Journal of the Academy
Asociación de Universidades del Perú, Lima, Perú.
ISSN 2707-0301
Núm. 15. Julio-Diciembre 2026
http://journalacademy.net/

PROPUESTA CONCEPTUAL DE UN MODELO ADAPTATIVO BASADO EN INTELIGENCIA ARTIFICIAL PARA LA DETECCIÓN DE FRAUDE EN TRANSACCIONES INTERNACIONALES: INNOVACIONES, RIESGOS EMERGENTES Y APLICABILIDAD EN PYMES DE LATINOAMÉRICA

A CONCEPTUAL AI-BASED ADAPTIVE MODEL FOR FRAUD DETECTION IN INTERNATIONAL TRANSACTIONS: INNOVATIONS, EMERGING RISKS, AND APPLICABILITY TO LATIN AMERICAN SMES

Vanessa SÁNCHEZ DEL CASTILLO[1]

Recibido	:	15/06/2026
Aprobado	:	28/09/2026
Publicado	:	13/09/2026

RESUMEN:
El crecimiento de las transacciones digitales internacionales ha incrementado la exposición de las pequeñas y medianas empresas (PYMES) al fraude financiero, en un contexto en el que la inteligencia artificial (IA) puede emplearse tanto para combatir como para perpetrar este delito. En México, se reportaron pérdidas por un monto aproximado de 17,400 millones de pesos durante el primer semestre de 2025. Este trabajo propone, a nivel conceptual, un modelo adaptativo de detección de fraude en la modalidad de Software como Servicio (SaaS), dirigido a PYMES latinoamericanas. Se empleó una metodología mixta de alcance exploratorio-descriptivo, que combinó una revisión documental narrativa, el análisis descriptivo de registros operativos agregados de una organización de remesas (de octubre de 2024 a enero de 2025) y el diseño de una arquitectura de entrada, procesamiento y salida de datos. En el caso analizado, 103.720 llamadas fueron derivadas al área de fraude y solo el 33,53 % se confirmaron como tales, lo que evidencia la fricción que generan los sistemas basados en reglas fijas. El modelo propuesto combina aprendizaje supervisado, aprendizaje no supervisado y revisión humana, y está diseñado para reducir esa fricción sin descuidar la seguridad. Su desempeño no fue medido en este estudio y requiere una validación empírica posterior.

Palabras clave: inteligencia artificial, fraude financiero, negocios internacionales, PYMES, aprendizaje automático, Industria 5.0.

[1] Estudiante de Lic. En Negocios Internacionales. Universidad del Valle de México (UVM), Campus Puebla, Camino Real a San Andrés Cholula No. 4002, Emiliano Zapata, San Andrés Cholula 72810, México A830111224@my.uvm.edu.mx - https://orcid.org/0009-0003-0085-5885

ABSTRACT:
The growth of digital international transactions has increased small and medium-sized enterprises' (SMEs) exposure to financial fraud, in a context where artificial intelligence (AI) can be used both to combat and commit this crime. In Mexico, losses of approximately 17.4 billion pesos were reported during the first half of 2025. This paper proposes, at a conceptual level, an adaptive fraud detection model under the Software as a Service (SaaS) modality aimed at Latin American SMEs. We used a mixed-methods approach with an exploratory-descriptive scope, combining a narrative documentary review, a descriptive analysis of aggregated operational records from a remittance organization (October 2024 to January 2025), and the design of an input, processing, and output data architecture. In the case analyzed, 103,720 calls were referred to the fraud department, and only 33,53% were confirmed as fraud, illustrating the friction created by fixed-rule systems. The proposed model combines supervised learning, unsupervised learning, and human review, and is designed to reduce that friction without weakening security. This study did not measure performance and requires subsequent empirical validation.

Keywords: artificial intelligence, financial fraud, international business, SMEs, machine learning, Industry 5.0.

INTRODUCCIÓN

La digitalización y la globalización han impulsado el crecimiento de las transacciones internacionales, en particular a través de plataformas digitales. Ese avance ha traído consigo un aumento del riesgo de fraude financiero. Según la Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros (CONDUSEF, 2025), al menos el 80 % de los usuarios mexicanos ha enfrentado algún intento de estafa; durante el primer semestre de 2025 se registraron 3.5 millones de reclamaciones, de las cuales el 59 % se confirmaron como fraude cibernético, con pérdidas aproximadas de 17,400 millones de pesos.

Las pequeñas y medianas empresas (PYMES) son especialmente vulnerables. La Organización para la Cooperación y el Desarrollo Económicos (OECD, 2021) advierte que estas empresas arrastran una brecha tecnológica importante en ciberseguridad y que su limitada disponibilidad de personal especializado dificulta responder a amenazas que cambian con rapidez. A ello se suma un problema de fondo: la inteligencia artificial (IA), que permite detectar anomalías en tiempo real, también se ha convertido en una herramienta para sofisticar el fraude, un rasgo que Kshetri (2021) describe como la dualidad de esta tecnología.

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

La experiencia acumulada en prevención del fraude muestra, además, que muchas víctimas no se dan cuenta de inmediato de que están siendo engañadas. En diversos casos documentados, los perpetradores recurren a técnicas de alta sofisticación —como la clonación de voz o la suplantación de personas de confianza— y aprovechan estados emocionales de vulnerabilidad para consumar el engaño.

