23 Julio 2026 Revista de la Carrera de Ingeniería de Sistemas

FONDO EDITORIAL ULIMA

INTERFASES

23 Julio 2026 Revista de la Carrera de Ingeniería de Sistemas
FONDO EDITORIAL ULIMA

INTERFASES

Interfases
Revista de la Carrera de Ingeniería de Sistemas de la Facultad de Ingeniería
de la Universidad de Lima
N.° 23, julio, 2026
https://doi.org/10.26439/interfases2026.n023
Lima, Perú

Fondo Editorial
Av. Javier Prado Este 4600
Urb. Fundo Monterrico Chico
Santiago de Surco, Lima, Perú
Código postal 15023
Teléfono (511) 437-6767, anexo 30131
fondoeditorial@ulima.edu.pe
www.ulima.edu.pe

Edición, diseño y diagramación: Fondo Editorial de la Universidad de Lima.

Correspondencia:
interfases@ulima.edu.pe

Periodicidad: semestral
Arbitraje editorial: revisión por pares doble ciego
Directorios y catálogos: DOAJ, Redalyc, Qualis CAPES, OpenAlex, Latindex, Dialnet y LatinREV

ISSN (en línea) 1993-4912
Hecho el depósito legal en la Biblioteca Nacional del Perú n.º 2020-09967

ÍNDICE

PRESENTACIÓN

https://doi.org/10.26439/interfases2026.n023.9107

El mundo atraviesa importantes cambios en distintos ámbitos, como el ambiental, social, tecnológico y económico, entre otros. Ante este escenario, nuestra sociedad requiere más que nunca del aporte estratégico de la academia para afrontar estos cambios y contribuir, desde la generación y difusión del conocimiento, al desarrollo de la sociedad.

En este contexto, la revista *Interfases*, fiel a su compromiso social, contribuye, desde el ámbito de las publicaciones científicas, al brindar un espacio de calidad a los autores que, como resultado de sus procesos de investigación, desean poner a disposición de la comunidad científica y del público en general sus contribuciones al conocimiento. Esta labor se encuentra alineada con los objetivos de la revista *Interfases* y, especialmente, con los valores de la Universidad de Lima orientados a promover la ciencia, lo que facilita a los autores un espacio de publicación de calidad y comprometido con la ciencia abierta.

En este sentido, la revista *Interfases*, editada por la carrera de Ingeniería de Sistemas de la Universidad de Lima, presenta la edición 23. A continuación, ofrecemos una breve descripción de los contenidos que conforman este número.

En esta edición, se recibieron catorce manuscritos, de los cuales siete fueron seleccionados para su publicación, lo que reafirma el compromiso de la revista con la excelencia académica.

Los siete artículos de la edición 23 de *Interfases* abordan temas relacionados con las tecnologías de la información, la inteligencia artificial, la ciencia de datos, la ciberseguridad, la ingeniería de *software*, la transformación digital y las tecnologías emergentes aplicadas a distintos sectores. Los artículos publicados fueron seleccionados mediante un riguroso proceso de evaluación por pares bajo la modalidad de revisión por doble ciego.

En este número se cuenta con la participación de autores provenientes de Costa Rica, México, Colombia, Ecuador y Perú, afiliados a diversas instituciones académicas, entre ellas la Universidad de Lima (Perú), la Universidad Nacional Mayor de San Marcos (Perú), la Corporación Universitaria Minuto de Dios (Colombia), el Centro Universitario Valle de Chalco de la Universidad Autónoma del Estado de México, la Universidad Autónoma del Estado de México, la Universidad Técnica de Machala (Ecuador) y la Universidad

Internacional de las Américas (Costa Rica). Cabe destacar que uno de los artículos fue presentado en inglés.

Agradecemos a todos los autores que enviaron sus manuscritos como contribución a la revista y por elegir *Interfases* como espacio para compartir los resultados de sus investigaciones. Gracias a estas propuestas y al trabajo articulado del Comité Editorial y de nuestro equipo de revisores científicos, fue posible realizar una cuidadosa selección de los trabajos que conforman esta edición, que se sustentó en un proceso de evaluación riguroso y objetivo.

Cabe resaltar que los revisores pertenecen a diversas instituciones académicas nacionales e internacionales, entre ellas la Universidad Autónoma de Baja California (México), la Universidad de Lima (Perú), la Universidad La Salle Victoria (México), el Instituto de Ciencias Agrarias (España), la Universidad Nacional Abierta y a Distancia (Colombia), la Universidad Nacional de San Antonio Abad del Cusco (Perú), la Fundación Universitaria de Popayán (Colombia), la Universidad Técnica Particular de Loja (Ecuador), la Pontificia Universidad Católica del Perú, la Universidad Autónoma de Sinaloa (México), la Universidad Militar Nueva Granada (Colombia), la Universidad Autónoma del Estado de México, la Universidad Nacional de Santiago del Estero (Argentina), la Universidad Veracruzana (México), la Federación de las Industrias del Estado de Bahía (Brasil), University of St. Thomas (Estados Unidos) y la Universidad Nacional del Altiplano de Puno (Perú).

Finalmente, expresamos nuestro más sincero agradecimiento al Fondo Editorial y al Área de Colecciones de la Biblioteca de la Universidad de Lima por su valiosa colaboración en los procesos de revisión de estilo, diagramación y publicación de la edición 23 de *Interfases*.

Dra. Nadia Katherine Rodríguez Rodríguez

Directora de *Interfases*

ARTÍCULOS DE INVESTIGACIÓN

ANÁLISIS COMPARATIVO DE LA APLICACIÓN DE LOS ALGORITMOS DE APRENDIZAJE AUTOMÁTICO *RANDOM FOREST* Y REGRESIÓN LOGÍSTICA EN LA CLASIFICACIÓN DE CONJUNTOS DE TEXTOS

Stalin Francisco Calva Vicente
https://orcid.org/0009-0009-1163-0271
scalva3@utmachala.edu.ec
Universidad Técnica de Machala, Ecuador

Angel Ricardo Vera Santos
https://orcid.org/0000-0002-2300-6005
avera9@utmachala.edu.ec
Universidad Técnica de Machala, Ecuador

Freddy Anibal Jumbo Castillo
https://orcid.org/0000-0002-5200-7162
fjumbo@utmachala.edu.ec
Universidad Técnica de Machala, Ecuador

Johnny Paul Novillo Vicuña
https://orcid.org/0000-0002-4915-3441
jnovillo@utmachala.edu.ec
Universidad Técnica de Machala, Ecuador

Recibido: 26 de marzo del 2026 / Aceptado: 10 de junio del 2026
doi: https://doi.org/10.26439/interfases2026.n023.8710

RESUMEN. Gestionar la inmensa cantidad de datos no estructurados que generamos hoy sería impensable sin la clasificación automática de textos. Esta herramienta se ha vuelto indispensable para tareas tan variadas como el análisis de sentimientos o la detección de noticias falsas. Sin embargo, surge un dilema técnico importante: a pesar de la enorme oferta de algoritmos diseñados para estos fines, todavía no hay un consenso claro sobre cuál es el más efectivo para cada conjunto de datos específico. Para aliviar este problema, la presente investigación evalúa la robustez de dos modelos. Estos modelos son *random forest*, que funciona con conjuntos de árboles, y la regresión logística, un modelo lineal. El corpus empleado es Symptom2Disease, un conjunto de 1200 registros textuales con descripciones de síntomas en lenguaje natural distribuidos equitativamente entre 24 categorías diagnósticas que abarca patologías como psoriasis, diabetes, dengue y malaria.

Para este propósito, se ha adoptado un diseño experimental factorial con preprocesamiento estándar, la adición de aumento de datos para reducir la limitación de muestras y análisis de representación a través de N-gramas. Los resultados muestran que la regresión logística funciona mejor que *random forest*, pues alcanza un máximo de F1-Score de 0,9797 cuando se utiliza junto con el aumento de datos y los unigramas. Para textos médicos cortos, la linealidad del modelo y el enriquecimiento sintético del corpus constituyen una solución más eficiente y precisa en comparación con los conjuntos no lineales.

PALABRAS CLAVE: aprendizaje automático / procesamiento del lenguaje natural / clasificación de textos clínicos / regresión logística / *random forest*

COMPARATIVE ANALYSIS OF THE APPLICATION OF THE RANDOM FOREST AND LOGISTIC REGRESSION MACHINE LEARNING ALGORITHMS FOR TEXT CORPUS CLASSIFICATION

ABSTRACT. Managing the vast amount of unstructured data we generate today would be unthinkable without automatic text classification. This tool has become indispensable for tasks as varied as sentiment analysis and combating fake news. However, a significant technical dilemma arises: despite the vast array of algorithms designed for these purposes, there is still no clear consensus on which is the most effective for each specific dataset. To address this issue, the present study compares the robustness of two models: Random Forest, which operates using a set of trees, and Logistic Regression, a linear model. The corpus used is Symptom2Disease, a dataset of 1,200 text records containing descriptions of symptoms in natural language, evenly distributed across 24 diagnostic categories, covering conditions such as psoriasis, diabetes, dengue, and malaria. For this purpose, a factorial experimental design was adopted, incorporating standard preprocessing, data augmentation to mitigate sample size limitations, and representation analysis via N-grams. In other words, it appears that Logistic Regression performs better than Random Forest, achieving a maximum F1 score of 0.9797 when used in conjunction with data augmentation and unigrams. For short medical texts, the linearity of the model and synthetic corpus enrichment prove to be a more efficient and accurate solution compared to non-linear models.

KEYWORDS: machine learning / natural language processing / clinical text classification / logistic regression / random forest

INTRODUCCIÓN

La digitalización ha disparado la creación de textos, lo que ha alterado profundamente los mecanismos de gestión de información. Hoy, la prioridad no es solo generar datos, sino entender cómo esa marea de contenido impacta en la toma de decisiones y en el uso de la información disponible (Kowsari et al., 2019). Hablamos de un ecosistema masivo de información no estructurada que abarca desde artículos de prensa y redes sociales hasta registros médicos o legales cuyo valor real es inmenso. Sin embargo, ese potencial solo se puede aprovechar si implementamos procesos de análisis y tratamiento de datos que sean realmente consistentes y rigurosos. Dado el punto anterior, la clasificación automática de textos, un tema central del procesamiento de lenguaje natural (NLP), representa una de sus tareas básicas para dar sentido a grandes cantidades de datos (Aggarwal & Zhai, 2012). Su aplicabilidad es transversal a múltiples dominios, lo que facilita desde la detección de noticias falsas (Báez et al., 2022) hasta el análisis de sentimientos (positivo, negativo, neutral) de un texto (Calero Sánchez et al., 2024). Ello evidencia el enorme impacto de la inteligencia artificial en el aprendizaje automático aplicado al procesamiento de texto.

A pesar de los avances logrados y del amplio abanico de algoritmos disponibles, persiste un problema central: la complejidad de elegir el modelo que mejor se adapte a un objetivo o *dataset* específico (Li et al., 2022). Desde modelos lineales tradicionales hasta conjuntos de algoritmos complejos, la literatura existente detalla diferencias significativas en el rendimiento en relación con el tamaño de los datos, la dimensión del corpus y el tipo de lenguaje (Cardoso et al., 2023). Tal variabilidad resalta la importancia de enfoques comparativos sólidos que guíen a los investigadores y profesionales en la selección de la mejor herramienta de utilidad para sus respectivos dominios.

En el ámbito clínico, la clasificación automática de textos médicos enfrenta desafíos adicionales en comparación con otros dominios de texto. La alta especificidad del vocabulario médico, la presencia de términos con significados variables según el contexto sintomático y la frecuente escasez de corpus etiquetados de gran escala condicionan tanto la elección del algoritmo como la estrategia de representación vectorial adoptada (Báez et al., 2022). A esto se suma que, en entornos de apoyo al diagnóstico, la interpretabilidad del modelo cobra un valor estratégico: los profesionales de la salud no solo requieren predicciones precisas, sino también comprensión de los factores que motivaron una clasificación determinada. Esta exigencia convierte la elección del algoritmo en una decisión que equilibra simultáneamente exactitud, eficiencia computacional e interpretabilidad clínica, tres dimensiones en las que los modelos clásicos de aprendizaje automático se distinguen de manera diferenciada.

La elección de *random forest* y regresión logística como objetos de comparación responde a una estrategia metodológica deliberada: ambos representan paradigmas opuestos en la taxonomía del aprendizaje supervisado. Uno es lineal y probabilístico,

el otro es no lineal y se basa en conjuntos de árboles. Ambos figuran entre los clasificadores de referencia más ampliamente utilizados en la literatura de clasificación de textos médicos (Pranckevičius & Marcinkevičius, 2017). Al contrastar específicamente estos dos enfoques, la investigación busca aportar evidencia empírica sobre cuál paradigma resulta más adecuado para el dominio de síntomas clínicos en lenguaje natural, una pregunta que, a pesar de su relevancia práctica, carece de respuesta sistemática en el contexto de corpus médicos de tamaño pequeño y equilibrado.

Para mitigar este problema, esta investigación realiza un análisis comparativo de los rendimientos de dos algoritmos de clasificación bien aceptados, basados en diferentes fundamentos: *random forest* y regresión logística. *Random forest* es un algoritmo de conjunto bien conocido que construye múltiples árboles de decisión y es altamente capaz de acomodar las necesidades de datos de alta dimensión y evita el sobreajuste en grandes conjuntos de datos (Pal, 2005). Sin embargo, la regresión logística es un modelo lineal probabilístico que es favorecido debido a su simplicidad, eficiencia computacional e interpretabilidad, y descrito como un modelo de referencia confiable en muchos de los problemas de clasificación de textos (Genkin et al., 2007).

La investigación adopta un método experimental basado en pruebas controladas y exhaustivas. Más allá del preprocesamiento tradicional (tokenización, lematización), se evalúan dos estrategias de optimización complementarias: el aumento del léxico del corpus y la implementación de N-gramas para la vectorización por frecuencia de término-frecuencia inversa de documento (TF-IDF), un método estadístico que pondera la relevancia de cada palabra según su capacidad discriminativa en el texto, con la intención de capturar el contexto sintáctico de los síntomas. Para ello, se crearon múltiples escenarios de prueba para separar el efecto de estas variables en el rendimiento de *random forest* y de regresión logística. Luego, cada modelo fue entrenado y se compararon ambos estadísticamente sobre la base de los siguientes indicadores establecidos en la industria y la academia: precisión y *recall*, así como matriz de confusión y F1 *score*. Con ello, se pudo proporcionar una evaluación objetiva y multidimensional sobre si ambos modelos tuvieron éxito en la tarea de clasificación.

En cuanto a trabajos relacionados, investigaciones recientes han explorado la clasificación de textos médicos y la predicción de enfermedades a partir de descripciones sintomáticas en lenguaje natural comparando modelos clásicos y avanzados. Hassan et al. (2024) aplicaron modelos de lenguaje y aprendizaje profundo sobre el mismo corpus Symptom2Disease utilizado en esta investigación, y lograron una exactitud del 99,58 % con el modelo de normalización de conceptos médicos-representaciones de codificador bidireccional de transformadores (MCN-BERT); no obstante, dicho desempeño se obtiene a expensas de una complejidad computacional significativamente mayor frente a los enfoques clásicos aquí propuestos. Ello refuerza la utilidad de los modelos más simples e interpretables en escenarios de recursos limitados.

Por su parte, Al-qarni & Algarni (2025) compararon técnicas de aprendizaje profundo (LSTM, CNN-LSTM, GRU) frente a métodos clásicos basados en TF-IDF para la predicción de enfermedades a partir de texto libre. Concluyeron que los clasificadores lineales tradicionales mantienen ventajas en interpretabilidad y eficiencia computacional cuando el volumen de datos es reducido. En la misma línea, Vyas et al. (2025) evaluaron el sistema SympTextML para la predicción multiclase de enfermedades mediante descripciones de síntomas en lenguaje natural empleando representaciones TF-IDF sobre 24 categorías diagnósticas, y hallaron que los modelos clásicos bien configurados pueden competir con arquitecturas más complejas en corpus de tamaño reducido. Estas investigaciones subrayan la pertinencia de comparar los algoritmos de regresión logística y *random forest* en el dominio clínico, especialmente en tareas en las que la interpretabilidad y la eficiencia constituyen criterios de selección tan determinantes como la exactitud de la clasificación.

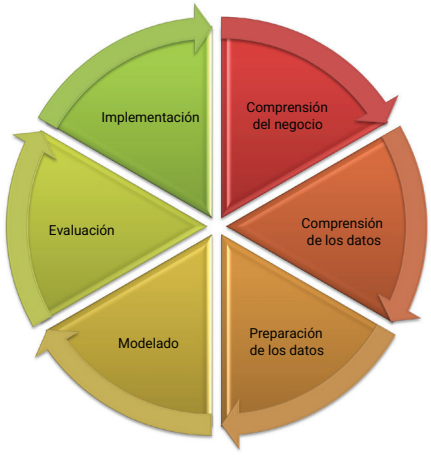
METODOLOGÍA

El estudio actual está diseñado bajo un marco cuantitativo, comparativo y experimental que evalúa el rendimiento de los algoritmos de aprendizaje automático *random forest* y regresión logística en la clasificación de textos en el dominio de síntomas y enfermedades.

En el desarrollo metodológico, adoptamos el modelo CRISP-DM (proceso estándar interindustrial para la minería de datos), el cual describe el marco de seis etapas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implementación.

Figura 1

Fases del modelo de referencia CRISP-DM



Nota. Los datos proceden de *CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc*, por P. Chapman, J. Clinton, R. Kerber, T. Reinartz, C. Shearer y R. Wirth, 2000, DaimlerChrysler (https://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/CRISP-DM.pdf).

Comprensión del negocio

El análisis automatizado de la información médica es un medio importante para mejorar los procesos de diagnóstico y la gestión del conocimiento clínico en la era digital. Con eso en mente, el estudio tiene como objetivo verificar si los algoritmos clasificadores *random forest* o regresión logística funcionan efectivamente en un texto médico que detalla síntomas y enfermedades. Además, busca proporcionar evidencia empírica que optimizará el análisis del sistema y la previsión en futuras investigaciones de salud digital.

En un sentido más amplio, el contexto de aplicación comprende sistemas de apoyo a la decisión clínica (*clinical decision support systems*, CDSS) que procesen narrativas escritas por pacientes en plataformas de telemedicina, portales de salud electrónica y sistemas de triaje remoto. La capacidad de clasificar automáticamente estas descripciones textuales en una categoría diagnóstica permitirá orientar al paciente hacia el especialista correspondiente, reducir tiempos de espera y optimizar la asignación de recursos en atención primaria. La validación experimental de modelos eficientes e interpretables para este fin constituye, por tanto, un aporte concreto a la infraestructura técnica de la salud digital (Calero Sánchez et al., 2024).

Comprensión de los datos

Para el desarrollo experimental, se utilizó el conjunto de datos Symptom2Disease, obtenido del repositorio público de ciencia de datos Kaggle y publicado originalmente por Barman et al. (2023). Este corpus fue organizado utilizando un formato estructurado, elegido por ser adecuado para la tarea de procesamiento de lenguaje natural involucrado en el dominio médico. Se compone de 1200 registros estructurados en dos columnas esenciales que caracterizan las variables de entrada y salida del modelo:

- label (variable dependiente): tipo de etiqueta categórica que representa el diagnóstico patológico. El conjunto cubre un total de 24 enfermedades como psoriasis, diabetes, dengue y COVID-19, etcétera.
- text (variable independiente): la descripción en lenguaje natural de los síntomas reportados por el paciente asociados con la enfermedad.

A modo de ilustración, la Tabla 1 presenta dos registros representativos del corpus que muestran la estructura de ambas variables tal y como fueron procesadas durante el experimento:

Tabla 1

Ejemplos representativos del corpus Symptom2Disease

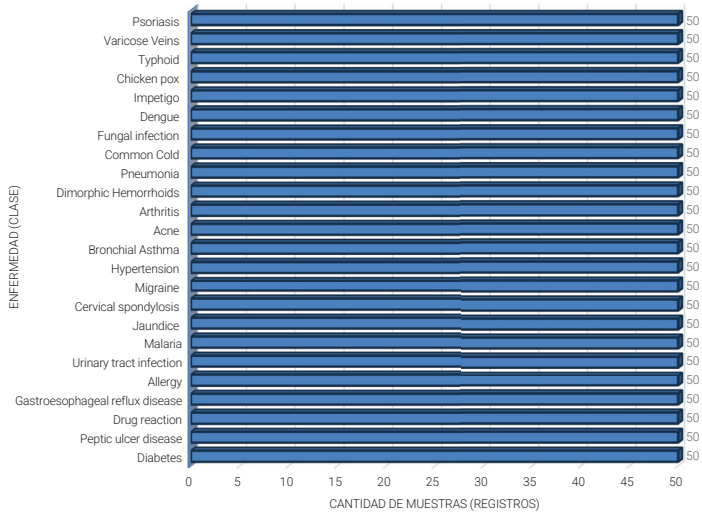
ID	label (diagnóstico)	text (descripción de síntomas)
0	Psoriasis	I have been experiencing a skin rash on my arms, legs, and torso for the past few weeks. It is red, itchy, and covered in dry, scaly patches.
600	Diabetes	I've been feeling really thirsty lately and urinating more than usual. I've also been losing weight unexpectedly. My vision has become blurry and I feel tired all the time.

Nota. Cada registro consta de una etiqueta de diagnóstico y una descripción de síntomas en inglés, en lenguaje coloquial. Los ejemplos proceden de *Symptom2Disease: Diseases and Natural Language Symptom Descriptions*, por N. Barman, F. Karim y K. Sharma, 2023.

Una clave para dar sentido a los datos es probar la distribución de las clases, ya que un desequilibrio puede sesgar el modelo hacia la mayoría de las clases. Como se demuestra en la Figura 2, el análisis exploratorio de datos verifica que los estos están equilibrados con 50 instancias de cada una de las 24 enfermedades con números correspondientes. Esta distribución uniforme también permite que los modelos de aprendizaje automático (*random forest* y regresión logística) aprendan los patrones típicos de cada patología de manera similar, por lo que no deberían requerir el uso inicial de métodos de submuestreo o sobremuestreo.

Figura 2

Distribución de muestras por enfermedad en el conjunto de datos

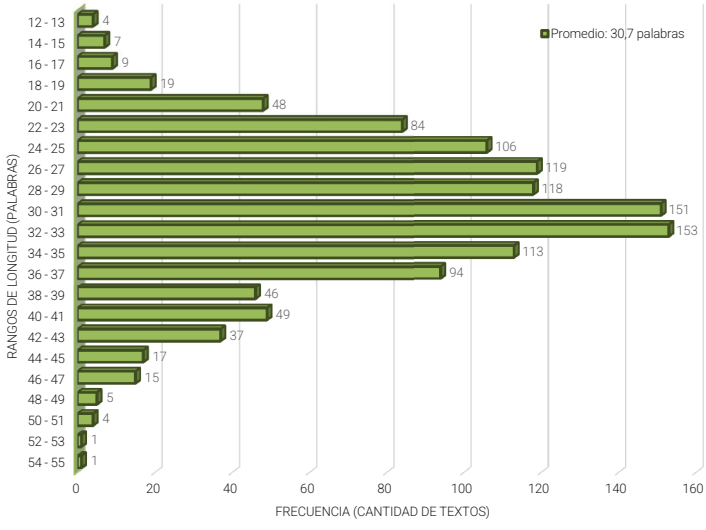


Nota. Se observa una distribución equitativa con 50 registros por clase, lo que asegura un entrenamiento balanceado. Los datos proceden de *Symptom2Disease: Diseases and Natural Language Symptom Descriptions*, por N. Barman, F. Karim y K. Sharma, 2023

Los textos también fueron analizados para determinar la longitud de las descripciones de los síntomas. Del histograma en la Figura 3, la longitud de los textos es aproximadamente igual a la normal (el promedio es de 30,7 palabras). Esto sugiere que los modelos funcionarían con textos cortos y directos, y la presencia de términos importantes particulares tendría un alto peso discriminativo, por lo que se justifica la exploración en representación basada en unigramas.

Figura 3

Distribución de la longitud de los textos (conteo de palabras)



Nota. Distribución del conteo de palabras por registro, con un promedio de 30,7 palabras por descripción. El volumen principal de los textos clínicos se concentra entre las 20 y 40 palabras. Los datos proceden de *Symptom2Disease: Diseases and Natural Language Symptom Descriptions*, por N. Barman, F. Karim y K. Sharma, 2023.

Finalmente, para caracterizar el dominio semántico del corpus, se generó una nube de palabras (WordCloud) que visualiza los términos más frecuentes. La Figura 4 destaca palabras como *skin* (piel), *pain* (dolor), *fever* (fiebre), *cough* (tos) y *discomfort* (malestar). La predominancia de estos sustantivos confirma que el vocabulario es altamente específico del dominio clínico sintomatológico, lo cual es ideal para el entrenamiento de clasificadores supervisados.

Figura 4

Nube de palabras de los términos médicos más frecuentes

Preparación de los datos

El corpus textual se preprocesó para lograr datos de alta calidad y uniformes antes del modelado. Las principales tareas trabajadas se hicieron de acuerdo con los principios de la minería de texto estándar presentados por Manning et al. (2008):

- Limpieza de datos: eliminar duplicados, caracteres nulos e irrelevantes.
- Unificación textual: todo el texto se ha reducido a minúsculas y se han eliminado los signos de puntuación.
- Tokenización: segmentación de los síntomas en palabras clave.
- Eliminación de stopwords: eliminar palabras sin semántica.
- Lematización: la normalización lingüística de los términos a su forma inicial.
- Aumento de datos (data augmentation): también utilizamos el método de intercambio aleatorio de palabras (random swap) con los datos de entrenamiento. Este enfoque está respaldado por Wei & Zou (2019) en su marco easy data augmentation (EDA), en el cual los textos originales se generan y utilizan como variantes sintéticas para hacer el modelo más sólido y manejar el posible sobreajuste que surge de su falta de datos por clase.
- Preservación del significado clínico: siguiendo la advertencia metodológica de Wei & Zou (2019), quienes señalan que el intercambio excesivo de palabras puede comprometer la validez de las etiquetas originales en textos breves, en el presente estudio se constató que el random swap preserva razonablemente

el significado clínico de cada registro. La Tabla 2 muestra ejemplos representativos del proceso de aumento, en el que se ilustra cómo el intercambio de posición de dos palabras mantiene el cuadro sintomático identificable.

Tabla 2

Ejemplos de textos originales y sus versiones aumentadas mediante random swap

Tipo	label	text
Original	Psoriasis	I have been experiencing a *skin rash* on my arms, legs, and torso. It is red, itchy, and covered in dry, scaly patches.
Aumentado (*random swap*)	Psoriasis	I have been experiencing a *rash skin* on my arms, legs, and torso. It is red, itchy, and covered in scaly, dry patches.

Nota. El intercambio de posición de dos palabras (señaladas en cursiva) no altera el diagnóstico subyacente ni elimina los términos clínicos clave. El corpus aumentado conserva la misma etiqueta original.

- Vectorización y representación: transformación de textos utilizando el esquema de frecuencia de término-inversa frecuencia de documento (TF-IDF). Este enfoque, que ha sido bien fundamentado en la literatura de clasificación de textos (Alsaeedi, 2020), pondera la relevancia de cada síntoma al reducir el peso de los términos no representativos que no expresan poderes discriminativos y, por lo tanto, para modelos lineales y de conjunto, proporciona el ajuste del modelo de entrada.

El diseño del *pipeline* de preprocesamiento respondió a las particularidades lingüísticas del corpus. La lematización desempeña un papel crítico en textos clínicos, pues permite unificar variantes morfológicas de términos diagnósticos: palabras como *itching*, *itches* e *itchy* se reducen al lema "itch", lo que disminuye la dispersión del espacio vectorial sin pérdida semántica. La eliminación de *stopwords* actúa como filtro de ruido léxico, priorizando los sustantivos y adjetivos que constituyen el núcleo de cada descripción sintomática (Manning et al., 2008).

En cuanto a la vectorización TF-IDF, su ventaja sobre una representación simple de frecuencias radica en ponderar los términos según su relevancia discriminativa en el corpus: un síntoma infrecuente, pero diagnósticamente determinante como *petechiae*, asociado al dengue hemorrágico, recibe un peso muy superior al de términos genéricos como *feeling* que aparecen en múltiples categorías. Esta propiedad es especialmente valiosa en corpus médicos, en los que el vocabulario específico de cada patología constituye el principal marcador de identidad diagnóstica (Alsaeedi, 2020).

Modelado

En esta sección, el proceso de modelado se ha realizado con la aplicación de dos algoritmos típicos de aprendizaje supervisado: regresión logística y *random forest*. La elección de estos modelos es una cuestión metodológica basada en su diferente enfoque, complejidad y literatura científica respecto a la clasificación de textos.

La regresión logística fue seleccionada como un modelo lineal probabilístico prevalente en clasificaciones binarias y multiclase en el que se puede interpretar el efecto de las variables independientes en la predicción final. Su bajo costo computacional, facilidad de implementación y adecuación en datos de texto vectorizados por TF-IDF lo convierten en un punto de referencia confiable para comparar modelos más sofisticados, según reportan Pranckevičius y Marcinkevičius (2017). Los autores han demostrado que el modelo produce con alta precisión en corpus de reseñas y textos cortos, incluso ha superado a clasificadores más grandes en estabilidad.

Desde una perspectiva matemática, en su extensión multiclase, la regresión logística opera bajo el esquema multinomial, en el que, para cada clase k, se estima la probabilidad condicional $P(y = k|x)$ mediante la función *softmax* aplicada a la combinación lineal de los vectores TF-IDF de entrada. El parámetro de regularización C controla la penalización sobre los coeficientes del modelo: valores bajos implican regularización fuerte, mientras que valores altos permiten mayor ajuste a los datos de entrenamiento. Con $C = 1$, la regularización L2 equilibra la capacidad de ajuste con la generalización, lo que previene el sobreajuste en espacios de alta dimensionalidad como los vectores TF-IDF, cuya cardinalidad puede superar los diez mil atributos activos (Genkin et al., 2007). Esta propiedad resulta determinante en un corpus de texto médico breve, en el que la mayoría de los términos del vocabulario son infrecuentes y el riesgo de colinealidad entre características es elevado.

Por el contrario, el *random forest* fue elegido como modelo de tipo conjunto que se construye alrededor de árboles de decisión, ya que es un método ideal para alta dimensionalidad, interacción no lineal y evasión de sobreajuste. Según Sun et al. (2020), la integración de múltiples árboles de decisión en un ensamble proporciona mayor estabilidad predictiva y mejora la generalización frente a datos complejos y no estructurados, lo cual es deseable para reducir la dispersión presente en el procesamiento de texto no estructurado.

Para proporcionar una evaluación sistemática del efecto de los métodos probados, se ha establecido un diseño factorial 2 x 2 x 2 que conduce a 8 experimentos controlados. Estas situaciones permiten determinar el alcance del efecto de los algoritmos, el enfoque de aumento de datos y la representación de N-gramas (Tabla 3).

Tabla 3

Diseño experimental

N.°	Modelo	Técnica de datos	Representación (TF-IDF)	Hipótesis/objetivo del experimento
1	Regresión logística	Originales	Unigramas (1-gram)	Línea base (*baseline*): rendimiento estándar sin mejoras
2	Regresión logística	Originales	N-gramas (1,2)	Evaluar si capturar frases como *dolor de pecho* mejora la detección con datos limitados
3	Regresión logística	*Data augmentation*	Unigramas (1-gram)	Evaluar si la variabilidad léxica (sinónimos) compensa la falta de contexto
4	Regresión logística	*Data augmentation*	N-gramas (1,2)	Modelo óptimo lineal: combinación de riqueza de datos y contexto semántico
5	*Random forest*	Originales	Unigramas (1-gram)	Línea base (ensamble): evaluar capacidad base del modelo no lineal
6	*Random forest*	Originales	N-gramas (1,2)	Verificar si aumentar la dimensionalidad (más *features* por los bigramas) afecta al RF
7	*Random forest*	*Data augmentation*	Unigramas (1-gram)	Evaluar la reducción de sobreajuste (*overfitting*) mediante más datos
8	*Random forest*	*Data augmentation*	N-gramas (1,2)	Modelo óptimo ensamble: prueba de estrés con máxima complejidad de datos y *features*

Nota. Resumen de los escenarios para el análisis comparativo, variando el algoritmo, la técnica de aumento de datos y la tokenización.

No se consideraron otros algoritmos, incluidos los SVM y los modelos de aprendizaje profundo, ya que el objetivo no es analizar sistemáticamente todos los métodos existentes, sino comparar dos enfoques representativos y contrastantes: la regresión logística (lineal y explicativa) y un modelo de conjunto no lineal (*random forest*). De ese modo, esta metodología permite obtener hallazgos interpretables y equilibrados sobre cómo la naturaleza del modelo afecta la clasificación de textos médicos.

Con el propósito de garantizar la reproducibilidad de los experimentos, la Tabla 4 detalla la configuración de los hiperparámetros empleados en cada componente del *pipeline*: los dos clasificadores y el vectorizador TF-IDF. Los valores indicados corresponden a los utilizados en la implementación experimental, ya sea que hayan sido especificados explícitamente en el código o correspondan al valor predeterminado de la biblioteca scikit-learn.

Tabla 4

Configuración de hiperparámetros de los modelos y el vectorizador TF-IDF

Modelo	Hiperparámetro	Valor utilizado
Regresión logística	solver	lbfgs
Regresión logística	C (regularización L2)	1,0 (predeterminado)
Regresión logística	max_iter	1000
Regresión logística	multi_class	auto (predeterminado)
Regresión logística	random_state	42
Random forest	n_estimators	100
Random forest	max_depth	none (sin límite)
Random forest	min_samples_split	2 (predeterminado)
Random forest	criterion	gini (predeterminado)
Random forest	random_state	42
TF-IDF Vectorizer	ngram_range	(1,1) o (1,2) según experimento
TF-IDF Vectorizer	max_features	none (sin límite)
TF-IDF Vectorizer	min_df	1 (predeterminado)

Nota. Los valores marcados como *(predeterminado)* corresponden a los valores por defecto de scikit-learn y fueron mantenidos sin modificación. La semilla aleatoria random_state = 42 se fijó en todos los componentes para garantizar la reproducibilidad de los experimentos. El hiperparámetro ngram_range varía según el escenario experimental definido en la Tabla 3.

Los dos modelos se entrenaron con las mismas configuraciones experimentales. Además, para evitar sesgos en la estimación del error de generalización, se llevó a cabo una validación cruzada estratificada (Stratified K-Fold con $k = 5$). Esta estrategia, alineada con las recomendaciones de Vabalas et al. (2019) para la validación de algoritmos con tamaños de muestra pequeños, mantiene la misma proporción de las clases en cada partición, lo que minimiza efectivamente el sesgo de selección y proporciona criterios de evaluación sólidos. Es fundamental precisar que el proceso de *data augmentation* fue aplicado exclusivamente sobre el conjunto de entrenamiento dentro de cada partición (*fold*), es decir, los textos sintéticos fueron generados a partir de los datos de entrenamiento de ese *fold* después de realizar la división, y el vectorizador TF-IDF fue ajustado únicamente sobre dichos datos aumentados. El conjunto de validación de cada *fold* permaneció intacto y no fue expuesto durante el proceso de aumento ni durante el ajuste del vectorizador, lo que garantizó que no se produzca filtración de datos (*data leakage*) que pudiera inflar artificialmente las métricas de desempeño reportadas.

Evaluación

Siguiendo a Powers (2008), el rendimiento de cada modelo se evalúa considerando métricas estándar en el aprendizaje supervisado para la evaluación de clasificadores en escenarios equilibrados:

- Precisión: proporción de predicciones correctas entre los casos positivos estimados
- Exhaustividad (recall): proporción de casos positivos correctamente identificados
- Puntuación F1: media armónica entre precisión y recall para balances de rendimiento
- Matriz de confusión: análisis de aciertos y errores por clase

Estas métricas permiten determinar qué algoritmo realiza un mejor trabajo en lograr una clasificación bien equilibrada de textos médicos.

Implementación

El experimento utilizó el entorno de programación Python con pandas, numpy, scikit-learn y nltk como las bibliotecas personalizadas para el procesamiento del lenguaje natural y modelado de los datos.

Los resultados establecen las posibles conclusiones sobre la aplicabilidad de cada modelo en el campo del análisis automatizado de información médica, lo que proporciona información útil para futuras investigaciones en clasificación de textos.

RESULTADOS

Para evaluar la eficiencia de los modelos propuestos, se llevaron a cabo los ocho escenarios experimentales especificados en la metodología utilizando validación cruzada estratificada (Stratified K-Fold) de $k = 5$. A continuación, el rendimiento de los algoritmos se presenta, brevemente, en términos de diferentes configuraciones de procesamiento de datos y representación vectorial.

Desempeño del modelo de regresión logística

En primera instancia, se evaluó la capacidad del modelo lineal utilizando el conjunto de datos original sin técnicas de generación sintética. Los resultados se detallan en la Tabla 5.

Tabla 5

Desempeño del modelo de regresión logística con conjuntos de datos originales

N.°	Modelo	Técnica de datos	Representación (TF-IDF)	Precisión	*Recall*	F1-Score
1	Regresión logística	Originales	Unigramas (1-gram)	0,9728	0,9700	0,9695
2	Regresión logística	Originales	N-gramas (1,2)	0,9705	0,9667	0,9663
		Promedio		0,9716	0,9684	0,9679

Nota. Evaluación de la línea base (*baseline*) utilizando representaciones de unigramas y N-gramas sin técnicas de aumento de datos.

Como se observa en la Tabla 5, la regresión logística establece una línea base muy alta incluso sin aumento de datos: alcanza un F1-Score promedio de 0,9679. Al comparar las representaciones, el modelo basado en unigramas (Exp. 1) obtuvo un rendimiento ligeramente superior (0,9695) frente al de bigramas (0,9663), lo que sugiere que, con los datos originales, la inclusión de pares de palabras no aportó una ganancia significativa, incluso introdujo un leve ruido en la clasificación.

Posteriormente, se evaluó el impacto de la técnica de *data augmentation* en el modelo lineal. Los resultados de esta segunda fase se presentan en la Tabla 6.

Tabla 6

Impacto del data augmentation *en el modelo de regresión logística*

N.°	Modelo	Técnica de datos	Representación (TF-IDF)	Precisión	*Recall*	F1-*Score*
3	Regresión logística	*Data augmentation*	Unigramas (1-gram)	0,9823	0,9800	0,9797
4	Regresión logística	*Data augmentation*	N-gramas (1,2)	0,9765	0,9733	0,9731
		Promedio		0,9794	0,9766	0,9764

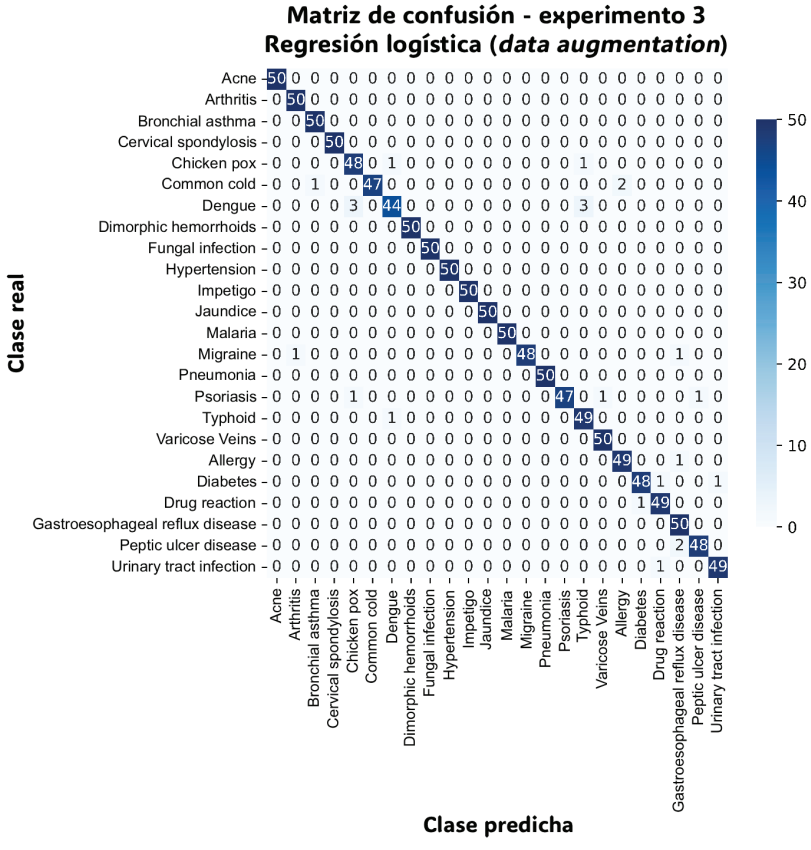
Nota. Comparación de métricas tras duplicar el corpus de entrenamiento mediante la técnica de intercambio aleatorio de palabras (*random swap*).

La Tabla 6 evidencia que la incorporación de datos sintéticos mejoró el rendimiento global del modelo: se elevó el F1-Score promedio a 0,9764. Es destacable que el experimento 3 (*data augmentation* con unigramas) alcanzó el desempeño máximo de todo el estudio con una precisión del 98,23 % y un F1-Score de 0,9797. Esto confirma que enriquecer la variabilidad léxica mediante el intercambio de palabras potencia la capacidad discriminativa del modelo lineal más que el aumento de la complejidad semántica (N-gramas).

Para ilustrar la precisión de este escenario óptimo, se presenta la matriz de confusión del experimento 3 (Figura 5).

Figura 5

Matriz de confusión del mejor modelo de regresión logística (Exp. 3)



Nota. Análisis de errores para la configuración óptima con *data augmentation* y unigramas, muestra la distribución de predicciones correctas (diagonal) y falsos positivos.

Desempeño del modelo *random forest*

A continuación, se presenta el análisis del comportamiento del algoritmo de ensamble (*random forest*). Primero, se evaluó su desempeño con los datos originales, tal como se muestra en la Tabla 7.

Tabla 7

Desempeño del modelo random forest *con conjuntos de datos originales*

N.°	Modelo	Técnica de datos	Representación (TF-IDF)	Precisión	*Recall*	F1-Score
5	*Random forest*	Originales	Unigramas (1-gram)	0,9594	0,9567	0,9555
6	*Random forest*	Originales	N-gramas (1,2)	0,9528	0,9475	0,9467
		Promedio		0,9561	0,9521	0,9511

Nota. Evaluación del algoritmo de ensamble bajo condiciones estándar contrastando el uso de unigramas frente a bigramas.

Los resultados de la Tabla 7 muestran que el *random forest*, en su estado base, tiene un rendimiento sólido, pero inferior a la regresión logística, con un F1-Score promedio de 0,9511. En cuanto a la representación, se repite el patrón observado anteriormente: los Unigramas (Exp. 5) superan a los Bigramas (Exp. 6). La caída en el rendimiento al usar N-gramas (0,9467) sugiere que el aumento de dimensionalidad afecta negativamente al modelo de árboles cuando se dispone de un conjunto de datos limitado (1200 registros), lo cual dificulta la generalización.

Finalmente, se aplicó la técnica de *data augmentation* al *random forest* para evaluar su capacidad de mejora. Los resultados se detallan en la Tabla 8.

Tabla 8

Impacto del data augmentation *en el modelo* random forest

N.°	Modelo	Técnica de datos	Representación (TF-IDF)	Precisión	*Recall*	F1-Score
7	*Random forest*	*Data augmentation*	Unigramas (1-gram)	0,9611	0,9583	0,9573
8	*Random forest*	*Data augmentation*	N-gramas (1,2)	0,9615	0,9575	0,9566
		Promedio		0,9613	0,9579	0,9570

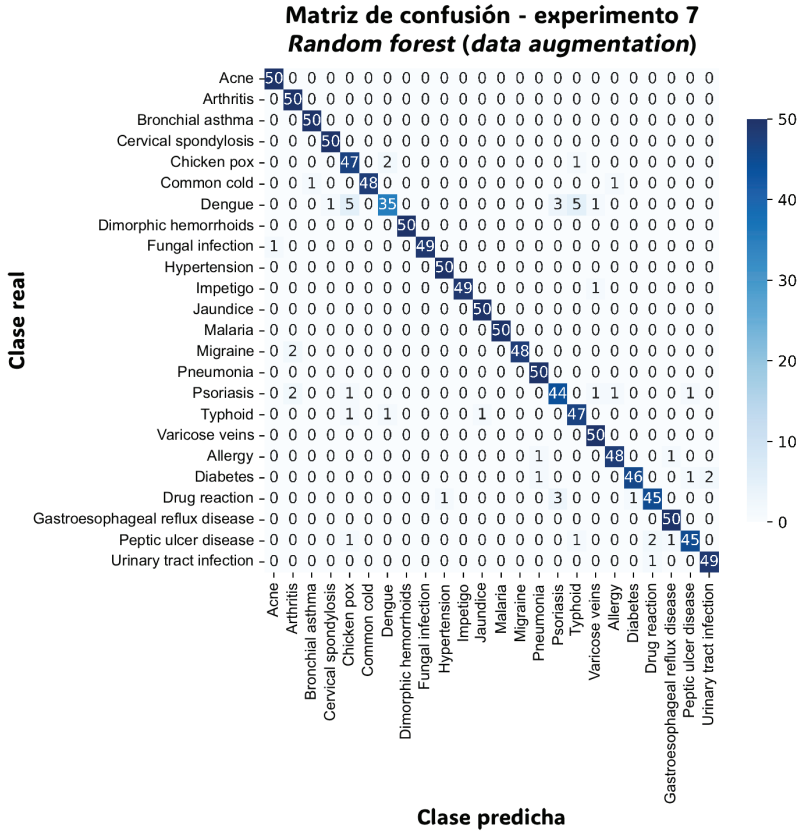
Nota. Métricas de rendimiento obtenidas tras la aplicación de datos sintéticos para mitigar el sobreajuste en el modelo de árboles.

Tal como se aprecia en la Tabla 8, el aumento de datos permitió al *random forest* superar sus métricas base: alcanzó un F1-Score promedio de 0,9570. El mejor escenario para este algoritmo fue el experimento 7 (*data augmentation* + unigramas), con un F1-Score de 0,9573. Aunque la mejora es consistente, la diferencia entre usar unigramas y bigramas se reduce significativamente en presencia de más datos (una diferencia de apenas 0,0007 en F1-Score), lo que indica que el *data augmentation* ayuda a estabilizar el modelo frente a la alta dimensionalidad.

A continuación, se presenta la matriz de confusión del mejor modelo de ensamble obtenido. La comparación entre la matriz de confusión del experimento 3 (regresión logística) y la del experimento 7 (*random forest*) revela un patrón consistente: el modelo de ensamble exhibe mayor dispersión de errores fuera de la diagonal principal, particularmente en enfermedades con vocabulario sintomático compartido. Este comportamiento sugiere que la estructura no lineal del *random forest* tiende a sobreajustar levemente los patrones de clases con pocas instancias cuando el corpus base es reducido (cincuenta registros por clase), incluso con la aplicación de *data augmentation*. Este hallazgo refuerza la recomendación de Pranckevičius & Marcinkevičius (2017) de evaluar sistemáticamente ambos tipos de clasificadores antes de seleccionar el modelo definitivo para aplicaciones clínicas.

Figura 6

Matriz de confusión del mejor modelo random forest *(experimento 7)*



Nota. Distribución de aciertos y confusiones para el algoritmo de ensamble optimizado con unigramas y aumento de datos.

DISCUSIÓN DE LOS RESULTADOS

El análisis comparativo de los ocho escenarios experimentales arroja hallazgos significativos sobre la interacción entre la naturaleza del algoritmo, la técnica de representación y el volumen de datos en el dominio de la clasificación de textos médicos.

En primer lugar, se evidencia una superioridad consistente de la regresión logística sobre el *random forest* para este conjunto de datos específico. El mejor modelo lineal (experimento 3) superó al mejor modelo de ensamble (experimento 7) en más de dos puntos porcentuales en el F1-Score (0,9797 frente a 0,9573). Este hallazgo es coherente con lo expuesto por Genkin et al. (2007), quienes sostienen que los modelos lineales suelen ser altamente efectivos en clasificación de textos, debido a que la alta dimensionalidad de los vectores (TF-IDF) a menudo hace que las clases sean linealmente separables. Mientras que *random forest* destaca por manejar interacciones no lineales complejas (Pal, 2005), en el contexto de descripciones cortas de síntomas, la complejidad del ensamble de árboles parece no aportar una ventaja decisiva frente a la eficiencia de los hiperplanos de decisión de la regresión logística.

Este resultado se alinea con el comportamiento observado en estudios recientes que operan sobre el mismo dominio clínico. Hassan et al. (2024), quienes trabajaron con el mismo corpus Symptom2Disease, reportaron que los modelos de lenguaje profundo como el modelo de normalización de conceptos médicos-representaciones de codificadores bidireccionales de transformadores (*medical concept normalization–bidirectional encoder representations from transformers*, MCN-BERT) superan en exactitud a los clasificadores clásicos; no obstante, los propios autores reconocen que esta ventaja conlleva un costo computacional significativamente mayor y una reducción en la interpretabilidad del modelo, aspecto crítico en entornos de apoyo diagnóstico.

En ese sentido, los resultados del presente estudio indican que, cuando los recursos computacionales son limitados o la transparencia del modelo es un requisito del sistema, la regresión logística representa una alternativa robusta y eficiente que compite favorablemente con arquitecturas más complejas en escenarios de datos pequeños. Esta tesis es consistente con lo reportado por Vyas et al. (2025), quienes encontraron que los modelos lineales bien calibrados con representaciones TF-IDF alcanzan rendimientos competitivos en corpus de tamaño reducido sin sacrificar la capacidad de explicar las predicciones a través de los pesos de sus coeficientes.

En segundo lugar, la aplicación de *data augmentation* demostró ser una estrategia transversalmente efectiva. En ambos algoritmos, el uso de datos sintéticos generados por *random swap* mejoró las métricas de evaluación respecto a la línea base. Esto valida la hipótesis de que, en corpus con un número limitado de muestras por clase (cincuenta registros), la variabilidad léxica artificial actúa como un regularizador potente que ayuda a los modelos a generalizar mejor y evitar el sobreajuste. Este resultado refuerza la

importancia del preprocesamiento robusto en tareas de procesamiento del lenguaje natural (PLN) clínico, en el que la disponibilidad de datos etiquetados suele ser escasa (Báez et al., 2022).

Finalmente, respecto a la representación semántica, los resultados desafiaron la hipótesis inicial de que los N-gramas (bigramas) mejorarían la detección de patologías. Se observó que el uso de unigramas (1-gram) fue marginalmente superior o equivalente al uso de bigramas. Esto sugiere que, en el *dataset* Symptom2Disease, las palabras clave individuales (por ejemplo: *vomiting*, *itching*) poseen una carga discriminativa suficiente por sí mismas. Tal como señalan Aggarwal y Zhai (2012), aumentar la dimensionalidad del espacio de características (agregando bigramas) sin un aumento proporcional en el volumen de datos puede diluir la densidad de información, lo que explica por qué los modelos más simples (unigramas) lograron mantener una precisión tan elevada sin el costo computacional adicional de procesar secuencias de palabras.

En conjunto, los resultados sugieren que para tareas de triaje o diagnóstico preliminar basado en textos breves, un modelo lineal bien calibrado con técnicas de aumento de datos puede ofrecer un rendimiento superior y más eficiente que arquitecturas más complejas. Desde una perspectiva práctica, los resultados de este estudio aportan evidencia empírica relevante para el diseño de sistemas de apoyo al diagnóstico en entornos con recursos limitados. La adopción de un modelo lineal como la regresión logística, combinado con un *pipeline* de preprocesamiento estándar y *data augmentation* mediante *random swap*, ofrece una arquitectura computacionalmente eficiente, interpretable y replicable que facilita la validación por profesionales médicos y su integración en flujos de trabajo clínicos reales.

Asimismo, la alta exactitud alcanzada en las 24 categorías diagnósticas sugiere que este tipo de enfoque podría extenderse a sistemas de triaje automatizado en entornos de telemedicina, en los que la entrada del usuario es una descripción textual libre de sus síntomas (Al-qarni & Algarni, 2025). Sin embargo, la validación de estos resultados en corpus clínicos en español, con mayor variabilidad léxica y solapamiento sintomático, constituye el siguiente paso natural de esta línea de investigación.

CONCLUSIONES

Los resultados obtenidos en este estudio evidencian el potencial y la viabilidad de los modelos clásicos de aprendizaje automático para abordar con eficacia la clasificación automática de textos médicos breves. Tanto la regresión logística como el *random forest* demostraron un desempeño notable, pues superaron el umbral del 95 % de exactitud global tras la aplicación de técnicas de optimización, lo que confirma que, incluso con conjuntos de datos limitados, es posible construir herramientas de apoyo al diagnóstico con alta fiabilidad mediante un preprocesamiento riguroso.

Desde una perspectiva práctica, los hallazgos de este estudio ofrecen una guía metodológica concreta para equipos de investigación que desarrollen herramientas de clasificación de texto médico con recursos limitados. Se recomienda priorizar modelos lineales como la regresión logística cuando el corpus disponible es reducido y el vocabulario diagnóstico es específico; complementar el conjunto de entrenamiento con técnicas de *data augmentation* léxica para mitigar el efecto del bajo número de muestras por clase; y optar por representaciones TF-IDF con unigramas como punto de partida, dado que la incorporación de bigramas puede deteriorar el rendimiento en corpus de pequeñas dimensiones. Estas recomendaciones son especialmente pertinentes en el contexto de la salud digital en economías en desarrollo, donde la escasez de grandes corpus médicos etiquetados obliga a maximizar el rendimiento de los datos disponibles (Calero Sánchez et al., 2024).

En particular, el modelo de regresión logística logró los mejores indicadores de rendimiento de toda la investigación, pues se destacó con un F1-Score de 0,9797 en su configuración óptima. Estos hallazgos confirman su idoneidad para manejar espacios vectoriales de alta dimensionalidad (TF-IDF), en los que la linealidad de sus fronteras de decisión resultó ser más eficaz que las estructuras no lineales de los árboles para discriminar síntomas específicos. Su capacidad para asignar pesos interpretables a términos clave refuerza su valor en aplicaciones biomédicas, en las que la eficiencia computacional es prioritaria.

Por su parte, el algoritmo *random forest* también obtuvo resultados sólidos, alcanzó un F1-Score de 0,9573, aunque mostró una leve inferioridad respecto al modelo lineal. Esta diferencia de rendimiento sugiere que, en el contexto de descripciones cortas y dispersas de síntomas, la complejidad del ensamble de árboles no logra capitalizar sus ventajas habituales sobre la linealidad y puede verse más afectada por la dispersión de los datos. No obstante, su desempeño se mantuvo estable y competitivo validándose como una alternativa robusta frente al ruido.

Un hallazgo transversal y significativo fue el impacto determinante de la técnica de *data augmentation*. El incremento de métricas observado en ambos modelos, al duplicar el corpus mediante intercambio léxico (*random swap*), corrobora que la variabilidad sintética actúa como un regularizador eficaz, pues mitiga el sobreajuste inherente a *datasets* pequeños. Asimismo, el análisis de la representación semántica reveló que los unigramas poseen suficiente carga discriminativa por sí mismos; la incorporación de N-gramas (bigramas) no tradujo mejoras sustanciales, lo que evidencia que la densidad terminológica de los síntomas individuales es el factor predictor dominante en este dominio.

Cabe señalar también un aspecto estructural relevante del corpus analizado: las 24 enfermedades que componen el conjunto de datos presentan cuadros clínicos considerablemente distintos entre sí, de modo que sus síntomas característicos resultan altamente específicos para cada condición. Por ejemplo, las erupciones cutáneas escamosas de la

psoriasis difícilmente se confunden con la fiebre, el dolor articular y el exantema hemorrágico del dengue. Esta especificidad léxica favorece que los modelos alcancen métricas de exactitud elevadas. Sin embargo, en escenarios clínicos reales, donde dos o más enfermedades comparten síntomas comunes (como fiebre, fatiga o dolor muscular), la tarea de clasificación se vuelve considerablemente más compleja y los clasificadores podrían registrar un desempeño inferior.

En consecuencia, la capacidad de generalización de los modelos presentados debe interpretarse en el contexto de un corpus con alta separabilidad interclase. Extender estos resultados a conjuntos de datos con mayor solapamiento sintomático constituye una limitación del presente trabajo y una línea de investigación futura prioritaria, que podría abordarse mediante la incorporación de corpus con mayor complejidad diagnóstica o la exploración de representaciones contextuales de mayor profundidad semántica.

En conjunto, estos resultados respaldan la implementación de flujos de trabajo basados en modelos lineales optimizados y enriquecimiento de datos como estrategia costo-efectiva para la clasificación de literatura médica. A su vez, ponen en evidencia que la clave para optimizar estos sistemas no reside necesariamente en aumentar la complejidad del algoritmo, sino en refinar las estrategias de curación de datos y representación de características sentando una base sólida para futuras investigaciones que busquen integrar estas herramientas en entornos clínicos reales.

