendiéndose incluso por meses. Por el contrario, los algoritmos tradicionales de aprendizaje automático suelen requerir menos tiempo de entrenamiento.

Sin embargo, una de las principales ventajas que presenta Crabtree (2024) sobre el aprendizaje profundo es que no necesita métodos manuales de extracción de características, ya que identifica automáticamente las más relevantes para cumplir con sus tareas, comenzando con las más simples y avanzando hacia representaciones más complejas. Por el contrario, el aprendizaje automático tradicional depende de características diseñadas manualmente, lo que exige un trabajo más detallado de ingeniería. Esto hace especialmente útil al DL para problemas complejos, como el análisis de imágenes o patrones de comportamiento no estructurados.

En resumen, mientras que el aprendizaje automático tradicional requiere menor volumen de datos y menor tiempo de entrenamiento, el aprendizaje profundo ofrece mayor capacidad de adaptación y autonomía en la extracción de características. Estas diferencias adquieren particular relevancia en escenarios donde es necesario identificar patrones complejos y dinámicos como el analizado en el presente trabajo.

## 2.5 Conclusión

En conclusión, los estudios analizados permiten observar que las redes sociales representan un entorno especialmente vulnerable frente a ataques de ingeniería social, debido a la exposición de información personal y a la explotación de factores emocionales por parte de los atacantes.

Los modelos basados en redes neuronales convolucionales, recurrentes y autoencoders han demostrado un desempeño prometedor en la identificación de comportamientos inusuales, incluso en escenarios complejos y dinámicos. No obstante, su efectividad depende de la disponibilidad de datos representativos, la adecuada configuración de hiperparámetros y la capacidad computacional requerida, factores que aún presentan desafíos para su aplicación generalizada.

La comparación entre enfoques de Machine Learning y Deep Learning revela que, si bien los métodos tradicionales pueden resultar más accesibles en términos de recursos y volumen de datos, las soluciones basadas en arquitecturas profundas destacan por su mayor capacidad para adaptarse a variaciones en el comportamiento y procesar información compleja en tiempo real.

Como resultado, se observa que existe un alto potencial en el uso del aprendizaje profundo en el campo estudiado. A través del análisis cuantitativo planteado en esta investigación, se espera

complementar la literatura relevada desde la perspectiva de los usuarios de redes sociales con perfiles orientados a la Tecnología.

# Capítulo 3: Marco teórico

## 3.1 Introducción

En este capítulo se expone la evolución de la ciberseguridad y su importancia en la protección de las redes sociales, donde como ya se ha mencionado, las amenazas y vulnerabilidades son cada vez más frecuentes.

Asimismo, se examina cómo la ingeniería social se ha consolidado como una de las principales estrategias del cibercrimen y de qué manera la Inteligencia Artificial, particularmente el aprendizaje profundo (Deep Learning), está transformando este campo.

## 3.2 Relevancia de la Ciberseguridad en el entorno digital

La ciberseguridad se consolida como una disciplina esencial en el contexto de un mundo cada vez más interconectado. Según Sancho Hirare (2017), esta se origina en respuesta al aumento del uso del ciberespacio como un nuevo ámbito para la interacción social, consecuencia de la revolución de las tecnologías de la información y comunicación (TIC), que ha impulsado la globalización y sorprende continuamente con su innovación.

Entender la ciberseguridad implica reconocer la amplitud de sus definiciones, que si bien es un término que suele utilizarse para referirse a la seguridad de los sistemas, tiene una amplia variedad de definiciones (Rea Guamán, 2020).

La ciberseguridad ha ganado protagonismo junto con términos como ciberdelincuencia, ciberterrorismo, ciberataque y ciberdefensa (Ballestero, 2020).

La definición expuesta anteriormente por Ballestero subraya la importancia tanto de la ciberseguridad como de los ciberdelitos en el contexto actual. En esta línea, Arteaga y Vera Vázquez (2022) explican que el delito cibernético ha avanzado junto con los desarrollos tecnológicos. Aunque las TICs ofrecen numerosos beneficios, también generan perjuicios, tanto materiales como intangibles, a causa de los programas maliciosos, que mantienen en alerta a autoridades, empresas y ciudadanos. En el mismo artículo se menciona que el ciberfraude, siendo

un tipo ciberdelito, genera pérdidas económicas significativas y afecta la estabilidad social debido a la inseguridad digital.

En cuanto a la vulnerabilidad de los usuarios, Ruiz Martínez (2021) enfatiza que debemos ser conscientes que cuanto mayor sea la interacción en internet, mayor es el riesgo de sufrir ataques, y aspirar a sistemas completamente seguros aún es un ideal inalcanzable.

Jarrin Rodríguez coincide en que "existe un alto riesgo de ataque a los sistemas informáticos que se encuentran interconectados a través de Internet. Esto ha generado la necesidad de desarrollar diversos mecanismos de seguridad que permitan prevenir, evadir y detectar estas irrupciones" (Jarrin Rodríguez, 2019, p.16)

En la misma línea, Domínguez Arteaga y Vera Vázquez (2022) destacan que el delito cibernético ha evolucionado junto con los avances tecnológicos. Aunque las Tecnologías de la Información y las Comunicaciones (TIC) ofrecen múltiples beneficios, su contraparte negativa, está representada por programas maliciosos, e implica riesgos significativos para las autoridades, empresas y la sociedad en general.

### 3.3 La Ingeniería Social como Amenaza en la Ciberseguridad

En el ámbito de la ciberseguridad, las estrategias de manipulación psicológica, comúnmente conocidas como ingeniería social, se consideran una de las amenazas más persistentes y efectivas. Algunos autores mencionan que tiene raíces históricas que se remontan a siglos atrás, donde era utilizada para obtener información confidencial con el propósito de manipular o extorsionar a las víctimas (Alzas Hernández, 2023).

Por otro lado, esta práctica también puede abordarse desde una perspectiva psicológica. Por ejemplo, algunos autores mencionan que la ingeniería social es:

> El conjunto de técnicas o estrategias sociales utilizadas de forma predeterminada por un usuario para obtener algún tipo de ventaja respecto de otro u otros. Por lo que de ingeniería tiene más bien poco, es más, de hecho, se acerca más a la piscología social o a la sociología de ventas. Ya que, para llevar a cabo ataques de ingeniería social, no tiene por qué tener conocimientos técnicos de ningún tipo. (Berenguer Serrato, 2018, p. 4)

En la misma investigación Berenguer Serrato (2018) menciona que esta táctica de

manipulación implica la creación de técnicas para influir psicológicamente en los usuarios de internet, con el objetivo de que revelen información confidencial o cometan errores al llevar a cabo operaciones inseguras, dejándolos vulnerables mediante este enfoque.

El sitio web Proof Point (s.f.) explica que engañar a los usuarios para que revelen información sensible no algo nuevo en ciberseguridad. Lo que ha evolucionado, en cambio, son los métodos empleados, las narrativas para obtener la información, y el nivel de sofisticación de los ataques. Aunque el término ingeniería social fue introducido en 1894 por el industrial holandés JC Van Marken, esta técnica comenzó a ser ampliamente utilizada como una herramienta de ciberataque a partir de la década de 1990.

En relación con esta evolución, Berenguer Serrato (2018) describe cómo los primeros métodos de ingeniería social se centraban en realizar llamadas telefónicas, en las que los estafadores se hacían pasar por entidades legitimas para obtener información confidencial de las víctimas. Sin embargo, con la llegada de Internet, comenzaron a proliferar nuevas formas de ataque. Actualmente, los ciberdelincuentes refinaron sus tácticas gracias a que los dispositivos móviles almacenan una gran cantidad de datos personales, como contactos, fotos, conversaciones y credenciales de redes sociales.

## 3.4 Inteligencia Artificial en Ciberseguridad

### *3.4.1 Introducción a la IA*

Para llevar adelante este apartado sobre la aplicación de la Inteligencia Artificial en la ciberseguridad, resulta fundamental proporcionar al lector una definición clara y precisa de esta disciplina, junto con un análisis detallado de su historia y evolución, poniendo especial énfasis en su influencia dentro del ámbito de la ciberseguridad.

Ponce Gallegos et al. (2014) en la primera edición de su libro Inteligencia Artificial, destacan que, desde sus inicios, esta tecnología ha enfrentado el reto de no contar con una definición precisa y unificada. No sorprende, entonces, que hasta la fecha no exista una definición única aceptada universalmente. Sin embargo, los autores explican que la IA es una rama de la informática que actualmente genera gran interés debido a su amplio rango de aplicaciones. La búsqueda de herramientas para comprenderla, crear modelos y simulaciones ha llevado a numerosos científicos a dedicarse a esta área de investigación.

A continuación, se presentan algunas interpretaciones claves propuestas por expertos en el campo, que ayudarán al lector a obtener una comprensión más clara.

McCarthy (2007) menciona en su artículo publicado por la Universidad de Stanford que la Inteligencia Artificial se refiere a la disciplina científica y tecnológica dedicada al desarrollo de máquinas inteligentes, especialmente programas de computadora capaces de imitar la inteligencia sin necesariamente replicar los mismos procesos que ocurren en los seres vivos.

Por otro lado, Rouhiainen (2018) amplía este concepto señalando que la IA implica el uso de algoritmos que aprenden de datos y aplican ese conocimiento en la toma de decisiones, emulando a los humanos, pero con la ventaja de procesar grandes volúmenes de información de manera simultánea.

En conclusión, aunque no se ha alcanzado una definición consensuada, la Inteligencia Artificial se comprende generalmente como la capacidad de las máquinas para imitar aspectos clave de la inteligencia humana, como el aprendizaje y la toma de decisiones, con la ventaja añadida de minimizar errores y procesar grandes cantidades de información.

### *3.4.2 Historia de la IA en el ámbito de la ciberseguridad*

La Inteligencia Artificial (IA) ha recorrido un largo camino, su historia está marcada por hitos importantes que influyeron en su desarrollo y popularización en el campo tecnológico.

De acuerdo con Ponce Gallegos et al. (2014), desde la antigüedad el ser humano ha intentado crear mecanismos que imiten su propio comportamiento, tanto en apariencia como en funciones. Un ejemplo temprano fue el Perceptrón de Rosenblatt, desarrollado en los años 50, que consistía en un sistema de reconocimiento de patrones. A pesar de que este proyecto tuvo cierto éxito inicial, su desarrollo no se sostuvo en el tiempo.

Un avance significativo surge en el año 1950, como explica Smith (2006), cuando Alan Turing propuso el Test de Turing. Este experimento sugería que una máquina podría considerarse inteligente si lograba imitar el comportamiento humano de manera convincente.

A lo largo de las décadas siguientes, esta disciplina experimenta una evolución constante, ampliando su aplicación a ámbitos cada vez más especializados, por ejemplo, según lo señalado en el sitio web Microsoft (s.f), a partir de los años 80, la IA comienza a tener aplicaciones más directas en la ciberseguridad, utilizando sistemas basados en reglas predefinidas para generar

alertas. A partir de los años 2000, con el auge del aprendizaje automático, es posible analizar grandes cantidades de datos y detectar patrones de comportamiento inusuales. En este sentido, Ayerbe destaca que:

> La IA puede utilizarse para ayudar a los profesionales de la seguridad a tratar la cada vez mayor complejidad de los sistemas modernos de IT, industria 4.0, infraestructura del Internet de las Cosas (IoT por sus siglas en inglés) …, así como la ingente cantidad de datos creados por ellos, e intentar estar por delante de los ciberatacantes. (Ayerbe, 2020, p.2).

En la actualidad, Microsoft (s.f) señala que la Inteligencia Artificial aplicada a la ciberseguridad analiza grandes volúmenes de datos provenientes de diversas fuentes para detectar patrones de actividad en una organización, tales como los momentos y lugares en los que los usuarios inician sesión o los niveles de tráfico. Al establecer un comportamiento típico, es capaz de detectar anomalías que podrían requerir investigación.

De este modo, es posible afirmar que la IA evoluciona de manera constante desde sus primeras concepciones, consolidándose como una herramienta clave para la detección y prevención de ataques digitales.

### *3.4.3 Detección de anomalías*

En el ámbito de la ciberseguridad identificar amenazas en tiempo real constituye un proceso clave en la mitigación de delitos informáticos.

Un elemento importante de la IA en la ciberdefensa es su habilidad para identificar anomalías. Los sistemas que utilizan esta tecnología tienen la capacidad de vigilar el tráfico de red de manera continua, reconociendo patrones atípicos que podrían señalar un delito. La Inteligencia Artificial puede mejorar su capacidad de predicción y adaptarse a nuevas amenazas aprendiendo de experiencias anteriores (Secretaría de Transformación Digital, 2024).

Asimismo, un avance significativo se evidencia también en la protección de datos y la prevención de ataques. Sin embargo, no basta con la detección, en este escenario también es importante detectar:

> Una de las líneas de desarrollo más prometedoras de la IA en ciberseguridad es la aplicación de la primera en la identificación de ataques sofisticados y en la anticipación de tácticas de ciberdelincuentes. Se está trabajando en el desarrollo de IAs que logren una comprensión más profunda de los patrones de comportamiento malicioso y que pueda detectar actividades anómalas de manera más ágil, incluso en entornos donde las tácticas de ataque evolucionan constantemente (Bustio Martínez, 2024, p.4).

En conclusión, la detección de anomalías a través de la Inteligencia Artificial se presenta como un pilar fundamental en el ámbito la ciberseguridad, dado que la capacidad para aprender de experiencias pasadas potencia su efectividad predictiva y adaptativa frente a amenazas emergentes.

## 3.5 Machine Learning y Deep Learning

### *3.5.1 Introducción al ML*

El Machine Learning (ML) es una rama de la Inteligencia Artificial que juega un papel clave en la ciberseguridad debido a que mejora la detección de amenazas.

McKinsey & Company (2024) menciona, en su sitio web, que el aprendizaje automático es capaz de identificar patrones y generar predicciones o recomendaciones mediante el análisis de datos y experiencias, esto sin necesidad de recibir instrucciones programadas de forma directa. Además, se ajustan de forma continua, mejorando su rendimiento con el tiempo.

Ruiz Martínez destaca en su investigación la importancia del papel que desempeña en la ciberseguridad:

> Gracias al aprendizaje automático, un sistema informático tiene la capacidad de poder localizar en grandes cantidades de datos, comportamientos extraños y situaciones anómalas que son conocidas como patrones. El Machine Learning es usado para detectar intrusiones o situaciones fuera de lo normal, que quieran infiltrarse en la red de un sistema. (Ruiz Martínez, 2021, p.46)

En el siguiente apartado, se examinan las distintas variantes del aprendizaje automático y las principales características empleadas para la detección de anomalías.

### *3.5.2 Clasificación de algoritmos de ML*

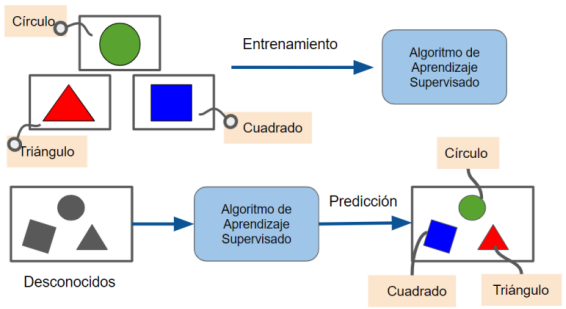
En el campo del aprendizaje automático, los métodos se organizan en distintas categorías diseñadas para abordar la resolución de diversos tipos de problemas. En este sentido, “el aprendizaje automático usa algoritmos para aprender de los patrones de datos” (Rouhiainen, 2018, p.20).

De acuerdo con la clasificación propuesta por Navarro (2024), existen cuatro tipos fundamentales: El aprendizaje supervisado, aprendizaje no supervisado, aprendizaje por refuerzo, aprendizaje semi-supervisado.

#### *3.5.1.1 Aprendizaje supervisado.*

En el aprendizaje supervisado, como expone Guatelli (2023), el algoritmo recibe un conjunto de datos en el que cada objeto tiene atributos y una etiqueta o categoría ya asignada. Su objetivo es aprender o construir un modelo que pueda prever la etiqueta de objetos que no ha visto antes. Como se muestra en la Figura 3 este tipo de aprendizaje se utiliza comúnmente en tareas de clasificación y regresión.