La literatura reciente ha avanzado en dos frentes. Por un lado, se han revisado las técnicas de aprendizaje automático para la detección de fraude financiero (Hernández Aros et al., 2024). Por otro lado, se ha discutido la explicabilidad de los modelos, es decir, la posibilidad de comprender y justificar sus decisiones (Černevičienė y Kabašinskas, 2024; Hassija et al., 2024). Sin embargo, en las fuentes consultadas para este trabajo se encontraron pocos planteamientos que integren en un mismo diseño tres elementos decisivos para las PYMES latinoamericanas: una solución accesible, sin infraestructura propia, bajo la modalidad de Software como Servicio (SaaS); la dimensión conductual del fraude, basada en la ingeniería social; y la fricción operativa que sufre el cliente legítimo cuando el sistema retiene sus operaciones. Este trabajo busca contribuir a cubrir ese vacío mediante una propuesta conceptual, apoyada en un caso operativo real del corredor de remesas entre Estados Unidos y Latinoamérica.

La pregunta que orienta el estudio es la siguiente: ¿cómo puede un modelo adaptativo basado en inteligencia artificial fortalecer la detección de fraude financiero en transacciones internacionales realizadas por PYMES latinoamericanas frente a nuevas amenazas, sin generar una fricción excesiva para los usuarios legítimos? El propósito central consiste en proponer, a nivel conceptual, un modelo adaptativo antifraude basado en inteligencia artificial para fortalecer la detección de amenazas emergentes. De manera complementaria, se busca fundamentar el impacto de la dualidad de la IA ante amenazas como los deepfakes y el phishing avanzado; describir, a partir de un caso operativo, las limitaciones de un sistema de detección basado en reglas fijas; estructurar el modelo en tres etapas —entrada, procesamiento y salida de datos—; y estimar de forma preliminar la viabilidad económica y operativa de soluciones antifraude SaaS para PYMES.

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

DESARROLLO

La incorporación del aprendizaje automático a los sistemas antifraude ha cambiado la forma de identificar operaciones ilícitas. La OECD (2021) señala que estas técnicas permiten procesar grandes volúmenes de datos y detectar patrones atípicos con mayor precisión y menos falsas alarmas, especialmente en la banca digital y en las transferencias. Hernández Aros et al. (2024), en su revisión de la literatura, destacan que los modelos de aprendizaje automático reconocen patrones complejos y favorecen la detección temprana de actividades fraudulentas.

Dentro de estas técnicas se distinguen dos familias. El aprendizaje supervisado se entrena con registros históricos etiquetados como legítimos o fraudulentos y resulta eficaz para reconocer patrones recurrentes. El aprendizaje no supervisado, en cambio, identifica comportamientos inusuales sin necesidad de ejemplos previos y, por ello, puede detectar nuevas modalidades de fraude. Samson-Onuorah (2023) sostiene que los mejores resultados se obtienen cuando el aprendizaje automático se combina con el monitoreo continuo y con procesos de gobernanza que permiten supervisar, validar y actualizar los modelos a medida que evolucionan las amenazas.

El Banco Mundial (World Bank) advierte que las mismas herramientas que fortalecen la defensa ya están en manos de los atacantes (Boeddu et al., 2021). Kshetri (2021) lo conceptualiza como una dualidad: la capacidad de la IA para reforzar los mecanismos de protección es equivalente a su potencial para complicar los ataques. Entre las modalidades que más preocupan se encuentran los deepfakes y la suplantación de identidad mediante voz e imágenes sintéticas capaces de evadir sistemas biométricos (Passos et al., 2024); el phishing avanzado, con mensajes personalizados generados por modelos de lenguaje (Boeddu et al., 2021); los bots automatizados que ejecutan operaciones fraudulentas en masa; y el fraude adaptativo, en el que los delincuentes aprenden a evadir los controles con rapidez (Samson-Onuorah, 2023).

En América Latina, el aumento de las transacciones digitales, las remesas y la actividad de las PYMES ha ido acompañado de un crecimiento proporcional en los casos de suplantación de identidad y de uso indebido de cuentas (OECD, 2021). Además, el fraude no es solo un

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

fenómeno técnico: gran parte de los ataques se basa en la manipulación de la confianza y la urgencia de la víctima, una dimensión conductual que cualquier modelo de detección debe considerar.

La IA forma parte de la Cuarta Revolución Industrial, impulsada por la integración de Big Data, la computación en la nube, la inteligencia artificial y la interconectividad. El Big Data es indispensable porque el volumen de transacciones diarias sería imposible de revisar manualmente; su valor radica en convertirlo en conocimiento útil para la toma de decisiones (Hernández Aros et al., 2024). La computación en la nube ha reducido las barreras tecnológicas para las PYMES al brindarles acceso a herramientas avanzadas de monitoreo sin requerir grandes inversiones en infraestructura propia (OECD, 2021). La interconectividad facilita el comercio internacional, pero también multiplica los puntos vulnerables.

La Industria 5.0 parte de ese mismo entorno tecnológico, pero desplaza el foco hacia el ser humano: la tecnología debe complementar sus capacidades y no sustituirlas. Schmalzried et al. (2023) sostienen que, aunque la IA procesa millones de registros en poco tiempo, los especialistas siguen siendo necesarios para interpretar contextos complejos y validar decisiones que pueden afectar a clientes legítimos. Tóth et al. (2023) señalan que la colaboración entre sistemas inteligentes y analistas produce procesos más confiables y resilientes que la automatización exclusiva. Esta postura dialoga con el debate sobre la transparencia de los modelos: a medida que crece la complejidad de los algoritmos, la precisión ya no basta; también importa poder comprender y justificar sus decisiones (Černevičienė y Kabašinskas, 2024; Hassija et al., 2024).

