



Fraudes Bancarios en tiempos de Inteligencia Artificial

Profesora de Trabajo Final: Marcela Samela

Alumno: Gastón Nicolás Echaury

Trabajo Final de Carrera presentado para obtener el título de
Licenciatura en Gestión de Tecnología informática

(julio, 2026)

Resumen

El presente trabajo de investigación aborda la problemática emergente en el sector financiero que deriva de la integración de la inteligencia artificial (IA), analizando su doble rol como herramienta defensiva y también como herramienta potencial para cometer fraudes sofisticados. La investigación se centró en la transformación de las metodologías delictivas, examinando la evolución desde técnicas físicas y analógicas hacia la ejecución de fraudes automatizados e hiperrealistas. Metodológicamente, se desarrolló un estudio de enfoque mixto que integra un análisis documental exhaustivo sobre las vulnerabilidades de los modelos de aprendizaje automático y la generación de contenido sintético, complementando con un relevamiento de campo compuesto por una encuesta a usuarios de servicios financieros y una entrevista a una experta en prevención de fraudes. El análisis de los datos recolectados permitió dimensionar el impacto de la ingeniería social potenciada por IA en el contexto argentino, evidenciando una asimetría crítica entre la percepción de seguridad de los clientes y su exposición real a nuevas amenazas como los deepfakes y la clonación de voz. Finalmente, el estudio logró caracterizar el fenómeno de la “industrialización del fraude”, identificando las limitaciones de los sistemas de autenticación biométrica actuales y estableciendo lineamientos estratégicos para la actualización de los protocolos de ciberdefensas en los bancos tradicionales.

Palabras Clave: banco, ciberseguridad, fraude bancario, identidad, inteligencia artificial, suplantación

Abstract

This research addresses the emerging problems in the financial sector stemming from the integration of artificial intelligence (AI), analyzing its dual role as both a defensive tool and a potential tool for carrying out sophisticated frauds. The research focused on the transformation of criminal methodologies, examining the evolution from physical and analog techniques to the execution of automated and hyperrealistic frauds. Methodologically, a mixed-methods approach was developed, integrating a comprehensive documentary analysis of the vulnerabilities of machine learning models and the generation of synthetic content, complemented by field research consisting of a survey of financial services users and an interview with a fraud prevention expert. The analysis of the collected data allowed for quantifying the impact of AI-enhanced social engineering in the Argentine context, revealing a critical asymmetry between customers' perception of security and their current exposure to new threats such as deepfakes and voice cloning. Finally, the study succeeded in characterizing the phenomenon of the “industrialization of fraud”, identifying the limitations of current biometric authentication systems and establishing strategic guidelines for updating cyber defense protocols in traditional banks.

Key words: artificial intelligence, bank, banking fraud, cybersecurity, identity, impersonation

Dedicatoria

Para comenzar, quiero agradecer a todas esas personas que me acompañaron durante todo este proceso de crecimiento tanto personal como profesional.

A mi papá, mi mamá y mis hermanas, gracias por ser el pilar principal en cada etapa de este camino. Por su paciencia, su ejemplo, y por enseñarme que la constancia y el compromiso siempre tienen su recompensa. Cada paso o decisión que tomé durante la carrera estuvo acompañado de su apoyo incondicional.

A mi novia, gracias por estar presente incluso en los momentos en los que no era fácil acompañarme. Por tus palabras de aliento en todo momento, por creer en mí aun cuando yo dudaba, y por darme siempre un motivo para seguir dando un poco más de mí. Este logro también es tuyo.

A mis amigos, los que siempre están, gracias por su apoyo constante, las charlas sinceras, y el aguante en los días largos. Ustedes hicieron que este recorrido tenga también momentos inolvidables.

A los profesores y profesoras, les agradezco su orientación constante y su disponibilidad para acompañarme en este desafío.

Y, por último, agradezco también a mis compañeros de trabajo, el equipo de Seguridad en Proyectos, por su apoyo constante, por compartir conocimientos y experiencias. Su colaboración y buena predisposición fueron clave para llevar a cabo esta etapa.

Reconocimientos

Quiero reconocer especialmente a la Universidad Abierta Interamericana, por brindarme una formación académica sólida, actualizada y comprometida con el desarrollo profesional. Cada materia, cada docente, y cada instancia de aprendizaje fueron fundamentales en la construcción de este trabajo final.

Índice General

Capítulo 1 - Introducción	12
1.1 Identificación del Problema	12
1.2 Planteamiento del Tema de Investigación	12
1.3 Pregunta de Investigación	13
1.4 Hipótesis.....	13
1.5 Objetivo general.....	13
1.6 Objetivos Particulares	13
1.7 Justificación del Tema	14
1.8 Metodología de Investigación.....	14
Capítulo 2 – Estado del Arte	16
2.1 Introducción al Estado del Arte.....	16
2.2 La Inteligencia Artificial como Herramienta Defensiva	16
2.3 Fraudes Bancarios en Argentina	19
2.4 IA Defensiva a Nivel Internacional	21
2.5 Investigaciones sobre Técnicas Ofensivas con IA	22
2.5.1 Vulnerabilidades mediante Adversarial Machine Learning	22
2.5.2 Ingeniería Social Potenciada por IA.....	23
2.5.3 Suplantación de Identidad con Deepfakes.....	23
2.6 Aportes de la Investigación	24
Capítulo 3 – Marco teórico.....	26
3.1 Introducción al Marco Teórico.....	26
3.2 Origen y métodos tradicionales de Fraude Bancario	26
3.2.1 Fraude en la Era Pre-Digital	26
3.2.2 Fraude Documental	27

3.2.3	<i>Fenómeno del Skimming</i>	27
3.3	Digitalización y el auge de los fraudes en línea	29
3.3.1	<i>Del Fraude Físico al Digital</i>	29
3.3.2	<i>El Auge del Phishing</i>	29
3.3.3	<i>Las Herramientas de Malware</i>	30
3.4	La Ingeniería Social y su Potenciación mediante IA	30
3.4.1	<i>Phishing como Métodos Clave</i>	30
3.4.2	<i>El impacto en Argentina</i>	31
3.5	La Automatización del Fraude mediante IA	32
3.5.1	<i>La Inteligencia Artificial</i>	32
3.5.2	<i>Introducción al Machine Learning</i>	33
3.5.3	<i>Redes Neuronales y Deep Learning</i>	34
3.6	Respuestas Bancarias y Adaptación de la Ciberseguridad	35
3.6.1	<i>La obligación de Seguridad del BCRA</i>	35
3.6.2	<i>La Primera Línea de Defensa</i>	35
3.6.3	<i>Adopción de IA para la Defensa</i>	38
3.7	Marco legal y Regulatorio del Fraude Bancario en Argentina	38
3.7.1	<i>Ley de Delitos Informáticos y el Fraude Informático</i>	38
3.7.2	<i>Normativa del BCRA</i>	39
3.8	Conclusiones del Marco Teórico	40
Capítulo 4 – Desarrollo Técnico		41
4.1	Introducción al Desarrollo Técnico	41
4.2	Ataques mediante Adversarial Machine Learning	41
4.2.1	<i>La Lógica del Modelo de IA</i>	41
4.2.2	<i>Ataques de Evasión</i>	42

4.2.3	<i>Ataques de Envenenamiento</i>	43
4.3	Ingeniería Social Automatizada	45
4.3.1	<i>El Factor Humano como Vulnerabilidad Central</i>	45
4.3.2	<i>Anatomía de un Ataque de Ingeniería Social</i>	45
4.3.3	<i>La Potenciación de los Ataques mediante IA</i>	46
4.4	Suplantación de Identidad Biométrica	47
4.4.1	<i>Videos Hiperrealistas o Deepfakes</i>	47
4.4.2	<i>Redes Generativas Antagónicas (GANs)</i>	47
4.4.3	<i>Aplicación Ofensiva en la Verificación Biométrica (KYC)</i>	48
4.5	Conclusión de Desarrollo Técnico	49
Capítulo 5 – Análisis de los Resultados		50
5.1	Introducción	50
5.2	Resultados de la Encuesta (Análisis Cuantitativo)	50
5.2.1	<i>Perfil de la Muestra</i>	51
5.2.2	<i>Percepción sobre el Impacto de la Inteligencia Artificial</i>	51
5.2.3	<i>El Factor Humano</i>	52
5.2.4	<i>Nivel de Confianza vs Nivel de Vulnerabilidad</i>	54
5.3	Resultados de la Entrevista (Análisis Cualitativo)	55
Capítulo 6 - Conclusiones y Sugerencias		57
Capítulo 7 – Líneas de Investigación Futura		59
Acrónimos		60
Referencias		61
Anexos		66
Anexo 1. Resultados de la Encuesta (Análisis Cuantitativo)		66
Anexo 2. Resultados de la Entrevista (Análisis Cualitativo)		70

Anexo 3. Ejemplos Adicionales de Phishing Sofisticado	74
Anexo 4. Ejemplos de Campañas de Concientización Bancaria.....	76

Índice de Figuras

Figura 1	Métodos más Utilizados para Hacer Skimming.....	28
Figura 2	Ejemplo de Phishing Sofisticado (Caso Real).....	31
Figura 3	Incidentes Informáticos Registrados en el 2022.....	32
Figura 4	Aprendizaje Automático.....	33
Figura 5	Ejemplo1 de Campaña de Concientización Bancaria.....	36
Figura 6	Ejemplo 2 de Campaña de Concientización Bancaria.....	37
Figura 7	Ejemplo 3 de Campaña de Concientización Bancaria.....	37
Figura 8	Ataque de Evasión.....	42
Figura 9	Ataque de Envenenamiento.....	44
Figura 10	Página Falsa Net Cash – Banca Empresa BBVA 1.....	74
Figura 11	Página Falsa Net Cash – Banca Empresa BBVA 2.....	75
Figura 12	Página Falsa – Centro de Atención VISA que afecta a Clientes BBVA.....	75
Figura 13	Ejemplo 4 de Campaña de Concientización Bancaria.....	76
Figura 14	Ejemplo 5 de Campaña de Concientización Bancaria.....	76
Figura 15	Ejemplo 5 de Campaña de Concientización Bancaria.....	77
Figura 16	Ejemplo 6 de Campaña de Concientización Bancaria.....	77
Figura 17	Ejemplo 7 de Campaña de Concientización Bancaria.....	78
Figura 18	Ejemplo 8 de Campaña de Concientización Bancaria.....	78

Índice de Tablas

Tabla 1	Frecuencia de Uso de la Banca Online	51
Tabla 2	Nivel de Dificultad de Detección de los Fraudes con IA.....	52
Tabla 3	Respuesta ante Ataque de Phishing	52
Tabla 4	Respuesta ante Ataque de Vishing.....	53
Tabla 5	Uso de la Misma Contraseña	53
Tabla 6	Aceptación de Permisos en Aplicaciones Móviles	53
Tabla 7	Nivel de Confianza en Home Banking	54
Tabla 8	Victima de Fraude Bancario con Pérdida Económica	55

Capítulo 1 - Introducción

1.1 Identificación del Problema

La evolución de la inteligencia artificial (IA) transformó fuertemente al sector bancario en todo el mundo. Si bien en sus inicios la IA fue adoptada como una herramienta para optimizar procesos y fortalecer las defensas contra amenazas cibernéticas, en la actualidad también ha sido incorporada por actores maliciosos como medio para desarrollar nuevas formas de fraude.

Este sector, históricamente vulnerable a ataques cibernéticos, enfrenta hoy un escenario más complejo, el uso de IA ha potenciado la capacidad de los atacantes para generar fraudes más efectivos, automatizados, y difíciles de detectar. Entre estas nuevas modalidades se encuentran el uso de deepfakes para suplantación de identidad, el phishing personalizado mediante algoritmos de aprendizaje automático, y ataques de manipulación de modelos como el model poisoning, que comprometen la integridad de los sistemas basados en IA.

La dificultad creciente para detectar y prevenir estos fraudes pone en evidencia una transformación en el paradigma de la ciberseguridad bancaria. Este trabajo de investigación planteó la necesidad de comprender, desde una perspectiva técnica y contextual, cómo la IA ha contribuido a modificar las formas en que se cometen fraudes bancarios. La identificación y análisis de estas prácticas resulta clave para dimensionar el alcance del problema y su evolución en el tiempo.

1.2 Planteamiento del Tema de Investigación

Este trabajo investigó el uso de la IA para cometer fraudes en entidades bancarias. Aunque la IA fue originalmente concebida para fortalecer la ciberseguridad, se constató que actualmente también es aprovechada por delincuentes para ejecutar ataques más rápidos, sofisticados y difíciles de detectar.

Este problema representa un desafío significativo para el sector bancario, que maneja información y fondos críticos de millones de personas. La sofisticación de estos fraudes conlleva consecuencias importantes, incluyendo pérdidas económicas, la pérdida de la confianza del cliente y el daño reputacional para las instituciones. Además, se observó que este tipo de ataques no se limita a una sola región, sino que es un fenómeno global que afecta a países con sistemas bancarios digitalizados.

La investigación tuvo como objetivo comprender en profundidad estas nuevas técnicas de fraude. Se espera que los hallazgos de este análisis sirvan para que las entidades bancarias y los profesionales de la ciberseguridad puedan desarrollar estrategias de protección más efectivas.

Aunque existen numerosas investigaciones sobre el uso de la IA en la defensa, este estudio se centró en su aplicación ofensiva, un área que ha recibido menos atención hasta el momento. Se examinaron las medidas que distintos países, tanto de Latinoamérica como de todo el mundo, ya han implementado para combatir este problema, como los sistemas automatizados de detección de fraudes. Sin embargo, se concluyó que aún existen brechas importantes en la seguridad, especialmente ante ataques avanzados que manipulan directamente los modelos de IA o utilizan deepfakes para suplantar identidades.

1.3 Pregunta de Investigación

¿Cómo está cambiando el fraude bancario a partir del uso ofensivo de inteligencia artificial, y qué desafíos plantea para los sistemas actuales de seguridad?

1.4 Hipótesis

El uso ofensivo de inteligencia artificial está cambiando las formas en que se cometen fraudes bancarios, superando las capacidades de los sistemas actuales de seguridad.

1.5 Objetivo general

Analizar cómo el uso ofensivo de inteligencia artificial está transformando el fraude bancario y qué desafíos plantea para los sistemas de seguridad actuales.

1.6 Objetivos Particulares

Identificar las principales técnicas de inteligencia artificial que se utilizan hoy en día para cometer fraudes bancarios.

Describir casos concretos en los que la IA fue utilizada para suplantar identidades o evadir los controles de seguridad.

Analizar cómo estas técnicas afectan la eficacia de los sistemas de seguridad actuales.

Evaluar los principales desafíos que enfrentan los bancos ante este tipo de fraudes.

Proponer recomendaciones para mejorar la seguridad frente a estas nuevas amenazas.

1.7 Justificación del Tema

Este trabajo de investigación adquiere relevancia ante el incremento de los fraudes cibernéticos en el sector bancario, fenómeno que se ha intensificado en paralelo con los procesos de digitalización del sistema financiero y la incorporación de tecnologías avanzadas, como la inteligencia artificial. El enfoque adoptado proporciona una perspectiva centrada en el impacto de la IA en la seguridad bancaria, permitiendo analizar los desafíos y riesgos que estas tecnologías representan en los entornos financieros actuales.

Asimismo, la elección del tema se vincula con la experiencia profesional desarrollada en el área de Seguridad en Proyectos dentro de una entidad bancaria. Esta inserción en el sector permitió identificar problemáticas concretas relacionadas con la gestión de riesgos tecnológicos y la prevención de fraudes en contextos altamente digitalizados. Dicha perspectiva contribuye a orientar el análisis hacia situaciones y desafíos vigentes en la práctica profesional del área de ciberseguridad.

1.8 Metodología de Investigación

Este trabajo se basó en una metodología de tipo mixta, combinando enfoques exploratorios, cualitativos y cuantitativos.