**Figura 3.**
Aprendizaje supervisado




*Nota:* Obtenido de Aprendizaje supervisado, de Guatelli, R., S, 2023, Detección de la Enfermedad de Parkinson basada en Máquinas de Aprendizaje Extremo. [Tesis de Maestría, Universidad Nacional de La Matanza]

Navarro (2024) destaca que es imprescindible entrenar a las máquinas utilizando una base de datos que esté etiquetada con precisión, lo cual permite que puedan predecir resultados

específicos. En otras palabras, se le indica a la máquina lo que debe aprender, y esta debe cumplir con esas indicaciones de manera exacta.

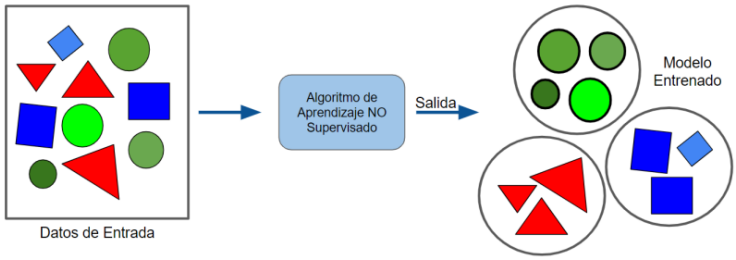
### *3.5.1.2 Aprendizaje no supervisado.*

El aprendizaje no supervisado, se distingue del anterior debido a que no recibe supervisión en ninguna de sus etapas. En este caso, la máquina actúa de manera autónoma y no requiere datos etiquetados, aunque sí necesita un conjunto de datos para generar resultados.

En este proceso, la máquina intentará organizar o agrupar los resultados en etiquetas basadas en las relaciones detectadas. Al mismo tiempo, el algoritmo debe identificar una forma de medir y establecer conexiones entre todos los resultados obtenidos (Navarro, 2024).

Guatelli (2023) señala que el objetivo del algoritmo es desarrollar un modelo que facilite la identificación de patrones o estructuras dentro de los datos. Por ejemplo, puede agrupar objetos similares en conjuntos o clusters, como se muestra en la Figura 4.

**Figura 4.**
Aprendizaje No supervisado




*Nota:* Obtenido de Aprendizaje no supervisado, de Guatelli, R., S, 2023, Detección de la Enfermedad de Parkinson basada en Máquinas de Aprendizaje Extremo. [Tesis de Maestría, Universidad Nacional de La Matanza]

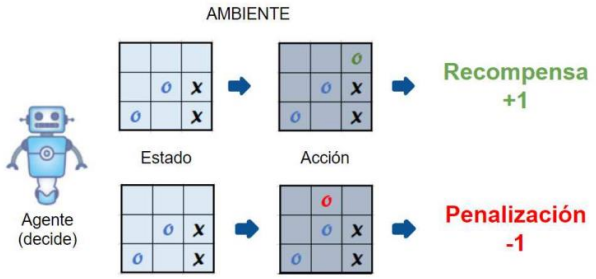
### *3.5.1.3 Aprendizaje por refuerzo.*

El aprendizaje por refuerzo, en cambio se fundamenta, según Navarro (2024), en premiar los comportamientos correctos y penalizar aquellos no deseados. Este método se sustenta en la retroalimentación, ya que la máquina va adquiriendo conocimientos de sus propias experiencias a través de su desarrollo y rendimiento.

Podemos resumir que “el algoritmo interactúa con un entorno y recibe recompensas o penalizaciones por sus acciones. El objetivo es aprender una política que maximice la recompensa total a lo largo del tiempo” (Guatelli ,2023, p.25).

En la Figura 5 podemos observar el sistema de recompensa-penalización.

**Figura 5.**
Aprendizaje por refuerzo




*Nota:* Obtenido de Aprendizaje por refuerzo, de Guatelli, R., S, 2023, Detección de la Enfermedad de Parkinson basada en Máquinas de Aprendizaje Extremo. [Tesis de Maestría, Universidad Nacional de La Matanza]

### *3.5.1.4 Aprendizaje semi-supervisado.*

El aprendizaje semi-supervisado se posiciona entre los dos anteriores, empleando una minoría de datos etiquetados junto con una mayoría de datos no etiquetados. Esta estrategia responde a la necesidad de reducir los costos asociados con el etiquetado completo de los datos. Aunque existe cierta supervisión sobre el funcionamiento de la máquina, esta no se aplica durante todo el proceso. Si bien algunos resultados deben etiquetarse de forma manual, otros serán generados automáticamente por el sistema de machine learning (Navarro, 2024).

En resumen, el Machine Learning, a través de sus diferentes algoritmos, facilita la detección de amenazas dentro de los grandes volúmenes de datos como los que se generan en las redes sociales. Asimismo, en estas plataformas, el análisis rápido de datos es esencial para mitigar ataques como los de ingeniería social. No obstante, a medida que las amenazas evolucionan y crecen en complejidad, se ha vuelto necesario recurrir a nuevos enfoques más sofisticados como el Deep Learning.

### *3.5.2 Introducción al Deep Learning*

El Deep Learning (DL) se ha convertido en una de las técnicas más avanzadas dentro del campo de la IA:

> Una de las aplicaciones más poderosas y de mayor crecimiento de la inteligencia artificial es el aprendizaje profundo (en inglés, deep learning). Se trata de un subcampo del aprendizaje automático que se utiliza para resolver problemas muy complejos y que normalmente implican grandes cantidades de datos. (Rouhiainen, 2018, pp.21-22)

Para una comprensión más precisa del concepto, el sitio web de Amazon (s.f-a) proporciona una explicación clara. En ella se describe que el DL permite a las computadoras procesar datos de manera similar al funcionamiento del cerebro humano. Los modelos de este tipo de aprendizaje pueden identificar patrones complejos en imágenes, textos, sonidos y otros tipos de datos, lo que les permite generar información y predicciones con gran precisión. Esta tecnología se puede aplicar para automatizar tareas que normalmente requieren inteligencia humana, como describir imágenes o convertir archivos de sonido en texto.

El DL emplea redes neuronales multicapa, conocidas como redes profundas, para imitar la capacidad de toma de decisiones del cerebro humano (IBM, 2024).

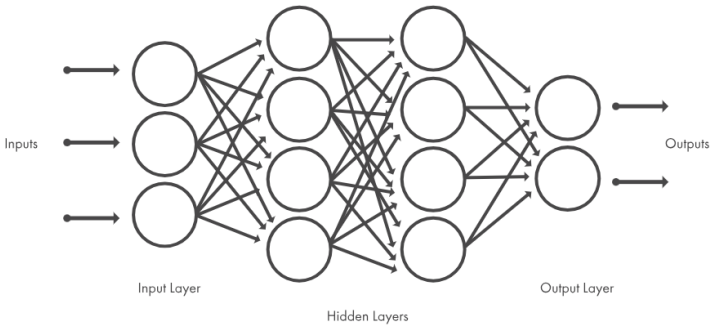
### *3.5.3 Las redes neuronales artificiales*

Tal como se explica en la introducción, el DL utiliza redes neuronales. Amazon (s.f-b) señala que de manera similar al cerebro que contiene millones de neuronas interconectadas que colaboran para aprender y procesar información, las redes neuronales artificiales en DL están compuestas por múltiples capas de neuronas artificiales que trabajan conjuntamente dentro de un sistema informático. Estas neuronas artificiales, también conocidas como nodos, son módulos de software que emplean cálculos matemáticos para procesar los datos. Estos nodos se utilizan para abordar problemas complejos.

En la Figura 6 podemos observar que la red neuronal, está formada por una capa de entrada, capas ocultas y capas de salida. Las ocultas o intermedias son las que aportan profundidad a la arquitectura de un modelo, que puede incluso poseer cientos de capas. Estas, están formadas por neuronas, que reciben múltiples entradas y las procesan para generar un único valor de salida. Por

otro lado, las unidades de entrada, solo se encargan de recoger las variables del modelo sin realizar ninguna transformación. Además, las flechas que conectan las capas tienen únicamente un sólo sentido, dirigiéndose hacia la capa de salida (González Muñiz, 2023).

**Figura 6.**
Ejemplo de arquitectura de una red profunda prealimentada o feedforward




*Nota:* Obtenido de *Introducción a Deep Learning,* por Math Works (s.f). https://la.mathworks.com/discovery/deep-learning.html

## 3.6 Tipos de Modelos de DL

El modelo multicapa permite la identificación y análisis de diversos tipos de redes neuronales, las cuales se detallan a continuación.

### *3.6.1 Redes convolucionales (CNN)*

Entre las distintas clases de redes neuronales que se utilizan en el campo del aprendizaje profundo (DL), se destacan las redes neuronales convolucionales (CNN). Como menciona en su investigación González Muñiz estas últimas:

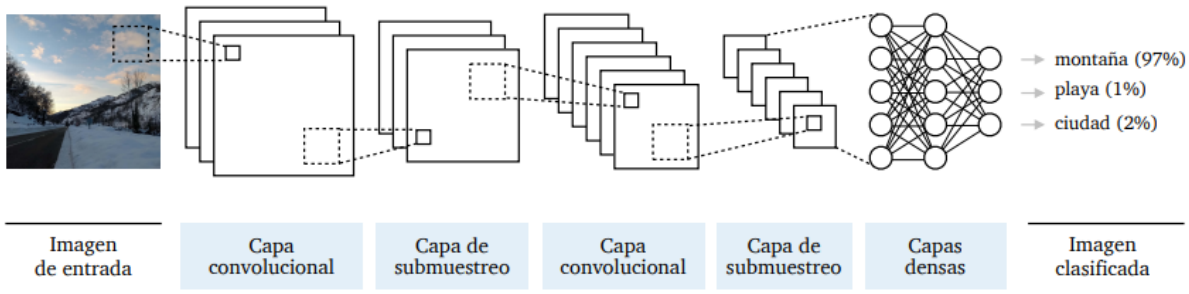
> Se han convertido en uno de los principales exponentes del deep learning, desatacando por su especial habilidad para extraer características de forma automática a partir de los datos de entrada, lo que se conoce como aprendizaje de características o feature learning. (González Muñiz, 2023, p.39)

Las CNN se emplean principalmente en áreas como la visión artificial y clasificación de imágenes. Pueden detectar características y patrones dentro de imágenes y videos, lo que permite

realizar tareas como detección de objetos, reconocimiento de imágenes, reconocimiento de patrones y facial (IBM, 2024).

En la Figura 7 se observan las distintas capas de una arquitectura de red CNN, en este caso puntual, aplicada en la clasificación de imágenes. Las capas convolucionales identifican patrones locales en los datos, a continuación, las capas de submuestreo agrupan características similares para reducir la complejidad y la dimensionalidad, haciéndolas más resistentes a pequeñas distorsiones. Finalmente, las capas densas conectan todas las neuronas para generar una representación global de las muestras. Así, las redes CNN realizan un análisis de alto nivel que permite la clasificación de imágenes (González Muñiz, 2023).

**Figura 7**.
Ejemplo de arquitectura de red CNN




*Nota:* Adaptado de Red Neuronal Recurrente con múltiples capas por A. González Muñiz, 2023, Universidad de Oviedo. https://hdl.handle.net/10651/70370

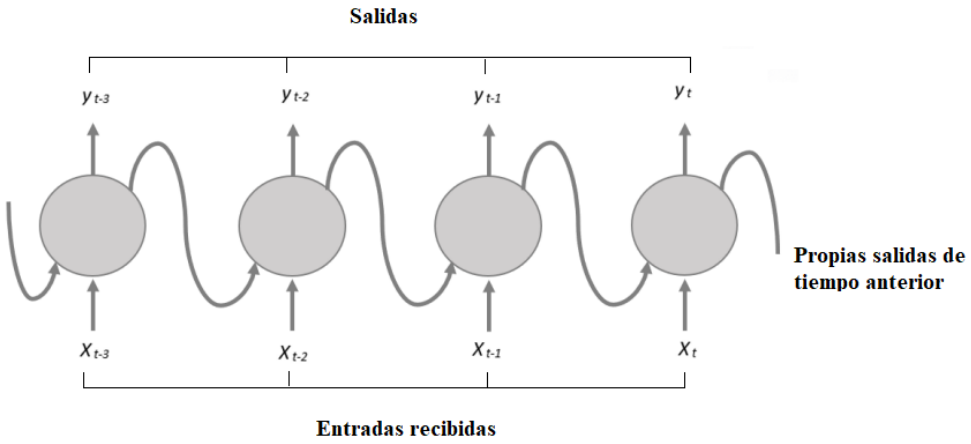
Es importante destacar que este tipo de redes no se limita exclusivamente a la clasificación de imágenes, aunque sobresalen particularmente en esta tarea.

### *3.6.2 Redes Neuronales Recurrentes (RNN)*

Las redes recurrentes (RNN) desempeñan un papel esencial y son ampliamente utilizadas en aplicaciones como el reconocimiento de voz y el procesamiento del lenguaje natural, ya que están diseñadas para manejar datos en secuencia o series temporales. Además, se caracterizan por tener bucles de retroalimentación, lo que les permite utilizar la información pasada para influir en las predicciones futuras, siendo particularmente útiles para realizar predicciones basadas en datos temporales (IBM, 2024).

En lo que respecta a su arquitectura, tal como detalle Gaset en su investigación, son un tipo de red diseñada para analizar datos en series temporales, lo que permite incorporar la dimensión del tiempo. En cada paso temporal, la neurona recurrente recibe tanto la entrada actual $x_t$ de la capa previa como su propia salida del paso temporal anterior $y_{t-1}$, para producir su nueva salida $y_t$ (Gaset, 2022). Esta arquitectura se representa gráficamente en la Figura 8.

**Figura 8.**
Arquitectura de Red Neuronal Recurrente




*Nota:* Adaptado de Red Neuronal Recurrente con múltiples capas por A. Gaset, 2022, Universidad de Torcuato Di Tella https://hdl.handle.net/10651/70370

### *3.6.3 Autocodificadores (Autoencoders)*

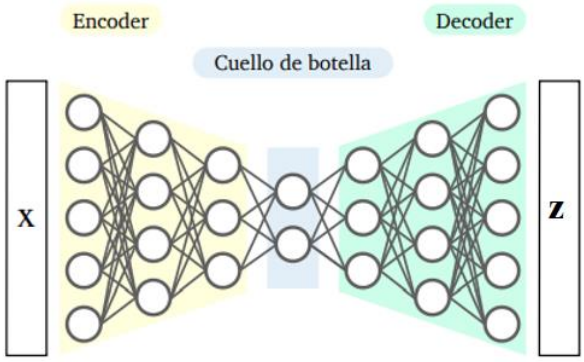
Continuando con las categorías de modelos de Deep Learning, este apartado se centrará en detallar los Autocodificadores.

Los Autoencoders fueron los primeros modelos en producir imágenes y voz realistas transformando datos no etiquetados en una representación comprimida y luego los reconstruyen en su forma original. Los autocodificadores simples se utilizaban para tareas como la restauración de imágenes dañadas, mientras que los autocodificadores variacionales añadieron la capacidad de generar nuevas variaciones a partir de los datos originales (IBM, 2024).

En cuanto a su funcionamiento, González Muñiz explica, tal como se muestra en la Figura 9, que la arquitectura consta de un encoder que genera una representación latente z a partir de los datos de entrada x, y un decoder que reconstruye los datos originales a partir de esa representación. Entre ambos se encuentra un cuello de botella, que evita que el modelo simplemente copie los

datos de entrada, forzándolo a aprender una representación compacta que capture la estructura esencial de los datos, manteniendo la información relevante y descartando la redundante o innecesaria para la reconstrucción (González Muñiz ,2023).

**Figura 9.**
Arquitectura de Autoencoders




*Nota:* Obtenido de Ejemplo de arquitectura de un Deep Autoencoders A. González, Muñiz, 2022, Universidad de Oviedo. https://hdl.handle.net/10651/70370

Una variación conocida son los Autoencoders Variacionales, según explica IBM (2024), incorporan una capacidad esencial que no solo permite reconstruir datos, sino también generar variantes a partir de los datos originales.

### 3.7 Conclusión

En este capítulo se evidencia la relevancia de la ciberseguridad en la era digital, con énfasis particular en las redes sociales. A medida que el cibercrimen evoluciona, también lo hacen las medidas de defensa, destacándose la Inteligencia Artificial (IA) como una herramienta fundamental para mitigar los riesgos.

El Machine Learning (ML) se posiciona como una tecnología preventiva de gran utilidad, ya que permite identificar patrones en grandes volúmenes de datos y mejorar continuamente en la detección de amenazas. Sin embargo, la creciente complejidad de los ciberataques ha impulsado la evolución hacia enfoques más avanzados, como el Deep Learning (DL).

El impacto específico del aprendizaje profundo en la detección de anomalías de comportamiento, así como su relación directa con el fortalecimiento de la ciberseguridad, constituye una cuestión por explorar desde distintas perspectivas.