Este trabajo adopta una postura intermedia, alineada con la Industria 5.0: automatizar la respuesta a las operaciones de bajo riesgo y reservar la intervención humana para los casos ambiguos o de alto riesgo. Los sistemas basados en reglas fijas —umbrales de monto o verificación de coincidencia de nombres— pierden eficacia cuando los defraudadores utilizan datos reales robados o fragmentan las cantidades. Los sistemas antifraude deben evolucionar hacia arquitecturas con aprendizaje continuo y actualización constante frente a

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

nuevas amenazas (Samson-Onuorah, 2023), una necesidad especialmente aguda en las PYMES (OECD, 2021).

Sobre esa base, se propone un modelo adaptativo basado en IA en modalidad SaaS, organizado en tres etapas: entrada de datos, procesamiento y salida de resultados. A diferencia de los sistemas tradicionales, que operan con reglas predefinidas, el modelo está diseñado para adaptarse a los cambios constantes en el fraude digital.

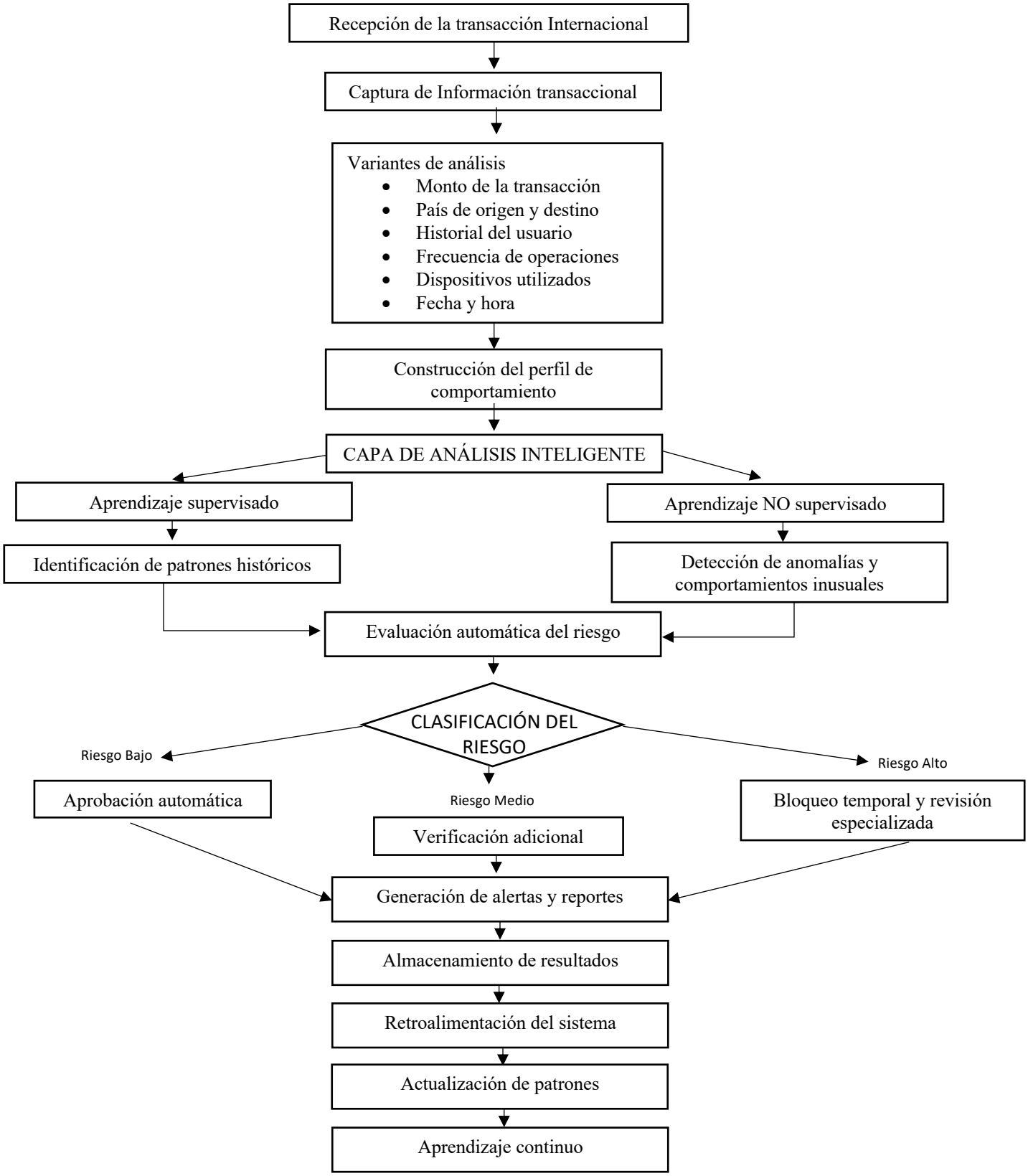
En la etapa de entrada se recopila información de cada transacción para construir un perfil de comportamiento del usuario. Las variables integradas son: el monto, la frecuencia de operaciones, la ubicación geográfica, los países de origen y de destino, el historial del usuario, el dispositivo utilizado y la fecha y la hora. Su combinación permite reconocer comportamientos habituales y detectar desviaciones asociadas a la suplantación de identidad, a accesos no autorizados o a patrones inusuales.

El procesamiento se organiza en tres capas interrelacionadas: adquisición de datos, análisis inteligente y gestión de resultados. En el análisis inteligente se propone un enfoque híbrido. El aprendizaje supervisado reconoce patrones repetitivos a partir de registros históricos etiquetados (Hernández Aros et al., 2024); el no supervisado identifica comportamientos inusuales no registrados previamente, útiles en transacciones internacionales donde surgen modalidades nuevas. Samson-Onuorah (2023) señala que los sistemas más eficaces combinan distintos mecanismos de análisis, monitoreo continuo y actualización permanente. A partir de esta combinación, el sistema identifica patrones habituales, detecta anomalías, evalúa el riesgo y aprende de forma continua. El rasgo central de la propuesta es su capacidad adaptativa: cada transacción, cada alerta confirmada o cada intento de fraude detectado puede ajustar progresivamente los parámetros de análisis (ver Figura 1). Al mismo tiempo, se reconoce el problema de las «cajas negras» (Hassija et al., 2024), por lo que la IA se plantea como una herramienta de apoyo a la toma de decisiones y no como sustituto del criterio humano (Schmalzried et al., 2023; Tóth et al., 2023).