Según los autores Hernández Sampieri et al. (2014), el valor que le da este enfoque a la investigación realizada es la complementariedad de sus componentes para obtener una comprensión integral. El análisis cuantitativo ofrece datos objetivos sobre variables medibles y, por otro lado, el análisis cualitativo aporta los detalles necesarios para interpretar esos datos. **(p. 569)**

Desde el enfoque cualitativo, se realizó un análisis documental de estudios técnicos, informes especializados, artículos académicos y casos reales de fraude bancario impulsado por IA. Se buscó comprender las estrategias ofensivas utilizadas, su evolución y el contexto en el que se desarrollan.

Desde el enfoque cuantitativo, se incorporaron datos estadísticos reales disponibles sobre incidentes de fraude, niveles de impacto económico, frecuencia de ataques y evolución de técnicas

ofensivas. Estos datos permitieron dimensionar el problema y contrastar la información obtenida en el análisis cualitativo.

La investigación fue de carácter exploratorio, ya que se trató un tema nuevo y poco estudiado en la región. El objetivo fue generar conocimiento que contribuya a futuras investigaciones y al desarrollo de estrategias de ciberseguridad más efectivas en el sector bancario.

Complementando el análisis documental, la investigación incorporó un estudio de campo para obtener datos primarios. Se distribuyó una encuesta para cuantificar la percepción del riesgo y el nivel de preparación ante estas nuevas amenazas. Asimismo, se realizó una entrevista en profundidad a una experta en fraudes bancarios para obtener una perspectiva cualitativa y práctica sobre los desafíos analizados.

Capítulo 2 – Estado del Arte

2.1 Introducción al Estado del Arte

Para el desarrollo de esta investigación, se consideró importante una comprensión profunda del contexto actual en torno al uso de la IA en el sector bancario, tanto desde una perspectiva defensiva como ofensiva. Por eso, en este capítulo se presenta un recorrido por los principales aportes académicos, técnicos y profesionales que han abordado la relación entre IA y fraudes bancarios.

Se revisaron estudios que explican cómo la IA ha sido implementada para detectar posibles ataques cibernéticos, así como trabajos más recientes que advierten sobre el uso malicioso de estas mismas tecnologías por parte de actores malintencionados. También se analizaron investigaciones internacionales que identifican nuevas amenazas, patrones de ataque y respuestas estratégicas adoptadas por entidades bancarias en distintos contextos.

El propósito de esta revisión fue detectar avances, vacíos y oportunidades que sirvan de base para dar contexto a la presente investigación. Esto permitió no solo fundamentar el enfoque elegido, sino también destacar el aporte diferencial que ofrece este trabajo.

2.2 La Inteligencia Artificial como Herramienta Defensiva

La detección de fraudes involucra a muchas industrias, pero sobre todo a los bancos. Una herramienta defensiva ante fraudes se encarga de identificar y bloquear toda actividad que se considere sospechosa, con el fin de evitar poner en riesgo los sistemas y el negocio. Antes de la llegada de la IA a los sistemas bancarios, el método más común que se aplicaba para la detección de posibles fraudes era el cruce de grandes cantidades de datos contra reglas preestablecidas capaces de detectar estos posibles ataques. Este método tradicional demandaba mucho tiempo de procesamiento y trabajo manual por parte de los analistas (uFlow, 2024).

La continua sofisticación y complejidad de los fraudes bancarios han sido los principales motivos para la rápida evolución de la IA en los últimos años. Tecnologías como el Machine Learning (ML), big data y la detección predictiva de anomalías se han vuelto herramientas clave en este escenario. Aunque el uso de IA no es para nada simple, los beneficios que aporta en la detección de riesgos son considerablemente mayores. De cara al futuro, la función de la IA en la

estrategia contra el fraude es cada vez más fundamental, equipando a las entidades financieras con los medios necesarios para asegurar sus recursos y la confianza de sus clientes (Monroy Gomez et al., 2024).

A continuación, se mencionan en orden cronológico cómo fueron evolucionando los sistemas de detección de fraudes bancarios más relevantes:

Al comienzo, durante la década del 90, se utilizaban sistemas de detección tradicionales. Los cuales se basaban en el uso de reglas predefinidas según experiencias pasadas. Este sistema era muy útil pero solo para identificar patrones que ya eran conocidos por los analistas, al no retroalimentarse no era posible detectar nuevos patrones de fraude (FasterCapital, 2024).

La necesidad de manejar un alto volumen de información en la década del 2000 impulsó a las entidades bancarias a adoptar sistemas estadísticos tradicionales. Herramientas como el análisis de correlación permitían identificar anomalías al descubrir relaciones complejas entre distintas variables. A pesar de esta ventaja, al igual que los sistemas de detección tradicional, estos sistemas presentaban una dependencia significativa del factor humano. Al ser incapaces de aprender de forma autónoma, requerían que los operadores los actualizaran manualmente con nuevos patrones de fraude para mantener su efectividad (Grapheverywhere, 2024).

Para la década del 2010, se comenzaron a implementar sistemas de aprendizaje automático, estos sistemas se destacan por integrar algoritmos de ML, los cuales aprenden a partir de datos históricos. Al entrenar el modelo con casos pasados de fraude y no fraude, se logra minimizar los falsos positivos y optimizar la exactitud de las reglas de riesgo, haciendo la detección más eficiente. Tal y como indica Tanant (2024) “puedes implementar estas reglas para bloquear o permitir ciertas acciones de usuario, como accesos sospechosos, robos de identidad o transacciones fraudulentas” (p.1).

El aprendizaje profundo o deep learning, implementado a través de redes neuronales profundas, es una rama de la inteligencia artificial que enseña a las computadoras a procesar datos de una manera que simula el funcionamiento neuronal del cerebro humano. Esta capacidad les permite identificar patrones complejos en diversas fuentes de datos, como imágenes, texto y sonido, para finalmente producir información valiosa y predicciones certeras (Amazon Web Services, Inc, 2023).

Los algoritmos de aprendizaje automático son muy importantes para analizar grandes volúmenes de datos en el sector bancario con el objetivo de identificar patrones de fraude. Para

minimizar el riesgo, los bancos implementan modelos robustos como redes neuronales, árboles de decisión y máquinas de soporte vectorial, los cuales son capaces de distinguir comportamientos anómalos de manera eficiente (Borrero-Tigrreros y Bedoya-Leiva, 2020).

A continuación, se mencionan algunos usos de la inteligencia artificial en el sector bancario:

La automatización de procesos, impulsada por la IA, es una estrategia clave para mejorar la eficiencia operativa en las organizaciones. Su implementación consiste en utilizar sistemas capaces de recopilar, procesar y analizar datos de forma continua y autónoma. Esto no solo facilita la toma de decisiones, sino que también elimina la necesidad de ejecutar manualmente tareas repetitivas, teniendo como resultado directo una minimización significativa del riesgo de errores humanos (McClintock, 2023). Como menciona McClintock (2023) “Al integrar la IA en la automatización, los sistemas pueden aprender y tomar decisiones basadas en datos, contexto y patrones” (p. 1).

El análisis predictivo se utiliza principalmente para identificar y mitigar riesgos. De acuerdo con SafetyCulture (2024), este enfoque combina datos actuales e históricos mediante el uso de aprendizaje automático, estadística y modelado de datos. Al procesar toda la información a su disposición, los modelos predictivos pueden anticipar posibles amenazas y ayudar a encontrar las soluciones más efectivas para bloquearlas.

El uso de algoritmos del ML es fundamental para el análisis de grandes volúmenes de información, ya que permite identificar patrones y anomalías asociados a actividades fraudulentas. Para los bancos, la mitigación del riesgo es prioridad, por lo que emplean técnicas avanzadas como redes neuronales, árboles de decisión y máquinas de soporte vectorial (SVM) para fortalecer sus sistemas de detección (Borrero-Tigrreros y Bedoya-Leiva, 2020). En el sector financiero, esta tecnología tiene múltiples aplicaciones clave, como, por ejemplo, evaluar la solvencia y analizar el comportamiento de pago de los clientes, también para la detección de fraude en solicitudes de crédito y en los procesos de autenticación en diversos canales. Adicionalmente, es fundamental en la apertura de productos on-line y en el análisis en tiempo real de transacciones sospechosas. La implementación de estas herramientas permite no solo proteger los activos, sino también aumentar la confianza de sus clientes (Fintech Americas, 2023).

Este detalle de la evolución tecnológica sirve como marco para entender los desafíos específicos que enfrentan los bancos en Argentina, donde la adopción de estas herramientas ha sido acelerada por un contexto social y económico particular.

2.3 Fraudes Bancarios en Argentina

El contexto argentino presenta una dinámica especial en la lucha contra el fraude bancario. Para entenderlo hay que comprender dos principales factores. Primero, la gente utiliza cada vez más las plataformas digitales de los bancos (home banking), y a ello se añade que el país tiene desafíos económicos propios que dificultan la situación. Gracias a investigaciones locales podemos entender un panorama que va desde el evento que detonó el problema hasta el por qué todavía existen frenos, tanto en la tecnología como en la gente, para usar la IA como principal mecanismo de defensa.

El punto de quiebre que evidenció la vulnerabilidad del sistema bancario argentino fue la pandemia de Covid-19. El estudio de Teixido (2021) sobre la situación en Mendoza revela que el confinamiento obligatorio forzó una migración masiva y abrupta de los usuarios hacia los canales digitales, creando un terreno fértil para el ciberdelito. Esta transición expuso especialmente a los grupos más vulnerables, como las personas mayores y jubilados, a nuevas y sofisticadas modalidades de estafa. Detalla también que los delitos de fraude, extorsión y suplantación de identidad no solo se incrementaron drásticamente, sino que evolucionaron, siendo el phishing a través de WhatsApp una de las nuevas modalidades más destacadas para vaciar cuentas bancarias. Los datos recopilados por Teixido (2021) confirman esta tendencia, mostrando que durante 2020 el phishing bancario y el fraude con tarjetas representaron el 39% y el 29% respectivamente de los delitos, siendo la transferencia de ahorros la operatoria más común.

Frente a este escenario de amenaza creciente, los bancos tradicionales privados en Argentina reconocen que la implementación de soluciones de aprendizaje automático es una oportunidad ineludible para mejorar su competitividad y eficiencia (Nocetti, 2019). La tesis de Nocetti de la Universidad de San Andrés, basada en entrevistas con directivos de bancos locales (2019), confirma que la detección de fraude es uno de los casos de uso más relevantes para la IA. Sin embargo, la misma investigación de Nocetti (2019) subraya que la adopción de estas tecnologías enfrenta barreras estructurales significativas. La principal barrera identificada es la cultura organizacional de los bancos, caracterizada por procesos burocráticos, lentitud en la toma

de decisiones y una baja tolerancia al fallo, en contraste con la agilidad de las Fintech. A esto se suman otros problemas. Por un lado, es un desafío técnico conectar la tecnología nueva con los sistemas antiguos. Y, por otro lado, faltan expertos en manejo de datos en el país. Finalmente, las leyes actuales generan dudas, especialmente porque no es fácil entender o explicar las decisiones que toma la IA.

El panorama actual refleja esta diferencia entre la necesidad de innovar y las dificultades para hacerlo. Por un lado, nativos digitales como Mercado Libre ya cuentan con equipos de más de 1.000 profesionales que utilizan IA para diversas soluciones, incluyendo la prevención de fraudes (Leguiza, 2024). Por otro lado, la percepción de los profesionales del sector contable y administrativo en Argentina es contradictoria. Una encuesta realizada por Bumeran y analizada por Leguiza (2024) revela que, si bien el 79% de los talentos argentinos cree que la IA impactará positivamente en su trabajo al incorporar nuevos conocimientos y facilitar tareas, un 68% admite tener miedo a ser reemplazado por esta tecnología. Esta preocupación es especialmente marcada en el sector de administración y finanzas, señalado por un 29% de los encuestados como uno de los más susceptibles a ser reemplazado (Leguiza, 2024). Esta dualidad de optimismo y temor define el estado actual de la transición hacia la IA en el ámbito financiero argentino, un proceso que avanza de manera desigual y que plantea tanto oportunidades de eficiencia como profundos interrogantes sobre el futuro del trabajo profesional (Leguiza, 2024).

En el contexto local, existe una coincidencia clara entre Teixido (2021) y Nocetti (2019), el principal problema en Argentina no es solo técnico, sino que está en el factor humano y en la cultura lenta de los bancos tradicionales. Sin embargo, surge una diferencia interesante al ver los datos de Leguiza (2024), donde se muestra que los empleados bancarios tienen sentimientos encontrados, ven a la IA como una ayuda para ser más eficientes, pero al mismo tiempo tienen miedo de ser reemplazados. Esta discusión muestra que, aunque la tecnología está disponible, su implementación en el país enfrenta barreras culturales y de capacitación que todavía no han sido resueltas.

Habiendo mencionado la problemática y las barreras estructurales en el contexto nacional, resulta importante analizar cómo otras organizaciones a nivel internacional están abordando desafíos similares, para así identificar tendencias y lecciones aprendidas.

2.4 IA Defensiva a Nivel Internacional

La evolución tecnológica ha transformado profundamente los mecanismos de protección en el sector financiero a nivel mundial. Como señalan Monroy Gomez et al. (2024), la IA, en particular, ha sido clave para detectar patrones irregulares, automatizar procesos de análisis y anticipar amenazas cibernéticas en tiempo real. Esta sección presenta tres investigaciones internacionales que permiten contextualizar y analizar cómo otras organizaciones han implementado la IA en la prevención de fraudes, sus beneficios y desafíos.

En un estudio sobre el Banco Colpatria de Colombia, Monroy Gomez et al. (2024), evidencian cómo el uso de la IA ha fortalecido los controles internos en la detección de fraudes. Los autores identifican que las tecnologías de IA más utilizadas son el reconocimiento facial y de voz, junto con el aprendizaje automático. Su investigación también señala que los principales beneficios percibidos por los expertos son un aumento en la eficiencia operativa y en la seguridad, lo que se tradujo en la detección de más de 10,000 intentos de fraude. Sin embargo, el estudio concluye que existen limitaciones importantes, como los elevados costos de implementación y una etapa inicial de madurez tecnológica dentro de la entidad (Monroy Gomez et al., 2024).

Por otro lado, una investigación de González Martínez et al. (2024) centrada en la experiencia del consumidor, aborda cómo la IA responde a las ineficiencias de los procesos bancarios tradicionales, como los largos tiempos de espera y el exceso de trámites. El estudio revela que la automatización mediante IA permite agilizar los análisis de crédito y personalizar la oferta de productos, mejorando la percepción de seguridad y la confianza del usuario. Este impacto positivo se refleja, según los autores (2024), en el crecimiento exponencial de los canales digitales, que para el año 2021 ya representaban el 75% de las operaciones bancarias en Colombia.

Finalmente, un análisis más estructural del Banco de España, realizado por Alonso-Robisco y Carbó (2022), aborda el impacto de la IA en los mercados financieros globales, explicando el contexto como una alianza estratégica definida por el debate entre la búsqueda de una mayor precisión predictiva y el desafío de la interpretabilidad de los modelos. Si bien su estudio confirma que la prevención del fraude es una de las áreas donde las ventajas del ML son más evidentes, también advierte sobre los riesgos asociados a la opacidad algorítmica o el problema de la caja negra, que puede ser un obstáculo en actividades reguladas. Por ello, concluyen que el uso de estas tecnologías exige mayores controles y nuevos marcos regulatorios (Alonso-Robisco y Carbó, 2022).

Al comparar lo que plantean Borrero-Tigeros y Bedoya-Leiva (2020) con el estudio de Alonso-Robisco y Carbó (2022), se nota un punto de discusión importante. Por un lado, los primeros destacan la gran precisión de los modelos de IA para detectar fraudes, y, por otro lado, el Banco de España advierte sobre el problema de la caja negra, donde no siempre es fácil explicar cómo toma sus decisiones la IA. Esto permite entender que el desafío actual de los bancos no es solo tener una tecnología que detecte el fraude, sino que también cumpla con las leyes que exigen transparencia en las operaciones financieras.