# Capítulo 4: Enfoque Metodológico

## 4.1 Metodología

El presente trabajo se enmarca en un enfoque cuantitativo de tipo descriptivo, cuyo propósito según Hernández Sampieri (2014) se basa en la recopilación de datos cuantificables que, mediante técnicas estadísticas, permiten comprobar la hipótesis y detectar patrones de comportamiento, contribuyendo así a la validación de teorías existentes.

En este estudio, la técnica utilizada para la recolección de datos es una encuesta estructurada, diseñada conforme a los lineamientos metodológicos propuestos el mismo autor en su obra Metodología de la Investigación (2014). El diseño adoptado es no experimental, dado que este tipo de estudios "se realizan sin la manipulación deliberada de variables y en los que solo se observan los fenómenos en su ambiente natural para analizarlos" (Hernández Sampieri, 2014, p.152).

El alcance es descriptivo, en tanto busca observar, caracterizar y documentar las percepciones y experiencias de los encuestados tal como se manifiestan en la realidad. Este tipo de estudio permite identificar y detallar las cualidades, atributos y perfiles de individuos, grupos u otros fenómenos sometidos a análisis (Hernández Sampieri, 2014).

En este sentido, el presente trabajo se orienta a relevar la percepción de los usuarios sobre la seguridad en las redes sociales, con el fin de analizar desde esa perspectiva cómo los métodos de IA, en particular el aprendizaje profundo, contribuyen a mejorar la detección de anomalías en el comportamiento.

## 4.2 Recolección de Datos

La recolección de datos se realiza mediante un cuestionario autoadministrado, elaborado con la herramienta Google Forms y distribuido por medios digitales. En el diseño se incluyen, tal como propone Hernández Sampieri (2014), una serie de preguntas formuladas en formato abierto y cerrado.

El formulario se trata de un instrumento "completado directamente por los propios participantes, sin intervención de terceros, siendo ellos mismos quienes registran sus respuestas" (Hernández Sampieri, 2014, p. 233).

### 4.3 Preguntas de Investigación

Las preguntas incluidas tienen formato de opción múltiple, valoración tipo Likert y casillas de verificación. Las mismas se encuentra agrupadas en cinco secciones.

La primera sección releva datos generales como edad, nivel educativo y ocupación de los participantes, lo que permite contextualizar las respuestas y analizar si las variables sociodemográficas influyen en la percepción de seguridad digital.

La segunda sección se centra en la experiencia directa de los encuestados con ataques de ingeniería social, permitiendo identificar si han sido víctimas de este tipo de amenazas y si logran reconocer sus características y mecanismos.

En la tercera se exploran las percepciones de los usuarios respecto de las medidas de seguridad automatizadas que se implementan en las redes sociales, así como la efectividad atribuida a dichas herramientas.

La cuarta sección indaga el nivel de familiaridad que poseen los encuestados con el uso de inteligencia artificial en contextos de ciberseguridad, y finalmente, la quinta se orienta a conocer las opiniones sobre conceptos más avanzados, como el de aprendizaje profundo, evaluando tanto el conocimiento declarado como el grado de confianza en estas tecnologías.

### 4.4 Selección de muestra

De acuerdo con la definición de Hernández Sampieri (2014), la muestra corresponde a un subconjunto claramente definido y delimitado de la población objetivo, sobre el cual se recolectan los datos. La selección de la muestra en esta investigación se realizó mediante un muestreo no probabilístico de tipo intencional, ya que “la elección de los elementos no depende de la probabilidad, sino de las características de la investigación” (Hernández Sampieri, 2014, p.176).

La muestra estuvo conformada por empleados de empresas tecnológicas, entre ellas Despegar.com, ITTI S.A., Pedidos Ya y Mercado Libre, así como por alumnos de la Universidad Abierta Interamericana (UAI) pertenecientes a las carreras de Licenciatura en gestión de la Tecnología Informática e Ingeniería en Sistemas, y docentes de dos instituciones de educación técnica: Escuela Nacional Técnica N° 11 Manuel Belgrano y el Colegio Técnico Otto Krause.

Esta combinación de perfiles fue considerada adecuada para contrastar las percepciones de diferentes actores del ecosistema tecnológico: profesionales en actividad con experiencia directa en el uso de herramientas de inteligencia artificial y sistemas de seguridad informática; estudiantes de tecnología, que representan a la nueva generación de futuros especialistas en sistemas; y docentes técnicos, que aportan una perspectiva formativa y pedagógica sobre la integración de estas herramientas emergentes.

Tal como sostiene Hernández Sampieri (2014), los participantes "deben situarse claramente por sus características de contenido, lugar y tiempo" (p. 174). En este sentido, la composición de la muestra contribuye a obtener una visión integral sobre el nivel de conocimiento, la percepción del riesgo y la confianza en soluciones basadas en Deep Learning aplicadas a la detección de anomalías en redes sociales. Asimismo, la heterogeneidad de los participantes fortalece la validez del estudio al incorporar miradas tanto académicas como profesionales, en coherencia con el propósito general de la investigación.

### 4.5 Análisis de los Datos

Los datos recolectados a través del cuestionario se procesan utilizando herramientas de análisis cuantitativo. En una primera instancia, se emplean los gráficos estadísticos generados automáticamente por la plataforma Google Forms, los cuales permiten una visualización preliminar de las respuestas. Posteriormente, se utilizan funciones y tablas dinámicas en Microsoft Excel para profundizar el análisis y establecer relaciones entre variables. Además, se aplica la prueba estadística de Chi-cuadrado con el objetivo de identificar posibles asociaciones significativas entre variables categóricas relevantes para el estudio.

# Capítulo 5: Análisis de Resultados

## 5.2 Presentación de resultados

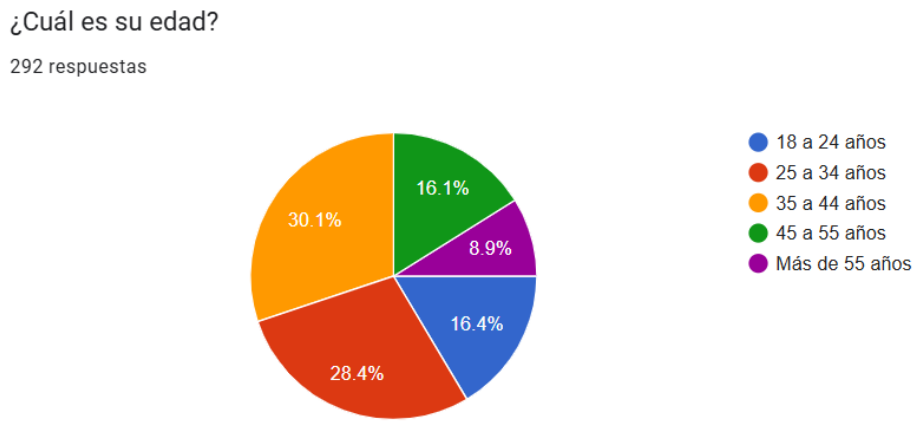
A continuación, se presentan los resultados obtenidos a partir del cuestionario autoadministrado, que ha sido respondido por un total de 292 participantes. Este apartado tiene como propósito brindar una descripción ordenada de los datos recolectados, incorporando un análisis preliminar de algunas preguntas clave. Con posterioridad, en el Capítulo 5, estos resultados son retomados y analizados en mayor profundidad.

### *5.2.1 Edad del Entrevistado*

Esta pregunta tiene como objetivo relevar los primeros datos vinculados al perfil de los encuestados, información necesaria para interpretar adecuadamente las respuestas posteriores, especialmente en relación con la forma en que los usuarios reaccionan ante intentos de ataques.

En la Figura 10 se presentan los resultados obtenidos.

**Figura 10.**
Resultados pregunta N° 1:¿Cuál es la edad del entrevistado?




*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada.

En el gráfico se observa una distribución relativamente equilibrada entre los distintos rangos etarios, con una mayor concentración en los grupos de 35 a 44 años (30,1%) y de 25 a 34

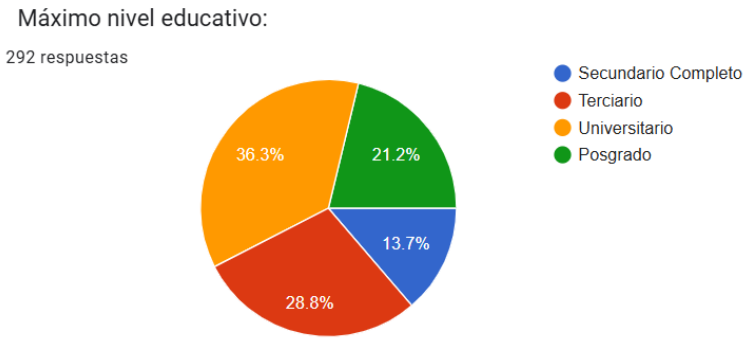
años (28,4%). Les siguen los grupos de 18 a 24 años (16,4%) y de 45 a 55 años (16,1%). Finalmente, el grupo de personas mayores de 55 años representa el porcentaje más bajo, con un 8,9%.

Además de su función descriptiva, este primer dato se retoma el siguiente capitulo para contextualizar las respuestas y analizar si existen patrones de comportamiento diferenciados según el rango etario.

### *5.2.2 Nivel Educativo*

Aunque esta variable no constituye un eje central del análisis, su incorporación permite contextualizar el perfil general de los encuestados. Conocer el nivel máximo de formación académica brinda un marco de referencia sobre el grado de instrucción formal, lo cual puede resultar útil para interpretar ciertos patrones o tendencias en las respuestas. Los resultados obtenidos se exponen en la Figura 11.

**Figura 11.**
Resultados pregunta N°2: Nivel educativo del entrevistado




*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada.

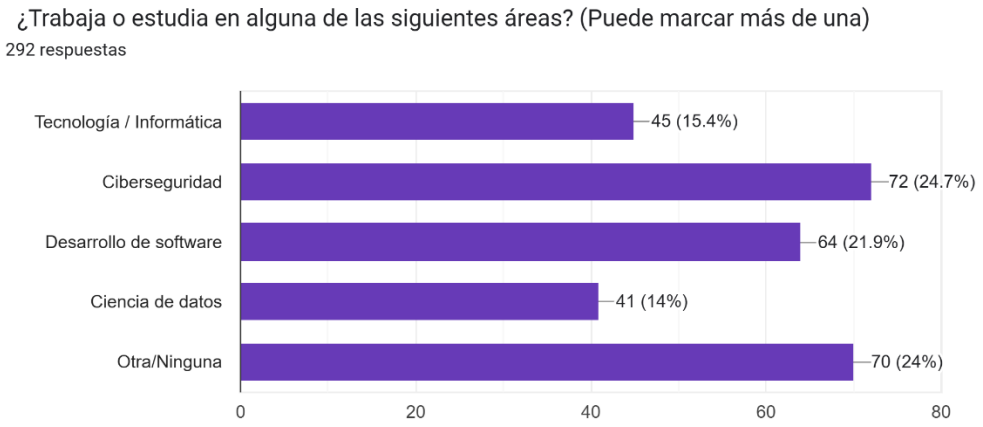
Los resultados indican que la mayoría de los participantes posee estudios universitarios completos (36,3%), seguidos por quienes finalizaron estudios terciarios (28,8%) y quienes cursaron posgrados (21,2%). El grupo con únicamente estudios secundarios completos representa el 13,7% del total. Este perfil refleja una población mayoritariamente formada, lo que podría estar

asociado a un nivel alto de alfabetización digital, algo relevante a la hora de interpretar su familiaridad con conceptos como phishing, detección de anomalías o aprendizaje profundo.

### *5.2.3 Área de Trabajo o Estudio*

Identificar si los participantes se desempeñan en campos como tecnología, ciberseguridad, desarrollo de software o ciencia de datos ofrece indicios sobre su posible familiaridad con los conceptos centrales abordados en este trabajo. Esta variable se tiene en cuenta en los análisis posteriores, con el fin de evaluar en qué medida la formación o el entorno laboral influyen en la percepción de los riesgos y en el nivel de confianza depositado en sistemas automatizados de prevención. A continuación, en la Figura 12 se presenta la evidencia obtenida.

**Figura 12.**
Resultados pregunta N° 3: Área de trabajo o estudio del entrevistado.




*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada.

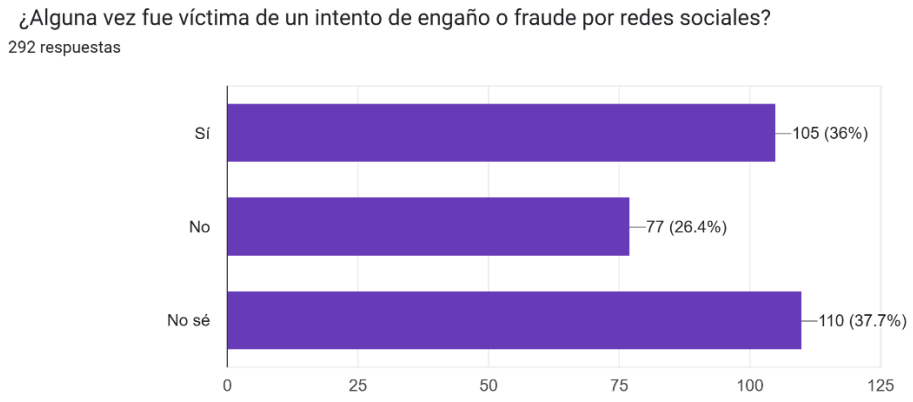
Los resultados indican que el 24,7% de los encuestados declaró trabajar o estudiar en ciberseguridad, el 21,9% en desarrollo de software, el 15,4% en tecnología/informática y el 14% en ciencia de datos. Por su parte, el 24 % manifestó no desempeñarse en ninguna de estas disciplinas.

Este dato adquiere especial relevancia, ya que quienes se encuentran vinculados a áreas técnicas probablemente cuenten con un mayor nivel de conocimiento sobre el funcionamiento de los algoritmos y la aplicación de aprendizaje profundo.

### *5.2.4 Experiencia con Fraudes en Redes Sociales*

La primera pregunta de la Sección B está orientada a introducir al participante en la temática específica de los ataques de ingeniería social. Su formulación, de carácter directo y con opciones de respuesta cerradas, tuvo como finalidad facilitar la participación y establecer una base empírica sobre la experiencia real de los usuarios frente a este tipo de amenazas.

**Figura 13.**
Resultados pregunta N°4: Experiencia del entrevistado con fraudes en redes sociales.



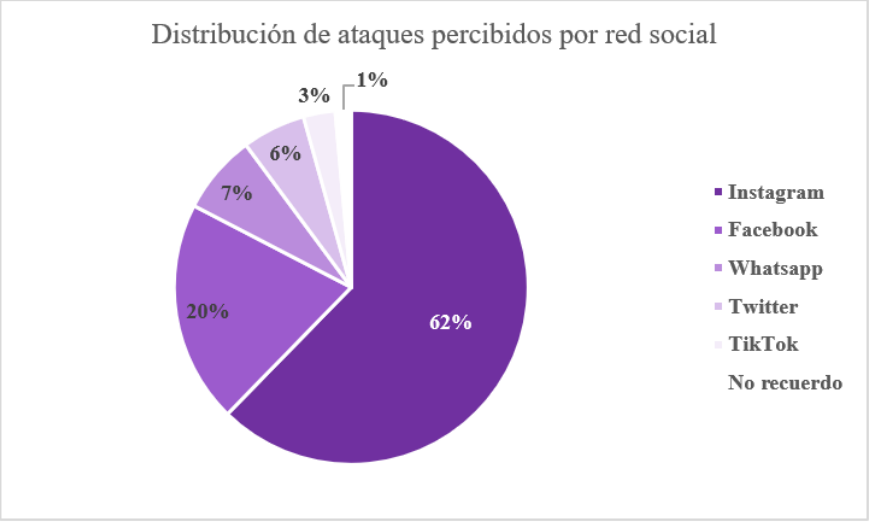

*Nota*: Reproducido de Google Forms generado a partir de la encuesta realizada.

Tal como se observa en la Figura el 36% de los encuestados afirma haber sido víctima de al menos un intento de fraude en redes sociales. Un 26,4% responde negativamente, mientras que el 37,7% indica estar seguro. Este resultado representa uno de los primeros indicios clave del estudio, ya que una proporción considerable de la muestra ha tenido contacto, directo o potencial, con situaciones de engaño.