REFERENCIAS

Aggarwal, C. C., & Zhai, C. (2012). A survey of text classification algorithms. En C. C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 163-222). Springer US. https://doi.org/10.1007/978-1-4614-3223-4_6

Al-qarni, S. S., & Algarni, A. (2025). Disease prediction from symptom descriptions using deep learning and NLP technique. *International Journal of Advanced Computer Science and Applications*, *16*(5), 416-426. https://doi.org/10.14569/IJACSA.2025.0160541

Alsaeedi, A. (2020). A survey of term weighting schemes for text classification. *International Journal of Data Mining, Modelling and Management*, *12*(2), Artículo 237. https://doi.org/10.1504/IJDMMM.2020.10028060

Báez, P., Arancibia, A. P., Chaparro, M. I., Bucarey, T., Núñez, F., & Dunstan, J. (2022). Procesamiento de lenguaje natural para texto clínico en español: el caso de las listas de espera en Chile. *Revista Médica Clínica Las Condes*, *33*(6), 576-582. https://doi.org/10.1016/J.RMCLC.2022.10.002

Barman, N., Karim, F., & Sharma, K. (2023). *Symptom2Disease: Diseases and natural language symptom descriptions*. https://www.kaggle.com/datasets/niyarrbarman/symptom2disease?resource=download

Calero Sánchez, M., González, J. C., Sánchez Berriel, I., Burillo-Putze, G., & Roda García, J. L. (2024). Natural language processing for reviewing search results from PubMed. *Revista Española de Urgencias y Emergencias*, *3*, 184-195. https://doi.org/10.55633/S3ME/REUE030.2024

Cardoso, F. E., Ferneda, E., & Botega, L. (2023). Clasificación de textos: un enfoque con uso de *machine learning*. *Revista EDICIC*, *3*(3), 1-17. https://doi.org/10.62758/re.v3i3.212

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc.* DaimlerChrysler. https://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/14.2/es/CRISP-DM.pdf

Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale bayesian logistic Regression for text categorization. *Technometrics*, *49*(3), 291-304. https://doi.org/10.1198/004017007000000245

Hassan, E., Abd El-Hafeez, T., & Shams, M. Y. (2024). Optimizing classification of diseases through language model analysis of symptoms. *Scientific Reports*, *14*(1), Artículo 1507. https://doi.org/10.1038/s41598-024-51615-5

Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. *Information*, *10*(4), Artículo 150. https://doi.org/10.3390/info10040150

Li, Q., Peng, H., Li, J., Xia, C., Yang, R., Sun, L., Yu, P. S., & He, L. (2022). A survey on text classification: From traditional to deep learning. *ACM Transactions on Intelligent Systems and Technology*, *13*(2), 1-41. https://doi.org/10.1145/3495162

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26* (1), 217-222. https://doi.org/10.1080/01431160412331269698

Powers, D. (2008). Evaluation: From precision, recall and F-factor to ROC, informedness, markedness & correlation. *Mach. Learn. Technol.*, *2*(1), 37-63. https://doi.org/10.48550/arXiv.2010.16061

Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of Naive Bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, *5*(2), 221-232. https://doi.org/10.22364/bjmc.2017.5.2.05

Sun, Y., Li, Y., Qingtao, Z., & Bian, Y. (2020). Application research of text classification based on random forest algorithm. En *2020 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)* (pp. 370-374). IEEE. https://doi.org/10.1109/AEMCSE50948.2020.00086

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. (2019). Machine learning algorithm validation with a limited sample size. *PLOS ONE*, *14*(11), 1-20. https://doi.org/10.1371/journal.pone.0224365

Vyas, D., Shah, M., Kantawala, H., Patel, B., Patel, T., & Enamala, J. (2025). SympTextML: Leveraging natural language symptom descriptions for accurate multi-disease prediction. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *7*(3), 911-924. https://doi.org/10.35882/jeeemi.v7i3.946

Wei, J., & Zou, K. (2019). EDA: Easy data augmentation techniques for boosting performance on text classification tasks. En K. Inui, J. Jiang, V. Ng & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 6382-6388). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1670

Evaluación psicométrica de los constructos predominantes en la adopción de ChatGPT en la educación superior: una muestra mexicana

María Azeneth Duana Guerra
https://orcid.org/0009-0008-1725-5751
mduanag001@alumno.uaemex.mx
Universidad Autónoma del Estado de México

Alejandra Morales Ramírez
http://orcid.org/0000-0002-8737-5985
amoralesr@uaemex.mx
Universidad Autónoma del Estado de México

Guadalupe Montserrat Figueroa Castañeda
https://orcid.org/0009-0005-2437-6541
gfigueroac001@alumno.uaemex.mx
Universidad Autónoma del Estado de México

Cuauhtémoc Hidalgo Cortés
http://orcid.org/0000-0001-6324-7180
chidalgoc@uaemex.mx
Universidad Autónoma del Estado de México

Juan de Jesús Amador Reyes
https://orcid.org/0000-0003-1925-2710
jjamadorr@uaemex.mx
Universidad Autónoma del Estado de México

Recibido: 29 de abril del 2026 / Aceptado: 15 de junio del 2026
doi: https://doi.org/10.26439/interfases2026.n023.8785

RESUMEN. El estudio evaluó las propiedades psicométricas de los constructos asociados a la adopción educativa de ChatGPT. El instrumento se diseñó sobre la base de Morales et al. (2025), quienes identificaron los constructos más recurrentes en la literatura, provenientes de modelos como el basado en el valor percibido (VAM), el modelo de aceptación de la tecnología (TAM) y la teoría unificada de aceptación y uso de la tecnología (UTAUT). Se seleccionaron los constructos utilidad percibida, influencia social, intención conductual y facilidad de uso, incorporando, además, el riesgo percibido por su relevancia emergente.

Participaron 579 universitarios asignados aleatoriamente en dos muestras (M1 = 281 y M2 = 298). El tiempo de uso de ChatGPT reportado por los participantes muestra una distribución equilibrada entre quienes tienen menos de un año de experiencia y quienes lo han utilizado más tiempo. Respecto al nivel de competencia, predomina el dominio básico, seguido del intermedio y una minoría reporta un manejo avanzado. Aunque ChatGPT es utilizado por todos los participantes, no es la única herramienta empleada; estos complementan su uso con otras IA generativas. Los principales patrones de uso se relacionan con la síntesis de información, la explicación de conceptos complejos y la identificación de tendencias; en contraste, su uso en tareas de evaluación crítica es menor. El análisis factorial exploratorio identificó una estructura de cinco factores que explica el 59,08 % de la varianza. El análisis factorial confirmatorio mostró índices de ajuste adecuados del modelo (RMSEA = 0,047; SRMR = 0,058; CFI = 0,942; TLI = 0,934 y PClose = 0,753). Se destaca que la adopción de ChatGPT está asociada con la utilidad percibida, la facilidad de uso, el entorno social y el riesgo percibido. Por tanto, el instrumento obtenido puede considerarse válido y confiable para evaluar la adopción de ChatGPT en estudiantes universitarios mexicanos.

PALABRAS CLAVE: ChatGPT / TAM / VAM / UTAUT / análisis factorial confirmatorio / análisis factorial exploratorio

PSYCHOMETRIC EVALUATION OF THE PREDOMINANT CONSTRUCTS IN THE ADOPTION OF CHATGPT IN HIGHER EDUCATION: A MEXICAN SAMPLE

ABSTRACT. The purpose of this study was to evaluate the psychometric properties of the constructs involved in the educational adoption of ChatGPT. The instrument was designed based on Morales et al. (2025), who identified the most recurrent constructs in the literature, derived from models such as TAM, VAM, and UTAUT. Perceived Utility, Social Influence, Behavioral Intention, and Ease of Use were selected, and Perceived Risk was also incorporated due to its emerging relevance. A total of 579 university students participated in the study, were randomly assigned to two samples (M1=281 and M2=298). The time used of ChatGPT by participants shows a balanced distribution between those with less than one year of experience and those who had used it for a longer period. However, regarding competence level, basic proficiency predominated, followed by intermediate proficiency, while only a minority reported advanced use of the tool. Although ChatGPT is used by all participants, it is not an exclusive tool, as they complement its use with other generative AI tools. The primary usage patterns involve information synthesis, explaining complex concepts, and providing examples or insights into current trends; in contrast, its use in critical evaluation tasks is less common. The exploratory factor analysis identified a five-factor structure that explained 59.08% of the variance. The confirmatory factor analysis showed adequate model fit indices (RMSEA=0.047; SRMR=0.058; CFI=0.942; TLI=0.934; and PClose=0.753). Highlighting that the use and adoption of ChatGPT are associated with

perceived usefulness, ease of use, the social environment, and perceived risk. Therefore, the resulting instrument can be considered valid and reliable for assessing ChatGPT adoption among Mexican university students.

KEYWORDS: ChatGPT / TAM / VAM / UTAUT / confirmatory factor analysis / exploratory factor analysis

INTRODUCCIÓN

La transformación digital en la educación superior se ha acelerado de manera significativa en los últimos años, impulsada por el desarrollo y la integración de tecnologías basadas en inteligencia artificial (IA). Este cambio comienza a transformar los procesos de enseñanza, aprendizaje, evaluación e investigación, lo que ha generado que las instituciones educativas empiecen a analizar y replantear sus prácticas pedagógicas, sus políticas de integridad académica y sus modelos de innovación educativa (Rizun et al., 2026; United Nations Educational, Scientific and Cultural Organization [Unesco], 2023). En este contexto, la IA no solo se concibe como una herramienta tecnológica, sino también como un recurso que puede influir en las creencias, actitudes y comportamientos de docentes y estudiantes respecto de su adopción y utilización en el entorno educativo.

Entre las aplicaciones emergentes de la IA, los sistemas de lenguaje generativo han adquirido gran importancia. Tal es el caso de ChatGPT, que fue lanzado públicamente a finales del 2022 y que está cobrando relevancia como una de las herramientas más destacadas en el ámbito educativo (Acosta-Enríquez et al., 2024) debido a su capacidad para generar textos coherentes y contextualizados, con un estilo conversacional similar al humano (Nemt-allah et al., 2024; Rizun et al., 2026; Vaswani et al., 2017).

Su rápida adopción ha alcanzado millones de usuarios en muy poco tiempo (García-Peñalvo et al., 2024) y ha generado un notable interés académico por entender su impacto en la educación superior, tanto por sus beneficios como por los desafíos éticos (Abdi et al., 2025), pedagógicos y cognitivos que presenta (Romero-Rodríguez et al., 2023), especialmente en relación con la manera en cómo los estudiantes emplean esta herramienta para sus actividades académicas.

Investigaciones recientes destacan que ChatGPT puede apoyar en diversas actividades escolares, como la producción de textos (Song & Song, 2023), la síntesis de información, la resolución de problemas, la programación, el aprendizaje personalizado junto al análisis y comparación de fuentes (García-Peñalvo et al., 2024). Aunado a que diversos estudios reportan percepciones positivas respecto a su utilidad y facilidad de uso, así como su contribución a la reducción de la carga cognitiva y de la ansiedad académica, lo cual favorece no solo su adopción, sino también su incorporación en prácticas académicas cotidianas.

No obstante, también se han señalado riesgos asociados a su uso, como problemas éticos (Rizun et al., 2026), dependencia excesiva, derechos de autor (Abdaljaleel et al., 2024; Abdi et al., 2025; Unesco, 2023), presencia de información inexacta o sesgada (Balaskas et al., 2025; Hsu & Silalahi, 2024; Kooli, 2023; Sallam et al., 2023), implicaciones para la integridad académica, entre otras preocupaciones.

Desde una perspectiva teórica, la adopción de ChatGPT en contextos educativos ha sido analizada mediante diversos modelos consolidados de adopción y aceptación

tecnológica, como el modelo basado en el valor percibido (*value-based model*, VAM) (Al-Abdullatif & Alsubaie, 2024), el modelo de aceptación de la tecnología (*technology acceptance model*, TAM) (Xu et al., 2024) y la teoría unificada de aceptación y uso de la tecnología (*unified theory of acceptance and use of technology*, UTAUT) (Bouteraa et al., 2024; Rana et al., 2024), y sus extensiones (Lai et al., 2023), los cuales se han empleado en diversos estudios, complementados con enfoques centrados en las actitudes, la motivación y el comportamiento del usuario. Estos marcos han permitido identificar constructos clave, como la utilidad percibida, la influencia social, la intención conductual, la facilidad de uso, entre otros (Morales et al., 2025), los cuales resultan relevantes para comprender tanto la disposición de los usuarios para utilizar la tecnología como su implementación efectiva en los contextos educativos. Sin embargo, a pesar del creciente volumen de estudios sobre la IA generativa en la educación superior, persiste una limitación importante relacionada con la disponibilidad de instrumentos psicométricamente validados que permitan evaluar los constructos asociados a su adopción educativa en diferentes contextos nacionales (Rizun et al., 2026).

Tomando en cuanto todo lo anterior, el objetivo de este estudio fue evaluar las propiedades psicométricas de los constructos predominantes que influyen en la adopción educativa de ChatGPT en una muestra de estudiantes universitarios.

METODOLOGÍA

Muestra

Se trabajó con una muestra total de 579 participantes, la cual se dividió aleatoriamente en dos muestras independientes de tipo no probabilístico, conformadas por estudiantes universitarios. La primera muestra incluyó a 281 participantes con edades entre 17 y 54 años: 114 hombres con una media de edad de 20,44 años (DE = 3,95) y 167 mujeres con una edad promedio de 19,75 años (DE = 1,71). La segunda muestra estuvo integrada por 298 estudiantes con edades entre 17 a 50 años: 188 fueron mujeres (M = 19,73; DE = 2,09) y 110 fueron hombres (M = 20,06; DE = 3,63). Todos los participantes cursaban estudios en una institución pública del Estado de México, pertenecientes a seis carreras universitarias y un posgrado en Ciencias Computacionales. La diversidad de las carreras universitarias proporciona una visión integral sobre la participación de estudiantes de distintas áreas del conocimiento.

Tabla 1

*Perfil sociodemográfico de la muestra (*n = *579)*

Características demográficas	Categorías	M1 (281 participantes)		M2 (298 participantes)	
		N	%	N	%
Género	Hombres	114	40,60	110	36,9
	Mujeres	167	59,40	188	63,1
Edad (años)	17-19	135	48,04	162	54,3
	20-22	120	42,70	109	36,5
	23-25	20	7,12	18	6,0
	Más de 25	6	2,14	9	3,2
Carrera universitaria	Contaduría	47	16,70	55	18,5
	Derecho	39	13,90	54	18,1
	Psicología	42	14,90	61	20,5
	Administración	56	19,90	48	16,1
	Ingeniería en Computación	50	17,80	33	11,1
	Informática Administrativa	45	16,10	43	14,4
	Maestría en Ciencias Computacionales	2	0,70	4	1,3
Total		281	100	298	100

Instrumento

El cuestionario aplicado se diseñó a partir de los hallazgos reportados por Morales et al. (2025), quienes identificaron los constructos empleados para explicar la adopción de ChatGPT en contextos académicos universitarios. Estos constructos se encuentran integrados en diversos modelos teóricos, tales como UTAUT, TAM, VAM y sus extensiones. En particular, para la presente investigación se seleccionaron los primeros cuatro constructos que mostraron una mayor presencia en la literatura científica: utilidad percibida, influencia social, intención conductual y facilidad de uso. Adicionalmente, se incorporó el constructo riesgo percibido debido a la creciente evidencia sobre los posibles efectos asociados al uso de ChatGPT en contextos educativos.

Con base en lo anterior, se estructuró un instrumento de tres secciones: la primera orientada a la recolección de datos demográficos (género, edad, carrera, semestre, promedio general y turno); la segunda destinada a evaluar la experiencia y el uso de la IA generativa específicamente de ChatGPT mediante seis preguntas; y la tercera enfocada en explorar la percepción estudiantil sobre la adopción de ChatGPT en el ámbito académico a través de 46 ítems (Tabla 2) tipo Likert de cinco opciones (Totalmente

en desacuerdo = 1, En desacuerdo = 2, Neutral = 3, De acuerdo = 4, Totalmente de acuerdo = 5), distribuidos en los siguientes constructos:

- Utilidad percibida (UP). Se conceptualiza como el grado en que los usuarios consideran que ChatGPT puede mejorar su rendimiento académico, productividad y resultados de aprendizaje (Rudolph et al., 2023). La literatura señala que la adopción de una tecnología aumenta cuando los usuarios la perciben como una herramienta útil para alcanzar sus objetivos educativos.
- Influencia social (IS). Hace referencia a la percepción de un individuo respecto a que personas significativas dentro de su entorno (familiares, pareja, amistades, docentes o compañeros) consideran que debería utilizar ChatGPT (Parveen et al., 2024). Este constructo refleja la presión social, tanto explícita como implícita, que puede fomentar o inhibir el uso de la herramienta como parte de las actividades académicas.
- Intención conductual (IC). Describe la probabilidad subjetiva de que un usuario continúe utilizando la tecnología en el futuro (Venkatesh et al., 2003), en este caso ChatGPT.
- Facilidad de uso (FU). Se define como el grado en que el usuario percibe que la interacción con ChatGPT es sencilla, clara y libre de esfuerzo (Abdullatif, 2024; Acosta-Enríquez et al., 2024). Este constructo se refleja en aspectos como la facilidad para comprender su funcionamiento, la accesibilidad de sus funciones y la simplicidad de su uso.
- Riesgo percibido (RP). Alude a que el uso inadecuado de ChatGPT puede generar diversas amenazas en el ámbito académico, por ejemplo, plagio, deshonestidad académica, sobredependencia tecnológica (Cotton et al., 2024), pérdida de pensamiento crítico y propagación de desinformación (Rudolph et al., 2023).

Tabla 2

Ítems representativos del instrumento

Constructo	Ejemplo de ítems	Fuente
Utilidad percibida	Usar ChatGPT me permite ser más productivo(a) y eficiente en las tareas académicas, al brindarme información precisa y relevante.	Abdalla (2024), Abdullatif (2024), Almogren et al. (2024), Güner et al. (2024), Mahmud et al. (2024)
Influencia social	Las personas que son importantes para mí creen que debería utilizar ChatGPT en mis cursos.	Güner et al. (2024)
Intención conductual	Estoy decidido a utilizar ChatGPT de manera constante en mis estudios.	Elshaer et al. (2024), Parveen et al. (2024)

(*continúa*)

(*continuación*)

Constructo	Ejemplo de ítems	Fuente
Facilidad de uso	Mi interacción con ChatGPT es clara y comprensible.	Bouteraa et al. (2024), Rana et al. (2024)
Riesgo percibido	Me preocupa que el uso de ChatGPT pueda ser malinterpretado como una falta académica (plagio o trampa).	Sallam et al. (2023)

Nota. La versión completa se encuentra en el Anexo A.

Procedimiento

Los estudiantes fueron invitados a participar en el estudio tras obtener la autorización por parte de las autoridades universitarias. En primera instancia, se explicó el objetivo de la investigación y se dieron las indicaciones necesarias para acceder y completar el cuestionario. Posteriormente, se les facilitó el formulario de consentimiento informado. Una vez leído y aceptado, se habilitaron las tres secciones del cuestionario para su respuesta. La participación fue voluntaria, con un tiempo estimado que osciló entre 5 y 10 minutos.

Análisis estadísticos

En una primera etapa, se realizó el análisis factorial exploratorio (AFE) en el programa Statistical Package for the Social Sciences (SPSS), versión 27. Para ello, se empleó el método de extracción por ejes principales y una rotación oblicua tipo Promax. El número de factores se determinó mediante el análisis paralelo de Horn, se conservaron únicamente los ítems cuya carga factorial fuera igual o superior a 0,40. Los reactivos que presentaron cargas cruzadas con una diferencia menor a 0,15 fueron descartados y aquellos cuya diferencia fue igual o mayor a 0,15 se mantuvieron dentro del factor en el que mostraron la mayor saturación (Costello & Osborne, 2005; Velicer & Fava, 1998; Worthington & Whittaker, 2006). La consistencia interna se evaluó mediante el coeficiente alfa de Cronbach.

En una segunda etapa, se realizó el análisis factorial confirmatorio (AFC) en el *software* SPSS Amos. En esta fase, se calcularon el estadístico χ^2, el índice de Tucker-Lewis (TLI), el índice de ajuste comparativo (CFI), la raíz cuadrada media residual (RMSR), el error cuadrático medio de aproximación (RMSEA), el índice PClose (Lai, 2021; Xia & Yang, 2019) y el coeficiente de fiabilidad compuesta (CR).

RESULTADOS

Tiempo de uso y nivel de habilidad en el uso de ChatGPT

En relación con el tiempo de uso de ChatGPT, el 51 % de los participantes indica que tiene menos de un año utilizando esta herramienta, mientras que el 49 % reporta un uso

superior a un año. Esta distribución relativamente equilibrada muestra que, aunque una parte importante de los jóvenes estudiantes es usuario reciente, una proporción similar ya cuenta con una experiencia más prolongada en la interacción con ChatGPT.

En cuanto al nivel de competencia, se observa que el 69 % de los universitarios reconoce tener un dominio básico de la herramienta, es decir, que probablemente utiliza ChatGPT para realizar tareas simples y de apoyo inmediato. Por su parte, el 21 % reporta tener un nivel intermedio, lo que sugiere que ellos tienen más experiencia en el manejo de las funcionalidades de la herramienta. En contraste, únicamente el 10 % afirma poseer un dominio avanzado.

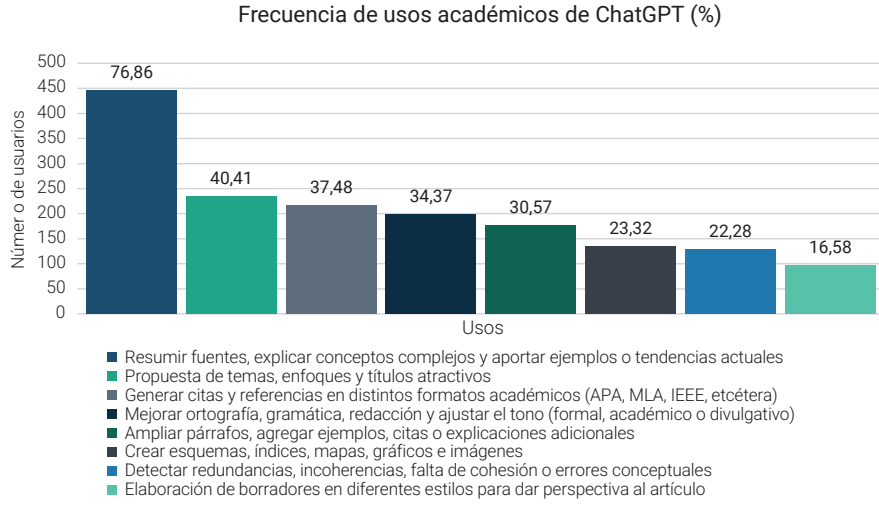
Usos académicos de ChatGPT

Con el fin de identificar los principales usos de ChatGPT, el instrumento incluyó una pregunta de opción múltiple para explorar las preferencias académicas de los participantes. Los resultados muestran (Figura 1) que el uso más frecuente se relaciona con la síntesis de fuentes de información, explicar conceptos complejos, aportar ejemplos y tendencias actuales (76,86 %). En segundo lugar, se identificó su uso para la generación de propuestas de temas, enfoques y títulos atractivos (40,41 %).

En contraste, los usos menos reportados por los universitarios fueron la detección de redundancias, incoherencias, falta de cohesión o errores conceptuales (22,28 %), así como la elaboración de borradores en diferentes estilos con el fin de ofrecer diversas perspectivas de redacción (16,58 %). Es decir, las funciones relacionadas con la revisión crítica son utilizadas con menor frecuencia por los estudiantes.

Figura 1

Distribución de los usos académicos de ChatGPT

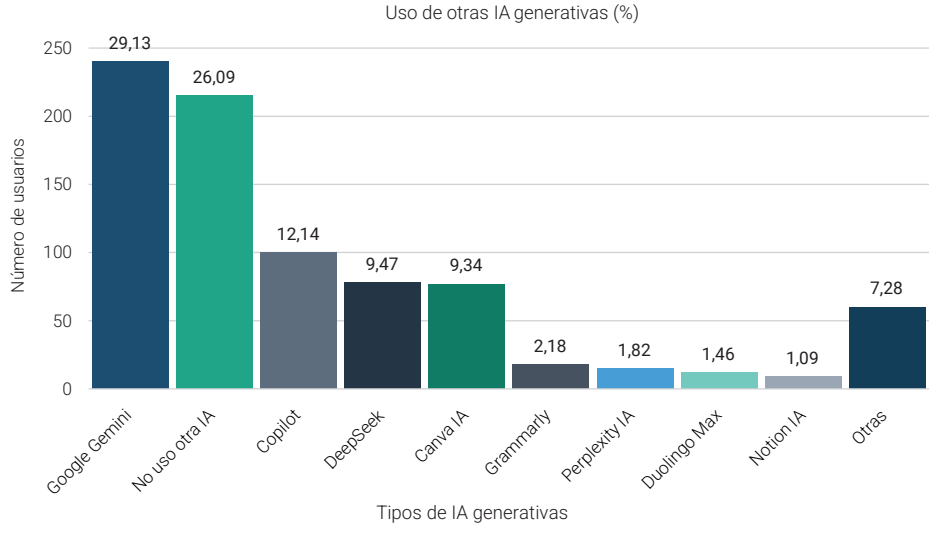


Uso de otras IA generativas

La Figura 2 muestra otras herramientas de IA generativa utilizadas por los estudiantes, además de ChatGPT. Entre ellas, Google Gemini destaca como la plataforma con mayor presencia, seguida por Copilot, DeepSeek y Canva IA. En contraste, Grammarly, Perplexity, Duolingo Max y Notion IA muestran niveles de adopción considerablemente menores. Asimismo, algunos estudiantes indicaron utilizar otras herramientas que no fueron especificadas en el instrumento. Aunque ChatGPT es la herramienta predominante entre los participantes, su uso no es exclusivo, los estudiantes la complementan con otras IA generativas, lo que refleja una inclinación hacia la adopción múltiple de estas tecnologías en el ámbito académico.

Figura 2

Uso de otras aplicaciones de IA generativa



Análisis factorial exploratorio

El AFE se realizó con la primera muestra de 281 participantes. La solución factorial inicial sugirió un modelo de nueve factores; no obstante, tras examinar el tamaño de las cargas factoriales (> 0,40) y las cargas cruzadas con una diferencia menor a 0,15, se procedió a eliminar 20 ítems que no cumplían con los criterios establecidos. Como resultado, se obtuvo un modelo final integrado por 26 ítems distribuidos en cinco factores (Tabla 2).

Los indicadores de adecuación muestral confirmaron la pertinencia de aplicar el AFE. En primer lugar, la medida Kaiser-Meyer-Olkin (KMO = 0,891) evidenció un nivel adecuado de adecuación de la muestra. Asimismo, la prueba de esfericidad de Bartlett resultó

estadísticamente significativa (χ^2 = 3046,99, *gl* = 325, $p < 0{,}001$), lo que indica la existencia de correlaciones suficientes entre las variables para realizar el análisis factorial.

En términos de varianza explicada, el primer factor explicó el 26,73 % de la varianza total; el segundo, el 16,03 %; el tercero, el 6,53 %; el cuarto, el 5,55 %; y el quinto, el 4,24 %. En conjunto, estos factores explican el 59,08 % de la varianza total. Dichos factores corresponden conceptualmente a los constructos teóricos incluidos en el instrumento: UP, IS, IC, FU y RP.

Finalmente, en cuanto a la consistencia interna de los constructos, los coeficientes alfa de Cronbach obtenidos para cada uno de los factores fueron superiores al umbral recomendado de 0,70, lo que demuestra una adecuada fiabilidad de las escalas y respalda la consistencia interna del instrumento (Tabla 3).

Tabla 3

*Análisis factorial exploratorio a cinco factores (*n = *281)*

Factor	Ítems (carga factorial)			α
Riesgo percibido	RP10 (0,859) RP2 (0,828)	RP9 (0,783) RP3 (0,765)	RP7 (0,730) RP16 (0,689)	0,869
Facilidad de uso	FU4 (0,799) FU3 (0,796)	FU8 (0,782) FU2 (0,647)	FU6 (0,597)	0,804
Utilidad percibida	UP2 (0,781) UP3 (0,764)	UP7 (0,717) UP1 (0,682)	UP5 (0,646) UP6 (0,589)	0,806
Influencia social	IS5 (0,853) IS6 (0,805)	IS3 (0,648)	IS2 (0,544)	0,720
Intención conductual	IC2 (0,821) IC4 (0,698)	IC1 (0,682) IC3 (0,662)	IC5 (0,605)	0,805

Análisis factorial confirmatorio

Para confirmar la estructura factorial de los cinco factores obtenidos previamente con el AFE, se realizó el AFC utilizando la segunda muestra compuesta por 298 participantes (Figura 3).

El análisis se llevó a cabo a partir de la matriz de correlaciones, empleando el método de estimación de máxima verosimilitud. Los resultados obtenidos fueron los siguientes: $\chi^2(289) = 477{,}19$, $p < 0{,}001$; $\chi^2/gl = 1{,}651$; CFI = 0,942; TLI = 0,934; RMSEA = 0,047; PClose = 0,753 y SRMR = 0,058.

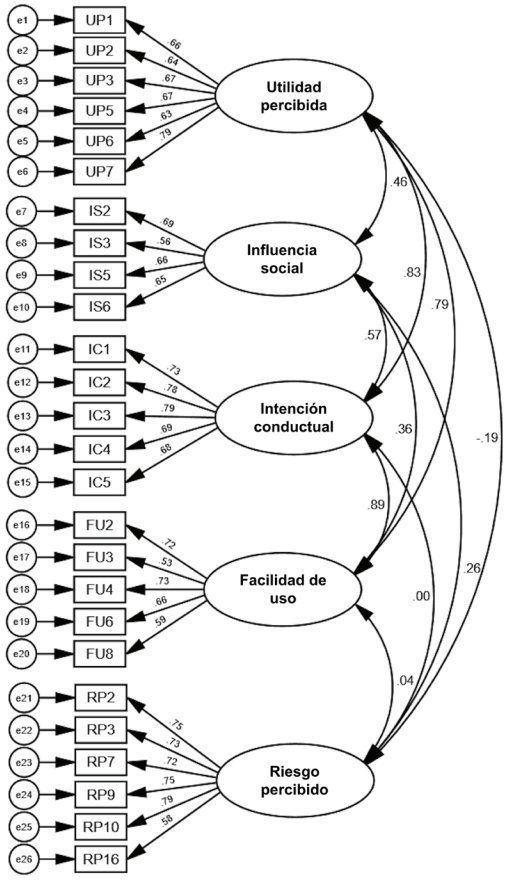
De acuerdo con los criterios comúnmente aceptados en la literatura metodológica, valores de CFI y TLI iguales o superiores a 0,90 denotan un ajuste adecuado, en tanto que valores superiores a 0,95 se asocian con un ajuste excelente. Asimismo, valores de RMSEA inferiores a 0,06 y SRMR menores a 0,08 se consideran indicativos de un buen ajuste del modelo (Jordan, 2021; Xia & Yang, 2019). Adicionalmente, el índice χ^2/gl

permite evaluar la relación entre el modelo teórico y los datos observados; valores entre 1 y 3 reflejan un buen ajuste, mientras que valores situados entre 3,1 y 5 se consideran aceptables (Hooper et al., 2008; Schumacker & Lomax, 2004).

Por su parte, el índice PClose evalúa la hipótesis de ajuste cercano del modelo basado en el RMSEA; valores superiores a 0,05 sugieren que no se rechaza la hipótesis de un ajuste cercano, lo que respalda la adecuación del modelo a los datos observados (Xia & Yang, 2019).

Figura 3

Modelo de medición con cargas factoriales



Respecto de la evaluación de los indicadores, las cargas factoriales estandarizadas reflejan la fortaleza de la relación entre cada ítem y su constructo correspondiente (Figura 3). Los resultados muestran que todos los ítems presentan cargas iguales o superiores a 0,50, además de ser estadísticamente significativas ($p < 0{,}001$), lo que indica que

la asociación observada entre los indicadores y sus respectivos factores no es atribuible al azar (Tabla 4).

Tabla 4

*Resultados del AFC (*n *= 298)*

Factores	Ítems	Cargas estandarizadas	t-valor	p
Utilidad percibida	UP1	0,655	-	-
	UP2	0,637	9,556	<0,001
	UP3	0,673	10,025	<0,001
	UP5	0,665	9,925	<0,001
	UP6	0,631	9,487	<0,001
	UP7	0,795	11,451	<0,001
Influencia social	IS2	0,691	-	-
	IS3	0,561	7,912	<0,001
	IS5	0,664	8,998	<0,001
	IS6	0,655	8,912	<0,001
Intención conductual	IC1	0,734	-	-
	IC2	0,777	13,067	<0,001
	IC3	0,788	13,309	<0,001
	IC4	0,690	11,590	<0,001
	IC5	0,685	11,478	<0,001
Fiabilidad de uso	FU2	0,724	-	-
	FU3	0,528	8,486	<0,001
	FU4	0,733	11,731	<0,001
	FU6	0,658	10,559	<0,001
	FU8	0,591	9,501	<0,001
Riesgo percibido	RP2	0,747	-	-
	RP3	0,733	12,167	<0,001
	RP7	0,719	11,879	<0,001
	RP9	0,753	12,514	<0,001
	RP10	0,878	13,073	<0,001
	RP16	0,582	9,571	<0,001

Nota. "-" = Ítem restringido con fines de identificación del modelo de medida.

Adicionalmente, como se muestra en la Tabla 5, se estimó el CR para cada uno de los constructos del modelo. Los valores obtenidos fueron los siguientes: UP (CR = 0,835), IS (CR = 0,739), IC (CR = 0,855), FU (CR = 0,784) y RP (CR = 0,867). En todos los casos, los coeficientes superan el umbral recomendado de 0,70, lo que indica una adecuada consistencia interna de los factores. Asimismo, los valores de la varianza extraída media (AVE) se situaron en niveles aceptables entre 0,501 y 0,541, lo que cumple satisfactoriamente con el criterio de superar el umbral mínimo recomendado de 0,50.

En términos de la normalidad de los datos, los coeficientes de asimetría (−0,068 a 0,394) y curtosis (−0,504 a 0,168) se encuentran dentro de rangos aceptables, lo que sugiere que las distribuciones no presentan desviaciones severas de la normalidad. Mientras que la prueba de Kolmogórov-Smirnov (K-S) resultó estadísticamente significativa en todos los factores (p < 0,05), lo que indicaría una desviación de la normalidad en sentido estricto. Sin embargo, considerando que la prueba K-S es altamente sensible al tamaño de la muestra, y dado que los valores de asimetría y curtosis se mantienen dentro de límites aceptables, podríamos asumir una normalidad aproximada de los datos (Tabla 5).

Tabla 5

Consistencia interna y normalidad inferencial de los factores

Factores	Ítems	CR	AVE	Asimetría	Curtosis	K-S	*p*
Utilidad percibida	6	0,835	0,516	-0,068	-0,367	0,063	0,006
Influencia social	4	0,739	0,501	0,124	-0,043	0,107	<0,001
Intención conductual	5	0,855	0,541	0,013	-0,504	0,078	<0,001
Facilidad de uso	5	0,784	0,509	0,215	0,168	0,098	<0,001
Riesgo percibido	6	0,867	0,522	0,394	-0,479	0,094	<0,001

Nota. K-S = prueba de Kolmogórov-Smirnov; p = valor de significancia.

Los resultados obtenidos en el AFC muestran que el modelo de medida compuesto por cinco factores presenta un nivel de ajuste adecuado, lo cual respalda empíricamente la estructura factorial propuesta para evaluar la adopción educativa de ChatGPT en la muestra analizada.

DISCUSIÓN Y CONCLUSIÓN

La investigación realizada permitió evaluar las propiedades psicométricas de los constructos predominantes en la adopción educativa de ChatGPT en una muestra de estudiantes universitarios mexicanos. Los resultados confirmaron una estructura robusta de cinco factores (UP, IS, IC, FU, RP), consistente con marcos teóricos internacionales como el VAM, TAM, UTAUT y sus extensiones adaptadas específicamente al contexto de

la IA generativa (Morales et al., 2025). Si bien el instrumento no fue diseñado a partir de un modelo teórico específico, la convergencia empírica de los factores identificados en estos marcos consolidados refuerza su coherencia conceptual y respalda su pertinencia para el análisis de la adopción de ChatGPT en el contexto educativo.

En cuanto a la validación factorial, se puede mencionar que los valores obtenidos de la AVE que fluctúan entre 0,501 y 0,541, junto con el CR superior a 0,70, confirman la validez convergente de los factores y su capacidad para explicar adecuadamente la varianza de sus indicadores, lo que respalda la solidez del modelo en la muestra mexicana.

En relación con el tiempo que llevan los estudiantes usando ChatGPT, se obtuvo una distribución equilibrada entre los que tienen menos de un año de experiencia y quienes lo han utilizado por más tiempo, sin embargo, respecto al nivel de competencia, predomina el dominio básico, seguido del intermedio, mientras que solo una minoría reporta un manejo avanzado y aunque ChatGPT es la herramienta que todos los participantes utilizan, no es exclusiva, ya que complementan su uso con otras IA generativas.

En lo que respecta a los patrones de uso, los universitarios reportaron que ChatGPT se emplea principalmente como apoyo para resumir fuentes, explicar conceptos complejos y aportar ejemplos, así como tendencias actuales (76,86 %). Este hallazgo coincide con lo reportado por Rizun et al. (2026), quienes identificaron la recuperación de información y el resumen de textos como prácticas predominantes en estudiantes europeos.

En este mismo sentido, Panggabean y Silalahi (2025) encontraron que la asistencia académica entendida como el apoyo en tareas, la comprensión de contenidos, la elaboración de trabajos y la resolución de dudas representan el uso principal de ChatGPT en estudiantes de Taiwán e Indonesia, aunque se manifiesta con distintos niveles de intensidad. En relación con ello, Chen et al. (2024) mencionan que, en el ámbito de la investigación, la influencia de ChatGPT está creciendo debido a que es capaz de apoyar a los estudiantes en diferentes tareas, tales como la redacción, la revisión y el resumen de la literatura.

De manera general, estos antecedentes muestran que el uso de ChatGPT entre estudiantes universitarios tiende a concentrarse en utilidades prácticas inmediatas, especialmente las que están vinculadas con la comprensión, síntesis y producción de información, lo que refuerza la relevancia de la UP en su adopción.

Por el contrario, el uso de ChatGPT para llevar a cabo la revisión crítica de textos como la detección de redundancia, incoherencias, falta de cohesión o errores conceptuales, entre otros, es considerablemente baja (22,28 %). Esta misma situación ha sido mencionada en otras investigaciones recientes, en donde se señala que el uso de ChatGPT puede fomentar una dependencia excesiva y afectar procesos cognitivos como el pensamiento crítico (Abdi et al., 2025), además de favorecer que los estudiantes deleguen en la herramienta la elaboración de sus tareas académicas mediante la generación automática de textos (Lo, 2023).

En cuanto a los factores identificados empíricamente, la validación de los constructos UP y FU en el contexto mexicano permite reafirman su vigencia dentro de los modelos TAM, UTAUT y sus extensiones; con ello, se muestra su capacidad para explicar la adopción de ChatGPT en espacios educativos (Abdullatif, 2024; Acosta-Enríquez et al., 2024; Almogren et al., 2024; Güner et al., 2024; Lai et al., 2023; Sallam et al., 2023). En particular, la facilidad con la que el estudiante interactúa con la herramienta, sin requerir un esfuerzo significativo, se consolida como un elemento relevante para su adopción, lo cual coincide con los hallazgos reportados por Sobaih et al. (2024) en Arabia Saudita y por Almogren et al. (2024) en Malasia.

No obstante, lo anterior no se replica en todos los escenarios. Por ejemplo, Acosta-Enríquez et al. (2024), en Perú, encontraron en su investigación que la UP no influyó significativamente en la intención de uso frecuente. Esta diferencia muestra que, en ciertos contextos latinoamericanos, la adopción podría estar más asociada a factores como la curiosidad, la preferencia personal o el efecto de novedad, más que en la percepción de beneficios académicos reales que el estudiante puede tener al utilizar la herramienta.

De igual manera, el RP también se presenta como otro factor importante dentro de la estructura del modelo en la muestra mexicana, debido a que los estudiantes asocian el uso de ChatGPT con riesgos que van más allá de preocupaciones tradicionales como las relacionadas con la dependencia tecnológica y la afectación de la autonomía en el aprendizaje.

Entre ellas se identifican inquietudes vinculadas con la pérdida de habilidades de pensamiento crítico, la sustitución de fuentes confiables, el deterioro de habilidades blandas y una posible deshumanización del proceso educativo. Estos resultados muestran coincidencia con la tipología guardianes de la tecnología ética (*ethical tech guardians*) identificada por Espartinez (2024) en el contexto filipino, en el que tanto estudiantes como docentes manifestaron una preocupación persistente por el uso ético de la IA generativa, particularmente en relación con la integridad académica, el plagio y la sobredependencia tecnológica.

Este panorama se alinea con la paradoja beneficios-riesgos planteada por Panggabean y Silalahi (2025), en la que el valor funcional de la herramienta coexiste con percepciones de riesgo que pueden limitar su uso. En esa línea, Balaskas et al. (2025) consideran que el RP actúa como un mediador crítico en la adopción de la IA generativa; por ello, si las instituciones no gestionan adecuadamente estas preocupaciones, especialmente aquellas relacionadas con el desarrollo cognitivo y formativo, la adopción de ChatGPT efectiva podría verse comprometida.

Por otra parte, la estructura factorial obtenida sustenta que la IS entendida como la influencia del entorno social es un constructo relevante para los estudiantes mexicanos.

Los ítems incorporados indican que el uso de ChatGPT está influido por la observación del entorno, las opiniones, las recomendaciones y las experiencias compartidas dentro de la comunidad educativa. Esta información coincide con la investigación de Sobaih (2024), realizada en Arabia Saudita, en la que la IS es uno de los principales predictores de la intención de uso. En contraste, Almogren et al. (2024) indican que la influencia que ejerce el entorno social no muestra un impacto significativo en la aceptación de ChatGPT.

IMPLICACIONES

El desarrollo de esta investigación aporta un instrumento válido y confiable que puede ser utilizado tanto en la investigación como en la práctica docente con la finalidad de medir la adopción de la IA generativa en estudiantes universitarios de diversos entornos educativos. Como señalan Rizun et al. (2026) y Kooli (2023), la prohibición del uso de herramientas como ChatGPT no es una solución viable. Es mejor comprender la percepción de utilidad y los riesgos asociados a su uso por parte de los estudiantes, a fin de contar con elementos que permitan diseñar políticas de integridad académica efectivas, orientadas a fomentar un uso responsable, crítico y ético.

CONTRIBUCIÓN

Este estudio amplía la evidencia empírica sobre la validez y estructura de los constructos que explican la adopción educativa de ChatGPT en un contexto latinoamericano, en el que aún existe una limitada disponibilidad de instrumentos validados. Asimismo, los hallazgos confirman la vigencia de los modelos consolidados como TAM, VAM, UTAUT y sus extensiones en el análisis de tecnologías emergentes como la IA generativa, al tiempo que demuestran la importancia de incluir constructos como la IS y el RP.

LIMITACIONES

El uso de un muestreo no probabilístico utilizado limita la generalización de los resultados a otros contextos educativos y la naturaleza transversal del estudio impide establecer relaciones causales entre los constructos analizados.

TRABAJOS A FUTURO

Es recomendable que en investigaciones futuras se amplíe el alcance a otros niveles educativos y contextos culturales mexicanos. Así como se sugiere integrar enfoques longitudinales que permitan comprender cómo evoluciona la dinámica de la adopción de la IA generativa en el tiempo.

REFERENCIAS

Abdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., Mohammed, A. H., Hassan, B. A., Wayyes, A. M., Farhan, S. S., El Khatib, S., Rahal, M., Sahban,

A., Abdelaziz, D., Mansour, N., AlZayer, R., Khalil, R., Fekih-Romdhane, F., Hallit, R., … Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports*, *14*(1), 1983. https://doi.org/10.1038/s41598-024-52549-8

Abdalla, R. A. M. (2024). Examining awareness, social influence, and perceived enjoyment in the TAM framework as determinants of ChatGPT. Personalization as a moderator. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(3), 100327. https://doi.org/10.1016/j.joitmc.2024.100327

Abdi, A.-N. M., Omar, A. M., Ahmed, M. H., & Ahmed, A. A. (2025). The predictors of behavioral intention to use ChatGPT for academic purposes: Evidence from higher education in Somalia. *Cogent Education*, *12*(1) 2460250. https://doi.org/10.1080/2331186X.2025.2460250

Abdullatif, A. M. (2024). Investigating influencing factors of learning satisfaction in AI ChatGPT for research: University students perspective. *Heliyon*, *10*(11), e32220. https://doi.org/10.1016/j.heliyon.2024.e32220

Acosta-Enríquez, B. G., Arbulú, M. A., Huamaní, O., López, C., & Saavedra, K. (2024). Analysis of college students' attitudes toward the use of ChatGPT in their academic activities: Effect of intent to use, verification of information and responsible use. *BMC Psychology*, *12*, 255. https://doi.org/10.1186/s40359-024-01764-z

Al-Abdullatif, A. M., & Alsubaie, M. A. (2024). ChatGPT in learning: Assessing students' use intentions through the lens of perceived value and the influence of AI literacy. *Behavioral Sciences*, *14*(9), 845. https://doi.org/10.3390/bs14090845

Almogren, A. S., Al-Rahmi, W. M., & Dahri, N. A. (2024). Exploring factors influencing the acceptance of ChatGPT in higher education: A smart education perspective. *Heliyon*, *10*(11), e31887. https://doi.org/10.1016/j.heliyon.2024.e31887

Balaskas, S., Tsiantos, V., Chatzifotiou, S., & Rigou, M. (2025). Determinants of ChatGPT adoption intention in higher education: Expanding on TAM with the mediating roles of trust and risk. *Information*, *16*(2), 82. https://doi.org/10.3390/info16020082

Bouteraa, M., Bin-Nashwan, A. S., Al-Daihani, M., Dirie, A. K., Benlahcene, A., Sadallah, M., Zaki, H. O., Lada, S., Ansar, R., Fook, L. M., & Chekima, B. (2024). Understanding the diffusion of AI-generative ChatGPT in higher education: Does students' integrity matter? *Computers in Human Behavior Reports*, *14*, 100402. https://doi.org/10.1016/j.chbr.2024.100402

Chen, S.-Y., Kuo, H. Y., & Chang, S.-H. (2024). Perceptions of ChatGPT in healthcare: Usefulness, trust, and risk. *Frontiers Public Health*, *12*, 1457131. https://doi.org/10.3389/fpubh.2024.1457131

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*, 7. https://doi.org/10.7275/jyj1-4868

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Dahri, N. A., Yahaya, N., Al-Rahmi, W. M., Aldraiweesh, A., Alturki, U., Almutairy, S., Shutaleva, A., & Soomro, R. B. (2024). Extended TAM based acceptance of AI-Powered ChatGPT for supporting metacognitive self-regulated learning in education: A mixed-methods study. *Heliyon*, *10*(8), e29317. https://doi.org/10.1016/j.heliyon.2024.e29317

Elshaer, I. A., Hasanein, A. M., & Sobaih, A. E. (2024). The moderating effects of gender and study discipline in the relationship between university students' acceptance and use of ChatGPT. *European Journal of Investigation in Health, Psychology and Education*, *14*(7), 1981-1995. https://doi.org/10.3390/ejihpe14070132

Espartinez, A. S. (2024). Exploring student and teacher perceptions of ChatGPT use in higher education: A Q-Methodology study. *Computers and Education: Artificial Intelligence*, *7*, 100264. https://doi.org/10.1016/j.caeai.2024.100264

García-Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2024). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *Revista Iberoamericana de Educación a Distancia*, *27*(1), 9-39. https://doi.org/10.5944/ried.27.1.37716

Güner, H., Er, E., Akçapinar, G., & Khalil, M. (2024). From chalkboards to AI-powered learning: Students' attitudes and perspectives on use of ChatGPT in educational settings. *Educational Technology & Society*, *27*(2), 386-404. https://www.jstor.org/stable/48766180

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53-60. https://doi.org/10.21427/D7CF7R

Hsu, W. L., & Silalahi, A. D. K. (2024). Exploring the paradoxical use of ChatGPT in education: Analyzing benefits, risks, and coping strategies through integrated UTAUT and PMT theories using a hybrid approach of SEM and fsQCA. *Computers and Education: Artificial Intelligence*, *7*, 100329. https://doi.org/10.1016/j.caeai.2024.100329

Jordan, F. M. (2021). Valor de corte de los índices de ajuste en el análisis factorial confirmatorio. *Psocial*, *7*(1). https://www.redalyc.org/journal/6723/672371335005/672371335005.pdf

Kooli, C. (2023). Chatbots in education and research: A critical examination of ethical implications and solutions. *Sustainability*, *15*(7), 5614. https://doi.org/10.3390/su15075614

Lai, C. Y., Cheung, K. Y., & Chan, C. S. (2023). Exploring the role of intrinsic motivation in ChatGPT adoption to support active learning: An extension of the technology acceptance model. *Computers and Education: Artificial Intelligence*, *5*, 100178. https://doi.org/10.1016/j.caeai.2023.100178

Lai, K. (2021). Fit difference between nonnested models given categorical data: Measures and estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(1), 99-120. https://doi.org/10.1080/10705511.2020.1763802

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, *13*(4), 410. https://doi.org/10.3390/educsci13040410

Mahmud, A., Sarower, A. H., Sohel, A., Assaduzzaman, M., & Bhuiyan, T. (2024). Adoption of ChatGPT by university students for academic purposes: Partial least square, artificial neural network, deep neural network and classification algorithms approach. *Array*, *21*, 100339. https://doi.org/10.1016/j.array.2024.100339

Morales, A., Figueroa, G., Duana, M., Hidalgo, C., & Amador, J. (2025). Acceptance and adoption of ChatGPT in higher education: a systematic review. *Revista Digital de Investigación en Docencia Universitaria*, *19*(2), e2119. https://doi.org/10.19083/ridu.2025.2119

Nemt-allah, M., Khalifa, W., Badawy, M., Elbably, Y., & Ibrahim, A. (2024). Validating the ChatGPT usage scale: Psychometric properties and factor structures among postgraduate students. *BMC Psychology*, *12*, 497. https://doi.org/10.1186/s40359-024-01983-4

Panggabean, E. M., & Silalahi, A. D. K. (2025). How do ChatGPT's benefit-risk-coping paradoxes impact higher education in Taiwan and Indonesia? *Computers and Education: Artificial Intelligence*, *8*, 100412. https://doi.org/10.1016/j.caeai.2025.100412

Parveen, K., Phuc, T. Q. B., Alghamdi, A. A., Hajjej, F., Obidallah, W. J., Alduraywish, Y. A., & Shafiq, M. (2024). Unraveling the dynamics of ChatGPT adoption and utilization through structural equation modeling. *Scientific Reports*, *14*, 23469. https://doi.org/10.1038/s41598-024-74406-4

Rana, M., Siddiqee, M. S., Sakib, N., & Ahamed, R. (2024). Assessing AI adoption in developing country academia: A trust and privacy-augmented UTAUT framework. *Heliyon*, *10*(18), e37569. https://doi.org/10.1016/j.heliyon.2024.e37569

Rizun, N., Bordean, O. N., Nikiforova, A., Beleiu, I. N., & Revina, A. (2026). Generative AI in higher education: Ethical and behavioral factors influencing students' intentions

to use ChatGPT. *Computers and Education Open*, *10*, 100336. https://doi.org/10.1016/j.caeo.2026.100336

Romero-Rodríguez, J.-M., Ramírez-Montoya, M.-S., Buenestado-Fernández, M., & Lara-Lara, F. (2023). Use of ChatGPT at university as a tool for complex thinking: Students' perceived usefulness. *Journal of New Approaches in Educational Research*, *12*(2), 323-339. https://doi.org/10.7821/naer.2023.7.1458

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching*, *6*, 342-363. https://doi.org/10.37074/jalt.2023.6.1.9

Sallam, M., Salim, N. A., Barakat, M., Al-Mahzoum, K., Al-Tammemi, A. B., Malaeb, D., Hallit, R., & Hallit, S. (2023). Assessing health students' attitudes and usage of ChatGPT in Jordan: Validation study. *JMIR Medical Education*, *9*, e48254. https://doi.org/10.2196/48254

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2.ª ed.). Lawrence Erlbaum Associates Publishers.

Shuhaiber, A., Kuhail, M. A., & Salman, S. (2025). ChatGPT in higher education - A student's perspective. *Computers in Human Behavior Reports*, *17*, 100565. https://doi.org/10.1016/j.chbr.2024.100565

Sobaih, A. E., Elshaer, I. A., & Hasanein, A. M. (2024). Examining students' acceptance and use of ChatGPT in Saudi Arabian higher education. *European Journal of Investigation in Health, Psychology and Education*, *14*(3), 709-721. https://doi.org/10.3390/ejihpe14030047

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

United Nations Educational, Scientific and Cultural Organization. (2023). *Guidance for generative AI in education and research*. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need* [Presentación de escrito]. 31st Conference on Neural Information Processing Systems, Long Beach, California, Estados Unidos. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*(2), 231-251. https://psycnet.apa.org/doi/10.1037/1082-989X.3.2.231

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425-478. https://doi.org/10.2307/30036540

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806-838. https://doi.org/10.1177/0011000006288127

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Method*, *51*, 409-428. https://doi.org/10.3758/s13428-018-1055-2

Xu, X., Su, Y., Zhang, Y., Wu, Y., & Xu, X. (2024). Understanding learners' perceptions of ChatGPT: A thematic analysis of peer interviews among undergraduates and postgraduates in China. *Heliyon*, *10*(4), e26239. https://doi.org/10.1016/j.heliyon.2024.e26239

ANEXO A. Instrumento

Ítems completos del instrumento

ID	Ítem	Fuente
UP1	Usar ChatGPT me permite ser más productivo(a) y eficiente en las tareas académicas, al brindarme información precisa y relevante.	Abdalla (2024), Abdullatif (2024), Almogren et al. (2024); Güner et al. (2024), Mahmud et al. (2024)
UP2	Usar ChatGPT para responder mis consultas académicas mejora mi eficacia.	Lai et al. (2023), Güner et al. (2024), Mahmud et al. (2024)
UP3	Usar ChatGPT mejora mi experiencia de aprendizaje.	Abdalla (2024), Almogren et al. (2024); Dahri et al. (2024)
UP4	ChatGPT es más útil para responder consultas en comparación con otras fuentes de información que he usado con anteriormente.	Sallam et al. (2023)
UP5	ChatGPT puede brindar respuestas de alta calidad a mis consultas, problemas e investigaciones.	Abdalla (2024), Abdullatif (2024)
UP6	ChatGPT me ayuda a comprender mejor los materiales de mis cursos.	Almogren et al. (2024)
UP7	Considero que ChatGPT es una herramienta beneficiosa para mi desarrollo educativo.	Dahri et al. (2024)
UP8	Usar ChatGPT mejora la probabilidad de alcanzar objetivos importantes en las actividades académicas.	Elshaer et al. (2024), Sobaih et al. (2024)
UP9	ChatGPT es una herramienta valiosa para mis actividades académicas.	Sobaih et al. (2024)
UP10	ChatGPT es una herramienta útil para resolver problemas.	Elshaer et al. (2024)
IS1	Las personas que son importantes para mí creen que debería utilizar ChatGPT en mis cursos.	Güner et al. (2024)
IS2	Es más probable el uso de ChatGPT cuando observo que personas de mí alrededor lo utilizan.	Abdalla (2024)
IS3	Las experiencias y comentarios de otros miembros de mi comunidad educativa han influido con respecto al uso de ChatGPT.	Dahri et al. (2024)
IS4	La recomendación de expertos en educación (investigadores, docentes o instructores,) ha influido en mi decisión de utilizar ChatGPT para fines académicos.	Dahri et al. (2024)
IS5	La opinión de mis compañeros ha influido en mi decisión de usar ChatGPT para fines académicos.	Dahri et al. (2024)
IS6	Uso ChatGPT con fines académicos, ya que otras personas de mi círculo social quieren que lo use.	Mahmud et al. (2024)
IS7	ChatGPT me ayuda a destacar entre mis compañeros.	Parveen et al. (2024)
IC1	Tengo la intención de seguir utilizando ChatGPT en el futuro.	Abdalla (2024), Elshaer et al. (2024), Parveen et al. (2024), Sobaih et al. (2024)

(*continúa*)

(*continuación*)

ID	Ítem	Fuente
IC2	Estoy decidido a utilizar ChatGPT de manera constante en mis estudios.	Elshaer et al. (2024); Parveen et al. (2024)
IC3	Me dedico a utilizar ChatGPT como herramienta para mis estudios.	Sobaih et al. (2024)
IC4	Suponiendo que tengo acceso a ChatGPT, tengo la intención de usarlo.	Güner et al. (2024)
IC5	Es probable que recomiende el uso de ChatGPT a otras personas con fines educativos.	Almogren et al. (2024)
FU1	Mi interacción con ChatGPT es clara y comprensible.	Bouteraa et al. (2024), Rana et al. (2024)
FU2	Me resulta fácil dominar el uso de ChatGPT.	Abdalla (2024), Bouteraa et al. (2024), Rana et al. (2024)
FU3	ChatGPT es una herramienta fácil de usar.	Abdalla (2024), Bouteraa et al. (2024), Rana et al., (2024)
FU4	Aprender a utilizar nuevas funcionalidades de ChatGPT a resultado fácil.	Elaboración propia
FU5	ChatGPT es una herramienta con una interfaz intuitiva y fácil de usar.	Elshaer et al. (2024), Parveen et al. (2024), Sobaih et al. (2024)
FU6	El uso de ChatGPT permite ahorrar tiempo.	Parveen et al. (2024)
FU7	Utilizar ChatGPT no requiere esfuerzo.	Elshaer et al. (2024)
FU8	Obtener experiencia en el uso de ChatGPT ha resultado muy sencillo.	Sobaih et al. (2024)
RP1	Me preocupa que el uso de ChatGPT pueda ser malinterpretado como una falta académica (plagio o trampa).	Sallam et al. (2023)
RP2	Adoptar ChatGPT puede reducir mi esfuerzo mental y la capacidad de pensamiento crítico.	Sobaih (2024)
RP3	Tengo miedo de volverme demasiado dependiente de ChatGPT.	Sallam et al. (2023)
RP4	Me preocupan los posibles riesgos de seguridad que supone utilizar ChatGPT (por ejemplo, el manejo inadecuado de datos personales).	Sallam et al. (2023)
RP5	Tengo dudas sobre la validez ética y la forma correcta de citar las respuestas generadas por ChatGPT.	Elaboración propia
RP6	Me causa duda la confiabilidad y exactitud de la información que proporciona ChatGPT.	Elaboración propia
RP7	Temo que el uso frecuente de ChatGPT sustituya el uso de otras fuentes confiables de información.	Elaboración propia
RP8	Me preocupa no saber cómo se almacenan, usan y protegen los datos que proporciono al interactuar con ChatGPT.	Elaboración propia

(*continúa*)

(*continuación*)

ID	Ítem	Fuente
RP9	Me preocupa que mi rendimiento y logros académicos se vean afectados por el uso de ChatGPT	Hsu (2024)
RP10	Me preocupa que el uso de ChatGPT obstaculice mis habilidades blandas.	Hsu (2024)
RP11	Usar ChatGPT implica riesgos técnicos (por ejemplo, inaccesibilidad de la herramienta).	Shuhaiber et al. (2025)
RP12	Me preocupa que ChatGPT tenga la posibilidad de incurrir en faltas legales.	Panggabean & Silalahi (2025)
RP13	Me preocupa que ChatGPT otorgue respuestas con información sesgada.	Panggabean & Silalahi (2025)
RP14	Me preocupa que ChatGPT me entregue las mismas respuestas que a mis compañeros.	Elaboración propia
RP15	Me preocupa que el uso de ChatGPT pueda generar errores en la documentación científica, lo que podría afectar la precisión de la información que utilizo para mis tareas.	Elaboración propia
RP16	El uso excesivo de ChatGPT podría llevar a una deshumanización del aprendizaje donde las respuestas de los estudiantes parezcan estandarizadas.	Sobaih (2024)

ARQUITECTURAS SUBAGÉNTICAS DE IA: HACIA LA ESPECIALIZACIÓN Y LA ORQUESTACIÓN JERÁRQUICA

Olda Bustillos Ortega
https://orcid.org/0000-0003-2822-3428
obustillos@uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Jorge Murillo Gamboa
https://orcid.org/0000-0001-5548-8283
jmurillo@uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Olman Núñez Peralta
https://orcid.org/0000-0001-6780-022X
onunez@edu.uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Fabián Rodríguez Sibaja
https://orcid.org/0009-0008-3276-9865
frodriguezsi@edu.uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Daniel Mena Bocker
https://orcid.org/0009-0002-1062-6675
dfmenab@edu.uia.ac.cr
Escuela de Ingeniería Informática,
Universidad Internacional de las Américas, Costa Rica

Recibido: 27 de marzo del 2026 / Aceptado: 18 de abril del 2026
doi: https://doi.org/10.26439/interfases2026.n023.8714

RESUMEN. Durante el segundo semestre del 2025, la investigación en inteligencia artificial (IA) ha evidenciado una transición hacia arquitecturas multiagente, en las cuales los subagentes adquieren un rol estructural central. A diferencia de los enfoques monolíticos, estas arquitecturas permiten descomponer tareas complejas en unidades funcionales

especializadas, coordinadas por un agente orquestador, lo que favorece la escalabilidad, el paralelismo y una gestión más eficiente del contexto. Se definen los sistemas multiagente, agentes web y subagentes. Se analiza el marco conceptual *multi-agent data mining* (MADM), la arquitectura del árbol de tareas complejas y la metáfora de la cocina. Se identifican beneficios en términos de eficiencia, resiliencia y adaptabilidad, tanto en entornos científicos como en plataformas comerciales. No obstante, persisten desafíos asociados a la orquestación, la seguridad, los costos computacionales, formación académica y la gobernanza ética. En este contexto, se examina el Índice Latinoamericano de Inteligencia Artificial (ILIA 2025) como guía para integrar arquitecturas subagénticas a modelos educativos con enfoque ético, interdisciplinario y sostenible, orientadas a impulsar una innovación regional más responsable e inclusiva.