A nivel mundial, la IA es una herramienta clave. Sin embargo, en Argentina esta tendencia se enfrenta a desafíos culturales y a un menor nivel de desarrollo, lo que justifica la importancia de este trabajo.

2.5 Investigaciones sobre Técnicas Ofensivas con IA

Mientras que la mayoría de los estudios académicos se enfocan en el uso defensivo de la IA, se ha identificado un potencial crecimiento de investigaciones que analizan su contraparte, el uso ofensivo. Un estudio realizado por Martín Martín de la Universidad Abierta de Cataluña (2024) resalta particularmente este doble papel que cumple la IA, reconociéndola como una herramienta que también puede ser utilizada por actores maliciosos. Los estudios de esta sección demuestran que la IA ofensiva no se limita a un único tipo de ataque, sino que abarca distintas técnicas que van desde la manipulación de los propios modelos de aprendizaje automático hasta la automatización a gran escala de fraudes basados en la ingeniería social (Martín Martín, 2024).

2.5.1 Vulnerabilidades mediante Adversarial Machine Learning

El fundamento teórico de los ataques más avanzados contra sistemas que emplean IA reside en el concepto de Adversarial Machine Learning. Según el estudio realizado por García Peña et al. (2018), este campo de estudio parte de la premisa de que los métodos basados en ML pueden ser vulnerados. En lugar de atacar la infraestructura de un sistema, estas técnicas se enfocan en explotar las vulnerabilidades inherentes a los modelos predictivos para engañarlos. El objetivo del atacante es manipular el proceso de toma de decisiones del modelo para que cometa errores que beneficien sus fines (García Peña et al., 2018).

Dentro de este campo, los estudios analizados distinguen principalmente dos tipos de amenazas:

Ataques de Evasión: En esta modalidad, el ataque no busca romper el sistema de seguridad, sino engañarlo. El atacante toma una entrada inesperable y maliciosa, como por ejemplo una transacción fraudulenta, y le aplica una serie de modificaciones mínimas e imperceptibles para el humano. El objetivo final es que el sistema catalogue la entrada como si fuese una operación benigna y permitirle de esta manera evadir todos los controles de seguridad (Fernandez et al., 2023).

Ataques de Envenenamiento: Esta técnica busca corromper un modelo de IA desde su fase de entrenamiento. Al introducir datos maliciosos en el conjunto de entrenamiento, el atacante puede degradar el rendimiento general del modelo o crear "puertas traseras" que puede explotar en el futuro (Fernandez et al., 2023).

2.5.2 Ingeniería Social Potenciada por IA

Una de las aplicaciones ofensivas de la IA más extendidas es su uso para potenciar ataques de ingeniería social. Respecto a esta técnica, se ha observado una tendencia creciente en los correos electrónicos maliciosos generados por IA, un fenómeno relacionado directamente con el lanzamiento de herramientas de IA generativa (Al Lawati, 2025).

Informes técnicos del sector de la ciberseguridad han cuantificado este impacto. Según el estudio realizado por Maithan Al Lawati (2025), se ha determinado que un mensaje de phishing convincente puede ser generado por una IA en tan solo cinco minutos.

2.5.3 Suplantación de Identidad con Deepfakes

Como último punto de esta sección, la IA ofensiva ha revolucionado las técnicas de suplantación de identidad mediante la generación de contenido multimedia sintético. Investigaciones sobre nuevas modalidades de estafas demuestran cómo los atacantes utilizan la IA para crear perfiles falsos y testimonios fraudulentos muy convincentes (Crypto.com, 2024).

Una de las herramientas más notables en este ámbito son los videos y audios falsos (*deepfakes*), utilizados específicamente para engañar a las víctimas en diversos esquemas de fraude. Esta tecnología representa una amenaza directa para los procesos de verificación de

identidad que están siendo adoptados por el sector bancario, como los sistemas de Conozca a su Cliente (KYC) basados en video o la autenticación biométrica por voz.

En conclusión, el análisis de esta literatura evidencia que el uso ofensivo de la IA no es una amenaza teórica, sino una realidad en constante evolución. La sofisticación, escalabilidad y accesibilidad de estas herramientas representan un desafío directo y fundamental para los modelos de seguridad tradicionales (Martín Martín, 2024).

Esto justifica la necesidad de investigaciones que, como la presente, se enfoquen en comprender en profundidad estas nuevas modalidades de fraude para poder desarrollar estrategias de ciberdefensa que estén a la altura del nuevo paradigma.

2.6 Aportes de la Investigación

A pesar del creciente interés por el uso de la IA en ciberseguridad, la mayoría de los estudios actuales se enfocan principalmente en sus aplicaciones defensivas dentro del sistema financiero. Existen numerosos trabajos que analizan cómo la IA puede ser utilizada para detectar fraudes, identificar patrones inusuales o reforzar medidas de autenticación. Sin embargo, son aún limitadas las investigaciones que examinan el uso ofensivo de estas tecnologías por parte de actores maliciosos.

Asimismo, gran parte de las publicaciones se centran en contextos internacionales, con especial foco en países de altos ingresos y marcos regulatorios avanzados. Esto genera una falta de evidencia empírica y análisis situados sobre cómo estas amenazas impactan a bancos tradicionales en regiones como América Latina, donde los niveles de digitalización, inversión en ciberdefensa y madurez regulatoria presentan características distintas.

Este trabajo busca aportar a ese vacío, proponiendo una mirada desde adentro del sector bancario, combinando experiencia profesional directa con una revisión académica. Se espera que este enfoque permita comprender las técnicas utilizadas para cometer fraudes con IA.

Al centrarse en el uso ofensivo de la IA desde una perspectiva regional, esta investigación no solo busca mapear las nuevas amenazas, sino también contextualizar el riesgo para los bancos tradicionales. De esta manera, se pretende generar un aporte práctico y localizado que sirva como punto de partida para que profesionales, desarrolladores y reguladores en Argentina puedan anticiparse a la próxima generación de fraudes.

Como se ha visto en este capítulo, el Estado del Arte identifica un creciente interés en estas técnicas ofensivas, pero se enfoca en sus implicaciones generales. Los siguientes capítulos tomarán estos conceptos, los definirán formalmente y desglosarán su funcionamiento técnico detallado para llenar el hueco analítico detectado.

Capítulo 3 – Marco teórico

3.1 Introducción al Marco Teórico

El objetivo principal de este capítulo es establecer los conceptos principales sobre los cuales se realizó el análisis de la investigación. Para tener un entendimiento completo y acertado sobre el impacto del uso ofensivo de la IA con el fin de cometer fraudes bancarios, es necesario primero definir y dar contexto a elementos claves que han configurado la situación actual del delito bancario.

El desarrollo de este marco teórico sigue un orden lógico, tanto en orden cronológico como tecnológico. El primer apartado se enfoca en describir cómo eran los métodos de fraudes en su era tradicional y física, para luego dar lugar al primer gran cambio que fue la digitalización y el auge de los fraudes en línea.

Luego se profundiza sobre el elemento central de la mayoría de los ataques, el factor humano, explicando los fundamentos de la ingeniería social.

Continúa con las definiciones tecnológicas que representan el objetivo de este trabajo de investigación, la IA, el Machine Learning y el Deep Learning. Con las definiciones antes mencionadas, se puede comprender cómo los bancos han tenido que adaptar sus estrategias de ciberseguridad en respuesta a la constante evolución de las amenazas.

Y como último punto se analizó el contexto local a través del marco legal y regulatorio argentino, definiendo las normativas vigentes que regulan esta materia.

Al finaliza este capítulo, se espera contar con las herramientas necesarias para comprender el análisis técnico de los ataques y las conclusiones que se presentarán en las secciones posteriores de este trabajo.

3.2 Origen y métodos tradicionales de Fraude Bancario

3.2.1 *Fraude en la Era Pre-Digital*

Los comienzos del fraude bancario fueron empleando métodos fundamentalmente físicos y tangibles. Previo a la era digital los métodos delictivos requerían de la manipulación y sustracción de elementos físicos. En algunos casos hasta era necesaria la presencia del atacante en una locación cercana (Moreno, 1966). Este tipo de delitos es a lo que Moreno (1966), en su tesis doctoral definió como “delito de astucia y engaño” (p. 6), donde el éxito del fraude no dependía

de la tecnología empleada por el atacante, sino de su habilidad para falsificar la realidad a través de medios materiales (Moreno, 1966).

Dado que los métodos de fraudes eran físicos imponía límites naturales a su alcance y velocidad. Cada ataque requería un esfuerzo y planificación muy dedicada y minuciosa, obligando en algunos casos al atacante a estar en un lugar físico específico, cerca de la víctima. Como consecuencia de esto, siempre quedaban rastros de evidencia, como documentos alterados o dispositivos manipulados, que luego permitirían una investigación por parte de los peritos forenses (Moreno, 1966).

3.2.2 Fraude Documental

Durante el siglo XX, el factor más importante de los fraudes bancarios fue el cheque. La seguridad de la veracidad y autenticación de estos documentos descansaba en la confianza y capacidad de detección que tenía cada persona. La principal amenaza era la falsificación, que podía ser creación de un documento enteramente falso o la adulteración de un documento verdadero (Moreno, 1966).

La defensa contra estos ataques era un proceso manual y reactivo. Los empleados del banco eran la primera línea de defensa, encargados de revisar visualmente cada firma. Esta total dependencia en el ojo humano era la mayor debilidad del sistema. La propia dinámica del trabajo bancario hacía casi imposible un análisis riguroso, una realidad que Moreno (1966) describió en detalle.

Esta dependencia del ojo humano y del análisis manual representa el principal problema que la IA defensiva vino a solucionar décadas más tarde, aunque, irónicamente, la IA ofensiva hoy busca explotar sesgos similares, pero a nivel algorítmico.

3.2.3 Fenómeno del Skimming

El auge de las tarjetas de crédito y débito provocó una gran evolución de los fraudes bancarios. La técnica más significativa de esta etapa fue el skimming, una modalidad delictiva que se consolidó como una de las principales amenazas para los usuarios en sus transacciones físicas (Banco Santander, 2023). Este método constituía el principal riesgo para los clientes bancarios

antes de que la pandemia acelerara de forma abrupta la migración hacia los canales digitales (Teixido, 2021), punto de inflexión que analizaremos en la próxima sección.

La forma de operar de skimming dependía enteramente de la manipulación del hardware. La técnica consistía en el robo de los datos de la tarjeta en el momento en que un cliente realizaba una transacción. Como podemos ver en la Figura 1, para lograrlo, los atacantes instalaban dispositivos electrónicos ocultos en los cajeros automáticos. Estos aparatos, conocidos como skimmers, tenían como único objetivo copiar la información contenida en la banda magnética de la tarjeta cuando el usuario la introducía en la ranura del cajero (Banco Santander, 2023).

Para que el fraude sea completo, el atacante necesitaba el Número de Identificación Personal (PIN) de la víctima. Por eso el modo de operar del atacante se complementaba con la instalación de microcámaras, estratégicamente instaladas para grabar el momento en el que el usuario digitaba su clave, o también colocando una falsa placa sobre el teclado original que registraba las pulsaciones (Banco Santander, 2023).

Figura 1

Métodos más Utilizados para Hacer Skimming



Nota. Tomado de "Skimming": ¿qué es y cómo protegerte contra este fraude?, por Banco Santander, 2023. (<https://www.santander.com/es/stories/skimming>)

Todo el proceso delictivo, desde la planificación y la instalación de los dispositivos hasta su posterior recuperación para descargar la información robada, estaba estrictamente ligado a una locación física. Esto demuestra que la presencia del criminal en el lugar del hecho era condición para poder cometer este tipo de fraudes (Teixido, 2021).

En este sentido, el skimming puede ser entendido como el método principal de fraude físico, donde la principal vulnerabilidad no residía en el software del banco ni en el comportamiento en línea del usuario, sino en la seguridad del hardware de acceso público.

3.3 Digitalización y el auge de los fraudes en línea

3.3.1 Del Fraude Físico al Digital

A principios del siglo XX, la transición a lo digital representó una de las transformaciones más significativas en la historia del sector financiero. La introducción de las aplicaciones web y móviles, expandió exponencialmente el campo de ataque para los actores maliciosos. El fraude ya no requiere de la presencia física del atacante ni la manipulación de objetos tangibles, el campo de batalla se trasladó por completo al ciberespacio (Teixido, 2021).

Según lo mencionado por Teixido (2021), este proceso de digitalización se vio drásticamente acelerado en Argentina por la pandemia de COVI-19. Las medidas de confinamiento obligatorio forzaron una migración masiva y abrupta de los usuarios a los canales digitales, lo que creó un entorno ideal para el crecimiento del ciberdelito. Esta transición no fue gradual ni planificada para la mayoría de la población, sino una necesidad impuesta por las circunstancias. Como resultado se expuso, a un nuevo grupo de usuarios con menos conocimientos técnicos como las personas mayores, a un ecosistema de amenazas completamente nuevo y desconocido para ellos.

3.3.2 El Auge del Phishing

Con la era digital de las entidades financieras, se consolidó una modalidad de ataque que se convertiría en la más emblemática del fraude en línea, el phishing. Esta técnica es parte de la ingeniería social, que busca manipular a las personas para que realicen acciones o divulguen información confidencial (Dirección Nacional de Ciberseguridad, 2023). Según Teixido (2021),

el phishing es un tipo de engaño que utiliza un correo electrónico de apariencia legítima con la intención de que la persona revele su información personal.

La forma de operar del phishing evolucionó rápidamente. Los atacantes comenzaron a crear sitios web falsos, réplicas exactas de los portales legítimos de los bancos, con el objetivo de que la víctima ingresara voluntariamente sus credenciales de acceso (Teixido, 2021). Esta modalidad introdujo el problema de la escalabilidad, donde un único ataque de phishing podía dirigirse a miles de usuarios simultáneamente con un costo y esfuerzo mínimo.

3.3.3 *Las Herramientas de Malware*

Según la Dirección Nacional de Ciberseguridad (2023), un malware es un software malicioso que está diseñado para infiltrarse en un dispositivo sin que el usuario se dé cuenta. Este tipo de software se convirtió en una herramienta fundamental para los atacantes, permitiéndoles tomar control de los dispositivos de las víctimas para el robo de información.

Dentro del ecosistema del malware, el troyano fue particularmente relevante para el fraude bancario en esta etapa. Este es un tipo de malware que se disfraza de software legítimo, como un programa útil o un documento, para engañar al usuario y lograr que lo instale (Dirección Nacional de Ciberseguridad, 2023). Una vez dentro del sistema, puede realizar una variedad de acciones maliciosas, como registrar datos, robar información financiera o proporcionar acceso remoto al atacante (Teixido, 2021).

Como respuesta a estas nuevas amenazas, las entidades bancarias comenzaron a implementar una nueva generación de defensas digitales, como la autenticación de doble factor (2FA), sentando las bases de la carrera armamentista tecnológica que define la ciberseguridad actual.

3.4 La Ingeniería Social y su Potenciación mediante IA

3.4.1 *Phishing como Métodos Clave*

Como se mencionó en el apartado anterior, el phishing es la manifestación más común de la ingeniería social en el entorno digital. Consiste en un intento de estafa que utiliza el correo electrónico como principal vehículo de ataque (Jefatura de Gabinete de Ministros, 2021). El

método de operación se basa en la suplantación de la identidad de una organización de confianza para solicitar a la víctima que haga click en un enlace o descargue un archivo malicioso (Jefatura de Gabinete de Ministros, 2021).

Esta técnica tiene diversas variantes que adaptan el canal de comunicación. Las más relevantes en el contexto argentino son, el smishing, que utiliza mensajes de texto (SMS) en lugar de correos electrónicos para enviar el enlace fraudulento, y el vishing, que se ejecuta a través de una llamada telefónica (voice phishing), donde el atacante se hace pasar por un empleado del banco para solicitar datos sensibles (Alzas Hernandez, 2023).