### *5.2.5 Red Social del Último Intento de Fraude*

Con el objetivo de identificar cuáles son las plataformas más propensas a intentos de ataques de ingeniería social, esta pregunta fue dirigida exclusivamente a quienes manifestaron haber sido víctimas de este tipo de incidentes en la pregunta anterior. Se formuló de manera abierta, sin un listado predefinido de opciones, con el fin de no limitar las respuestas y captar una mayor diversidad de experiencias. Sin embargo, el carácter abierto de la pregunta generó una amplia diversidad en la forma en que los encuestados consignaron el nombre de las plataformas. Por ejemplo, para referirse a Instagram se utilizaron expresiones como “IG”, “Insta” o “Instagram de manera similar, para Facebook se registraron respuestas como “facebook” o “face”. Ante esta dispersión, se procedió a la normalización de la información mediante la herramienta Excel, creando una nueva columna de datos estandarizados que permitió agrupar respuestas equivalentes bajo una misma categoría, lo cual facilitó un análisis ordenado sobre el cual se generó el gráfico que se muestra en la Figura 14.

**Figura 14.**
Resultados pregunta N° 5: Red social del último intento de fraude el entrevistado.




*Nota:* Elaboración propia realizada en herramienta Excel a partir de la normalización de los datos obtenidos en la encuesta.

En el grafico se puede observar que Instagram fue mencionado por el 62% de los encuestados, posicionándose como la red social con mayor frecuencia se perciben intentos de fraude o engaño. En segundo lugar, Facebook concentra el 20% de las respuestas. WhatsApp es identificada en el 7% de los casos, seguida por X (anteriormente denominada Twitter) con un 6%. TikTok representa un 3% de las menciones, mientras que el 1% de las personas no recuerda en qué plataforma ocurrió el intento.

Este resultado coincide con lo planteado por Lentini (2024) quien sostiene que, en lugar de dirigir sus ataques a grandes organizaciones, los ciberdelincuentes apuntan a personas individuales, a quienes les sustraen sus identidades a través de redes sociales como Instagram, Facebook y WhatsApp.

### *5.2.6 Medidas de Seguridad Reconocidas*

La importancia de esta pregunta la explica Kozar (2023), quien sostiene que el uso de inteligencia artificial en ciberseguridad analiza continuamente el tráfico de red y el comportamiento de los usuarios para identificar amenazas incluso desconocidas, lo que refuerza la importancia de comprender qué herramientas automatizadas resultan más visibles.

La pregunta es de carácter abierto, por lo tanto, se ha aplicado el procedimiento de codificación propuesto por Sampieri (2014), que implica identificar patrones recurrentes en las respuestas, asignarles una denominación común y, posteriormente, representarlos mediante valores numéricos o símbolos con el fin de facilitar su análisis comparativo.

En este caso luego de la denominación común se asignan once categorías, sobre las cuales se generó una tabla dinámica en la herramienta Excel para obtener la frecuencia de mención de cada categoría. Dichos resultados se muestran a continuación en Tabla 1.

**Tabla 1.**
Resultados pregunta N° 6: Medidas de seguridad reconocidas por los encuestados

| Código | Categoria | Frecuencia de mención |
|---|---|---|
| 1 | Alertas de comportamiento sospechoso | 66 |
| 2 | Verificación de identidad / autenticación | 42 |
| 3 | Bloqueo automático de cuenta o sesión | 37 |
| 4 | Otras | 34 |
| 5 | No lo sabe / No recuerda | 26 |
| 6 | Detección de comportamiento anómalo con IA | 25 |
| 7 | Detección de fraude | 18 |
| 8 | Control de acceso | 18 |
| 9 | econocimiento de patrones y amenazas específic | 13 |
| 10 | Clasificación automatizada de contenido | 8 |
| 11 | Sistemas de reputación y evaluación de perfiles | 5 |

*Nota:* Elaboración propia realizada con la herramienta Excel los datos obtenidos en la encuesta.

Entre las respuestas procesadas, emergen la detección de comportamiento anómalo con IA (25 menciones), la detección de fraude (18 menciones), así como la segmentación y el control de acceso (18 menciones).

Estas respuestas refuerzan lo desarrollado en el Capítulos 3, donde se explora cómo las redes neuronales y el aprendizaje profundo se están utilizando como herramientas avanzadas para prevenir amenazas digitales. En este sentido, se ha señalado que los sistemas de IA aplicados a la ciberseguridad permiten no solo identificar patrones maliciosos complejos, sino también anticipar tácticas de ataque incluso en entornos altamente dinámicos (Bustio Martínez, 2024).

### *5.2.7 Frecuencia de Mensajes Sospechoso*

La frecuencia con la que los usuarios logran identificar mensajes potencialmente maliciosos en redes sociales constituye un indicador tanto del nivel de exposición a este tipo de amenazas como de su capacidad para reconocer intentos de engaño. A continuación, en la Figura 15 se presentan los resultados obtenidos.

**Figura 15.**
Resultados de pregunta N° 7: Frecuencia en que los encuestados reciben mensajes sospechosos.




*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada.

En los resultados se evidencia una distribución relativamente equilibrada entre las diferentes frecuencias reportadas. Un 22,3 % de los encuestados indica que rara vez detecta mensajes sospechosos, seguido por un 21,2 % los recibe ocasionalmente, un 18,5 % los identifica con frecuencia, un 18,8 % los percibe muy frecuentemente y un 19,2 % afirma no haberlos detectado nunca.

Este panorama se relaciona con lo planteado por Bodnar (2020), quien explica que las redes sociales han permitido a los atacantes perfeccionar sus estrategias, creando perfiles falsos que simulan ser legítimos o incluso imitan identidades reales, lo que contribuye a que los intentos de engaño sean más creíbles y difíciles de identificar.

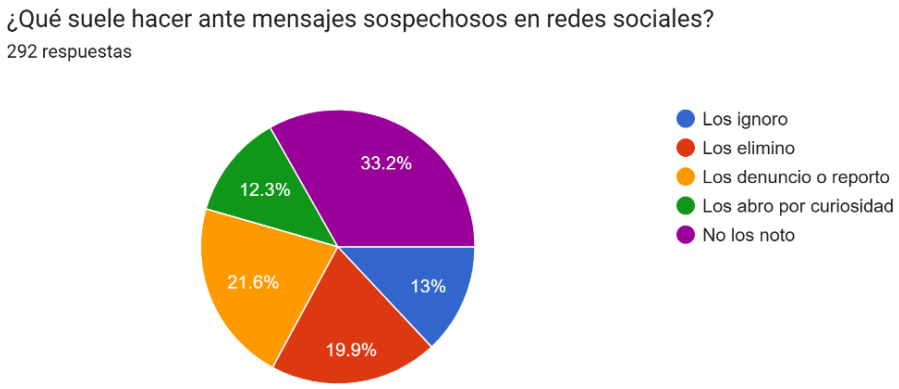
En resumen, la dispersión de respuestas pone de manifiesto que la percepción de riesgo no es homogénea entre los usuarios, lo que refuerza la necesidad de combinar acciones de formación con tecnologías de apoyo que fortalezcan la capacidad de detección frente a amenazas cada vez más disimuladas.

### *5.2.8 Reacción ante Mensajes Sospechosos*

Esta pregunta busca comprender la rección de los usuarios de redes sociales frente a la aparición de mensajes sospechosos, con el fin de identificar las conductas más frecuentes ante posibles amenazas. Los datos obtenidos permiten analizar si las acciones adoptadas contribuyen

o no al fortalecimiento de las estrategias de ciberseguridad, ya sea de forma directa (mediante la denuncia o eliminación del contenido) o indirecta (a través de la omisión o exposición a riesgos). En la Figura 16 se detallan los resultados.

**Figura 16.**
Resultados pregunta N°8: Reacción de los encuestados ante mensajes sospechosos.




*Nota*: Reproducido de Google Forms generado a partir de la encuesta realizada.

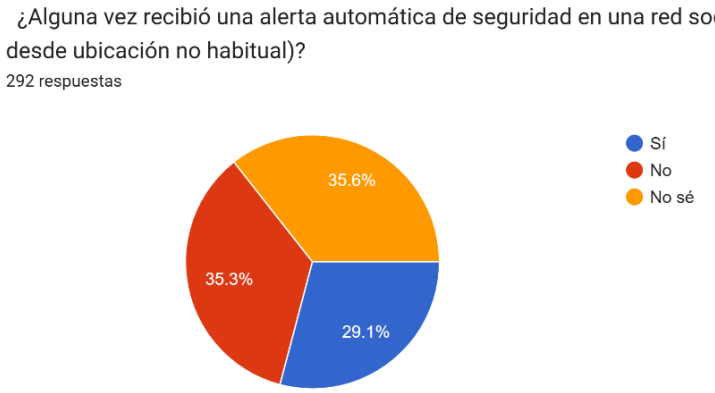
Los resultados muestran que el 33,2 % de los encuestados no percibe los mensajes sospechosos; el 21,6 % afirma que los denuncia o reporta; el 19,9 % opta por eliminarlos; el 13 % los ignora; y el 12,3 % reconoce que los abre por curiosidad. Esta distribución refleja un panorama heterogéneo: por un lado, casi un tercio de los usuarios no logra identificar amenazas potenciales, mientras que otro grupo significativo adopta actitudes pasivas o directamente riesgosas. Pero también se destaca un segmento que actúa de manera preventiva al denunciar o eliminar este tipo de mensajes.

Este conjunto de respuestas se vincula con lo planteado por Berenguer Serrato (2018), quien sostiene que la forma más efectiva de prevenir los ataques de ingeniería social es aprender a reconocerlos cuando ocurren. Desde esta perspectiva, que un 33,2% de los encuestados no detecte señales de alerta, y que un 12,3 % interactúe con estos mensajes por curiosidad evidencia una fragilidad considerable en la capacidad de identificación temprana. Si bien este comportamiento no se analiza en profundidad en este capítulo, se considera un aspecto relevante en el siguiente.

### *5.2.9 Percepción sobre Ciberseguridad e IA*

Esta pregunta introduce la sección orientada a relevar la percepción de los encuestados sobre medidas de seguridad automatizadas, en contraste con la anterior que se enfocaba en las amenazas percibidas. Su foco está en la comparación entre métodos tradicionales y soluciones basadas en inteligencia artificial. De este modo, es posible evaluar no solo la existencia de dichos mecanismos, sino también su nivel de visibilidad y comprensión desde la perspectiva del usuario. A continuación, se detallan los resultados en la Figura 17.

**Figura 17.**
Resultados de pregunta N° 9: Percepción que tiene los encuestados sobre la Ciberseguridad e Inteligencia Artificial




*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada.

Los resultados muestran que el 37,9 % de los encuestados indica haber recibido alguna vez este tipo de alerta, mientras que el 30,7 % responde negativamente y el 31,4 % señala no estar seguro. Esta distribución sugiere que, si bien una parte significativa de los usuarios ha sido efectivamente notificada ante comportamientos inusuales, persiste un amplio margen de desconocimiento o falta de conciencia respecto de la existencia de estas medidas automatizadas.

La relevancia de este hallazgo en el marco de esta investigación se refleja en lo señalado por Kozar (2023), quien sostiene que “el uso de inteligencia artificial en ciberseguridad analiza continuamente el tráfico de red y el comportamiento de los usuarios para identificar amenazas incluso desconocidas, lo que refuerza la importancia de comprender qué herramientas

automatizadas resultan más visibles para los usuarios.” En otras palabras, si bien estas tecnologías operan de forma efectiva en segundo plano, su impacto puede verse reducido si los usuarios no las reconocen o no saben interpretarlas adecuadamente.

### *5.2.10 Preguntas relacionadas a ciberseguridad*

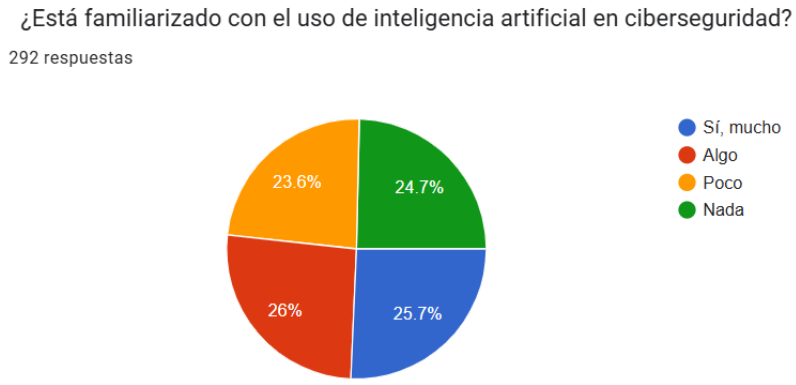
Entre las preguntas del cuestionario también se incluyen cuatro afirmaciones que buscan indagar las percepciones de los participantes sobre aspectos clave del estudio, como la utilidad de las medidas automatizadas, la capacidad de las plataformas para detectar fraudes, y el rol de la inteligencia artificial en este ámbito. Estas afirmaciones fueron evaluadas a través de una escala de Likert de cinco puntos, que permitió captar distintos niveles de acuerdo o desacuerdo. Dado que estas variables están directamente vinculadas con los objetivos e hipótesis del trabajo, su análisis se desarrollará en el apartado 5.1:Percepción de la eficacia sobre medidas de seguridad.

### *5.2.11 Familiaridad con IA en Ciberseguridad*

Relevar la familiaridad de los encuestados con el uso de inteligencia artificial aplicada a la ciberseguridad permite contextualizar las respuestas posteriores, en las que se abordan aspectos más específicos vinculados a tecnologías de aprendizaje profundo en la detección de anomalías.

A continuación, en la Figura 18, se detallan los datos obtenidos en este caso:

**Figura 18.**
Resultados pregunta N°12: Familiaridad con el uso de la IA en ciberseguridad.




*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada.

En el gráfico se observa una distribución equilibrada entre las cuatro categorías de respuesta. El 25,7 % de los encuestados indica estar muy familiarizado con el tema, mientras que un 26 % tiene algún conocimiento. En contraste, el 23,6 % está poco familiarizado, mientras que el 24,7 % afirma no tener ningún conocimiento al respecto.

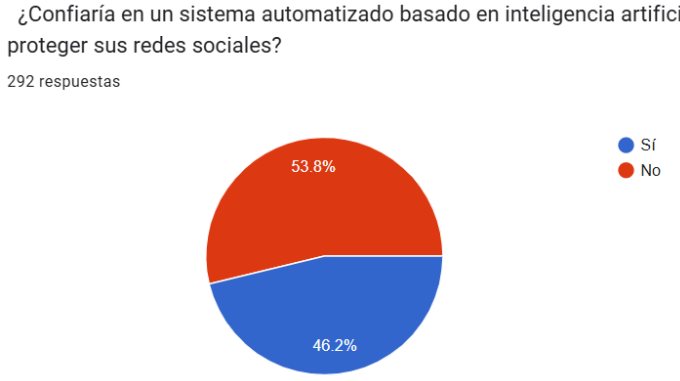
Esta distribución refleja que, si bien una proporción importante de los participantes presenta cierto nivel de familiaridad, más del 48 % se encuentra en los niveles bajos o nulos de conocimiento sobre inteligencia artificial aplicada a la ciberseguridad.

Este resultado se vincula con lo desarrollado por Lubeck (2021), quien advierte que la falta de formación y concientización de los usuarios sigue siendo un factor clave en la exposición a riesgos digitales. La escasa familiaridad con conceptos como el phishing o con herramientas de protección automatizadas puede limitar la capacidad de los usuarios para reconocer amenazas y utilizar de forma consciente las soluciones disponibles.