PALABRAS CLAVE: agentes / arquitecturas / ILIA / inteligencia artificial / subagentes

SUB-AGENTIC AI ARCHITECTURE: A CONCEPTUAL REVIEW OF SPECIALIZATION AND HIERARCHICAL ORCHESTRATION

ABSTRACT. During the second half of 2025, artificial intelligence (AI) research has shown a shift towards multi-agent architectures, in which sub-agents acquire a central structural role. Unlike monolithic approaches, these architectures allow complex tasks to be broken down into specialized functional units, coordinated by an orchestrating agent, which promotes scalability, parallelism, and more efficient context management. This article defines multi-agent architectures, web agents, and sub-agents. The conceptual framework Multi-Agent Data Mining (MADM), the architecture of the complex task tree, and the kitchen metaphor are analyzed. Benefits in terms of efficiency, resilience, and adaptability are identified, both in scientific environments and on commercial platforms. However, challenges remain related to orchestration, security, computational costs, academic training, and ethical governance. In this context, the Latin American Artificial Intelligence Index (ILIA 2025) is examined as a guide for integrating sub-agentic architectures into educational models with an ethical, interdisciplinary, and sustainable focus, aimed at fostering more responsible and inclusive regional innovation.

KEYWORDS: agents / architecture / artificial inteligence / ILIA / subagents

INTRODUCCIÓN

Multiagentes

En el ámbito de las ciencias de la computación, los sistemas multiagente (*multi-agent data mining*, MAS) constituyen una subárea de la inteligencia artificial (IA) cuyo propósito es resolver problemas que superan las capacidades de un solo agente. Estos sistemas se han consolidado como uno de los campos más relevantes de investigación en IA, ya que permiten abordar tareas complejas mediante la cooperación, la coordinación o, incluso, la competencia entre agentes, lo que permite una mayor eficacia y adaptabilidad del sistema en su conjunto (Elmahalawy, 2015; Gonçalves et al., 2015).

De acuerdo con Carrascosa et al. (2005), un sistema MAS puede adoptar distintos comportamientos o asumir diferentes roles. En el ejemplo de un entorno típico de comercio electrónico, un mismo agente puede actuar tanto como comprador o como vendedor, alternando entre estos dos comportamientos en función del tipo de transacción, incluso cuando sus objetivos y tareas resulten opuestos o potencialmente contradictorios.

Una de las principales razones de este enfoque radica en la creciente necesidad de automatizar procesos complejos, que abarcan desde la investigación científica y el análisis de grandes volúmenes de datos hasta la atención al cliente. Plataformas recientes evidencian que la coordinación jerárquica de agentes, mediante orquestadores que supervisan conjuntos de subagentes especializados, permite superar limitaciones asociadas al manejo del contexto, la latencia y la eficiencia cognitiva de los modelos.

De acuerdo con Rother et al. (2025), y respecto del entrenamiento de estos multiagentes, surgió el aprendizaje por refuerzo multiagente (*multi-agent reinforcement learning*, MARL), aplicado en entornos con agentes previamente definidos y a través de tareas principalmente cooperativas. El aprendizaje se apoyó inicialmente en esquemas de entrenamiento o control centralizado, aunque con un alcance limitado. Posteriormente, se desarrollaron enfoques descentralizados que permitieron a cada agente aprender su propia política, como el Independent Q-Learning. No obstante, la mayoría de estas técnicas asumen conjuntos de agentes con tareas relativamente fijas. En contraste, enfoques más recientes plantean la incorporación o eliminación dinámica de agentes y la adaptación a tareas cambiantes, lo que amplía el alcance del MARL.

WebAgents o agentes web

A medida que la World Wide Web (WWW), o red informática mundial, evoluciona, el acceso a la información y a los servicios digitales se transforma de forma sustancial. Actividades cotidianas como realizar compras y pagos en línea, registrar nuevas cuentas o completar formularios requieren la introducción reiterada de datos personales, lo que incrementa la carga operativa de los usuarios. En respuesta a este escenario, han surgido

los WebAgents como sistemas basados en técnicas de IA diseñados para automatizar tareas rutinarias mediante la emulación de procesos cognitivos humanos, con el objetivo de optimizar la productividad y mejorar la calidad de vida.

De acuerdo con Ning et al. (2025), la interacción entre el usuario y un WebAgent puede realizarse mediante una única instrucción en lenguaje natural, por ejemplo, solicitar una reunión: "Programa una reunión con León en Starbucks el día 23 de noviembre de 2025 a las 4:00 p. m. mediante correo electrónico". Estos agentes son capaces de ejecutar de manera autónoma flujos completos de trabajo, desde obtener la dirección electrónica de León, redactar el mensaje, incluir la reunión en la agenda y, finalmente, enviarlo por correo. El WebAgent lleva a cabo la interacción con aplicaciones hasta la ejecución de acciones finales, lo que incrementa significativamente la eficiencia y la comodidad del usuario. No obstante, su adopción plantea desafíos asociados a la transparencia, la propagación de sesgos, la protección de la privacidad y el desempeño frente a situaciones imprevistas.

De acuerdo con Chevrot et al. (2025), los agentes autónomos web (*autonomous web agents*, AWA) constituyen una alternativa prometedora para reducir la fragilidad y el alto costo de mantenimiento de la automatización de tareas tradicionales. Estos agentes interactúan con aplicaciones web mediante ciclos iterativos de observación, análisis y acción guiados por descripciones de tareas en lenguaje natural, sin requerir conocimiento previo del sistema.

Mediante la construcción y el ajuste dinámico de *prompts*, los AWA son capaces de interpretar los escenarios de prueba como tareas autónomas, ejecutar las acciones correspondientes y aplicar modificaciones de manera adaptativa. Esta capacidad de interpretación contribuye a fortalecer la robustez y la flexibilidad de los procesos al reducir la dependencia de *scripts* estáticos y mitigar la vulnerabilidad de estos frente a cambios en la interfaz o en la lógica de las aplicaciones.

Subagentes de IA

Las arquitecturas subagénticas de IA pueden entenderse como aquellos sistemas distribuidos compuestos por agentes especializados que cooperan tanto en los procesos de toma de decisiones como en la ejecución de tareas y que están articulados en torno a modelos de lenguaje LLM (*large language models*). Estas arquitecturas se fundamentan en la delegación estructurada de tareas desde un agente orquestador hacia subagentes especializados, los cuales operan de manera autónoma y colaborativa para cumplir objetivos específicos dentro del sistema distribuido.

De acuerdo con Claude Code Docs (2025), los subagentes posibilitan una resolución de problemas más eficiente al operar con configuraciones de tareas específicas que incluyen indicaciones de sistema personalizadas, herramientas dedicadas y ventanas de

contexto independientes. En plataformas de programación como Claude Code[1], estos subagentes se asemejan a asistentes de IA especializados que pueden ser invocados para gestionar tipos concretos de tareas.

A modo de ejemplo, cuando el orquestador de Claude Code identifica una tarea alineada con la especialización de un subagente, procede a delegarla a dicho subagente, el cual ejecuta de manera autónoma las acciones necesarias y retorna los resultados correspondientes. Entonces, una vez preconfigurados, los subagentes reciben encargos definidos y ejecutan de manera autónoma las acciones correspondientes dentro de su ámbito de especialización:

- Tienen un propósito específico y un área de experiencia funcional.
- Utilizan su propia ventana de contexto separada de la conversación principal.
- Pueden configurarse para que utilice herramientas específicas.
- Incluyen un indicador personalizado que guía su comportamiento.

Este enfoque de subagentes de IA plantea una mejora en el rendimiento técnico y favorece la modularidad, escalabilidad y adaptabilidad de los sistemas, lo que posiciona a los agentes y subagentes como un recurso estratégico para el desarrollo productivo. Sin embargo, su impacto efectivo depende de la adecuada gestión de desafíos asociados a la orquestación, la seguridad y la gobernanza ética.

Para organizar e integrar las funciones de subagentes de IA, Zhu et al. (2025) proponen tres categorías: identificadores de agentes, monitoreo en tiempo real y registro de actividades. Estas están destinadas a incrementar la visibilidad operativa de los sistemas de IA, medidas que pueden interpretarse como funciones especializadas dentro de una arquitectura subagéntica.

En una arquitectura jerárquica, las funciones de identificación, monitoreo y registro pueden delegarse a subagentes especializados, lo que permite una supervisión coordinada y una rendición de cuentas efectiva. Esta distribución de funciones permite coordinar de manera coherente la supervisión, el control y la rendición de cuentas del sistema, lo que refuerza la tesis de que las arquitecturas subagénticas, al asignar responsabilidades específicas a módulos coordinados jerárquicamente, ofrecen un marco robusto para la transparencia, la seguridad y la gestión fiable del comportamiento de agentes autónomos.

[1] Claude Code es un entorno especializado de programación y asistencia al desarrollo creado por Anthropic con IA enfocada en tareas de codificación, depuración, análisis de archivos y ejecución controlada de código, similar a un espacio de trabajo optimizado para programadores. Para visitar la página web, puede ingresar aquí: https://website.claude.com/product/overview

Aplicaciones con subagentes de IA

- Movilidad y transporte. De acuerdo con Samdani et al. (2025), la aplicación de subagentes de IA en movilidad y transporte, especialmente en vehículos autónomos, ha transformado la interacción entre sistemas inteligentes, infraestructura digital y seres humanos, con el objetivo principal de reducir el error humano, una de las causas predominantes de accidentes viales. En este contexto, los subagentes permiten la especialización funcional en tareas como percepción del entorno, razonamiento, planificación y control, integrando enfoques basados en reglas y aprendizaje automático dentro de arquitecturas distribuidas.
- Inversiones financieras. De acuerdo con Babaei y Giudici (2025), las arquitecturas subagénticas permiten mejorar el análisis de inversiones financieras mediante la coordinación de agentes especializados, lo que aumenta la trazabilidad, la explicabilidad y la precisión de las decisiones. En este enfoque jerárquico, el modelo principal de riesgo-retorno se complementa con módulos explicativos basados en valores. Así, la especialización funcional y la coordinación multinivel incrementan tanto la precisión como la transparencia, lo que consolida a las arquitecturas subagénticas como un marco relevante para la toma de decisiones financieras robustas y verificables.
- Sector de salud. De acuerdo con Samdani et al. (2025), en el ámbito de la salud, las arquitecturas subagénticas de IA pueden potenciar las capacidades del personal clínico y administrativo mediante la especialización de tareas clínicas y no clínicas. No obstante, debido a la sensibilidad del dominio sanitario, su adopción exige una estrecha colaboración interdisciplinaria y una gobernanza ética sólida que garantice un uso responsable de sistemas híbridos entre lo humano y la IA.
- Sector académico. En este sector, al introducir estas arquitecturas de multiagente, WebAgents y subagentes, los estudiantes de distintos niveles, desde la secundaria hasta la universitaria, tienen acceso a plataformas educativas adaptativas, lo que reduce las brechas de aprendizaje y potencia el desarrollo de competencias en áreas como Matemáticas, Programación o Ciencias Aplicadas. Los docentes pueden apoyarse en agentes y subagentes para automatizar sus tareas administrativas, tales como generar material didáctico individualizado, analizar múltiples referencias y recibir recomendaciones de parte de un agente de IA en torno a estrategias pedagógicas basadas en datos del desempeño estudiantil (Babaei & Giudici, 2025).

En síntesis, las arquitecturas subagénticas de IA representan un modelo evolutivo de inteligencia distribuida, capaz de potenciar la productividad y la sostenibilidad, siempre

que su desarrollo se acompañe de estrategias coordinadas en orquestación técnica, seguridad digital, financiamiento de la infraestructura y regulación ética.

En este artículo, se examina el marco conceptual del *multi-agent data mining framework* (MADM), el cual integra técnicas de minería de datos (*data mining*, DM) con sistemas multiagente (*multi-agent systems*, MAS), con el objetivo de automatizar, distribuir y potenciar el proceso de descubrimiento de conocimiento en grandes volúmenes de información.

Otra referencia utilizada para ilustrar la dinámica entre agentes y subagentes de IA es la metáfora de la brigada de cocina profesional. En este contexto, el *chef de cuisine* asume la responsabilidad de coordinar de manera integral las actividades de la brigada, por lo que desempeña un rol análogo al del agente orquestador dentro de una arquitectura de IA. Por su parte, Los *chefs de partie* representan a los agentes especializados encargados de áreas o tareas específicas, mientras que los roles de apoyo —como los *cuisiniers* y *commis*— pueden asociarse a subagentes que ejecutan funciones operativas de asistencia bajo un esquema de colaboración jerárquica.

El árbol de tareas complejas constituye otra referencia conceptual —la cual será analizada más adelante— que permite comprender el ciclo de búsqueda de información (*information retrieval*) orientado a la resolución de tareas de alta complejidad. Este enfoque modela dicho proceso mediante una estructura jerárquica en forma de árbol, sustentada en el uso coordinado de agentes y subagentes de IA.

Asimismo, se incluye la iniciativa del Índice Latinoamericano de Inteligencia Artificial 2025 (ILIA 2025) como un esfuerzo pionero orientado a medir de forma sistemática el avance de la inteligencia artificial en América Latina y el Caribe. La experiencia de los países de la región sugiere que la articulación entre talento humano, una institucionalidad sólida y la adopción responsable de la IA puede posicionar a América Latina como un entorno favorable para el desarrollo de sistemas de IA escalables, transparentes y orientados al bien común.

OBJETIVOS DE LA INVESTIGACIÓN

El objetivo principal fue analizar el desarrollo y la relevancia de las arquitecturas subagénticas de inteligencia artificial, destacando sus fundamentos conceptuales, aplicaciones emergentes, desafíos técnicos y consideraciones éticas, con el fin de comprender su impacto en la evolución de los sistemas agénticos y en la automatización de procesos complejos.

Por su parte, los objetivos específicos de la investigación son los siguientes:

- Definir el concepto de subagente y diferenciarlo de los modelos de agente único y de las arquitecturas multiagente tradicionales

- Examinar los principales enfoques arquitectónicos que sustentan a los subagentes, incluyendo jerarquía, paralelismo e interacción colaborativa
- Identificar aplicaciones prácticas de los subagentes en investigación científica, educación, industria y robótica
- Analizar los desafíos técnicos asociados a la coordinación, eficiencia computacional, seguridad y gobernanza de los sistemas subagénticos
- Explorar las perspectivas futuras de la investigación en subagentes y su potencial integración con otras tecnologías disruptivas

Asimismo, se buscó responder a la siguiente pregunta de investigación: ¿cómo contribuyen las arquitecturas subagénticas de IA a impulsar el desarrollo productivo, considerando los desafíos en la orquestación, la seguridad y la gobernanza ética?

METODOLOGÍA

La presente investigación se desarrolló mediante un enfoque de revisión sistemática de la literatura, orientado a analizar las arquitecturas subagénticas en IA y su impacto en diversos sectores de aplicación. El proceso metodológico se diseñó siguiendo lineamientos de revisiones en ingeniería y ciencias computacionales, a fin de garantizar transparencia y reproducibilidad.

La estrategia de búsqueda de información se realizó entre enero del 2020 y marzo del 2026, con el objetivo de capturar los avances más recientes en el área. Para ello, se consultaron bases de datos académicas indexadas de alto impacto, incluyendo ACM Digital Library, IEEE Xplore, Google Académico, ProQuest Dissertations and Theses y Academia.edu.

Las ecuaciones de búsqueda se construyeron combinando operadores booleanos y términos clave en español y en inglés:

- “sub-agent architectures” OR “multi-agent systems”
- “AI agents” AND “coordination”
- “hierarchical agents” OR “agent orchestration”
- “autonomous agents” AND “AI systems design”
- “distributed artificial intelligence” AND “applications”

Estas expresiones se adaptaron a la sintaxis específica de cada base de datos. De igual manera, como criterios de inclusión se establecieron los siguientes:

- Publicaciones entre 2020 y 2026
- Artículos en revistas indexadas (Q1-Q3) y congresos reconocidos en informática (por ejemplo, IEEE, ACM)

- Estudios que aborden fundamentos teóricos, arquitecturas, aplicaciones o implicaciones éticas de sistemas basados en subagentes
- Documentos en idioma inglés o español

Por su lado, los criterios de exclusión se definieron de la siguiente manera:

- Estudios centrados exclusivamente en herramientas específicas sin aporte arquitectónico
- Publicaciones no revisadas por pares (salvo informes institucionales relevantes)
- Trabajos enfocados únicamente en ChatGPT, lenguajes de programación o bases de datos sin relación directa con arquitecturas subagénticas

Como fuentes complementarias y con el fin de contextualizar tendencias industriales y aplicaciones reales, se incorporaron informes recientes de organizaciones y empresas tecnológicas, tales como McKinsey & Company, Deloitte, Gartner, World Economic Forum, IBM y Microsoft. Asimismo, se analizaron estudios de caso asociados a plataformas como Amazon Bedrock (AgentCore) y Claude, así como el documento estratégico ILIA 2025 y sus recomendaciones sobre IA para la región latinoamericana.

Los estudios seleccionados fueron analizados mediante un enfoque de síntesis cualitativa temática, clasificando los hallazgos en tres dimensiones principales: fundamentos y modelos arquitectónicos; aplicaciones y casos de uso; e implicaciones éticas, organizacionales y técnicas.

RESULTADOS

Marco conceptual *multi-agent data mining* (MADM)

Para comprender el funcionamiento de los sistemas multiagente de IA, se retoma el marco conceptual MADM propuesto por Chaimontree et al. (2010). Este integra técnicas de minería de datos (*data mining*, DM) con sistemas multiagente (*multi-agent systems*, MAS). Este enfoque tiene como objetivo automatizar, distribuir y potenciar de manera inteligente el proceso de descubrimiento de conocimiento en grandes volúmenes de datos.

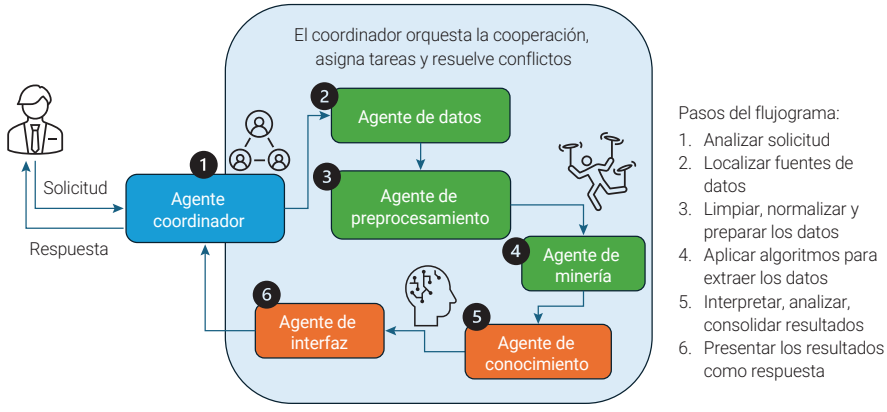
En este entorno, un conjunto de agentes coopera para abordar tareas de minería de datos específicas, con el fin de mejorar el proceso de acceso a los datos mediante agentes inteligentes cooperativos, los cuales pueden ejecutar de manera autónoma, distribuida y coordinada las distintas etapas del proceso, el cual se encuentra subdividido en tres partes:

- Preprocesamiento (limpieza, integración y selección de datos)
- Minería propiamente dicha (aplicación de algoritmos de aprendizaje o *clustering*)
- Posprocesamiento (evaluación e interpretación de los resultados)

Para facilitar la comprensión de su funcionamiento, en la Figura 1 se muestra un flujograma que ilustra la interacción entre el usuario y el sistema: el usuario formula una solicitud al agente, el orquestador distribuye y coordina las tareas entre los agentes especializados, y, finalmente, se integra y devuelve la respuesta correspondiente.

Figura 1

Flujograma de interacción de multiagentes de IA basado en modelo MADM



Nota. Los datos proceden de "Multi-Agent Based Clustering: Towards Generic Multi-Agent Data Mining", por S. Chaimontree, K. Atkinson y F. Coenen, 2010, en P. Perner (Ed.), *Advances in Data Mining: Applications and Theoretical Aspects*, Springer (https://doi.org/10.1007/978-3-642-14400-4_9).

El análisis de este flujograma de seis pasos es el siguiente:

1. El agente coordinador analiza la solicitud, orquesta la cooperación entre los demás agentes, asigna tareas y resuelve conflictos dentro de su área de acción.
2. Los agentes de datos gestionan las fuentes de datos, acceden a información relevante usando distintas bases y repositorios.
3. Los agentes de preprocesamiento limpian, transforman, normalizan y preparan los datos para ser extraídos apropiadamente.
4. Los agentes de minería se encargan en extraer los datos aplicando algoritmos apropiados según el tipo de análisis.

5. Los agentes de conocimiento interpretan y analizan patrones, y consolidan los resultados como información final.
6. Los agentes de interfaz, como especializados en la interacción con el usuario, son los que dan formato a la información y presentan los resultados al coordinador, quien, finalmente, envía la respuesta al solicitante.

La secuencia inicia desde el momento en que el usuario formula la solicitud. Luego, el proceso se desarrolla de manera cooperativa, adaptativa y paralela, bajo la supervisión de un agente coordinador que actúa como orquestador de todo el ciclo. Conforme al modelo MADM, y tomando como referencia el proceso de minería de datos, diversos tipos de agentes especializados intervienen con funciones específicas, cuya interacción es gestionada por un coordinador general (Tabla 1).

Tabla 1

Funciones de multiagentes según modelo MADM

Tipo de agente	Función principal
Agente de coordinación (*coordination agent*)	Orquesta la cooperación entre los demás agentes, asigna tareas y resuelve conflictos.
Agente de datos (*user agent*)	Gestiona fuentes de datos distribuidas, accede y selecciona la información relevante. Proporciona una interfaz gráfica de usuario entre los agentes de usuario y los agentes de tarea. Conduce hacia la minería de datos genérica multiagéntica.
Agente de preprocesamiento (*task agent*)	Es responsable de gestionar y programar las solicitudes de minería de datos. Facilitan la ejecución de las tareas de minería de datos: limpian, transforman y normalizan los datos.
Agente de datos (*data agent*)	Posee metadatos sobre un conjunto de datos específico, lo que les permite acceder a ellos. Existe una relación directa entre los agentes de datos y los conjuntos de datos.
Agente minero o analítico (*mining agent*)	Es responsable de realizar la minería de datos y generar resultados. El resultado se devuelve a un agente de tareas. Aplica algoritmos de minería de datos (por ejemplo, árboles de decisión, redes neuronales, *clustering*, reglas de asociación).
Agente de validación (*validation agent*)	Es responsable de evaluar y valorar la validez de los resultados de minería de datos. Identifica varios subtipos de agentes de validación, como subtipos identificados y asociados con diferentes tareas genéricas de minería de datos.
Agente de conocimiento (*knowledge agent*)	Interpreta y almacena los patrones descubiertos, y gestiona la base de conocimiento.
Agente de interfaz o usuario (*user interface agent*)	Facilita la interacción con el analista o usuario final.

Nota. Los datos proceden de "Multi-Agent Based Clustering: Towards Generic Multi-Agent Data Mining", por S. Chaimontree, K. Atkinson y F. Coenen, 2010, en P. Perner (Ed.), *Advances in Data Mining: Applications and Theoretical Aspects*, Springer (https://doi.org/10.1007/978-3-642-14400-4_9).

Desde esta perspectiva, las arquitecturas multiagente de IA contribuyen al desarrollo productivo cuando permiten la automatización y distribución inteligente de tareas, tal como lo demuestra el marco conceptual MADM, orientado a optimizar procesos analíticos en entornos complejos y de gran escala. En este contexto, los multiagentes cooperan para ejecutar tareas específicas de minería de datos, lo que facilita el acceso, el análisis y el procesamiento de la información mediante una ejecución autónoma, distribuida y coordinada de las distintas etapas del proceso (Tabla 1).

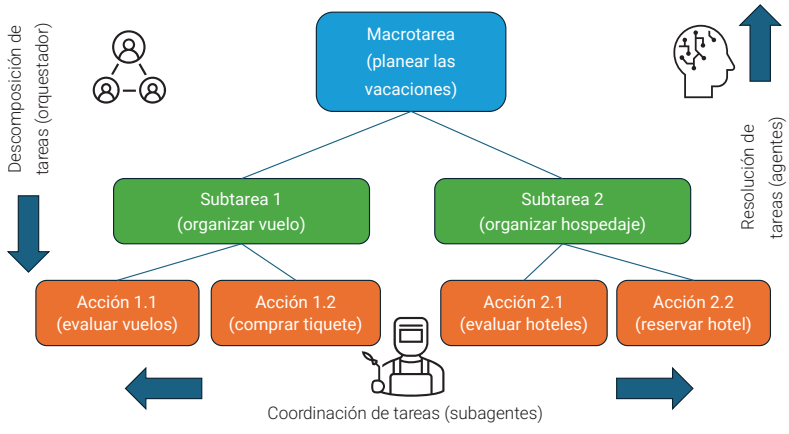
Árbol de tareas complejas

Los motores de búsqueda han existido durante décadas y desempeñan un papel fundamental al ofrecer acceso casi inmediato a respuestas y recursos para una amplia gama de consultas. Sin embargo, el fortalecimiento de los procesos de búsqueda mediante agentes de inteligencia artificial exige situar a las tareas como el eje central del *information retrieval*, en la medida en que estas articulan los objetivos, las acciones y los contextos que guían la interacción entre los usuarios y los sistemas de búsqueda.

De acuerdo con White (2024), los sistemas tradicionales continúan basándose, principalmente, en la observación de acciones explícitas del usuario, lo que limita la inferencia de los objetivos subyacentes, especialmente en escenarios caracterizados por una elevada complejidad. Estas tareas se caracterizan por ser mal definidas, multifásicas y distribuidas a lo largo de múltiples consultas y sesiones, además de exigir elevados niveles de esfuerzo cognitivo, estrechamente vinculados con la experiencia del usuario cuando interactúa con la IA. No obstante, su integración con arquitecturas de agentes de IA organizadas jerárquicamente permite trascender los modelos tradicionales centrados en consultas aisladas.

Figura 2

Árbol de tareas complejas utilizando agentes y subagentes de IA



Nota. Los datos proceden de "Advancing the Search Frontier with AI Agents", por R. W. White, 2024, *Communications of the ACM*, *67*(9), p. 3 (https://doi.org/10.1145/3655615).

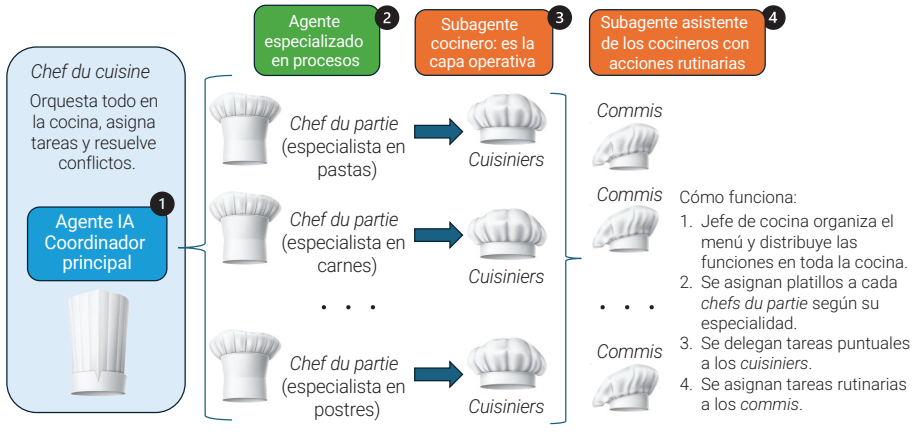
Un agente orquestador, apoyado por subagentes especializados, permite una comprensión más precisa de las intenciones del usuario y la coordinación de acciones avanzadas para la resolución progresiva de tareas complejas. En la Figura 2, se muestra un proceso de búsqueda de información orientado a la resolución de tareas complejas, modelado mediante una estructura jerárquica en forma de árbol.

Esta propuesta permite la descomposición de tareas complejas en macrotareas, subtareas y acciones, que representan objetivos de alto nivel (planear las vacaciones), componentes intermedios (evaluar el vuelo y el hospedaje) y, finalmente, operaciones observables (comprar un tiquete y reservar un hotel). Para abordar la complejidad inherente a tareas de búsqueda, White (2024) sugiere el modelo de árbol de tareas, sustentado en agentes y subagentes coordinados jerárquicamente, en el que los agentes especializados asumen la ejecución de subtareas específicas bajo la supervisión de un agente orquestador (subtareas 1 y 2). Este enfoque facilita la descomposición estructurada de dichas subtareas (acciones 1.1, 1.2, 2.1 y 2.2).

La coordinación eficiente y la resolución progresiva de tareas complejas consolidan a las arquitecturas subagénticas como un mecanismo fundamental para mejorar la efectividad de los sistemas de búsqueda y para apoyar la toma de decisiones (White, 2024).

Figura 3

Metáfora de la cocina utilizando agentes y subagentes de IA



Nota. Los datos proceden de "Creando asistentes, agentes y plataformas colaborativas de inteligencia artificial", por O. Bustillos Ortega, J. Murillo Gamboa, O. Nuñez Peralta y F. Rodríguez Sibaja, 2025, *Interfases*, (21), p. 46 (https://doi.org/10.26439/interfases2025.n021.7801).

Metáfora de la cocina

Otra forma de comprender el funcionamiento de los agentes y asistentes de IA es a través de la organización jerárquica de una brigada de cocina tradicional, considerando al

personal compuesto por un chef principal, chefs especializados, cocineros y asistentes de cocina. Esta metáfora permite ilustrar cómo los distintos roles humanos pueden asociarse conceptualmente con agentes y subagentes de IA, cada uno encargado de ejecutar tareas diferenciadas y coordinadas dentro de un entorno operativo común (Figura 3).

El ciclo operacional de la cocina, según la Figura 3, se articula de la siguiente manera:

1. *Chef de cuisine*: define el menú del día y establece las directrices generales relativas para la preparación y presentación de los platillos. Luego, descompone estas directrices en objetivos y tareas específicas, que son asignadas a los *chefs* conforme a su especialización funcional.
2. *Chefs de partie* especializados: coordinan con sus respectivos subagentes (*cuisiniers* y *commis*), quienes ejecutan las operaciones concretas requeridas para la elaboración de cada platillo, a fin de asegurar la coherencia, la eficiencia y la alineación con los objetivos globales del sistema (Bustillos Ortega et al. 2025).
3. *Cuisiniers*: se enfocan en elaborar cada componente del platillo.
4. *Commis*: asisten y ayudan a los *cuisiniers* y a los *chefs de partie*.

En esta metáfora, el *chef de cuisine* actúa como el responsable de la coordinación global de todas las actividades, por lo que cumple un rol equivalente al de un orquestador en un sistema de IA. Los *chefs de partie* son los responsables de la preparación de platillos específicos —como pastas, carnes o postres—, por lo que pueden considerarse agentes de IA especializados, cada uno focalizado en un dominio funcional o en un conjunto de tareas bien delimitadas. Los cocineros (*cuisiniers*) y los ayudantes de cocina (*commis*) conforman la capa operativa de soporte, que es análoga a la de los subagentes, encargados de ejecutar acciones más acotadas, rutinarias y de menor nivel de abstracción, siguiendo las directrices establecidas por los agentes. principales.

Soluciones con agentes y subagentes de IA

Los agentes de IA pueden, entre otras cosas, ayudar a los usuarios a alcanzar objetivos, maximizar la utilidad y realizar funciones automatizadas. Por ejemplo, los asistentes digitales personales Siri de Apple y Alexa de Amazon, que pueden responder preguntas y ayudar con la gestión de tareas personales.

Con respecto a las plataformas de desarrollo de soluciones de IA, existe GitHub Copilot (Github, s. f.), el cual reduce el esfuerzo requerido al desarrollador de aplicaciones, facilita la ejecución exitosa de las tareas y acelera significativamente la finalización de soluciones automatizadas. También está el Auto-GPT (s. f.), un agente totalmente autónomo que puede descomponer las tareas en subtareas y ejecutarlas de forma independiente en nombre del usuario para facilitar la consecución de objetivos.

En ambos casos, la capa de orquestación de IA gestiona los flujos de información internos, activa, establece y ejecuta herramientas o complementos, y procesa sus respuestas, entre otras cosas. Todo esto se ejecuta sobre una infraestructura de IA a gran escala alojada en la nube en plataformas, tales como Microsoft Azure, Google Cloud y Amazon Web Services. Se sustenta en un fuerte compromiso con una IA responsable que garantiza la seguridad, protección y transparencia de los agentes (White, 2024).

Índice Latinoamericano de IA (ILIA2025)

De acuerdo con Durán et al. (2025), la iniciativa del Índice Latinoamericano de Inteligencia Artificial (ILIA) es una iniciativa elaborada en el 2023 por el Centro Nacional de Inteligencia Artificial (CENIA) de Chile y el Observatorio de Desarrollo Digital de la Comisión Económica para América Latina y el Caribe (CEPAL), con el apoyo de diversas organizaciones académicas, públicas y privadas. El ILIA identifica una transición en la región, basado en el análisis de ochenta y cinco indicadores en doce áreas clave para la transformación digital y con la visión de incorporar de manera estratégica estas tecnologías emergentes para acelerar un desarrollo productivo, inclusivo y sostenible, y asegurar, al mismo tiempo, su uso ético, responsable y orientado al bien común.

Los resultados del ILIA 2025 evidencian que América Latina avanza progresivamente desde una fase de adopción inicial hacia la consolidación de capacidades propias en investigación, formación avanzada y desarrollo de talento especializado en inteligencia artificial, aunque este progreso se caracteriza por marcadas asimetrías entre países. El conocimiento y la producción científica continúan concentrándose en un grupo reducido de naciones —principalmente Brasil, México, Chile, Colombia y Argentina—, mientras que gran parte de la región mantiene ecosistemas científicos incipientes y una limitada proyección internacional. No obstante, experiencias como las de Costa Rica y Perú demuestran que la articulación entre talento humano, institucionalidad sólida y adopción responsable de la IA puede sentar las bases para ecosistemas escalables, auditables y orientados al bien común, en los que la inteligencia distribuida actúe como motor del desarrollo sostenible. En el ámbito de la formación académica, se observa un crecimiento sostenido de los programas de maestría y doctorado en IA, junto con la expansión de iniciativas de ciencia abierta y colaboración regional que favorecen el desarrollo de soluciones locales y transparentes. Sin embargo, persiste un "embudo formativo": mientras se amplía la alfabetización en IA en los niveles básico y técnico, continúa siendo insuficiente la generación de desarrolladores e investigadores altamente especializados (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023).

En tal sentido, se considera que la formación futura deberá integrar ética, explicabilidad, sostenibilidad y multidisciplinariedad, incorporando la IA generativa como herramienta pedagógica y de democratización del conocimiento. Asimismo, el informe recomienda fortalecer centros regionales de excelencia, programas interdisciplinarios de

posgrado y consorcios de infraestructura compartida en cómputo y datos abiertos. Solo así la región latinoamericana podrá pasar de ser consumidora a productora de innovación responsable, lo que garantizará soberanía digital y desarrollo sostenible.

El ILIA 2025 destaca el avance de América Latina hacia ecosistemas de IA más integrados, en los que los sistemas compuestos por agentes cooperativos y autónomos emergen como un instrumento clave para impulsar el desarrollo productivo sostenible. Estas arquitecturas distribuyen funciones de percepción, razonamiento y acción, lo que mejora el desempeño y la escalabilidad de los procesos industriales, educativos y públicos mediante la optimización de recursos y la adaptabilidad operativa. Sin embargo, su implementación enfrenta desafíos en orquestación, seguridad, costo computacional y gobernanza ética.

Se proyecta que el desarrollo futuro de la IA en América Latina estará condicionado por la capacidad de sus sistemas educativos y científicos para formar talento avanzado, promover investigación abierta y aplicada, y articular redes regionales de conocimiento. En este sentido, el tránsito desde una alfabetización básica hacia una especialización profunda en IA se configura como un punto decisivo, ya que solo a través de una educación con enfoque ético, interdisciplinario y sostenible la región podrá superar su rol de consumidora de tecnología y consolidarse como productora de innovación responsable (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023). En la Tabla 2, se presentan recomendaciones clave agrupadas según el eje estratégico del ILIA2025.

Tabla 2

Recomendaciones estratégicas para la academia según ILIA 2025

Eje estratégico	Recomendación clave	Objetivo
Política educativa	Incluir IA en todos los niveles educativos, con enfoque en equidad y ética	Democratizar el acceso y generar interés temprano
Investigación aplicada	Vincular universidades con sectores productivos y públicos	Alinear la IA con desafíos nacionales (salud, sostenibilidad, educación)
Formación avanzada	Fortalecer programas de maestría y doctorado interdisciplinarios	Elevar la especialización científica y tecnológica
Infraestructura académica	Crear consorcios regionales de cómputo y datos abiertos	Garantizar soberanía digital e investigación de frontera
Colaboración internacional	Participar en redes globales y normas ISO/IEEE de IA	Aumentar visibilidad y transferencia de conocimiento

Nota. Los datos proceden de *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025*, por R. Durán, A. Moreno, S. Adasme, S. Rovira, V. Jordán y L. Poveda (Coords.), 2025, Comisión Económica para América Latina y el Caribe (https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025).

El ILIA 2025 confirma que América Latina transita de una etapa de adopción incipiente hacia la construcción de capacidades propias en investigación y formación avanzada en inteligencia artificial. Sin embargo, este proceso avanza de manera desigual y todavía enfrenta limitaciones estructurales en inversión, especialización y articulación científica (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023) (Tabla 3).

Tabla 3

Índice Latinoamericano ILIA 2025 estado actual y tendencias en investigación

ILIA (indicadores)	Estado actual y tendencias
Concentración geográfica del conocimiento	La investigación en IA sigue altamente concentrada. Brasil y México reúnen el 68 % de los investigadores activos y —junto con Colombia, Chile y Argentina— concentran el 90 % de las publicaciones científicas. Solo siete países participan en conferencias internacionales de alto impacto, con predominio de Chile y Brasil, lo que deja a la mayoría con baja visibilidad.
Crecimiento del ecosistema formativo en Latinoamérica	Se observa un aumento de programas de posgrado en IA (magíster y doctorado), especialmente en Chile, Brasil, México, Colombia y emergentes como Costa Rica y Perú, que han incorporado IA en currículos universitarios y de educación técnica. Pese a este avance, once de los diecinueve países aún no ofrecen programas de doctorado específicos en IA.
Apertura y colaboración	El informe destaca el modelo de código abierto y la ciencia colaborativa como motores del desarrollo científico regional. Honduras, El Salvador y Cuba son ejemplos del potencial del *open source* para crear soluciones locales y fomentar comunidades académicas activas. Esta tendencia podría ser clave para democratizar la investigación, superar barreras económicas y promover la transparencia algorítmica.
Desafíos estructurales	Escasa inversión en I+D. La región representa apenas el 1,12 % de la inversión global en IA, pese a concentrar el 8,8 % de la población mundial. Falta de redes interuniversitarias sólidas y limitado intercambio de investigadores impiden generar una masa crítica regional. Poca participación en estándares y foros internacionales, lo que reduce la influencia latinoamericana en la gobernanza global de la IA.
Hacia una alfabetización avanzada	La alfabetización en IA en la región Latinoamericana se ha expandido —ya presente en los currículos escolares de seis países—, pero la formación especializada sigue rezagada. La proporción de profesionales con habilidades técnicas en IA es cuatro veces menor que la de alfabetización general. El informe advierte un "embudo formativo": se forma una base amplia de usuarios, pero pocos desarrolladores o investigadores.
Competencias clave para el futuro	El ILIA 2025 propone impulsar nuevos perfiles académicos centrados en IA responsable y ética, con formación en sesgos, transparencia y explicabilidad, así como en ciencia de datos aplicada a políticas públicas y desarrollo sostenible. Asimismo, se recomienda promover la interdisciplinariedad mediante la integración de ingeniería, ciencias sociales y humanidades en programas de IA, así como desarrollar capacidades para la IA generativa y multimodal, considerada un campo democratizador y de alto impacto.

(*continúa*)

(*continuación*)

ILIA (indicadores)	Estado actual y tendencias
Formación conectada a la soberanía digital	La escasez de infraestructura (HPC y GPU) limita la investigación avanzada; por ello, el ILIA recomienda alianzas regionales para uso compartido de cómputo y datos. La educación debe orientarse a fortalecer capacidades endógenas, a fin de evitar la dependencia tecnológica y sesgos en los modelos importados.
Modelos educativos colaborativos	Se promueve la creación de centros regionales de excelencia en IA y plataformas abiertas de aprendizaje que repliquen experiencias exitosas como "Hazlo con IA" en Chile o el LatamGPT colaborativo. Estas iniciativas demuestran que la formación masiva puede acelerar la adopción responsable y reducir la brecha de talento.

Nota. Los datos proceden de *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025*, por R. Durán, A. Moreno, S. Adasme, S. Rovira, V. Jordán y L. Poveda (Coords.), 2025, Comisión Económica para América Latina y el Caribe (https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025).

Riesgos y ética

Los riesgos relacionados con las tecnologías de IA ponen de manifiesto la necesidad de diseñar arquitecturas que incorporen mecanismos de visibilidad, trazabilidad y control capaces de habilitar una supervisión granular del comportamiento de los agentes. La literatura coincide en que el aumento de la autonomía de los agentes inteligentes incrementa su exposición a amenazas técnicas y éticas, lo que exige el diseño de arquitecturas jerárquicas basadas en subagentes especializados que refuercen la supervisión, la resiliencia y la gobernanza interna.

Para mitigar riesgos asociados a la privacidad, el control, la explicabilidad y la asignación de responsabilidades, resulta fundamental incorporar desde etapas tempranas principios éticos, mecanismos robustos de seguridad, así como procesos de auditoría y trazabilidad que garanticen un despliegue confiable, seguro y responsable de estas arquitecturas en entornos productivos (Deng et al. 2025; Ning et al. 2025; Samdani et al. 2025). De acuerdo con Chan et al. (2024), los agentes de IA presentan riesgos significativamente mayores que los sistemas tradicionales, debido a su capacidad de actuar autónomamente y de ejecutar acciones en el mundo digital sin supervisión constante. En tal sentido, destacan como principales riesgos los siguientes:

- La manipulación de los objetivos del agente, que puede llevar a perseguir fines no alineados con los establecidos por los desarrolladores
- La ejecución de acciones dañinas o inseguras, derivadas de interpretaciones erróneas, instrucciones ambiguas o fallos en sus cadenas de planificación
- La susceptibilidad a ataques externos, en los que actores maliciosos pueden redirigir, engañar o explotar al agente

- La opacidad operativa que dificulta detectar desviaciones de comportamiento, fallos de seguridad o actividades no autorizadas

En conjunto, las arquitecturas subagénticas presentan riesgos significativos asociados a la seguridad, la gobernanza, la propagación de sesgos y los costos computacionales derivados de la coordinación entre múltiples subagentes. Frente a estas vulnerabilidades, resulta esencial promover diseños que integren explícitamente el bienestar humano y fomenten una colaboración humano-IA segura, transparente y responsable. A futuro, los sistemas de IA colaborativos basados en subagentes deberán orientarse a tareas especializadas y centradas en el usuario y avanzar hacia modelos de control compartido humano-IA (Samdani et al. 2025).

Desde una perspectiva de riesgos, los resultados del ILIA 2025 evidencian que la limitada inversión en investigación y desarrollo constituye una vulnerabilidad estructural crítica para el avance de la inteligencia artificial en América Latina. A pesar de albergar el 8,8 % de la población mundial, la región concentra apenas el 1,12 % de la inversión global en IA, lo que incrementa el riesgo de dependencia tecnológica, rezago científico y pérdida de competitividad. Este escenario se ve agravado por la debilidad de las redes interuniversitarias y el escaso intercambio de investigadores, factores que obstaculizan la conformación de una masa crítica regional y limitan la capacidad de generar conocimiento, innovación y liderazgo científico sostenible. Frente a este riesgo, la ausencia de centros regionales de excelencia en inteligencia artificial compromete la formación continua de talento especializado y la adopción coordinada, ética y responsable de arquitecturas subagénticas de IA, lo que incrementa la probabilidad de una implementación fragmentada y poco sostenible de estas tecnologías (Durán et al. 2025; Soto Arriaza & Durán Rojas, 2023).

DISCUSIÓN

Los resultados del análisis evidencian que las arquitecturas subagénticas de inteligencia artificial representan un avance significativo frente a los enfoques monolíticos tradicionales, pues permiten la especialización funcional, la descomposición de tareas complejas y la coordinación jerárquica mediante agentes orquestadores. Esta organización distribuida se traduce en mejoras sustanciales en términos de eficiencia operativa, escalabilidad, paralelismo y adaptabilidad, particularmente en contextos caracterizados por grandes volúmenes de información y procesos multifásicos.

El estudio del marco conceptual MADM demuestra que la integración de técnicas de minería de datos con sistemas multiagente facilita la automatización y optimización del descubrimiento de conocimiento, pues asigna funciones específicas a agentes especializados en preprocesamiento, análisis, validación e interpretación. Este enfoque evidencia que la inteligencia distribuida no solo acelera los flujos analíticos, sino que también

mejora la trazabilidad y la calidad de los resultados, lo que la consolida como un modelo eficaz para entornos productivos y científicos a gran escala.

Asimismo, el modelo del árbol de tareas complejas confirma que la estructuración jerárquica de tareas mediante agentes y subagentes permite una comprensión más precisa de las intenciones del usuario y una resolución progresiva de problemas complejos. La capacidad de descomponer macrotareas en subtareas y acciones observables fortalece los procesos de búsqueda de información y soporte a la toma de decisiones, lo que supera las limitaciones de los sistemas tradicionales basados en consultas aisladas.

La metáfora de la brigada de cocina refuerza conceptualmente estos hallazgos al ilustrar cómo la coordinación entre roles especializados, bajo la supervisión de un orquestador, resulta clave para garantizar la coherencia, la eficiencia y la calidad de los resultados. Esta analogía facilita la comprensión de la dinámica operativa de las arquitecturas subagénticas y subraya la importancia de una orquestación clara para evitar redundancias, conflictos y sobrecarga computacional.

En cuanto a las aplicaciones prácticas, los casos analizados en movilidad y transporte, finanzas, salud y educación evidencian que las arquitecturas subagénticas potencian la productividad al distribuir responsabilidades, aumentar la explicabilidad de los procesos y favorecer la colaboración entre humanos e IA. No obstante, los resultados también ponen de manifiesto desafíos persistentes asociados a la orquestación técnica, la seguridad, los costos computacionales y la gobernanza ética, especialmente en sistemas con altos niveles de autonomía.

La discusión resalta que la autonomía creciente de los subagentes incrementa los riesgos de opacidad, sesgos, vulnerabilidades de seguridad y desviaciones de comportamiento, lo que exige la incorporación de mecanismos de monitoreo, auditoría, trazabilidad y control jerárquico como componentes estructurales del diseño. En este sentido, la delegación de funciones de supervisión y seguridad a subagentes especializados emerge como una estrategia efectiva para reforzar la resiliencia y la confiabilidad de los sistemas.

Finalmente, el análisis del ILIA 2025 contextualiza estos resultados en el ámbito regional, lo que evidencia que América Latina cuenta con avances relevantes en alfabetización y adopción de IA, pero enfrenta limitaciones estructurales en inversión, infraestructura y formación avanzada. Los hallazgos sugieren que las arquitecturas subagénticas pueden convertirse en un catalizador del desarrollo productivo regional, siempre que se articulen con políticas educativas, marcos éticos sólidos y estrategias de cooperación interinstitucional orientadas a la soberanía digital y al bien común.

CONCLUSIONES

Con respecto a la pregunta de investigación —¿cómo contribuyen las arquitecturas subagénticas de IA a impulsar el desarrollo productivo, considerando los desafíos en

la orquestación, la seguridad y la gobernanza ética?—, las arquitecturas subagénticas de inteligencia artificial contribuyen de manera decisiva al desarrollo productivo al posibilitar la descomposición de tareas complejas en unidades funcionales especializadas, coordinadas por un agente orquestador dentro de sistemas distribuidos. Este enfoque incrementa la eficiencia operativa, la escalabilidad y la adaptabilidad de los procesos en sectores clave, tales como la industria, la educación, la salud, las finanzas y la gestión del conocimiento, al permitir la ejecución paralela, la reutilización de capacidades especializadas y una mejor gestión del contexto en entornos intensivos en datos. Marcos como el MADM, el árbol de tareas complejas y las arquitecturas de WebAgents demuestran que la inteligencia distribuida facilita la automatización avanzada y la toma de decisiones informadas en escenarios de alta complejidad.

No obstante, el impacto productivo de estas arquitecturas depende críticamente de la orquestación técnica, ya que la coordinación jerárquica entre agentes y subagentes requiere mecanismos robustos para la asignación de tareas, la resolución de conflictos y la integración coherente de resultados. Una orquestación deficiente puede generar ineficiencias, sobrecostos computacionales y pérdida de control operativo, lo que subraya la necesidad de diseños modulares, monitoreo en tiempo real y estrategias de control distribuido.

En paralelo, los riesgos para la seguridad emergen como un desafío estructural, dado que la autonomía de los subagentes amplía la superficie de ataque y los riesgos asociados a acciones no alineadas, fallos de interpretación o interferencias maliciosas. El documento evidencia que la incorporación de subagentes dedicados a funciones de identificación, monitoreo, auditoría y registro fortalece la resiliencia del sistema y permite aplicar principios de defensa en profundidad, lo que reduce vulnerabilidades sin sacrificar la flexibilidad operativa.

La gobernanza ética se posiciona como un eje transversal para garantizar que el desarrollo productivo impulsado por arquitecturas subagénticas sea sostenible y socialmente legítimo. La necesidad de transparencia, trazabilidad, explicabilidad y rendición de cuentas exige integrar marcos normativos, auditorías continuas y supervisión entre humanos e IA desde las fases iniciales de diseño. En este sentido, el ILIA 2025 destaca que la articulación entre talento humano, institucionalidad sólida y adopción responsable de la IA puede convertir a estas arquitecturas en un motor de innovación inclusiva, especialmente en contextos latinoamericanos caracterizados por brechas de infraestructura y especialización.

El documento concluye que las arquitecturas subagénticas de IA impulsan el desarrollo productivo al combinar especialización funcional, coordinación jerárquica y automatización inteligente, siempre que se acompañen de estrategias integrales de orquestación, seguridad y gobernanza ética. Su verdadero valor no reside únicamente en la eficiencia técnica, sino en su capacidad para habilitar ecosistemas productivos más

transparentes, resilientes y orientados al bien común, sustentados en una formación avanzada y políticas públicas responsables.

En conclusión, el desafío central para la región consiste en transformar la alfabetización tecnológica en innovación científica sustentada en disciplinas STEM, promoviendo una adopción responsable de la inteligencia artificial, el desarrollo sostenible y una participación activa en la producción global de soluciones basadas en IA. Este objetivo requiere una formación continua y especializada, así como la implementación estratégica de arquitecturas subagénticas en los distintos sectores productivos y en los sistemas educativos, de modo que la inteligencia distribuida se consolide como un motor de competitividad, equidad y bienestar social.