Figura 1

Arquitectura conceptual del modelo antifraude basado en inteligencia artificial para PYMES

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7



Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

La clasificación del riesgo contempla tres niveles. El riesgo alto implica un bloqueo temporal y una revisión manual especializada, sin cancelación definitiva de la operación. El riesgo medio requiere autenticación adicional o un breve contacto de verificación. El riesgo bajo permite la aprobación inmediata. Esta clasificación concentra la supervisión en las operaciones potencialmente riesgosas sin afectar significativamente la experiencia de los usuarios legítimos.

Como salida, el modelo genera alertas de riesgo en tiempo real, clasificaciones de transacciones, respuestas automatizadas y reportes para análisis posterior. La principal aportación radica en el cambio de la validación estática a un análisis conductual: en lugar de un bloqueo definitivo, el sistema deriva la operación sospechosa a una lista de revisión prioritaria, donde un analista puede contactar al cliente para una verificación breve; si no hay respuesta, se envían mensajes automáticos. Así se busca equilibrar la seguridad y la experiencia del usuario. Todo ello son propiedades de diseño que deberán verificarse empíricamente.

METODOLOGÍA

La investigación es de tipo exploratorio-descriptivo, con un enfoque mixto de predominio cualitativo. Se desarrolló en tres momentos complementarios: una revisión documental de la literatura sobre fraude financiero e inteligencia artificial; el análisis descriptivo de los registros operativos de una organización dedicada a las remesas; y el diseño conceptual de un modelo antifraude adaptativo bajo la modalidad SaaS, contrastado teóricamente con la evidencia reunida.

Conviene precisar el alcance desde el inicio. En este trabajo no se desarrolló, entrenó ni probó ningún algoritmo con datos reales. Por lo tanto, el desempeño del modelo —precisión, reducción de falsos positivos, tiempos de respuesta— no fue medido. Lo que se presenta son características de diseño y beneficios esperados, sujetos a validación empírica posterior. Los únicos resultados observados provienen de la revisión documental y del caso organizacional.

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

La revisión documental fue de carácter narrativo, no sistemática. La selección de las fuentes se orientó con el apoyo de un profesor asesor y se basó en el conocimiento previo de la autora y en su experiencia profesional en prevención del fraude. No se aplicó un protocolo formal con bases de datos, ecuaciones de búsqueda ni un diagrama de flujo de selección; en consecuencia, el estudio no afirma la exhaustividad ni la reproducibilidad de la búsqueda. Se consultaron artículos científicos con DOI publicados en revistas especializadas en finanzas, tecnología y gestión del riesgo, así como documentos de organismos internacionales y entidades reguladoras: OECD (2021), World Bank (Boeddu et al., 2021), Organización Mundial del Comercio (2024), Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC, 2022) y CONDUSEF (2025). Los criterios aplicados fueron la pertinencia temática, la actualidad —prioridad a publicaciones de los últimos cinco años— y la posibilidad de identificar la autoría y de verificar la fuente. Los informes institucionales se utilizaron para dimensionar el problema y contextualizar las cifras, no como evidencia del desempeño de los algoritmos.

El material se organizó en cuatro ejes: técnicas de aprendizaje automático aplicadas a la detección de fraude; uso de la IA con fines fraudulentos; explicabilidad y supervisión humana en el marco de la Industria 5.0; y condiciones de adopción tecnológica en las PYMES. Se contrastaron posturas divergentes: Hernández Aros et al. (2024) destacan la capacidad del aprendizaje automático para identificar patrones complejos, mientras que Hassija et al. (2024) y Černevičienė y Kabašinskas (2024) advierten que la complejidad compromete la transparencia. Esa discusión respaldó la decisión de plantear la IA como apoyo al criterio humano y no como su sustituto. Una revisión sistemática con protocolo PRISMA se plantea como una línea de trabajo futura.

En cuanto al caso organizacional, se analizaron registros operativos de una organización transnacional dedicada a las remesas, que opera desde Estados Unidos hacia México, Colombia, Perú y Chile. Por razones de confidencialidad, su identidad se mantiene en reserva. Los registros corresponden al periodo comprendido entre octubre de 2024 y el 15 de enero de 2025 —aproximadamente tres meses y medio, que incluyen la temporada de fin de año—. La autora consolidó los reportes mensuales hasta obtener los totales del periodo.

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

Trabajó más de un año en la organización, en el área de atención y prevención de fraude, y fue responsable de la elaboración y edición de dichos reportes. Para este estudio, un colaborador que continúa en la organización le proporcionó únicamente los datos estrictamente necesarios, en forma agregada y sin nombres, cuentas ni información de clientes. Por su carácter reservado, los registros originales no pueden divulgarse ni verificarse de forma independiente, lo cual se reconoce como una limitación.