### *5.2.12 Confianza en la IA para proteger sus redes sociales.*

En este caso, se realiza una pregunta simple cuyo objetivo es introducir a los encuestados en la temática de la eficiencia de la inteligencia artificial aplicada a la protección de redes sociales. A continuación, en la Figura 19, se detallan los datos obtenidos en este caso:

**Figura 19.**
Resultados pregunta N°14: Confianza en Inteligencia Artificial para proteger redes sociales.



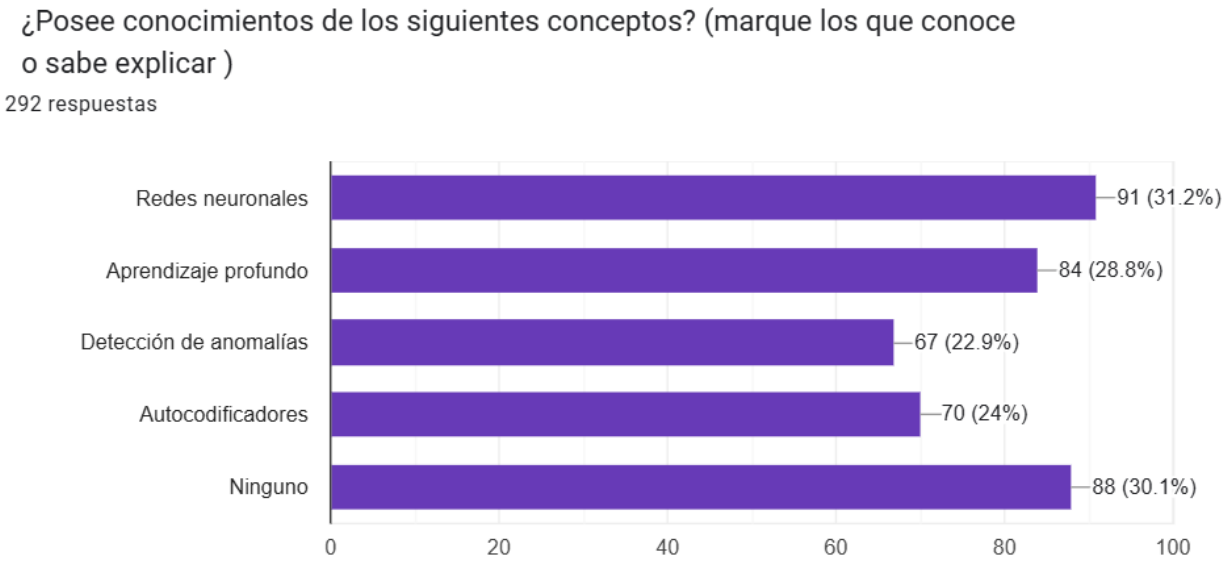

*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada.

Como puede observarse, de forma sorprendente, casi el 54 % de los encuestados no confía en un sistema automatizado basado en inteligencia artificial para proteger sus redes sociales. Esta percepción negativa requiere un análisis más profundo que permita comprender el contexto y las posibles causas de esta desconfianza. Dicho análisis será abordado en el apartado 5.2 del siguiente capítulo.

### *5.2.13 Conocimiento de Conceptos de Aprendizaje Profundo*

Esta pregunta marca la primera instancia en la que se introduce el aprendizaje profundo dentro del análisis empírico del trabajo. Se indaga sobre la familiaridad de los participantes con una serie de conceptos, como redes neuronales, aprendizaje profundo, autocodificadores y detección de anomalías. También se ofrece la opción de no conocer ninguno de estos términos. A continuación, se detallan los resultados en la Figura 20.

**Figura 20.**
Resultados pregunta N°15: Conocimiento de los Conceptos de Inteligencia Artificial.




*Nota*: Reproducido de Google Forms generado a partir de la encuesta realizada

Los resultados obtenidos muestran que el 31.2 % de los encuestados reconoce el concepto de redes neuronales, seguido por un 28.8 % que identifica el término aprendizaje profundo.

Un 24 % conoce los autocodificadores, mientras que la detección de anomalías es mencionada por el 22.9 % de los participantes. Por otro lado, el 30.1 % expresó no haber escuchado ninguno de los conceptos mencionados.

Este panorama sugiere que, si bien hay un sector de la muestra con cierto nivel de familiaridad técnica, el conocimiento sobre aprendizaje profundo no está ampliamente difundido entre todos los encuestados. El hecho de que más de un tercio de los participantes no reconozca estos términos plantea un posible desafío en la comprensión y aceptación de tecnologías automatizadas basadas en este tipo de enfoques.

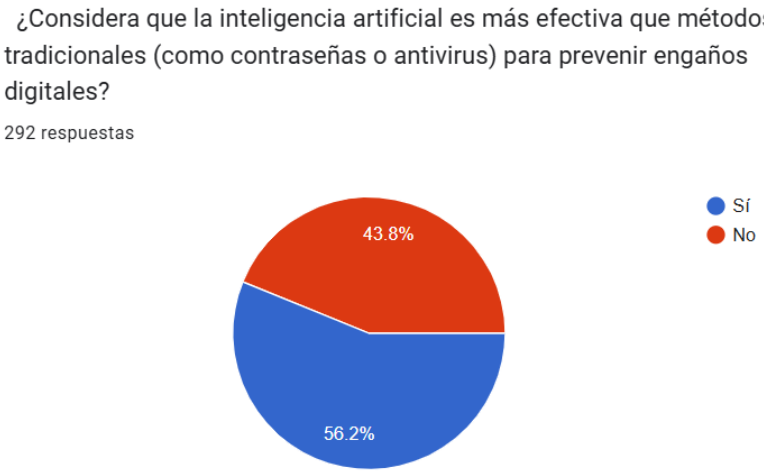
Estos resultados ofrecen una primera base para explorar en qué medida la familiaridad con estos conceptos incide luego en la percepción, confianza y actitud frente a herramientas de detección de anomalías y prevención de fraudes en redes sociales.

En el siguiente capitulo se profundizará está pregunta para entender si aquellos que tienen mayor conocimiento técnico además también confían más en estos sistemas como métodos de detección.

#### *5.2.14 Comparación entre IA y Métodos Tradicionales*

En la última instancia del cuestionario se indaga cómo perciben los encuestados la efectividad de las tecnologías tradicionales de seguridad en comparación con soluciones basadas en inteligencia artificial. En la Figura 21 se exponen los resultados.

**Figura 21.**
Resultados pregunta N°17: Comparación de los métodos de Inteligencia Artificial con los tradicionales




*Nota:* Reproducido de Google Forms generado a partir de la encuesta realizada

Los resultados muestran que el 56,2 % de los participantes considera que la IA es más efectiva que los métodos tradicionales para prevenir engaños digitales.

Esta diferencia evidencia una tendencia moderada pero definida hacia una valoración positiva de los mecanismos basados en inteligencia artificial.

# Capítulo 6: Análisis relacional de la encuesta

Con los resultados del cuestionario ya expuestos en el capítulo anterior, este apartado se dedica a profundizar el análisis de los datos, con el fin de identificar posibles relaciones entre variables que permitan una comprensión más completa del fenómeno estudiado. De este modo, se busca aportar una mirada más integrada que enriquezca los aportes de la investigación.

## 6.1 Percepción de eficacia sobre medida de seguridad automatizadas

A continuación, se analizan las respuestas correspondientes a las afirmaciones incluidas en el apartado N°4.1.9, “Percepción sobre Ciberseguridad en Inteligencia Artificial”. Para este bloque de preguntas se empleó una escala de Likert que, según Hernández Sampieri (2014), está compuesta por un conjunto de afirmaciones ante las cuales los encuestados debían expresar su grado de acuerdo eligiendo entre distintas categorías de respuesta. Con estas cuatro afirmaciones se busca interpretar si los encuestados perciben a la inteligencia artificial como una herramienta eficaz en la detección de comportamientos anómalos y en la prevención de ataques de ingeniería social dentro de las redes sociales. En este caso, la puntuación mínima posible fue de 4 y la máxima de 20. A continuación, se presentan en la Tabla 2, las cuatro afirmaciones utilizadas, junto con las opciones de respuesta ofrecidas a los encuestados.

**Tabla 2:**
Preguntas N° 11, 12, 14, 16 del cuestionario.

| | |
|---|---|
| **¿Considera útiles las medidas automatizadas que brindan las redes sociales ante comportamientos sospechosos?** | |
| 1 = Totalmente en desacuerdo | 4 = De acuerdo |
| 2 = En desacuerdo | 5 = Totalmente de acuerdo |
| 3 = Ni de acuerdo ni en desacuerdo | |
| **¿Cree que las plataformas mejoraron su capacidad para detectar fraudes o comportamientos inusuales?** | |
| 1 = Totalmente en desacuerdo | 4 = De acuerdo |
| 2 = En desacuerdo | 5 = Totalmente de acuerdo |
| 3 = Ni de acuerdo ni en desacuerdo | |
| **¿Qué tan de acuerdo está con la siguiente afirmación? "La inteligencia artificial puede ayudar a detectar comportamientos extraños en redes sociales y prevenir ataques."** | |
| 1 = Totalmente en desacuerdo | 4 = De acuerdo |
| 2 = En desacuerdo | 5 = Totalmente de acuerdo |
| 3 = Ni de acuerdo ni en desacuerdo | |
| **¿Qué tan de acuerdo está con la siguiente afirmación? "Los sistemas que detectan comportamientos anómalos pueden prevenir ataques de ingeniería social."** | |
| 1 = Totalmente en desacuerdo | 4 = De acuerdo |
| 2 = En desacuerdo | 5 = Totalmente de acuerdo |
| 3 = Ni de acuerdo ni en desacuerdo | |

*Nota*: Tabla generada a partir de cuatro preguntas realizadas en formulario con sus respuestas que fueron presentadas en formado de checkbox.

Para calcular las puntuaciones se implementa el procedimiento descripto por el mismo autor, quien señala que “la escala de Likert se construye con una serie de afirmaciones y se asignan valores a las respuestas para luego sumar las puntuaciones de cada sujeto” (Sampieri et al., 2014, p. 252). De este modo, se calcula una puntuación total (PT) individual para cada participante, a partir de la suma de los valores asignados a las cuatro afirmaciones.

Posteriormente, se clasifican las respuestas según su nivel de favorabilidad, y los datos se presentan mediante una tabla dinámica elaborada en Excel. Esta herramienta permite visualizar de forma ordenada la distribución general de las respuestas que se observan en la Tabla 3.

**Tabla 3.**
Resultados de respuestas de preguntas N°11, 12, 14 y 17.

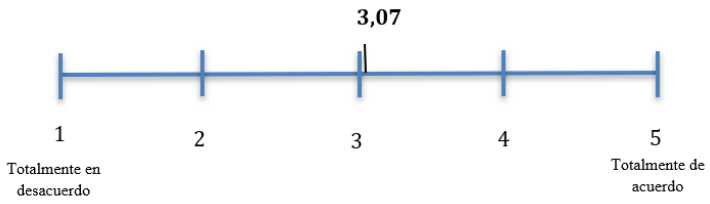
| | Totalmente en desacuerdo (1) | En desacuerdo (2) | Ni de acuerdo ni en desacuerdo (3) | De acuerdo (4) | Totalmente de acuerdo (5) |
|---|---|---|---|---|---|
| Pregunta Nro11 sobre si considera útiles las medidas automatizadas. | 48 | 60 | 56 | 59 | 69 |
| Pregunta Nro 12 sobre si cree que las plataformas mejoraron su capacidad, | 56 | 60 | 65 | 58 | 53 |
| Pregunta Nro 14 sobre si la IA puede detectar comportamientos extraños y prevenir ataques. | 48 | 61 | 58 | 60 | 65 |
| Pregunta Nro 17 sobre los si los sistemas que detectan anomalías pueden prevenir ingeniería social. | 53 | 55 | 56 | 75 | 53 |

Nota: Tabla generada a partir de cuatro preguntas realizada en base a las respuestas del cuestionario.

Una vez obtenidas las puntuaciones individuales de los 292 encuestados, se calcula el promedio general sumando todas las puntuaciones (3589) y dividiendo ese valor por el total de casos (292), lo que arrojó un puntaje promedio de 12,29. Para facilitar su interpretación, se utiliza la fórmula del promedio resultante que "en la escala mediante la sencilla fórmula *PT*/*NT* (donde *PT* es la puntuación total en la escala y *NT* es el número de afirmaciones"(Sampieri, 2014,p.243).

En este caso, el número total de afirmaciones (NT) fue de 4, por lo que al dividir el puntaje promedio (12,29) entre dicha cantidad se obtuvo un promedio de 3,07 por afirmación.

**Figura 22.**
Representación gráfica del NT de las preguntas analizadas




*Nota:* Elaboración propia realizada a partir de los resultados del análisis de las preguntas Nro 11,12, 14 y 17.

Como se puede observar en la Figura 22, el valor 3,07 corresponde a una actitud neutral o levemente favorable hacia las herramientas de seguridad digital impulsadas por la IA. Esto puede estar asociado a que efectividad de las medidas automatizadas puede verse restringida cuando los atacantes utilizan técnicas de manipulación emocional que resultan difíciles de detectar por los

sistemas automáticos. Tal como explica Burdova, “los estafadores crean escenarios falsos pero verosímiles, donde interpretan un papel convincente para manipular emocionalmente al individuo y persuadirlo para que brinde información privada” (2023). Esta capacidad de engaño reduce la eficacia percibida de los mecanismos de protección, especialmente cuando no son comprendidos o reconocidos como tales por los usuarios.

Luego del análisis general sobre la percepción de eficacia de las medidas automatizadas, resulta pertinente indagar si esta valoración varía cuando se consideran únicamente aquellos participantes que manifestaron conocimientos específicos sobre técnicas de aprendizaje profundo, como redes neuronales, redes convolucionales, autocodificadores y detección de anomalías. Este recorte adquiere particular interés en el marco del presente estudio, ya que permite focalizar el análisis en tecnologías capaces de identificar comportamientos atípicos mediante modelos avanzados de aprendizaje automático.

Con este fin, se ha seleccionado la cuarta afirmación de la escala de Likert “Los sistemas que detectan comportamientos anómalos pueden prevenir ataques de ingeniería social” y se ha relacionado con las variables correspondientes a las medidas de seguridad reconocidas (pregunta N°4.1.6: “Medidas de Seguridad Reconocidas”), luego se filtraron aquellos participantes que declararon poseer conocimiento técnico en Deep Learning (pregunta N°4.1.13: Conocimiento de Conceptos de Aprendizaje Profundo). A continuación, Tabla 4 se muestra el resultado del análisis:

**Tabla 4.**
Confianza y detección de alertas según conocimientos en Deep Learning.

| Normalización de Alertas | 1 = Totalmente en desacuerdo | 2 = En desacuerdo | 3 = Ni de acuerdo ni en desacuerdo | 4 = De acuerdo | 5 = Totalmente de acuerdo |
|---|---|---|---|---|---|
| Alertas de comportamiento sospechoso | 15.91% | 34.09% | 9.09% | 22.73% | 18.18% |
| Bloqueo automático de cuenta o sesión | 14.81% | 22.22% | 14.81% | 25.93% | 22.22% |
| Clasificación automatizada de contenido | 50.00% | 0.00% | 0.00% | 25.00% | 25.00% |
| Detección de comportamiento anómalo con IA | 23.08% | 7.69% | 7.69% | 46.15% | 15.38% |
| Modelado predictivo y detección de fraude | 37.50% | 12.50% | 25.00% | 25.00% | 0.00% |
| No lo sabe / No recuerda | 21.43% | 21.43% | 7.14% | 35.71% | 14.29% |
| Otras | 12.90% | 19.35% | 35.48% | 9.68% | 22.58% |
| Reconocimiento de patrones y amenazas específicas | 18.18% | 18.18% | 27.27% | 9.09% | 27.27% |
| Segmentación y control de acceso inteligente | 0.00% | 33.33% | 0.00% | 66.67% | 0.00% |
| Sistemas de reputación y evaluación de perfiles | 50.00% | 50.00% | 0.00% | 0.00% | 0.00% |
| Verificación de identidad / autenticación | 20.83% | 16.67% | 16.67% | 29.17% | 16.67% |
| **Total general** | **18.18%** | **22.46%** | **16.04%** | **25.67%** | **17.65%** |

Nota: Tabla generada a partir de las respuestas del cuestionario.

Entre los resultados resaltan aquellas categorías que refieren a tecnologías como la detección de comportamiento anómalo con IA, el modelado predictivo y detección de fraude. Estas medidas resultan especialmente pertinentes para las variables consideradas en la hipótesis de esta investigación. En este caso se destaca uno de los datos más relevantes para la investigación, el 61,53 % de los encuestados de este grupo particular, es decir con conocimientos en aprendizaje profundo y sus algoritmos, manifestó estar de acuerdo o totalmente de acuerdo con la afirmación: "Los sistemas que detectan comportamientos anómalos pueden prevenir ataques de ingeniería social". Si bien el análisis general evidenció una actitud moderadamente favorable, este recorte permite advertir que la valoración de estas tecnologías mejora cuando los usuarios logran reconocer con mayor claridad su funcionamiento y propósito.