REFERENCIAS

Auto-GPT. (s. f.). *Auto-GPT – The next evolution of data driven Chat AI*. https://auto-gpt.ai/

Babaei, G., & Giudici, P. (2025). Explainable artificial intelligence (XAI) in investment decision-making. *Academia AI and Applications*, *1*(2). https://doi.org/10.20935/AcadAI8017

Bustillos Ortega, O., Murillo Gamboa, J., Núñez Peralta, O., & Rodríguez Sibaja, F. (2025). Creando asistentes, agentes y plataformas colaborativas de inteligencia artificial. *Interfases*, (21), 35-57. https://doi.org/10.26439/interfases2025.n021.7801

Carrascosa, C., Terrasa, A., Fabregat, J., & Botti, V. (2005). Gestión de comportamientos en agentes de tiempo real. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, *9*(25), 39-48. https://dialnet.unirioja.es/servlet/articulo?codigo=1212985

Chaimontree, S., Atkinson, K., & Coenen, F. (2010). Multi-agent based clustering: Towards generic multi-agent data mining. En P. Perner (Ed.), *Advances in data mining: Applications and theoretical aspects* (pp. 115-127). Springer. https://doi.org/10.1007/978-3-642-14400-4_9

Chan, A., Ezell, C., Kaufmann, M., Wei, K., Hammond, L., Bradley, H., Bluemke, E., Rajkumar, N., Krueger, D., Kolt, N., Heim, L., & Anderljung, M. (2024). Visibility into AI agents. En *FAccT '24: The 2024 ACM conference on fairness, accountability, and transparency* (pp. 958-973). Association for Computing Machinery. https://doi.org/10.1145/3630106.3658948

Chevrot, A., Vernotte, A., Falleri, J.-R., Blanc, X., Legeard, B., & Cretin, A. (2025). Are autonomous web agents good testers? *Proceedings of the ACM on Software Engineering*, *2*(ISSTA), 206-228. https://doi.org/10.1145/3728879

Claude Code Docs. (2025). *Crear subagentes personalizados*. https://code.claude.com/docs/es/sub-agents

Deng, Z., Guo, Y., Han, C., Ma, W., Xiong, J., Wen, S., & Xiang, Y. (2025). AI agents under threat: A survey of key security challenges and future pathways. *ACM Computing Surveys*, *57*(7), Artículo 182. https://doi.org/10.1145/3716628

Durán, R., Moreno, A., Adasme, S., Rovira, S., Jordán, V., & Poveda, L. (Coords.). (2025). *Índice Latinoamericano de inteligencia artificial (ILIA) 2025*. Comisión Económica para América Latina y el Caribe. https://www.cepal.org/es/publicaciones/82514-indice-latinoamericano-inteligencia-artificial-ilia-2025

Elmahalawy, A. M. (2015). A new approach in learning for intelligent multi agent systems. En *Proceedings 21st European Conference on Modelling and Simulation*. European Council for Modelling and Simulation. https://www.academia.edu/70025062/Intelligent_Agent_Technology_Intelligent_Multi_Agent

Github. (s. f.). *Github Copilot. Domina tu oficio*. https://github.com/features/copilot?locale=es-419

Gonçalves, E. J. T., Cortés, M. I., Campos, G. A. L., Lopes, Y. S., Freire, E. S. S., Da Silva, V. T., De Oliveira, K. S. F., & De Oliveira, M. A. (2015). MAS-ML 2.0: Supporting the modelling of multi-agent systems with different agent architectures. *Journal of Systems and Software*, *108*, 77-109. https://doi.org/10.1016/j.jss.2015.06.008

Ning, L., Liang, Z., Jiang, Z., Qu, H., Ding, Y., Fan, W., Wei, X. Y., Lin, S., Liu, H., Yu, P. S., & Li, Q. (2025). A survey of web agents: Towards next-generation AI agents for web automation with large foundation models. En *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 6140-6150). Association for Computing Machinery. https://doi.org/10.1145/3711896.3736555

Rother, D., Pajarinen, J., Peters, J., & Weisswange, T. H. (2025). Open-ended coordination for multi-agent systems using modular open policies. *Autonomous Agents and Multi-Agent Systems*, *39*, Artículo 40. https://doi.org/10.1007/s10458-025-09723-7

Samdani, G., Viswanathan, G., & Dasu Jegadeesh, A. (2025). Human-AI collaboration: Balancing agentic ai and autonomy in hybrid systems. *International Journal on Cloud Computing: Services and Architecture*, *15*(1). https://doi.org/10.5121/ijccsa.2025.15101

Soto Arriaza, A., & Durán Rojas, R. (Dirs.). (2023). *Índice latinoamericano de inteligencia artificial*. Centro Nacional de Inteligencia Artificial. https://indicelatam.cl/wp-content/uploads/2023/08/ILIA-2023.pdf

White, R. W. (2024). Advancing the search frontier with AI agents. *Communications of the ACM*, *67*(9), 54-65. https://doi.org/10.1145/3655615

Zhu, Z., Tan, Y., Yamashita, N., Lee, Y. C., & Zhang, R. (2025, 26 de abril). The benefits of prosociality towards AI agents: Examining the effects of helping AI agents on human well-being. En *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1-18). Association for Computing Machinery. https://doi.org/10.1145/3706598.3713116

GESTIÓN DE RIESGO DE MODELOS DE CRÉDITO CON *MACHINE LEARNING*: DESARROLLO, EXPLICABILIDAD Y MONITOREO AUTOMATIZADO

Eduardo Rafael Jauregui Romero
https://orcid.org/0009-0008-0248-2601
eduardo.jauregui1@unmsm.edu.pe
Universidad Nacional Mayor de San Marcos, Perú

Recibido: 26 de noviembre del 2025 / Aceptado: 23 de junio del 2026
doi: https://doi.org/10.26439/interfases2026.n023.8406

RESUMEN. La adopción de modelos de aprendizaje automático en el sector bancario peruano enfrenta el desafío crítico de garantizar su estabilidad y transparencia en entornos productivos, conforme a las exigencias de la regulación vigente sobre gestión de riesgo de modelos. Este estudio presenta no solo el desarrollo de un modelo de calificación crediticia basado en el algoritmo HistGradientBoosting optimizado, sino, principalmente, la implementación de una arquitectura integral de monitoreo automatizado desplegada en servicios de computación en la nube. Utilizando una muestra histórica de 50 206 registros, se diseñó un flujo de trabajo que integra ingeniería de características y validación cruzada y que ha logrado un indicador de Gini del 67,4 %. La principal contribución radica en el despliegue de una plataforma web interactiva que permite la vigilancia en tiempo real de métricas de desempeño y estabilidad. A través de este sistema, se evaluó la consistencia del modelo mediante el índice de estabilidad poblacional y se obtuvo un valor de 0,4 %, lo que confirma la ausencia de degradación predictiva o de cambios estructurales en las variables a lo largo del tiempo. Este enfoque demuestra cómo la integración de técnicas de *machine learning operations* (MLOps) y herramientas de visualización avanzada facilita la auditoría continua y la detección temprana de desviaciones, lo que supera las limitaciones de los reportes estáticos tradicionales.

PALABRAS CLAVE: riesgo de crédito / *machine learning* / monitoreo de modelos / gestión de riesgo de modelos

CREDIT MODEL RISK MANAGEMENT WITH MACHINE LEARNING: DEVELOPMENT, EXPLAINABILITY, AND AUTOMATED MONITORING

ABSTRACT. The adoption of machine learning models in the Peruvian banking sector faces the critical challenge of ensuring stability and transparency in production environments, in

compliance with current model risk management regulations. This study presents not only the development of a credit scoring model based on the optimized HistGradientBoosting algorithm but primarily the implementation of a comprehensive automated monitoring architecture deployed on cloud computing services. Using a historical sample of 50,206 records, a workflow integrating feature engineering and cross-validation was designed, achieving a Gini coefficient of 67.4%. The main contribution lies in the deployment of an interactive web platform that enables real-time monitoring of performance and stability metrics. Through this system, the model's consistency was evaluated using the Population Stability Index (PSI), obtaining a value of 0.4%, which confirms the absence of predictive degradation or structural changes in variables over time. This approach demonstrates how the integration of Machine Learning Operations (MLOps) techniques and advanced visualization tools facilitates continuous auditing and early detection of drifts, surpassing the limitations of traditional static reporting.

KEYWORDS: credit risk / machine learning / model monitoring / model risk management

INTRODUCCIÓN

En la actualidad, los modelos de inteligencia artificial (IA) están presentes en todos los sectores económicos. El sector bancario peruano no es ajeno a esta tendencia. Su importancia ha sido tal que se han emitido normas que promueven el uso de la inteligencia artificial en el sistema financiero del país (Superintendencia de Banca, Seguros y AFP [SBS], 2025). El uso de la IA se refleja en aumentos concretos de productividad. Estudios han documentado que la adopción de IA en empresas bancarias creció un 100 % durante la pandemia (Aguirre Felix Díaz et al., 2022); sin embargo, este crecimiento trajo consigo cuestionamientos sobre sesgo, discriminación, transparencia e interpretación que pueden afectar directamente a los usuarios (Carbó-Valverde & Rodríguez-Fernández, 2024).

Antes de la llegada masiva del *machine learning*, el sector bancario usaba algoritmos estadísticos fáciles de interpretar. El *score* crediticio era producto de una regresión logística, cuyos coeficientes podían analizarse directamente. Los modelos de *scoring* clásicos catalogan a los clientes según su nivel de riesgo en función de un conjunto de datos (Enriquez Montes & Orbezo Huamancari, 2022). Con la llegada del *machine learning*, quedó claro que estos algoritmos superan a los métodos tradicionales: la *support vector machine* (SVM) detectó un 20 % más de clientes riesgosos que la regresión logística (Cervantes Artavia et al., 2023), mientras que *random forest* junto con redes neuronales superaron a la regresión en 17 y 20 puntos de *area under the curve* (AUC), respectivamente (Valenzuela Sánchez, 2025).

Esta realidad llevó a la industria bancaria a adoptar algoritmos como XGBoost, KNN, Naive Bayes y Decision Tree, que forman parte del estado del arte en modelos de *scoring* (Tocto Reyes et al., 2022). Una ventaja central de estos métodos es que capturan relaciones no lineales entre variables, adaptándose mejor a un entorno cambiante (Cosme Cervera et al., 2025). A pesar de ello, esta ganancia en rendimiento tiene un costo: los modelos se vuelven difíciles de interpretar.

Frente a esto, la SBS emitió un reglamento que obliga a las instituciones a garantizar la gestión del riesgo de modelos dentro de las organizaciones bancarias, de seguros y de las AFP. Las debilidades en el desarrollo y seguimiento de modelos, como metodologías inadecuadas o supuestos incorrectos, pueden generar pérdidas y consecuencias adversas (Resolución S.B.S. 00053-2023). Para responder a estas exigencias, surgieron técnicas de explicabilidad como *shapley additive exPlanations* (SHAP) y *local interpretable model-agnostic explanations* (LIME), que permiten identificar el aporte de cada variable dentro de un algoritmo complejo (García García, 2024). Complementariamente, un correcto análisis exploratorio de datos y la ingeniería de variables facilitan la interpretabilidad desde la base del modelo (Erazo Laínez, 2024).

No obstante, entrenar bien un modelo no es suficiente. Una vez desplegado en producción, su comportamiento cambia conforme cambia la población que analiza. Por

lo tanto, el seguimiento debe cubrir dos frentes: el rendimiento, que mide qué tan bien distingue el modelo, y la estabilidad, que detecta variaciones en la predicción y en las variables predictoras (Cordova Benitez et al., 2024). Para evaluar el rendimiento se usan métricas como el coeficiente de Gini y las derivadas de la matriz de confusión (Dos Santos et al., 2025). Para evaluar la estabilidad, el índice de estabilidad poblacional (*population stability index*, PSI) permite medir el cambio en la distribución de la población respecto de una referencia y definir umbrales de alerta ante posibles cambios en la distribución (*drifts*) (Reyes Morales & Sosa Castro, 2022).

A pesar de los avances alcanzados por separado en cada uno de estos frentes, existe una brecha concreta en la literatura: los estudios de *scoring* se enfocan en el desarrollo del modelo, los de explicabilidad lo tratan como un análisis estático y los de *machine learning operations* (MLOps) abordan el monitoreo como un problema de infraestructura genérica. Ninguno integra estos tres componentes dentro de un marco orientado al cumplimiento regulatorio del sistema financiero peruano. Entonces, este artículo propone un flujo de trabajo que cubre el desarrollo, la explicabilidad mediante SHAP y el monitoreo automatizado en producción, considerando los lineamientos de la Resolución 00053-2023 de la SBS. Se comparan múltiples algoritmos con validación cruzada, se aplica un ajuste de hiperparámetros y se despliega una plataforma web de seguimiento en tiempo real que toma el conjunto de entrenamiento como referencia para detectar desviaciones futuras.

ESTADO DEL ARTE

Modelos de *machine learning* para riesgo crediticio

La predicción del incumplimiento crediticio ha dejado de depender exclusivamente de la regresión logística. Noriega et al. (2023), en una revisión de cincuenta y dos estudios sobre riesgo crediticio en microfinanzas, documentaron que los modelos de tipo *boosting* (XGBoost, AdaBoost, LightGBM) son los más efectivos y los más investigados, cuyos principales obstáculos recurrentes son el desequilibrio de clases y la baja representatividad de los datos. Por su parte, Suhadolnik et al. (2023) confirmaron esta tendencia con evidencia empírica sólida: evaluaron diez algoritmos sobre una cartera con más de 2,5 millones de registros y encontraron que los modelos de ensamble, en particular XGBoost, superan ampliamente a los métodos tradicionales en la clasificación de créditos.

En línea con lo anterior, Rahmani et al. (2024) propusieron un flujo de trabajo sistemático que incorpora la codificación *weight of evidence* para el tratamiento de valores atípicos y faltantes, junto con la optimización de hiperparámetros mediante algoritmos genéticos multiobjetivo, con lo que se obtuvieron modelos más robustos desde el punto de vista financiero. Yang et al. (2024), por su parte, aplicaron ensamble de modelos con LightGBM y XGBoost sobre el dataset Home Credit, a fin de gestionar el desbalance con SMOTE y mostrar que las variables de historial crediticio externo dominan la predicción.

En el contexto peruano, Castañeda Hilario et al. (2025) desarrollaron un modelo explicable para tarjetas de crédito de bancos como BBVA e Interbank mediante LightGBM e integraron SHAP, con un AUC de 0,94. Con ello, se identificó el uso intensivo de la línea de crédito y la morosidad acumulada como los predictores de mayor peso.

Explicabilidad e inteligencia artificial responsable

El alto rendimiento predictivo de los modelos complejos se enfrenta a una exigencia regulatoria concreta: las instituciones deben poder explicar sus decisiones. Hermitaño Castro (2022), en una revisión sistemática publicada en esta misma revista, señaló que la caja negra es la desventaja central de los algoritmos de *machine learning* (ML) en crédito y recomendó el uso de SHAP como vía para transparentar los resultados sin sacrificar rendimiento. De Lange et al. (2022) demostraron esta afirmación en un banco noruego real: un modelo LightGBM combinado con SHAP superó al modelo logístico interno del banco, identificando la volatilidad del saldo de crédito y la duración de la relación con el cliente como variables determinantes, y facilitando la interpretación ante reguladores. Weber et al. (2024), en una revisión de sesenta artículos sobre XAI en finanzas, observaron que el campo ha migrado de modelos inherentemente transparentes hacia técnicas de explicabilidad *post-hoc* sobre algoritmos más complejos, de las cuales SHAP fue la más utilizada. Cepeda Cavero y Véliz Soto (2025) analizaron las brechas entre principios éticos y gobernanza algorítmica real en la banca entre 2020 y 2025, y encontraron que reducir el sesgo puede costar hasta un 9 % de precisión y que la equidad estadística no siempre coincide con la equidad percibida por el usuario. Estos hallazgos refuerzan que la explicabilidad no es solo técnica sino también organizacional y regulatoria.

Monitoreo de modelos y MLOps

Entrenar un modelo no es suficiente, pues su comportamiento en producción cambia conforme cambia la población que analiza; además, detectar esa deriva a tiempo es parte del ciclo de vida del modelo. Dar y Cavus (2026) propusieron el método *profile drift detection* (PDD), que usa perfiles de dependencia parcial para identificar no solo cuándo el modelo se deteriora, sino también qué variable ha cambiado su relación con el objetivo, integrando XAI directamente al monitoreo. Bhagavathula (2025) desarrolló una plataforma de observabilidad en tiempo real basada en OpenTelemetry compatible con AWS, Azure y GCP, que detecta la deriva con una precisión de 94,2 % y menos de 5 % de falsos positivos combinando pruebas estadísticas como KS, PSI y chi-cuadrado de forma simultánea. Shcherbinin (2026) abordó el reentrenamiento adaptativo en entornos industriales mediante la orquestación con Apache Airflow y redujo el tiempo de detección de incidentes en un 60 % y la carga manual de preparación de datos en un 40 %. Estos trabajos muestran que el monitoreo efectivo no se reduce a calcular una métrica periódica, sino que requiere una arquitectura tecnológica diseñada para operar en producción de forma continua y auditable.

Marco regulatorio

El contexto normativo es inseparable del problema técnico. En el Perú, la Superintendencia de Banca, Seguros y AFP emitió la Resolución 00053-2023, que establece el Reglamento de Gestión de Riesgos de Modelo y obliga a las instituciones financieras a garantizar la estabilidad, la interpretabilidad y el monitoreo de sus modelos en producción (Zegarra Jara, 2025). Internacionalmente, el marco SS1/23 de la Autoridad de Regulación Prudencial del Reino Unido (True North Partners, 2023) clasifica el riesgo de modelo como una disciplina independiente y exige validación independiente, gobernanza liderada por el directorio y controles sobre el desarrollo y la implementación. Joshi (2025) advierte que los marcos MRM tradicionales no son suficientes para los modelos de IA actuales, dado que sus características estocásticas y la falta de trazabilidad dificultan la validación estándar, y propone integrar simulaciones de Monte Carlo y alinearse con estándares como NIST AI RMF e ISO/IEC 23894. Este estudio responde directamente a estos tres niveles regulatorios: local, regional e internacional.

La revisión de la literatura permite identificar una brecha concreta. Los estudios sobre modelos de ML para crédito se concentran en el desarrollo y en la comparación de algoritmos; los trabajos sobre XAI abordan la explicabilidad como un análisis estático posterior al entrenamiento; y los de MLOps tratan el monitoreo como un problema de infraestructura genérica. Ninguno de estos ejes se ha integrado en un solo flujo de trabajo orientado al cumplimiento regulatorio específico del sistema financiero peruano. Este estudio cubre esa brecha al proponer un marco que conecta el desarrollo del modelo, la explicabilidad mediante SHAP y el monitoreo automatizado en producción dentro de los lineamientos de la Resolución 00053-2023 de la SBS, lo que convierte el ciclo de vida del modelo en una unidad auditable y reproducible.

METODOLOGÍA

En este apartado se explicará a detalle las fases del desarrollo y seguimiento del modelo propuesto.

Dataset

Para el presente estudio se tomó una muestra de clientes del portafolio de una institución financiera peruana interesada en analizar el comportamiento de impago de su cartera. Los periodos disponibles comprenden desde el 202301 hasta el 202403, con un total de 50 206 registros y 44 variables que combinan información externa del reporte crediticio consolidado (RCC) con información interna. El reporte RCC permite conocer la salud financiera de un individuo natural o de una persona jurídica en las diferentes entidades del sistema financiero, lo que otorga una visión más amplia para el análisis (Romero Carmen, 2024). En la Tabla 1, se muestran los grupos de variables disponibles y su descripción general.

Tabla 1

Conjunto de variables del dataset *de la institución financiera*

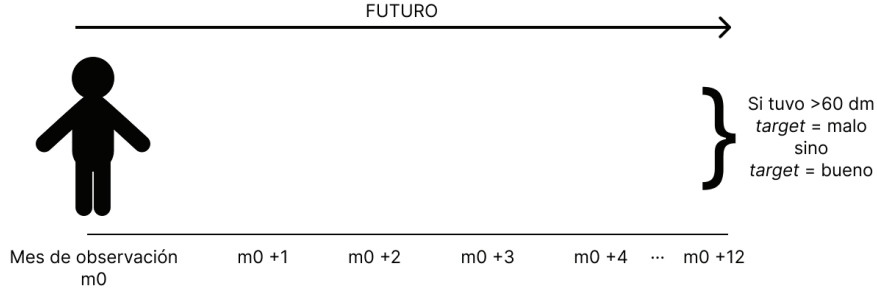
Fuente	# variables	Descriptivo general de variables
Stock	5	Diferencia de meses de buen y mal comportamiento. DIF_BUE_MA
RCC	1	*Flag* deuda vencida últimos 24 meses. FLG_DVNCD_24M
RCC	3	Incremento de deuda / incremento de entidades. INC_X
Stock	4	Máximo número de días con atraso. MAX_ATRA_X
RCC	3	Máxima deuda reportada / máximas entidades. MAX_DEU_ENT_X
Stock	4	Numero de meses con atrasos mayores a X días. NMES_UATR_X
Stock	5	Meses con buen comportamiento. N_BU_X
RCC	7	Numero de meses con deuda o normal o diferente normal. N_DEU_X
RCC	9	Máximos, promedios y ratios de deudas activas o hipotecarias
RCC	3	Variaciones de calificaciones, deudas y entidades. VAR_X

Nota. Las variables fueron agrupadas por relación. Esto no significa que sean exactamente iguales, ya que tienen diferentes ventanas de tiempo y casuísticas especiales.

En cuanto a los aspectos éticos, el *dataset* utilizado fue extraído en el marco de las funciones del autor dentro de la institución financiera y se encuentra completamente anonimizado, con los identificadores personales encriptados, por lo que ningún registro puede vincularse a un individuo específico. El estudio no involucra recolección de datos primarios ni contacto directo con personas, por lo que no requiere aprobación de un comité de ética. Dado que los modelos de *scoring* crediticio pueden afectar el acceso de personas a productos financieros, este trabajo enfatiza la interpretabilidad y explicabilidad del modelo como mecanismos para reducir el riesgo de decisiones arbitrarias o sesgadas, en línea con los principios de gobernanza algorítmica responsable (Cepeda Cavero & Véliz Soto, 2025).

Figura 1

Diagrama conceptual de la construcción del target



Nota. Para la creación del *target* se toma el máximo días de mora en los siguientes doce meses.

Análisis exploratorio de datos (EDA)

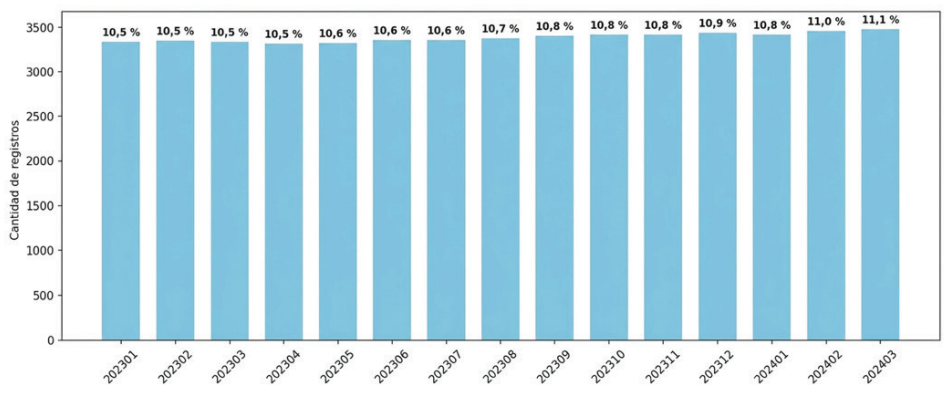
*Concepto y estabilidad de la variable objetivo (*target*)*

Para este artículo se definió que el *target* tenga dos posibles valores: cliente con *default* (malo) y cliente sin *default* (bueno). Los clientes malos son aquellos que tienen más de sesenta días de mora en los meses siguientes. En la Figura 1, se muestra en detalle el concepto del *target*.

Una vez definido el *target*, es importante analizar su distribución y su estabilidad a lo largo del tiempo. En la Figura 2, se muestra la cantidad de registros y la tasa de malos (TM) para cada uno de los periodos. La tasa de malos se define como el porcentaje de malos respecto del total de clientes; este indicador refleja la calidad de la cartera y puede presentar valores fluctuantes según el tipo de *stock* que se analice (Reyes Morales & Sosa Castro, 2022). Como se puede observar, el *dataset* se considera desbalanceado, con una proporción de 10 a 1. Este problema será abordado más adelante con hiperparámetros internos del modelo final, así que no es necesario aplicar una técnica de *over-/subsampling*.

Figura 2

Evolutivo de tasa de malos y cantidad de clientes



Nota. Se aprecia que la tasa de malos es estable a lo largo de los periodos.

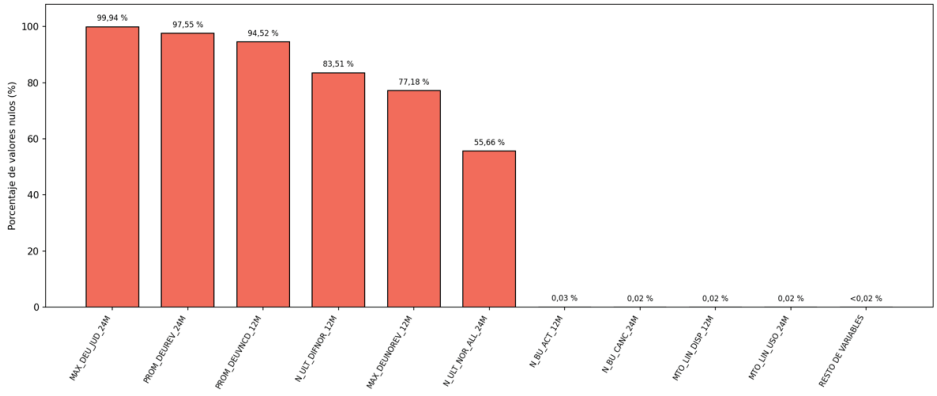
Análisis de valores faltantes en las variables

Durante esa fase, se analizaron las 44 variables, a fin de reconocer cuáles de ellas presentaban un alto nivel de valores faltantes (*missings*). En el caso de que las variables presentaran más del 90 % de *missings*, se retiraban del *dataset*, ya que su aporte será ínfimo al presentar grandes cantidades de valores en blanco. Asimismo, los *missings*

pueden imputarse; sin embargo, al tener una gran concentración de valores nulos, su efecto puede decantar en entrenar un modelo de manera incorrecta. Imputar variables con muchos *missings* puede causar sesgos y relaciones incorrectas (Moneta Pizarro et al., 2022). En la Figura 3, se muestran tres variables que poseen más del 90 % de *missings*, las cuales serán retiradas del *dataset*, por lo que tras este primer filtro, permanecen 41 de las 44 variables iniciales.

Figura 3

Porcentaje de missings *en las cuarenta y cuatro variables iniciales*



Nota. Figura creada con la librería missingno y matplotlib en Python.

Poder predictivo de variables

Para no entrenar el modelo con variables innecesarias, es importante realizar un correcto filtro sobre la base de su importancia. Una de las métricas más conocidas para medir el impacto y la relación que tienen las variables con respecto al *target* es el coeficiente de Gini. Esta métrica es la más usada, ya que su interpretación es sencilla debido a que indica el grado de diferenciación que se establece según las hojas de un árbol. Sus valores oscilan entre -1 y 1; los valores más extremos indican una mejor relación indirecta o directa, respectivamente (Gantman, 2020). En la Figura 4, se observan aquellas variables que no pasaron el filtro de importancia. Para el presente estudio, se eliminaron aquellas variables con un coeficiente de Gini menor del 5 % en valor absoluto.

Figura 4

Variables con coeficiente de Gini menor al 5 %

	variable	gini
33	NMES_UATR15_I_U6K_24M	0.043617
34	N_NEG_24M	0.042934
35	INC_MAX_DTOTSH_ACTU_24M	0.036731
36	N_DEU_ACTSH_18M	0.035886
37	INC_MAX_ENT_ACTU_24M	0.031567
38	RMAX_DREV_DDIR_24M	0.019963
39	R_DREV_DDIR_24M	0.019903
40	MAX_DEUNOREV_12M	0.017121

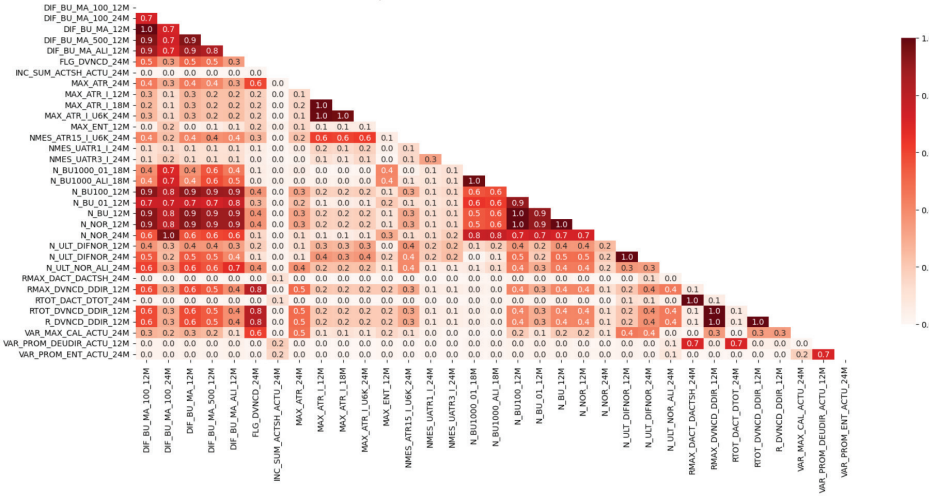
Nota. Ocho variables fueron eliminadas del *dataset* debido a su bajo poder predictivo.

Correlación de variables

Hasta este punto se ha analizado cada variable de manera individual, pero se ha obviado una parte fundamental: estas variables pueden estar relacionadas, por lo que se estaría incurriendo en redundancia al incluir variables correlacionadas entre sí. Frente a ello, la correlación de Pearson nos permite conocer la dirección y la intensidad de la relación entre variables (Pérez Monsalve, 2024). Debido a esto, en la Figura 5, se presenta el mapa de calor de las variables restantes en función a la correlación de Pearson.

Figura 5

Mapa de calor con correlaciones para las variables

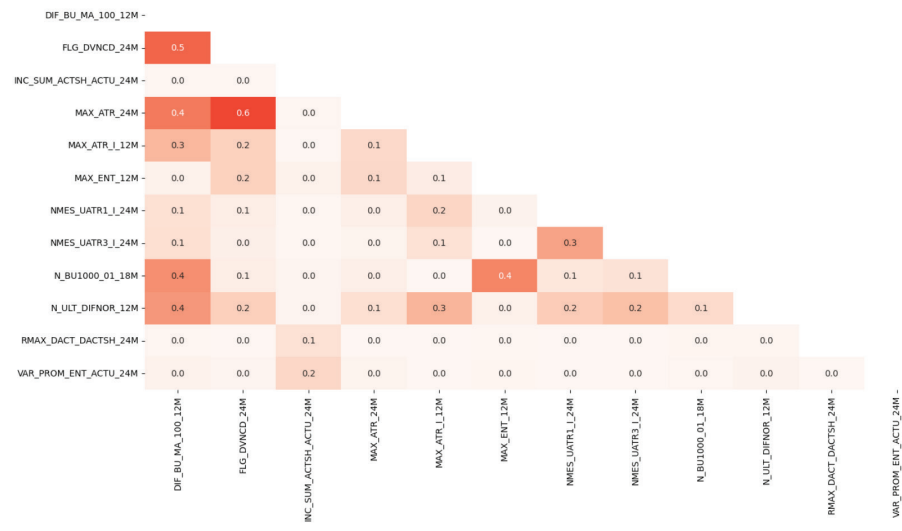


Nota. Los valores mostrados son absolutos para una mejor compresión del grado de correlación.

El mapa de calor (Figura 5) muestra que muchas variables están correlacionadas entre sí, por lo que es necesario fijar un umbral para eliminar la redundancia. Una correlación se considera moderada cuando es menor a 0,5 en valor absoluto (Polo Romero, 2024); sin embargo, se eligió 0,6 por dos motivos. El primero es que bajar a 0,5 hubiera eliminado variables que ya pasaron el filtro del coeficiente de Gini; es decir, variables con poder predictivo comprobado. El segundo es que tener variables muy correlacionadas complica el análisis bivariado, porque, al mostrar tasas de malos similares, se vuelve difícil distinguir qué aporta cada una al negocio. Por estas razones, el umbral de 0,6 equilibra la reducción de redundancia sin perder información relevante, lo que resulta en doce variables finalistas (Figura 6).

Figura 6

Mapa de calor con correlaciones para las variables finalistas



Nota. Estas variables representan el 27 % del set inicial luego de pasar por los diversos filtros establecidos.

Ingeniería de variables

En esta etapa se realizó un último análisis para observar la relación bivariada que existe entre las variables finalistas y el *target*; de ser el caso, sobre la base de este análisis, se realizó una imputación de datos con base en criterios de experto para las variables que contienen *missings*.

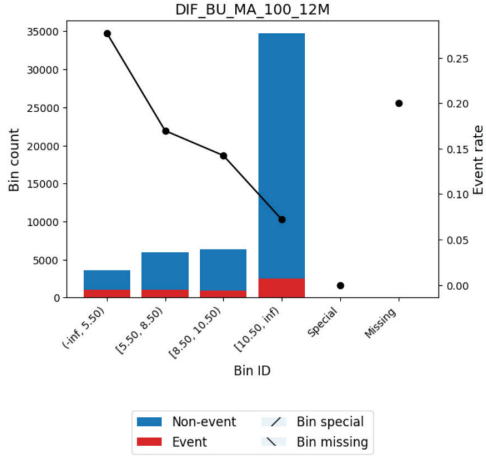
Análisis bivariado

Uno de los problemas que tiene el uso de algoritmos de *machine learning* es entender cómo interactúan las variables con la probabilidad predicha. Debido a esto, es fundamental

realizar un análisis para comprender el sentido que tienen; es decir, si premian o castigan la probabilidad de impago. El *binning* es una de las técnicas más usadas para agrupar variables, además nos permite conocer, de manera sencilla, el sentido monótono que tienen en el riesgo crediticio (Wheeler & Carrol, 2023). En las figuras 7 y 8, se pueden ver los cortes óptimos tomados por la librería a través del uso de un algoritmo de programación convexa que maximiza el *information value* (IV), bajo la hipótesis de que la variable es monótona.

Figura 7

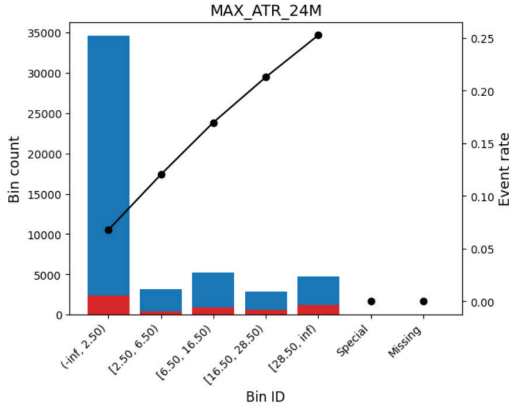
Gráfico bivariado para la variable DIF_BU_MA_100_12M



Nota. Se observa que, en esta variable relación inversa, a mayor valor de la variable, menor es el riesgo.

Figura 8

Gráfico bivariado para la variable MAX_ATR_24M



Nota. Se observa que, en esta variable relación directa, a mayor valor de la variable, mayor es el riesgo.

De este análisis, se obtuvieron las conclusiones que se observan en la Tabla 2. En esta se muestra la variable, su descripción y el sentido en relación con el *target*.

Tabla 2

Diccionario de variables finalistas

Fuente	Descripción	Interpretación	Sentido
DIF_BU_MA_100_12M	Diferencia entre meses de buen y mal comportamiento en los últimos doce meses	A mayor diferencia positiva (más meses buenos), menor riesgo.	Inverso
FLG_DVNCD_24M	Flag deuda vencida de los últimos veinticuatro meses	El valor de 1 indica mayor riesgo	Directo
INC_SUM_ACTSH_ACTU_24M	Incremento de la deuda total en los últimos veinticuatro meses	Si la deuda crece mucho, el riesgo aumenta; por tanto, el incremento alto empeora el perfil.	Inverso
MAX_ATR_24M	Máximo número de días de atraso en los últimos veinticuatro meses	Más días de atraso, mayor riesgo.	Directo
MAX_ATR_I_12M	Máximo atraso en los últimos doce meses	Un atraso más severo en el último año incrementa riesgo.	Directo
MAX_ENT_12M	Máximo de entidades acreedoras en los últimos doce meses	Más entidades implican mayor apalancamiento y probabilidad de sobreendeudamiento.	Directo
NMES_UATR1_I_24M	Meses desde el último atraso ≥ 1 día (últimos veinticuatro meses)	Cuanto más tiempo ha pasado desde un atraso, menor es el riesgo.	Inverso
NMES_UATR3_I_24M	Meses desde el último atraso ≥ 3 días (últimos veinticuatro meses)	Más meses sin atrasos relevantes, menor riesgo.	Inverso
N_BU1000_01_18M	Meses de buen comportamiento (últimos dieciocho meses)	Más meses sin atrasos reflejan mejor disciplina de pago y menor riesgo.	Inverso
N_ULT_DIFNOR_12M	Meses desde la última clasificación distinta a "normal" (últimos doce meses)	Más tiempo sin una mala clasificación indica recuperación y menor riesgo.	Inverso
RMAX_DACT_DACTSH_24M	Ratio entre la máxima deuda activa y la máxima deuda activa sin hipoteca (últimos veinticuatro meses)	Una mayor proporción implica niveles más altos de deuda activa, mayor riesgo	Directo
VAR_PROM_ENT_ACTU_24M	Variación en el número de entidades acreedoras (últimos dieciocho meses)	Una reducción en entidades indica des apalancamiento, aumento implica mayor riesgo.	Inverso

Nota. El orden directo indica que, a mayor valor, mayor es el riesgo, mientras que el inverso indica que, a mayor valor, menor es el riesgo, y viceversa.

Imputación de variables

Durante esa sección se detectó que existía un porcentaje ínfimo de variables con *missings* dentro de las variables finalistas (<1 %). Debido a esto se realizó una imputación por

criterio de experto, ya que los *missings* tienen un significado a nivel de negocio. En la Tabla 3, se aprecian las imputaciones realizadas sobre nuestras variables finalistas.

Tabla 3

Imputaciones de variable por criterio de experto

Grupo de variables	Fundamento	Valor imputado
Variables de comportamiento (atrasos, entradas, variaciones)	Sin ocurrencia de evento, sin registros. Ausencia de comportamiento riesgoso	Se imputa por el valor de 0, mantiene lógica de riesgo bajo.
Variables de tiempo o frecuencia	Para clientes nuevos	Imputado por 99, valor alto. Sin eventos.

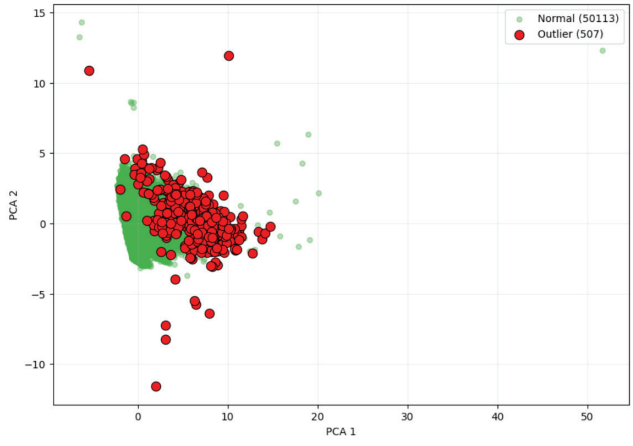
Nota. Tras realizar esta imputación tenemos un *dataset* con *missings* de 0 % para todos los valores.

Detección de outliers

Otro paso importante para contar con un *dataset* correcto es identificar los valores *outliers*. Los valores atípicos pueden tener un efecto significativo a la hora de obtener resultados (Dash et al., 2023). Para identificar estos valores atípicos, existe una multitud de técnicas apoyadas en conceptos estadísticos y otras más avanzadas que usan modelos de inteligencia artificial. Dentro de las técnicas más usadas y confiables, se cuenta con la de Isolation Forest. Este algoritmo no supervisado permite aislar y detectar *outliers*, y se caracteriza por su *performance* y confiabilidad en grandes conjuntos de datos (Banchero et al., 2021). En la Figura 9, se puede observar el número de *outliers* detectados en todo nuestro conjunto de entrenamiento, teniendo en cuenta que el grado de contaminación indicado es del 1 %.

Figura 9

Outliers *detectados en* dataset

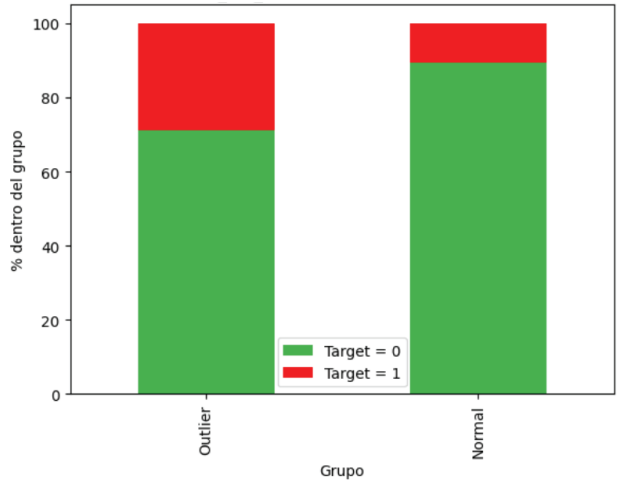


Nota. El gráfico mostrado se elaboró a partir de un aplanamiento de análisis de componentes principales (PCA) para visualizar en 2D los valores atípicos.

Si bien los algoritmos basados en árboles de decisión presentan una tolerancia natural a los valores atípicos, debido a que sus particiones se basan en ordenamiento y no en distancias, la eliminación de *outliers* responde a un criterio de calidad del *dataset*. En datos financieros, los valores atípicos suelen corresponder a errores de registro o comportamientos que no representan a la población objetivo de la cartera analizada, por lo que su inclusión puede introducir ruido en las relaciones entre variables y distorsionar el análisis bivariado. La Figura 10 respalda esta decisión, pues muestra que la distribución de malos y buenos es similar entre el grupo de *outliers* y el grupo normal, lo que confirma que su retiro no genera un sesgo en la variable objetivo y que el *dataset* resultante mantiene la representatividad de la cartera.

Figura 10

Distribución de target *en* outliers



Nota. Las distribuciones de malos rondan el 30 % y el 10 % para grupos *outliers* y normales, respectivamente. No existen grandes diferencias en su distribución.

Entrenamiento de modelo

Para establecer el conjunto de *train*/validación, se consideraron los meses comprendidos desde el 202301 hasta el 202312, mientras que para la base de fuera tiempo (*out of time*) se utilizó la data correspondiente del periodo 202401-202403. Con el fin de obtener resultados transparentes, se aplicó validación cruzada estratificada con un pliegue de *folds* igual a 5. La validación cruzada permite mejorar la robustez del modelo, lo que favorece su aplicación (Zuleta Fuerte et al., 2025). En la Tabla 4, se observan todos los algoritmos entrenados con sus parámetros por defecto, la validación cruzada y las métricas obtenidas.

Tabla 4

Métricas obtenidas para todos los modelos

Modelo	AUC_CV_ mean (%)	Gini_CV_ mean (%)	KS_CV_ mean (%)	AUC_ OOT (%)	Gini_ OOT (%)	KS_OOT (%)	Variación Gini
HistGradient-Boosting	82,9	65,8	51,0	83,7	67,3	52,0	0,1791
LightGBM	82,3	64,7	50,6	83,0	66,0	51,1	0,1829
GradientBoost-ing	82,5	65,0	50,5	82,8	65,6	50,3	0,1778
CatBoost	82,0	64,0	50,4	82,4	64,8	50,9	0,1844
AdaBoost	81,9	63,8	49,8	82,4	64,7	49,9	0,1862
LogisticRegres-sion	81,4	62,7	48,8	81,7	63,5	48,4	0,1900
XGBoost	80,5	61,0	47,6	81,4	62,8	48,7	0,2045
SVC	81,1	62,2	50,4	81,4	62,8	49,7	0,1923
SGDClassi-fier_log	78,0	56,1	43,6	80,6	61,1	45,9	0,2447
RandomForest	79,8	59,6	46,9	80,0	60,0	46,6	0,2043
ExtraTrees	78,7	57,5	45,4	79,3	58,6	45,1	0,2183
BernoulliNB	77,0	54,1	45,4	77,3	54,7	44,2	0,2329
GaussianNB	76,3	52,7	42,6	76,9	53,9	43,7	0,2424
Bagging_DT	75,4	50,7	41,8	75,2	50,4	45,0	0,2448
KNN	72,5	44,9	40,0	71,4	42,7	45,8	0,2645
DecisionTree	61,3	22,6	23,7	59,5	18,9	43,9	0,3684

Nota. Las métricas CV hacen referencia a las *cross validation,* mientras que la muestra OOT es el periodo que nunca se tomó en cuenta a la hora de entrenar/validar.

Según las métricas obtenidas, se observó que el algoritmo HistGradientBoosting obtuvo los mejores indicadores: un mejor rendimiento de casi el 2 % por encima del coeficiente de Gini en su OOT respecto del conjunto CV. Este algoritmo potencia su forma original de *boosting* utilizando la técnica del *binning*, lo que reduce la complejidad computacional y mejora la eficiencia del modelo (Ferro Rugeles et al., 2023). Además, este algoritmo se caracteriza por presentar una gran robustez y controlar de manera eficiente las interacciones entre variables. Como lo señala Betancourt Ludeña (2025), este algoritmo puede superar inclusive a redes neuronales profundas en problemas de clasificación y gana no solo en *performance*, sino también en interpretabilidad. En el presente estudio se optó por algoritmos de *machine learning* y no de *deep learning*, precisamente porque se quiso mejorar el concepto de interpretabilidad a la hora de implementar el modelo.

Optimización de hiperparámetros

Un paso crítico es ajustar (optimizar) el modelo. En la sección anterior comparamos los modelos bajo el supuesto de los hiperparámetros por defecto. Optuna es una librería que realiza una búsqueda inteligente para encontrar la mejor configuración de hiperparámetros y se basa en algoritmos de optimización de árboles en búsqueda de bandas (Martínez Martínez, 2024). En la Figura 11, se aprecian los hiperparámetros que se buscaron dentro de la optimización, así como el espacio de búsqueda.

Figura 11

Búsqueda de hiperparámetros mediante Optuna

```
def objective(trial):
    params = {
        "learning_rate": trial.suggest_float("learning_rate", 0.01, 0.3, log=True),
        "max_iter": trial.suggest_int("max_iter", 100, 500),
        "max_depth": trial.suggest_int("max_depth", 2, 8),
        "min_samples_leaf": trial.suggest_int("min_samples_leaf", 20, 200),
        "max_bins": trial.suggest_int("max_bins", 64, 255),
        "l2_regularization": trial.suggest_float("l2_regularization", 1e-4, 10.0, log=True),
        "random_state": SEED,
        "early_stopping": False,
        "loss": "log_loss"
    }
    kf = StratifiedKFold(
        n_splits=N_SPLITS,
        shuffle=True,
        random_state=SEED
    )
    gini_folds = []
    for train_idx, val_idx in kf.split(X_train, y_train):
        X_tr = X_train.iloc[train_idx]
        X_val = X_train.iloc[val_idx]
        y_tr = y_train.iloc[train_idx]
        y_val = y_train.iloc[val_idx]
        sw_tr = np.where(y_tr == 1, w_pos, w_neg)
        model = HistGradientBoostingClassifier(**params)
        model.fit(X_tr, y_tr, sample_weight=sw_tr)
        proba_val = model.predict_proba(X_val)[:, 1]
        gini = gini_score(y_val, proba_val)
        gini_folds.append(gini)
    mean_gini = float(np.mean(gini_folds))
    return mean_gini
```

Nota. El espacio de búsqueda fue definido a partir de criterio experto.

Tras realizar la búsqueda con Optuna en cincuenta escenarios, se obtuvieron los hiperparámetros definidos en la Tabla 5. Si bien cincuenta iteraciones parecen un número reducido, Optuna no realiza una búsqueda aleatoria, sino una optimización bayesiana que aprende de cada evaluación anterior para dirigir la búsqueda hacia zonas de mayor rendimiento del espacio de hiperparámetros (Martínez Martínez, 2024). Esto implica que cincuenta iteraciones son suficientes para alcanzar la convergencia en espacios de búsqueda de dimensión moderada, como el definido en este estudio. Los resultados respaldan esta decisión: la métrica del coeficiente de Gini en la CV mejoró casi un punto porcentual respecto del modelo con parámetros por defecto, mientras que en la muestra OOT se mantuvo en 67,3 %, lo que indica que la búsqueda encontró una configuración estable y de buen rendimiento.

Tabla 5

Mejores hiperparámetros obtenidos por Optuna

Parámetro	Descripción	Valor
learning_rate	Controla la velocidad con la que el modelo aprende en cada iteración; valores más bajos hacen el aprendizaje más gradual.	0,106
max_iter	Número máximo de iteraciones que realizará el algoritmo para ajustar el modelo.	125
max_depth	Profundidad máxima de cada árbol; limita la complejidad y ayuda a evitar sobreajuste.	2
min_samples_leaf	Mínimo de muestras requeridas en una hoja para asegurar que los cortes no se basen en grupos pequeños.	180
max_bins	Máxima cantidad de bins usados para discretizar valores continuos antes del entrenamiento.	207
l2_regularization	Penalización L2 que ayuda a estabilizar el modelo y reducir sobreajuste.	0,.0005228

Nota. Los hiperparámetros tuneados son los que mayormente se personalizan y son los más relevantes.

Adicionalmente, para el tratamiento del desbalance de clases identificado en el apartado "Dataset" de la sección "Metodología", se configuró el parámetro class_weight = 'balanced' del algoritmo HistGradientBoosting. Este parámetro ajusta automáticamente los pesos de cada clase según su frecuencia en el *dataset*, lo que penaliza más los errores sobre la clase minoritaria (malos) sin necesidad de aplicar técnicas externas de remuestreo como SMOTE u *oversampling*.

Importancia de variables

También es importante identificar las variables que más contribuyen a la predicción. En tal sentido, el modelo se caracteriza por ser de alto rendimiento, pero de poca interpretabilidad. Una de las técnicas para lograr la explicabilidad de nuestros modelos es SHAP. Esta está basada en la teoría de juegos de manera cooperativa y otorga el valor marginal que tiene cada variable a la hora de predecir (Buergo Lagunas, 2025). Este método tiene un fundamento matemático, por lo que es de implementación rápida y, sobre todo, es muy interpretable. Al usar la técnica de SHAP, la institución financiera que interpreta el modelo es capaz de conocer la influencia de cada variable a la hora de dar un seguimiento (Marchant Contreras, 2022). En la Tabla 6, se muestra el *ranking* de las variables según la importancia SHAP.

Tabla 6

Importancia SHAP del modelo ajustado

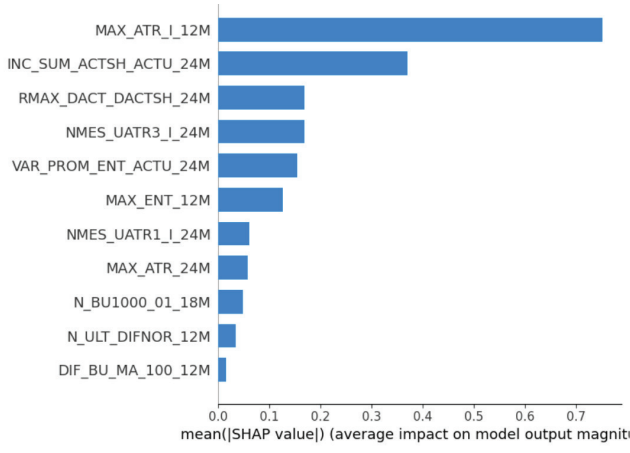
Variable	Importancia SHAP	SHAP relativo (%)
MAX_ATR_I_12M	0,752125	38,5
INC_SUM_ACTSH_ACTU_24M	0,369670	18,9
RMAX_DACT_DACTSH_24M	0,168870	8,6
NMES_UATR3_I_24M	0,168777	8,6
VAR_PROM_ENT_ACTU_24M	0,154431	7,9
MAX_ENT_12M	0,125298	6,4
NMES_UATR1_I_24M	0,059864	3,1
MAX_ATR_24M	0,056815	2,9
N_BU1000_01_18M	0,048253	2,5
N_ULT_DIFNOR_12M	0,033692	1,7
DIF_BU_MA_100_12M	0,014554	0,7
FLG_DVNCD_24M	0	0,0

Nota. El valor de SHAP relativo es el cálculo producto de la suma del valor SHAP entre la suma de todos valores SHAP.

En la Tabla 6, se observa que la variable flag_dvncd_24m no tiene importancia en la predicción, por lo que resulta innecesaria dentro del *dataset*. Debido a esto, se procedió a retirar esta variable y a reentrenar el modelo con los mismos hiperparámetros, pero solo con once variables, conservando las mismas métricas mostradas anteriormente. En la Figura 12, se aprecia el *ranking* de las variables SHAP, acompañado de sus respectivos valores.

Figura 12

Diagrama de barras de importancia SHAP

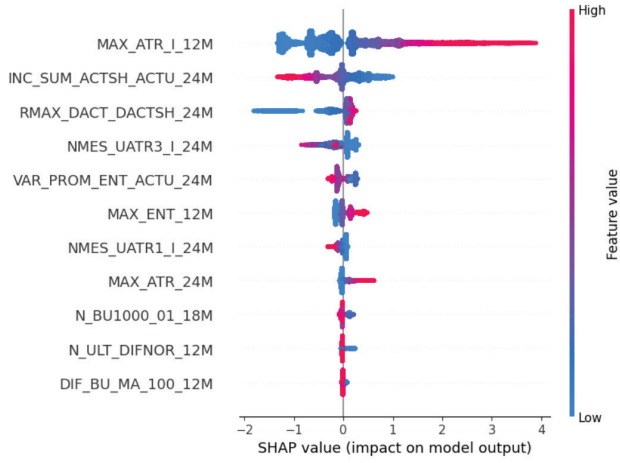


Nota. La variable más importante para nuestro modelo es la de máximo días de atraso en los últimos doce meses (top 1).

En la Figura 13, se observa la influencia que tiene cada característica en la predicción de manera resumida, mientras que, en las figuras 14 y 15, se aprecia su contribución con mayor detalle. Los gráficos de dependencia parcial (DP) permiten visualizar el impacto de las variables en la probabilidad, ya que esta técnica usa la contribución individual de cada variable y su relación con la variable objetivo (Jaimes et al., 2023).

Figura 13

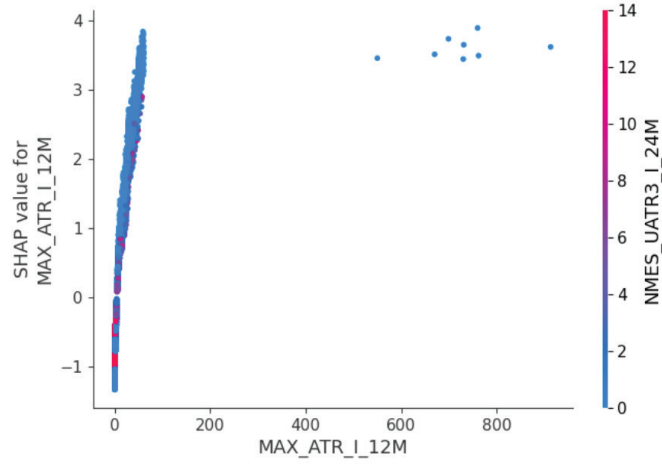
Diagrama de barras de importancia SHAP



Nota. El color rojo representa mayor probabilidad de riesgo y el azul representa menor probabilidad.

Figura 14

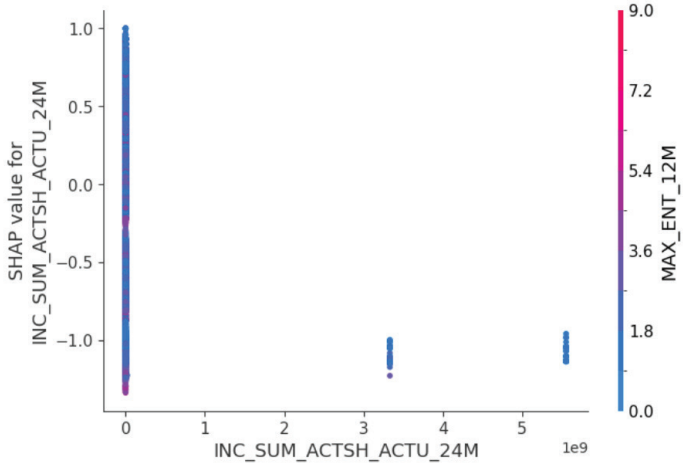
Diagrama de dependencia parcial variable Max_atr_i_12m



Nota. La figura nos indica que, a mayor valor de la variable, la probabilidad de caer en *default* aumenta.

Figura 15

Diagrama de dependencia parcial variable Inc_sum_actsh_actu_24m



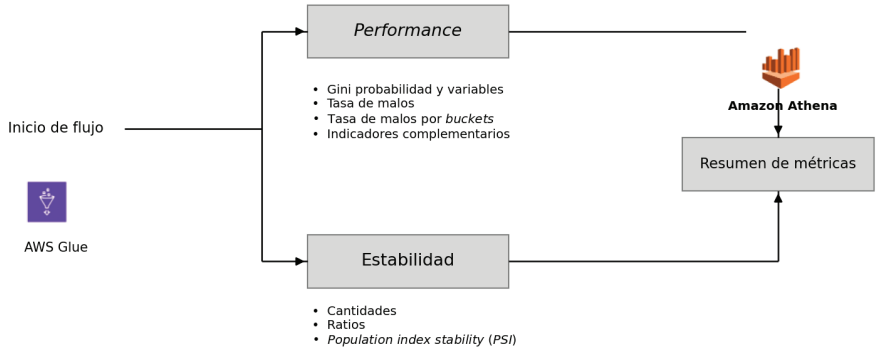
Nota. La figura nos indica que, a menor valor de la variable, aumenta la probabilidad de caer en *default*.

Seguimiento del modelo (monitoreo)

Para poder medir la salud de nuestro modelo, se construyó una serie de *pipelines* que evalúan el modelo en función de dos frentes: el *performance* y la estabilidad de la variable a predecir (probabilidad de *default*) y las variables predictoras. En la Figura 16, se aprecia el flujo seguido para el cálculo de métricas que fue desplegado en ambiente de Amazon Web Services (AWS).

Figura 16

Arquitectura general del flujo para monitoreo



Nota. Los servicios fueron orquestados mediante Step Function.

Cálculo de indicadores de referencias

Un punto clave para realizar un correcto monitoreo es establecer una referencia, es decir, identificar el *dataset* con el que se estará comparando constantemente el modelo. Para el presente estudio, se escogió como *dataset* de referencia todo el conjunto de *train* del periodo 202301-202312. A partir de este, se tomaron los indicadores de *performance* como el valor base, así como la cantidad y los porcentajes de las variables, los que representan los parámetros utilizados para el seguimiento mensual. En la Tabla 7, se observa el detalle del cálculo de referencia obtenido para realizar el seguimiento/monitoreo del modelo.