Las tres cifras del caso comparten la misma unidad de análisis, la llamada. Se entiende por interacción la llamada atendida en los canales generales de servicio al cliente o de soporte; por derivación a fraude, la llamada monitoreada cuyo caso se transfirió al área de fraude, donde se realizaba un cotejo para validar la posible irregularidad; y por caso confirmado, la derivación en la que el cotejo confirmó el fraude o en la que el reclamo se formalizó, según el criterio de la propia organización. Las derivaciones que no derivaron en fraude confirmado no se consideran, en su totalidad, falsos positivos. De acuerdo con la experiencia de la autora, este grupo incluía inconformidades con el servicio, errores operativos, falsas alarmas de fraude y usuarios que abandonaron la verificación por falta de tiempo o por la demora en la respuesta. Como no se dispone de un desglose cuantitativo, este grupo se denomina «derivaciones no confirmadas».

Se revisaron, además, más de 100 reclamaciones para identificar tipologías de estafa recurrentes y patrones estacionales. Se trata de una revisión de casos con fines exploratorios, no de una muestra estadísticamente representativa. Las observaciones derivadas de la experiencia operativa de la autora se utilizaron únicamente para interpretar comportamientos recurrentes y complementar los registros, sin sustituirlos.

A partir de lo anterior se diseñó el modelo adaptativo. Las variables de entrada se eligieron por su recurrencia en la literatura revisada y por su relación con las modalidades de fraude observadas en el caso. La arquitectura se organizó en tres etapas —entrada, procesamiento y salida—, con un enfoque híbrido de aprendizaje supervisado y no supervisado, respaldado por Samson-Onuorah (2023) y Hernández Aros et al. (2024). El contraste consistió en comprobar la correspondencia entre cada vector de fraude identificado en el caso, cada componente del modelo y lo que sostienen la literatura y los organismos especializados. Se

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

trata de una evaluación de coherencia teórica y no de una validación predictiva o empírica. La estimación de costos se elaboró con precios públicos de proveedores tecnológicos consultados durante el primer semestre de 2025, bajo el supuesto de una PYME de aproximadamente 20 colaboradores, y constituye una aproximación preliminar, no un análisis costo-beneficio.

RESULTADOS

La organización analizada funciona como un corredor de flujos financieros de alto volumen desde Estados Unidos hacia Latinoamérica. Entre octubre de 2024 y el 15 de enero de 2025 se registraron las cifras que resume la Tabla 1.

Tabla 1

Indicadores operativos del caso (octubre de 2024 al 15 de enero de 2025)

Indicador	Registros	Proporción
Interacciones atendidas en canales de servicio	256,943	100 %
Derivaciones al área de fraude para cotejo	103,720	40.37 % de las interacciones
Casos de fraude confirmados	34,782	33.53 % de las derivaciones; 13.54 % de las interacciones
Derivaciones no confirmadas como fraude	68,938	66.47 % de las derivaciones

Nota. Elaboración propia con datos agregados y anonimizados de una organización de remesas cuya identidad se mantiene en reserva.

Cuatro de cada diez llamadas monitoreadas se derivaron al área de fraude, pero solo una de cada tres derivaciones se confirmó como fraude. Este dato revela una ineficiencia estructural: el sistema de reglas de la organización retenía un gran número de casos que, tras el cotejo, no constituían fraude. Como se explicó en la metodología, ese grupo no equivale por completo a falsos positivos, porque incluye inconformidades con el servicio, errores operativos y usuarios que abandonaron el proceso.

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

El abandono merece atención propia. Según la observación de la autora, varios clientes retenidos decidían no continuar con la verificación por falta de tiempo o por la lentitud de la respuesta. En cuanto a los plazos, la organización comunicaba al personal una política de reembolso de 21 días, pero el área que daba seguimiento a los reembolsos solía tardar más de dos meses en validar su procedencia. Estos tiempos provienen de la experiencia directa de la autora y no de los registros analizados.

El análisis cualitativo de las reclamaciones mostró que las redes delictivas no buscaban vulnerar el código del sistema, sino explotar la psicología y los sesgos del usuario. Se identificaron cuatro tipologías recurrentes de ingeniería social. La primera, conocida como relationship scam, consiste en una manipulación afectiva prolongada en la que el estafador construye un vínculo de confianza para solicitar transferencias bajo el pretexto de emergencias médicas o de la compra de boletos de viaje. La segunda es el fraude en el comercio electrónico, que consiste en sitios fraudulentos que capturan dinero mediante ofertas de productos inexistentes. La tercera corresponde a los falsos descuentos, que presionan mediante amenazas de corte inmediato de servicios esenciales si no se realiza un depósito rápido a cambio de una supuesta rebaja. La cuarta, y más agresiva, es el scam de inmigración: los defraudadores usan datos personales reales, obtenidos de filtraciones previas, para suplantar a funcionarios de agencias gubernamentales y forzar transferencias inmediatas bajo amenaza de deportación.

Los montos defraudados no seguían un patrón específico; se diversificaban para evadir los límites establecidos. En cambio, sí se observó estacionalidad, con repuntes en fechas de alta carga transaccional y emocional, como Navidad, el Día de las Madres, el Super Bowl, así como en coyunturas políticas y sociales de cada país de destino. Cabe precisar que el periodo cuantitativo analizado abarca únicamente el cierre de año; los repuntes de otras fechas son observaciones de la autora que no se cuantificaron en este estudio.

La Tabla 2 relaciona cada vector de fraude observado con la limitación del sistema tradicional y con el componente del modelo propuesto diseñado para abordarlo.