### 6.2 Conocimiento de conceptos de IA vs. percepción de efectividad de la IA

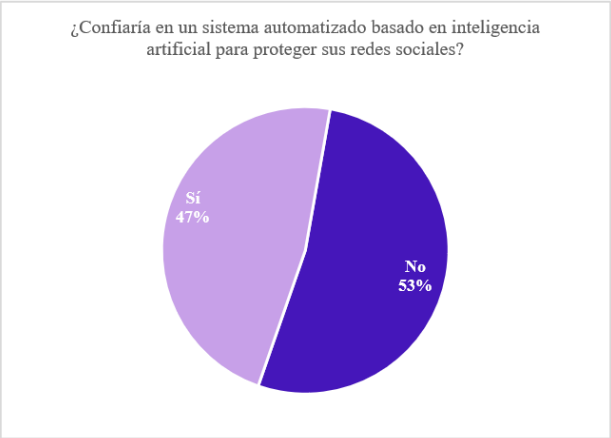
A continuación, se continúa explorando la familiaridad de los encuestados con conceptos relacionados al Deep learning, pero en este caso abordando su relación con el nivel de confianza que manifiestan frente al uso de inteligencia artificial para la protección de sus redes sociales. Para ello, nuevamente se tomaron los datos correspondientes a la pregunta del punto N°4.1.12: "Conocimiento de Conceptos de Aprendizaje Profundo" seleccionando únicamente a aquellos participantes que indicaron conocer al menos uno de estos conceptos. El filtrado de respuestas se realiza mediante tablas dinámicas en Excel. De las 292 respuestas totales, 109 casos en los que los encuestados no reconocieron ninguno de los tres conceptos mencionados, por lo tanto, quedaron fuera de este análisis.

Posteriormente, se cruzaron las respuestas de la población objetivo con la pregunta detallada en el apartado N° 4.1.12, "Confianza en la IA para proteger sus redes sociales".

La información resultante se visualiza en el gráfico de la Figura 23.

**Figura 23.**

Confianza en un sistema basado en IA para proteger las redes sociales.

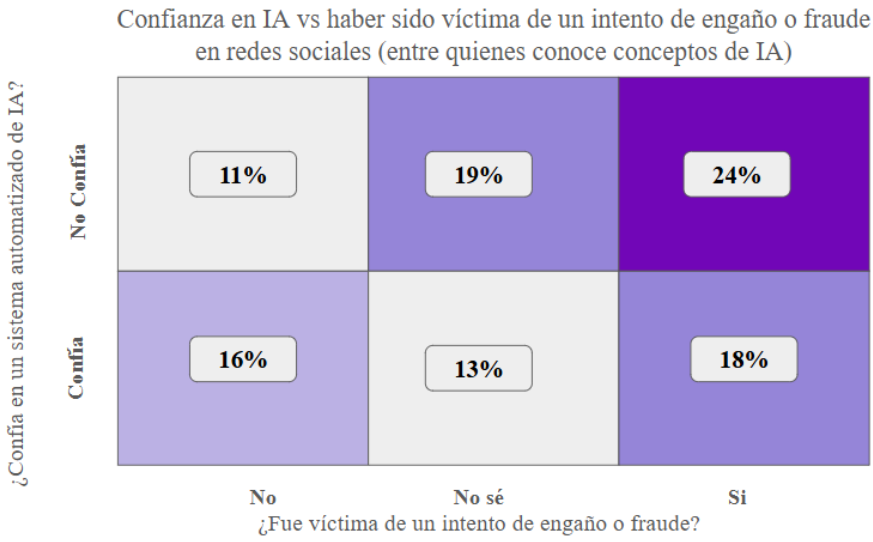



*Nota*: Elaboración propia realizado a partir del cruce respuestas de la encuesta.

La relación entre el conocimiento técnico y la confianza en sistemas automatizados basados en inteligencia artificial arroja un resultado significativo. Entre los encuestados que conocen conceptos como redes neuronales, aprendizaje profundo y autocodificadores, el 53 % no confiaría en este tipo de sistemas para proteger sus redes sociales. Este dato resulta particularmente relevante, ya que pone de manifiesto que la familiaridad con los fundamentos técnicos de la inteligencia artificial no necesariamente se traduce en una mayor aceptación o confianza en su aplicación práctica, especialmente en ámbitos sensibles como la ciberseguridad personal.

Con el propósito de profundizar en la interpretación del resultado anterior, es necesario incorporar una tercera variable correspondiente a la pregunta detallada en el apartado N°4.1.4 sobre la experiencia previa como víctima de un intento de fraude o engaño a través de redes sociales. Esto ayudara a comprender si el hecho de haber atravesado una situación de vulnerabilidad digital puede influir de manera significativa en la percepción sobre la efectividad de los sistemas automatizados. Esta nueva variable fue integrada en la tabla dinámica elaborada en Excel, y los resultados obtenidos fueron representados visualmente en la Figura 24 mediante un mapa de calor, diseñado en una presentación de Google Slides.

**Figura 24.**
Mapa de calor sobre la confianza en los sistemas de IA.




*Nota*: Elaboración propia realizado a partir del cruce respuestas de la encuesta.

El análisis relacional representado en la Figura 24 revela un hallazgo significativo para la investigación. Entre los encuestados que declararon poseer conocimientos sobre conceptos clave vinculados al Deep Learning, se observa que un 43% de quienes no confían en los sistemas automatizados también fueron víctimas (24%) o no están seguros (19%) de haber sufrido un intento de engaño o fraude en redes sociales.

Para validar estadísticamente esta relación observada, se ha realizado una prueba de Chi-cuadrado de independencia entre las variables "experiencia previa como víctima de un intento de fraude o engaño " y "Confianza en un sistema basado en IA para proteger las redes sociales". El análisis fue realizado mediante la plataforma online DataTab, a la cual se ha adjuntado el conjunto de respuestas correspondiente a ambas preguntas. Los resultados del análisis se presentan a continuación en las Tablas 5, 6, y 7.

**Tabla 5.**
Frecuencias observadas para variables perfectamente independientes

| | | IA para prevenir ataques | | | | | |
|---|---|---|---|---|---|---|---|
| | | 4 = De acuerdo | 2 = En desacuerdo | 3 = Ni de acuerdo ni en desacuerdo | 1 = Totalmente en desacuerdo | 5 = Totalmente de acuerdo | Total |
| Victima de ataque | Sí | 22 | 26 | 14 | 11 | 32 | 105 |
| | No | 15 | 9 | 14 | 23 | 16 | 77 |
| | No sé | 23 | 26 | 30 | 14 | 17 | 110 |
| | Total | 60 | 61 | 58 | 48 | 65 | 292 |

*Nota:* Reproducido de DataTab. (s.f). *Prueba Chi-cuadrado.* https://datatab.es/tutorial/chi-square-test

En la Tabla 6 se muestran las frecuencias esperadas bajo el supuesto de independencia entre las variables, es decir, lo que se esperaría si no existiera relación alguna entre la experiencia de fraude y la percepción sobre IA.

**Tabla 6.**
Frecuencias esperadas para variables perfectamente independientes

| | | IA para prevenir ataques | | | | | |
|---|---|---|---|---|---|---|---|
| | | 4 = De acuerdo | 2 = En desacuerdo | 3 = Ni de acuerdo ni en desacuerdo | 1 = Totalmente en desacuerdo | 5 = Totalmente de acuerdo | Total |
| Victima de ataque | Sí | 21.58 | 21.93 | 20.86 | 17.26 | 23.37 | 105 |
| | No | 15.82 | 16.09 | 15.29 | 12.66 | 17.14 | 77 |
| | No sé | 22.6 | 22.98 | 21.85 | 18.08 | 24.49 | 110 |
| | Total | 60 | 61 | 58 | 48 | 65 | 292 |

*Nota*: Reproducido de DataTab. (s.f). *Prueba Chi-cuadrado*. https://datatab.es/tutorial/chi-square-test

Por último, en la Figura 25 se presenta el resultado estadístico final de la prueba de Chi-cuadrado, que incluye el valor del estadístico calculado, los grados de libertad y el p-valor obtenido.

**Figura 25.**
Resultado prueba chi-cuadrado

Prueba Chi-cuadrado

| | $Chi^2$ | df | p |
|---|---|---|---|
| Victima de ataque - IA para prevenir ataques | 26.93 | 8 | .001 |

Nota: Reproducido de DataTab. (s.f). *Prueba Chi-cuadrado*. https://datatab.es/tutorial/chi-square-test

El resultado de la prueba es estadísticamente significativo, con un valor de $X^2$=26.93, grados de libertad = 8, y un p-valor = 0.001. Esto indica que existe una relación significativa entre haber sido víctima de un intento de fraude digital y la percepción de la eficacia de los sistemas de inteligencia artificial para prevenir este tipo de ataques.

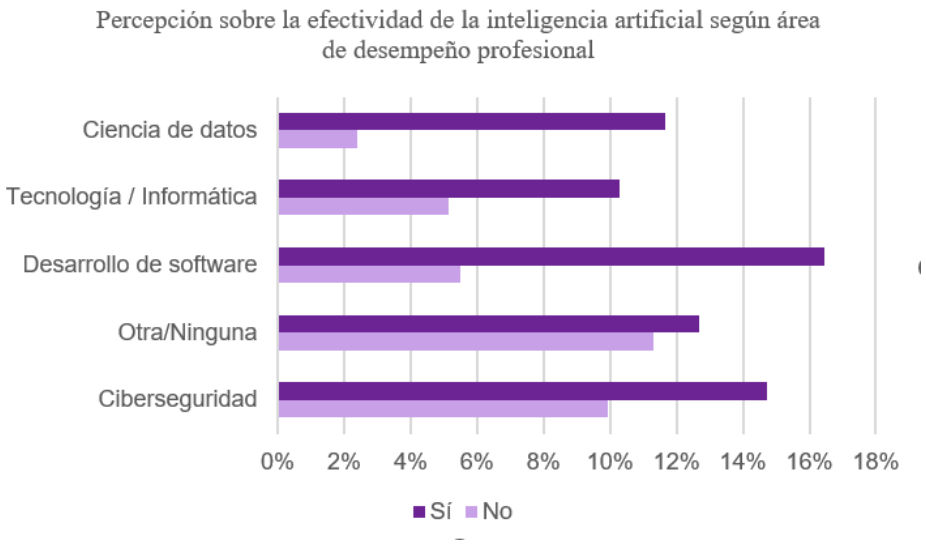
Este nuevo contexto permite replantear la conclusión preliminar de este apartado, que sugería que un mayor conocimiento técnico podría estar asociado a una menor confianza en la inteligencia artificial. Lo que realmente se pone de manifiesto es que la variable más influyente en la percepción de eficacia de estos sistemas es la experiencia directa de haber atravesado una situación de vulnerabilidad digital. Esto revela una barrera adicional vinculada a la percepción: incluso si un sistema funciona correctamente, su efectividad puede verse limitada si no es percibido como confiable.

### 6.3 Ocupación en tecnología o ciberseguridad vs. percepción sobre la capacidad de la IA para detectar comportamientos anómalos

En este apartado el análisis examina la percepción sobre la efectividad de la inteligencia artificial en función del área profesional en la que se desempeñan los encuestados. La comparación resulta pertinente ya que aporta evidencia empírica sobre el grado de aceptación de las soluciones considerando el contexto formativo y profesional del usuario.

Se utilizó nuevamente la herramienta de tabla dinámica de Excel, a fin de cruzar las respuestas correspondientes a la pregunta detallada en el apartado N° 4.1.3: Área de trabajo o estudio y la pregunta N° 4.1.14 Comparación entre IA y métodos tradicionales. En la Figura 25 se exponen los resultados.

**Figura 26.**
Gráfico sobre la percepción de la efectividad de la IA por área profesional.




*Nota:* Elaboración propia realizado a partir del cruce respuestas de la encuesta.

Se observa que los mayores niveles de aceptación hacia la inteligencia artificial se concentran en los ámbitos de ciencia de datos, ciberseguridad y desarrollo de software (43% en total).

En particular, dentro del área de ciencia de datos, se registra una diferencia significativa entre quienes confían en la IA (12 %) y quienes no (2 %). Este cruce de variables permite inferir que, en los sectores más vinculados al desarrollo tecnológico y a la ciberseguridad, la inteligencia artificial es percibida como una solución más eficaz frente a los enfoques convencionales.

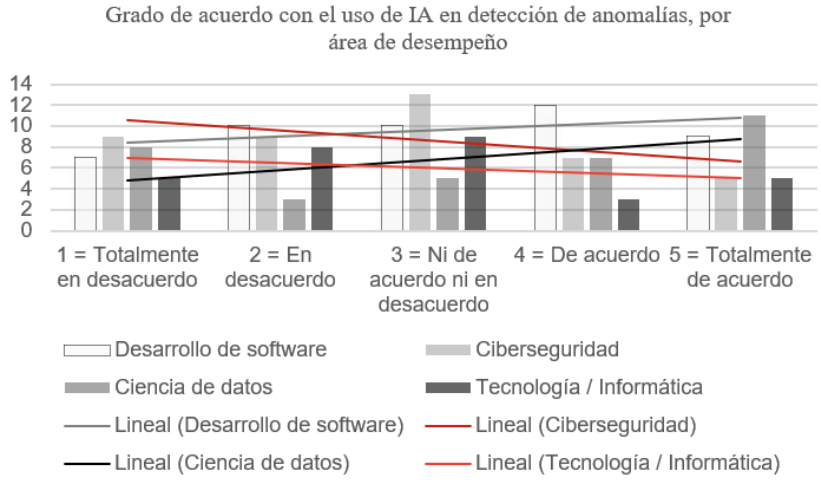
Los resultados sugieren que los perfiles profesionales con mayor exposición al uso del aprendizaje profundo y a sistemas automatizados de análisis de datos tienden a reconocer con mayor claridad los beneficios de estas tecnologías.

Con el fin de explorar este contexto, se incorpora la tercera afirmación de la escala Likert presentada en el apartado N°5.1 "La inteligencia artificial puede ayudar a detectar comportamientos extraños en redes sociales y prevenir ataques".

Para incluir esta tercera variable, se aplica un filtro en la tabla dinámica de Excel que permite conservar únicamente a los encuestados que manifestaron confianza en la inteligencia

artificial. A partir de esta segmentación, se elaboró el siguiente gráfico (Figura 27), en el que se destacan las líneas de tendencia según el área profesional de los participantes.

**Figura 27**.
Gráfico de acuerdo del uso de la IA en detección de anomalías por área de desempeño.




*Nota:* Elaboración propia realizado a partir del cruce respuestas de la encuesta.

En la figura 27 se puede evidenciar que los profesionales de ciberseguridad y tecnología/informática presentan una inclinación hacia el desacuerdo con la afirmación de que la IA puede ayudar a detectar comportamientos extraños en redes sociales. Esta postura podría explicarse por una evaluación crítica fundamentada en la experiencia profesional: si bien reconocen que la inteligencia artificial puede resultar más efectiva que los métodos tradicionales, consideran que aún no ha alcanzado el nivel de madurez necesario para detectar delitos cibernéticos o prevenir ataques con la precisión requerida.

En contraposición, los profesionales vinculados a las áreas de ciencia de datos y desarrollo de software manifiestan una tendencia más favorable respecto de la utilidad de la inteligencia artificial en la prevención de ataques digitales. Esta diferencia podría atribuirse a su mayor proximidad con los modelos predictivos y con el desarrollo de sistemas basados en IA, lo que genera una percepción más positiva sobre su aplicación concreta en el ámbito de la ciberseguridad.

## 6.4 Recepción de alertas automáticas, reacción ante mensajes sospechosos y frecuencia de exposición

En esta sección el enfoque se orienta hacia el comportamiento que adoptan los usuarios frente a ataques sufridos. Para ello, se analizan conjuntamente tres variables extraídas del cuestionario: la frecuencia con la que los participantes detectan mensajes o contenidos sospechosos (apartado N°4.1.7) , la reacción declarada ante estos casos (apartado N°4.1.8), y la experiencia previa de haber recibido alertas automáticas de seguridad por parte de alguna red social (apartado N°4.1.5).

Estas variables, evaluadas de forma conjunta, permiten incorporar las respuestas concretas que los usuarios adoptan ante recomendaciones o advertencias del sistema. Tal como se señaló en el Marco Teórico, este aspecto resulta relevante según lo expuesto por McKinsey & Company (2024) sostiene que los sistemas de machine learning mejoran su rendimiento con el tiempo a partir del análisis de datos y experiencias. En este caso la reacción de los usuarios frente a los posibles ataques.

A continuación, se presentan tres tablas, cada una agrupa a los participantes según su respuesta sobre la recepción de alertas automáticas de seguridad: quienes afirmaron haberlas recibido, quienes no, y quienes no recuerdan haberlo hecho, combinando su rección en estas situaciones.