Tabla 7

Cálculo de referencia

Tipo de monitoreo	Descripción	Valor
Performance	Gini referencia	68,4 %
	Tasa de malos referencia	10,5 %
	Indicadores complementarios	*Acuraccy*: 79 Sensibilidad: 73 Especificidad: 80 %
Estabilidad	*Score* (probabilidad transformada)	Se hace cortes en deciles
	Variables	Se hacen cortes con base en los cortes óptimos del *optimal binning* o en su defecto cortes por quintiles

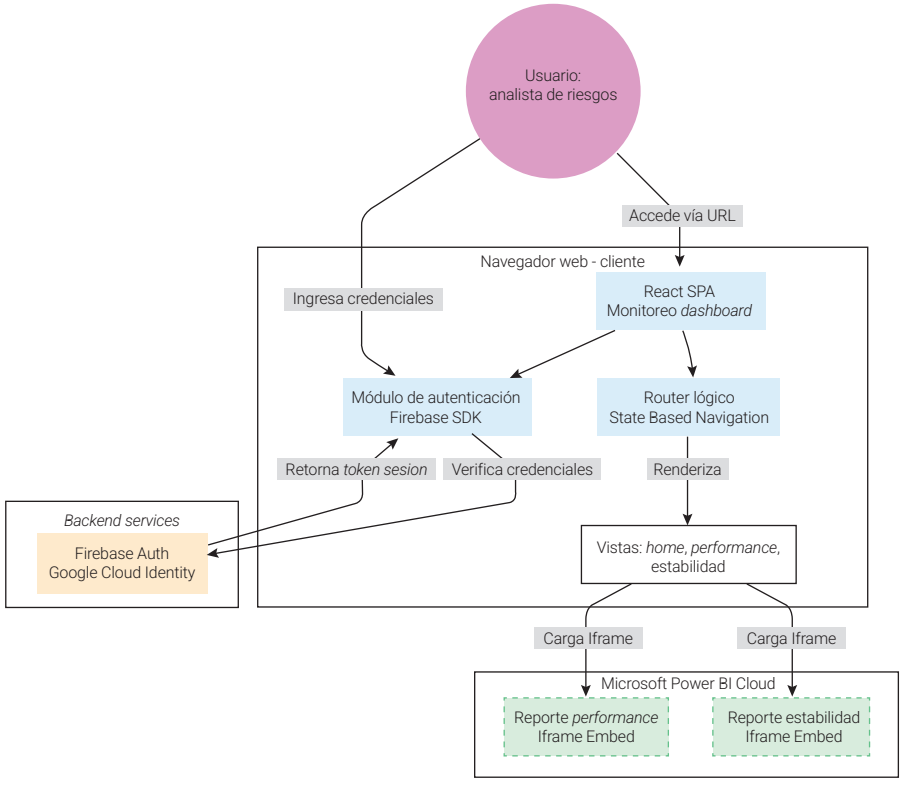
Nota. El *score* es la transformación de la probabilidad, representa solo un escalado de su complemento, que oscila entre 0-1000, donde un número mayor indica menor riesgo. Se realizó este escalado para comprender mejor su interpretación.

Despliegue de la plataforma de seguimiento/monitoreo

Para poder realizar un monitoreo integral del modelo, se desplegó una plataforma que permitiera realizar una interacción directa con los resultados en tiempo real. En la Figura 17, se aprecia la arquitectura de la app web desplegada.

Figura 17

Arquitectura general de la app web desplegada



Nota. La aplicación fue desplegada en Vercel.

RESULTADOS

Cabe precisar que el presente estudio se enfoca en la evaluación del desempeño estadístico y la estabilidad del modelo, por lo que la cuantificación del impacto financiero, como la estimación de pérdida esperada o beneficios operativos derivados de su implementación, queda fuera del alcance de esta investigación y se plantea como una línea de trabajo futuro.

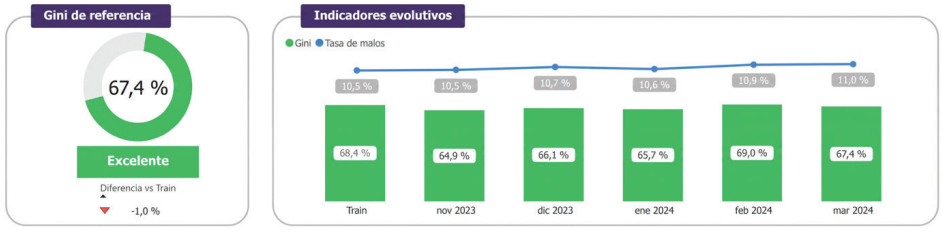
Performance general

Para verificar los resultados obtenidos del modelo, se analizaron los reportes desplegados en la plataforma web de monitoreo. En la Figura 18, se puede observar que, para los últimos meses del *dataset*, el coeficiente de Gini se mantuvo en un 67,4 %. Si comparamos este indicador versus nuestra referencia de *train*, solo es un punto porcentual por

debajo. Este Gini puede ser catalogado como excelente, ya que supera el umbral del 60 % (Putri et al., 2021). Además de tener este indicador sobresaliente, se evidencia que es constante a lo largo del tiempo con ligeras oscilaciones, lo cual indica que es un modelo robusto.

Figura 18

Métricas de performance *del modelo final*



Nota. Diagrama extraído de la plataforma web.

Performance complementario

También es importante verificar que el modelo cuente con una estabilidad en las métricas derivadas de la matriz de confusión (métricas auxiliares). En la Figura 19, se aprecia que las métricas auxiliares se mantuvieron muy similares a lo evidenciado en el *train*, lo que refuerza lo mencionado anteriormente: el modelo cuenta con una robustez importante a lo largo del tiempo. La métrica de *accuracy* que nos indica el porcentaje de acierto del modelo se mantiene en aproximadamente el 80 % en el último mes, con tendencia a subir.

Figura 19

Métricos de auxiliares del modelo final

Indicadores de la matriz de confusión

Etiqueta_Indicador	Train	nov 2023	dic 2023	ene 2024	feb 2024	mar 2024
Accuracy	79%	78%	79%	79%	79%	80%
Especificidad	80%	79%	80%	80%	80%	81%
F1 Score	42%	40%	41%	42%	44%	43%
Odds	10,5	8,4	8,8	9,9	11,1	9,9
Prec. Neg.	96%	96%	96%	96%	96%	96%
Prec. Pos.	29%	28%	29%	30%	31%	31%
Sensibilidad	73%	69%	69%	71%	74%	70%

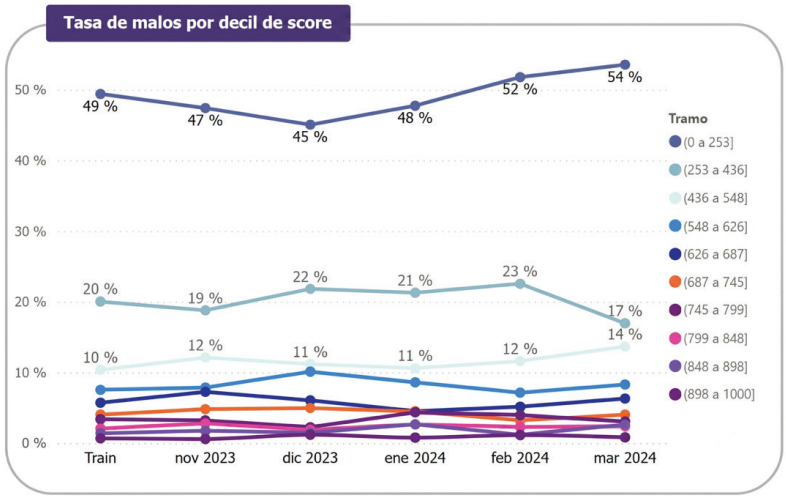
Nota. Diagrama extraído de la plataforma web.

Métricas de negocio

En las secciones anteriores se analizaron métricas de un carácter más estadístico y matemático; sin embargo, no se debe dejar de lado las métricas de negocio. En su mayoría están asociadas a ver si la diferenciación de los clientes predichos como malos y buenos se sigue manteniendo como estaba en *train*, y si, además de esto, se sigue manteniendo un ordenamiento coherente en función del sentido de la variable de *score*. En la Figura 20, se aprecia el evolutivo de la tasa de malos por decil de *score*, ordenados de menor a mayor.

Figura 20

Tasa de malos por decil de cortes de score

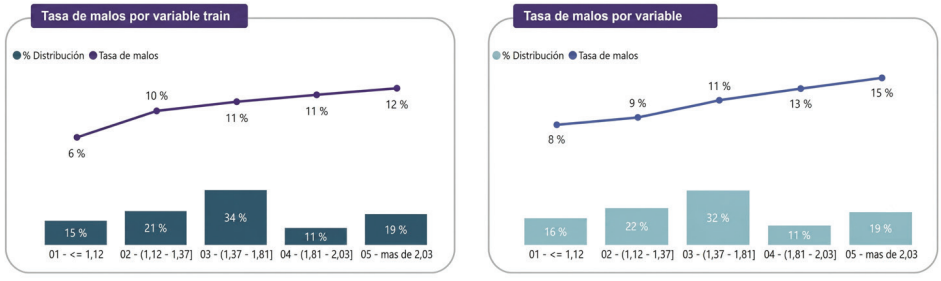


Nota. Diagrama extraído de la plataforma web.

Así como se vio el resultado de la predicción (*score*), también es importante analizar cada variable. En la Figura 21, se aprecia la tasa de malos para la variable más importante (top 1) en su periodo de *train* y para el periodo 202403 que corresponde al último monitoreo. Se observa que, a pesar de que han pasado varios meses, la tendencia de la variable mantiene sentido directo, es decir, a mayor valor, mayor probabilidad de riesgo, lo cual es resultado de la correcta selección de variables realizada en las secciones anteriores.

Figura 21

Tasa de malos de la variable top 1 - ratio de deuda máxima y máxima deuda activa sin hipotecario



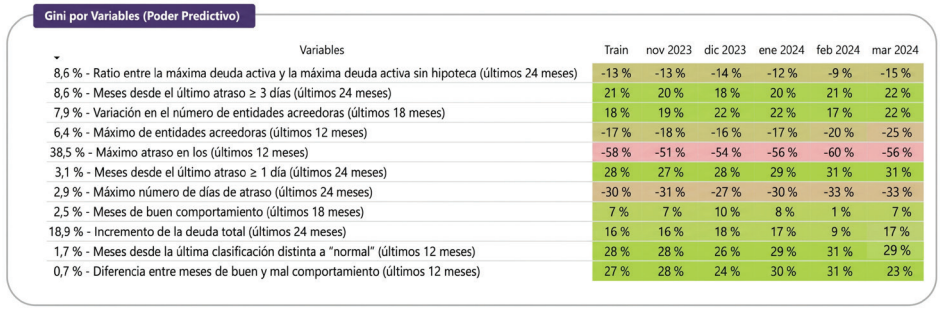
Nota. Diagrama extraído de la plataforma web.

Poder predictivo de variables

Como último frente dentro de la *performance*, también se evaluó que se mantuviera el poder predictivo de las variables. En la Figura 22, se observa que, a lo largo del tiempo, las variables mantuvieron, en su mayoría, valores cercanos a los de *train*, oscilando en hasta ±4 puntos versus su versión de *train*. Esto es un indicador de que el modelo viene performando bien y que no existe ningún cambio de relación con el *target*, lo que garantiza un modelo robusto, interpretable y que refleja confianza en los usuarios finales.

Figura 22

Evolutivo de Gini por variables ordenadas por peso



Variables	Train	nov 2023	dic 2023	ene 2024	feb 2024	mar 2024
8,6 % - Ratio entre la máxima deuda activa y la máxima deuda activa sin hipoteca (últimos 24 meses)	-13 %	-13 %	-14 %	-12 %	-9 %	-15 %
8,6 % - Meses desde el último atraso ≥ 3 días (últimos 24 meses)	21 %	20 %	18 %	20 %	21 %	22 %
7,9 % - Variación en el número de entidades acreedoras (últimos 18 meses)	18 %	19 %	22 %	22 %	17 %	22 %
6,4 % - Máximo de entidades acreedoras (últimos 12 meses)	-17 %	-18 %	-16 %	-17 %	-20 %	-25 %
38,5 % - Máximo atraso en los (últimos 12 meses)	-58 %	-51 %	-54 %	-56 %	-60 %	-56 %
3,1 % - Meses desde el último atraso ≥ 1 día (últimos 24 meses)	28 %	27 %	28 %	29 %	31 %	31 %
2,9 % - Máximo número de días de atraso (últimos 24 meses)	-30 %	-31 %	-27 %	-30 %	-33 %	-33 %
2,5 % - Meses de buen comportamiento (últimos 18 meses)	7 %	7 %	10 %	8 %	1 %	7 %
18,9 % - Incremento de la deuda total (últimos 24 meses)	16 %	16 %	18 %	17 %	9 %	17 %
1,7 % - Meses desde la última clasificación distinta a "normal" (últimos 12 meses)	28 %	28 %	26 %	29 %	31 %	29 %
0,7 % - Diferencia entre meses de buen y mal comportamiento (últimos 12 meses)	27 %	28 %	24 %	30 %	31 %	23 %

Nota. Diagrama extraído de la plataforma web.

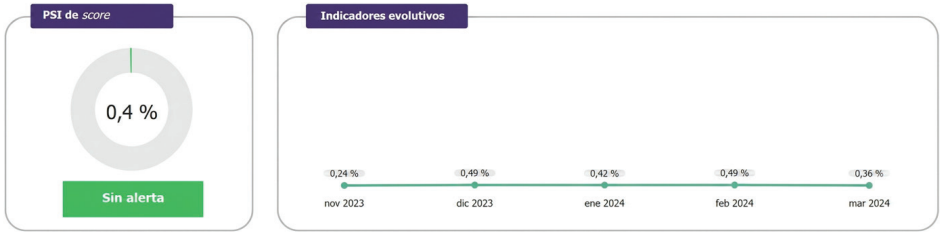
Estabilidad general

Ya se ha comprobado que el modelo está mostrando un desempeño satisfactorio, pero no se sabe aún si su población ha cambiado con el paso del tiempo. Para ello, es necesario

usar el indicador de PSI para establecer umbrales a partir de los cuales se generará una alerta. Un valor menor que 10 % indica que no existe cambio, valores mayores que 10 % y menores que 25 % indican pequeños cambios, y valores mayores que 25 % reflejan cambios significativos (Reyes Morales & Sosa Castro, 2022). En la Figura 23, se aprecia que el valor de PSI para la variable de predicción (*score*) se ha mantenido en valores cercanos a 0 %, lo que indica una estabilidad notable y se cataloga como una situación sin alerta.

Figura 23

Evolutivo PSI del score

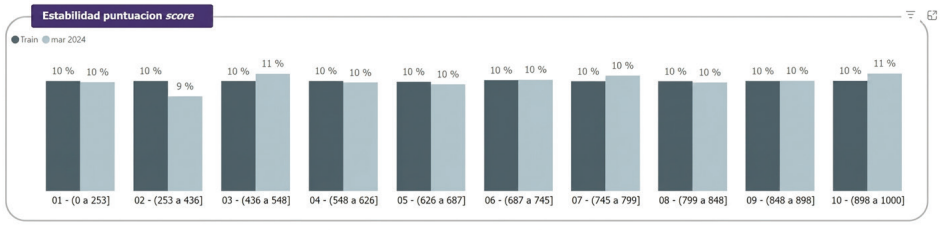


Nota. Diagrama extraído de la plataforma web.

La Figura 24 es consistente con lo mencionado anteriormente. Al observar la distribución, se aprecia que los valores son muy similares a los de la muestra de *train*, por lo que indica que la población no ha sufrido cambios significativos.

Figura 24

Distribución del score *vs.* train



Nota. Diagrama extraído de la plataforma Web.

Así como se han observado los resultados del *score*, es importante verificar la estabilidad de las variables predictoras. En la Figura 25, se observa un cuadro resumen del PSI para todas las variables, en el que ninguna de estas presenta alerta. En línea con esta idea para la variable más importante (top 1), en la Figura 26 se presenta su distribución, la cual tiene solo pequeñas variaciones con respecto a la de *train*.

Figura 25

PSI por variables ordenadas de mayor a menor peso

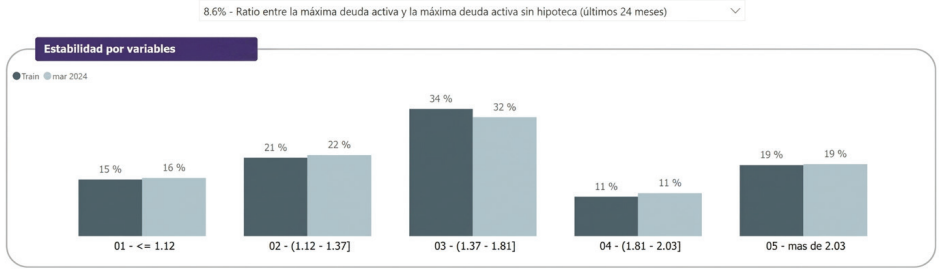
PSI por variables

Variables	nov 2023	dic 2023	ene 2024	feb 2024	mar 2024
8,6 % - Ratio entre la máxima deuda activa y la máxima deuda activa sin hipoteca (últimos 24 meses)	0,23 %	0,22 %	0,05 %	0,27 %	0,24 %
8,6 % - Meses desde el último atraso ≥ 3 días (últimos 24 meses)	0,22 %	0,11 %	0,15 %	0,28 %	0,73 %
7,9 % - Variación en el número de entidades acreedoras (últimos 18 meses)	0,05 %	0,13 %	0,19 %	0,04 %	0,06 %
6,4 % - Máximo de entidades acreedoras (últimos 12 meses)	0,00 %	0,16 %	0,05 %	0,13 %	0,08 %
38,5 % - Máximo atraso en los (últimos 12 meses)	0,20 %	0,05 %	0,28 %	0,08 %	0,78 %
3,1 % - Meses desde el último atraso ≥ 1 día (últimos 24 meses)	0,35 %	0,82 %	0,15 %	0,29 %	0,38 %
2,9 % - Máximo número de días de atraso (últimos 24 meses)	0,03 %	0,06 %	0,02 %	0,01 %	0,00 %
2,5 % - Meses de buen comportamiento (últimos 18 meses)	0,01 %	0,03 %	0,16 %	0,09 %	0,11 %
18,9 % - Incremento de la deuda total (últimos 24 meses)	0,01 %	0,14 %	0,15 %	0,14 %	0,04 %
1,7 % - Meses desde la última clasificación distinta a 'normal' (últimos 12 meses)	0,01 %	0,00 %	0,08 %	0,00 %	0,00 %
0,7 % - Diferencia entre meses de buen y mal comportamiento (últimos 12 meses)	0,10 %	0,29 %	0,07 %	0,13 %	0,18 %

Nota. Diagrama extraído de la plataforma web.

Figura 26

Distribución de la variable top 1



Nota. Diagrama extraído de la plataforma web.

DISCUSIONES

El coeficiente de Gini de 67,4 % obtenido en el periodo de seguimiento es competitivo dentro del contexto del *scoring* crediticio con datos reales de una sola institución. Estudios como el de Suhadolnik et al. (2023), en el que se trabajó con más de 2,5 millones de registros de una cartera diversificada, reportaron métricas superiores con XGBoost, lo cual es esperable dado el volumen y la variedad de datos disponibles. En cambio, este estudio se realizó con 50 206 registros de un portafolio específico, lo que restringe naturalmente el techo de rendimiento, pero, a la vez, hace que el modelo sea más representativo del comportamiento real de esa cartera. El resultado es coherente con lo reportado por De Lange et al. (2022), quienes también encontraron que los modelos desarrollados a partir de carteras reales y acotadas producen métricas más modestas pero más estables en producción.

Respecto de la explicabilidad, los valores SHAP identificaron el máximo atraso en los últimos doce meses como la variable de mayor peso, lo cual es consistente con la literatura. Castañeda Hilario et al. (2025) encontraron que la morosidad acumulada domina la predicción en carteras peruanas y De Lange et al. (2022) reportaron que el historial de comportamiento de pago es determinante en bancos europeos. El hecho de que dos contextos tan distintos converjan en el mismo tipo de variable refuerza la idea de que el modelo captura una señal genuina y no un artefacto del *dataset*.

El aspecto más diferenciador de este trabajo frente a la literatura revisada es el monitoreo automatizado. La mayoría de estudios sobre *scoring* crediticio concluyen con la validación del modelo en una muestra *out of time*, pero no abordan qué pasa después del despliegue. Bhagavathula (2025) y Shcherbinin (2026) tratan el monitoreo como problema de infraestructura genérica, mientras que este estudio lo integró directamente al ciclo de vida del modelo crediticio bajo el marco regulatorio de la SBS, lo que representa una contribución aplicada concreta.

No obstante, el estudio tiene limitaciones que deben reconocerse. La validación se realizó con datos de una sola institución financiera peruana, lo que limita la generalización de los resultados a otras carteras o entidades. El periodo de monitoreo abarca solo tres meses, un intervalo insuficiente para capturar ciclos económicos adversos o eventos de estrés como los que ocurrieron durante la pandemia. Finalmente, el modelo fue evaluado en condiciones de relativa estabilidad macroeconómica, por lo que su comportamiento ante escenarios de crisis continúa siendo una pregunta abierta para investigaciones futuras.

CONCLUSIONES

La presente investigación logró desarrollar e implementar un sistema integral de calificación crediticia basado en el algoritmo HistGradientBoosting, optimizado mediante técnicas de búsqueda bayesiana. El modelo final alcanzó un coeficiente de Gini del 67,4 % en la etapa de seguimiento, manteniendo una diferencia mínima de un punto porcentual respecto al conjunto de entrenamiento. Este nivel de desempeño, catalogado como excelente, demuestra que la adopción de algoritmos de aprendizaje automático tipo *boosting*, combinada con una rigurosa ingeniería de características mediante *binning*, supera en robustez y precisión a los enfoques tradicionales, sin sacrificar la estabilidad temporal del riesgo calculado.

Los resultados evidenciaron que el equilibrio entre poder predictivo e interpretabilidad es alcanzable mediante el uso de valores SHAP, lo cual permitió identificar que el comportamiento histórico de pagos —específicamente, el máximo atraso en los últimos doce meses— constituye el factor determinante en la predicción del *default*. Asimismo, la implementación de un esquema de monitoreo continuo validó la estabilidad del modelo,

registrando un índice de estabilidad poblacional (PSI) de 0,004 (0,4 %) para el *score*, cifra que se ubica muy por debajo del umbral de alerta de 0,10. Esto confirma que, a pesar de la volatilidad del entorno económico, la estructura de las variables predictoras y la distribución del riesgo se mantuvieron constantes durante el periodo de evaluación.

Más allá de la precisión métrica, este estudio revela el valor práctico de integrar el ciclo de vida del modelo con una arquitectura tecnológica moderna basada en servicios en la nube (AWS) y visualización web. La capacidad de desplegar tableros de control en tiempo real para la vigilancia de métricas de *performance* (coeficiente de Gini, tasa de malos) y estabilidad (PSI) responde directamente a las exigencias regulatorias de la SBS sobre la gestión de riesgos de modelos. Esta aplicación permite a las instituciones financieras pasar de revisiones estáticas y periódicas a una gestión de riesgos dinámica y proactiva, lo que facilita la detección temprana de desviaciones o *drifts* en la cartera.

Como líneas de trabajo futuro se identifican cuatro frentes. Primero, incorporar métricas de impacto financiero, tales como la pérdida esperada y el beneficio operativo derivado de la correcta clasificación de clientes, lo que permitiría traducir el rendimiento estadístico del modelo en valor tangible para la institución. Segundo, someter el modelo a pruebas de estrés bajo escenarios macroeconómicos adversos, como crisis financieras o periodos de alta morosidad, a fin de evaluar su robustez fuera de condiciones estables. Tercero, explorar el monitoreo de variables de negocio adicionales como tasas de aprobación, rentabilidad por segmento y tasa de recuperación, para ampliar el seguimiento más allá de las métricas estadísticas. Cuarto, evaluar la incorporación de fuentes de datos no tradicionales para enriquecer el perfilamiento de clientes con historial crediticio limitado y explorar mecanismos de reentrenamiento automático activados por las alertas del sistema de monitoreo, lo que permitiría completar el ciclo de MLOps.

Finalmente, este trabajo documenta que los modelos de *machine learning* de tipo *boosting* superan a la regresión logística clásica en el contexto bancario peruano y establecen un marco metodológico replicable para su auditoría y explicabilidad. Los resultados muestran que es posible implementar algoritmos de alta complejidad cumpliendo con los principios de transparencia y gestión de riesgos exigidos por el regulador, siempre que el ciclo de vida del modelo incluya la interpretabilidad, una validación robusta y un monitoreo continuo en producción.

REFERENCIAS

Aguirre Felix Díaz, I., Argomedo Sotelo, G. Y., Monzon Ñañez, J. A., & Tuesta Izaguirre, C. A. (2022). *Impacto de la adopción de inteligencia artificial como estrategia de negocio en las empresas del sector servicios durante la época de pandemia en el Perú* [Tesis de maestría, Pontificia Universidad Católica del Perú]. Repositorio Institucional PUCP. http://hdl.handle.net/20.500.12404/21241

Banchero, S., Verón, S., Petek, M., Sarrailhe, S., & De Abelleyra, D. (2021). *Detección de outliers en muestras de entrenamiento generadas mediante interpretación visual* [Documento de conferencia]. 50 Jornadas Argentinas de Informática e Investigación Operativa, XIII Congreso de Agroinformática (CAI 2021), Argentina. https://sedici.unlp.edu.ar/bitstream/handle/10915/140745/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

Betancourt Ludeña, I. C. (2025). *Aplicación de modelos de predicción para identificar patrones de abandono escolar en los datos de unidades educativas del sistema nacional* [Tesis de maestría, Universidad de las Américas]. Repositorio Digital UDLA. https://dspace.udla.edu.ec/handle/33000/18132

Bhagavathula, M. B. (2025). ML pipeline monitor: A comprehensive OpenTelemetry—Based observability platform for production machine learning systems. *International Journal of AI, BigData, Computational and Management Studies*, *6*(2), 109-112. https://doi.org/10.63282/3050-9416.IJAIBDCMS-V6I2P112

Buergo Lagunas, J. R. (2025). *Cálculo eficiente de SHAP para modelos de redes bayesianas gaussianas* [Tesis de maestría, Universidad de Cantabria]. Repositorio Abierto UCrea. https://repositorio.unican.es/xmlui/handle/10902/37755

Carbó-Valverde, S., & Rodríguez-Fernández, F. (2024). *Inteligencia artificial en banca: situación y perspectivas*. Funcas. https://www.funcas.es/libro/inteligencia-artificial-en-banca-situacion-y-perspectivas/

Castañeda Hilario, A., Mas Azahuanche, G. A., Mendoza Arenas, R. D., Castillo Paredes, O. T. A., Reinoso Palacios, A. R., Delgado Baltazar, M. P., & Mendoza Delgado, R. S. (2025). Explainable machine learning for credit card default prediction using web-scraped financial data: A case study in the Peruvian banking sector. En *Entrepreneurship with purpose: Social and Technological Innovation in the age of IA* (pp. 1-10). LACCEI. *https://laccei.org/LEIRD2025-Cartagena/meta/FP451.html*

Cepeda Cavero, L. E., & Véliz Soto, M. P. (2025). Ética y gobernanza algorítmica en el sector financiero: revisión sistemática de principios, gobernanza y brechas de implementación. *Interfases*, (22), 159-184. https://doi.org/10.26439/interfases2025.n022.8230

Cervantes Artavia, J., Monge Cordonero, M., & Sabater Guzmán, D. (2023). *Score crediticio con regresión logística y máquinas de soporte vectorial*. Universidad de Costa Rica, Escuela de Matemática. https://serengueti.fce.ucr.ac.cr/images/2023/03/16/Articulos/Score_crediticio_con_Regresion_Logistica_y_Maquinas_de_Soporte_Vectorial.pdf

Cordova Benites, L. S., Gonzales Collantes, L. Y., Leiva Cruz, F. D. R., & Navarro Espinoza, L. E. (2024). *Aplicación de la inteligencia artificial a la evaluación crediticia en el sistema financiero peruano* [Tesis de maestría, Universidad ESAN]. Repositorio Institucional ESAN. https://hdl.handle.net/20.500.12640/4249

Cosme Cervera, S., Rodríguez Arellano, S., Rocha Barrón, L. R., & Angel Angel, J. D. J. (2025, noviembre). Qué tan vulnerable es la regresión lineal ante los datos del mundo real. En D. Tinoco Varela (Ed.), *Memorias del Congreso Estudiantil de Inteligencia Artificial Aplicada a la Ingeniería y Tecnología* (pp. 118-124). Universidad Nacional Autónoma de México. https://www.researchgate.net/profile/Jose-Angel-Angel/publication/397461440

Dar, U., & Cavus, M. (2026). *From XAI to MLOps: Explainable concept drift detection with profile drift detection*. arXiv. *https://doi.org/10.48550/arXiv.2412.11308*

Dash, C. S. K., Behera, A. K., Dehuri, S., & Ghosh, A. (2023). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal*, *6*, 100164. https://www.sciencedirect.com/science/article/pii/S2772662223000048

De Lange, P. E., Melsom, B., Vennerød, C. B., & Westgaard, S. (2022). Explainable AI for credit assessment in banks. *Journal of Risk and Financial Management*, *15*(12), 556. https://doi.org/10.3390/jrfm15120556

Dos Santos, A. B., Dos Santos, I. D. B., & Toma, K. M. (2025). Detecção de transações fraudulentas com cartão de crédito: uma abordagem comparativa baseada em aprendizado de máquina. *Revista Foco*, *18*(5), e8328. https://doi.org/10.54751/revistafoco.v18n5-037

Enriquez Montes, J., & Orbezo Huamancari, D. A. M. (2025). *Modelo de* scoring *de clientes para la gestión eficiente de productos financieros mediante regresión logística y* data analytics*: un enfoque moderno para la industria financiera en Perú* [Trabajo de suficiencia profesional, Universidad Peruana de Ciencias Aplicadas]. Repositorio Académico UPC. https://repositorioacademico.upc.edu.pe/handle/10757/684767

Erazo Laínez, G. (2024). *Machine learning para la estimación de riesgo de crédito de una cartera de consumo* [Tesis de maestría, Universidad Complutense de Madrid]. Docta Complutense. https://docta.ucm.es/entities/publication/fc512956-4b59-4f40-ab93-18850315b2a3

Ferro Rugeles, R. D., Cossio Escobar, G., & Fernández Moncada, N. (2023). *Pronóstico de ventas a través de una aplicación web aplicando machine learning y variables exógenas* [Trabajo de grado, Escuela Colombiana de Ingeniería Julio Garavito]. Repositorio Escuela Colombiana de Ingeniería Julio Garavito. https://repositorio.escuelaing.edu.co/handle/001/2815

Gantman, E. R. (2020). Apertura comercial, pobreza y desigualdad: un análisis mediante árboles de regresión. *RInERS: Revista de Investigación en Economía y Responsabilidad Social*, *1*(4), 1-14. https://repositoriocyt.unlam.edu.ar/handle/123456789/1298

García García, P. (2024). *Benchmarking de técnicas de explicabilidad para la inteligencia artificial* [Tesis de maestría, Universidad Politécnica de Madrid]. Archivo Digital UPM. https://oa.upm.es/82899/1/TFM_PABLO_GARCIA_GARCIA.pdf

Hermitaño Castro, J. A. (2022). Aplicación de *machine learning* en la gestión de riesgo de crédito financiero: una revisión sistemática. *Interfases*, (15), 160-178. https://doi.org/10.26439/interfases2022.n015.5898

Jaimes, D., González, C. A., Llanos, L., Estrada, O., Barrios, C., Agudelo, D., Navarro Racines, C., Jiménez, D., Gardeazabal, A., & Ramírez Villegas, J. (2023). *Aplicación en la agricultura de técnicas de explicabilidad del aprendizaje de máquinas: aclarando la caja negra*. Alliance Bioversity & CIAT; CGIAR; CIMMYT. https://hdl.handle.net/10568/135694

Joshi, S. (2025). *Model risk management in the era of generative AI: Challenges, opportunities, and future directions* (MPRA Paper 125221). Munich Personal RePEc Archive. https://mpra.ub.uni-muenchen.de/125221/

Marchant Contreras, V. M. (2022). *Un modelo predictivo interpretable para la estimación del ingreso monetario de clientes bancarios basado en XGBoost y SHAP* [Tesis de licenciatura, Universidad de Concepción]. Repositorio UdeC. https://repositorio.udec.cl/server/api/core/bitstreams/fcf68aa4-ddb0-4ef7-9cb4-661b16a7b231/content

Martínez Martínez, G. (2024). *Optimización bayesiana con multifidelidad para la búsqueda de hiperparámetros en algoritmos de aprendizaje por refuerzo* [Tesis de maestría, Universidad Politécnica de Madrida]. Archivo Digital UPM. https://oa.upm.es/82959/

Moneta Pizarro, A. M., Juárez, M. A., Camilo Caro, L. N., & Allub, M. D. H. (2022). Una revisión del supuesto MAR en datos faltantes de la EPH. *Documentos de Trabajo de Investigación de la Facultad de Ciencias Económicas (DTI-FCE)*, *7*, 1-20. https://revistas.unc.edu.ar/index.php/DTI/article/view/37746

Noriega, J. P., Rivera, L. A., & Herrera, J. A. (2023). Machine learning for credit risk prediction: A systematic literature review. *Preprints.org*. https://doi.org/10.20944/preprints202308.0947.v1

Pérez Monsalve, L. O. (2024). *Predicción de la vida útil de las baterías de un Nissan Leaf usando datos reales de conducción mediante el uso de Machine Learning* [Trabajo

de grado, Universidad de Los Andes]. Repositorio Institucional Séneca. https://hdl.handle.net/1992/75130

Polo Romero, V. J. (2024). Modelo predictivo basado en Naive Bayes a través de *machine learning supervised* y la deserción estudiantil, en centros de educación tecnológicos públicos de la región La Libertad. *Sciéndo Ingenium*, 59-71. https://revistas.unitru.edu.pe/index.php/PGM/article/view/6169

Putri, N. H., Fatekurohman, M., & Tirta, I. M. (2021). Credit risk analysis using support vector machines algorithm. *Journal of Physics: Conference Series*, *1863*, 012039. https://doi.org/10.1088/1742-6596/1836/1/012039

Rahmani, R., Parola, M., & Cimino, M. G. C. A. (2024). *A machine learning workflow to address credit default prediction*. arXiv. *https://doi.org/10.48550/arXiv.2403.03785*

Resolución S.B.S. 00053-2023 [Superintendencia de Bancas, Seguros y AFP]. Reglamento de Gestión de Riesgos de Modelo. 6 de enero del 2023. https://intranet2.sbs.gob.pe/dv_int_cn/2240/v1.0/Adjuntos/0053-2023.R.pdf

Reyes Morales, M. A., & Sosa Castro, M. M. (2022). Modelo de puntuación crediticia para tarjeta de crédito en México: una aproximación logística. *Ensayos Revista de Economía*, *41*(1), 17-52. https://doi.org/10.29105/ensayos41.1-2

Romero Carmen, J. A. (2024). *El margen financiero como medida de riesgo de incumplimiento de los préstamos personales del Perú entre 2019 y 2022* [Tesis de maestría, Pontificia Universidad Católica del Perú]. Repositorio Institucional PUCP. https://tesis.pucp.edu.pe/items/f0ab9c35-89a8-4b54-973a-2c1a1a8ce82e

Shcherbinin, A. (2026). Adaptive mechanisms for model retraining in production: Drift triggers, retraining prioritisation, and reduced data preparation time through automated orchestration. *Universal Library of Innovative Research and Studies*, *3*(1), 94-100. https://doi.org/10.70315/uloap.ulirs.2026.0301013

Suhadolnik, N., Ueyama, J., & Da Silva, S. (2023). Machine learning for enhanced credit risk assessment: An empirical approach. *Journal of Risk and Financial Management*, *16*(12), 496. https://doi.org/10.3390/jrfm16120496

Superintendencia de Banca, Seguros y AFP. (2025). *Oficio 02569-2025-SBS: opinión sobre el Proyecto de Ley 8969*. https://www.sbs.gob.pe/Portals/0/jer/opinion_proy_leg/2024/1-OFICIO-02569-2025-SBS-PL8969.pdf

Tocto Reyes, G., Giraldo Coral, X. A., & Hirpo Marchinares, A. E. (2022). Aplicación de *machine learning* para campañas de *marketing* en la banca comercial. *Interfases*, *2*(16), 185-198. https://www.redalyc.org/pdf/7301/730180375008.pdf

True North Partners. (2023). *PRA SS1/23 - Principios de gestión de riesgo de modelo. Nuevos requerimientos regulatorios. https://www.clubgestionriesgos.org/wp-content/uploads/TNP-Riesgo-de-Modelo-PRA-SS1-23_CGR.pdf*

Valenzuela Sánchez, M. S. (2025). *Predicción del riesgo crediticio en clientes del sector retail mediante redes neuronales: caso de estudio de la empresa Credicorp* [Tesis de maestría, Universidad de las Américas]. Repositorio Digital UDLA. https://dspace.udla.edu.ec/handle/33000/18134

Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in Finance—A systematic review of finance, information systems, and computer science literature. *Management Review Quarterly*, *74*, 867-907. https://doi.org/10.1007/s11301-023-00320-0

Wheeler, R., & Carroll, F. (2023). An explainable AI solution: Exploring extended reality as a way to make artificial intelligence more transparent and trustworthy. En C. Onwubiko, P. Rosati, A. Rege, A. Erola, X. Bellekens, H. Hindy & M. G. Jaatun (Eds.), *Proceedings of the International Conference on Cybersecurity, Situational Awareness and Social Media* (pp. 255-276). Springer. https://link.springer.com/chapter/10.1007/978-981-19-6414-5_15

Yang, S., Huang, Z., Xiao, W., & Shen, X. (2024). *Interpretable credit default prediction with ensemble learning and SHAP*. arXiv.

Zegarra Jara, N. (Dir.). (2025). *Microfinanzas: la revista del sistema financiero peruano* (Edición 236). https://statuscomunicaciones.pe/microfinanzas/M236.pdf

Zuleta Fuerte, D. F., Bru Cordero, O. E., & Pastor Sierra, K. S. (2025). Validación cruzada: una herramienta crucial para mejorar la eficiencia de modelos de clasificación con datos biomédicos. *Comunicaciones en Estadística*, *18*(1), 51-85. https://revistas.usantotomas.edu.co/index.php/estadistica/article/view/11214

INTEGRACIÓN DE HERRAMIENTAS DIGITALES PARA EL DISEÑO, SIMULACIÓN E IMPLEMENTACIÓN DE CIRCUITOS RESISTIVOS

Edgar Serrano Pérez
https://orcid.org/0000-0003-4012-9709
eserranop_s@uaemex.mx
Centro Universitario - UAEMEX Valle de Chalco, México

Anabelem Soberanes Martin
https://orcid.org/0000-0002-1101-8279
asoberanesm@uaemex.mx
Centro Universitario - UAEMEX Valle de Chalco, México

Marisol Hernández Hernández
https://orcid.org/0000-0002-9100-1775
mhernandezh@uaemex.mx
Centro Universitario - UAEMEX Valle de Chalco, México

Jose Luis Castillo Mendoza
https://orcid.org/0000-0002-5668-0602
jlcastillom@uaemex.mx
Centro Universitario - UAEMEX Valle de Chalco, México

Recibido: 13 de febrero del 2026 / Aceptado: 23 de junio del 2026
doi: https://doi.org/10.26439/interfases2026.n023.8612

RESUMEN. En este artículo se presenta la implementación de un conjunto de ejercicios prácticos sobre circuitos resistivos en configuración serie, paralela y mixta mediante el uso de una *protoboard*. Este compendio de ejercicios se ha desarrollado como parte introductoria del curso de Circuitos Eléctricos del área de Ingeniería en Computación. Por ello, se analizan en detalle los resultados más relevantes de la experiencia obtenida por los estudiantes. Asimismo, se ofrecen comentarios útiles para quienes se inician en el uso de la *protoboard*, con el propósito de reducir la curva de aprendizaje y evitar los errores más comunes durante su utilización. También se describen las características físicas más relevantes de la *protoboard* y su influencia en la aparición de problemas de conexión entre los dispositivos eléctricos y sus contactos internos. De este modo, este trabajo, que comienza con una descripción teórica y simbólica de los circuitos resistivos, se complementa con la verificación mediante un sistema de simulación virtual y con una guía práctica de conexión

física, apoyada en descripciones gráficas que facilitan la visualización de los aspectos clave durante su implementación.

PALABRAS CLAVE: circuitos / simulación / aprendizaje activo / *protoboard*

INTEGRATION OF DIGITAL TOOLS FOR THE DESIGN, SIMULATION, AND IMPLEMENTATION OF RESISTIVE CIRCUITS

ABSTRACT. It is presented the implementation of a set of practical exercises on resistive circuits in series, parallel, and mixed configurations using a breadboard. The exercises were developed as an introductory component of an Electrical Circuits course within the Computer Engineering program, and the most relevant results from the students' experience are discussed in detail. Useful and important feedback is provided for novice students using breadboards, aiming to reduce the learning curve and avoid common errors. A description of the relevant physical characteristics of the breadboard is provided, as well as their influence on the origin of connection problems between electrical devices and the breadboard's internal contacts. Thus, this work, which begins with a theoretical and symbolic description of resistive circuits, is complemented by verification through a virtual simulation system and provides a practical guide for physical connections through graphical descriptions that facilitate the visualization of key details during implementation.

KEYWORDS: circuits / simulation / learning / protoboard

INTRODUCCIÓN

La síntesis de circuitos eléctricos y electrónicos constituye una parte fundamental de diversas ramas de la ingeniería. Con frecuencia, estos circuitos permiten solucionar distintas problemáticas de la vida cotidiana mediante el uso de dispositivos conectados en configuraciones en serie, paralela o mixta. Por su parte, los circuitos lógicos digitales y los microcontroladores hacen posible el desarrollo de soluciones a medida que, por lo general, se orientan a resolver problemas que requieren una mayor capacidad de cómputo, lo que genera sistemas embebidos (Han, 2024). Tanto en el caso de circuitos simples como en el de circuitos complejos, durante las primeras etapas de diseño, en las que se materializa la integración y conexión de múltiples dispositivos sobre una base de montaje, la tablilla de prototipos, más conocida como *protoboard*, constituye una herramienta fundamental para su desarrollo (Alessandrini, 2023; Sáenz et al., 2024; Xu, 2021). De esta manera, durante el proceso de aprendizaje y entrenamiento de los estudiantes en la síntesis de circuitos, la *protoboard* facilita la implementación de circuitos eléctricos sin necesidad de realizar operaciones de montaje y soldadura sobre una tablilla de circuito impreso, lo que permite reutilizar el material en otras prácticas y aplicaciones (Fox, 2023; Saiz-Vela et al., 2020).

Una de las principales ventajas de la *protoboard* es que permite montar y desmontar los dispositivos de forma rápida y sencilla. Sin embargo, su uso adecuado requiere el desarrollo de múltiples habilidades, tanto técnicas como cognitivas, las cuales demandan dedicación y paciencia. Por ejemplo, técnicamente, es necesario identificar e insertar correctamente los dispositivos que se van a utilizar. Entre las habilidades cognitivas, destaca el razonamiento lógico y secuencial, que permite al estudiante comprender el funcionamiento del circuito y desarrollar una secuencia ordenada de pasos para organizar su implementación. Asimismo, el uso de la *protoboard* en la síntesis de circuitos eléctricos y electrónicos requiere de una estrecha coordinación entre el tacto y la vista. El circuito resultante, con sus dimensiones, distribución y geometrías, es único e irrepetible, ya que el resultado final depende tanto de las características de los materiales y dispositivos utilizados como de las habilidades, destrezas y experiencia del usuario que lo implementa.

REVISIÓN DE LA LITERATURA

Para los estudiantes que recién se inician en el prototipado de circuitos, la síntesis física de circuitos resistivos constituye un gran desafío, debido a la falta de una guía que les permita agilizar el proceso de aprendizaje y a la existencia de múltiples factores y detalles a considerar durante el uso de la *protoboard*. Esta situación acentúa la brecha existente entre la teoría y la práctica. En la literatura, los estudiantes pueden consultar las ecuaciones que permiten obtener las resistencias equivalentes de los circuitos resistivos; sin embargo, generalmente se abordan casos de estudio desde un punto de vista teórico y con un enfoque general (Christensen, 2022; Urone & Hinrichs, 2016). Cabe destacar que

en los casos de estudio que se presentan, se utilizan valores ideales de resistores que no pueden adquirirse comercialmente en las tiendas de electrónica, con el propósito de analizarlos de acuerdo con los modelos teóricos y contrastarlos posteriormente con los resultados obtenidos mediante su implementación física.

Además, aunque en la literatura pueden encontrarse las ecuaciones y los métodos de análisis de circuitos, los estudiantes suelen presentar dificultades para sintetizarlos físicamente. Esto se debe a que la implementación de un circuito resistivo a partir de un diagrama o esquema, se requiere el desarrollo de habilidades espaciales y de interpretación de símbolos y representaciones esquemáticas, aspectos que, por lo general, no son descritos en los trabajos publicados. Con el propósito de reducir la brecha entre los modelos teóricos y la implementación práctica, se han reportado trabajos relacionados con el desarrollo de un kit experimental (Schiavon et al., 2022) y de un laboratorio virtual de aprendizaje (Da Silva et al., 2021), ambos enfocados en la temática de circuitos resistivos.

Con los avances tecnológicos actuales, se identifica la necesidad de incorporar herramientas modernas de diseño y simulación, como EasyEda, Fritzing y TinkerCad, con el fin de abordar la temática de circuitos resistivos desde una perspectiva computacional, lo que favorece tanto la visualización de los diagramas eléctricos como sus múltiples representaciones y conexiones. De esta manera, se busca establecer un puente entre la teoría y la implementación práctica mediante el uso de tecnologías computacionales para la visualización de circuitos resistivos.

Asimismo, se considera importante brindar una guía con detalles y consideraciones relevantes para los estudiantes que se inician en el prototipado de circuitos resistivos utilizando una *protoboard*. En este contexto, el presente trabajo se enfoca en el análisis, la simulación y la implementación de circuitos resistivos a través de herramientas digitales, además de ofrecer una guía práctica de implementación con *protoboard* que describe diversos aspectos de conexión que deben considerarse al iniciar el prototipado de estos circuitos. También se presentan los errores más comunes que suelen cometerse y se proponen recomendaciones para lograr conexiones satisfactorias.

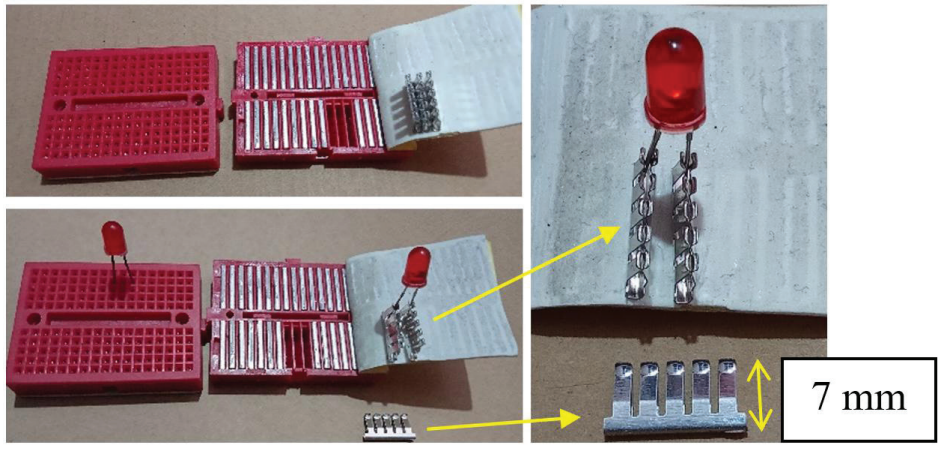
MATERIALES Y MÉTODOS

La parte externa de la *protoboard* está constituida por una estructura aislante de material polimérico que proporciona aislamiento, protección y soporte a los elementos conductores ubicados en su interior. Existen múltiples tamaños y configuraciones (Harrold, 2020), por lo que es posible encontrar modelos con rieles de alimentación o bornes de conexión externa, característica común en las versiones de mayor tamaño. La mayoría de las *protoboards* cuenta en sus extremos con pestañas que facilitan la unión de dos o más unidades en forma paralela, lo que incrementa el área disponible para la conexión de múltiples dispositivos.

Independientemente de su tamaño, la gran mayoría dispone de un canal central que la divide en dos secciones. Esta característica es importante, ya que facilita el montaje de circuitos integrados con encapsulado de tipo dual (*dual in line package*, DIP). Asimismo, en cada sección, se distribuyen filas con cinco perforaciones en cada una. Cada perforación de la superficie superior coincide con la ubicación de un contacto metálico perteneciente a un riel conductor alojado en el interior de la estructura polimérica. De esta manera, es posible insertar la terminal de un dispositivo electrónico en cualquiera de estas perforaciones para establecer la conexión eléctrica, dado que cada fila está integrada internamente por un riel conductor que interconecta sus cinco contactos, como se observa en la Figura 1.

Figura 1

Vista en detalle de la estructura interna y externa de una protoboard



Los rieles conductores, conformados por cinco contactos eléctricos interconectados, presentan diversas particularidades que deben considerarse al utilizar la *protoboard* en la síntesis de circuitos eléctricos. Por ejemplo, la altura de cada contacto es de alrededor de 7 mm, mientras que el espesor de la estructura exterior de plástico es de aproximadamente 8 mm. Asimismo, la separación entre rieles conductores adyacentes es de 2,54 mm. Es importante tomar en cuenta que las reducidas dimensiones de las *protoboards* comerciales dificultan la visualización de las conexiones (Scott, 2004), por lo que se debe prestar particular atención y cuidado durante la síntesis de los circuitos.

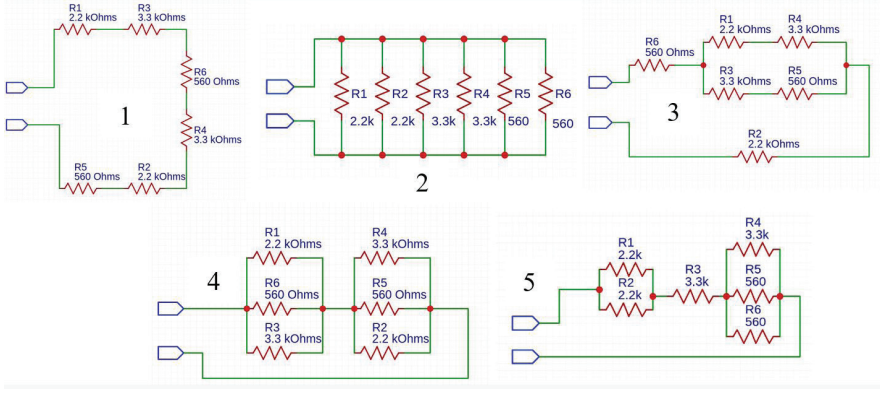
Por otra parte, los contactos están formados por dos láminas metálicas delgadas ubicadas de manera paralela, que adoptan la forma de una pinza y se abren al introducir los terminales o los cables conductores entre sus superficies internas. Dependiendo de los materiales y de la geometría empleada en su fabricación, la fuerza de sujeción que ejercen

sobre las terminales de los dispositivos puede variar de un fabricante a otro, de modo que algunas *protoboards* ofrecen una mayor firmeza, mientras que otras presentan una mayor holgura. Por supuesto, este comportamiento también se encuentra influenciado por el espesor de las terminales de los componentes utilizados. Este aspecto es de suma importancia, ya que algunos componentes electrónicos poseen terminales de espesor delgado, como ciertas resistencias o cables de reducida sección transversal, que al insertarse en contactos con holgura pueden deslizarse y desprenden con facilidad, lo que da lugar a falsos contactos y a la pérdida de continuidad en las conexiones. Por otro lado, cuando los contactos son demasiado firmes y los terminales son de pequeña sección transversal, es común que estas se doblen o deformen sin lograr introducirse entre las láminas que conforman el contacto interno de la *protoboard*. Así, es de suma importancia evitar insertar terminales de cables o dispositivos con una longitud superior a 7 mm, ya que, al exceder esta longitud, el cable puede sobrepasar el riel conductor y entrar en contacto con alguno de los rieles adyacentes, lo que da lugar a conexiones no deseadas.

En el curso de Circuitos Eléctricos y Electrónicos, impartido en el quinto semestre de la carrera de Ingeniería en Computación, se abordó el tema de las conexiones en serie, en paralelo y mixtas de circuitos resistivos utilizando la *protoboard*. El objetivo fue que los estudiantes iniciaran el análisis de circuitos resistivos, inicialmente de forma teórica y matemática; luego, verificar dicho análisis mediante simulación virtual; y, posteriormente, validarlo físicamente a través de la implementación de los circuitos en una *protoboard*. De esta manera, para la asignatura, se diseñaron cinco circuitos resistivos utilizando seis resistencias de valores comerciales. El primer circuito representa una conexión en serie, el segundo una conexión en paralelo, y los circuitos 3, 4 y 5 corresponden a configuraciones mixtas, con el fin de que los estudiantes puedan observar las distintas resistencias equivalentes que se obtienen al realizar diferentes combinaciones entre ellas. El diseño de los diagramas de conexión se realizó con el editor en línea EasyEDA, tal como se observa en la Figura 2.

Figura 2

Diagramas de conexión de los circuitos resistivos

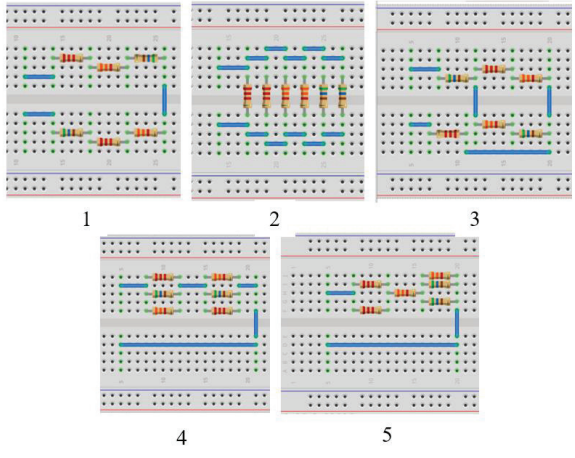


Cabe señalar que existen múltiples formas de obtener la conexión de los circuitos resistivos presentados en la *protoboard*, dado que las resistencias eléctricas pueden colocarse y distribuirse en distintas filas o regiones, con lo que se obtiene la misma resistencia equivalente. Un resultado satisfactorio se obtiene siempre que se ejecuten y mantengan correctamente las conexiones eléctricas del circuito, las cuales incluyen, por un lado, las conexiones internas con los contactos de la *protoboard* y, por otro lado, la conexión con los cables o dispositivos externos. Sin embargo, con el fin de favorecer la visualización física de las conexiones en la *protoboard* para los estudiantes, se diseñaron los esquemas de conexión en dos dimensiones mediante el *software* Fritzing (versión 0.9.3), para que sean utilizados como guía de conexión.

Esta herramienta resulta particularmente útil para los estudiantes que no han tenido experiencia previa en la conexión de circuitos eléctricos o en la implementación física con *protoboard*. Además, es importante resaltar que en el sistema de Fritzing se puede visualizar de forma gráfica una línea punteada de color verde en cada fila de perforaciones que cuenta con una conexión activa. De esta manera, una vez que se conecta algún dispositivo o cable, dicha fila se ilumina en color verde, como indicativo de la conexión interna entre todos los puntos de contacto de esa fila (cinco), lo que facilita la visualización de las conexiones eléctricas entre los distintos componentes que se utilizan (Fox, 2023), tal como se observa en los diagramas de la Figura 3.

Figura 3

Diagramas de conexión en la protoboard *utilizando la aplicación Fritzing*



De esta forma, es posible utilizar el *software* para enfatizar aspectos relevantes en las conexiones realizadas en la *protoboard*. Por ejemplo, permite ejemplificar de manera gráfica y visual las situaciones en las que no existe conexión entre las terminales de los dispositivos —es decir, cuando se encuentran sin continuidad—, así como aquellas en

las que sí existe la conexión o continuidad entre las terminales. Otros aspectos importantes que amerita resaltar desde un enfoque didáctico son la posibilidad de visualizar la orientación, ubicación y distribución de los componentes, lo que facilita la planificación de la conexión del circuito. Además, se favorece la visualización de las características intrínsecas de los componentes (forma, color y tamaño), por lo que se han preparado diagramas de conexión que contribuyen a la comprensión de los escenarios en los que existe o no conexión entre los dispositivos, tal como se observa en la Figura 4.