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

Tabla 2

Vectores de fraude observados y componentes del modelo propuesto

Vector observado	**Limitación del sistema de reglas**	**Componente del modelo propuesto (diseño)**
Suplantación con datos reales e inmigración scam	Verificación manual de coincidencia de nombres, inútil si el delincuente ya tiene los datos reales	Evaluación multifactorial en tiempo real (geolocalización IP, metadatos del dispositivo, historial); lista de revisión prioritaria y contacto breve con el cliente
Montos diversificados	Umbrales monetarios fijos	Perfil conductual individual: se evalúa la desviación respecto del patrón habitual del usuario
Picos estacionales	Saturación y fatiga del personal en fechas críticas	Aprendizaje supervisado que usa la estacionalidad histórica; aprobación expedita de operaciones de bajo riesgo
Estafas con envío voluntario (relationship scam)	Las reglas fijas no generan alerta	Aprendizaje no supervisado que calcula desviaciones del comportamiento ordinario
Usuarios nuevos sin historial	No hay referencia de comparación	Perfil construido desde la primera operación; referencia inicial disponible desde la segunda
Abandono por fricción	Revisión manual lenta y bloqueo indiscriminado	Clasificación por niveles de riesgo, revisión prioritaria y mensajes automáticos de contacto

Nota. Elaboración propia a partir del caso analizado y de la literatura revisada.

Dos aspectos merecen explicarse. Primero, cuando el algoritmo detecta algo extraño, la operación no se cancela automáticamente: pasa a una lista de revisión prioritaria, donde un analista puede contactar al cliente con una minientrevista de verificación de no más de cinco minutos; si no hay respuesta, el sistema envía correos o mensajes automáticos para que el

usuario se comunique, con lo cual se protegen los fondos sin generar un rechazo definitivo. Segundo, ante el llamado «arranque en frío», el modelo empieza a construir el perfil desde la primera transacción y, para la segunda, ya cuenta con un marco de referencia. Esa base se enriquece con la interacción humana, porque durante las entrevistas los clientes suelen revelar voluntariamente el propósito de la operación, dato que el analista registra y con el que retroalimenta al algoritmo.

Para dimensionar los costos se revisaron precios públicos de proveedores tecnológicos durante el primer semestre de 2025 (Amazon Web Services, 2026; Bitdefender, 2026; CrowdStrike, 2026; Forter, 2025; Google Cloud, s.f., Microsoft Swcurity, s.f; Mimecast, 2026; Proofpoint, 2026; Riskified, 2026; SEON, s.f.; SentinelOne, s.f.), y se contrastaron con reportes de mercado (McKinsey, 2022; Statista, 2025) y con estudios recientes (Samson-Onuorah, 2023; Hernández Aros et al., 2024). El escenario de referencia es una PYME de aproximadamente 20 colaboradores. La Tabla 3 resume el resultado.

Tabla 3

Costos estimados de protección frente a riesgos de ciberseguridad en PYMES (USD)

Concepto	**Costo estimado**	**Riesgo si no se implementa**	**Impacto en negocios internacionales**
Protección de endpoints (EDR)	15–25 por usuario/mes	Malware, robo de datos	Pérdida de información crítica y de clientes
Seguridad en correo (antiphishing)	5–15 por usuario/mes	Phishing, fraude financiero	Transferencias internacionales fraudulentas
Autenticación multifactor (MFA)	Bajo o gratuito	Suplantación de identidad	Accesos no autorizados a sistemas financieros
Respaldo en la nube	100–300 por mes	Pérdida total de datos	Interrupción de operaciones internacionales
SaaS antifraude (detección con IA)	20–100 por mes, según uso	Fraude no detectado	Pérdidas económicas directas

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

Monitoreo continuo	1,000 por mes	Ataques prolongados sin detección	Daño reputacional y financiero
Costo de un incidente de fraude	No aplica	10,000–100,000 o más por incidente	Pérdida de competitividad global o quiebra

Nota. Elaboración propia con base en OECD (2021), World Bank (Boeddu et al., 2021) y en los precios públicos de los proveedores consultados durante el primer semestre de 2025. Los valores son estimaciones de referencia y varían según el proveedor, el plan y el volumen de uso. El rango del costo de un incidente se toma de World Bank (Boeddu et al., 2021) y de CONDUSEF (2025).

Con esos supuestos, un paquete que incluya todos los conceptos de la tabla, salvo el costo por incidente, sumaría entre 1.520 y 2.200 dólares mensuales para una PYME de 20 colaboradores. Ese rango excluye los costos de integración, capacitación y mantenimiento, e incluye el monitoreo continuo, que es el componente más costoso. Una PYME que se limite a la protección básica —autenticación multifactor, correo electrónico seguro y SaaS antifraude— reduciría considerablemente esa cifra. Aun así, el rango de referencia del costo de un solo incidente supera con creces la inversión estimada. Se trata de una comparación ilustrativa: no constituye un análisis costo-beneficio ni demuestra, por sí misma, la viabilidad económica del modelo completo.

DISCUSIÓN DE RESULTADOS

Los resultados del caso y la literatura revisada apuntan a que la IA ha contribuido a mejorar la detección del fraude financiero, pero al mismo tiempo ha facilitado la aparición de nuevas modalidades de ataque. Esto coincide con la dualidad descrita por Kshetri (2021) y con las advertencias del World Bank (Boeddu et al., 2021). El caso analizado añade un matiz: incluso sin sofisticación tecnológica por parte del atacante, el fraude prospera porque explota la confianza, la urgencia y la vulnerabilidad emocional de las víctimas, y porque un sistema de reglas fijas termina afectando también al cliente legítimo, quien abandona el proceso.

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

Desde el punto de vista teórico, el estudio sugiere que la discusión sobre los modelos antifraude no debería limitarse al desempeño predictivo ni a la explicabilidad. La fricción que sufre el usuario legítimo y la dimensión conductual del fraude son variables de diseño que la literatura consultada aborda poco y que aquí se articulan con el enfoque de la Industria 5.0 (Schmalzried et al., 2023; Tóth et al., 2023). Se propone, además, distinguir entre alerta, derivación, reclamo y fraude confirmado, distinción necesaria para no interpretar como falsos positivos todo lo que no se confirme.