**Tabla 7.**
Comportamiento ante mensajes sospechosos. Usuarios que recibieron alertas automáticas.

| | Actitud | | | | |
|---|---|---|---|---|---|
| Frecuencia | Los abro por curiosidad | Los denuncio o reporto | Los elimino | Los ignoro | No los noto |
| A veces | 4 | 3 | 5 | 1 | 1 |
| Frecuentemente | 1 | 7 | 3 | 3 | 2 |
| Muy frecuentemente | 2 | 3 | 1 | | 2 |
| Nunca | 1 | | 1 | 3 | 9 |
| Rara vez | 4 | 6 | 10 | 5 | 8 |
| Total general | 12 | **19** | **20** | 12 | 22 |

*Nota*: Elaboración propia a partir de los datos de la encuesta realizada.

**Tabla 8.**
Comportamiento ante mensajes sospechosos. Usuarios que no recibieron alertas automáticas.

| | Actitud | | | | |
|---|---|---|---|---|---|
| Frecuencia | Los abro por curiosidad | Los denuncio o reporto | Los elimino | Los ignoro | No los noto |
| A veces | 4 | 4 | 6 | 5 | 8 |
| Frecuentemente | 3 | 5 | 4 | 6 | 4 |
| Muy frecuentemente | 1 | 4 | 5 | 6 | 6 |
| Nunca | 0 | 1 | 2 | 0 | 16 |
| Rara vez | 0 | 4 | 4 | 3 | 2 |
| Total general | 8 | 18 | 21 | 20 | 36 |

*Nota:* Elaboración propia a partir de los datos de la encuesta realizada.

**Tabla 9.**
Comportamiento ante mensajes sospechosos. Usuarios que no saben si recibieron alertas automáticas.

| | Actitud | | | | |
|---|---|---|---|---|---|
| Frecuencia | Los abro por curiosidad | Los denuncio o reporto | Los elimino | Los ignoro | No los noto |
| A veces | 4 | 4 | 6 | 5 | 8 |
| Frecuentemente | 3 | 11 | 4 | 0 | 4 |
| Muy frecuentemente | 1 | 8 | 5 | 2 | 6 |
| Nunca | 0 | 1 | 2 | 0 | 16 |
| Rara vez | 0 | 4 | 4 | 3 | 2 |
| Total general | 8 | 28 | 21 | 10 | 36 |

*Nota*: Elaboración propia a partir de los datos de la encuesta realizada.

El análisis comparativo entre los tres grupos evidencia que la interacción previa con sistemas de seguridad automatizados influye de manera directa en el comportamiento de los usuarios. Aquellos que afirman haber recibido alertas tienden a adoptar respuestas más activas y preventivas, como denunciar o eliminar los mensajes sospechosos, tal como se resalta en los resultados generales de la Tabla 7. En contraste, entre quienes no han recibido alertas o no recuerdan haberlas recibido, se observa una mayor proporción de respuestas pasivas o de inacción, tales como ignorar el contenido o no detectarlo (Tabla 8, resultados resaltados).

Mientras que en la Tabla 9 resulta coherente que quienes indican no saber si recibieron alertas de seguridad presenten,  presenten un número elevado de respuestas que reflejan una falta de percepción.

Los resultados obtenidos en este estudio permiten suponer que el contacto con alertas automatizadas no solo cumple una función informativa, sino que también promueve una respuesta más comprometida por parte del usuario frente al riesgo.

A continuación, se ha incorporada una nueva variable al apartado, (Capitulo 4, sección 4.1.4) el uso de las plataformas de redes sociales en las cuales los usuarios reconocieron haber sido víctimas de un engaño o fraude con la recepción de alertas de seguridad. Esta incorporación permite contextualizar los comportamientos analizados, ya que cada red social presenta distintos niveles de exposición, mecanismos de detección y formas de interacción que pueden influir en la percepción y respuesta de los usuarios frente a alertas de seguridad.

En la Tabla 10, se presentan los resultados expresados en porcentajes, agrupados según el total de respuestas normalizadas por cada plataforma.

**Tabla 10.**
Tabla sobre Red Social y recepción de alerta automática.

| Red social | Alerta automatica de seguridad | | |
|---|---|---|---|
| | No | No sé | Sí |
| Instagram | 28% | 44% | 28% |
| Facebook | 36% | 36% | 29% |
| Whatsapp | 60% | 20% | 20% |
| Twitter | 50% | 25% | 25% |
| TikTok | 0% | 50% | 50% |
| Total general | 33% | 39% | 28% |

*Nota*: Elaboración propia a partir de los datos de la encuesta realizada.

Los resultados obtenidos muestran que la red social Instagram presenta un porcentaje elevado (72 %) de usuarios que no recibieron o no recuerdan haber recibido alertas automáticas de seguridad. Este dato cobra especial relevancia si se considera que, en una pregunta anterior, los encuestados identificaron a Instagram como la plataforma en la que más recordaban haber sido víctimas de intentos engaño.

Facebook, WhatsApp y Twitter reflejan una tendencia similar, aunque con diferencias porcentuales menores. En contraste, TikTok exhibe un comportamiento distinto: el 50 % de los usuarios declaró haber recibido alertas automáticas de seguridad, lo que representa el valor más alto entre las plataformas analizadas.

Esta dualidad de resultados inferir que la efectividad de las estrategias de prevención como las notificaciones por accesos desde ubicaciones inusuales o la detección de mensajes sospechosos, no puede evaluarse de manera uniforme. Es necesario considerar las particularidades de cada red

social, ya que las características propias de la plataforma y los hábitos de uso de sus usuarios influyen directamente en la percepción, recepción y respuesta frente a los mecanismos de seguridad automatizados.

### 6.5 Edad y Reacción ante mensajes sospechosos

Con el objetivo de analizar si el comportamiento ante intentos de fraude varía según el rango etario del participante, se relacionan en este apartado, la variable "edad"(Capitulo 4, sección N°4.1.1) con las respuestas brindadas a la pregunta sobre reacciones frente a mensajes sospechosos (Capítulo 4, sección 4.1.8).

En la Tabla 11 se presentan los resultados.

**Tabla 11.**
Cruce de edad con reacción ante mensajes sospechosos.

| Edad | Los abro por curiosidad | Los denuncio o reporto | Los elimino | Los ignoro | No los noto |
|---|---|---|---|---|---|
| 18 a 24 años | 5 | 9 | 9 | 12 | 13 |
| 25 a 34 años | 8 | 24 | 15 | 8 | 28 |
| 35 a 44 años | 6 | 16 | 20 | 10 | 36 |
| 45 a 55 años | 11 | 10 | 8 | 6 | 12 |
| Más de 55 años | 6 | 4 | 6 | 2 | 8 |
| Total general | 36 | 63 | 58 | 38 | 97 |

*Nota:* Tabla generada en base a las respuestas obtenidas del formulario.

Los datos permiten concluir que las reacciones ante mensajes sospechosos varían significativamente según la edad. Los adultos entre 35 y 44 años presentan los comportamientos más preventivos, siendo quienes más eliminan o denuncian mensajes. En cambio, los jóvenes de 18 a 24 años y los mayores de 55 años son los grupos más vulnerables: los primeros por su alta exposición combinada con baja detección, y los segundos por su pasividad frente a amenazas. El grupo de 25 a 34 años presenta un perfil ambiguo, ya que, aunque es el que más reporta comportamientos sospechosos, también es el que tiende en mayor medida a ignorar posibles riesgos. Finalmente, los adultos de 45 a 55 años son los que más abren mensajes por curiosidad, lo cual representa una conducta de alto riesgo frente a la ingeniería social.

Considerando los resultados se puede interpretar que las reacciones ante posibles amenazas no son homogéneas entre los grupos etarios. Este hallazgo permite inferir que los sistemas automatizados de detección y alerta deberían adaptarse al perfil del usuario, considerando la edad como una variable relevante para personalizar la experiencia de seguridad en redes sociales.

# Capítulo 7: Conclusiones

A lo largo de este trabajo se ha relevado que la eficacia en la prevención de ataques de ingeniería social en las distintas plataformas digitales se relaciona de forma directa con la aplicación de métodos de aprendizaje profundo, en particular redes neuronales convolucionales, recurrentes y autocodificadores. A partir del análisis de 292 respuestas obtenidas de la encuesta distribuida y del estudio entre variables, se recolectaron datos que permiten validar parcialmente la hipótesis formulada.

Los resultados confirman lo planteado en la justificación de esta investigación, donde se advertía un crecimiento sostenido de los ataques de ingeniería social. En línea con esto, el 36 % de los encuestados afirmó haber sido víctima de al menos un intento de fraude en redes sociales y un 37,7 % manifestó no estar seguro, lo que sugiere que una parte significativa pudo haber sido víctima incluso sin detectarlo.

Uno de los hallazgos más alentadores fue que el 61,53 % de los participantes expresó estar de acuerdo o totalmente de acuerdo con que los sistemas automatizados basados en IA pueden prevenir exitosamente ataques mediante la detección de comportamientos anómalos. Además, quienes se desempeñan en áreas como ciberseguridad, ciencia de datos y desarrollo de software manifestaron mayores niveles de confianza en estas herramientas, especialmente dentro del área de ciencia de datos, donde la diferencia fue significativa (12 % confía frente a 2 % que no).

Sin embargo, la relación entre conocimiento técnico y confianza no resultó favorable en la percepción de eficacia. El 53 % de quienes reconocen conceptos como redes neuronales, aprendizaje profundo y autocodificadores indicó que no confiaría en estos sistemas para proteger sus perfiles de redes sociales. Al incorporar la variable “experiencia previa con fraude”, se observó que la desconfianza es mayor entre quienes fueron víctimas, lo cual indica que la percepción de eficacia está más influida por el historial de exposición al riesgo que por el conocimiento técnico, y que esta puede verse limitada si el sistema no es percibido como confiable.

Otro descubrimiento relevante es que quienes afirmaron haber recibido alertas automáticas de seguridad adoptaron respuestas más preventivas, denunciaron o eliminaron los mensajes sospechosos con mayor frecuencia que quienes no recibieron advertencias. Estos últimos, ignoraron los mensajes o no detectaron el contenido riesgoso. Este patrón de comportamiento

evidencia que la interacción previa con sistemas de seguridad influye positivamente en la respuesta frente a amenazas, lo que lo demuestra en parte su eficiencia.

También en relación con las alertas de seguridad se evidenció que es necesario considerar las particularidades de cada red social ya que el hábito que los usuarios desarrollando en cada una de ellas también influye en la percepción y respuesta frente a los mecanismos de seguridad.

También se identificó que la edad influye en las reacciones frente a intentos de fraude: los adultos entre 35 y 44 años adoptan mayoritariamente respuestas preventivas, como eliminar o reportar los mensajes sospechosos, mientras que los jóvenes y los adultos mayores tienden a incurrir en conductas más riesgosas o pasivas, como abrir los mensajes por curiosidad o no advertir su carácter engañoso. Este hallazgo refuerza lo señalado en el Estado del Arte, donde se destaca que muchos ataques de ingeniería social comienzan con un error humano, y pone de manifiesto la importancia de considerar el perfil etario en el diseño de estrategias de ciberseguridad adaptadas a distintos segmentos de usuarios.

A partir de estos resultados, y en conjunto con los antecedentes teóricos presentados en el Capítulo 2, el estudio ha cumplido con los objetivos propuestos. Se observó que los algoritmos de aprendizaje profundo presentan ventajas frente a técnicas tradicionales en términos de adaptabilidad, precisión y automatización. En particular, las RNN mostraron gran capacidad de detección en tiempo real y bajo nivel de falsos positivos, las CNN resultan especialmente útiles en la detección visual y los autoencoders alcanzan un rendimiento elevado siempre que sus parámetros sean correctamente ajustados. Sin embargo, su implementación depende de factores técnicos exigentes, como la calidad de los datos y el ajuste de hiperparámetros.

Además, la revisión teórica permitió identificar y clasificar distintos tipos de ataques de ingeniería social en plataformas digitales, detectándose que la manipulación emocional constituye un factor común en la mayoría de ellos. Finalmente, al comparar estos enfoques con las técnicas tradicionales de seguridad, se constató que los métodos basados en aprendizaje profundo ofrecen ventajas concretas en términos de precisión, adaptabilidad y capacidad de respuesta automatizada.

En síntesis, los hallazgos permiten afirmar que la hipótesis se valida parcialmente. Las técnicas de aprendizaje profundo contribuyen a mejorar la detección de comportamientos anómalos vinculados a ataques de ingeniería social, especialmente cuando su aplicación se acompaña de mecanismos de respuesta automatizada y adecuada interacción con el usuario. Sin

embargo, la confianza en estos sistemas no es homogénea, incluso entre perfiles técnicamente calificados, y está fuertemente condicionada por experiencias previas y variables personales.

Esto revela que la eficacia real de estas tecnologías no depende solo de su desempeño técnico, sino también de factores humanos, como la percepción, la confianza y la experiencia del usuario. Validar completamente la hipótesis requeriría un alineamiento entre eficacia técnica, aceptación generalizada y uso consciente, lo cual no se observó de manera uniforme en todos los grupos analizados.

# Capítulo 8: Líneas futuras de investigación

En función de los resultados obtenidos, surgen diversas líneas que podrían ser abordadas en investigaciones futuras con el fin de profundizar en la comprensión de los factores que condicionan la eficacia de los sistemas de aprendizaje profundo en la detección y prevención de ataques de ingeniería social en redes sociales.

Una primera línea de análisis apunta al comportamiento de los adultos de entre 45 y 55 años, quienes, según los resultados, son el grupo etario que con mayor frecuencia abre mensajes sospechosos por curiosidad. Esta conducta representa un riesgo elevado frente a intentos de ingeniería social. En este sentido, una futura línea de investigación podría centrarse en el desarrollo de modelos de aprendizaje profundo que incorporen variables sociodemográficas, como la edad, para personalizar la detección de comportamientos anómalos y adaptar la respuesta automatizada al perfil de cada usuario.

Otra línea para explorar es el nivel de comprensión sobre el funcionamiento de las medidas automatizadas. La actitud general de los encuestados frente a estas herramientas fue moderadamente favorable. Este hallazgo sugiere la necesidad qué factores influyen en su aceptación por parte del usuario final. Como resultado de ese análisis se podrían mejorar los algoritmos de detección.

Además, los resultados mostraron que los usuarios que no recibieron alertas automáticas de seguridad tienden a responder con menor nivel de prevención ante contenidos sospechosos, ya sea ignorándolos o sin lograr detectarlos. Investigar a este grupo específico permitiría identificar cuáles son los puntos de mejora que tiene hoy las alertas que debilitan la capacidad de respuesta del usuario y de los métodos para detectar comportamientos sospechosos.