Figura 4

Dispositivo presentado a los estudiantes en el que se observa desconexión (rojo) y conexión (verde) entre las terminales de las resistencias

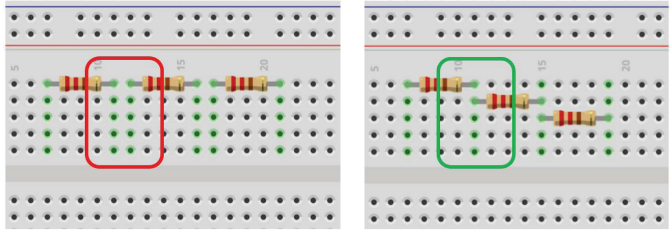
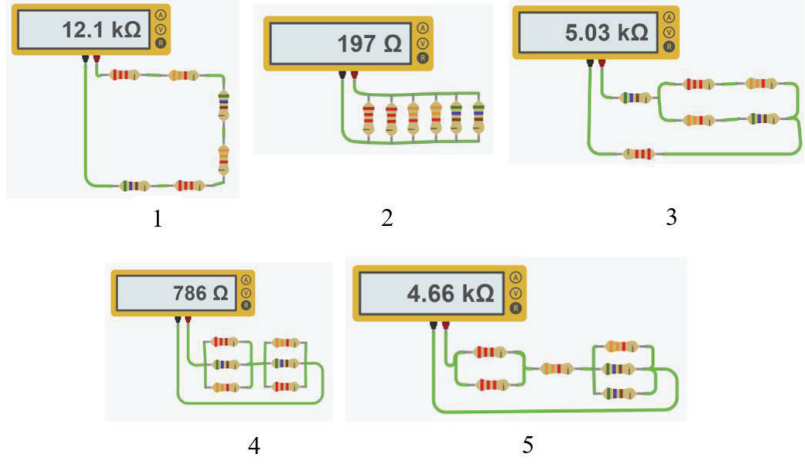


Figura 5

Obtención de la resistencia equivalente mediante un multímetro para cada uno de los circuitos resistivos utilizados



Dado que la temática de los circuitos resistivos incluye la obtención de la resistencia equivalente, se desarrollaron modelos de simulación utilizando el *software* Tinkercad. Los resultados obtenidos con el simulador permiten contrastarlos y relacionarlos con

aquellos obtenidos a partir de los cálculos matemáticos desarrollados en los modelos teóricos. Se siguió una estructura de conexión similar a la observada en el diagrama eléctrico con representación simbólica, lo que permite al estudiante correlacionar el símbolo eléctrico de la resistencia con el elemento resistivo representado de forma virtual en el simulador. De esta manera, la simulación virtual es una herramienta que facilita la comprensión de la relación entre la representación abstracta y simbólica del resistor y el dispositivo físico con sus características intrínsecas, como sus dimensiones y su código de bandas de colores (véase la Figura 5).

Como se observa en la Figura 5, es posible agregar un multímetro en la modalidad de medición de resistencia, con el fin de verificar la resistencia equivalente del circuito. En la actualidad, el *software* Tinkercad puede utilizarse a través de un navegador con conexión a internet, por lo que es accesible para los estudiantes dentro y fuera del aula de clases, lo que facilita la verificación de los resultados teóricos con aquellos obtenidos en el modelo de simulación, prácticamente desde cualquier lugar con acceso a internet.

RESULTADOS Y DISCUSIÓN

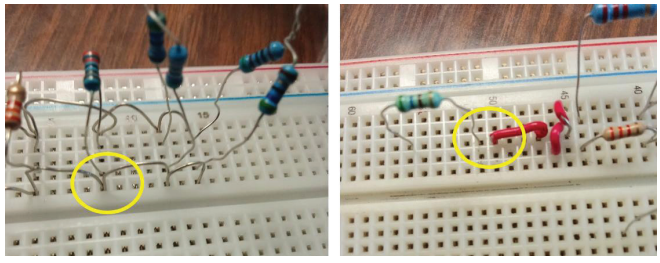
Una vez que los estudiantes contaron con los diagramas de conexión, se procedió a la implementación utilizando la *protoboard*. Al finalizar las conexiones necesarias, se utilizó el multímetro en su modalidad de óhmetro, con la finalidad de verificar los resultados teóricos, de simulación y experimentales. Durante la etapa de revisión, se observó que, en algunos casos, los resultados obtenidos no fueron satisfactorios, pues no se encontró correspondencia entre los valores numéricos teóricos y los indicados en la pantalla del multímetro. Por ello, se llevó a cabo una etapa de detección de fallas en el armado de los circuitos. Asimismo, es importante resaltar que, al implementar un circuito y verificar su funcionamiento, es común que no se obtengan resultados satisfactorios en el primer intento, lo que puede generar molestia y frustración entre los estudiantes. Esta dificultad para encontrar una solución rápida es una situación habitual en la resolución de problemas en el ámbito de la ingeniería, por lo que debe prevalecer la perseverancia y paciencia en la búsqueda de la solución.

En este sentido, se indicó a los estudiantes que deben evitar colocar más de un cable o terminal en cada una de las perforaciones individuales de la *protoboard*, ya que este procedimiento puede provocar deformaciones en los contactos internos, lo que reduce la fuerza de sujeción y ocasiona posibles desconexiones debido a fuerzas o movimientos externos. Cabe señalar que esta situación es una práctica común durante las primeras sesiones de uso de la *protoboard*. Además, se ha observado que algunos estudiantes colocan las terminales de los dispositivos en filas paralelas entre sí, lo cual no genera una conexión eléctrica entre ellos; por ello, es importante representar de forma clara la manera en que dos elementos deben conectarse, asegurando que compartan la misma fila de conexión.

Otra situación relevante observada es que los dispositivos adquiridos por los estudiantes en tiendas de electrónica no siempre son verificados previamente. Por ejemplo, se identificó el uso de resistencias con valores diferentes a los solicitados, por lo que los estudiantes intentan ajustar el valor requerido mediante la conexión de resistencias en serie. Este tipo de comportamiento ya ha sido reportado en la literatura y se basa en la adición de componentes que pueden alterar la dinámica del sistema (Booth et al., 2016), añadiendo fuentes de error y ruido innecesarios al circuito, tal como se observa en la Figura 6.

Figura 6

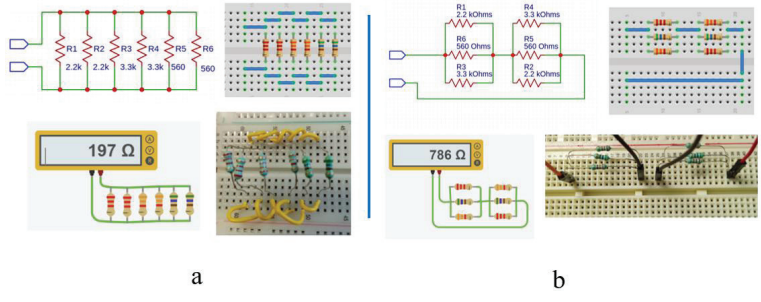
Procedimientos más comunes que realizan los estudiantes al iniciar el uso de la protoboard



Durante la síntesis de los circuitos eléctricos, es importante resaltar una de las ventajas de la *protoboard*, la cual permite corregir errores de forma rápida una vez que han sido detectados, ya sea verificando la correcta conexión en un punto o reemplazando un cable o dispositivo defectuoso. De este modo, al realizar los ajustes necesarios y utilizar los diagramas de conexión en Fritzing, los estudiantes comenzaron a realizar la síntesis de los circuitos de manera más fluida, lo que les permitió contar con una referencia de conexión para los dispositivos en una representación cercana a los dispositivos físicos, como se observa en los circuitos eléctricos resultantes en la Figura 7.

Figura 7

Resultados de conexión obtenidos por algunos equipos de la asignatura



Nota. Se observan las conexiones físicas correspondientes a los circuitos 2 (a) y 4 (b).

En concordancia con la literatura, las principales fallas identificadas se atribuyen, por una parte, a un deficiente montaje de los dispositivos y a su conexión con los contactos internos de la *protoboard* y, por otra, a la ausencia o a la mala conexión de los cables auxiliares (Booth et al., 2016; Lee et al, 2021). El origen de los problemas en un circuito eléctrico es amplio, y cabe resaltar que la mayoría de ellos son difíciles de identificar o detectar, ya que, por lo general, la causa puede encontrarse fuera del alcance de la vista humana o derivarse de una conexión incorrecta como resultado de algún descuido. Por ello, es importante señalar, como guía para los estudiantes, las siguientes consideraciones:

- Evitar el uso de protoboards cuyos materiales de fabricación sean de baja calidad, ya que los contactos internos tienden a doblarse, debilitarse y desgastarse con mayor rapidez.
- Evitar usar dispositivos o cables con una excesiva longitud de inserción en los contactos internos, pues puede provocar falsos contactos o cortocircuitos con otros elementos en el interior de la tablilla y, al encontrarse fuera del campo visual, estos son difíciles de identificar o detectar.
- Es posible que los cables o jumpers utilizados en prácticas pasadas no tengan continuidad; es decir, que se encuentren cortados o dañados en el interior del recubrimiento plástico, lo cual tampoco es visible desde el exterior de la protoboard.
- Es posible que los dispositivos utilizados presenten defectos de fabricación o hayan sido dañados durante manipulaciones previas, lo cual también es difícil de determinar únicamente mediante inspección visual.
- Es importante evitar doblar y deformar las terminales de los dispositivos, ya que esto impide obtener una conexión estable con los contactos internos de la tablilla.
- Utilizar los componentes con las características requeridas y especificaciones requeridas para el circuito.
- Planificar la distribución, orientación y ubicación de los componentes en la superficie de la protoboard antes de iniciar el montaje de los dispositivos.
- Realizar la conexión de los dispositivos por etapas, verificando el correcto funcionamiento de cada nuevo elemento a medida que se incorpora, así como su adecuada interacción con los componentes previamente conectados.

Adicionalmente, al realizar una consulta en el *chatbot* ChatGPT, basado en GPT-5.2, mediante el *prompt* "los principales fallos de estudiantes principiantes en nivel universitario al comenzar a utilizar la *protoboard*", se obtuvo la información que se detalla a continuación.

Principales fallos al usar la protoboard

Estudiantes principiantes - Nivel universitario

1. No comprender la estructura interna de la protoboard
   Desconocer cómo se conectan internamente las filas centrales y las líneas de alimentación laterales.
2. Interpretar mal el esquema eléctrico
   No traducir correctamente el diagrama al montaje físico en la protoboard.
3. Colocar incorrectamente los circuitos integrados
   Ignorar la ranura central o la orientación del pin 1 del IC.
4. Errores en la alimentación del circuito
   - Invertir Vcc y GND
   - No alimentar ambos rieles laterales
   - Aplicar un voltaje incorrecto
5. Omitir componentes esenciales
   Como resistencias limitadoras, pull-up/pull-down o capacitores de desacoplo.
6. No respetar la polaridad de los componentes
   Especialmente LEDs, diodos y capacitores electrolíticos.
7. Cableado largo, desordenado o sin criterio
   Uso excesivo de jumpers que introduce ruido y dificulta la depuración.
8. No verificar conexiones antes de energizar
   No comprobar continuidad ni posibles cortocircuitos con el multímetro.
9. Exceder las limitaciones de la protoboard
   Intentar trabajar con corrientes altas o señales de alta frecuencia.
10. Depender del ensayo-error sin análisis previo
    No contrastar el montaje con el esquema ni con valores teóricos esperados.

Se observa que existen puntos de concordancia entre la interacción con el *chatbot* y los problemas identificados en este proyecto. En relación con el primer punto, se coincide en que es fundamental que los estudiantes conozcan la estructura interna de la *protoboard*, la cual en este trabajo ha sido descrita de forma gráfica en la Figura 1. El segundo punto, que implica una posible malinterpretación del esquema eléctrico, puede abordarse mediante los diagramas de conexión física en la *protoboard*, como los mostrados en la

Figura 3, los cuales favorecen la relación entre el esquema simbólico y el montaje físico. Los puntos 4, 5 y 6, aunque son importantes, por el momento quedan fuera del análisis, dado que, en esta propuesta, no se emplean circuitos integrados, tampoco se energizan los circuitos con fuentes de alimentación ni se utilizan arreglos auxiliares; además, las resistencias no presentan polaridad. Sin embargo, sí es importante considerar la mención del canal divisorio de la *protoboard*.

El punto 7, que implica el cableado largo y desordenado, también ha sido abordado en este proyecto y se vincula con las características dimensionales de los contactos internos de la *protoboard*, los cuales miden aproximadamente 7 mm, por lo que el cable que se inserta en las perforaciones no debería exceder dicha dimensión. El punto 8 es particularmente uno de los más útiles e importantes, ya que implica la verificación de continuidad mediante el uso del multímetro, con el fin de comprobar el buen estado de los cables auxiliares o de las conexiones internas de la *protoboard*. Con respecto al punto 9, es importante señalar que no deben excederse corrientes superiores a 1 amperio, con el fin de evitar el sobrecalentamiento de los contactos internos. Por su parte, el décimo punto menciona la importancia de contar con diagramas y esquemas de conexión que faciliten el montaje, así como una referencia de los valores esperados, a los que se puede acceder mediante simuladores virtuales, como los diagramas mostrados en la Figura 5.

Dado que gran parte de las fallas son difíciles de detectar mediante una inspección visual, el multímetro constituye una herramienta fundamental para detectar y localizar posibles fallas en los circuitos resistivos. Por ejemplo, permite determinar directamente el valor nominal de una resistencia. Además, es posible medir el valor de la resistencia equivalente en un circuito en serie, paralelo o mixto, y la función de continuidad permite verificar si los cables o los contactos se encuentran en buen estado. Esto facilita el descarte sistemático de posibles fallas durante la búsqueda de la causa que origina el funcionamiento inadecuado del circuito.

Una vez que los circuitos resistivos se encuentran correctamente conectados, en las siguientes temáticas se analizan los circuitos resistivos conectados a una fuente de alimentación, iniciando con fuentes de corriente continua. En este sentido, para abordar las siguientes prácticas, se sugiere la adquisición de una tarjeta tipo Arduino Uno, la cual ofrece la ventaja de integrar en la misma placa una fuente de alimentación de 5 voltios. Adicionalmente, también proporciona una salida de 3,3 voltios para alimentar otros dispositivos. Es importante señalar que se trata de una fuente de alimentación regulada, característica fundamental para energizar distintos dispositivos eléctricos y electrónicos, así como circuitos integrados y sensores que requieren distintos voltajes de alimentación.

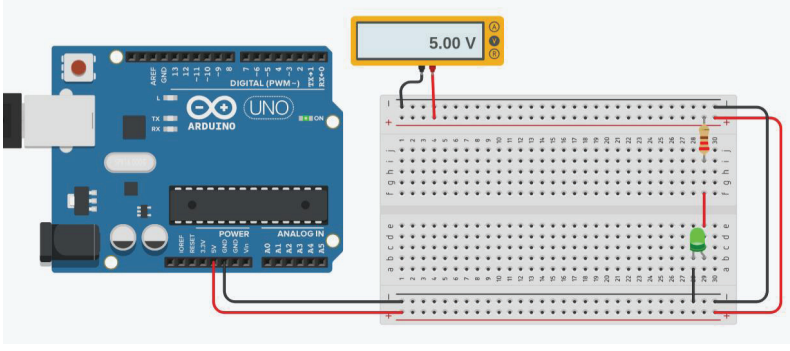
El suministro de energía a la tarjeta tipo Arduino Uno puede realizarse mediante la conexión a un puerto USB de una computadora personal o portátil. Además, es posible

alimentarla con una batería de 9 voltios, por lo que constituye una fuente de alimentación compacta y portátil que puede utilizarse prácticamente en cualquier lugar, tanto dentro como fuera del laboratorio. En este sentido, su uso resulta especialmente útil en situaciones que requieren sesiones a distancia o en línea, en los que el acceso al equipo de laboratorio de las instituciones es limitado. En estos casos, la fuente de alimentación integrada en la tarjeta con microcontrolador representa una alternativa práctica y conveniente. Asimismo, debe considerarse que la adquisición de una tarjeta con microcontrolador constituye una inversión a corto y mediano plazo, ya que puede utilizarse en diversas asignaturas posteriores de la carrera de Ingeniería en Computación, como Sistemas Digitales, Robótica y Sistemas Embebidos.

En relación con la conexión de la tarjeta tipo Arduino Uno con la *protoboard*, se sugiere utilizar cables de cobre o *jumpers*. Así, como buena práctica para identificar las conexiones, se recomienda emplear un cable de color rojo para la conexión al positivo de la fuente —es decir, a la terminal de +5 voltios—. En contraste, para la conexión al pin GND de la placa, se utilizar un cable de color negro. En la tarjeta con microcontrolador, el pin GND constituye la referencia de 0 voltios del circuito. Así, se establece una convención de colores en los conductores, en la que el rojo y el negro identifican las líneas de alimentación eléctrica de los circuitos, tal como se observa en la Figura 8.

Figura 8

Conexión de la fuente de alimentación de una tarjeta tipo Arduino Uno con una protoboard



Es importante señalar que, por lo general, la *protoboard* cuenta con dos rieles de alimentación ubicados en los extremos laterales de la tablilla, los cuales, en algunos casos, son desmontables, a fin de permitir la unión de dos o más *protoboards*. Para facilitar su identificación visual, algunos fabricantes incorporan un marcado mediante líneas continuas de color rojo y negro, sugiriendo al usuario que las utilice como líneas de alimentación de los circuitos, lo que facilita su posterior identificación y uso para la distribución de la energía eléctrica. Asimismo, algunas cuentan con un indicativo del signo

positivo (+) y otro del signo negativo (-) en sus esquinas, lo que facilita la identificación de las líneas de conexión destinadas a la fuente de alimentación. Por lo general, los rieles de alimentación ubicados en los laterales de la *protoboard* se encuentran eléctricamente aislados entre sí. Por ello, para obtener una continuidad entre ambos, es necesario realizar conexiones adicionales, como las que se observan en la parte derecha de la Figura 9.

En algunos otros modelos, los rieles laterales de la *protoboard* se encuentran divididos en secciones, por lo que también es necesario realizar conexiones adicionales si se requiere garantizar la continuidad de los rieles de alimentación y energizar otras regiones de la *protoboard*. Asimismo, existen modelos pequeños y compactos que incluso no cuentan con rieles destinados a líneas de alimentación, por lo que se utilizan para arreglos más particulares y específicos. En el contexto de prácticas de laboratorio educativas, se recomienda adquirir una *protoboard* de mayor tamaño, que permita desarrollar los circuitos necesarios de acuerdo con su dimensión proyectada y el número de componentes a utilizar, los cuales en algunos casos pueden llegar a ser bastante extensos, especialmente en asignaturas como Circuitos Digitales. De esta manera, cuando se requiere la alimentación de circuitos integrados, es necesario utilizar una *protoboard* con rieles destinados a energizar las distintas regiones de la placa, así como los múltiples dispositivos que pueden utilizarse como entradas o salidas del sistema.

El uso de los colores rojo y negro para los cables de alimentación es prácticamente una convención generalizada, ya que resulta visualmente intuitiva para los usuarios. De hecho, se utiliza en muchos aparatos y dispositivos eléctricos; por ejemplo, puede observarse en los cables de las baterías de la mayoría de los automóviles, con el fin de evitar errores de conexión que podrían derivar en la quema de fusibles, daños en componentes eléctricos críticos y, en algunos casos graves, en el sobrecalentamiento de los conductores, lo que podría llegar a provocar un incendio, dada la alta capacidad energética de las baterías que los alimentan.

También resulta importante mencionar que, en algunas áreas de la ingeniería, como la neumática, se utiliza el color azul para identificar cables que transportan señales eléctricas, situación que también se observa en diversos sectores industriales en los que se manejan distintas señales eléctricas y se requiere diferenciarlas de los conductores de alimentación. De esta manera, los colores negro y rojo facilitan tanto la identificación de la polaridad correcta en un circuito que se energiza como la verificación de las conexiones en caso de que se presente algún fallo y se requiera realizar un mantenimiento correctivo. Se considera, entonces, que el uso de esta convención de colores permite prevenir errores y agilizar la revisión o el mantenimiento del circuito eléctrico.

Otro aspecto que amerita mencionarse es el calibre de los conductores utilizados en las conexiones. Aunque para el trabajo con *protoboard* se recomienda un calibre 22 AWG, es necesario seleccionar un calibre que brinde un buen equilibro entre firmeza y

facilidad de inserción en las perforaciones de la placa. El uso de un calibre mayor, como 24 AWG, aunque funcional, puede resultar flojo y propenso a desconexiones accidentales durante la manipulación o el traslado del circuito. Por el contrario, un calibre menor, como 20 AWG, aunque ofrece un ajuste más firme, con el tiempo tiende a desgastar y aflojar los contactos internos de la *protoboard*, lo que reduce la capacidad de sujeción del conector y dificulta la estabilidad de la conexión, situación que además es muy difícil de identificar visualmente dada la limitada visibilidad del interior de la mayoría de las *protoboards*.

El calibre de los conductores también determina la capacidad de corriente eléctrica que pueden soportar: a menor calibre numérico, mayor es la corriente que pueden conducir. En este sentido, para fines prácticos educativos en la *protoboard*, los circuitos de baja potencia no deben superar 1 amperio, por lo que el calibre 22 AWG resulta adecuado para esta aplicación. Para verificar que los rieles de alimentación de una *protoboard* se encuentran en funcionamiento y correctamente conectados, se puede utilizar un multímetro para medir el voltaje suministrado desde la tarjeta tipo Arduino Uno. En este sentido, es importante mantener la misma polaridad del instrumento, es decir, conectar el terminal positivo del multímetro con el positivo de la fuente de alimentación de la tarjeta tipo Arduino Uno, a fin de obtener una lectura correcta.

Con frecuencia, se inscribe una gran cantidad de alumnos en la asignatura, lo que dificulta proveer la cantidad suficiente de instrumentos de medición para cada uno de los equipos formados. Por ello, se sugiere colocar un LED con una resistencia en conexión serie, a fin de verificar visualmente que la fuente de alimentación se encuentra en funcionamiento y que provee la energía necesaria para el circuito que se va a energizar. Incluso, puede utilizarse uno de los seis pines analógicos de la tarjeta tipo Arduino Uno (A0-A5) para monitorear el voltaje de la fuente de alimentación, mediante una conversión analógico-digital y su visualización a través del monitor serial del entorno de programación; sin embargo, su descripción queda fuera del alcance del presente trabajo.

Es importante resaltar que los cables de conexión deben ser lo más cortos posible, por lo que resulta necesario proyectar la mínima distancia entre los componentes sobre la superficie de la *protoboard*. Los cables largos tienden a captar mayor cantidad de ruido electromagnético, lo que en algunos casos puede provocar modificaciones en la señal que transportan. Además, dificultan el seguimiento visual de la continuidad del circuito, ya que pueden enredarse y generar confusión con el seguimiento en la identificación de las trayectorias de conexión, lo que dificulta la detección de errores. En este sentido, al utilizar una *protoboard* se busca que las conexiones externas sean lo más cortas y ordenadas posibles, a fin de facilitar el mantenimiento y la identificación de fallos de conexión en el circuito. Dadas las bondades del multímetro como instrumento de medición para la verificación de características eléctricas de los circuitos, en sesiones posteriores se diseñaron recursos educativos destinados a favorecer su comprensión y uso, los cuales serán reportados en una siguiente comunicación.

CONCLUSIONES

Como parte de las actividades iniciales de la asignatura de Circuitos Eléctricos y Electrónicos de la carrera de Ingeniería en Computación, se presentó a los estudiantes una serie de ejercicios de circuitos resistivos en configuración serie, paralela y mixta. El análisis incluyó el cálculo de la resistencia equivalente mediante fórmulas y procedimientos teóricos. Se utilizaron programas de simulación electrónica como Tinkercad y Fritzing, los cuales permiten al estudiante visualizar una forma adecuada de interconexión entre los componentes utilizando una *protoboard*. En concordancia con la literatura, se resalta su papel en la facilitación del reconocimiento e identificación de los componentes electrónicos (Knörig et al., 2009; Pedersen et al., 2021; Watkiss, 2016), y en el caso de este proyecto con resistores, se favorece la visualización de las dimensiones, geometría y las bandas de colores, características fundamentales que determinan los valores nominales de las resistencias. Además, la iluminación en color verde que se activa al colocar las terminales de las resistencias en la *protoboard* permite al estudiante observar la distribución de las conexiones en el riel de contactos, lo que favorece la comprensión de la conexión o el aislamiento de las conexiones.

Es importante destacar el papel de la simulación de circuitos mediante simuladores virtuales en la web, ya que permite cometer errores sin dañar los componentes físicos (Faiña, 2023), lo que facilita que los estudiantes aprendan a partir de dichas fallas y realicen las correcciones necesarias. Este proceso contribuye a una mejor comprensión de la dinámica de los circuitos una vez que han sido correctamente conectados. La implementación física de los circuitos presentados se llevó a cabo utilizando resistencias de valores comerciales, cables de conexión y una *protoboard*, todos ellos de bajo costo y accesibles para la comunidad estudiantil. Durante la etapa de verificación de resultados experimentales utilizando un multímetro en su modalidad de ohmímetro, se identificaron diversos errores de conexión, así como distintas causas que originan falsos contactos, los cuales pueden deberse a la mala calidad de los componentes utilizados o una deficiente interconexión entre los dispositivos y las pistas conductoras de la *protoboard*.

Ante esta situación, se presentó un listado de consideraciones útiles para estudiantes principiantes en la implementación de circuitos físicos, con el fin de favorecer el desarrollo de habilidades técnicas y cognitivas. Dentro de la guía se incluyen una serie de recomendaciones para realizar conexiones físicas de manera satisfactoria, así como la descripción de características físicas de la *protoboard*, a fin de considerar sus detalles dimensionales y de fabricación. Se incluye además una sección destinada al tema de las conexiones con una fuente de alimentación, la cual se sugiere realizar a través de una tarjeta con microcontrolador. Asimismo, se recomienda utilizar la codificación de colores rojo y negro para facilitar su identificación, así como para distinguir las señales eléctricas de los dispositivos que se conectan en un circuito.

La integración del conjunto de herramientas digitales presentadas, de carácter tecnológicamente innovador, tiene un impacto positivo en el estado de ánimo del estudiante, al reducir la frustración y el enfado que algunos experimentan al no obtener resultados satisfactorios inmediatos o al enfrentar dificultades para lograr conexiones adecuadas entre los componentes. De esta manera, la ruta de análisis teórico, verificación virtual y validación experimental, junto con la integración de las herramientas digitales presentadas, permite a los estudiantes consolidar gradualmente la comprensión de la dinámica de los circuitos eléctricos, su interconexión y el desempeño del sistema en el mundo real.

AGRADECIMIENTOS

Se reconoce el apoyo de la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) a través del Sistema Nacional de Investigadoras e Investigadores (SNII).

REFERENCIAS

Alessandrini, A. (2023). A study of students engaged in electronic circuit wiring in an undergraduate course. *Journal of Science Education and Technology*, *32*(1), 78-95. https://doi.org/10.1007/s10956-022-09994-9

Booth, T., Stumpf, S., Bird, J., & Jones, S. (2016). Crossed wires: Investigating the problems of end-user developers in a physical computing task. En *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)* (pp. 3485-3497). Association for Computing Machinery. https://doi.org/10.1145/2858036.2858533

Christensen, D. A. (2022). *Introduction to biomedical engineering: Biomechanics and bioelectricity - Part II*. Springer. https://doi.org/10.1007/978-3-031-01638-7_5

Da Silva, J. B., Machado, L. R., Bilessimo, S. M. S., & Da Silva, I. N. (2021, November). Remote teaching of electrical circuits: Proposal for the use of online laboratories in secondary education. En *2021 World Engineering Education Forum/Global Engineering Deans Council (WEEF/GEDC)* (pp. 594-600). IEEE. https://doi.org/10.1109/WEEF/GEDC53299.2021.9657239

Faiña, A. (2023). Learning hands-on electronics from home: A simulator for Fritzing. En J. M. Cascalho, M. O. Tokhi, M. F. Silva, A. Mendes, K. Goher & M. Funk, M. (Eds.), *Robotics in natural settings (CLAWAR 2022)* (Lecture notes in networks and systems, Vol. 530, pp. 483-492). Springer. https://doi.org/10.1007/978-3-031-15226-9_38

Fox, J. (2023). *Beginning breadboarding: Physical computing and the basic building blocks of computers*. Apress. https://doi.org/10.1007/978-1-4842-9218-1_3

Han, W. P. (2024). *A trilogy for teaching and learning digital electronics and microprocessors* [Ponencia]. 2024 ASEE Annual Conference & Exposition, Portland, Oregon. https://10.18260/1-2--46504

Harrold, C. (2020). *Practical smart device design and construction: Understanding smart technologies and how to build them yourself*. Apress. https://doi.org/10.1007/978-1-4842-5614-5_11

Knörig, A., Wettach, R., & Cohen, J. (2009). Fritzing: A tool for advancing electronic prototyping for designers. En N. Villar, S. Izadi, M. Fraser & S. Benford (Eds.), *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction (TEI '09)* (pp. 351-358). Association for Computing Machinery. https://doi.org/10.1145/1517664.1517735

Lee, W., Prasad, R., Je, S., Kim, Y., Oakley, I., Ashbrook, D., & Bianchi, A. (2021). VirtualWire: Supporting rapid prototyping with instant reconfigurations of wires in breadboarded circuits. En *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21)* (pp. 1-12). Association for Computing Machinery. https://doi.org/10.1145/3430524.3440623

Pedersen, B. K. M. K., Larsen, J. C., & Nielsen, J. (2021). From diagram to breadboard: Limiting the gap and strengthening the understanding. En W. Lepuschitz, M. Merdan, G. Koppensteiner, R. Balogh & D. Obdržálek (Eds.), *Robotics in Education (RiE 2020)* (pp. 339-353). Springer. https://doi.org/10.1007/978-3-030-67411-3_31

Sáenz, J., De la Torre, L., Chaos, D., & Dormido, S. (2024). Hands-on lab practices on control in a distance education university. *IFAC-PapersOnLine*, *58*(26), 37-42. https://doi.org/10.1016/j.ifacol.2024.10.267

Saiz-Vela, A., Fontova, P., Pallejà, T., Tresanchez, M., Garriga, J. A., & Roig, C. (2020). Plataforma de desarrollo de bajo coste para implementar circuitos digitales en FPGAs mediante hardware y software libre. En *Actas del XIV Congreso de Tecnologías Aplicadas a la Enseñanza de la Electrónica (TAEE 2020)* (pp. 419–428). Instituto Superior de Engenharia do Porto. https://doi.org/10.1109/TAEE46915.2020.9163730

Schiavon, G. J., Santos, O. R., Batista, M. C., Braga, W. S., & Bratti, V. M. (2022). Experimental didactic kit for teaching resistors, capacitors and RC timing circuits. *Physics Education*, *57*(5), 055019. https://doi.org/10.1088/1361-6552/ac7cb3

Scott, T. C. (2004, 20-23 de junio). *Versatile, low cost electronics lab protoboard* [Ponencia]. ASEE Annual Conference & Exposition, Salt Lake City, UT, United States. https://doi.org/10.18260/1-2--13627

Urone, P. P., & Hinrichs, R. (2016). 21.1 *College physics*. OpenStax. https://openstax.org/books/college-physics/pages/21-1-resistors-in-series-and-parallel

Watkiss, S. (2016). Let the innovation begin: Designing your own circuits. En S. Watkiss (Ed.), *Learn electronics with Raspberry Pi: Physical computing with circuits, sensors, outputs, and projects* (pp. 247-261). Apress. https://doi.org/10.1007/978-1-4842-1898-3_11

Xu, C. (2021, julio). *Redesigned electrical circuit lab course to face the challenges of remote learning* [Ponencia]. 2021 ASEE Virtual Annual Conference Content Access. https://doi.org/10.18260/1-2--37645

FRAMEWORK CIBERFÍSICO Y ORIENTADO A DATOS PARA LA GESTIÓN DE PROYECTOS DE CONSTRUCCIÓN: DISEÑO Y VALIDACIÓN EN CONTEXTOS REALES

Sergio Zabala-Vargas
https://orcid.org/0000-0001-5803-1123
Sergio.zabala@uniminuto.edu
Corporación Universitaria Minuto de Dios, Colombia

María Jaimes-Quintanilla
https://orcid.org/0009-0000-5735-1379
Maria.jaimes.q@uniminuto.edu
Corporación Universitaria Minuto de Dios, Colombia

Recibido: 30 de abril del 2026 / Aceptado: 16 de junio del 2026
doi: https://doi.org/10.26439/interfases2026.n023.8789

RESUMEN. La industria de la construcción enfrenta desafíos persistentes asociados a la fragmentación de la información, la baja interoperabilidad entre sistemas y la limitada capacidad para gestionar datos en tiempo real, lo que afecta la toma de decisiones en proyectos complejos. En este contexto, el presente estudio tiene como objetivo diseñar y validar un *framework* orientado a integrar tecnologías emergentes y enfoques basados en datos para fortalecer la gestión de proyectos. Esto se realiza a través de un enfoque mixto, en el que se combina una validación cuantitativa mediante el método Delphi con la participación de cinco expertos internacionales y una evaluación cualitativa mediante grupos focales con representantes de treinta y cuatro empresas del sector de la construcción. El *framework* se estructura a partir de una arquitectura multicapa, un flujo de datos continuo y seis dimensiones operativas que articulan la captura, procesamiento, análisis y uso de la información. Los resultados evidencian una alta aceptación del *framework*, con una media global de 4,51 sobre 5,00, un coeficiente de variación promedio de 0,019, un alfa de Cronbach de 0,87 y un coeficiente de Kendall de 0,79, lo que indica un elevado nivel de consenso y confiabilidad. Asimismo, la evaluación cualitativa confirma su aplicabilidad en contextos reales y destaca mejoras en la toma de decisiones, la optimización de recursos y la anticipación de riesgos, aunque se identifican desafíos asociados a la interoperabilidad, la inversión tecnológica y la disponibilidad de talento especializado. Se concluye que el *framework* constituye una propuesta viable para avanzar hacia modelos de gestión más inteligentes, integrados y orientados a datos en el sector construcción.

PALABRAS CLAVES: gestión de proyectos / sector construcción / transformación digital / inteligencia artificial / analítica de datos / *framework*

CYBER-PHYSICAL AND DATA-ORIENTED FRAMEWORK FOR CONSTRUCTION PROJECT MANAGEMENT: DESIGN AND VALIDATION IN REAL-WORLD CONTEXTS

ABSTRACT. The construction industry faces persistent challenges associated with information fragmentation, low interoperability between systems, and limited capacity to manage data in real time, which affects decision-making in complex projects. In this context, this study aims to design and validate a framework focused on integrating emerging technologies and data-driven approaches to strengthen project management. The research adopts a mixed-method approach, combining quantitative validation through the Delphi method with the participation of five international experts and qualitative evaluation through focus groups with representatives from 34 construction sector companies. The framework is structured based on a multi-layer architecture, a continuous data flow, and six operational dimensions that articulate data capture, processing, analysis, and use. The results show high acceptance of the framework, with a global mean of 4.51 out of 5.00, an average coefficient of variation of 0.019, a Cronbach's alpha of 0.87, and a Kendall's coefficient of concordance of 0.79, indicating high levels of consensus and reliability. Additionally, the qualitative evaluation confirms its applicability in real contexts, highlighting improvements in decision-making, resource optimization, and risk anticipation, although challenges related to interoperability, technological investment, and availability of specialized talent are identified. It is concluded that the framework represents a viable proposal to advance toward more intelligent, integrated, and data-driven management models in the construction sector.

KEYWORDS: project management / construction sector / digital transformation / artificial intelligence / data analytics / framework

INTRODUCCIÓN

La industria de la construcción enfrenta desafíos persistentes asociados a la complejidad operativa, la incertidumbre inherente a los proyectos y la fragmentación de la información entre actores, procesos y tecnologías. Estas condiciones se traducen en desviaciones en costos y cronogramas, así como en limitaciones para gestionar riesgos de manera oportuna y basada en evidencia. Aunque metodologías consolidadas de dirección de proyectos y herramientas digitales como el *building information modeling* (BIM) han mejorado la coordinación y la planificación, su adopción aislada no ha sido suficiente para resolver problemas estructurales de interoperabilidad, integración de datos y analítica en tiempo real (Akbari et al., 2018; Jakobson et al., 2025; Kerzner, 2025; Sacks et al., 2020).

Desde la perspectiva tecnológica, la literatura reciente ha documentado el potencial de diversas herramientas digitales para transformar la gestión de proyectos. En este sentido, la convergencia de tecnologías emergentes, tales como la inteligencia artificial (IA), el internet de las cosas (*internet of things*, IoT), el *big data* y los sistemas ciberfísicos (*cyber-physical systems*, CPS), ha abierto nuevas posibilidades para transformar la gestión de proyectos hacia enfoques integrados, dinámicos y orientados a datos. Estas tecnologías permiten capturar información desde el entorno físico, procesarla mediante infraestructuras escalables y analizarla con modelos capaces de identificar patrones, predecir eventos y recomendar acciones. Así, la transición hacia esquemas *data-driven* favorece el paso de prácticas reactivas a modelos predictivos y prescriptivos basados en el análisis continuo de datos (Alotaibi et al., 2024; Alves et al., 2025; Baghalzadeh Shishehgarkhaneh et al., 2022; Cui et al., 2023; Huang et al., 2021; Sivarajah et al., 2024; Zabala-Vargas et al., 2023).

A pesar de estos avances, persisten importantes limitaciones para su implementación integrada en entornos reales, debido a que su incorporación en el sector construcción ha sido fragmentada y desarticulada. En muchos casos, algunas soluciones como BIM, plataformas IoT o herramientas de analítica avanzada se implementan de forma independiente, sin arquitecturas que aseguren interoperabilidad ni flujos de datos que transformen la información en valor estratégico. Esta desconexión limita el aprovechamiento de la digitalización, genera silos de información y dificulta la toma de decisiones informadas (Chathuranga et al., 2023; Wu & AbouRizk, 2023).

Un elemento crítico es la ausencia de *frameworks* técnicos que articulen coherentemente los componentes tecnológicos, metodológicos y operativos necesarios para una gestión inteligente de proyectos. Aunque la literatura reporta avances en áreas específicas, como la aplicación de aprendizaje automático (*machine learning*) para la predicción de riesgos, el uso de drones para monitoreo de obra y la implementación de gemelos digitales, persiste una brecha en la integración de estas capacidades dentro de modelos unificados a lo largo del ciclo de vida del proyecto (Parsamehr et al., 2023; Shokouhi & Bachari, 2025).

En este contexto, los sistemas ciberfísicos adquieren un papel central cuando permiten la integración en tiempo real de sistemas físicos y digitales mediante el uso de sensores, datos y mecanismos de control inteligente. A través de dispositivos conectados y plataformas de comunicación, estos sistemas facilitan la captura continua de datos desde el entorno de obra, los cuales pueden integrarse con estrategias como BIM y gemelos digitales para generar representaciones dinámicas del estado del proyecto. Esta capacidad mejora la visibilidad, el control y la simulación de escenarios, lo que contribuye a la optimización de recursos y a la anticipación de eventos con mayor precisión. Sin embargo, su implementación requiere arquitecturas robustas que soporten la integración, el procesamiento y la explotación de los datos (Qian et al., 2021; Sacks et al., 2020; Sompolgrunk et al., 2024; Wu & AbouRizk, 2023).

La literatura reciente coincide en que la transformación digital del sector no depende únicamente de la adopción tecnológica, sino de su integración dentro de modelos de gestión articulados. La combinación de inteligencia artificial, *big data* y analítica avanzada ha mostrado impactos positivos en la predicción de riesgos, la optimización de cronogramas y el control de costos, especialmente cuando se integra con metodologías como BIM y entornos de datos conectados. No obstante, estos beneficios se ven limitados en ausencia de estructuras integradoras, lo que reduce el aprovechamiento del valor de los datos (Pan et al., 2022; Wu & AbouRizk, 2023).

Asimismo, se evidencian barreras organizacionales que condicionan la adopción tecnológica, como la baja madurez digital, la resistencia al cambio, la falta de talento especializado y los problemas de interoperabilidad. A ello se suman desafíos asociados a la calidad e integración de los datos, que dificultan el desarrollo de modelos analíticos robustos y escalables. En consecuencia, la evolución hacia modelos inteligentes requiere no solo inversión tecnológica, sino de *frameworks* que articulen arquitectura, datos y procesos para soportar decisiones en tiempo real.

Considerando las brechas identificadas en la literatura y en la práctica profesional, surge la necesidad de propuestas integradoras que articulen la tecnología, los datos y la gestión. En este contexto, el presente estudio propone el diseño y la validación de un *framework* ciberfísico y orientado a datos para la gestión de proyectos de construcción, el cual integra una arquitectura multicapa, un flujo continuo de datos y un conjunto de dimensiones operativas alineadas con el ciclo de vida del proyecto. El enfoque busca superar la fragmentación mediante una propuesta estructurada que conecte capacidades tecnológicas con procesos de gestión.

Finalmente, este artículo se organiza de la siguiente forma: en primer lugar, se presenta el marco conceptual; posteriormente, se presenta el diseño del *framework*, incluyendo la arquitectura, el flujo de datos y las dimensiones; a continuación, se presenta la metodología, que integra la validación cuantitativa mediante método Delphi y la evaluación cualitativa mediante grupos focales; seguidamente, se presentan los resultados y su

integración a través de triangulación metodológica; y, por último, se presentan la discusión, las conclusiones, las limitaciones y líneas futuras de investigación.

METODOLOGÍA

En esta sección se describe el enfoque metodológico utilizado para el diseño, la validación y la ruta de integración de los resultados del *framework* propuesto.

Enfoque y diseño de la investigación

La investigación adopta un enfoque mixto con predominio cuantitativo, orientado al diseño y validación de un *framework* ciberfísico y basado en datos para la gestión de proyectos de construcción. El componente cuantitativo se fundamenta en la aplicación del método Delphi para evaluar la pertinencia del modelo, mientras que el componente cualitativo permite analizar su aplicabilidad en contextos reales mediante grupos focales. Este enfoque integrador permite abordar fenómenos complejos en los que convergen dimensiones tecnológicas y organizacionales (Creswell & Clark, 2017). El diseño metodológico se desarrolla en cuatro fases: conceptualización del modelo, estructuración del *framework*, validación cuantitativa mediante Delphi y validación cualitativa mediante grupo focal. Finalmente, se implementó una estrategia de triangulación metodológica para integrar los resultados y fortalecer la validez del estudio.

Diseño del *framework* CPD-FCPM

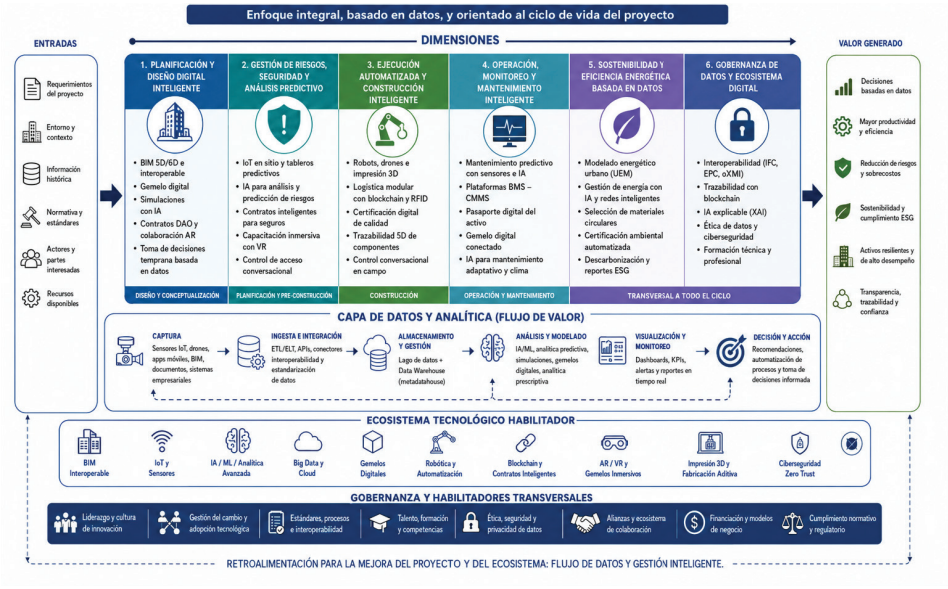
En esta sección se revisa el diseño del *framework* propuesto. Debido a que el objetivo principal de la investigación se asocia al diseño y validación de dicho *framework*, se comprende la importancia de justificar su desarrollo como parte de la metodología de construcción del artefacto investigativo. Como eje central del estudio se desarrolló el *framework cyber-physical and data-driven for construction project management* (CPD-FCPM), concebido como un modelo integrador que articula sistemas ciberfísicos, analítica de datos e inteligencia artificial para la gestión de proyectos de construcción. Su propósito es transformar información del entorno físico y digital en conocimiento útil para la toma de decisiones, mediante una arquitectura estructurada, un flujo de datos continuo y su aplicación en distintas fases del ciclo de vida del proyecto. Este enfoque busca mejorar la eficiencia operativa, optimizar recursos, anticipar riesgos y fortalecer la gestión basada en datos en entornos complejos.

El *framework* se compone de tres elementos principales: una arquitectura multicapa que organiza la captura y procesamiento de datos; un flujo de datos continuo que permite su transformación en conocimiento; y un conjunto de dimensiones operativas que vinculan estas capacidades con el ciclo de vida del proyecto. Adicionalmente, se soporta en un ecosistema de herramientas digitales que facilita su implementación, lo

que garantiza la interoperabilidad y la generación de valor a partir de los datos. En la Figura 1, se presenta un diagrama general que sintetiza el modelo.

Figura 1

Diagrama general del framework CPD-FCPM para la gestión de proyectos



Arquitectura multicapa del framework *CPD-FCPM*

La arquitectura del CPD-FCPM se concibe como una estructura multicapa que organiza la captura, gestión, análisis y explotación de datos en proyectos de construcción. Este enfoque responde a la necesidad de integrar el entorno físico y digital, lo que facilita la interacción entre tecnologías y procesos. La arquitectura se compone de cuatro niveles interrelacionados que soportan una gestión basada en datos a lo largo del ciclo de vida del proyecto (Sacks et al., 2020; Sompolgrunk et al., 2024).

En el primer nivel (N1), la infraestructura ciberfísica constituye la base del modelo y se encarga de la captura de datos desde el entorno operativo. Integra sensores IoT, dispositivos inteligentes, maquinaria conectada y sistemas BIM, lo que permite obtener información en tiempo real y en modo *batch*. Su función es garantizar datos confiables y actualizados para el monitoreo continuo y la detección temprana de eventos (Sompolgrunk et al., 2024). El segundo nivel (N2), correspondiente a la gestión e integración de datos, consolida, depura y almacena la información capturada. Incluye procesos de calidad, transformación y normalización que aseguran coherencia e interoperabilidad, además de mecanismos de gobernanza que regulan el acceso y uso de los datos (Sivarajah et al., 2024).

En el tercer nivel (N3), la capa analítica y la inteligencia artificial transforman los datos en conocimiento útil. Se aplican técnicas de analítica descriptiva, predictiva y prescriptiva, apoyadas en modelos de aprendizaje automático que permiten comprender el comportamiento del proyecto, anticipar riesgos y optimizar decisiones (Greeshma & Edayadiyil, 2022). Finalmente, el cuarto nivel (N4), orientado a la aplicación y la toma de decisiones, facilita la interpretación de la información mediante *dashboards*, indicadores y sistemas de soporte. Esta capa permite generar alertas y recomendaciones, lo que fortalece la capacidad de respuesta del proyecto (Sacks et al., 2020).

En conjunto, esta arquitectura garantiza un flujo coherente de información desde su origen hasta su uso estratégico, lo que promueve una gestión integral basada en datos en entornos complejos.

Flujo de datos del framework *CPD-FCPM*

El flujo de datos del CPD-FCPM constituye el eje dinámico que articula los niveles de la arquitectura, lo que permite transformar datos del entorno físico y digital en conocimiento útil para la toma de decisiones. Este flujo se concibe como un proceso continuo y cíclico, estructurado en cinco etapas interdependientes: captura de datos, procesamiento inicial, análisis de datos, visualización y toma de decisiones. Su diseño garantiza trazabilidad, coherencia y oportunidad en el manejo de la información, en línea con enfoques contemporáneos de analítica de datos y sistemas de soporte a decisiones (Shmueli et al., 2019).

En la fase de captura de datos, se integran múltiples fuentes de información, incluyendo sensores IoT, dispositivos de monitoreo, sistemas BIM, registros operativos y plataformas digitales. Esta diversidad permite obtener datos en tiempo real y en modo *batch*, lo que facilita una visión integral del estado del proyecto. La literatura destaca que la integración de datos provenientes de entornos ciberfísicos es fundamental para mejorar la visibilidad y el control en sistemas complejos (Sompolgrunk et al., 2024).

En la etapa de procesamiento inicial, los datos son sometidos a procesos de depuración, validación, normalización e integración para garantizar su calidad y consistencia. Este proceso es clave para evitar errores en etapas posteriores y asegurar la interoperabilidad entre sistemas, un aspecto crítico en entornos heterogéneos (Sivarajah et al., 2024). En la fase de análisis de datos, se aplican técnicas de analítica descriptiva, predictiva y prescriptiva, apoyadas en modelos de inteligencia artificial y aprendizaje automático que permiten identificar patrones, anticipar riesgos y generar recomendaciones, lo que contribuye a una toma de decisiones más informada (Shmueli et al., 2019). Posteriormente, en la etapa de visualización, los resultados son representados mediante *dashboards* interactivos, indicadores clave de desempeño y herramientas de monitoreo, a fin de facilitar su interpretación por parte de los actores del proyecto. Esta representación visual es fundamental para reducir la complejidad y mejorar la comprensión de los datos.

Finalmente, en la fase de toma de decisiones, el conocimiento generado se traduce en acciones orientadas a la optimización del proyecto, como ajustes en cronogramas, asignación de recursos o mitigación de riesgos. Este proceso incorpora retroalimentación continua, a fin de permitir adaptar las decisiones a la evolución del proyecto y consolidar un enfoque de gestión basado en datos, coherente con la transformación digital en la construcción (Duan et al., 2019; Sivarajah et al., 2024).

En conjunto, este flujo garantiza la transformación coherente de la información desde su origen hasta su explotación, lo que promueve una gestión integral, adaptativa y basada en datos en proyectos de construcción.

Dimensiones operativas del framework *CPD-FCPM*

El *framework* CPD-FCPM se articula mediante seis dimensiones operativas que integran las capacidades tecnológicas del modelo con las fases del ciclo de vida de los proyectos de construcción. Estas dimensiones no operan de forma aislada, sino que se interrelacionan y se soportan en la arquitectura multicapa y el flujo de datos, a fin de facilitar una gestión integral, adaptativa y basada en información (Alotaibi et al., 2024; Duan et al., 2019; Shokouhi & Bachari, 2025). Las dimensiones se sintetizan de la siguiente manera:

- Dimensión de planificación y diseño digital inteligente: integra herramientas como BIM y analítica de datos para la simulación de escenarios, estimación de tiempos y costos, y optimización de decisiones en etapas tempranas. Este enfoque mejora la precisión en la planificación y reduce la incertidumbre mediante BIM 4D/5D/6D, modelado paramétrico y algoritmos de inteligencia artificial para predicción y optimización.
- Dimensión de gestión de riesgos, seguridad y análisis predictivo: incorpora analítica avanzada para identificar y mitigar riesgos. El uso de modelos predictivos y datos en tiempo real permite anticipar eventos críticos y fortalecer la seguridad en obra, apoyado con aprendizaje automático (*machine learning*), análisis de series temporales, sensores IoT y sistemas de alerta temprana.
- Dimensión de ejecución automatizada y construcción inteligente: integra tecnologías como drones, robótica y monitoreo en tiempo real para mejorar la eficiencia operativa y el control del avance. Estas capacidades reducen errores y optimizan recursos mediante visión computacional, sistemas ciberfísicos y plataformas digitales integradas con BIM.
- Dimensión de operación, monitoreo y mantenimiento inteligente: se enfoca en el uso de modelos predictivos y gemelos digitales para el seguimiento del desempeño de los activos. Permite anticipar fallos, optimizar el mantenimiento y prolongar la vida útil de los activos mediante sensores IoT, mantenimiento basado en condición y plataformas de gestión de activos.

- Dimensión de sostenibilidad y eficiencia energética basada en datos: integra herramientas analíticas para optimizar el consumo de recursos y evaluar el impacto ambiental. Esta dimensión impulsa modelos de construcción sostenibles mediante simulación energética, modelado de emisiones y plataformas de análisis de datos.
- Dimensión de gobernanza de datos y ecosistema digital: establece mecanismos para la gestión, la calidad, la seguridad y el uso ético de la información, a fin de garantizar la interoperabilidad y la confiabilidad. Se apoya en arquitecturas de datos, estándares, tecnologías *blockchain* y políticas de ciberseguridad.

Estas dimensiones permiten operacionalizar el *framework* CPD-FCPM, conectando la arquitectura técnica y el flujo de datos con la gestión práctica de los proyectos, y facilitando su implementación en entornos complejos.

Ecosistema tecnológico de soporte al framework *CPD-FCPM*

La implementación del *framework* CPD-FCPM requiere un ecosistema tecnológico que permita operacionalizar sus componentes, el flujo de datos y las dimensiones del modelo en entornos reales de la construcción. Este ecosistema no se concibe como un conjunto aislado de soluciones, sino como una infraestructura integrada que facilita la captura, el procesamiento, el análisis y la explotación de la información a lo largo del ciclo de vida del proyecto.

Este entorno articula distintos tipos de *software* y plataformas, incluyendo entre ellos sistemas BIM, herramientas de captura de datos mediante IoT, infraestructuras de almacenamiento y gestión de datos, soluciones de analítica avanzada e inteligencia artificial, así como plataformas de visualización y soporte a la toma de decisiones. Su selección responde a criterios de interoperabilidad, escalabilidad y alineación con un enfoque *data-driven*, a fin de garantizar la coherencia entre los niveles del modelo.

A continuación, se presenta una síntesis del ecosistema de herramientas tecnológicas:

- Infraestructura ciberfísica y captura de datos: incluye plataformas IoT como Azure IoT Hub, sensores inteligentes, dispositivos *wearables* y maquinaria conectada. Se complementa con herramientas geoespaciales como drones, escáneres LiDAR y *software* de fotogrametría (DJI Terra, Pix4D), que permiten la captura de datos en tiempo real y la generación de representaciones digitales del entorno físico.
- Gestión, integración y almacenamiento de datos: comprende entornos comunes de datos como Autodesk Construction Cloud o Bentley ProjectWise,

junto con plataformas como Data Lake y Data Warehouse (Azure Data Lake, Amazon Redshift, Google BigQuery). Se integran tecnologías como Apache Kafka y *blockchain* para asegurar trazabilidad, calidad e interoperabilidad.

- Analítica avanzada e inteligencia artificial: incluye herramientas como Python, R, Scikit-learn, XGBoost, TensorFlow y PyTorch, así como SPSS y MATLAB. Estas permiten desarrollar modelos predictivos, identificar patrones y optimizar procesos mediante análisis de datos.
- Visualización, simulación y soporte a decisiones: se destacan plataformas como Power BI, Tableau y Qlik Sense para dashboards interactivos, junto con herramientas de simulación como AnyLogic o Simio. También se integran sistemas de soporte a decisiones que generan recomendaciones automatizadas.
- Gestión del ciclo de vida del proyecto: incluye herramientas como Primavera P6, Microsoft Project, Procore, PlanGrid e IBM Maximo, que permiten integrar la información del modelo con la ejecución, operación y mantenimiento del proyecto.

En conjunto, este ecosistema permite conectar el análisis con la acción, lo que facilita la implementación del *framework* en entornos reales y su aplicación en proyectos caracterizados por alta complejidad.

Validación cuantitativa del *framework* CPD-FCPM mediante método Delphi

En esta sección se presenta la estrategia para evaluar la pertinencia, la coherencia y la viabilidad del *framework* CPD-FCPM. Para esto, se implementó un proceso de validación mediante el método Delphi bajo un enfoque cuantitativo, seleccionado por su capacidad para estructurar el juicio experto y generar consenso en contextos complejos. Se conformó un panel de cinco expertos internacionales, los cuales fueron seleccionados por su experiencia superior a ocho años en gestión de proyectos de construcción, transformación digital, inteligencia artificial, BIM y sistemas ciberfísicos, así como por su formación de posgrado y participación en proyectos de alcance internacional.

Vale precisar que el tamaño del panel se definió siguiendo recomendaciones metodológicas para estudios Delphi especializados, en las que se privilegia la experiencia y heterogeneidad de los participantes sobre el número absoluto de expertos. En este sentido, un panel reducido resulta válido cuando el fenómeno analizado es altamente específico y los participantes poseen alta experticia en el campo de estudio, como se sustenta en Belton et al. (2019) y Niederberger y Spranger (2020).

El instrumento consistió en un cuestionario estructurado que evaluó cuatro componentes del modelo (dimensiones, niveles de arquitectura, flujo de datos y ecosistema tecnológico) mediante cinco criterios (relevancia, factibilidad, innovación, impacto y claridad),

utilizando una escala Likert de cinco puntos. El proceso se desarrolló en dos rondas. En la primera, los expertos evaluaron los componentes y se calcularon medidas descriptivas, como la media, la desviación estándar y el coeficiente de variación. En la segunda, se compartieron los resultados para ajustar las valoraciones y favorecer la convergencia.

El análisis incluyó indicadores de consenso y confiabilidad. La media permitió valorar la aceptación, mientras que la desviación estándar y el coeficiente de variación permitieron medir el acuerdo, considerando valores inferiores a 0,20 como indicativos de un alto consenso. Adicionalmente, se calcularon el alfa de Cronbach y el coeficiente de Kendall para evaluar la consistencia interna y la concordancia del panel.

Evaluación cualitativa de la aplicabilidad del *framework* mediante grupos focales

En esta sección se presenta un ejercicio cualitativo basado en la técnica de grupo focal, orientado a evaluar la aplicabilidad del modelo en contextos reales del sector construcción. Este enfoque permite explorar percepciones, experiencias y valoraciones de los actores, lo que facilita la comprensión del *framework* desde una perspectiva operativa (Krueger & Casey, 2014).

Participaron representantes de treinta y cuatro empresas del sector construcción del nororiente colombiano, incluyendo organizaciones de obras civiles, edificaciones, interventoría y consultoría. La selección se realizó mediante muestreo intencional, considerando la experiencia en gestión de proyectos y la disposición hacia la innovación. Los participantes incluyeron a directores de proyectos, ingenieros residentes, coordinadores técnicos y gerentes, lo que permitió una visión integral del modelo.