La sofisticación de este tipo de campañas se evidencia en ejemplos reales detectados por la industria. Como se observa en la Figura 2, este correo de phishing fue detectado en un entorno operativo real, simulando una alerta de seguridad urgente utilizando lenguaje visual y el tono exacto de una comunicación bancaria legítima. No posee errores de redacción o de diseño, lo cual antes delataba el fraude. Esta verosimilitud extrema es precisamente lo que anula las defensas del usuario y demuestra como la IA permite escalar la personalización y la credibilidad del engaño. Para ver más ejemplos de phishing ver Anexo 3. Ejemplos Adicionales de Phishing Sofisticado.

Figura 2

Ejemplo de Phishing Sofisticado (Caso Real)



Nota. BBVA. (2025). *Material de concientización sobre ciberseguridad*. <https://www.bbva.com.ar>

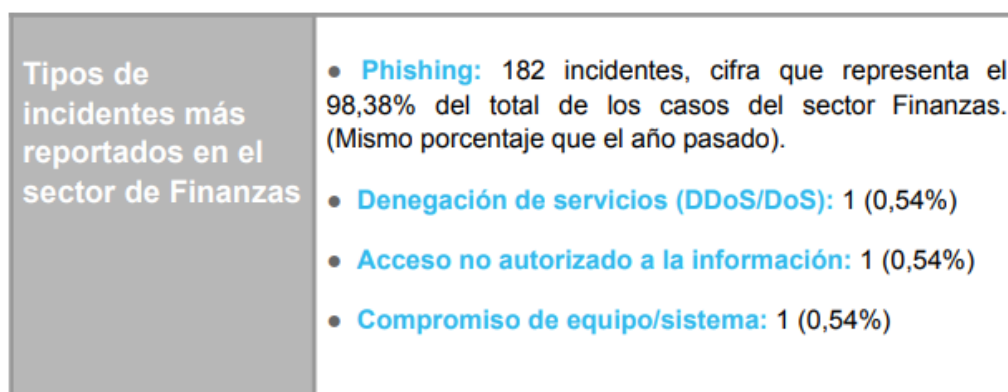
3.4.2 *El impacto en Argentina*

Los ataques basados en ingeniería social en argentina están documentados en informes oficiales. El CERT.ar (2023) en su informe anual de 2022, clasificó la publicación de información sensible, a menudo obtenida a través de phishing, como uno de los incidentes más reportados, y el phishing en sí mismo como una de las categorías principales de incidentes gestionados, tal y como se detalla

en la Figura 3. Esto demuestra que la ingeniería social no es una amenaza teórica, sino una de las causas más importantes de incidentes de seguridad reales en el país (CERT.ar, 2023).

Figura 3

Incidentes Informáticos Registrados en el 2022



Nota. Tomado de *Incidentes Informáticos, Informe anual de incidentes de seguridad informática registrados en el 2022*, por CERT.ar, 2023. (https://www.argentina.gob.ar/sites/default/files/2023/02/informe_cert_2022.docx.pdf)

El impacto de estas técnicas se vio magnificado durante la pandemia, afectando desproporcionadamente a los grupos más vulnerables. La migración digital forzada expuso a personas no familiarizadas con estas amenazas a un aluvión de estafas, principalmente a través de plataformas masivas como WhatsApp (Teixido, 2021).

3.5 La Automatización del Fraude mediante IA

3.5.1 La Inteligencia Artificial

El fraude en su constante evolución ha adoptado las herramientas tecnológicas más avanzadas de cada época. En el escenario actual la tecnología disruptiva que está redefiniendo las capacidades tanto de los atacantes como de los defensores es la IA. En su conceptualización más amplia, la IA es una rama de la informática cuyo objetivo es crear máquinas y sistemas que puedan simular la inteligencia humana para realizar tareas de forma autónoma. Su paradigma operativo consiste en procesar grandes cantidades de datos para aprender, razonar y tomar decisiones, una

capacidad que los ciberdelincuentes han comenzado a explorar para automatizar y perfeccionar sus ataques (Rouhiainen, 2018).

Para comprender su impacto en el fraude, es indispensable distinguir dos de sus subcampos más relevantes, que representan diferentes niveles de complejidad, el Machine Learning y el Deep Learning.

3.5.2 Introducción al Machine Learning

El Machine Learning (ML), o aprendizaje automático, es el subcampo más extendido de la IA. Su principal característica es que los sistemas aprenden directamente a partir de los datos, sin necesidad de que un programador escriba reglas explícitas para cada tarea. La calidad y cantidad de los datos con los que se entrena al modelo son cruciales, ya que su rendimiento y futuras predicciones dependen directamente de ellos (Rouhiainen, 2018). En función de la metodología de aprendizaje que se emplee, es posible diferenciar tres enfoques principales, como se muestra en la Figura 4.

Figura 4

Aprendizaje Automático



Nota. Tomado de *Inteligencia Artificial, 101 cosas que debes saber hoy sobre nuestro futuro*, por Lasse Rouhiainen, 2018. (https://planetadelibrosar0.cdnstatics.com/libros_contenido_extra/40/39307_Inteligencia_artificial.pdf)

En el Aprendizaje Supervisado, el algoritmo se entrena con un conjunto de datos que ha sido previamente etiquetado por humanos, donde cada ejemplo tiene una respuesta correcta conocida. El objetivo es que el modelo aprenda a predecir la etiqueta correcta para datos nuevos que no ha visto antes. Este método es muy utilizado para problemas de clasificación, como decidir si una transacción bancaria es fraudulenta o legítima basándose en ejemplos pasados (Rouhiainen, 2018).

A diferencia del anterior, el Aprendizaje No Supervisado, trabaja con un conjunto de datos sin etiquetar. El algoritmo aprende explorando los datos por su cuenta para identificar patrones, estructuras o anomalías ocultas. En el ámbito de la seguridad, se emplea para tareas de agrupación (clustering), como detectar comportamientos de gasto inusuales que no siguen un patrón conocido que podrían indicar un fraude (Rouhiainen, 2018).

Por último, el Aprendizaje por Refuerzo, se basa en un modelo de prueba y error, similar a como aprende un ser humano. El algoritmo aprende a tomar decisiones interactuando con un entorno, y recibe recompensas por las acciones correctas y penalizaciones por las incorrectas. A través de este proceso, el agente aprende cuál es la estrategia óptima para maximizar su recompensa a largo plazo (Rouhiainen, 2018).

3.5.3 *Redes Neuronales y Deep Learning*

El Deep Learning o aprendizaje profundo, constituye una rama avanzada del ML, que ha ganado gran popularidad en los últimos años. Su principal característica es que utiliza redes neuronales artificiales con múltiples capas para analizar datos de una forma muy compleja. Estas redes están inspiradas en la estructura y el funcionamiento del cerebro humano, permitiendo que el modelo aprenda patrones extremadamente abstractos y sutiles a partir de volúmenes masivos de datos (Rouhiainen, 2018).

Gracias a esta arquitectura de capas interconectadas, los modelos de Deep Learning son especialmente potentes para tareas como reconocimiento de imágenes y el procesamiento de lenguaje natural (Rouhiainen, 2018). Es precisamente esta capacidad la que los ciberdelincuentes están comenzando a utilizar para crear fraudes más sofisticados, como los deepfakes, que requieren un entendimiento profundo y generativo de patrones visuales y auditivos.

3.6 Respuestas Bancarias y Adaptación de la Ciberseguridad

3.6.1 La obligación de Seguridad del BCRA

Frente a la constante evolución de las amenazas, la estrategia de ciberdefensa de las entidades financieras en Argentina no es opcional, sino que parte de un marco regulatorio estricto impuesto por el Banco Central de la República Argentina. La principal normativa en esta materia es la Comunicación “A” 7181 (2020), que establece los requisitos mínimos de gestión, implementación y control de los riesgos tecnológicos que todos los bancos deben cumplir.

Esta regulación impone a los bancos una obligación de seguridad proactiva en la cual no solo deben reaccionar ante los incidentes, sino que están obligados a desarrollar e implementar una política de seguridad de la información robusta y documentada.

Sin embargo, este marco no es estático. En respuesta directa a la creciente sofisticación de los ataques, el BCRA ha profundizado estas exigencias a través de normativas posteriores. Se destaca la Comunicación “A” 7724 (2023), que actualiza la norma original introduciendo requisitos específicos sobre ciberresiliencia, gestión de incidentes y la obligación de reportar ataques.

Al imponer este estándar mínimo de diligencia, el BCRA crea el contexto normativo dentro del cual se inscriben las respuestas bancarias que se analizarán a continuación, como la educación al cliente y la adopción de tecnologías avanzadas como la IA para la defensa. En esencia la normativa actúa como el motor fundamental que empuja al sector hacia una mejora continua de sus capacidades de ciberdefensa. (Banco Central de la República Argentina, 2020).

Los detalles específicos de estas exigencias se abordarán en la sección 3.7.2 correspondiente al marco legal.

3.6.2 La Primera Línea de Defensa

Más allá de las defensas técnicas, una parte fundamental de la estrategia de seguridad de los bancos en Argentina se centra en la educación y la concientización del cliente, reconociéndolo como la primera línea de defensa. El propio Banco Central, a través de campañas de difusión reportadas por medios de comunicación, establece una serie de recomendaciones para que los usuarios puedan protegerse de las estafas virtuales (Diario Clarin, 2025).

Estas recomendaciones se centran en fortalecer el eslabón más débil, el factor humano. Se instruye a los clientes a nunca compartir información personal ni bancaria, como claves o códigos de verificación, ya que los bancos nunca solicitan estos datos. Asimismo, se enfatiza la importancia de no operar bajo presión, de cortar cualquier llamada sospechosa y de contactar al banco por sus canales oficiales para verificar cualquier comunicación. Medidas técnicas al alcance del usuario, como el uso de contraseñas seguras y la activación del 2FA, son también pilares de esta estrategia de prevención (Diario Clarin, 2025).

Esta estrategia de educación activa se evidencia en diversas campañas de difusión implementadas por las entidades financieras. Las Figuras 5, 6 y 7 muestran ejemplos concretos de este tipo de material, donde se observa que el mensaje central se enfoca en advertir al usuario sobre las tácticas comunes de los estafadores, como solicitar claves o generar urgencia, y reforzar la recomendación clave: nunca compartir información sensible y contactar siempre al banco por canales oficiales. Este enfoque preventivo, dirigido directamente al usuario, respalda la visión de que la concientización es un pilar fundamental. Para ver más ejemplos de capacitaciones, ver Anexo 4. Ejemplos de Campañas de Concientización Bancaria.

Figura 5

Ejemplo1 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 6

Ejemplo 2 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 7

Ejemplo 3 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

3.6.3 *Adopción de IA para la Defensa*

En paralelo a la regulación y la educación, los bancos argentinos han comenzado a adoptar tecnologías más avanzadas para hacer frente a la sofisticación de los ataques. La IA se ha posicionado como herramienta clave en esta nueva etapa de la ciberdefensa.

En la investigación académica basada en entrevistas con directivos de bancos tradicionales y privados en Argentina de Nocetti (2019), se revela que la detección de fraude es considerada uno de los casos de uso más relevantes y con mayor potencial para la aplicación de IA y ML en el sector. Las entidades reconocen que la implementación de estas soluciones es una oportunidad ineludible para mejorar no solo la seguridad, sino también su competitividad y eficiencia frente a las nuevas amenazas (Nocetti, 2019). Esta visión interna del sector confirma que la adaptación de la ciberseguridad ya no pasa solo por cumplir con la normativa, sino por una carrera tecnológica en la que la IA es la principal herramienta de defensa.

3.7 Marco legal y Regulatorio del Fraude Bancario en Argentina

3.7.1 *Ley de Delitos Informáticos y el Fraude Informático*

El marco legal que rige el fraude bancario en su faceta digital en Argentina tiene como pilar fundamental la Ley 26.388, conocida comúnmente como Ley de Delitos Informáticos, sancionada en el año 2008. La sanción de esta ley representó un punto de inflexión en la legislación del país, ya que, hasta ese momento, no existían figuras delictivas diseñadas específicamente para el entorno digital. Los fiscales y jueces debían forzar los hechos, como el robo de credenciales o las transferencias no consentidas, para que encajaran en la figura de la estafa tradicional del artículo 172 del Código Penal, lo que generaba enormes dificultades probatorias y de interpretación judicial (Ribón, 2019).

La Ley 26.388 optó por una reforma integral, modificando e incorporando diversos artículos al Código Penal de la Nación para dar respuesta a las nuevas modalidades delictivas (Ribón, 2019). Para el objeto de esta investigación, la incorporación más relevante es el artículo 173, inciso 16, que detalla de manera específica el fraude informático. “El que defraudare a otro mediante cualquier técnica de manipulación informática que altere el normal funcionamiento de un sistema informático o la transmisión de datos” (Ministerio de Justicia de la Nación, 2008, p. 1).

Esta figura penal es de suma importancia, ya que su redacción amplia permite abarcar la mayoría de las modalidades de fraude en línea que surgieron con la digitalización. El tipo penal no se centra en un método específico, sino en el resultado. Este concepto es lo suficientemente abarcador para incluir tanto los ataques de ingeniería social, como el phishing, donde se manipula al usuario para que ingrese sus datos en un sistema fraudulento, como los ataques basados en malware, que alteran directamente el funcionamiento del dispositivo de la víctima (Ribón, 2019).

3.7.2 Normativa del BCRA

Más allá del marco penal que sanciona al delincuente, la responsabilidad de las entidades financieras está rigurosamente regulada por el BCRA, tal y como se menciona en el apartado 3.6.1, la normativa central de esta materia es la Comunicación “A” 7181. Sin embargo, en respuesta directa a la evolución y sofisticación de las ciberamenazas, esta norma fue actualizada y profundizada por la Comunicación “A” 7724.

Estas comunicaciones establecen una obligación de seguridad que recae sobre las entidades financieras, un deber que va mucho más allá de una simple recomendación. Exige a los bancos desarrollar, implementar y controlar una política de seguridad de la información robusta y documentada. Las normas reconocen implícitamente la asimetría en la relación entre el banco y el cliente (Banco Central de la República Argentina, 2023). Según el BCRA (2023), el banco, como proveedor profesional del servicio, posee el conocimiento y los recursos técnicos, y, por lo tanto, tiene el deber agravado de garantizar la seguridad de las operaciones.

Entre las múltiples exigencias detalladas en su anexo, la Comunicación “A” 7181 (2020) sentó las bases obligando a las entidades financieras a gestionar los riesgos tecnológicos, identificando, evaluando y monitoreando los riesgos de manera continua, implementar controles de acceso seguro. La Comunicación “A” 7724 (2023) profundiza este mandato, introduciendo explícitamente el concepto de ciberresiliencia y ajustando los requisitos sobre los planes de respuesta a incidentes, el monitoreo y el reporte de ataques, adaptándose al nuevo paradigma de amenazas.

Esta regulación es de suma importancia porque establece un estándar mínimo de diligencia. Si un cliente sufre un fraude y se demuestra que el banco no cumplió con estas exigencias, la entidad podría ser considerada responsable por los daños sufridos.

3.8 Conclusiones del Marco Teórico

A lo largo del Marco Teórico, se establecieron las bases de conocimiento para comprender la evolución del fraude bancario. Comenzando desde la descripción del fraude en su era tradicional, un delito de naturaleza física y tangible, limitado por barreras geográficas y logísticas, cuyos métodos emblemáticos fueron la falsificación de cheques y clonación de tarjetas mediante skimming.

Luego se analizó la primer gran disrupción, la digitalización. La llegada del home banking transformó el paradigma, dando lugar al auge de los fraudes en línea, cómo el phishing y el malware, que permitieron a los delincuentes operar a una escala global y masiva. Se investigó como la ingeniería social se consolidó como el principal vehículo para estos ataques, explotando el factor humano como vulnerabilidad central del sistema.

Por último, se introdujo la tecnología que define el escenario actual, la IA. Se definieron sus subcampos, el Machine Learning y el Deep Learning, como las herramientas que han iniciado una nueva guerra tecnológica. Los bancos han comenzado a adoptar estas tecnologías para robustecer sus defensas, mientras que los atacantes las utilizan para crear fraudes más sofisticados, automatizados y difíciles de detectar.