En el plano práctico, para una PYME, el modelo sugiere que la protección puede empezar con soluciones SaaS y medidas de bajo costo, sin necesidad de construir infraestructura propia. Para los proveedores de estos servicios, el caso muestra la utilidad de registrar el motivo por el cual una derivación no se confirma —error, inconformidad, abandono, falsa alarma—, ya que ese dato permitiría medir el falso positivo real y entrenar mejor el modelo. Como paso siguiente, sería conveniente realizar una prueba piloto con datos reales que mida la precisión, la sensibilidad, la tasa de falsos positivos, la tasa de abandono y el tiempo de resolución frente a un sistema de reglas.

En México y Latinoamérica, las PYMES muestran una alta vulnerabilidad frente a fraudes sofisticados impulsados por IA, principalmente debido a limitaciones en la infraestructura tecnológica y al acceso limitado a sistemas avanzados (OECD, 2021). Las plataformas SaaS antifraude ofrecen una alternativa accesible: permiten detectar movimientos inusuales, monitorearlos en tiempo real y generar alertas automáticas sin necesidad de infraestructura propia. Se espera que la adopción de modelos inteligentes y adaptativos contribuya a reducir las pérdidas financieras y a mejorar la eficiencia operativa; esta expectativa, no obstante, deberá contrastarse con evidencia empírica.

Dado que el fraude no se debe únicamente a fallas tecnológicas, los modelos basados en IA deben complementarse con estrategias de concientización, formación digital y gestión integral de riesgos. En materia de capacitación, se recomiendan programas que aborden phishing e ingeniería social, ética y cumplimiento normativo, prevención de fraudes empresariales y detección de deepfakes y de fraude asistido por IA. También conviene desarrollar competencias digitales y financieras para identificar señales de alerta, como la

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

inducción de miedo, la solicitud de códigos de verificación, las ofertas inverosímiles y las comunicaciones con tono de urgencia.

Existen, además, estrategias complementarias de bajo costo, como la capacitación básica en ciberseguridad del personal, la autenticación multifactor y políticas simples de control de accesos. Esta tendencia coincide con lo documentado por PwC (2024) sobre el aumento de la inversión en ciberseguridad de las PYMES a nivel global. Samson-Onuorah (2023) añade que, más allá del costo inicial, estos sistemas requieren ciclos de retroalimentación para actualizarse conforme cambian las tácticas de los defraudadores.

CONCLUSIONES

Respecto de la pregunta de investigación, se concluye que un modelo adaptativo que combine aprendizaje supervisado, aprendizaje no supervisado y revisión humana constituye una alternativa conceptualmente sólida para fortalecer la detección de fraude en las transacciones internacionales de las PYMES latinoamericanas. Esta conclusión se apoya en la coherencia entre el diseño propuesto, la literatura revisada y el caso operativo analizado, y no en una prueba de desempeño.

Lo observado en el estudio es lo siguiente. En un periodo de tres meses y medio, el 40.37 % de las interacciones de una organización de remesas se derivaron al área de fraude y solo el 33.53 % de esas derivaciones se confirmaron como tales, mientras que una parte del resto se perdió por abandono. Las estafas identificadas explotaban más el factor humano que la infraestructura. Lo propuesto, en cambio, es una arquitectura de entrada, procesamiento y salida, con clasificación de riesgo por niveles y revisión prioritaria, diseñada para reducir la fricción con los clientes legítimos. Su efectividad y su viabilidad económica requieren pruebas empíricas posteriores.

La IA es hoy un recurso estratégico para prevenir el fraude en los negocios internacionales, pero su carácter dual constituye un desafío estructural: fortalece la defensa y, a la vez, es aprovechada por actores maliciosos para atacar con mayor sofisticación. Esta situación coloca a las PYMES en una posición de vulnerabilidad particular, aunque el modelo SaaS

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

abre una vía más accesible para acceder a mecanismos de seguridad sin comprometer en exceso sus recursos. En un contexto donde el fraude evoluciona sin pausa, la articulación entre soluciones tecnológicas, la protección de los datos personales y el criterio analítico humano seguirá siendo un desafío central para la sostenibilidad operativa de las PYMES en el comercio internacional.

El modelo no fue desarrollado, entrenado ni probado con datos reales; por ello, sus beneficios son esperados y no demostrados, y el contraste realizado es de coherencia teórica. Los datos del caso provienen de una sola organización, corresponden a un periodo de tres meses y medio y tienen carácter reservado, por lo que no pueden verificarse de forma independiente ni generalizarse. Las observaciones sobre la estacionalidad fuera del cierre de año y sobre los tiempos de reembolso se basan en la experiencia directa de la autora. La revisión documental fue narrativa y no siguió un protocolo sistemático. El trabajo de campo fue acotado: no fue posible entrevistar a un grupo amplio de personas y el conocimiento directo se limita a pocas PYMES que han sufrido fraude.

La autora laboró anteriormente en la organización cuyos datos agregados se analizan en el caso de estudio y actualmente no mantiene ninguna relación con ella. Declara no tener otro conflicto de intereses. Esta investigación no ha recibido financiamiento. Durante la elaboración del artículo se utilizaron herramientas de apoyo a la redacción; las ideas, el análisis y las conclusiones son responsabilidad de la autora.