Para la ejecución de los grupos focales se conformaron cuatro grupos: dos integrados por nueve participantes y dos por ocho participantes, respectivamente. El instrumento consistió en una guía de preguntas semiestructurada y organizada en cinco bloques:

- Pertinencia del modelo:
  - ¿En qué medida el *framework* CPD-FCPM responde a las necesidades actuales de la gestión de proyectos de construcción?
  - ¿Qué problemáticas del sector considera que el modelo aborda de manera adecuada?
- Dimensiones operativas
  - ¿Cómo evalúa la aplicabilidad de las seis dimensiones propuestas en el modelo?
  - ¿Considera que alguna dimensión requiere fortalecimiento o ajuste para su implementación?

- Arquitectura del *framework*
  - ¿Qué tan comprensible resulta la estructura multicapa del modelo?
  - ¿Qué tan viable considera su implementación en su organización?
- Flujo de datos
  - ¿Qué tan clara y útil resulta la secuencia de captura, procesamiento, análisis, visualización y toma de decisiones?
  - ¿Qué dificultades identifica para implementar este flujo en su contexto organizacional?
- Ecosistema tecnológico
  - ¿Qué tan viable considera la adopción de las herramientas tecnológicas propuestas?
  - ¿Cuáles son las principales barreras tecnológicas u organizacionales para su implementación?

Los grupos focales se desarrollaron en sesiones estructuradas que incluyeron la presentación del *framework* y una discusión guiada, y se promovió la participación activa. Las sesiones fueron moderadas por el equipo investigador y registradas mediante notas estructuradas. El análisis se realizó mediante enfoque temático. Se identificaron patrones y se agruparon en categorías asociadas a pertinencia, viabilidad, impacto y limitaciones. Este proceso complementó los resultados del método Delphi, pues aportó evidencia cualitativa sobre la aplicabilidad del modelo en entornos reales. Finalmente, cabe resaltar que la matriz de sistematización cualitativa se encuentra disponible en el siguiente enlace: https://zenodo.org/records/20842885

Estrategia de integración de resultados

La validación del *framework* CPD-FCPM se sustentó en una estrategia de triangulación metodológica orientada a integrar evidencia cuantitativa y cualitativa, con el fin de fortalecer la validez del estudio. Este enfoque permite contrastar y complementar resultados mediante distintos métodos, a fin de reducir sesgos y aportar una comprensión más integral del fenómeno (Creswell & Clark, 2017). La triangulación se estructuró a partir de dos fuentes principales: la validación técnica mediante el método Delphi y la evaluación aplicada mediante grupos focales. El método Delphi permitió analizar la consistencia del modelo, evaluando sus componentes en términos de relevancia, factibilidad, innovación, impacto y claridad, con base en el juicio de expertos internacionales y el uso de indicadores como la media, el coeficiente de variación, el alfa de Cronbach y el coeficiente de Kendall. Por su parte, el grupo focal permitió examinar la aplicabilidad del modelo en contextos reales del sector de la construcción, incorporando la perspectiva de actores empresariales y facilitando la identificación de beneficios, percepciones y barreras para su implementación.

La integración se realizó mediante un proceso de comparación y convergencia de resultados, y se identificaron coincidencias y divergencias entre la valoración experta y la percepción empresarial. Este proceso permitió validar la coherencia conceptual del modelo y ajustar su interpretación en función de su viabilidad práctica. En conjunto, esta estrategia fortalece la robustez del estudio, al validar el *framework* desde una perspectiva técnica y aplicada, lo que lo consolida como una propuesta viable para entornos complejos.

RESULTADOS

Esta sección presenta los principales hallazgos de la investigación, destacando los resultados de la validación realizada con el método Delphi, los elementos obtenidos de los grupos focales y la integración de los resultados.

Resultados de la validación cuantitativa mediante el método Delphi

En cuanto al proceso de validación Delphi, se evidencia una valoración altamente favorable del *framework* CPD-FCPM por parte del panel de expertos. La Tabla 1 presenta la síntesis de resultados de la aplicación de la evaluación con los cinco expertos internacionales.

Tabla 1

Resultados de la validación Delphi del framework *CPD-FCPM*

ID	Componente evaluado	Relevancia	Factibilidad	Innovación	Impacto	Claridad	Media global
D1	Planificación y diseño digital inteligente	4,6	4.4	4.6	4.5	4.5	4.52
D2	Gestión de riesgos, seguridad y análisis predictivo	4,7	4.4	4.5	4.6	4.4	4.52
D3	Ejecución automatizada y construcción inteligente	4,5	4.3	4.6	4.5	4.4	4.46
D4	Operación, monitoreo y mantenimiento inteligente	4,6	4.4	4.5	4.6	4.5	4.52
D5	Sostenibilidad y eficiencia energética basada en datos	4,5	4,3	4,5	4,6	4,4	4,46
D6	Gobernanza de datos y ecosistema digital	4,6	4,4	4,6	4,5	4,4	4,50
A1	Infraestructura ciberfísica	4,6	4,4	4,5	4,6	4,5	4,52
A2	Gestión e integración de datos	4,5	4,4	4,5	4,5	4,5	4,48
A3	Capa analítica e inteligencia artificial	4,7	4,4	4,7	4,6	4,5	4,58
A4	Capa de aplicación y toma de decisiones	4,6	4,5	4,5	4,6	4,6	4,56
F1	Captura de datos	4,5	4,4	4,4	4,5	4,6	4,48
F2	Procesamiento inicial de datos	4,5	4,4	4,4	4,5	4,5	4,46

(*continúa*)

(*continuación*)

ID	Componente evaluado	Relevancia	Factibilidad	Innovación	Impacto	Claridad	Media global
F3	Análisis de datos	4,7	4,5	4,7	4,6	4,5	4,60
F4	Visualización de datos	4,5	4,5	4,4	4,5	4,6	4,50
F5	Toma de decisiones	4,7	4,5	4,6	4,7	4,6	4,62
H	Ecosistema de herramientas *software*	4,5	4,3	4,4	4,5	4,4	4,42

La media global del modelo fue de 4,51 sobre 5,00, lo que indica un alto nivel de aceptación en términos de relevancia, factibilidad, innovación, impacto y claridad.

A nivel de componentes, las dimensiones presentaron valores entre 4,46 y 4,52, entre los que se destacaron las dimensiones relacionadas con la gestión de riesgos, el monitoreo inteligente y la planificación digital. Estos resultados sugieren que el modelo aborda adecuadamente las necesidades clave del sector, particularmente en relación con la anticipación de riesgos y la integración de tecnologías en la toma de decisiones. En cuanto a los niveles de la arquitectura, se observaron valoraciones entre 4,48 y 4,58, siendo la capa analítica y de inteligencia artificial la mejor evaluada. Este resultado es consistente con el enfoque *data-driven* del *framework*, lo que evidencia que los expertos reconocen el valor de la analítica avanzada como elemento central para la gestión inteligente de proyectos.

Respecto al flujo de datos, las etapas presentaron medias entre 4,46 y 4,62, destacándose la fase de toma de decisiones como la de mayor valoración. Esto indica que el modelo logra articular de manera efectiva la transformación de datos en acciones, lo que consolida un enfoque coherente desde la captura hasta la explotación de la información. Por su parte, el ecosistema de herramientas tecnológicas obtuvo una media de 4,42, ligeramente inferior a la del resto de componentes. Este resultado puede interpretarse como una valoración positiva, aunque más cauta, asociada a posibles desafíos en la implementación, tales como los costos, la interoperabilidad y los niveles de madurez digital organizacional.

Desde el punto de vista del consenso, los resultados muestran una desviación estándar promedio de 0,09 y un coeficiente de variación promedio de 0,019, lo que indica una baja dispersión en las respuestas y, por tanto, un alto nivel de acuerdo entre los expertos. Estos valores se encuentran dentro de los rangos considerados como de alto consenso en estudios Delphi, lo que refuerza la robustez de los resultados. En términos de confiabilidad, el instrumento alcanzó un alfa de Cronbach de 0,87, lo que evidencia una buena consistencia interna en la medición de los criterios evaluados. Asimismo, el coeficiente de concordancia de Kendall (W = 0,79) indica un alto nivel de acuerdo global entre los expertos, lo que confirma la estabilidad de las valoraciones.

En conjunto, estos resultados permiten concluir que el *framework* CPD-FCPM es percibido como una propuesta técnicamente sólida, innovadora y con alto potencial de impacto en la gestión de proyectos de construcción. La convergencia de opiniones entre los expertos, junto con los altos niveles de aceptación y consistencia, respaldan la validez del modelo y su aplicabilidad en contextos reales.

Resultados de la evaluación cualitativa mediante grupos focales

Los resultados del grupo focal permitieron analizar la percepción de los actores del sector construcción sobre la aplicabilidad del *framework* en contextos reales. En términos generales, se evidenció una alta valoración de la pertinencia del modelo, destacándose su alineación con los procesos de transformación digital que actualmente atraviesa el sector. A partir del análisis realizado de los aportes de los participantes, se identificaron cuatro categorías recurrentes: pertinencia del *framework*, aplicabilidad operativa, beneficios esperados y barreras para la implementación.

En relación con la pertinencia del *framework*, los participantes coincidieron en que la propuesta responde a varias dificultades habituales del sector construcción. Entre ellas, mencionaron la dispersión de la información, la limitada conexión entre herramientas y la toma de decisiones apoyada en reportes incompletos o poco oportunos. Como lo expresó uno de los participantes: "En los proyectos muchas veces la información está dispersa entre planos, informes, contratistas y plataformas; un modelo que conecte esos datos puede ayudar a tomar decisiones con mayor oportunidad" (P3, comunicación personal, 2025). Esta intervención permite reconocer que el valor del *framework* no está únicamente en integrar tecnologías, sino en ofrecer una ruta para ordenar la información y convertirla en insumos útiles para la gestión del proyecto.

En cuanto a las dimensiones operativas, los participantes señalaron que estas guardan coherencia con el ciclo de vida de los proyectos de construcción. Las dimensiones de planificación digital, gestión de riesgos y monitoreo inteligente fueron percibidas como las de aplicación más inmediata, debido a su relación directa con situaciones frecuentes en obra. En palabras de uno de los representantes empresariales: "Las dimensiones de planificación, riesgos y monitoreo son las que más rápido podrían implementarse, porque son problemas que vivimos todos los días en obra: retrasos, reprocesos y decisiones tardías" (P8, comunicación personal, 2025). No obstante, también se reconoció que la sostenibilidad y la gobernanza de datos requieren mayores capacidades organizacionales. Al respecto, otro participante indicó: "La sostenibilidad y la gobernanza son importantes, pero muchas empresas todavía no tienen procesos claros para medir datos ambientales o controlar la calidad de la información" (P12, comunicación personal, 2025).

Sobre la arquitectura del *framework*, los participantes valoraron positivamente su organización por niveles, pues consideran que facilita la comprensión del papel que

cumple cada tecnología dentro del modelo. Sin embargo, también señalaron que su implementación no debería asumirse como un proceso inmediato ni uniforme para todas las organizaciones. Un participante lo explicó de la siguiente manera: "La arquitectura por niveles ayuda a entender dónde entra cada tecnología; sin embargo, para una empresa pequeña sería necesario implementarla por etapas y con acompañamiento técnico" (P17, comunicación personal, 2025). Esta apreciación muestra que la viabilidad del *framework* depende tanto de su coherencia técnica como de las condiciones reales de adopción en las empresas, especialmente en aquellas con menor disponibilidad de recursos o capacidades digitales.

Respecto al flujo de datos, los participantes destacaron que la secuencia propuesta resulta clara y útil para organizar la gestión de la información. No obstante, advirtieron que el principal reto está en lograr que los datos provenientes de distintas fuentes puedan integrarse de manera efectiva. Uno de ellos afirmó: "El flujo es claro, pero el reto está en que los sistemas se comuniquen; muchas empresas tienen información en Excel, *software* de obra, correos y documentos físicos" (P21, comunicación personal, 2025). Esta observación permite comprender que la interoperabilidad no es un aspecto secundario, sino una condición necesaria para que el enfoque basado en datos pueda operar en contextos reales.

En relación con el ecosistema tecnológico, los participantes reconocieron el potencial de las herramientas propuestas, aunque también señalaron barreras prácticas para su adopción. Los costos de implementación, la disponibilidad de personal capacitado y la integración con los procesos existentes fueron mencionados como aspectos críticos. Uno de los participantes sintetizó esta preocupación al señalar lo siguiente: "La tecnología existe y puede ser muy útil, pero el problema es el costo, la capacitación del personal y la capacidad de integrar esas herramientas con los procesos reales de la empresa" (P26, comunicación personal, 2025). Esta valoración muestra que el ecosistema tecnológico fue percibido como pertinente, pero condicionado por factores organizacionales, económicos y técnicos.

De manera transversal, los participantes resaltaron que el *framework* puede aportar a una mejor toma de decisiones, al aumento de la eficiencia operativa y a la anticipación de riesgos. Uno de los representantes lo expresó así: "Si el sistema permite alertas tempranas y muestra el avance real del proyecto, se podrían reducir decisiones reactivas y actuar antes de que el problema crezca" (P31, comunicación personal, 2025). Al mismo tiempo, se reiteró que su implementación exige inversión, gestión del cambio y fortalecimiento de capacidades digitales. Por ello, más que una adopción inmediata y total, los hallazgos sugieren la conveniencia de una implementación progresiva, ajustada al nivel de madurez tecnológica de cada organización. La matriz completa de sistematización cualitativa se encuentra disponible en el repositorio abierto indicado en la sección metodológica.

En conjunto, los resultados cualitativos muestran que el *framework* es aplicable al sector construcción, siempre que su incorporación se acompañe de condiciones habilitantes. Las voces de los participantes permiten observar una relación clara entre el valor técnico del modelo y las necesidades operativas de las empresas; sin embargo, también dejan ver que su adopción requiere una estrategia gradual, interoperabilidad entre sistemas, inversión tecnológica y fortalecimiento del talento humano.

Finalmente, las citas incorporadas en esta sección fueron anonimizadas mediante códigos alfanuméricos asignados a los participantes, con el fin de proteger su identidad y la de las organizaciones representadas. Estas citas provienen de la matriz de sistematización cualitativa elaborada a partir de las notas estructuradas del grupo focal, en la cual se organizaron las intervenciones según categorías temáticas asociadas a pertinencia, aplicabilidad operativa, beneficios esperados y barreras de implementación.

Integración de resultados y validación del modelo

La integración de los resultados cuantitativos y cualitativos evidencia una alta convergencia entre la valoración experta y la percepción empresarial, lo que fortalece la validez del *framework* CPD-FCPM desde una perspectiva tanto técnica como aplicada. Desde el enfoque cuantitativo, el modelo alcanzó una media global de 4,51 sobre 5, con valores superiores a 4,45 en la mayoría de sus componentes, lo que indica una valoración consistentemente alta en términos de relevancia, factibilidad, innovación, impacto y claridad. Asimismo, el coeficiente de variación promedio de 0,019 refleja un elevado nivel de consenso entre los expertos, mientras que el alfa de Cronbach de 0,87 y el coeficiente de concordancia de Kendall (W = 0,79) evidencian una adecuada consistencia interna y un grado significativo de acuerdo, parámetros que se consideran robustos en estudios Delphi aplicados a contextos tecnológicos complejos (Hsu & Sandford, 2007; Okoli & Pawlowski, 2004).

En cuanto al análisis por componentes, se destaca que la capa analítica e inteligencia artificial, así como la etapa de toma de decisiones dentro del flujo de datos, obtuvieron las valoraciones más altas, con medias cercanas a 4,60. Este resultado refuerza el enfoque *data-driven* del modelo, en el que la generación de conocimiento a partir de datos y su traducción en decisiones operativas constituye el núcleo de valor del *framework*. De igual manera, las dimensiones relacionadas con la gestión de riesgos, la planificación digital y el monitoreo inteligente presentaron altos niveles de aceptación, lo que evidencia su alineación con las necesidades prioritarias del sector construcción.

Desde la perspectiva cualitativa, los resultados del grupo focal con representantes de treinta y cuatro empresas corroboran esta valoración, destacando una percepción positiva frente a la pertinencia y utilidad del modelo. Los participantes reconocieron que el *framework* ofrece una estructura clara para integrar tecnologías emergentes en la gestión de proyectos, particularmente en lo relacionado con la mejora en la toma de

decisiones, la optimización de recursos y la anticipación de riesgos. Esta coincidencia entre la valoración experta y la percepción empresarial refuerza la validez del modelo desde una doble perspectiva: técnica y contextual.

No obstante, la integración de resultados también permitió identificar brechas relevantes. En el análisis cuantitativo, el ecosistema de herramientas tecnológicas obtuvo una media de 4,42, ligeramente inferior al resto de componentes, lo que sugiere una percepción más cauta frente a su implementación. Este hallazgo se ve respaldado por los resultados cualitativos, en los que los participantes identificaron como principales desafíos la inversión requerida, la interoperabilidad entre sistemas y la disponibilidad de talento especializado. Estas limitaciones coinciden con lo reportado en la literatura sobre la adopción de tecnologías digitales en la construcción, donde factores organizacionales y tecnológicos condicionan su implementación efectiva (Omokhua et al., 2025; Shojaei et al., 2023; Zabala-Vargas et al., 2023).

En conjunto, la triangulación metodológica confirma que el *framework* CPD-FCPM no solo presenta una estructura conceptualmente sólida y coherente, sino también una viabilidad práctica significativa, al ser reconocido como una herramienta potencialmente aplicable en entornos reales. Esta convergencia entre la evidencia cuantitativa y cualitativa posiciona el modelo como una propuesta integral que articula capacidades tecnológicas avanzadas con las dinámicas operativas del sector construcción, lo que contribuye a su transformación hacia esquemas de gestión más inteligentes, adaptativos y basados en datos.

DISCUSIÓN

Los resultados evidencian que el *framework* CPD-FCPM responde de manera consistente a las principales limitaciones identificadas en la literatura sobre la gestión de proyectos de construcción, especialmente a la fragmentación de datos, la baja interoperabilidad y la limitada capacidad de análisis en tiempo real. Los altos niveles de aceptación obtenidos con el método Delphi (media global de 4,51) coinciden con estudios que destacan el potencial de los enfoques *data-driven* y la inteligencia artificial para mejorar la toma de decisiones en entornos complejos (Datta et al., 2024; Huang et al., 2021). Asimismo, el énfasis en la capa analítica refuerza lo planteado por Duan et al. (2019), quienes destacan la capacidad de la inteligencia artificial para transformar datos en acciones estratégicas.

En términos de arquitectura, el enfoque multicapa se alinea con las tendencias en la integración de sistemas ciberfísicos, BIM y gemelos digitales (Alves et al., 2025; Sacks et al., 2020; Sompolgrunk et al., 2024). A diferencia de estudios que abordan estas tecnologías de forma aislada, el CPD-FCPM propone una integración estructurada que articula captura, gestión, análisis y explotación de datos, lo que contribuye a cerrar la brecha identificada por Parsamehr et al. (2023) y Shokouhi & Bachari (2025) sobre la

falta de *frameworks* integradores. Respecto al flujo de datos, se destaca la alta valoración de la etapa de toma de decisiones (media de 4,62), lo que evidencia la importancia de estructurar procesos analíticos que transformen datos en conocimiento accionable. Este resultado es coherente con investigaciones sobre *pipelines* de datos y analítica continua (Shmueli et al., 2019; Sivarajah et al., 2024). No obstante, persisten desafíos en la integración de datos procedentes de múltiples fuentes, lo que coincide con lo reportado por Wu & AbouRizk (2023) y Sompolgrunk et al. (2024), quienes señalan la interoperabilidad como una barrera clave.

Las dimensiones operativas del modelo muestran una alta alineación con las necesidades del sector, especialmente en la gestión de riesgos, la planificación digital y el monitoreo inteligente, en concordancia con estudios sobre inteligencia artificial y análisis predictivo (Craveiro & Domingues, 2025; Feng, 2022). Sin embargo, las dimensiones de sostenibilidad y gobernanza de datos presentan un menor grado de madurez, lo que refleja una tendencia reportada en la literatura (Cui et al., 2023; Shokouhi & Bachari, 2025). Un hallazgo relevante es la menor valoración del ecosistema tecnológico (4,42), lo que evidencia desafíos en su implementación asociados a factores organizacionales, económicos y técnicos. Esto coincide con estudios que señalan que la adopción tecnológica depende de la madurez organizacional, la disponibilidad de talento y la integración de sistemas (Shojaei et al., 2023; Shokouhi & Bachari, 2025). En este sentido, el CPD-FCPM no elimina estas barreras, pero las visibiliza y propone una estructura para su abordaje.

La triangulación metodológica permitió validar el modelo tanto en términos de su consistencia técnica como de su aplicabilidad en contextos reales, según lo planteado por Creswell & Clark (2017) sobre la pertinencia de enfoques mixtos. La convergencia entre expertos y actores del sector evidencia un equilibrio entre rigor conceptual y viabilidad práctica. Finalmente, el principal aporte del estudio radica en la formulación de un *framework* integrador que supera enfoques fragmentados y articula arquitectura, datos, analítica y gestión de proyectos. Este modelo facilita la transición hacia una gestión más inteligente y basada en datos, aunque su implementación requiere maduración organizacional, inversión tecnológica y desarrollo de capacidades, lo que abre líneas futuras de investigación en distintos contextos del sector construcción.

CONCLUSIONES

El presente estudio tuvo como propósito diseñar y validar el *framework* CPD-FCPM, orientado a fortalecer la gestión de proyectos de construcción mediante la integración de tecnologías emergentes, analítica de datos y sistemas ciberfísicos. Los resultados evidencian que el modelo constituye una alternativa robusta para abordar las limitaciones estructurales del sector, especialmente en términos de fragmentación de la información, interoperabilidad y análisis en tiempo real.

Desde el punto de vista técnico, el *framework* presenta una alta coherencia estructural, validada mediante el método Delphi con una valoración global de 4,51/5 y adecuados niveles de confiabilidad (α = 0,87; W = 0,79). Estos resultados confirman su pertinencia en términos de relevancia, factibilidad e innovación, destacando la arquitectura multicapa, el enfoque *data-driven* y el flujo de datos como elementos clave para la toma de decisiones. En el ámbito aplicado, la evaluación cualitativa con treinta y cuatro empresas del sector permitió corroborar su viabilidad en contextos reales. Los participantes resaltaron su utilidad para mejorar la toma de decisiones, optimizar recursos y anticipar riesgos, aunque también se identificaron desafíos asociados a la inversión tecnológica, la interoperabilidad y la disponibilidad de talento especializado.

El principal aporte del estudio radica en la formulación de un modelo integrador que articula de manera coherente la captura, la gestión, el análisis y el uso de datos dentro de una arquitectura estructurada. A diferencia de enfoques fragmentados, el *framework* CPD-FCPM vincula capacidades tecnológicas con dimensiones operativas del ciclo de vida del proyecto, lo que facilita una gestión más adaptativa y basada en evidencia. Asimismo, la triangulación metodológica permitió validar el modelo desde una doble perspectiva: técnica y contextual. Esto fortalece su robustez y contribuye a reducir la brecha entre desarrollo conceptual y aplicación práctica. No obstante, el estudio presenta limitaciones relacionadas con el número de expertos, el alcance geográfico de la validación cualitativa y la ausencia de una implementación longitudinal.

Futuras investigaciones deberían enfocarse en la aplicación del *framework* en proyectos reales, el análisis de su impacto en indicadores de desempeño y en su adaptación a distintos niveles de madurez digital. En conclusión, el CPD-FCPM se consolida como una propuesta viable para avanzar hacia modelos de gestión más inteligentes, integrados y orientados a datos en el sector construcción.

REFERENCIAS

Akbari, S., Khanzadi, M., & Gholamian, M. R. (2018). Building a rough sets-based prediction model for classifying large-scale construction projects based on sustainable success index. *Engineering, Construction and Architectural Management*, *25*(4), 534-558. https://doi.org/10.1108/ECAM-05-2016-0110

Alotaibi, B. S., Shema, A. I., Ibrahim, A. U., Abuhussain, M. A., Abdulmalik, H., Dodo, Y. A., & Atakara, C. (2024). Assimilation of 3D printing, artificial intelligence (AI) and internet of things (IoT) for the construction of eco-friendly intelligent homes: An explorative review. *Heliyon*, *10*(17), e36846. https://doi.org/10.1016/j.heliyon.2024.e36846

Alves, J. L., Palha, R. P., & Filho, A. T. de A. (2025). Towards an integrative framework for BIM and artificial intelligence capabilities in smart architecture, engineering,

construction, and operations projects. *Automation in Construction*, *174*, 106168. https://doi.org/10.1016/j.autcon.2025.106168

Baghalzadeh Shishehgarkhaneh, M., Keivani, A., Moehler, R. C., Jelodari, N., & Roshdi Laleh, S. (2022). Internet of things (IoT), building information modeling (BIM), and digital twin (DT) in construction industry: A review, bibliometric, and network analysis. *Buildings*, *12*(10), 1503. https://doi.org/10.3390/buildings12101503

Belton, I., MacDonald, A., Wright, G., & Hamlin, I. (2019). Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process. *Technological Forecasting and Social Change*, *147*, 72-82. https://doi.org/10.1016/j.techfore.2019.07.002

Chathuranga, S., Jayasinghe, S., Antucheviciene, J., Wickramarachchi, R., Udayanga, N., & Weerakkody, W. S. (2023). Practices driving the adoption of agile project management methodologies in the design stage of building construction projects. *Buildings*, *13*(4), 1079. https://doi.org/10.3390/buildings13041079

Craveiro, M., & Domingues, L. (2025). Artificial intelligence on project management performance domains. *Procedia Computer Science*, *256*, 1583-1590. https://doi.org/10.1016/j.procs.2025.02.294

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage.

Cui, Y., Ma, Z., Wang, L., Yang, A., Liu, Q., Kong, S., & Wang, H. (2023). A survey on big data-enabled innovative online education systems during the COVID-19 pandemic. *Journal of Innovation & Knowledge*, *8*(1), 100295. https://doi.org/10.1016/j.jik.2022.100295

Datta, S. D., Islam, M., Sobuz, M. H. R., Ahmed, S., & Kar, M. (2024). Artificial intelligence and machine learning applications in the project lifecycle of the construction industry: A comprehensive review. *Heliyon*, *10*(5), e26888. https://doi.org/10.1016/j.heliyon.2024.e26888

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data–Evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63-71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Feng, N. (2022). The influence mechanism of BIM on green building engineering project management under the background of big data. *Applied Bionics and Biomechanics*, 8227930. https://doi.org/10.1155/2022/8227930

Greeshma, A. S., & Edayadiyil, J. B. (2022). Automated progress monitoring of construction projects using machine learning and image processing approach. *International*

Conference on Advances in Construction Materials and Structures, *65*, 554-563. https://doi.org/10.1016/j.matpr.2022.03.137

Hsu, C.-C., & Sandford, B. A. (2007). The delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation*, *12*(1), 10. https://doi.org/10.7275/PDZ9-TH90

Huang, Y., Shi, Q., Zuo, J., Pena-Mora, F., & Chen, J. (2021). Research status and challenges of data-driven construction project management in the big data context. *Advances in Civil Engineering*, 6674980. https://doi.org/10.1155/2021/6674980

Jakobsen, S. M., Momme, M. B., & Kadenic, M. D. (2025). Exploring strategies for successful implementation of IT projects: Integrating project management practices and human factors. *Procedia Computer Science*, *256*, 1493-1504. https://doi.org/10.1016/j.procs.2025.02.283

Kerzner, H. (2025). *Project management: A systems approach to planning, scheduling, and controlling*. John Wiley & Sons.

Krueger, R. A., & Casey, M. A. (2014). *Focus groups: A practical guide for applied research* (5.ª ed.). Sage.

Niederberger, M., & Spranger, J. (2020, 21 de septiembre). Delphi technique in health sciences: A map. *Frontiers in Public Health*, *8*, Article 457. https://doi.org/10.3389/fpubh.2020.00457

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management*, *42*(1), 15-29. https://doi.org/10.1016/j.im.2003.11.002

Omokhua, D., Mayouf, M., Ashayeri, I., Ekanayake, E., & Zalloom, B. (2025). Adoption of BIM in architectural firms in Nigeria: A survey of current practices, challenges and enablers. *Buildings*, *15*(24), 4547. https://doi.org/10.3390/buildings15244547

Pan, M., Yang, Y., Zheng, Z., & Pan, W. (2022). Artificial intelligence and robotics for prefabricated and modular construction: A systematic literature review. *Journal of Construction Engineering and Management*, *148*(9), 03122004. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002324

Parsamehr, M., Perera, U. S., Dodanwala, T. C., Perera, P., & Ruparathna, R. (2023). A review of construction management challenges and BIM-based solutions: Perspectives from the schedule, cost, quality, and safety management. *Asian Journal of Civil Engineering*, *24*, 353-389. https://link.springer.com/article/10.1007/s42107-022-00501-4

Qian, Z., Yang, X., Xu, Z., & Cai, W. (2021). Research on key construction technology of building engineering under the background of big data. *Journal of Physics: Conference Series*, *1802*, 032003. https://doi.org/10.1088/1742-6596/1802/3/032003

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., & Girolami, M. (2020). Construction with digital twin information systems. *Data-Centric Engineering*, *1*, e14. https://doi.org/10.1017/dce.2020.16

Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2019). *Data mining for business analytics: Concepts, techniques and applications in Python*. John Wiley & Sons.

Shojaei, R. S., Oti-Sarpong, K., & Burgess, G. (2023). Enablers for the adoption and use of BIM in main contractor companies in the UK. *Engineering, Construction and Architectural Management*, *30*(4), 1726-1745. https://doi.org/10.1108/ECAM-07-2021-0650

Shokouhi, M., & Bachari, M. S. (2025). An overview of the aspects of sustainability in project management. *Progress in Engineering Science*, *2*(1), 100048. https://doi.org/10.1016/j.pes.2024.100048

Sivarajah, U., Kumar, S., Kumar, V., Chatterjee, S., & Li, J. (2024). A study on big data analytics and innovation: From technological and business cycle perspectives. *Technological Forecasting and Social Change*, *202*, 123328. https://doi.org/10.1016/j.techfore.2024.123328

Sompolgrunk, A., Banihashemi, S., Golzad, H., & Nguyen, K. L. (2024). Strategic alignment of BIM and big data through systematic analysis and model development. *Automation in Construction*, *168*, 105801. https://doi.org/10.1016/j.autcon.2024.105801

Wu, L., & AbouRizk, S. (2023). Towards construction's digital future: A roadmap for enhancing data value. En S. Walbridge, M. Nik-Bakht, K. T. W. Ng, M. Shome, M. S. Alam, A. el Damatty & G. Lovegrove (Eds.), *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021* (pp. 225-238). CSCE. https://doi.org/10.13140/RG.2.2.29564.67205

Zabala-Vargas, S., Jaimes-Quintanilla, M., & Jimenez-Barrera, M. H. (2023). Big data, data science, and artificial intelligence for project management in the architecture, engineering, and construction industry: A systematic review. *Buildings*, *13*(12), 2944. https://doi.org/10.3390/buildings13122944

ARTÍCULOS DE REVISIÓN

Post-Quantum Security Frameworks for Internet of Things Systems: A Layered Narrative Review of Architectures, Protocols, Trust, and Emerging Challenges

Rodrigo Jara Espinoza
https://orcid.org/0009-0004-0811-9206
20224280@aloe.ulima.edu.pe
Laboratorio SAP, Universidad de Lima, Perú

Yohamin Nafit Pimentel Alarcon
https://orcid.org/0009-0006-8068-2874
20215176@aloe.ulima.edu.pe
Laboratorio SAP, Universidad de Lima, Perú

Angelo Taco-Jimenez
https://orcid.org/0000-0001-9806-5379
ataco@ulima.edu.pe
Universidad de Lima, Perú

Fabricio Martin Chavez Rodriguez
https://orcid.org/0009-0002-6080-7642
20224387@aloe.ulima.edu.pe
Laboratorio SAP, Universidad de Lima, Perú

Received: May 16, 2026 / Accepted: June 15, 2026
doi: https://doi.org/10.26439/interfases2026.n023.8810

ABSTRACT. Quantum computing poses a significant threat to classical asymmetric cryptography, which is essential for ensuring confidentiality, authentication, and key exchange in contemporary digital infrastructures. Although post-quantum cryptography (PQC) provides mechanisms that resist quantum attacks, its implementation in Internet of Things (IoT) systems is challenged by constrained resources, including limitations in computation, memory, energy, latency, and bandwidth, and the heterogeneity of devices. This paper offers a comprehensive narrative review of PQC approaches applicable to IoT, systematically organizing 30 peer-reviewed studies published between 2022 and 2026 across four layers: device, communication, distributed trust, and application. Additionally, the review examines two cross-cutting dimensions, privacy and side-channel resistance. The analysis indicates a significant prevalence of lattice-based schemes, hybrid strategies, and integrations with blockchain technology, zero-knowledge proofs, federated learning, homomorphic

encryption, AI, and Zero Trust architectures. Notably, key gaps remain in side-channel evaluation, migration pathways, deployment costs, and real-world validation—issues that are particularly critical given the long lifecycles of IoT devices and the ongoing threat of "harvest now, decrypt later" attacks.

Keywords: post-quantum cryptography / Internet of Things / lattice-based cryptography / side-channel attacks / privacy

MARCOS DE SEGURIDAD POSCUÁNTICA PARA SISTEMAS DE LA INTERNET DE LAS COSAS: UNA REVISIÓN NARRATIVA ESTRATIFICADA DE ARQUITECTURAS, PROTOCOLOS, CONFIANZA Y DESAFÍOS EMERGENTES

RESUMEN. La computación cuántica representa una amenaza para la criptografía asimétrica clásica, que sustenta la confidencialidad, autenticación e intercambio de claves en las infraestructuras digitales modernas. Si bien la criptografía poscuántica (PQC) ofrece mecanismos resistentes a adversarios cuánticos, su despliegue en sistemas IoT se ve dificultado por recursos limitados —cómputo, memoria, energía, latencia y ancho de banda— y la heterogeneidad de dispositivos. Este artículo presenta una revisión narrativa de enfoques PQC para IoT, organizando treinta estudios revisados por pares (2022-2026) en cuatro capas (dispositivo, comunicación, confianza distribuida y aplicación) y dos dimensiones transversales (privacidad y resistencia a ataques de canal lateral). El análisis revela una predominancia de esquemas basados en retículos, estrategias híbridas e integraciones con blockchain, pruebas de conocimiento cero, aprendizaje federado, cifrado homomórfico, IA y arquitecturas Zero Trust. Las brechas abiertas más relevantes incluyen la evaluación de canales laterales, rutas de migración, costos de despliegue y validación en entornos reales —aspectos críticos dado el largo ciclo de vida de los dispositivos IoT y la amenaza activa de ataques *harvest now, decrypt later*.

PALABRAS CLAVE: criptografía poscuántica / internet de las cosas / criptografía basada en retículas / ataques de canal lateral / privacidad

1. INTRODUCTION

Quantum computing presents a significant challenge to the security of digital systems, primarily due to its potential to undermine classical asymmetric cryptography. In response to this threat, post-quantum cryptography has emerged as a viable approach to maintaining essential guarantees such as confidentiality, authentication, and integrity in the face of adversaries equipped with quantum capabilities. Transitioning to post-quantum cryptography within Internet of Things (IoT) systems is particularly complex, as these devices and networks frequently operate under constraints related to computation, memory, energy, latency, and bandwidth. Furthermore, the high degree of heterogeneity in platforms, protocols, and application domains exacerbates this challenge. As a result, migrating to post-quantum cryptography in IoT encompasses not only cryptographic concerns but also significant architectural and operational complexities.

Recent literature examines post-quantum security in IoT from a variety of perspectives. These include key encapsulation mechanisms, digital signatures designed for constrained hardware, the integration of post-quantum cryptography (PQC) into protocols and handshakes, post-quantum authentication, blockchain-based trust, privacy-preserving computation, federated learning, homomorphic encryption, and zero-knowledge proofs, among others. However, these contributions remain scattered across various technical layers, application domains, platforms, and evaluation criteria. This diversity underscores that the transition to PQC involves more than merely replacing vulnerable algorithms; it necessitates a thorough analysis of how new cryptographic primitives impact the device, communication, trust, and application levels of IoT systems. Consequently, a narrative synthesis is essential for organizing these approaches, identifying common patterns, and determining existing research gaps.

This review presents a narrative survey rather than a systematic literature review. It intentionally avoids exhaustive coverage or a formal screening protocol, aiming instead for a thematically coherent synthesis of recent peer-reviewed studies. The review organizes post-quantum security approaches into a layered analytical architecture comprising four functional layers: the device, communication, distributed trust, and application layers. Additionally, it incorporates the cross-cutting dimensions of privacy and resistance to side-channel attacks. This organization does not constitute a definitive taxonomy; rather, it serves as an analytical framework for comparing diverse contributions within a unified structure.

The contributions of this review are threefold. First, it systematically organizes recent studies on PQC-IoT based on the functional levels at which post-quantum mechanisms are implemented. Second, it delineates the primary cryptographic families and enabling technologies addressed in the literature, including lattice-based, hash-based, code-based, hybrid, blockchain-based, zero-knowledge, federated learning, and homomorphic encryption approaches. Third, it highlights persistent gaps concerning side-channel resistance,

longitudinal migration, operational deployment costs, and validation within real-world IoT environments.

This article employs a narrative review design in Section 2 and presents a conceptual framework in Section 3. In Sections 4 through 6, the authors describe the layered analysis, and in Section 7, they provide a comparative synthesis. The conclusions are explained in Section 8.

2. NARRATIVE REVIEW DESIGN AND CORPUS CONSTRUCTION

This study utilizes a narrative survey design rather than a systematic review. It does not employ a reproducible screening protocol; instead, it constructs a thematically selected corpus to organize diverse contributions and identify recurring research gaps. This approach is suitable given the field's diversity, which encompasses optimization of cryptographic primitives, blockchain technology, federated learning, artificial intelligence, edge computing, homomorphic encryption, and zero-knowledge proofs.

An exploratory search was conducted in the databases Scopus, IEEE Xplore, and Web of Science (initially accessed February 10, 2026), focusing on publications from 2021 to 2026. The search parameters included articles published in English, categorized under the Computer Science subject area, and addressing topics related to post-quantum cryptography and IoT security, such as authentication, key exchange, digital signatures, architectures, frameworks, and reviews. This search resulted in 210 references from Scopus, 153 from IEEE Xplore, and 105 from Web of Science. The collected references were organized in Rayyan for preliminary screening of titles, keywords, and abstracts, followed by a manual thematic refinement process.

The final corpus comprised 30 peer-reviewed articles published between 2022 and 2026, which were selected using purposive thematic criteria. This selection prioritized studies that explicitly linked PQC with IoT environments, including the Industrial Internet of Things (IIoT), Internet of Medical Things (IoMT), edge computing, fog/cloud IoT, wireless IoT, and embedded systems. Additionally, the corpus included articles addressing pertinent security functions such as authentication, key exchange, digital signatures, access control, privacy-preserving computation, blockchain-based trust, protocol integration, and experimental evaluation on constrained platforms.

Each article underwent systematic organization through the use of a documentary extraction form that captured essential bibliographic data, identified the security focus, categorized the cryptographic algorithm, specified the IoT domain, noted the evaluation platform, and reported metrics, results, advantages, limitations, and future directions. Additionally, a complementary structured record of technical evidence was created, including key tables and figures that detailed performance, energy consumption, memory usage, latency, throughput, security levels, and implementation constraints. This

structured approach facilitated the cross-checking of incorporated information, thereby mitigating the risk of unsupported comparisons or unverified attributions.

The studies were organized according to the layered analytical architecture described in Section 3.3, and assigned to the layer that corresponded to their primary contribution. The analysis also addressed any cross-layer overlaps.

The corpus does not provide a comprehensive overview of PQC–IoT research. The deliberate selection criteria and restrictions regarding language, source, document type, and time period may have resulted in the exclusion of pertinent studies. Additionally, the variability in platforms, protocols, metrics, security levels, and experimental conditions hinders direct numerical comparison across studies. Consequently, this analysis predominantly employs a qualitative and comparative approach, with quantitative data presented only in the context of the studies that reported them.

3. CONCEPTUAL FRAMEWORK FOR POST-QUANTUM SECURITY IN LAYERED IOT SYSTEMS

3.1 Quantum Threat and Post-Quantum Cryptography

The advancement of quantum computing requires a reassessment of the security assumptions that underpin contemporary digital services. Two quantum algorithms are pivotal to understanding this threat model. Shor's algorithm poses a significant risk to classical public-key schemes that rely on integer factorization and discrete logarithms, including RSA, ECC, ECDSA, and Diffie–Hellman. Meanwhile, Grover's algorithm effectively reduces the security of symmetric encryption and hash functions by a quadratic factor, thereby necessitating larger key sizes to maintain equivalent security levels. These risks are further exacerbated by the "*harvest now, decrypt later*" scenario, in which encrypted data collected today may be decrypted in the future if sufficiently advanced quantum capabilities become available. This concern is particularly relevant for systems with extended operational lifespans, such as IoT deployments (Halak et al., 2024; Shim, 2024).

Post-quantum cryptography (PQC) encompasses public-key mechanisms that are grounded in mathematical problems for which no efficient quantum algorithms are currently known. Within the examined corpus, lattice-based mechanisms are prominent in key encapsulation mechanisms (KEMs), digital signatures, authentication, and access control. Notable examples include Kyber, Dilithium, Falcon, Falcon-M, NTRU, Rudraksh, and Ring-ExpLWE (Castiglione et al., 2025; Halak et al., 2024; Kerimbayeva et al., 2025; Kundu et al., 2025).

Other cryptographic families serve more specialized functions: hash-based signatures, such as SPHINCS+, XMSS, and LMS, are primarily employed in acceleration and embedded blockchain applications (Marchsreiter, 2025; Wu et al., 2025); code-based

mechanisms, including Classic McEliece, BIKE, and HQC find applications in hybrid Transport Layer Security (TLS) protocols and comparative KEM evaluations (Cruz-Piris et al., 2025; Zafar & Iqbal, 2025); and MPC-in-the-Head signatures are represented by AIMer as a non-lattice alternative (Kwon et al., 2025).

The corpus discusses multivariate and isogeny-based approaches primarily as comparative references rather than as leading evaluated options for IoT hardware. In particular, the treatment of SIKE reflects caution in light of recent cryptanalytic advances that have affected its security (Cruz-Piris et al., 2025; Shim, 2024). Hybrid approaches that combine post-quantum cyptography (PQC) with classical mechanisms, such as X25519 or AES, play a crucial role in ensuring interoperability during the migration process (Ojetunde et al., 2025; Zafar & Iqbal, 2025). While quantum key distribution (QKD) is recognized as a related quantum-security technology, it does not fall under the strict definition of PQC, as it relies on specific physical infrastructure rather than on post-quantum public-key primitives (Aneesh Kumar et al., 2025). Table 1 consolidates these various families of approaches.

Table 1

Cryptographic Families and Related Approaches Identified in the Corpus

Family or approach	Examples mentioned in the corpus	Main relevance for PQC-IoT
Lattice-based	Kyber, Dilithium, Falcon, Falcon-M, NTRU, Rudraksh, Ring-ExpLWE	KEMs, signatures, authentication, access control
Hash-based	SPHINCS+, XMSS, LMS	Signatures, blockchain, acceleration
Code-based	Classic McEliece, BIKE, HQC	Hybrid TLS, KEM comparison, migration
MPC-in-the-Head	AIMer	Non-lattice signature alternative
Multivariate	Rainbow	Comparative reference with cryptanalytic caution; limited role in the corpus
Isogeny-based	SIKE	Historical/comparative reference with cryptanalytic caution
Hybrid PQC + classical approaches	PQC with X25519, AES, or other classical mechanisms	Interoperability during migration
QKD	Quantum key distribution	Adjacent quantum-security technology; not PQC in the strict sense

Note. The table serves solely for conceptual orientation purposes. The summary of cryptographic families presented is intended as a conceptual overview, rather than a comparison of performance.

3.2 IoT Systems: Constraints, Attack Surfaces, and Traditional Architectural Models

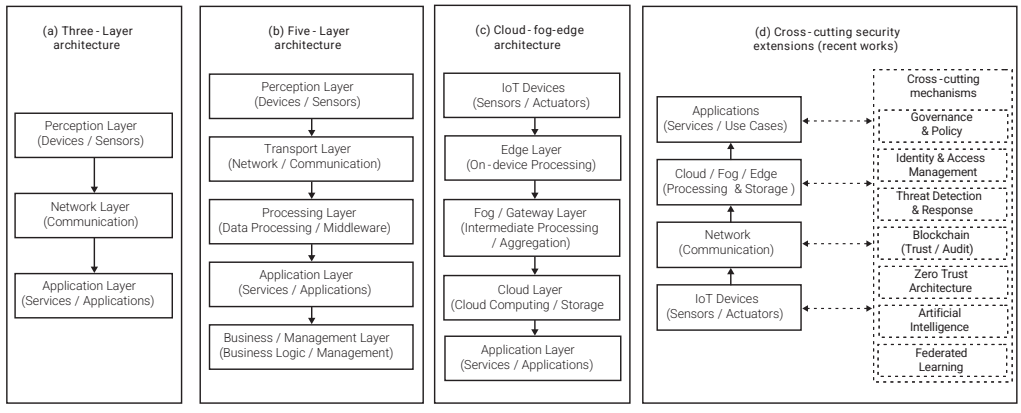
The adoption of PQC in the IoT differs from its implementation in general-purpose infrastructures due to the limitations faced by end nodes, which often operate with restricted RAM, flash memory, computational capacity, battery life, and bandwidth. These constraints

are particularly significant, as many PQC mechanisms use larger keys, ciphertexts, and signatures than classical schemes, and computational procedures that may impose considerable demands on constrained microcontrollers, sensors, and embedded devices (Halak et al., 2024; Kundu et al., 2025).

The limitations associated with IoT devices are exacerbated by their broad attack surface. These devices are often deployed in physically accessible environments and communicate over exposed or low-power channels. Consequently, they may encounter various types of attacks, including logical attacks such as man-in-the-middle, replay, and impersonation attacks, as well as physical attacks such as side-channel leakage (Marchsreiter, 2025; Shim, 2024). Therefore, the adoption of PQC in IoT must consider not only algorithmic security but also robustness of implementations and the context of deployment.

Figure 1

Classical Architectural Models of IoT and Contemporary Cross-Cutting Security Extensions



Note. The first three panels depict general architectural models for IoT, while the fourth panel demonstrates mechanisms such as blockchain, Zero Trust, artificial intelligence, federated learning, identity management, governance, and threat detection. These mechanisms operate across multiple layers rather than adhering to a single architectural level.

IoT systems have traditionally been represented through layered models. These include three-layer architectures focused on perception, network, and application functions; five-layer architectures that incorporate transport, processing, or business levels; and cloud–fog–edge models that efficiently distribute computation across devices, gateways, edge nodes, and cloud services. Recent studies have expanded these models by integrating cross-cutting mechanisms for governance, identity, threat detection, blockchain-based trust, Zero Trust architectures, artificial intelligence, and federated learning (Alatawi, 2025; Aleisa, 2025; Elkhodr, 2025). However, the original designs did not take into account post-quantum constraints, where concerns such as cryptographic overhead,

migration interoperability, and implementation security must now be prioritized. Figure 1 presents a summary of three general IoT architectural models along with one cross-cutting extension identified in the recent literature.

Consequently, the migration to PQC within IoT systems involves more than merely substituting vulnerable cryptographic primitives with quantum-resistant alternatives. The adoption of larger keys, ciphertexts, and signatures may lead to increased handshake size, latencies, and bandwidth consumption. Furthermore, the computational and memory overhead associated with these changes can influence whether operations are conducted locally or delegated to gateways or edge nodes. Additionally, existing legacy infrastructures require hybridization strategies during the migration process. These complexities underscore the need for an analytical framework that evaluates PQC approaches based on their functional impact on IoT systems, rather than viewing them solely as isolated algorithms (Cruz-Piris et al., 2025; Halak et al., 2024).

3.3 Layered Analytical Architecture for Organizing the Review

Drawing on these considerations, this review adopts a layered analytical architecture to systematically organize post-quantum security approaches for IoT. This architecture employs traditional IoT models as a foundational basis and reconfigures them to address the specific challenges posed by PQC, including constrained implementation, protocol migration, distributed trust, application-level requirements, privacy, and side-channel resistance. Rather than serving as a definitive taxonomy of the field, this framework aims to provide an interpretive lens for situating diverse contributions within a cohesive structure.

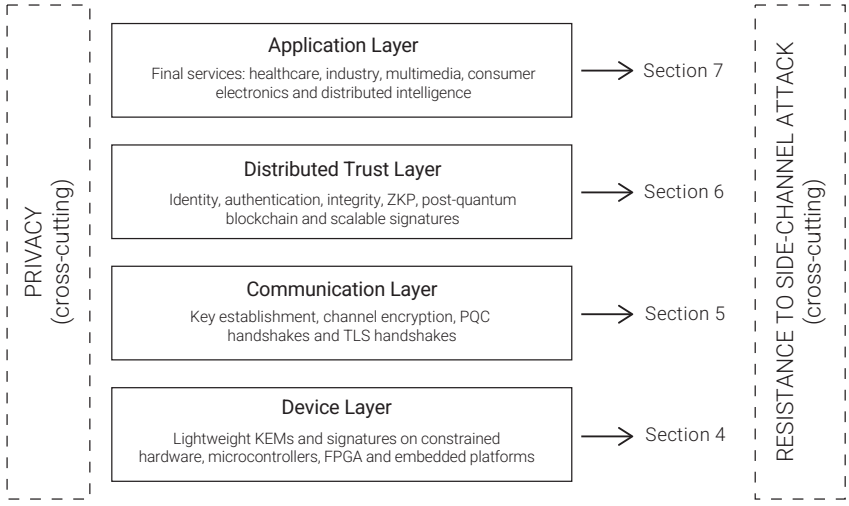
The architecture consists of four functional layers. The device layer encompasses end nodes and the local execution of PQC primitives on constrained hardware, where limitations in memory, energy, and computation are most pronounced. The communication layer facilitates secure data exchange among devices, gateways, edge nodes, and remote services, incorporating key establishment, channel encryption, transport protocols, and hybrid handshakes. The distributed trust layer incorporates mechanisms for identity, authentication, authorization, integrity, traceability, and decentralized verification, utilizing methods such as post-quantum blockchain, zero-knowledge proofs, access control, and scalable signatures. Finally, the application layer encompasses the IoT application domains that benefit from these security guarantees, including industrial, medical, wireless, multimedia, smart-grid, and AI-enabled environments.

Two cross-cutting dimensions enhance these layers. Privacy refers to guarantees that extend beyond channel confidentiality, including identity protection, privacy-preserving computation, federated learning, homomorphic encryption, differential privacy, and zero-knowledge mechanisms. Resistance to side-channel attacks pertains to the robustness of implementations against physical leakage, particularly in constrained

devices where cryptographic operations may be vulnerable to timing, power, electromagnetic, or fault-based analysis. These dimensions are categorized as cross-cutting because they influence multiple layers rather than being confined to a single system function. Figure 2 synthesizes the analytical architecture, illustrating the four functional layers alongside the two cross-cutting dimensions.

Figure 2

Layered Analytical Architecture for Organizing Post-Quantum Security Approaches in IoT Environments



Note. The four layers delineate the functional levels at which PQC affects IoT systems. Furthermore, privacy and resistance to side-channel attacks serve as cross-cutting dimensions within this framework. This model provides an overview to help readers compare and conceptualize the framework layers. It does not prescribe a deployment architecture.

This framework establishes the central organizing principle for the remainder of the review. Section 4 examines device-layer studies focused on lightweight Key Encapsulation Mechanisms (KEMs) and signatures tailored for constrained hardware. Section 5 analyzes communication-layer proposals that incorporate PQC stacks, protocols, and handshake processes. Section 6 engages in a discussion of distributed trust, authentication, blockchain, and mechanisms for perserving privacy. Section 7 provides a synthesis of research spanning various cryptographic families, application domains, and layers while also identifying exisiting research gaps. Finally, Section 8 delineates the conclusions drawn from this review.

4. DEVICE-LAYER APPROACHES: KEMS AND SIGNATURES ON CONSTRAINED HARDWARE

The device layer includes IoT end nodes and serves as the point where the costs associated with PQC become most evident. Low-end microcontrollers (MCUs), battery-powered sensors, embedded devices, and field-programmable gate arrays (FPGAs) operate within modest constraints of computational power, memory capacity, and energy consumption. In contrast, PQC primitives require larger key sizes, ciphertexts, and signatures compared to their classical counterparts. This corpus addresses the inherent tension by examining KEMs, digital signatures, and comparative evaluations conducted on embedded hardware.

Table 2 organizes the device-layer studies according to their contributions, platforms, and reported metrics. This table serves a descriptive purpose rather than a comparative one due to the heterogeneity of devices, implementations, and measurement settings. Figure 3 categorizes these studies into KEMs, digital signatures, and comparative evaluations.

Table 2

Assessment of Platforms and Metrics Reported in Studies within the Device Layer Corpus

Main contribution	Evaluated platform	Platform type	Reported metric categories	Studies
KEM (Rudraksh)	Virtex-7 and Artix-7 FPGA	FPGA	Area (LUT, FF, BRAM, DSP), latency (KG/Enc/Dec), frequency, time–area product	Kundu et al. (2025)
KEM (NTRU on Contiki-NG)	EFR32MG12 (Cortex-M4)	32-bit MCU	Cycles, stack, flash, power, energy, ciphertext size	Señor et al. (2022)
KEM (NTRUEncrypt)	Raspberry Pi 3B+ (Cortex-A53)	ARM Cortex-A SoC	Encryption and decryption times, key sizes, security levels	Al-Doori & Al-Gailani (2023)
KEM (Ring-ExpLWE)	Cortex-M0/M3 (software); Spartan-6 and Virtex-7 FPGA (hardware)	MCU + FPGA	Execution time, cycles, key and ciphertext sizes	Xu et al. (2022)
Signature (Falcon-M)	PC with Intel Core i7-9700K; Cortex-M4 and RISC-V as targets (not evaluated)	General-purpose PC; 32-bit MCU (target)	Key generation, signing, and verification times; RAM estimates	Kerimbayeva et al. (2025)
Signature (AIMer)	Cortex-M4 (Nucleo-L4R5ZI), Raspberry Pi 5, and MacBook Pro M4 (AArch64)	MCU + ARM SoC	KG/Sign/Verify cycles, stack, peak memory, code size, key and signature sizes	Kwon et al. (2025)

(*continues*)

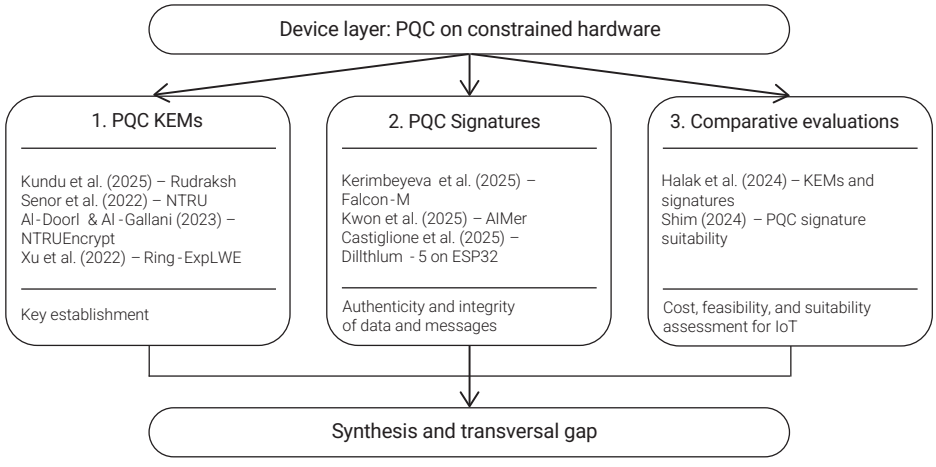
(*continued*)

Main contribution	Evaluated platform	Platform type	Reported metric categories	Studies
Signature (Dilithium-5)	ESP32 / ESP-WROOM-32D	Embedded platform with light-weight SoC	Key generation, signing, and verification times; key and signature sizes	Castiglione et al. (2025)
Comparative evaluation (KEMs and signatures)	STM32F469NI (Cortex-M4); RPi 3B+ and RPi Pico W as auxiliary nodes	32-bit MCU and ARM SoC	TLS/handshake latency, heap, stack, energy	Halak et al. (2024)
Review and comparative evaluation (PQC signatures)	Cortex-M3/M4, AArch64 (Cortex-A72, Apple M1), AVX2	Mixed platform	KG/Sign/Verify cycles, key and signature sizes, stack, energy; SCA (qualitative)	Shim (2024)

Note: SCA = side-channel attack.

Figure 3

Integration of PQC Approaches at the Device Layer



Note. The figure presents a consise overview of the analytical organization of device-layer studies categorized into KEMs, digital signatures, and comparative evaluations. It is not a performance ranking of the algorithms, authors, or platforms.

4.1 PQC KEMs for Constrained Hardware

The KEM studies employ three complementary strategies: designing compact mechanisms for lightweight hardware, implementing existing primitives on constrained platforms, and exploring algorithmic variants that are adapted for embedded execution.