El siguiente capítulo se enfocará en el desarrollo técnico para analizar en profundidad el modo de operar de estas nuevas amenazas impulsadas por la IA.

Capítulo 4 – Desarrollo Técnico

4.1 Introducción al Desarrollo Técnico

El presente apartado se adentrará en el desarrollo técnico de las modalidades de fraude bancario más sofisticadas que utilizan la IA como herramienta ofensiva. Si el Marco Teórico respondió a la pregunta ¿qué son estas amenazas?, este capítulo tiene como objetivo desglosar el modo de operación de los ataques, respondiendo a la pregunta fundamental de ¿cómo funcionan exactamente?

Para ello, se tomarán los conceptos ya definidos, como el Adversarial Machine Learning, la Ingeniería Social y los Deepfakes, para analizarlos en su aplicación práctica en el ecosistema bancario. Se describirán los procesos técnicos que siguen los atacantes, desde las estrategias para manipular los modelos de IA de los propios bancos hasta las tácticas de suplantación de identidad biométrica que vulneran los nuevos sistemas de verificación.

Al finalizar este capítulo, se habrá presentado una visión detallada de la mecánica interna de las amenazas que definen el nuevo paradigma de la ciberseguridad financiera, sentando las bases para el posterior análisis de los resultados y la formulación de conclusiones y recomendaciones.

4.2 Ataques mediante Adversarial Machine Learning

4.2.1 *La Lógica del Modelo de IA*

Como ya hemos visto en el Marco Teórico de este trabajo, los sistemas de detección de fraude basados en ML son una de las principales defensas de las entidades financieras. Sin embargo, la creciente sofisticación de los ciberdelincuentes ha provocado que estos mismos sistemas de defensa se conviertan en un nuevo y complejo campo de batalla (Fernandez et al., 2023). En lugar de intentar vulnerar la infraestructura de red del banco o el software tradicional, ha surgido una disciplina ofensiva conocida como Adversarial Machine Learning (AML), cuyo objetivo es atacar las vulnerabilidades intrínsecas de la propia lógica de los modelos de IA (Centro Criptológico Nacional, 2023).

El propósito de estos ataques es manipular el proceso de toma de decisiones del modelo para que cometa errores deliberados, como clasificar una transacción fraudulenta como si fuera legítima. En esencia, el atacante utiliza las propias herramientas y principios de la IA en su contra, buscando engañar al algoritmo en lugar de romper una barrera de seguridad convencional (Fernandez et al., 2023). La literatura académica y los informes técnicos utilizados distinguen dos grandes familias de ataques adversarios, clasificados según el momento en que se ejecuta el ataque: los ataques de evasión y los de envenenamiento.

4.2.2 Ataques de Evasión

Los ataques de evasión son la forma más común de ataque adversario y se producen cuando el modelo de IA ya ha sido entrenado y está en pleno funcionamiento (Centro Criptológico Nacional, 2023). El objetivo del atacante no es alterar el modelo en sí, sino crear una entrada de datos maliciosa conocida como ejemplo adversario, que ha sido sutilmente modificada para ser clasificada incorrectamente por el sistema (Fernandez et al., 2023). En el contexto bancario, esto equivaldría a tomar los datos de una transacción fraudulenta y aplicarle pequeñas modificaciones imperceptibles para que el sistema de IA del banco la catalogue como legítima y la apruebe, tal y como se muestra en la Figura 8.

Figura 8

Ataque de Evasión



Nota. Tomado de *Ataques a un Sistema de Detección de Intrusiones mediante Redes Generativas Adversarias*, por Fernandez, 2023, Trabajo Fin de Grado, Universidad Complutense de Madrid.

(<https://docta.ucm.es/rest/api/core/bitstreams/6eefcdad-6d29-4ebf-97b0-99a5c8acba3b/content>)

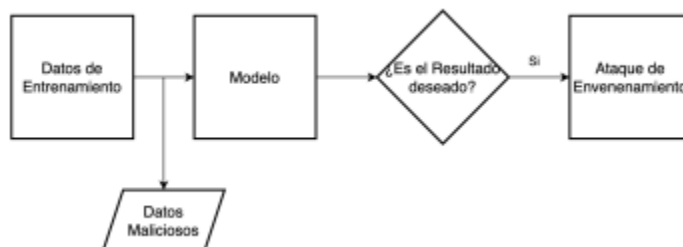
La complejidad y el método de estos ataques dependen fundamentalmente del nivel de conocimiento que el atacante tenga sobre el modelo de la víctima, lo que da a lugar a dos escenarios principales:

Ataques de Caja Blanca o White Box: En este escenario, se asume que el atacante tiene un conocimiento completo y transparente del modelo de IA de banco (Centro Criptológico Nacional, 2023). Esto incluye su arquitectura, los parámetros, los algoritmos utilizados y, en algunos casos, hasta los datos con los que fue entrenado. Con esta información privilegiada, el atacante puede calcular con precisión las modificaciones mínimas y óptimas necesarias para crear un ejemplo adversario que garantice el engaño del sistema (Fernandez et al., 2023). Aunque es el escenario menos probable para un atacante externo, es muy estudiado a nivel académico para entender las vulnerabilidades teóricas de los modelos.

Ataques de Caja Negra o Black Box: Este es el escenario más realista y peligroso para un ciberdelincuente que ataca a una entidad financiera desde el exterior (Centro Criptológico Nacional, 2023). El atacante no tiene ningún conocimiento sobre el modelo del banco, solo puede interactuar con él como un usuario más, enviando datos y observando los resultados (aprobado o rechazado). Para ejecutar el ataque en estas condiciones, el delincuente puede emplear una técnica sofisticada que consiste en entrenar un modelo local sustituto que busque imitar el comportamiento del modelo del banco. Al consultar repetidamente al sistema real con distintas entradas y observar sus respuestas, el atacante puede inferir su lógica y replicarla en su propio modelo. Una vez que tenga este modelo sustituto bajo control, lo utiliza para generar los ejemplos adversarios que también serán efectivos contra el sistema real del banco (Fernandez et al., 2023).

4.2.3 Ataques de Envenenamiento

A diferencia de los ataques de evasión, tal y como se muestra en la Figura 9, los ataques de envenenamiento de datos (data poisoning attacks) son más engañosos y estratégicos. Su objetivo no es engañar a un modelo ya entrenado, sino corromper el propio proceso de aprendizaje. Los modelos de ML de los bancos no son estáticos, a menudo se reentrenan con nuevos datos de transacciones para adaptarse a las nuevas modalidades de fraude. Es en esta fase de reentrenamiento donde el atacante busca la oportunidad para inyectar datos maliciosos en el conjunto de datos de entrenamiento (Fernandez et al., 2023).

Figura 9*Ataque de Envenenamiento*

Nota. Tomado de *Ataques a un Sistema de Detección de Intrusiones mediante Redes Generativas Adversarias*, por Fernandez, 2023, Trabajo Fin de Grado, Universidad Complutense de Madrid.

(<https://docta.ucm.es/rest/api/core/bitstreams/6eefcdad-6d29-4ebf-97b0-99a5c8acba3b/content>)

Al introducir estas muestras envenenadas, el atacante puede lograr dos objetivos principales con consecuencias muy diferentes:

Degradar el Rendimiento del Modelo: El objetivo más simple es introducir datos anómalos o mal etiquetados para reducir la precisión general del sistema de detección. Esto provoca que el modelo cometa más errores, tanto falsos positivos (alerta por transacciones legítimas) como falsos negativos (no detectar fraudes reales), volviéndolo menos confiable y en última instancia, inútil para la entidad (Fernandez et al., 2023).

Crea una Puerta Trasera (Backdoor): Este es el ataque más sofisticado y peligroso. El objetivo no es romper el modelo, sino enseñarle de manera encubierta una vulnerabilidad específica que solo el atacante conoce. Por ejemplo, el atacante podría inyectar en el set de entrenamiento una serie de transacciones fraudulentas que comparten un patrón sutil e irrelevante (como un comentario específico en la descripción o un monto decimal particular). Si el ataque tiene éxito, el modelo aprenderá a considerar que cualquier transacción futura que contenga ese patrón debe ser clasificada siempre como legítima. De esta manera, el atacante crea una puerta trasera que le permitirá realizar fraudes sin ser detectado en el futuro, simplemente añadiendo esa llave secreta a sus operaciones (Fernandez et al., 2023).

4.3 Ingeniería Social Automatizada

4.3.1 El Factor Humano como Vulnerabilidad Central

Más allá de las vulnerabilidades técnicas en el software o el hardware, una gran parte de los fraudes bancarios exitosos no se originan en la rotura de un código, sino en la explotación de la confianza humana. Este enfoque se conoce como ingeniería social, una disciplina que puede ser definida como un conjunto de técnicas utilizadas para manipular a usuarios legítimos con el fin de que realicen acciones de interés para un atacante, como divulgar información confidencial. El principio fundamental de la ingeniería social no es técnico, sino psicológico, basándose en la habilidad del atacante para generar un vínculo de confianza con la víctima y explotar sesgos cognitivos universales (Berenguer Serrato, 2018).

En su informe sobre phishing, la Jefatura de Gabinete de Ministros (2021) la describe como una técnica que busca manipular a las personas para que realicen acciones, como divulgar información personal, o para que otorguen acceso a sistemas informáticos. La premisa fundamental de la ingeniería social es que, a menudo, el ser humano es el eslabón más débil en la cadena de seguridad. Los ataques de este tipo no son efectivos por fallas de programación, sino sesgos cognitivos y principios psicológicos como el sentido de urgencia, el miedo, la curiosidad o el respeto a la autoridad (Jefatura de Gabinete de Ministros, 2021).

La efectividad de estos ataques radica en su capacidad para apelar a principios de la persuasión humana. Entre ellos se destacan la autoridad (simular ser una figura legítima, como un empleado del banco), la urgencia (crear la necesidad de una acción inmediata para evitar una consecuencia negativa), la simpatía (generar una conexión con la víctima) y la prueba social (hacer creer a la víctima que otros ya han realizado la acción solicitada) (Berenguer Serrato, 2018).

4.3.2 Anatomía de un Ataque de Ingeniería Social

Un ataque de ingeniería social no es un acto improvisado, sino un proceso metodológico que, según lo investigado, se desarrolla en fases bien definidas. La primera fase es la investigación, donde el atacante recopila la mayor cantidad de información posible sobre su objetivo, utilizando a menudo fuentes de acceso público como las redes sociales o bases de datos filtradas. El objetivo

es encontrar datos personales o de contexto que le permitan construir un engaño creíble (Alzas Hernandez, 2023).

La segunda fase consiste en desarrollar la confianza con la víctima. El atacante utiliza la información recopilada para construir un pretexto y presentarse como alguien legítimo, con el fin de disminuir las defensas naturales de la persona (Alzas Hernandez, 2023).

La última fase es la explotación, donde el atacante finalmente realiza la solicitud maliciosa, como pedir una clave, solicitar una transferencia o convencer a la víctima de que haga click en un enlace fraudulento (Alzas Hernandez, 2023).

4.3.3 La Potenciación de los Ataques mediante IA

Si bien la ingeniería social tradicional ya es efectiva, la irrupción de la IA Generativa ha actuado como el impulso que la vuelve exponencialmente más peligrosa. La IA permite a los atacantes superar las limitaciones de un estafador humano, principalmente en términos de escala, personalización y credibilidad de los engaños (Alzas Hernandez, 2023).

Una de las aplicaciones más alarmantes es la clonación de voz para ataques de vishing. Sin embargo, la IA de código abierto, como CoquiTTS y Tortoise, revelan un panorama complejo. El estudio de Alzas Hernández (2023) demuestra que la clonación de voz creíble a través de estas herramientas gratuitas presenta barreras significativas para un atacante promedio, pues requiere una gran cantidad de audios de la víctima, una calidad de audio mínima y un largo tiempo de entrenamiento.

De hecho, las pruebas realizadas por el autor (2023) con pocas muestras de audio arrojaron resultados no satisfactorios, considerados insuficientes para un ataque de ingeniería social convincente. No obstante, el mismo estudio advierte que este escenario cambia radicalmente con herramientas de pago más avanzadas, que parecen ofrecer resultados excelentes con un esfuerzo mínimo, democratizando la amenaza para quien pueda costearla (Alzas Hernandez, 2023).

Por lo tanto, si bien un atacante con recursos limitados enfrentaría dificultades técnicas, uno con acceso a servicios de IA pago podría obtener muestras de audio de videos públicos en redes sociales o llamadas grabadas y generar un mensaje de voz sintético suplantando la identidad de un familiar o de un supuesto empleado del banco, ejecutando así la estafa de una manera mucho más creíble que con su propia voz.

Este salto evolutivo representa uno de los mayores desafíos actuales, ya que difumina la línea entre una comunicación auténtica y un intento de fraude sofisticado.

4.4 Suplantación de Identidad Biométrica

4.4.1 *Videos Hiperrealistas o Deepfakes*

La evolución de la IA ofensiva ha culminado con la creación de una de sus herramientas más potentes y socialmente alarmantes, los deepfakes. Un deepfake puede ser definido como un video hiperrealista que ha sido manipulado digitalmente mediante técnicas de IA para representar a personas diciendo o haciendo cosas que en realidad nunca sucedieron. Estas representaciones sintéticas, que alcanzan un nivel de verosimilitud asombroso, plantean serias amenazas para la privacidad y han inaugurado un nuevo y complejo escenario en el que se incrementan drásticamente los riesgos derivados de las suplantaciones de identidad (García-Ull, 2020).

A diferencia de otras técnicas de fraude, la peligrosidad del deepfake radica en su capacidad para engañar no solo a la percepción humana, sino también a los sistemas de seguridad automatizados que dependen del análisis biométrico. Con la creciente sofisticación de la tecnología subyacente, resulta cada vez más complicado para un observador no entrenado o incluso para un sistema informático, detectar si un video o una declaración de una persona es real o, por lo contrario, es el resultado de una manipulación generada por IA (García-Ull, 2020).

Esto representa un desafío directo para la confianza en la evidencia digital, con implicancias profundas en la seguridad financiera.

4.4.2 *Redes Generativas Antagónicas (GANs)*

El motor tecnológico que hace posible la creación de deepfakes se conoce como Redes Generativas Antagónicas o GANs. Esta es una arquitectura de Deep Learning que, en su forma más básica, consiste en dos redes neuronales que se entrenan de forma simultánea y compiten entre sí en un juego de suma cero (Calcagni, 2020).

Las dos redes que componen una GAN son:

El Generador: Esta red neuronal tiene como objetivo aprender la distribución de los datos de un conjunto de entrenamientos para poder crear nuevos datos sintéticos a partir de ellos. En el

contexto de los deepfakes, su función es generar una imagen de un rostro falso que sea lo más realista posible (Calcagni, 2020).

El Discriminador: Esta segunda red actúa como un juez o un perito. Su función es recibir tanto datos reales (imágenes de rostros verdaderos del set de entrenamiento) como los datos falsos creados por el generador, y debe aprender a distinguir unos de otros, emitiendo una probabilidad de que la imagen sea real (Calcagni, 2020).

El proceso de entrenamiento consiste en un ciclo de retroalimentación en el que ambas redes se enfrentan y mejoran mutuamente. El discriminador se vuelve cada vez más experto en detectar los fraudes del generador, y el generador aprende de los rechazos del discriminador para producir imágenes cada vez más indistinguibles de las reales (Calcagni, 2020).

Es a través de este proceso de competencia antagónica que las GANs logran producir contenido sintético de un nivel de realismo excepcional.

La investigación en este campo ha avanzado a pasos agigantados. Arquitecturas más avanzadas, como SystleGAN, desarrollada por investigadores de NVIDIA, han perfeccionado esta técnica permitiendo no solo generar rostros sintéticos de alta calidad, sino también manipular su espacio latente para alterar rasgos específicos y expresiones faciales con un control muy preciso, un área de investigación activa incluso en el ámbito académico argentino (Aranda et al., 2022).