REFERENCIAS

Amazon Web Services. (2026). *AWS Backup pricing*. https://aws.amazon.com/backup/pricing/

Bitdefender. (2026). *Está explorando las soluciones de seguridad de Bitdefender GravityZone*. https://www.bitdefender.com/es-es/business/campaign/new-compare50

Boeddu, G. L., Chien, J., Grady, R. C. e Istuk, I. (2021). *Consumer Risks in Fintech - New Manifestations of Consumer Risks and Emerging Regulatory Approaches: Policy Research Paper (English)*. World Bank. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/515771621921739154

Černevičienė, J. y Kabašinskas, A. (2024). Explainable artificial intelligence (XAI) in finance: a systematic literature review. *Artificial Intelligence Review*, *57*(216). https://doi.org/10.1007/s10462-024-10854-8

Comisión Nacional para la Protección y Defensa de los Usuarios de Servicios Financieros - CONDUSEF. (2025). *Comercio Electrónico — Estadísticas*. https://www.condusef.gob.mx/index.php/necesitas-ayuda?p=estadisticas

CrowdStrike. (2026). *Tailored bundles to stop breaches*. https://www.crowdstrike.com/pricing/

Forter. (2025). *Contact Sales – Forter*. https://www.forter.com/pricing/

Google Cloud. (s.f.). *Precios de Cloud Storage*. https://cloud.google.com/storage/pricing

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. y Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Hernández Aros, L., Bustamante Molano, L. X., Gutiérrez-Portela, F., Moreno Hernández, J. J. y Rodríguez Barrero, M. S. (2024). Financial fraud detection through the application of machine learning techniques: a literature review. *Humanities and Social Sciences Communications*, *11*(1130). https://doi.org/10.1057/s41599-024-03606-0

Kshetri, N. (2021). The role of artificial intelligence in promoting financial inclusion in developing countries. *Journal of Global Information Technology Management*, *24*(1), 1–6. https://doi.org/10.1080/1097198X.2021.1871273

McKinsey. (2022). *The State of AI in 2022—and a Half Decade in Review*. QuantumBlack. https://www.mckinsey.com/~/media/mckinsey/business%20functions/quantumblack/our%20insights/the%20state%20of%20ai%20in%202022%20and%20a%20half%20decade%20in%20review/the-state-of-ai-in-2022-and-a-half-decade-in-review.pdf?shouldIndex=false

Microsoft Security. (s.f.). *Why Microsoft Security Secure from end to end*. https://www.microsoft.com/en-us/security

Mimecast. (2026). *Mimecast Plans to Mitigate Human Risk*. https://www.mimecast.com/products/mimecast-plans/

Oficina de las Naciones Unidas contra la Droga y el Delito - UNODC. (2022). *Compendio de Ciberdelincuencia Organizada*. Naciones Unidas. https://www.unodc.org/documents/Cybercrime/publications/compendio_de_delincuencia_organizada_es.pdf?v=3.6digest_cyber_organized_crime_2nd_ed.pdf?v=3.6

Organización Mundial del Comercio. (2024). *Informe sobre el comercio mundial 2024 — Comercio e inclusividad: Por un comercio que beneficie a todos*. https://www.wto.org/spanish/res_s/publications_s/wtr24_s.htm

Organización para la Cooperación y el Desarrollo Económicos. (2021). *The digital transformation of SMEs*. OECD Publishing. https://doi.org/10.1787/bdb9256a-en

Passos, L. A., Jodas, D., Costa, K. A. P., Souza Júnior, L. A., Rodrigues, D., Del Ser, J., Camacho, D. y Papa, J. P. (2024). A review of deep learning-based approaches for deepfake content detection. *Expert Systems*, *41*(8), e13570. https://doi.org/10.1111/exsy.13570

Proofpoint. (2026). *Complete Microsoft 365 Security for MSPs*. https://www.proofpoint.com/us/partners/msp

PwC. (2024). *Global Economic Crime Survey 2024. Meeting tomorrow's challenges, embracing risk intelligently*. PricewaterhouseCoopers. https://www.pwc.com/gx/en/services/forensics/economic-crime-survey.html

Riskified. (2026). *AI fraud prevention fueling ecommerce success.* https://www.riskified.com/

Samson-Onuorah, C. I. (2023). Data-driven fraud detection frameworks integrating machine learning, transaction monitoring, and governance across modern banking platforms. *International Journal of Research in Finance and Management*, *6*(2), 318–329. https://doi.org/10.33545/26175754.2023.v6.i2c.670

Schmalzried, D., Hurst, M., Wentzien, M. y Gräser, M. (2023). Analyse der Rolle Künstlicher Intelligenz für eine menschenzentrierte Industrie 5.0. *HMD Praxis der Wirtschaftsinformatik*, *60*, 1143–1155. https://doi.org/10.1365/s40702-023-01001-y

SentinelOne. (s.f.). *PRICING & PACKAGES. Secure More. Spend Less*. https://www.sentinelone.com/pricing/

SEON. (s.f.). *Pricing Plans. Select a Plan That Best Fits Your Needs*. https://seon.io/pricing/

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7

Statista. (2025). *E-commerce fraud* - statistics & facts. https://www.statista.com/topics/9240/e-commerce-fraud/

Tóth, A., Nagy, L., Kennedy, R., Bohuš, B., Abonyi, J. y Ruppert, T. (2023). The human-centric Industry 5.0 collaboration architecture. *MethodsX*, *11*, 102260. https://doi.org/10.1016/j.mex.2023.102260

Cómo citar: Sánchez del Castillo, V. (2026). Propuesta conceptual de un modelo adaptativo basado en inteligencia artificial para la detección de fraude en transacciones internacionales: innovaciones, riesgos emergentes y aplicabilidad en PYMES de Latinoamérica. *Journal of the Academy,* (15), 94-114. https://doi.org/10.47058/joa15.7