## Referencias


Alzas Hernández, J. (2023) *Estudio de fraudes basados en la técnica de Ingeniería Social.* [Tesis de Maestría, Universidad Oberta de Catalunya]. http://hdl.handle.net/10609/148147

Amazon Web Services (s.f-a). *What is deep-learning?* [¿Qué es el Deep learning*?] ¿Por qué es importante el aprendizaje profundo?* Amazon Web Services. Recuperado 26 de septiembre de 2024 de https://aws.amazon.com/what-is/deep-learning/

Amazon Web Services (s.f-b). *What is deep-learning?* [¿Qué es el Deep learning?] ¿Cómo funciona el aprendizaje profundo? Amazon Web Servicies. Recuperado 26 de septiembre de 2024 de https://aws.amazon.com/what-is/deep-learning/

Ayerbe, A. (2020). *La ciberseguridad y su relación con la inteligencia artificial.* Real Instituto Elcano. https://www.realinstitutoelcano.org/analisis/la-ciberseguridad-y-su-relacion-con-la-inteligencia-artificial/

Ballestero, F. (2020). La ciberseguridad en tiempos difíciles: ¿Nos ocupamos de ella o nos preocupamos por ella? *Boletín Económico De ICE*, (3122), 39-48. https://doi.org/10.32796/bice.2020.3122.6993

Banco Santander. (s.f). *¿Qué es el spear phishing y cómo funciona?* Banco Santander. Recuperado el 12 de octubre de 2024 de https://www.bancosantander.es/glosario/spear-phishing

Barcenas Medina, A. L., & Alcaraz Cancio, A. J. (2025). Identificación de Amenazas Zero-Day en Entornos Empresariales con Modelos de Aprendizaje Profundo. *RICT Revista De Investigación Científica, Tecnológica E Innovación*, *3*(5), 46–54. https://doi.org/10.5281/zenodo.15149234

Benavidez Astudillo, E., Fuertes Díaz, W., Sanchez Gordon, S. (2020). Un experimento para crear conciencia en las personas acerca de los ataques de Ingeniería Social. *Revista Ciencia UNEMI.* Volumen 13(32),27-40. https://ojs.unemi.edu.ec/index.php/cienciaunemi/article/view/1028/1003

Berenguer Serrato, D. (2018). *Estudio de metodologías de Ingeniería Social*. [Tesis de Maestría, Universidad Oberta de Catalunya]. http://hdl.handle.net/10609/81273

Bodnar, D. (29 de octubre de 2020). *Ingeniería social y cómo protegerse*. Avast Academy. https://www.avast.com/es-es/c-social-engineering

Borbón Sanabria, J.S (2012). Redes sociales, entre la ingeniería social y los riesgos de la privacidad. *Revista Seguridad y Cultura de prevención de TI.* Volumen 12. https://revista.seguridad.unam.mx/numeros/numero-12

Bustio Martínez, L., Herrera Sements, V., Alvarez Carmona, M.A, González Ordiano J.A. (2024) La inteligencia artificial en la ciberseguridad. *ReinvenTec, Revista de Ciencia y Tecnología del ITTLA*. Volumen 2(2024), 2-48. https://reinventec.tlalnepantla.tecnm.mx/index.php/revista/issue/view/3

Burdova, C. (19 de mayo de 2023). *Pretexting: definición, ejemplos y cómo prevenirlo.* AVG Signal. https://www.avg.com/es/signal/what-is-pretexting

Crabtree, M. (11 de septiembre de 2024). A*prendizaje Profundo (AD) vs Aprendizaje Automático (AM): Guía comparativa.* Datacamp. https://www.datacamp.com/es/tutorial/machine-deep-learning

Coronel Ayala, F. M. y López Sevilla, G. M. (2023). Mapeo del panorama actual de la ciberseguridad en la era moderna digital. *RECIMUNDO*, Volumen 7(2), 441-452. https://www.researchgate.net/publication/373758501_Mapeo_del_panorama_actual_de_la_ciberseguridad_en_la_era_moderna_digital

DataTab. (s.f). *Prueba Chi-cuadrado*. Recuperado el 28 de mayo de 2025 de https://numiqo.com/tutorial/chi-square-test

Domínguez Arteaga, R., A y Vera Vázquez, R. (2022). Análisis espacial del ciberfraude al comercio electrónico: consideraciones en agenda política Tamaulipeca. *PODIUM*, 41, 21–40. https://dialnet.unirioja.es/servlet/articulo?codigo=8621006

Faster Capital (s.f). *Aprendizaje Semisupervisado*. Faster Capital. Recuperado el 26 de octubre de 2024 de https://fastercapital.com/es/contenido/Reconocimiento-de-imagenes-por-ordenador--como-reconocer-y-clasificar-imagenes-con-ordenadores.html#aprendizaje-supervisado--no-supervisado-y-semisupervisado.html

Federal Bureau of Investigation(FBI). (2023). *Internet Crime Report*. [Reporte de crimen en Internet]. https://www.ic3.gov/AnnualReport/Reports/2023_IC3Report.pdf

Gaset, A. (2022). *Redes Neuronales recurrentes aplicadas a series de tiempo: Predicción de pasajeros de rutas regionales en Argentina.* [Tesis doctoral, Universidad de Torcuato Di Tella]. https://repositorio.utdt.edu/handle/20.500.13098/11575

González, M. P., & Quevedo-Sacoto, A. S. (2025). Tendencias actuales en ataques de Ingeniería social. Revisión de literatura. *MQRInvestigar*, *9*(1). https://doi.org/10.56048/MQR20225.9.1.2025.e203

González Muñiz, A. (2023). *Aplicación de técnicas de aprendizaje profundo (deep learning) al análisis y mejora de la eficiencia en sistemas de ingeniería*. [Tesis doctoral, Universidad de Oviedo]. https://hdl.handle.net/10651/70370

González López, F., Mora, A., Magán-Carrión, R.(21-23 de junio 2023*). Influencia de la selección de hiper-parámetros en el rendimiento de autoencoders para la detección de ataques en red.* Conference: Jornadas Nacionales de Investigación en Ciberseguridad. https://www.researchgate.net/publication/371001009_Influencia_de_la_seleccion_de_hiper-

Guatelli, R., S. (2023). Detección de la Enfermedad de Parkinson basada en Máquinas de Aprendizaje Extremo. [Tesis de Maestría, Universidad Nacional de La Matanza] https://repositoriocyt.unlam.edu.ar/bitstream/123456789/1588/1/MI%20Guatelli.pdf

Hernández Sampieri, R.., Collado, C. F., & Lucio, P. B. (2014). *Metodología de la investigación*. Mc Graw Hill Education.

IBM (17 de junio de 2024). *What is deep learning?* [¿Qué es Deep learning?]. https://www.ibm.com/topics/deep-learning

IBM (19 de diciembre de 2023). *Anomaly detection in machine learning: Finding outliers for optimization of business functions*. [Detección de anomalías en machine learning: Búsqueda de valores atípicos para la optimización de las funciones empresariales]. https://www.ibm.com/think/topics/machine-learning-for-anomaly-detection

Jarrin Rodríguez, D. A. (2019). *Estudio comparativo entre modelos de aprendizaje profundo, desarrollados a partir de redes neuronales recurrentes y redes neuronales convolucionales, para la detección de intrusos* [Tesis de maestría, Universidad Internacional SEK Ecuador]. https://repositorio.uisek.edu.ec/bitstream/123456789/3532/1/Tesis_DAVID_JARRIN_VERSION_FINAL.pdf

Kaspersky. (s.f) *What is Social Engineering*? [¿Qué es la ingeniería social?] Recuperado el 07 de octubre de 2024 de https://www.kaspersky.com/resource-center/definitions/what-is-social-engineering

Kozar, E. (19 de septiembre de 2023). *Efectos de la inteligencia artificial en la ciberseguridad.* https://www.ventasdeseguridad.com/mas-a-fondo/otros-enfoques/23619-efectos-de-la-inteligencia-artificial-en-la-ciberseguridad.html

Lentini, N. (30 de junio de 2024). Ingeniería Social: la evolución del "Baiting". *Ámbito.* https://www.ambito.com/opiniones/ingenieria-social-la-evolucion-del-baiting-n6022928

Lubeck, L. (07 de enero de 2021). *En 2020 se duplicaron las detecciones de ataques de ingeniería social*. https://www.welivesecurity.com/la-es/2021/01/07/2020-duplico-detecciones-ataques-ingenieria-social/

Math Works (s.f). *Introducción a Deep Learning.* Recuperado el 15 de octubre de 2024 de https://la.mathworks.com/discovery/deep-learning.html

McCarthy, J. (2007). What is artificial intelligence? [Qué es la inteligencia artificial?]. Standford University. http://jmc.stanford.edu/articles/whatisai.html

Mckinsey & Company. (30 de abril de 2024). *What is machine learning*? [¿Qué es el aprendizaje automático?]. Mckinsey. https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-machine-learning

Microsoft (s.f). *What is AI for cybersecurity?* [¿Qué es la IA para la ciberseguridad?] Recuperado el 22 de septiembre de 2024 de https://www.microsoft.com/en-us/security/business/security-101/what-is-ai-for-cybersecurity

Moreno Barea, F.J. (2023). *Desarrollo de técnicas de aumento de datos para la aplicación de aprendizaje profundo en problemas de bioinformática* [Tesis Doctoral, Universidad de Málaga] https://riuma.uma.es/xmlui/handle/10630/26074

Navarro, S. (10 de abril de 2024). 4 tipos de aprendizaje automático. *Keepcoding.* https://keepcoding.io/blog/tipos-de-aprendizaje-automatico/

Ponce Gallegos, J., Quezada Aguilera, F., Silva Sprock, A., Martínez Flor, E., Casali, A., Scheihing, E., Túpac Valdivia, Y., Torres Soto, M.,D., Ornelas Zapata,F., Hernández, J., Zavala, C., Vakhnia, N.y Pedreño O.(2014). *Inteligencia Artificial.* LATIn. http://hdl.handle.net/2133/17686

Proof Point. (s.f) *¿Qué es la ingeniería social*? Proof Point. Recuperado el 29 de septiembre de 2024 de https://www.proofpoint.com/es/threat-reference/social-engineering

Quirumbay Yagual, D., Castillo Yagual, C., & Coronel Suárez, I. (2022). Una revisión del Aprendizaje profundo aplicado a la ciberseguridad. *Revista Científica Y Tecnológica UPSE*, *9*(1), 57-65. https://doi.org/10.26423/rctu.v9i1.671

Rathod, T., Jadav N.K, Tanwar, S., Alabdulatif, A., Garg, D., Singn, A. (2025). A comprehensive survey on social engineering attacks, countermeasures, case study, and research challenges.[ Un estudio exhaustivo sobre ataques de ingeniería social, contramedidas, estudios de casos y desafíos de investigación.]. Revistar Procesamiento y gestión de la información. Volumen 62 (1). https://doi.org/10.1016/j.ipm.2024.103928

Rea Guamán, A., M (2020). *Madurez en la identificación y evaluación de riesgos en ciberseguridad* [Tesis de Doctorado, Universidad Politécnica de Madrid]. https://doi.org/10.20868/UPM.thesis.65871

Rouhiainen, L., (2018). Inteligencia artificial: 101 cosas que debes saber hoy sobre nuestro futuro. Aliente editorial. https://proassetspdlcom.cdnstatics2.com/usuaris/libros_contenido/arxius/40/39307_Inteligencia_artificial.pdf

Rudra, A. (22 de febrero de 2023) ¿Qué es un ataque con cebo y cómo prevenirlo? *Power D Mark.* https://powerdmarc.com/es/what-is-a-baiting-attack/

Ruiz Martínez, W. (2021). Análisis de técnicas de Machine Learning aplicadas a la ciberseguridad informática para mejorar la detección de intrusiones y comportamientos anómalos en la Web. *Hashtag Revista especializada en Ingeniería*, (17), 44-60. https://doi.org/10.52143/2346139X.829

Salman Nadeem, C. (22 de abril de 2022). Ingeniería Social: ataques de Quid pro Quo. *Mailfence*. https://blog.mailfence.com/es/ataques-de-quid-pro-quo/

Sancho-López, C., Cuenca-Tapia, J., y Ortega-Castro, J. (2023). Ingeniería social y sus consecuencias en la población académica tecnológica de la provincia de El Oro. *MQRInvestigar, Revista Multidisciplinaria de investigación científica*, 7(4), 610-633. https://doi.org/10.56048/MQR20225.7.4.2023.610-633

Sancho Hirare, C. (2017). Ciberseguridad. Presentación del dossier. *URVIO, Revista Latinoamericana De Estudios De Seguridad*, (20), 8–15. https://doi.org/10.17141/urvio.20.2017.2859

Sarenet (23 de enero de 2024). La revolución del Deep Learning en la Ciberseguridad: Un avance imprescindible. *Sarenet*. https://blog.sarenet.es/la-revolucion-del-deep-learning-en-la-ciberseguridad-un-avance-imprescindible/

Secretaría de Transformación Digital, Universidad Nacional de Cuyo. (12 de enero de 2024). *El papel crucial de la Inteligencia Artificial en la defensa cibernética.* https://www.uncuyo.edu.ar/transformaciondigital/el-papel-crucial-de-la-inteligencia-artificial-en-la-defensa-cibernetica

Smith, C., Mcguire, B., Huang, T. Yang, G. (2006). *The History of Artificial Intelligence.* [La Historia de la Inteligencia Artificial]. Universidad de Washington. https://es.scribd.com/document/826322295/1-the-History-of-Artificial-Intelligence-Author-Chris-Smith-Brian-McGuire-Ting-Huang

Trevino, A. (2 de junio de 2023). *¿Qué es un ataque de pretexto?* Keeper. https://www.keepersecurity.com/blog/es/2023/06/02/what-is-a-pretexting-attack/

Unidad Fiscal Especializada en Ciberdelincuencia (2023). *Informe de gestión de la Unidad Fiscal Especializada en Ciberdelincuencia 2022-2023*. https://www.fiscales.gob.ar/wp-content/uploads/2023/12/UFECI_informe-de-Gestion_23_15-12.pdf

Wazid, M., Kumar, A., Chamola,V., Park,Y. (2022). Uniting cyber security and machine learning: Advantages, challenges and future Research. ICT Express. Volume 8, Issue 3. https://doi.org/10.1016/j.icte.2022.04.007

# Anexo

En el presente anexo se incluye el cuestionario distribuido a los participantes de la investigación, orientado a relevar su percepción sobre el uso de inteligencia artificial en la prevención de ataques de ingeniería social en redes sociales.

## Anexo 1: Encuesta distribuida

Descripción: Esta encuesta forma parte de un trabajo final de carrera de la Licenciatura en Sistemas. Su objetivo es conocer percepciones y experiencias sobre el uso de inteligencia artificial para prevenir ataques de ingeniería social en redes sociales. Las respuestas son anónimas y se utilizarán exclusivamente con fines académicos.

Formulario:

Sección A – Preguntas generales

1. ¿Cuál es su edad?

a) 18 a 24 años

b) 25 a 34 años

c) 35 a 44 años

d) 45 a 55 años

e) Más de 55 años

2. Máximo nivel educativo alcanzado:

a) Secundario completo

b) Terciario

c) Universitario

d) Posgrado

3. ¿Trabaja o estudia en alguna de las siguientes áreas? (marque todas las opciones que correspondan)

a) Tecnología / Informática

b) Ciberseguridad

c) Desarrollo de software

d) Ciencia de datos

e) Otra / Ninguna

Sección B – Experiencia con ataques de ingeniería social

4. ¿Alguna vez fue víctima de un intento de engaño o fraude por redes sociales?

a) Sí

b) No

c) No sé

5. En caso de haber respondido “Sí” en la anterior, ¿Recuerda en qué red social ocurrió el último intento de engaño o fraude?

6. ¿Qué medida automática de seguridad conoce o recuerda haber visto en una red social? (Ej.: verificación en dos pasos, bloqueo de cuenta, alerta de acceso, etc.)

7. ¿Con qué frecuencia nota mensajes sospechosos en sus redes sociales?

a) Nunca

b) Rara vez

c) A veces

d) Frecuentemente

e) Muy frecuentemente

8. ¿Qué suele hacer ante mensajes sospechosos en redes sociales?

a) Los ignoro

b) Los elimino

c) Los denuncio o reporto

d) Los abro por curiosidad

e) No los noto

Sección C – Percepción de medidas automatizadas de seguridad

9. ¿Alguna vez recibió una alerta automática de seguridad en una red social (por ejemplo, acceso desde ubicación no habitual)?

a) Sí

b) No

c) No sé

10. ¿Considera útiles las medidas automatizadas que brindan las redes sociales ante comportamientos sospechosos?

a) Totalmente en desacuerdo

b) En desacuerdo

c) Ni de acuerdo ni en desacuerdo

d) De acuerdo

e) Totalmente de acuerdo

11. ¿Cree que las plataformas han mejorado su capacidad para detectar fraudes o comportamientos inusuales?

a) Totalmente en desacuerdo

b) En desacuerdo

c) Ni de acuerdo ni en desacuerdo

d) De acuerdo

e) Totalmente de acuerdo

Sección D – Inteligencia artificial en seguridad

12. ¿Está familiarizado con el uso de inteligencia artificial en ciberseguridad?

a) Sí, mucho

b) Algo

c) Poco

d) Nada

13. ¿Qué tan de acuerdo está con la siguiente afirmación?

"La inteligencia artificial puede ayudar a detectar comportamientos extraños en redes sociales y prevenir ataques."

a) Totalmente en desacuerdo

b) En desacuerdo

c) Ni de acuerdo ni en desacuerdo

d) De acuerdo

e) Totalmente de acuerdo

14. ¿Confiaría en un sistema automatizado basado en inteligencia artificial para proteger sus redes sociales?

a) Sí

b) No

15. ¿Posee conocimientos de los siguientes conceptos? (marque los que conoce o sabe explicar )

a) Redes neuronales

b) Aprendizaje profundo

c) Detección de anomalías

d) Autocodificadores

e) Ninguno

16. ¿Qué tan de acuerdo está con la siguiente afirmación? "Los sistemas que detectan comportamientos anómalos pueden prevenir ataques de ingeniería social."

a) Totalmente en desacuerdo

b) En desacuerdo

c) Ni de acuerdo ni en desacuerdo

d) De acuerdo

e) Totalmente de acuerdo

17. ¿Considera que la inteligencia artificial es más efectiva que métodos tradicionales (como contraseñas o antivirus) para prevenir engaños digitales?

a) Sí

b) No