4.4.3 Aplicación Ofensiva en la Verificación Biométrica (KYC)

La capacidad de generar rostros y videos sintéticos realistas tiene una aplicación directa y muy peligrosa en el fraude bancario, el ataque a los sistemas de verificación de identidad biométrica.

Con el auge de la banca digital, muchas entidades financieras han implementado procesos de alta de clientes conocidos como “Conozca a su Cliente” (KYC). Estos procesos remotos se basan fundamentalmente en la tecnología de reconocimiento facial, utilizada con fines de seguridad y validación, donde se le pide al usuario que se grabe en video mostrando su rostro y su documento de identidad para confirmar quién es durante la apertura de una cuenta o para autorizar una operación de alto riesgo (García-Ull, 2020).

Un atacante podría utilizar un deepfake para vulnerar estos sistemas de seguridad. Al haber obtenido previamente imágenes o videos de la víctima (por ejemplo, de sus redes sociales), el

delincuente puede utilizar una GAN para generar un video falso donde el rostro de la víctima es animado para que parezca que está hablando o siguiendo las instrucciones del sistema de verificación (como girar la cabeza o parpadear). La sofisticación de esta tecnología representa una amenaza directa para los sistemas de reconocimiento facial, que podrían ser engañados por estas representaciones sintéticas al no poder distinguir entre una persona real frente a la cámara y una simulación generada por IA (García-Ull, 2020).

Esta amenaza no es teórica, el uso de la IA para delitos financieros ya es una realidad. García-Ull (2020) documenta como los estafadores ya han utilizado audios falsos para suplantar la identidad de un ejecutivo y solicitar una “transferencia de efectivo urgente” (p. 8).

Este tipo de ataque, conocido como ataque de representación, difumina la línea entre la presencia real de una persona y una simulación digital, representando uno de los desafíos más complejos y urgentes para el futuro de la ciberseguridad bancaria (García-Ull, 2020).

4.5 Conclusión de Desarrollo Técnico

Al finalizar este capítulo, se ha detallado la mecánica interna de las amenazas que definen el nuevo paradigma del fraude. Se ha respondido al “cómo” funcionan los ataques de AML para engañar a los propios bancos, cómo la IA potencia la ingeniería social y como los deepfakes buscan vulnerar la identidad biométrica.

Sin embargo, este análisis técnico de la amenaza quedaría incompleto sin contrastarlo con la realidad empírica.

¿Son el phishing hiperrealista, la clonación de voz y los deepfakes una preocupación real y tangible en el sector bancario argentino? ¿Cómo es la percepción del usuario final frente a estas nuevas estafas?

El siguiente capítulo se centrará en responder estas preguntas, validando la teoría técnica con los datos de campo obtenidos en la encuesta a usuarios y a la entrevista a una experta del sector de prevención de fraudes bancarios.

Capítulo 5 – Análisis de los Resultados

5.1 Introducción

Como se mencionó en el apartado de Metodología de Investigación, se realizó un estudio de campo para poder complementar y validar lo investigado. Este estudio de campo es de carácter mixto, el cual se compone de una encuesta realizada a 137 usuarios de servicios financieros y se realizó una entrevista a la Gerenta del Departamento de Fraude de un reconocido banco en Argentina.

Es de suma importancia señalar que, debido al tamaño y alcance de la muestra, el análisis realizado posee un carácter exploratorio. Si bien los hallazgos permiten identificar tendencias valiosas y contrastar la percepción real de los usuarios con el marco teórico, los resultados no pretenden ser generalizables al a totalidad del sistema bancario ni al conjunto de los usuarios. Esta delimitación metodológica sitúa al estudio como una base sólida para comprender fenómenos específicos y abrir nuevas líneas de investigación, más que para establecer conclusiones de valides universal.

El objetivo principal fue obtener datos sobre la percepción real de los usuarios bancarios frente a las nuevas modalidades de fraude. En este capítulo se presentan y analizan los resultados obtenidos relacionando la perspectiva de los usuarios con los conceptos desarrollados en los apartados anteriores. Para ver el detalle de las preguntas y respuestas obtenidas en la encuesta ver Anexo 1. Resultados de la Encuesta (Análisis Cuantitativo).

5.2 Resultados de la Encuesta (Análisis Cuantitativo)

La encuesta realizada define cuatro secciones claves:

- El perfil del usuario.
- La percepción sobre el impacto de la IA.
- El comportamiento del usuario o factor humano frente una amenaza.
- La brecha entre la confianza en los sistemas de los bancos y el nivel de vulnerabilidad.

5.2.1 Perfil de la Muestra

La mayoría de los encuestados se caracterizan por tener un nivel de formación académica y por ser usuarios activos de las aplicaciones de los bancos. El 89,8% se encuentra en un rango de edad de entre 26 y 60 años (de 26 a 40 años el 50,7% y de 41 a 60 años el 39,1%), y el 86,9% posee estudios terciarios, universitarios o posgrado. Un alto porcentaje representado por el 85,4% utiliza el home banking o banca online todos los días o varias veces por semana, lo que significa un perfil de usuario que recurrentemente realiza operaciones digitales.

Tabla 1

Frecuencia de Uso de la Banca Online

Frecuencia	%
Todos los días	48,9
Varias veces por semana	36,5
Algunas veces al mes	10,9
Casi nunca	3,6

Nota. Elaboración propia.

5.2.2 Percepción sobre el Impacto de la Inteligencia Artificial

Uno de los resultados más marcados en la encuesta es que los usuarios reconocen que la IA es una de las herramientas principales para cometer fraudes más sofisticados. El 83,2% de los encuestados respondió que la IA hace que las estafas bancarias sean, “mucho más difíciles de detectar” con un 40,1% o “un poco más difíciles de detectar” para un usuario común con un 43,1%. Este resultado valida la hipótesis central de esta investigación, ya que demuestra que la percepción del usuario se encuentra alineada con la creciente sofisticación de las amenazas.

Tabla 2*Nivel de Dificultad de Detección de los Fraudes con IA*

Nivel de Dificultad	%
Mucho más difíciles	40,1
Un poco más difíciles	43,1
No influye	4,4
No estoy seguro/a	12,4

Nota. Elaboración propia.

5.2.3 El Factor Humano

Sorprendentemente se presenta una contradicción entre el conocimiento teórico de los usuarios sobre cómo actuar para prevenir un fraude, y lo que ponen en práctica. Por un lado, existe gran conocimiento sobre cómo reaccionar dando como resultado un 95,6% de usuarios que actuarían correctamente ante un correo de phishing y un 94,9% actuaría correctamente también ante una llamada de vishing.

Tabla 3*Respuesta ante Ataque de Phishing*

Acción	%
Hacer click	1,5
Pedir más información	2,9
Ignorar el correo	56,2
Eliminar el Correo	39,4

Nota. Elaboración propia.

Tabla 4*Respuesta ante Ataque de Vishing*

Acción	%
Dar la clave	0,7
Pedir más datos	2,2
Cortar y contactar al banco	94,9
Pedir que llame luego	2,2

Nota. Elaboración propia.

Contradictoriamente, los resultados a la encuesta dieron una falla significativa en la práctica. Un 45,9% de los usuarios utiliza la contraseña de su home banking en otros servicios o sistemas, y un 20,4% (uno de cada cinco), acepta los permisos de las aplicaciones móviles sin leerlos.

Tabla 5*Uso de la Misma Contraseña*

Contraseñas	%
Si, para casi todo	13,1
Para algunas cosas	32,8
Uso contraseñas diferentes	54

Nota. Elaboración propia.

Tabla 6*Aceptación de Permisos en Aplicaciones Móviles*

Aceptación de Permisos	%
Acepto todo	20,4
Reviso los permisos	36,5
No instalo la app	43,1

Nota. Elaboración propia.

Esta diferencia entre el saber lo que tengo que hacer y lo que realmente hago confirma que el factor humano sigue siendo la vulnerabilidad más crítica para los fraudes bancarios.

5.2.4 Nivel de Confianza vs Nivel de Vulnerabilidad

El resultado más crítico que dio la encuesta es la diferencia que se presenta entre la confianza que tienen los usuarios con la seguridad de la banca online y el impacto real de los fraudes. Un 81,8% de los encuestados se siente “bastante seguro” o “muy seguro” al operar en home banking.

Tabla 7

Nivel de Confianza en Home Banking

Nivel de Confianza	%
Muy seguro/a	16,8
Bastante seguro/a	65
Poco seguro/a	14,6
Nada seguro/a	3,6

Nota. Elaboración propia.

Este resultado se contrapone fuertemente con la tasa de victimización reportada, donde un 48,2%, siendo casi la mitad de los encuestados, afirma que ellos o un familiar cercano sufrieron algún tipo de estafa bancaria con pérdida económica. Esta contraposición supone una posible subestimación del riesgo personal y una falsa sensación de seguridad del home banking que los atacantes pueden explotar.

Tabla 8*Victima de Fraude Bancario con Pérdida Económica*

Víctimas de Fraude	%
Si	48,2
No	51,8

Nota. Elaboración propia.

5.3 Resultados de la Entrevista (Análisis Cualitativo)

Con el objetivo de profundizar en los temas abordados y también contar con una visión directa de la industria se realizó una entrevista a Agustina Ana Quevedo, quien se desempeña en el puesto de Financial Crime Prevention Senior Manager en el Banco BBVA Argentina.

Según lo relatado en la entrevista por la experta, el cambio más significativo no es una sola técnica, sino que menciona la “Industrialización del fraude”, un concepto que resume la transición de ataques aislados a bandas organizadas para cometer fraudes bancarios, quienes emplean una combinación de ingeniería social, automatización e IA para gestionar el proceso completo. El resultado es un fraude más rápido, menos costoso y mucho más personalizado que antes.

Respecto al contexto argentino, la entrevistada confirmó que las amenazas impulsadas por la IA son una preocupación. Destaca que los phishing hiperrealistas son casi imposibles de distinguir de las comunicaciones legítimas, y que los correos corporativos escritos por IA ya carecen de los errores ortográficos o de formato que antes servían como señal de alarma.

Respecto al volumen de ataques, confirma que existe un aumento significativo de ataques de vishing con clonación de voz. Las voces clonadas son cada vez más creíbles. El método de engaño más común sigue siendo la suplantación del departamento de seguridad del banco, solicitando un token para cancelar una falsa transferencia.

Sobre los ataques con deepfakes para KYC, la entrevistada comentó que, si bien son pocos los casos confirmados de suplantación de identidad en la apertura de cuentas, la amenaza se considera real. La visión es contundente, dejando en claro que, en un corto plazo, no se va a poder confiar en los motores biométricos por si solos, lo que obligará a reforzar las medidas de seguridad con factores adicionales o revisiones humanas.

Continuando con las preguntas realizadas, la experta en fraude describió cómo los atacantes usan la IA para probar variaciones y ver qué combinaciones logran evadir los modelos de detección, y también cómo utilizan los bots que imitan patrones humanos para enmascarar sus ataques. En cuanto a la defensa, si bien la IA se utiliza para analizar los patrones de comportamiento y transacciones en tiempo real, la entrevistada admite una realidad preocupante donde el delincuente está mucho más avanzado en el uso de IA para delinquir que los especializados en seguridad del banco.

La visión de la entrevistada valida uno de los hallazgos clave de la encuesta, la brecha de percepción. Según comentó en la entrevista, considera que existe una brecha muy grande. El público sigue pensando que cuando algo es falso, se detecta a simple vista y eso ya no es cierto. Esta falsa confianza es la vulnerabilidad principal que explotan los atacantes hiperrealistas.

Este último punto, sustenta la importancia de la educación posicionándola como un pilar fundamental de la estrategia de prevención. Sin embargo, la entrevistada reconoce que las campañas tradicionales ya no son suficientes, y que la educación dentro de las mismas aplicaciones, con mensajes contextuales y alertas preventivas, funcionan mejor.

Como último aspecto conversado en la entrevista, de cara al futuro, la mayor amenaza anticipada son los fraudes completamente conversacionales, donde una IA mantendrá un diálogo en tiempo real por voz o video imitando a empleados o familiares. En este escenario, concluye la experta, “la línea entre lo real y lo falso se va a volver muy fina”.

Para mayor detalle de las preguntas realizadas y sus respectivas respuestas ver Anexo 2. Resultados de la Entrevista (Análisis Cualitativo).

Capítulo 6 - Conclusiones y Sugerencias

La presente investigación permitió concluir que el sector financiero atraviesa un cambio de paradigma que trasciende a solo una evolución tecnológica. A través del análisis mixto realizado, se determinó que la transición desde métodos físicos como el skimming hacia la era digital del phishing ha culminado en lo que este trabajo identifica como la "industrialización del fraude". Este fenómeno, validado por la evidencia de campo, demuestra que la IA permite a las bandas organizadas ejecutar ataques con una velocidad, economía y personalización que superan ampliamente las defensas tradicionales y la capacidad de detección del usuario promedio.

Este estudio aporta una visión situada en la realidad argentina, donde se comprendió que la efectividad del fraude no reside solo en la técnica, sino en la anulación de la percepción de riesgo del cliente. El estudio de campo reveló una contradicción crítica: el 81,8% de los usuarios se siente seguro operando online, pero el 48,2% ha sido víctima de estafas. Esta "falsa sensación de seguridad" es el hallazgo central del trabajo; se obtuvo evidencia de que el conocimiento teórico (el 94,9% sabe cómo reaccionar ante una amenaza) no se traduce en un comportamiento seguro, ya que el 45,9% aún reutiliza contraseñas en sitios críticos.

Este análisis resulta relevante para el contexto local porque demuestra que el factor humano sigue siendo la vulnerabilidad principal, pero ahora potenciada por una IA que elimina los errores ortográficos o visuales que antes servían de alerta en Argentina.

En respuesta a la pregunta de investigación planteada al principio del trabajo, se concluye que el fraude bancario está cambiando, dejando de ser genérico, convirtiéndose en hiperrealista. Los fraudes actuales utilizan un lenguaje local impecable y coherencia entre canales (mail, SMS, llamada), lo que anula las señales de alerta tradicionales. La IA no solo se usa para engañar a humanos, sino también es utilizada para atacar a los sistemas de defensa de los bancos basados en IA y para vulnerar los pilares de la identidad digital a través de deepfakes y clonación de voz. El principal desafío es que los atacantes siempre están más avanzados, tal y como lo describió la experta en fraudes en la entrevista.

Por lo tanto, se valida la hipótesis planteada. La evidencia demuestra que el uso de la IA no solo está modificando la modalidad de fraude, sino que efectivamente está superando las capacidades de los sistemas de seguridad. Esta afirmación está sustentada por un 83,2% de los

usuarios encuestados, quienes afirman que la IA ha hecho que las estafas sean más difíciles de detectar, y también por la alta tasa de victimización real del 48,2% de los encuestados.

Como primer punto se recomienda a las entidades financieras implementar un modelo de "confianza cero" para la biometría, complementando la validación facial o de voz con factores de comportamiento o revisiones humanas (backoffice). Asimismo, desarrollar sistemas resistentes a ataques de evasión y envenenamiento que protejan la integridad de los propios modelos de IA del banco.

A los usuarios de servicios financieros se recomienda adoptar una cultura de escepticismo activo, priorizando la verificación por canales oficiales y el uso de herramientas técnicas como gestores de contraseñas y doble autenticación (2FA).

También a los entes reguladores, se les recomienda la normativa vigente (Comunicación "A" 7181) para incluir la gestión de riesgos específicos asociados a la suplantación de identidad con contenido sintético y ataques dirigidos al Machine Learning.

Como reflexión final, el fraude bancario ha cambiado para siempre y estará en constante cambio. Ya no se trata de estafadores solitarios o de correos mal escritos. La investigación realizada demuestra que hemos entrado en una nueva era: la "industrialización del fraude". Gracias a la IA, bandas criminales organizadas ahora producen ataques en masa, que son más rápidos, baratos y, sobre todo, mucho más personales y creíbles.

El problema es que, mientras los ataques se volvieron hiperrealistas, la percepción del público no ha avanzado al mismo ritmo. Esta investigación encontró una contradicción peligrosa: la gente se siente segura operando por internet, pero la realidad es que existe una gran cantidad de personas que han sido víctimas de una estafa. El usuario promedio sigue buscando las viejas señales de un fraude que ya no existen, creyendo que el engaño se nota, cuando la IA lo ha vuelto prácticamente invisible.

Esta batalla, además, se libra en dos frentes. No solo se engaña al cliente, sino que se ataca directamente la tecnología de los bancos. Esta guerra tecnológica es asimétrica, y hoy la ventaja la tienen los atacantes.

La consecuencia es una pérdida de la confianza en las herramientas que creíamos seguras. La amenaza de los deepfakes y las voces clonadas es tan seria que, pronto no podremos confiar únicamente en los sistemas biométricos para verificar una identidad.

La conclusión es clara, la seguridad bancaria ya no consiste en construir muros, sino en una adaptación constante. El futuro, con la llegada de "fraudes completamente conversacionales" por voz y video, nos llevará a un escenario donde la línea entre lo real y lo falso se va a volver muy fina. En este nuevo paradigma, la única defensa real será una cultura de escepticismo activo, donde la capacidad de dudar y verificar sea más importante que cualquier contraseña.

Capítulo 7 – Líneas de Investigación Futura

A partir de los hallazgos obtenidos en este estudio, se identificaron líneas que podrían ser exploradas en investigaciones futuras.

En primer lugar, profundizar en la amenaza anticipada por la experta en fraudes, analizando la efectividad de las IA conversacionales (voz y video) para ejecutar estafas en tiempo real.

También, evaluar la eficacia de las nuevas tecnologías de detección de pruebas de vida (liveness) y contenido sintético para determinar su fiabilidad frente a los deepfakes de última generación.

Adicionalmente, un estudio futuro podría medir y comparar el impacto en el comportamiento del usuario de las campañas tradicionales versus las alertas dentro de la aplicación bancaria.

Acrónimos

#

2FA: Autenticación de Dos Factores (Two – Factor Authentication)

A

AML: Adversarial Machine Learning

B

BCRA: Banco Central de la República Argentina

D

DL: Aprendizaje Profundo (Deep Learning)

G

GANs: Redes Generativas Antagónicas

I

IA: Inteligencia Artificial

K

KYC: Conozca a su Cliente (Know Your Customer)

M

ML: Aprendizaje Automático (Machine Learning)

P

PIN: Número de Identificación Personal

S

SVM: Máquinas de Soporte Vectorial (Support Vector Machines)

U

UFECI: Unidad Fiscal Especializada en Ciberdelincuencia

Referencias

- Al Lawati, M. (2025, julio 22). *Phishing de correo electrónico y estadísticas*. PowerDMARC: <https://powerdmarc.com/es/email-phishing-dmarc-statistics/>
- Alonso-Robisco, A., & Carbó, J. M. (2022). *Inteligencia artificial y finanzas: Una alianza estratégica*. Banco de España: <https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosOcasionales/22/Fich/do2222.pdf>
- Alzas Hernandez, J. (2023). *Estudio de fraudes basados en la técnica de Ingeniería Social*. Tesis de Maestría. Universitat Oberta de Catalunya, Catalunya. Repositorio Institucional UOC: <https://openaccess.uoc.edu/items/8890ba12-0a51-473f-b1b8-616d97656bde#page=1>
- Amazon Web Services, Inc. (2023). *¿Cuál es la diferencia entre el aprendizaje profundo y las redes neuronales?* Amazon Web Services: <https://aws.amazon.com/es/compare/the-difference-between-deep-learning-and-neural-networks/>
- Aranda, D., Gorla, J., Negri, P., Sandalinas, F., & Suffern, M. (2022). *Manipulación de expresiones faciales vía espacio latente de Red Generativa Antagónica (GAN)*. Trabajo de Grado. Universidad Nacional de La Plata, Buenos Aires. Universidad Nacional de La Plata: https://sedici.unlp.edu.ar/bitstream/handle/10915/151743/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- Banco Central de la República Argentina. (2020, diciembre 17). *Comunicación "A" 7181*. <https://www.bcra.gob.ar/pdfs/comytexord/a7181.pdf>
- Banco Central de la República Argentina. (2023). *Comunicación "A" 7724*. <https://www.bcra.gob.ar/pdfs/comytexord/A7724.pdf>
- Banco Santander. (2023, enero 30). *"Skimming": ¿qué es y cómo protegerte contra este fraude?* <https://www.santander.com/es/stories/skimming>
- Berenguer Serrato, D. (2018). *Estudio de metodologías de ingeniería social*. Tesis de Maestría. Universitat Oberta de Catalunya, Catalunya. Repositorio Institucional UOC: <https://openaccess.uoc.edu/items/546286bd-5c70-4efc-8393-1b4948810e98#page=1>

- Borrero-Tigreros, D., & Bedoya-Leiva, O. (2020). Predicción de riesgo crediticio en Colombia usando técnicas de inteligencia artificial. *UIS Ingenierías*, 19(4), 37-52.
<https://doi.org/https://doi.org/10.18273/revuin.v19n4-2020004>
- Calcagni, L. R. (2020). *Redes Generativas Antagónicas y sus aplicaciones*. Tesis de Posgrado. Universidad Nacional de La Plata, Buenos Aires. Repositorio SEDICI:
https://sedici.unlp.edu.ar/bitstream/handle/10915/101507/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y
- Centro Criptológico Nacional. (2023, octubre). *Aproximación a la inteligencia artificial y la ciberseguridad (CCN-CERT BP/30)*. <https://www.ccn-cert.cni.es/es/informes/informes-de-buenas-practicas-bp/7190-ccn-cert-bp-30-aproximacion-a-la-inteligencia-artificial-y-la-ciberseguridad/file.html>
- CERT.ar. (2023, febrero). *Incidentes informáticos: Informe anual de incidentes de seguridad informática registradas en el 2022*. Portal oficial del Estado argentino:
https://www.argentina.gob.ar/sites/default/files/2023/02/informe_cert_2022.docx.pdf
- Crypto.com. (2024, octubre 28). *Las estafas de criptomonedas con IA: consejos para protegerte*. <https://crypto.com/es/university/ai-scams-in-cryptocurrency>
- Diario Clarín. (2025, agosto 8). Las 10 recomendaciones del Banco Central para evitar ser víctima de estafas virtuales. *Clarín*. Clarín: https://www.clarin.com/sociedad/10-recomendaciones-banco-central-evitar-victima-estafas-virtuales_0_DH1MG3f8Hh.html
- Díaz Gutiérrez, L. F., González Martínez, N. P., Pamplona Ramírez, S. P., & Pedraza Jaramillo, S. (2024). *Impacto de la inteligencia artificial en el sector bancario al consumidor*. Tesis de Grado. Universidad EAN, Bogotá. Repositorio Institucional Ean:
<https://repository.universidadean.edu.co/server/api/core/bitstreams/1b1db5ec-8c13-451a-a32b-28c0a299e1bf/content>
- Dirección Nacional de Ciberseguridad. (2023). *Glosario de Ciberseguridad*. Jefatura de Gabinete de Ministros: <https://servicios.infoleg.gob.ar/infolegInternet/anexos/325000-329999/328599/res1523-2.pdf>
- FasterCapital. (2024, junio). *Sistemas de detección de fraude en Regtech: Descubrir crímenes financieros*. <https://fastercapital.com/es/contenido/Sistemas-de-deteccion-de-fraude-en-Regtech--descubrir-crimenes-financieros.html>

- Fernandez, M. M., Nuñez, A. E., & Portillo Torres, R. (2023). *Ataques a un sistema de detección de intrusiones mediante Redes Generativas Adversarias*. Tesis de Grado. Universidad Complutense de Madrid, Madrid. Repositorio Docta Complutense: <https://docta.ucm.es/rest/api/core/bitstreams/6eefcdad-6d29-4ebf-97b0-99a5c8acba3b/content>
- Fintech Americas. (2023, noviembre 15). *6 buenas prácticas para prevenir el fraude financiero utilizando Inteligencia Artificial*. <https://blog.fintechamericas.co/6-buenas-practicas-para-prevenir-el-fraude-financiero-utilizando-inteligencia-artificial>
- García Peña, M., Navarro Cadavid, A., Osorio Quintero, J. L., & Urcuqui López, C. C. (2018). *Ciberseguridad: un enfoque*. Tesis de Posgrado. Universidad ICESI, Cali. Universidad ICESI: https://la.regions.comsoc.org/wp-content/uploads/sites/68/2020/04/navarro_ciberseguridad_ciencia_2018.pdf
- García-Ull, F. J. (2020). *Deepfakes: El próximo reto en la detección*. Tesis de Grado. Universidad Europea de Valencia, Valencia. Repositorio Institucional: https://d1wqtxts1xzle7.cloudfront.net/67828796/3378_14983_2_PB-libre.pdf?1625129327=&response-content-disposition=inline%3B+filename%3DDeepfakes_el_proximo_reto_en_la_deteccio.pdf&Expires=1759264663&Signature=U6GsQ1ZBbU0KGFKkNG6R9wIB9V5XEINavprb3QfYbCtijU
- Grapheverywhere. (2024). *Fraude bancario | Métodos tradicionales de detección*. <https://www.grapheverywhere.com/fraude-bancario-metodos-tradicionales-de-deteccion/>
- Hernández Sampieri, R., Baptista Lucio, P., & Fernandez Collado, C. (2014). *Metodología de la Investigación* (6ta ed.). Mc Graw Hill Education.
- Jefatura de Gabinete de Ministros. (2021, octubre). *Phishing: Una guía y un glosario para conocer sus modalidades y prevenirlas*. Portal oficial del Estado argentino: https://www.argentina.gob.ar/sites/default/files/2021/10/informe_phishing_1.pdf
- Leguiza, C. G. (2024). *Inteligencia artificial “aplicada a los procesos contables”*. Tesis de Grado. Universidad Nacional de San Martín, Buenos Aires. Repositorio Institucional UNSAM: <https://ri.unsam.edu.ar/bitstream/123456789/2597/1/TFPP%20EEYN%202024%20LCG.pdf>

- Martín Martín, M. L. (2024). *Inteligencia artificial: Un estudio de su impacto*. Tesis de Grado. Universitat Oberta de Catalunya, Catalunya. Repositorio Institucional UOC:
<https://openaccess.uoc.edu/server/api/core/bitstreams/068e07b9-c15b-4207-a38e-3e097a81fde8/content>
- McClintock, M. (2023, septiembre 27). *Cómo implantar la automatización de procesos y la IA: Mejores prácticas y dificultades que hay que evitar*. Processmaker:
<https://www.processmaker.com/es/blog/how-to-implement-process-automation-and-ai-best-practices-and-pitfalls-to-avoid/>
- Ministerio de Justicia de la Nación. (2008, junio 4). *Ley 26.388: Modificación del Código Penal sobre delitos informáticos*.
<https://servicios.infoleg.gob.ar/infolegInternet/anexos/140000-144999/141790/norma.htm>
- Monroy Gomez, J. L., Uricoechea Salazar, E. G., & Urrutia Bonilla, J. A. (2024). Evolución de la inteligencia artificial en la prevención de fraudes bancarios en Colpatria. *Título de Grado*. Universidad EAN, Bogotá.
- Moreno, H. T. (1966). *Los fraudes en documentos*. Tesis Doctoral. Universidad de Buenos Aires, Buenos Aires. Repositorio Digital Econ-UBA:
http://bibliotecadigital.econ.uba.ar/download/tesis/1501-0849_MorenoHT.pdf
- Nocetti, N. (2019). *Promesas de la inteligencia artificial y el aprendizaje automático para la banca tradicional privada. Un análisis de oportunidades, aplicaciones, barreras y riesgos*. Tesis de Maestría. Universidad de San Andrés, Buenos Aires. Repositorio Institucional UdeSA:
<https://dspaceapi.live.udesa.edu.ar/server/api/core/bitstreams/2309518e-9e30-43bf-ba52-1a845d18b238/content>
- Ribón, M. M. (2019, enero 23). La estafa informática en el Código Penal Argentino. *Revista Pensamiento Penal*.
<https://www.pensamientopenal.com.ar/system/files/2019/01/doctrina47322.pdf>
- Rouhiainen, L. (2018). *Inteligencia atificial 101 cosas que debes saber hoy sobre nuestro futuro*. Alienta Editorial.
https://planetadelibrosar0.cdnstatics.com/libros_contenido_extra/40/39307_Inteligencia_artificial.pdf

- SafetyCulture. (2024, enero 15). *Análisis predictivo*. <https://safetyculture.com/es/temas/analisis-predictivo/>
- Tanant, F. (2024). *Machine learning para detectar fraude*. SEON: <https://seon.io/es/recursos/machine-learning-para-detectar-fraude/>
- Teixido, G. (2021). Ciberdelito y vulnerabilidad del Sistema. *Tesina*. Universidad Nacional de Cuyo, Mendoza. Repositorio BDigital: https://bdigital.uncu.edu.ar/objetos_digitales/18493/teixido-ciberdelito.pdf
- uFlow. (2024). *Detección de fraude ¿Cómo usar el aprendizaje automático (IA) en bancos y fintechs?* <https://uflow.biz/blog/deteccion-de-fraude-usar-el-aprendizaje-automatico-ia-en-bancos-fintechs/>

Anexos

Anexo 1. Resultados de la Encuesta (Análisis Cuantitativo)

El siguiente Anexo presenta los resultados obtenidos a partir de la encuesta realizada a 137 personas. La encuesta se organizó en 3 secciones claves: El perfil del usuario, conocimiento sobre amenazas y experiencias personales.

Link:

<https://docs.google.com/forms/d/e/1FAIpQLSckqKMmCc8GHzeY0IMxljDGgAMMTmxzltV82vDZe4X5r-Zfrg/viewform?usp=header>

Sección A: El perfil del usuario

1. ¿Cuál es su rango de edad?

18 a 25 años = 7,3%

26 a 40 años = 50,4%

41 a 60 años = 39,4%

Más de 60 años = 2,9%

2. ¿Cuál es su máximo nivel de estudios alcanzado?

Primario = 0,7%

Secundario = 12,4%

Terciario / Universitario = 73%

Posgrado = 13,9%

3. ¿Con qué frecuencia utiliza la banca online (home banking o apps del banco)?

Todos los días = 48,9%

Varias veces por semana = 36,5%

Algunas veces al mes = 10,9%

Casi nunca = 3,6%

Sección B: Conocimientos sobre amenazas

4. Usted recibe un correo electrónico de su banco que le informa sobre un "bloqueo de su cuenta" y le pide hacer clic en un enlace para verificar su identidad. ¿Qué acción considera más segura?

Hacer clic en el enlace y verificar mis datos para desbloquearla rápido = 1,5%

No hacer clic, pero responder el correo pidiendo más información = 2,9%

Ignorar el correo, no hacer clic en nada y contactar al banco por sus canales oficiales = 56,2%

Eliminar el correo inmediatamente = 39,4%

5. Recibe una llamada telefónica. La persona se presenta como un empleado del banco y le dice que necesita su clave para cancelar una compra sospechosa que usted no realizó. ¿Qué debería hacer?

Darle la clave, ya que es para cancelar un fraude en mi contra = 0,7%

Pedirle datos para asegurarse de que es un empleado real = 2,2%

Cortar la llamada y contactar al banco por mi cuenta para verificar la situación = 94,9%

Pedirle que me llame en 5 minutos para pensarlo = 2,2%

6. ¿Ha oído hablar de la tecnología "deepfake", que permite crear videos o audios falsos de una persona diciendo o haciendo algo que nunca hizo?

Sí, la conozco bien = 25,5%

He oído hablar de ella, pero no sé mucho = 43,1%

No, nunca había oído hablar de ella = 31,4%

7. ¿Cuál de las siguientes considera que es la forma más segura de acceder al home banking?

A través de un enlace guardado en mis "Favoritos" del navegador = 32,8%

Escribiendo la dirección del banco (ej:www.mibanco.com.ar) directamente en el navegador = 61,3%

Buscando el nombre del banco en Google y haciendo clic en el primer resultado = 5,8%

8. Si una aplicación en su celular le pide permisos para acceder a sus contactos, micrófono y mensajes para funcionar. ¿Qué hace usted generalmente?

Acepto todos los permisos sin leerlos para usar la app = 20,4%

Reviso los permisos y solo acepto los que me parecen necesarios = 36,5%

No instalo la aplicación si me pide permisos que no parecen tener relación con su función = 43,1%

Sección C: Experiencias Personales

9. En los últimos 12 meses, ¿ha recibido algún correo electrónico o mensaje de WhatsApp/SMS que le pareciera sospechoso y que aparentara ser de su banco?

Sí, varias veces = 21,9%

Sí, una o dos veces = 30,7%

No, que yo recuerde = 46%

No estoy seguro/a = 1,5%

10. Si alguna vez recibió un mensaje sospechoso, ¿qué hizo? (Marque la opción más frecuente)

Lo eliminé sin abrirlo = 58,4%

Lo abrí, pero no hice clic en ningún enlace = 36,5%

No he recibido mensajes sospechosos = 9,5%

11. ¿Alguna vez ha recibido una llamada telefónica sospechosa de alguien que decía ser de su banco o de una billetera virtual?

Si = 46%

No = 39,4%

No estoy seguro/a = 14,6%

12. En general, ¿qué tan seguro/a se siente al realizar operaciones bancarias por internet?

Muy seguro/a = 16,8%

Bastante seguro/a = 65%

Poco seguro/a = 14,6%

Nada seguro/a = 3,6%

13. ¿Cree usted que la Inteligencia Artificial hace que las estafas bancarias sean más difíciles de detectar para un usuario común?

Sí, mucho más difíciles = 40,1%

Sí, un poco más difíciles = 43,1%

No, no creo que influya = 4,4%

No estoy seguro/a = 12,4%

14. ¿Utiliza el mismo usuario y contraseña para su home banking que para otras cuentas (redes sociales, email, etc.)?

Sí, para casi todo = 13,1%

A veces, para algunas cosas = 32,8%

No, uso contraseñas diferentes para cada sitio importante = 54%

15. ¿Ha sido usted o un familiar cercano víctima de una estafa bancaria en la que hayan perdido dinero?

Si = 48,2%

No = 51,8%

Anexo 2. Resultados de la Entrevista (Análisis Cualitativo)

Con el objetivo de enriquecer los datos obtenidos en la encuesta, se realizó una entrevista a una experta del Sector de Fraude de uno de los bancos más reconocidos de Argentina. El objetivo fue obtener una perspectiva cualitativa sobre las amenazas y defensas. Las preguntas a continuación fueron diseñadas para profundizar en la realidad operativa del fraude potenciado por IA, los desafíos prácticos de la industria y las estrategias de mitigación implementadas.

1. Desde tu perspectiva, ¿cuál ha sido el cambio más significativo en las modalidades de fraude bancario en los últimos 5 años?

En estos años el cambio más notorio fue la industrialización del fraude. Pasamos de ataques aislados a bandas organizadas que combinan ingeniería social, automatización e incluso IA para orquestar todo el proceso: desde el contacto con la víctima hasta el cash-out.

Hoy el fraude es más rápido, más barato y mucho más personalizado que antes.

2. En el contexto argentino, ¿cuáles son las amenazas impulsadas por IA que más preocupan actualmente (phishing personalizado, vishing con voz sintética, deepfakes, etc.)?

Lo que más preocupa hoy son los phishing hiperrealistas. Se ven correos casi imposibles de distinguir de los reales. También empiezan a aparecer intentos con deepfakes en validaciones de identidad o apertura de cuentas, y correos corporativos escritos con IA sin errores ni señales típicas de fraude, sumado a que hay situaciones que las empresas tienen como validas (gestión de encargos y

mensajería) que utilizan exactamente lo mismo que el defraudador.

3. ¿Qué tan grande es la brecha entre el conocimiento del público sobre estas nuevas estafas y las amenazas reales que ustedes gestionan a diario?

Es grande. La mayoría de las personas sigue pensando que “se nota” cuándo algo es falso, y eso ya no es cierto. Con unas pocas muestras de fotos o un posteo en redes, se puede utilizar una foto para realizar un buen video y usurpar identidad (deepfake). Nosotros lo vemos todos los días, pero el usuario promedio sigue creyendo que el peligro está solo en los mails mal escritos.

4. Phishing Sofisticado: Sin revelar información sensible, ¿podría describir qué características hacen que una campaña de phishing moderna, potenciada por IA, sea tan efectiva y difícil de detectar para un cliente?

Lo que los vuelve tan efectivos es que usan lenguaje local, horarios habituales y coherencia entre canales. Por ejemplo, un mail que llega justo después de que el cliente usó el home banking, seguido de un SMS y una llamada “de verificación”. Además, las páginas falsas están tan bien logradas que incluyen hasta el candado de seguridad o un chat en vivo con IA.

5. Vishing y Clonación de Voz: ¿Han detectado un aumento en los intentos de estafa que utilizan IA para clonar voces? ¿Qué tan convincentes son estos audios y cuál es el “cuento del tío” más recurrente que utilizan?

Sí, hubo un aumento. Las voces clonadas son cada vez más creíbles, aunque por suerte, no estamos sufriendo en nuestra entidad (todavía) sobre todo si la víctima ya escuchó antes a ese supuesto “empleado del banco”. El cuento más común sigue siendo el del “departamento de seguridad” que detectó un acceso o transferencia y necesita validar el token. Otras veces se hacen pasar por familiares pidiendo ayuda urgente.

6. Deepfakes y Suplantación de Identidad: El uso de deepfakes para vulnerar procesos de validación de identidad (como en la apertura de una cuenta), ¿es una amenaza que ya están viendo en la práctica o es un riesgo emergente que anticipan?

Todavía son pocos los casos confirmados, pero es una amenaza real y en crecimiento. Ya hubo fraudes usando vídeos manipulados para validaciones de identidad u onboarding remoto. Hoy se considera un riesgo emergente, en crecimiento y actual; y se están reforzando los controles de liveness y textura de imagen para prevenirlo. Es de vital importancia contar con contingencias en cuanto al monitoreo transaccional para mitigar estos casos.

En un corto plazo, no se va a poder confiar en los motores biométricos por sí solos, se tendrán que ver reforzados o bien por un factor adicional, o por un backoffice que revise los videos selfies, o hasta se pueden demorar las acciones posteriores para verificar la identidad, esto hace 4 años no era así.

7. Ataques a los Sistemas Internos: ¿Cómo intentan los atacantes utilizar la IA para “engañar”; o evadir los sistemas del banco?

Los atacantes usan IA para probar variaciones y ver qué combinaciones logran evadir modelos o reglas. También generan tráfico automatizado con bots que imitan patrones humanos y se conectan desde IPs residenciales, para parecer clientes reales. Algunos llegan a usar modelos de lenguaje para revisar sus propios correos antes de enviarlos, buscando sonar más creíbles.

8. ¿De qué manera el banco utiliza la Inteligencia Artificial para defenderse de estas nuevas amenazas?

Del lado de la defensa, la IA se usa para analizar comportamiento, dispositivo, red y patrones transaccionales en tiempo real. Considero que los delincuentes están mucho más avanzados en el uso de IA para delinquir que nosotros para defendernos

o mitigar.

9. ¿Cuál considera que es el mayor desafío técnico al que se enfrentan para poder detectar un fraude generado por IA?

El principal desafío es distinguir a una persona real de un contenido sintético sin generar fricción. Es decir, proteger al cliente sin hacerlo pasar por un proceso engorroso. Cada año los atacantes logran suplantaciones más naturales, así que el equilibrio entre precisión y experiencia de usuario es lo más difícil.

10. ¿Qué rol juega la educación y concientización del cliente en la estrategia general de prevención del banco?

Es un pilar fundamental. Ningún sistema puede prevenir todo si el cliente no está informado. En ese sentido, las campañas de concientización —tanto las del BCRA como las propias de cada entidad— son esenciales. La educación dentro de las mismas apps, con mensajes contextuales y alertas preventivas, funciona mejor que los folletos tradicionales.

11. ¿Cuál crees que será la mayor amenaza de fraude bancario que enfrentaremos en los próximos 5 años?

Probablemente veamos fraudes completamente conversacionales, con IA que mantiene un diálogo en tiempo real por voz o video, imitando a empleados o familiares. Es un escenario donde la línea entre lo real y lo falso se va a volver muy fina. Por eso es clave invertir en detección de este tipo de amenazas y en reforzar la educación del usuario.

12. Finalmente, ¿qué recomendación clave le daría a un cliente para protegerse en este nuevo y complejo escenario?

La regla más simple sigue siendo la más efectiva: ante cualquier pedido de claves,

OTP o transferencias, cortá y volvé a comunicarte al número oficial. Nadie del banco ni del BCRA te va a pedir un código ni que instales una app por fuera del store oficial. Activar la doble autenticación y mantener límites diarios también ayuda muchísimo.

Anexo 3. Ejemplos Adicionales de Phishing Sofisticado

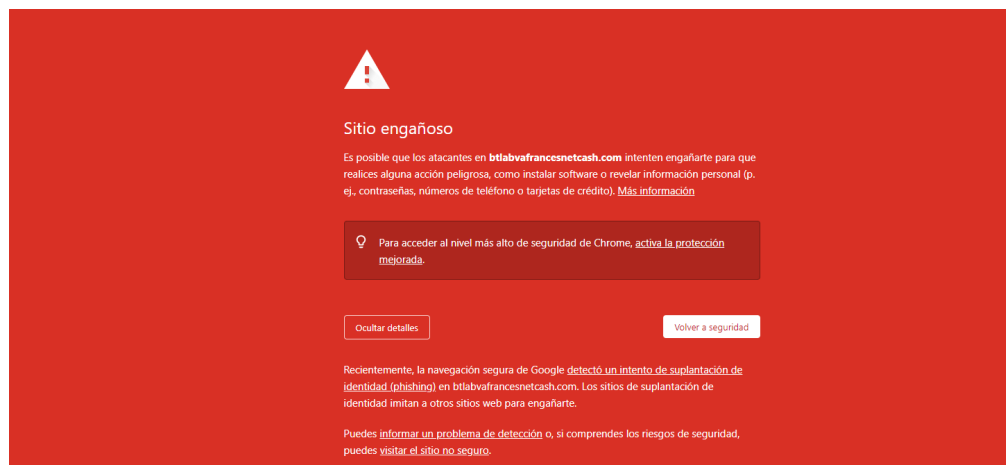
Este anexo complementa a la Figura 2 presentando ejemplos visuales adicionales de correos electrónicos de phishing detectados en el entorno bancario real. Estas capturas fueron enviadas por la experta entrevistada para esta investigación.

Figura 100

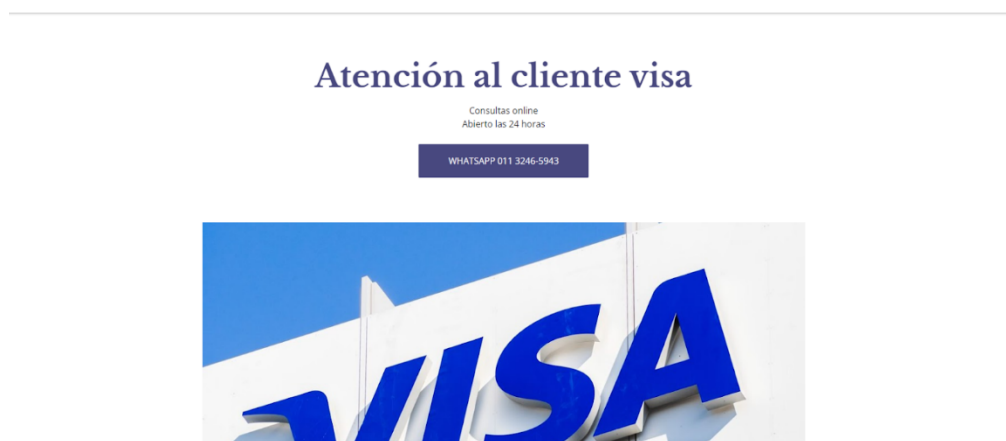
Página Falsa Net Cash – Banca Empresa BBVA 1



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 111*Página Falsa Net Cash – Banca Empresa BBVA 2*

Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 122*Página Falsa – Centro de Atención VISA que afecta a Clientes BBVA*

Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Anexo 4. Ejemplos de Campañas de Concientización Bancaria

Este anexo reúne material utilizado por entidades bancarias en sus campañas de concientización y educación dirigidas a clientes. Estas imágenes completan a las Figuras 5, 6 y 7.

Figura 133

Ejemplo 4 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 144

Ejemplo 5 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 155

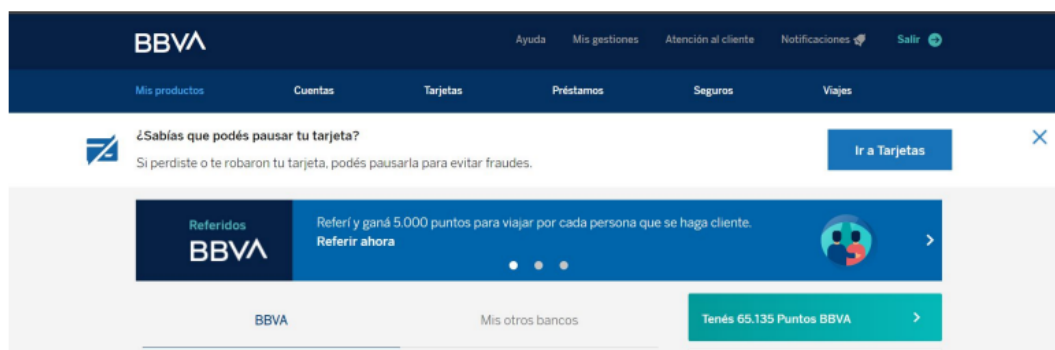
Ejemplo 5 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 166

Ejemplo 6 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 177

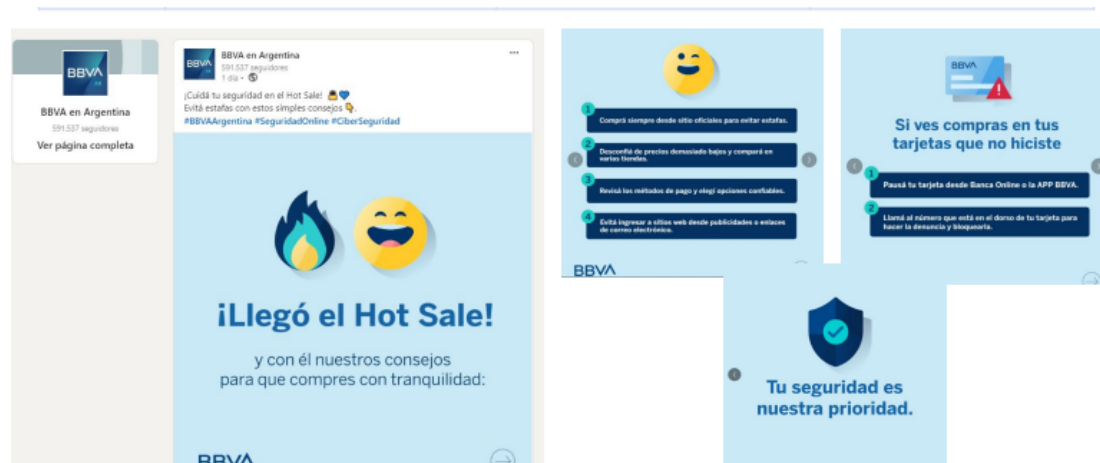
Ejemplo 7 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>

Figura 188

Ejemplo 8 de Campaña de Concientización Bancaria



Nota. BBVA. (2025). Material de concientización sobre ciberseguridad. <https://www.bbva.com.ar>