Kundu et al. (2025) introduce Rudraksh, a compact MLWE-KEM designed specifically for lightweight hardware, which they evaluate on Virtex-7 and Artix-7 FPGAs. This design integrates lightweight modular arithmetic and employs ASCON-XOF as a pseudorandom function, while also conducting parameter exploration against Kyber and NewHope at NIST security level 1. The findings indicate that Rudraksh has a smaller area footprint compared to selected optimized references from Kyber.

Two studies assess the performance of NTRU-family mechanisms on ARM platforms. Señor et al. (2022) successfully integrate ntruhps2048509 into a Cortex-M4 microcontroller running Contiki-NG, optimizing it for IIoT-oriented WSNs and reporting metrics such as cycles, stack memory usage, flash requirements, power consumption, and energy usage. Concurrently, Al-Doori and Al-Gailani (2023) evaluate NTRUEncrypt on a Raspberry Pi 3B+ to facilitate edge/fog integration, providing insights into encryption and decryption times, as well as key sizes across different security levels. Collectively, these studies illustrate the application of the same cryptographic family within varied device-layer contexts and metric profiles.

Xu et al. (2022) introduce Ring-ExpLWE, a variant of Ring-LWE that employs a discrete exponential error distribution to enhance the security profile without incurring additional hardware costs. The authors evaluate this proposal using software implementations on Cortex-M0 and Cortex-M3, as well as through hardware evaluations on Spartan-6 and Virtex-7 FPGAs.

Overall, these studies affirm the feasibility of KEM at the device layer under controlled conditions; however, they identify a persistent limitation that remains to be addressed. This limitation pertains to the evaluation of side-channel attacks on actual hardware, which remains insufficient, as further discussed in Section 4.3.

4.2 PQC Signatures on Microcontrollers and Embedded Platforms

Digital signatures typically pose greater challenges than KEMs for constrained devices. This body of research identifies three strategies to mitigate this bottleneck: Falcon-M and Dilithium-5 within the lattice-based paradigm, and AIMer as an alternative based on the MPC-in-the-Head model.

Kerimbayeva et al. (2025) introduce Falcon-M, a lightweight variant of the Falcon signature scheme that employs random polynomials and FFT, intended to simplify key generation and reduce the memory footprint. While the targeted platforms include Cortex-M4 and RISC-V, the evaluation was conducted on an Intel Core i7-9700K, with preliminary RAM estimates indicating potential for embedded deployments. Consequently, this study provides an algorithmic profile that necessitates further validation on the intended hardware.

Kwon et al. (2025) optimize AIMer, a signature scheme utilizing the MPC-in-the-Head approach grounded in the AIM2 symmetric primitive. They evaluated their implementation on the Cortex-M4 and extended the analysis to AArch64 platforms, including Raspberry Pi 5 and Apple Silicon, reporting metrics such as cycles, stack usage, peak memory consumption, code size, and key/signature sizes. This AIMer work expands the discussion at the device-layer to incorporate analyses beyond lattice-based signatures.

Castiglione et al. (2025) integrate Dilithium-5 into an ESP32 microcontroller for a biomedical blockchain application. By employing PQClean and implementing polynomial optimizations via Kronecker substitution, the study reports that key generation, signing, and verification times range from tens to hundreds of milliseconds, contingent upon the specific configuration used.

Collectively, these studies indicate that the feasibility of signatures is heavily influenced by the choice of algorithmic family, implementation strategy, and platform. Dilithium-5 demonstrates successful deployment on commercially available, low-cost hardware, while Falcon-M highlights the benefits of algorithmic simplification, albeit with embedded validation still pending. Furthermore, AIMer broadens the design landscape beyond lattice-based signatures. Additionally, hash-based signatures such as SPHINCS+ are explored in various contexts, including acceleration and embedded blockchain applications.

4.3 Comparative Evaluations and Suitability Reviews for IoT

Halak et al. (2024) and Shim (2024) offer complementary comparative perspectives on PQC. Halak et al. (2024) evaluate PQC schemes on the STM32F469NI microcontroller, examining KEMs such as Kyber, Saber, and FrodoKEM, as well as signatures including Dilithium, Falcon, SPHINCS+, and SIKE. They also reference classical algorithms such as ECDHE/ECDSA with secp256r1 and AES-256. Their analysis includes metrics such as TLS and handshake latency, heap and stack memory consumption, and energy usage, revealing that lattice-based KEMs typically offer better performance than PQC signatures on constrained hardware.

In contrast, Shim (2024) focuses on PQC signatures for IoT devices, evaluating performance across Cortex-M3, Cortex-M4, AArch64, and AVX2 architectures. This study encompasses the schemes Dilithium, Falcon, SPHINCS+, and UOV, while also noting Rainbow —which has been compromised through algebraic and MinRank attacks (Kerimbayeva et al., 2025)— as well as XMSS, XMSSMT, and LMS in a broader comparative framework.

Together, these studies highlight a significant gap at the device layer. Although several PQC mechanisms can be implemented on constrained devices, experimental evaluations of side-channel leakage in real hardware are infrequently conducted. This gap is addressed further in Section 7.

5. APPROACHES AT THE COMMUNICATION LAYER: PQC STACKS, PROTOCOLS, AND HANDSHAKES

The communication layer facilitates secure data exchange among devices, gateways, and remote services, encompassing key establishment, channel encryption, transport protocols, and legacy stacks. PQC impacts this layer by increasing handshake size, latency, and bandwidth consumption due to the utilization of larger keys, ciphertexts, and signatures. This corpus examines the resulting challenges from three perspectives: the integration of PQC within the existing TLS protocol, the deployment of PQC through ad-hoc combinations of KEM, signatures, and lightweight authenticated encryption over wireless and low-power wide-area networks (LPWAN), and the incorporation of PQC within legacy industrial protocols.

Table 3 categorizes the six communication-layer studies based on protocol or stack, security mechanisms, evaluation environments, and reported metrics. Consistent with the previous section, this table serves a descriptive purpose and should not be interpreted as a direct numerical comparison across diverse networks, platforms, and experimental conditions.

Table 3

Protocols, Security Mechanisms, and Documented Metrics within the Communication Layer

Protocol or stack	Security algorithms and mechanisms used	Evaluation environment	Reported metric categories	Studies
KEMTLS in IIoT	KEMs: BIKE, Kyber, FrodoKEM, HQC; signatures: Dilithium, Falcon, SPHINCS+, HAWK	Raspberry Pi 2 and 3 (IIoT scenarios)	TLS/handshake latency, energy, KG/Enc/Dec and signature times	Cruz-Piris et al. (2025)
Hybrid TLS 1.3 (PQC + classical)	Classic McEliece, BIKE, Kyber; X25519, RSA/ECC; QRNG, MFA	Intel Core i7-9700K PC; emulated IoT scenarios	Handshake latency, CPU, memory, throughput, QRNG entropy	Zafar & Iqbal (2025)
Lightweight PQC stack over NB-IoT (CoAP)	Kyber-768, SPHINCS+-128f, ChaCha20-Poly1305	ESP32-WROOM-32 + SIM7080G	Per-layer latency, current, peak RAM, firmware size	Do & Tran (2026)
IEEE 802.11n; PQC standalone and PQC-AES modes	Kyber, BIKE, HQC; Falcon; AES-256; HMAC-SHA256; HKDF	Client–server over IEEE 802.11n	KEM times, signature times, latency, throughput	Ojetunde et al. (2025)
Hybrid KEM + authenticated encryption framework	Kyber-512 + ASCON	Controlled testbed, 50 runs	Encryption time, RAM, CPU, energy, entropy	Mahdi & Abdullah (2025)

(*continues*)

(*continued*)

Protocol or stack	Security algorithms and mechanisms used	Evaluation environment	Reported metric categories	Studies
Post-SECS/GEM (PQC extension)	Kyber + Dilithium + AES-GCM-256 + HMAC-SHA3	Raspberry Pi 4; Ethernet; emulated network	Encryption/decryption, signing/verification, and key establishment. Evaluation metrics: CPU usage, memory consumption, and energy consumption	Al-Mekhlafi et al. (2026)

Note. This table provides a descriptive overview and does not constitute a direct performance comparison across diverse protocols, networks, and platforms. The following acronyms are defined for clarity: NB-IoT = narrowband IoT; CoAP = Constrained Application Protocol; KEMTLS = key-encapsulation-mechanism TLS; QRNG = quantum random number generator; MFA = multi-factor authentication; HKDF = HMAC-based key derivation function.

5.1 PQC in TLS and Transport Handshakes

Two studies examine post-quantum migration in TLS using complementary strategies: restructuring the handshake process and maintaining hybrid interoperability with classical infrastructure.

Cruz-Piris et al. (2025) evaluate PQC in IIoT by implementing KEMTLS on Raspberry Pi 2 and 3. Their analysis includes various KEMs such as BIKE, Kyber/ML-KEM, FrodoKEM, and HQC, along with digital signatures like Dilithium/ML-DSA, Falcon, SPHINCS+, and HAWK. Their findings indicate that KEMTLS handshakes introduce overhead when compared to classical TLS. Zafar and Iqbal (2025) prioritize interoperability by integrating Classic McEliece and BIKE with X25519, RSA/ECC, QRNG, and MFA within a hybrid TLS 1.3 framework designed for gradual migration.

Collectively, these studies demonstrate that post-quantum transport migration involves more than mere primitive selection; it necessitates deliberate protocol-level decisions regarding handshake structure, hybrid coexistence, and integration with legacy systems.

5.2 PQC in Wireless Networks, LPWAN, and IoT Channels

Three studies in the corpus present initial evidence for deploying PQC stacks in standard wireless networks, LPWAN channels, and generic IoT communications.

In their 2026 study, Do and Tran propose a lightweight hybrid stack for NB-IoT that integrates CRYSTALS-Kyber-768, SPHINCS+-SHA256-128f, and ChaCha20-Poly1305 on the ESP32-WROOM-32 platform using the SIM7080G and CoAP. This approach utilizes session persistence and buffer reuse strategies to mitigate the costs associated with handshakes.

Ojetunde et al. (2025) implement secure IEEE 802.11n communication modes, including standalone PQC and hybrid PQC-AES. They evaluate various algorithms, including Kyber/ML-KEM, BIKE, HQC, Falcon/FN-DSA, AES-256, and HMAC-SHA256. Their findings indicate that lattice-based KEMs show lower execution times compared to BIKE and HQC, and that hybrid PQC-AES configurations become increasingly advantageous as the volume of transmitted data increases.

Mahdi and Abdullah (2025) combine Kyber-512 and ASCON within a controlled testbed, reporting reductions in time, memory usage, CPU consumption, and energy expenditure compared to a configuration that relies solely on Kyber, all while maintaining near-maximum entropy.

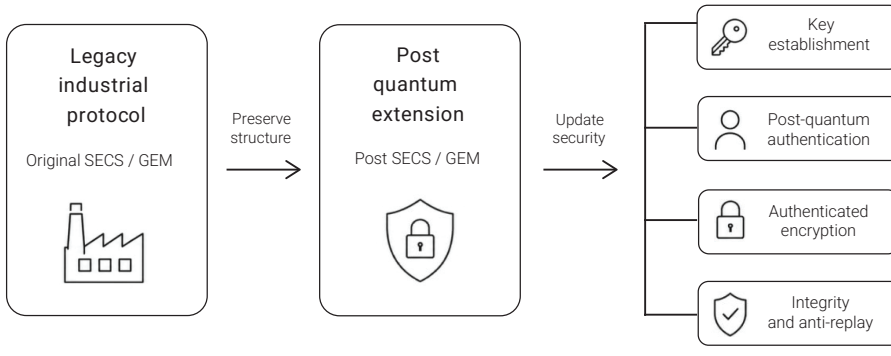
These studies indicate that lattice-based KEMs, when combined with lightweight authenticated encryption, represent a practical solution for constrained channels. In contexts such as Narrowband Internet of Things and broader wireless networks, where session persistence, reuse and compactness are critical, this approach effectively addresses the issue of repeated PQC handshakes that can significantly increase communication costs.

5.3 PQC in Legacy Industrial Protocols

Legacy industrial protocols require the development of migration strategies that enhance security without compromising operational behavior. This corpus addresses this challenge through a specific contribution centered on SECS/GEM in industrial automation.

Figure 4

Adaptation of Legacy Industrial Protocols to the Post-Quantum Paradigm



Note. The figure illustrates the adaptation of an established industrial protocol via a post-quantum extension, which maintains the integrity of the operational structure while enhancing security mechanisms.

Al-Mekhlafi et al. (2026) introduce Post-SECS/GEM, a post-quantum extension that integrates CRYSTALS-Kyber for encapsulation and session-key derivation,

CRYSTALS-Dilithium for authentication, AES-GCM-256 for authenticated encryption, and HMAC-SHA3 for integrity. The evaluation of this framework was conducted on a Raspberry Pi 4 under switched Ethernet and emulated network conditions. The framework effectively combines authentication, confidentiality, integrity, and safeguards against replay and DoS attacks while maintaining the operational structure of SECS/GEM. Figure 4 illustrates this adaptation logic.

This case underscores the necessity of coexistence with legacy infrastructure. In the context of industrial IoT, the migration to PQC typically involves enhancing security protocols rather than replacing them entirely.

6. DISTRIBUTED TRUST, AUTHENTICATION, AND POST-QUANTUM PRIVACY

This section analyzes 14 studies on distributed trust, authentication, and post-quantum privacy. Unlike investigations focused solely on the device and communication layers, this discussion encompasses not only algorithm execution and transport protocols but also mechanisms that uphold identity, authorization, traceability, verification, and privacy at the system level. The studies are characterized into three main themes: authentication and access control; blockchain, scalable signatures, and acceleration; and integrated frameworks for trust and operational privacy.

Table 4 summarizes these studies, detailing their focus, mechanisms, domain, type of contribution, and supporting evidence. This table serves as a navigational tool rather than a direct comparison among proposals that differ in objectives, platforms, and validation methodologies.

Table 4

Methodological Approaches, Mechanisms, and Forms of Contribution in Research Pertaining to the Distributed Trust, Authentication, and Privacy Layer

Focus	Mechanisms and technologies used	Domain	Contribution	Studies
PAKE	LWE (MP-SPHF)	General IoT	Algorithmic analysis	Li & Wang (2022)
Three-party mutual authentication	RLWE	IIoT	Protocol with simulation	Kim et al. (2026)
Three-factor authentication	RLWE + fuzzy extractor	IoMT	Protocol with evaluation	Ahmad et al. (2026)
Access control	RLWE-CP-ABE + Dilithium + AES-GCM	Cloud IoT, healthcare	Experimental evaluation	Poomekum et al. (2025)

(*continues*)

(*continued*)

Focus	Mechanisms and technologies used	Domain	Contribution	Studies
Scalable signatures	Fractional Verkle Trees (FVDS)	Blockchain, IoT firmware	PC-based prototype	Iavich et al. (2025)
Post-quantum blockchain	Dilithium, Falcon, SPHINCS+, XMSS-MT + PKR	IoT, energy, vehicular	Evaluation on embedded devices	Marchsreiter (2025)
Signature acceleration	SPHINCS+ + GPU batching	Cloud IoT, AIoT	GPU benchmarks	Wu et al. (2025)
Signature acceleration	HAETAE + GPU kernel fusion	IIoT, edge/cloud	GPU benchmarks	Wu et al. (2026)
Authentication with ZKP	Schnorr-ZKP + HE + blockchain	General IoT	Simulation	Tawfik et al. (2025)
ZTA + blockchain + ZKP	DQN + blockchain + ZKP	General IoT	Dataset-based simulation	Aleisa (2025)
AI + blockchain + PQC framework	Q-learning + blockchain + hybrid encryption + FL + DT	General IoT	Simulation	Elkhodr (2025)
6G edge + FL + blockchain	CNN-LSTM + Hyperledger + Kyber + Dilithium	Industry 5.0, smart grid	Edge-based evaluation	Alatawi (2025)
Privacy-preserving FL	FedAvg + GRU + HE + DP	IoT networks	Simulation with CICIDS2017	Rahmati & Pagano (2025)
Private inference	PDTE + SWHE (RLWE)	IIoT, healthcare	PC-based evaluation	Kjamilji (2024)

Note. This table provides a descriptive overview and should not be interpreted as a direct comparison of proposals that pursue varying objectives. The following acronyms are defined for clarity: ZKP = zero-knowledge proof; ZTA = Zero Trust architecture; FL = federated learning; HE = homomorphic encryption; DP = differential privacy; PKR = public-key recovery; DQN = deep Q-network; SWHE = somewhat homomorphic encryption; DT = digital twin; CNN = convolutional neural network; LSTM = long short-term memory; GRU = gated recurrent unit.

6.1 Post-Quantum Authentication, Key Management, and Access Control

Four studies contribute to advancing post-quantum authentication, key agreement, and access control mechanisms. Li and Wang (2022) propose a one-round lattice-based PAKE protocol that utilizes the MP-SPHF framework, effectively reducing communication costs for constrained nodes, although the evaluation remains purely analytical. Kim et al. (2026) develop a robust three-party RLWE-based mutual authentication protocol tailored for IIoT, complemented by a comprehensive security analysis and NS-3 simulations. Ahmad et al. (2026) integrate biometrics, passwords, and smart card factors within the RLWE framework for IoMT, employ a fuzzy extractor to mitigate biometric noise, and present results on computational efficiency, communication overhead, and scalability. Poomekum et al. (2025) broaden the discussion of fine-grained access control by employing RLWE-CP-ABE,

Dilithium-2, AES-256-GCM, attribute-hash revocation, salts, epoch-based blacklists, and fog computing assistance to alleviate the burden on end nodes.

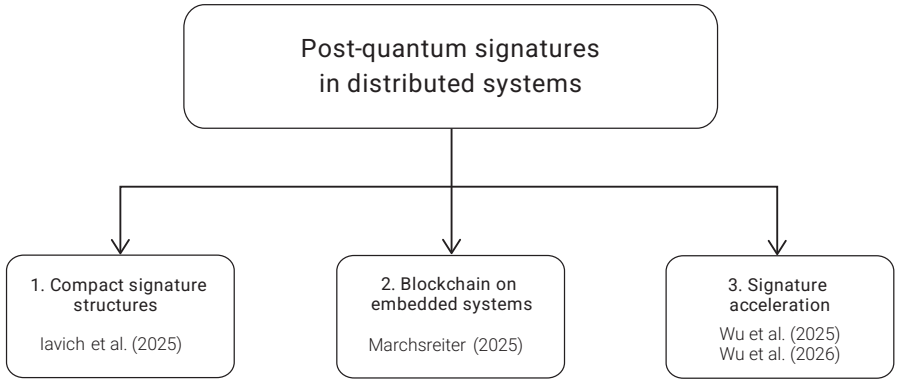
These proposals, while sharing a foundational lattice/RLWE structure, address distinct trust relationships—including password-authenticated key exchange, three-party mutual authentication, multifactor authentication, and attribute-based access control. Consequently, the migration to PQC at this layer is best interpreted as a diverse repertoire of mechanisms tailored to various trust models rather than a singular primitive approach.

6.2 Post-Quantum Signatures for Blockchain, Traceability, and Scalable Processing

In distributed systems such as blockchain, trust is primarily established through signatures that verify records, transactions, firmware updates, or identities. In the context of post-quantum systems, signature size, verification costs, and aggregate processing volume may pose limitations across the entire IoT stack, ranging from constrained end devices to gateways, edge nodes, and cloud services. The research discussed in this subsection addresses these challenges by developing compact signature structures, deploying embedded blockchain solutions, and implementing acceleration techniques for high-volume verification. Figure 5 provides a summary of these three strategies.

Figure 5

Emerging Research Trends in Post-Quantum Signatures for Distributed Systems



Note. This figure categorizes the studies in this subsection based on their primary area of focus, which includes compact signature structures, blockchain applications in embedded systems, and signature acceleration. It is important to clarify that the figure does not serve as a performance comparison among the various proposals presented.

Iavich et al. (2025) propose Fractional Verkle Trees and FVDS as a compact alternative to post-quantum Merkle-tree approaches for scenarios involving numerous signers. Their evaluation is conducted through a Python prototype on a personal computer. Marchsreiter

(2025) assesses post-quantum blockchain signatures on embedded systems, examining variants such as Dilithium, Falcon, SPHINCS+, and XMSS-MT. Additionally, he introduces public-key recovery to reduce on-chain credential storage, utilizing experiments on an Intel Core i7 and a Raspberry Pi 3B. Wu et al. (2025) enhance the performance of SPHINCS+ by implementing CUDA-based intra-block batch processing on RTX GPUs. Furthermore, Wu et al. (2026) optimize HAETAE on the Jetson Xavier and RTX 4090 through techniques such as kernel fusion and coarse-grained parallelism. These studies on acceleration are particularly significant for gateways, edge nodes, and cloud services that handle large volumes of signatures.

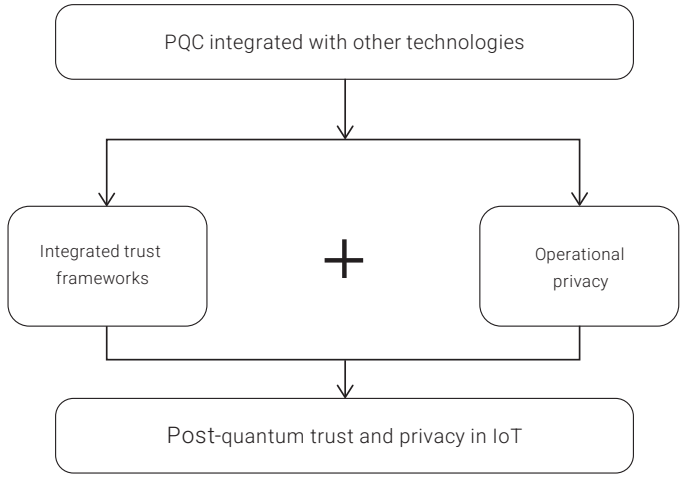
Collectively, these research efforts demonstrate that achieving post-quantum integrity and traceability in distributed IoT systems necessitates a focus on compactness, the feasibility of embedded systems, and the capacity for high-throughput verification.

6.3 Integrated Frameworks for Post-Quantum Trust and Operational Privacy

Six studies examine PQC as an integral component of broader architectures rather than as an isolated primitive. These studies can be divided into two complementary groups: integrated trust frameworks, which combine blockchain, ZKP, Zero Trust, AI, or federated learning to facilitate system-level coordination; and operational privacy mechanisms that safeguard data during training, detection, or inference. Figure 6 illustrates this relationship.

Figure 6

Exploring the Relationship between Integrated Trust Frameworks and Operational Privacy



Note: This figure delineates integrated trust frameworks from operational privacy. Integrated trust frameworks are designed to coordinate security at the system level, whereas operational privacy focuses on protecting data during training, detection, or inference. Both approaches are conceptualized as complementary rather than mutually exclusive categories.

Tawfik et al. (2025) integrate Schnorr-type zero-knowledge proofs over Module-SIS framework with homomorphic encryption and blockchain technologies to facilitate traceable and privacy-preserving authentication. Aleisa (2025) introduces QBC-ZKPAF, a Zero Trust architecture that combines ZKP, blockchain, and reinforcement learning, and evaluates it on the Edge-IIoTset. Elkhodr (2025) presents IASF-IoT, which orchestrates artificial intelligence, blockchain, post-quantum hybrid encryption, digital twins, and federated learning through a Q-learning-based security orchestrator across more than 1,000 simulated heterogeneous devices. Alatawi (2025) employs a similar convergence to Industry 5.0 by developing EdgeGuard-IoT, which integrates secure federated learning, Hyperledger Fabric, Kyber, Dilithium, and Zero Trust authentication on Raspberry Pi 4 and Jetson Nano edge nodes. While these frameworks share a compositional logic, they differ in their respective domain, evaluation methodologies, and the realism of their deployment scenarios.

Operational privacy is illustrated through the findings of two studies. Rahmati and Pagano (2025) integrate federated learning, homomorphic encryption, and differential privacy for IoT threat detection, utilizing the CICIDS2017 dataset in conjunction with a GRU model. Their research reports enhanced detection metrics and a reduction in membership inference success. In another study, Kjamilji (2024) introduces the PDTE scheme, which is based on the Ring-LWE approach and operates under the decision-RLWE model. This scheme is evaluated on benchmarks in healthcare, spam detection, and image recognition, using an Intel Core i9 processor.

Overall, the corpus indicates that achieving post-quantum privacy relies on the composition of multiple primatives rather than on a single primitive. However, the transferability across different domains and the behavior in large-scale production deployments remain insufficiently documented.

7. COMPARATIVE SYNTHESIS, APPLICATION DOMAINS, AND RESEARCH AGENDA

This section synthesizes the findings from various cryptographic families, application domains, and layered architectures, and identifies research gaps based on the studies analyzed in Sections 4 to 6.

7.1 Cryptographic Families Present in the Corpus

The corpus clearly demonstrates the predominance of lattice-based mechanisms, which serve multiple functions. These mechanisms are integral to KEMs such as Kyber and NTRU, digital signatures including Dilithium and Falcon, authentication protocols such as RLWE, and access control systems such as RLWE-CP-ABE.

Hash-based mechanisms are represented by SPHINCS+, XMSS-MT, and LMS within communication stacks, scalable signatures, acceleration, and blockchain applications.

Code-based mechanisms are exemplified by Classic McEliece and BIKE in hybrid TLS, along with HQC in comparative evaluations. Additionally, MPC-in-the-Head signatures manifest through AIMer across various platforms. Multivariate and isogeny-based families remain marginal, primarily serving as comparative references, while SIKE is approached with caution due to its documented cryptanalytic vulnerabilities. QKD is included as an adjacent technology in one multimedia encryption proposal; however, it is not treated as PQC in the strictest sense. Table 5 consolidates this distribution.

Table 5

Cryptographic Families identified in the Corpus

Family	Base problem	Algorithms in the corpus	Relative maturity in the corpus
Lattice-based	LWE, RLWE, NTRU, MLWE, SIS	Kyber, Dilithium, Falcon, Falcon-M, NTRU, Rudraksh, Ring-ExpLWE, MP-SPHF, RLWE-CP-ABE, HAETAE	Dominant; covers KEMs, signatures, authentication, and access control
Hash-based	Resistance of hash functions	SPHINCS+, XMSS-MT, LMS	Consolidated; present in scalable signatures and blockchain
Code-based	Decoding of linear codes	Classic McEliece, BIKE, HQC	Limited presence; present in hybrid TLS and comparative evaluations
MPC-in-the-Head	Multiparty computation over symmetric primitives	AIMer	Emerging; one evaluated proposal in the corpus
Multivariate	Nonlinear systems over finite fields	Rainbow, broken via cryptanalytic attacks, and other multivariate schemes	Marginal; comparative reference only
Isogeny-based	Isogenies between supersingular curves	SIKE, with warning due to cryptanalytic break	Limited presence; appears with warning due to cryptanalytic break
Complementary quantum cryptography	Quantum principles	QKD	Adjacent; not PQC in the strict sense

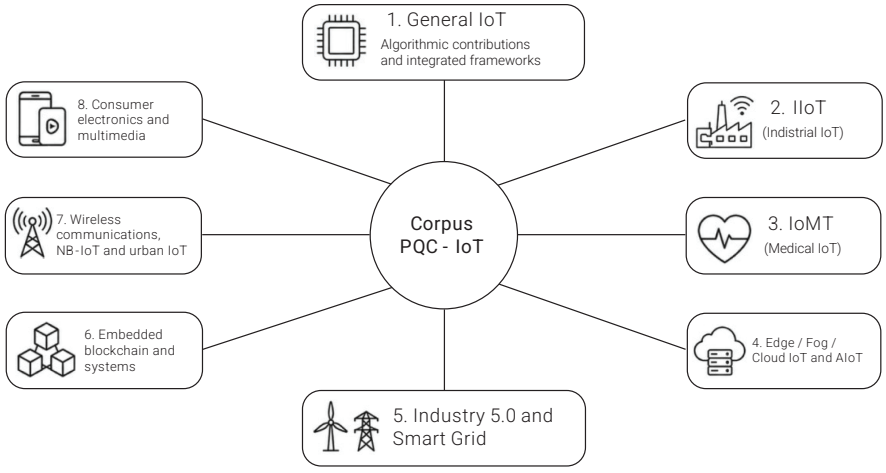
Note: This table provides a comprehensive summary of the families and associated categories identified within the corpus. It is important to note that hybrid combinations are excluded from classification as an independent family, as the integration of PQC + classical hybridization is considered a transitional strategy.

7.2 IoT Application Domains in the Corpus

The corpus encompasses various IoT domains, though their representation is not uniform. Figure 7 illustrates these areas for orientation purposes.

Figure 7

Identification of IoT Application Domains within the Corpus



Note: The figure provides a summary of the application domains identified within the corpus. It is important to note that the positioning of these domains does not reflect precise quantitative prevalence, as multiple studies may contribute to more than one domain.

The domain of the IoT encompasses both algorithmic advancements and comprehensive integrated frameworks. Notable contributions include PAKE and lattice-based authentication (Li & Wang, 2022), KEMs along with algorithmic variants designed for constrained hardware (Kundu et al., 2025; Xu et al., 2022), and signatures optimized for deployment on constrained platforms (Castiglione et al., 2025; Kerimbayeva et al., 2025; Kwon et al., 2025). Furthermore, suitability evaluations conducted on embedded platforms are reported (Halak et al., 2024; Shim, 2024), alongside integrated frameworks leveraging blockchain technomoby, ZKP, the Zero Trust model, and artificial intelligence (Aleisa, 2025; Elkhodr, 2025; Tawfik et al., 2025).

The Industrial Internet of Things (IIoT) serves as another critical area of focus, which including WSNs tailored for industrial settings (Señor et al., 2022), implementations of three-party mutual authentication (Kim et al., 2026), evaluations of KEMTLS (Cruz-Piris et al., 2025), private inference techniques (Kjamilji, 2024), GPU acceleration for enhanced aggregate processing (Wu et al., 2026), and the migration of legacy industrial protocols (Al-Mekhlafi et al., 2026).

IoMT manifests more specifically in the context of RLWE-based multifactor authentication (Ahmad et al., 2026) and healthcare components integrated into broader access-control or private-inference frameworks (Kjamilji, 2024; Poomekum et al., 2025).

Furthermore, the domains of edge, fog, cloud IoT, and AIoT encompass various technologies, including NTRU on ARM within edge and fog environments (Al-Doori & Al-Gailani, 2023), fog-assisted access control (Poomekum et al., 2025), GPU acceleration for signature processing in cloud and edge contexts (J. Wu et al., 2025; W. Wu et al., 2026), and the application of federated learning for threat detection (Rahmati & Pagano, 2025).

More specific applications arise in Industry 5.0 and smart grid systems, supported by a 6G-enabled edge intelligence framework (Alatawi, 2025). Additionally, developments in embedded blockchain technology focus on scalable signatures and public-key recovery (Iavich et al., 2025; Marchsreiter, 2025). In the realm of wireless communications and NB-IoT, IEEE 802.11n and LPWAN proposals offer innovative solutions (Do & Tran, 2026; Ojetunde et al., 2025). Finally, advancements in consumer electronics and multimedia are represented by an image-encryption proposal that integrates NTRU, QKD, and Falcon (Aneesh Kumar et al., 2025).

Two patterns emerge from this analysis. First, several studies contribute to multiple domains, indicating that the primary assignments listed in Supplementary Material S1 should not be considered exclusive. Second, the density of domainsrepresented in this corpus reflects specific contributions rather than the overall prevalence in the field. Areas such as smart homes, critical infrastructure, and the metaverse are underrepresented in this corpus, though they are not necessarily absent from the broader research landscape of PQC in the IoT.

7.3 Mapping the Corpus onto the Layered Analytical Architecture

The layered analytical architecture presented in Section 3 situates each study according to its primary contribution. Table 6 summarizes this mapping across the device, communication, distributed trust, and application layers.

Table 6

Mapping the Corpus onto the Layered Analytical Architecture

Layer or dimension	Key functions	Studies
Device layer	Lightweight KEMs and signatures on constrained hardware	Al-Doori & Al-Gailani (2023); Castiglione et al. (2025); Halak et al. (2024); Kerimbayeva et al. (2025); Kundu et al. (2025); Kwon et al. (2025); Señor et al. (2022); Shim (2024); Xu et al. (2022)
Communication layer	Key establishment, channel encryption, PQC stacks, and handshakes	Al-Mekhlafi et al. (2026); Cruz-Piris et al. (2025); Do & Tran (2026); Mahdi & Abdullah (2025); Ojetunde et al. (2025); Zafar & Iqbal (2025)

(*continues*)

(*continued*)

Layer or dimension	Key functions	Studies
Distributed trust layer	Authentication, identity, integrity, blockchain, ZKP, scalable signatures, and operational privacy	Ahmad et al. (2026); Alatawi (2025); Aleisa (2025); Elkhodr (2025); Iavich et al. (2025); Kim et al. (2026); Kjamilji (2024); Li & Wang (2022); Marchsreiter (2025); Poomekum et al. (2025); Rahmati & Pagano (2025); Tawfik et al. (2025); Wu et al. (2025); Wu et al. (2026)
Application layer	Final services built on the guarantees provided by the previous layers	Aneesh Kumar et al. (2025)
Privacy (cross-cutting)	Guarantees beyond channel confidentiality	Aleisa (2025); Elkhodr (2025); Kjamilji (2024); Poomekum et al. (2025); Rahmati & Pagano (2025); Tawfik et al. (2025)
Resistance to side-channel attacks (cross-cutting)	Robustness against physical leakage	Identified as a gap by Halak et al. (2024); Kerimbayeva et al. (2025); Kundu et al. (2025); Marchsreiter (2025); Señor et al. (2022); and Shim (2024).

This mapping supports two key observations. First, the device and communication layers present the most detailed algorithmic and experimental evidence, aligning with their identification as significant bottlenecks in the migration to PQC. Second, the distributed trust layer encompasses the most comprehensive array of integrated mechanisms, which underscores the systemic importance of identity, integrity, traceability, and privacy within the IoT.

Several studies exhibit intersections across these layers. For instance, Castiglione et al. (2025) establish a connection between the device and distributed trust layers by implementing Dilithium-5 on ESP32 within a blockchain use case. Al-Mekhlafi et al. (2026) link the communication and distributed trust layers by combining Kyber, Dilithium, and AES-GCM in SECS/GEM. Additionally, Alatawi (2025) bridges the distributed trust and application domains through the use of EdgeGuard-IoT for Industry 5.0. Table 6 categorizes each study according to its primary contribution, while discussions of overlaps occur in the relevant sections.

The cross-cutting dimensions exhibit distinct behaviors. Privacy emerges as a significant operational contribution across six studies, primarily through techniques such as federated learning, homomorphic encryption, differential privacy, ZKP, blockchain, and access control. In contrast, resistance to side-channel attacks is predominantly presented as an open limitation rather than as an evaluated property. This asymmetry represents one of the most notable gaps within the existing literature.

7.4 Common Gaps and Research Agenda

The cross-sectional analysis identifies three recurring gaps. The first and most prominent gap is the insufficient experimental evaluation of side-channel resistance on actual hardware. While several studies demonstrate the feasibility of PQC on constrained platforms, the resistance to physical leakage is largely presented as a limitation or future research direction across KEMs, digital signatures, and post-quantum blockchain implementations (Halak et al., 2024; Kerimbayeva et al., 2025; Kundu et al., 2025; Marchsreiter, 2025; Señor et al., 2022; Shim, 2024).

The second gap concerns sustained migration within real IoT ecosystems. Most evaluations predominantly depend on simulations, controlled testbeds, laboratory platforms, or emulated scenarios. As a result, evidence regarding longitudinal deployment costs, system stability, maintenance needs, hardware upgrades, interoperability with heterogeneous networks, and overall total cost of ownership remains limited. Furthermore, the technical metrics reported, such as latency, energy consumption, memory capacity, throughput, key size, and signature size, are seldom linked to sector-specific deployment decisions.

The third gap pertains to the transferability of integrated frameworks that combine PQC with federated learning, homomorphic encryption, differential privacy, zero-knowledge proofs, blockchain, or artificial intelligence. Current evaluations are often constrained to specific datasets, simulations, or platforms, raising unresolved questions regarding the behavior of these frameworks in large-scale production environments and their generalizability across different domains (Aleisa, 2025; Kjamilji, 2024; Rahmati & Pagano, 2025; Tawfik et al., 2025). Furthermore, this limitation is exacerbated by inconsistent reporting of standardization status by NIST, as well as by variations in security levels and comparisons with AES-equivalent security.

The resulting agenda is threefold: first, evaluating PQC against side-channel attacks on real hardware; second, studying longitudinal migrations while considering operational costs and critical sectors; and third, validating composite trust and privacy frameworks in production environments. Future reviews should aim to expand the corpus to include underrepresented families, platforms, and network technologies.

8. CONCLUSIONS

This narrative review demonstrates that the migration to PQC in the IoT entails more than merely replacing vulnerable asymmetric primitives. It significantly impacts constrained hardware, communication protocols, identity and authentication processes, privacy-preserving measures, and application services. Consequently, these transitions should be understood as a systemic transformation of IoT security.

The reviewed research articles predominantly feature lattice-based mechanisms, while hash-based, code-based, and MPC-in-the-Head families serve complementary roles. The exploration of hybrid approaches and compositions involving blockchain, zero-knowledge proofs, federated learning, homomorphic encryption, AI, and Zero Trust demonstrates that achieving post-quantum security in IoT relies on tailored combinations suited to specific functional layers and application domains.

The most significant gaps persist in the limited validation of side-channel resistance on actual hardware, the scarcity of longitudinal migration studies, the weak correlation between technical metrics and operational deployment costs, and the inadequate validation of integrated trust and privacy frameworks within large-scale heterogeneous environments.

This review contributes to the organization of a heterogeneous field by implementing a layered analytical architecture, identifying recurrent patterns within the included studies, and delineating existing research gaps. It is important to interpret these conclusions within the context of a narrative, non-exhaustive corpus, rather than as a comprehensive coverage of the PQC-IoT field.

Data Availability Statement

The comprehensive characterization matrix of the 30-study corpus can be accessed as Supplementary Material S1 at the following link: https://docs.google.com/spreadsheets/d/18fGUvDLxUFv6TFWoEqILFE75plqhGSZjstl_Fmk_304

REFERENCES

Ahmad, A., Jagatheswari, S., & Praveen, R. (2026). Quantum-secure lightweight fuzzy extractor based user authentication scheme for Internet of Medical Things. *Soft Computing, 30*, 787-808. https://doi.org/10.1007/s00500-025-11021-z

Alatawi, M. N. (2025). EdgeGuard-IoT: 6G-enabled edge intelligence for secure federated learning and adaptive anomaly detection in Industry 5.0. *Computers, Materials & Continua, 85*(1), 695-727. https://doi.org/10.32604/cmc.2025.066606

Al-Doori, M. J., & Al-Gailani, M. F. (2023). Securing IoT networks with NTRU cryptosystem: A practical approach on ARM-based devices for edge and fog layer integration. *International Journal of Intelligent Engineering & Systems, 16*(5), 462-472. https://doi.org/10.22266/ijies2023.1031.40

Aleisa, M. A. (2025). Blockchain-enabled zero trust architecture for privacy-preserving cybersecurity in IoT environments. *IEEE Access, 13*, 18660-18676. https://doi.org/10.1109/ACCESS.2025.3529309

Al-Mekhlafi, Z. G., Altmemi, J. M. H., Al-Shareeda, M. A., Al-Al-Hchaimi, A. A. J., Gaber, T., Homod, R. Z., Mohammed, B. A., Alshammari, G., Al-Dhlan, K. A., Alrashdi, R., & Alkhabra, Y. A. (2026). A post-quantum secure solution for SECS/GEM communications in Industrial Internet of Things (IIoT) application. *IEEE Open Journal of the Communications Society, 7*, 670-684. https://doi.org/10.1109/OJCOMS.2026.3654010

Aneesh Kumar, K. B., Mohith, L. S., Jain, K., Krishnan, P., Venkatachalam, N., & Buyya, R. (2025). Post-quantum cryptography-based multimedia encryption communication scheme in IoT consumer electronics. *IEEE Transactions on Consumer Electronics*, *71*(2), 4995-5006. https://doi.org/10.1109/TCE.2025.3572949

Castiglione, A., Esposito, J. G., Loia, V., Nappi, M., Pero, C., & Polsinelli, M. (2025). Integrating post-quantum cryptography and blockchain to secure low-cost IoT devices. *IEEE Transactions on Industrial Informatics, 21*(2), 1674-1683. https://doi.org/10.1109/TII.2024.3485796

Cruz-Piris, L., Marín-López, A., Álvarez-Campana, M., Sanz, M., Moreno, J. I., & Arroyo, D. (2025). Measuring the impact of post quantum cryptography in Industrial IoT scenarios. *Internet of Things, 34*, Article 101793. https://doi.org/10.1016/j.iot.2025.101793

Do, T.-B., & Tran, Q.-H. (2026). A hybrid lightweight and post-quantum communication framework for NB-IoT networks. *Microsystem Technologies, 32*, Article 37. https://doi.org/10.1007/s00542-026-06031-2

Elkhodr, M. (2025). An AI-driven framework for integrated security and privacy in Internet of Things using quantum-resistant blockchain. *Future Internet, 17*(6), Article 246. https://doi.org/10.3390/fi17060246

Halak, B., Gibson, T., Henley, M., Botea, C.-B., Heath, B., & Khan, S. (2024). Evaluation of performance, energy, and computation costs of quantum-attack resilient encryption algorithms for embedded devices. *IEEE Access, 12*, 8791-8805. https://doi.org/10.1109/ACCESS.2024.3350775

Iavich, M., Kuchukhidze, T., & Bocu, R. (2025). Fractional Verkle trees for scalable post-quantum signatures. *IEEE Access, 13*, 209921-209937. https://doi.org/10.1109/ACCESS.2025.3640288

Kerimbayeva, A., Iavich, M., Begimbayeva, Y., Gnatyuk, S., Tynymbayev, S., Temirbekova, Z., & Ussatova, O. (2025). A lightweight variant of Falcon for efficient post-quantum digital signature. *Information, 16*(7), Article 564. https://doi.org/10.3390/info16070564

Kim, C., Kwon, D., Park, Y., & Park, Y. (2026). Quantum-resistant three-party mutual authentication protocol for industrial IoT environments. *IEEE Internet of Things Journal*. Advance online publication. https://doi.org/10.1109/JIOT.2026.3677406

Kjamilji, A. (2024). Privacy-preserving zero-sum-path evaluation of decision trees in postquantum industrial IoT. *IEEE Transactions on Industrial Informatics, 20*(8), 10178-10191. https://doi.org/10.1109/TII.2024.3384523

Kundu, S., Ghosh, A., Karmakar, A., Sen, S., & Verbauwhede, I. (2025). Rudraksh: A compact and lightweight post-quantum key-encapsulation mechanism. *IACR Transactions on Cryptographic Hardware and Embedded Systems, 2025*(2), 647-680. https://doi.org/10.46586/tches.v2025.i2.647-680

Kwon, J., Lee, S., Lee, B., Seo, H., & Cho, J. (2025). Efficient implementations of AIMer post-quantum signature scheme for low-end to high-end IoT devices. *IEEE Internet of Things Journal, 12*(22), 46817-46837. https://doi.org/10.1109/JIOT.2025.3597415

Li, Z., & Wang, D. (2022). Achieving one-round password-based authenticated key exchange over lattices. *IEEE Transactions on Services Computing, 15*(1), 308-321. https://doi.org/10.1109/TSC.2019.2939836

Mahdi, L. H., & Abdullah, A. A. (2025). A hybrid post-quantum cryptographic framework integrating Kyber-512 and ASCON for secure IoT communications. *Engineering, Technology & Applied Science Research, 15*(5), 26527-26533. https://doi.org/10.48084/etasr.12471

Marchsreiter, D. (2025). Towards quantum-safe blockchain: Exploration of PQC and public-key recovery on embedded systems. *IET Blockchain, 5*(1), Article e12094. https://doi.org/10.1049/blc2.12094

Ojetunde, B., Kurihara, T., Yano, K., Sakano, T., & Yokoyama, H. (2025). A practical implementation of post-quantum cryptography for secure wireless communication. *Network, 5*(2), Article 20. https://doi.org/10.3390/network5020020

Poomekum, P., Suriyawong, A., & Fugkeaw, S. (2025). Fine-grained and lightweight quantum-resistant access control system with efficient revocation for IoT cloud. *IEEE Open Journal of the Communications Society, 6*, 8652-8666. https://doi.org/10.1109/OJCOMS.2025.3620094

Rahmati, M., & Pagano, A. (2025). Federated learning-driven cybersecurity framework for IoT networks with privacy preserving and real-time threat detection capabilities. *Informatics, 12*(3), Article 62. https://doi.org/10.3390/informatics12030062

Señor, J., Portilla, J., & Mujica, G. (2022). Analysis of the NTRU post-quantum cryptographic scheme in constrained IoT edge devices. *IEEE Internet of Things Journal, 9*(19), 18778-18790. https://doi.org/10.1109/JIOT.2022.3162254

Shim, K.-A. (2024). On the suitability of post-quantum signature schemes for Internet of Things. *IEEE Internet of Things Journal, 11*(6), 10648-10665. https://doi.org/10.1109/JIOT.2023.3327400

Tawfik, M., Abdelhaliem, A. H., & Fathi, I. (2025). Quantum-resistant privacy-preserving IoT authentication via zero-knowledge proofs and blockchain integration. *Statistics, Optimization & Information Computing, 14*(3), 1374-1402. https://doi.org/10.19139/soic-2310-5070-2399

Wu, J., Yu, Y., Chen, Z., Yang, H., Li, C., & Liu, Z. (2025). CBPSPX: A CUDA-based batch parallel optimization of post-quantum signature SPHINCS+. *IEEE Internet of Things Journal, 12*(18), 37898-37911. https://doi.org/10.1109/JIOT.2025.3585558

Wu, W., Dong, J., Hou, Y., Liu, M., Li, L., & Dong, Z. (2026). RIGHT: GPU-optimized parallel PQC HAETAE for high-throughput cryptographic acceleration. *IEEE Transactions on Industrial Informatics, 22*(1), 532-542. https://doi.org/10.1109/TII.2025.3613305

Xu, D., Wang, X., Hao, Y., Zhang, Z., Hao, Q., Jia, H., Dong, H., & Zhang, L. (2022). Ring-ExpLWE: A high-performance and lightweight post-quantum encryption scheme for resource-constrained IoT devices. *IEEE Internet of Things Journal, 9*(23), 24122-24134. https://doi.org/10.1109/JIOT.2022.3189210

Zafar, A., & Iqbal, S. S. (2025). Integrating code-based post-quantum cryptography into SSL TLS protocols through an interoperable hybrid framework. *Discover Computing, 28,* Article 202. https://doi.org/10.1007/s10791-025-09735-7

DATOS DE LOS AUTORES

JUAN DE JESÚS AMADOR REYES

Es doctor en Ciencias de la Computación. Actualmente, se desempeña como profesor de tiempo completo del Centro Universitario Ecatepec de la Universidad Autónoma del Estado de México, donde imparte clases en las carreras profesionales de Informática Administrativa y de Ingeniería en Computación, así como en la maestría en Ciencias de la Computación. Cuenta con perfil deseable PRODEP y ha publicado artículos en congresos y revistas científicas referentes a temas de realidad virtual, inteligencia artificial, computación gráfica y minería de procesos. Es miembro del Sistema Nacional de Investigadoras e Investigadores de México.

OLDA BUSTILLOS ORTEGA

Es magíster en Informática con especialidad en Auditoría Informática por la Universidad Latina de Costa Rica. Actualmente, se desempeña como directora de la Escuela de Ingeniería Informática y docente investigadora en la Universidad Internacional de las Américas. Ha participado en la XI Conferencia "Servicios creativos y modernos para el comercio y desarrollo sostenible" de la Red Latinoamericana y Caribeña para la Investigación en Servicios (REDLAS), en The 2022 International Conference on Information Technology & Systems (ICITS'22) y fue distinguida como *honorary academician of social sciences* por la International Academy of Social Sciences de Palm Beach, Florida, Estados Unidos. Sus intereses de investigación se centran en la inteligencia artificial, la ciberseguridad y el desarrollo de sistemas inteligentes.

STALIN FRANCISCO CALVA VICENTE

Es estudiante de pregrado de Tecnologías de la Información de la Facultad de Ingeniería Civil de la Universidad Técnica de Machala, Ecuador. Su vinculación académica con

una facultad orientada a la ingeniería y la tecnología refleja su interés por las ciencias aplicadas y por la búsqueda de soluciones para el desarrollo tecnológico y estructural de la región.

JOSE LUIS CASTILLO MENDOZA

Es doctor en Ciencias de la Educación por el Colegio de Posgraduados de la Ciudad de México, maestro en Administración por la Universidad del Valle de México y licenciado en Contaduría por la Escuela Bancaria y Comercial. Se desempeña como profesor de las carreras profesionales de Informática Administrativa y de Diseño Industrial. Es integrante del Sistema Nacional de Investigadoras e Investigadores (SNII) de la SECIHTI. Sus intereses actuales de investigación abarcan la formación de capital humano y la tecnología educativa, centrada en el desarrollo de recursos humanos y tecnológicos para optimizar los procesos de enseñanza-aprendizaje.

FABRICIO MARTIN CHAVEZ RODRIGUEZ

A Systems Engineering student at the University of Lima, where they also work in the University's SAP Laboratory. Their academic and professional interests focus on information technologies, with particular emphasis on artificial intelligence, automation, the Internet of Things (IoT), information analytics, systems auditing, and digital transformation.

MARÍA AZENETH DUANA GUERRA

Es estudiante de Informática Administrativa en el Centro Universitario Ecatepec de la Universidad Autónoma del Estado de México. Cuenta con experiencia en la producción académica orientada a la revisión sistemática de literatura y, actualmente, desarrolla trabajos de investigación sobre la aplicación de instrumentos de recolección de datos en entornos digitales. Se encuentra en constante formación técnica mediante certificaciones en Power BI y estudios del idioma inglés. Sus principales intereses se encuentran en las áreas de análisis de datos y gestión de proyectos.

GUADALUPE MONTSERRAT FIGUEROA CASTAÑEDA

Es estudiante de Informática Administrativa en el Centro Universitario Ecatepec de la Universidad Autónoma del Estado de México. A lo largo de su formación académica, ha desarrollado diferentes habilidades en áreas como la gestión de tecnologías de la información, análisis de sistemas, seguridad informática y desarrollo de proyectos tecnológicos.

Ha participado en diversos proyectos académicos. Su interés profesional se centra en la administración de tecnologías de la información, la consultoría tecnológica y la seguridad de la información.

MARISOL HERNÁNDEZ HERNÁNDEZ

Es doctora en Ciencias en Sistemas Computacionales y Electrónicos por la Universidad Autónoma de Tlaxcala, maestra en Tecnología Educativa por el Instituto Tecnológico de Estudios Superiores de Monterrey y licenciada en Informática por el Instituto Tecnológico de Apizaco. Es técnica académica de tiempo completo en la UAEMéx e integrante del Sistema Nacional de Investigadoras e Investigadores (SNII) de la SECIHTI. Sus intereses de investigación se centran en la realidad aumentada, la inteligencia artificial, la ingeniería de *software* y los modelos basados en lógica difusa.

CUAUHTÉMOC HIDALGO CORTÉS

Es doctor en Sistemas Computacionales. Desde el 2008, se desempeña como profesor de tiempo completo del Centro Universitario Ecatepec de la Universidad Autónoma del Estado de México, en donde imparte clases en la carrera profesional de Ingeniería en Computación y en la Maestría en Ciencias de la Computación. Es integrante del cuerpo académico TIC y Dispositivos Electrónicos con registro y nivel consolidado ante la Secretaría de Educación Pública. Simultáneamente, ha dirigido tesis de nivel licenciatura y maestría. Es miembro del Sistema Nacional de Investigadoras e Investigadores de México.

MARÍA JAIMES-QUINTANILLA

Es magíster e ingeniera en Calidad y Gestión Integral. Posee más de diez años de experiencia en docencia, investigación y dirección académica y empresarial. Es investigadora *junior* reconocida por el Ministerio de Ciencias, Tecnología e Innovación de Colombia. Ha participado en proyectos de investigación relacionados con inteligencia artificial, analítica de datos y gestión de proyectos en el sector construcción. Ha desempeñado cargos como directora de maestría, docente de pregrado y posgrado, y líder en procesos de aseguramiento de la calidad académica. Su trayectoria incluye experiencia en liderazgo organizacional, formación en educación virtual y desarrollo de capacidades en transformación digital. Asimismo, registra producción científica en revistas indexadas y participación en proyectos financiados, lo que consolida un perfil orientado a la integración de tecnologías emergentes y la optimización de procesos organizacionales.

RODRIGO JARA ESPINOZA

A Systems Engineering student at the University of Lima, with experience in the University of Lima SAP Laboratory. Their academic and professional profile is focused on software engineering and DevOps, with a particular interest in web development, process automation, and the continuous improvement of technological solutions.

EDUARDO RAFAEL JAUREGUI ROMERO

Es ingeniero de *software* por la Universidad Nacional Mayor de San Marcos. Actualmente, se desempeña como *data scientist advanced*. Obtuvo el segundo lugar en el *hackathon* del Summit of AI in Latin America (SALA) con el proyecto AI for a Fossil-Free Galápagos. Sus principales intereses de investigación se centran en inteligencia artificial, ciencia de datos, MLOps e inteligencia artificial generativa.

FREDDY ANÍBAL JUMBO CASTILLO

Posee un máster universitario en Inteligencia Artificial, un máster en Ciencia de Datos Aplicada, un *master of science* en Geographical Information Science & Systems (SIG) y una maestría en Educación Superior. Es ingeniero de sistemas. En el ámbito laboral, se desempeña actualmente como docente e investigador en la carrera profesional de Tecnologías de la Información en la Universidad Técnica de Machala, Ecuador. Cuenta con una amplia experiencia en el desarrollo de proyectos de investigación y vinculación con la sociedad para fortalecer el ecosistema tecnológico regional. A nivel de reconocimientos y aportes, destaca por sus múltiples publicaciones de libros y artículos científicos en revistas indexadas, en los que aborda temáticas innovadoras como el desarrollo de *chatbots* para orientación vocacional, el impacto de los sesgos y la equidad algorítmica en el aprendizaje automático, y la aplicación de los SIG para la delimitación de cuencas hidrográficas. Sus intereses principales se centran en la intersección de la educación y la tecnología, promoviendo el uso ético de la inteligencia artificial, la analítica de datos para el bienestar social y la formación de profesionales íntegros y comprometidos con su entorno.

DANIEL MENA BOCKER

Es magíster por la Universidad Magister. Actualmente, se desempeña como docente investigador. Sus intereses de investigación se centran en la ciberseguridad y la inteligencia artificial.

ALEJANDRA MORALES RAMÍREZ

Es doctora en Sistemas Computacionales. Desde el 2009, se desempeña como profesora de tiempo completo en el Centro Universitario Ecatepec de la Universidad Autónoma del Estado de México, donde imparte las asignaturas de Comunicación entre Computadoras, Protocolos de Red, Bases de Datos, Investigación I y Temas Selectos de Computación. Cuenta con perfil deseable PRODEP y ha publicado artículos en congresos y revistas científicas referentes a temas de educación, minería de procesos, desarrollo de *software* y sistemas de información, así como de circuitos analógicos. Conjuntamente, ha dirigido tesis de nivel licenciatura y maestría. Es miembro del Sistema Nacional de Investigadoras e Investigadores de México.

JORGE MURILLO GAMBOA

Es magíster en Computación con especialidad en Telemática por el Instituto Tecnológico de Costa Rica. Actualmente, se desempeña como docente investigador. Ha recibido los reconocimientos Valued Member and Supporter (2016) y Certificate of Appreciation (2014), otorgados por la IEEE Computer Society. Sus intereses de investigación se centran en la inteligencia artificial, la ciberseguridad y el desarrollo de habilidades y competencias digitales según el modelo SFIA.

JOHNNY PAÚL NOVILLO VICUÑA

Posee una maestría en Educación Superior. Es ingeniero eléctrico y candidato a doctor (Ph. D.) en Tecnologías de la Información y Comunicación por la Universidad de La Coruña, España. En el ámbito laboral, cuenta con una sólida experiencia en el sector privado dedicada a la consultoría y construcción de obras eléctricas, mientras que en el sector educativo se desempeña como docente titular e investigador en la Universidad Técnica de Machala, institución en la que también ha ejercido como coordinador de las carreras profesionales de Ingeniería de Sistemas y de Tecnologías de la Información. Su destacada producción académica incluye la publicación de libros de texto fundamentales para la formación universitaria, como la serie de volúmenes de *Fundamentos de los sistemas microprocesados* y *Arduino y el internet de las cosas,* además de diversos artículos científicos en revistas indexadas. Sus principales líneas de investigación y áreas de interés abarcan el desarrollo de sistemas con internet de las cosas aplicados a la automatización agrícola y acuícola, la criptografía homomórfica para la seguridad de datos en la nube de las pymes, y la aplicación de redes neuronales y visión artificial para la interpretación en tiempo real del lenguaje de señas ecuatoriano.

OLMAN NÚÑEZ PERALTA

Es magíster en Base de Datos por la Universidad Cenfotec. Actualmente, se desempeña como docente investigador. Sus intereses de investigación se centran en bases de datos, *machine learning* e inteligencia artificial.

YOHAMIN NAFIT PIMENTEL ALARCON

A Systems Engineering student at the University of Lima, Peru. They currently work at the University of Lima SAP Laboratory. Their academic and professional profile focuses on cybersecurity, with research interests in cybersecurity and the Internet of Things (IoT).

FABIÁN RODRÍGUEZ SIBAJA

Es magíster en Administración de Proyectos por la Universidad para la Cooperación Internacional. Actualmente, se desempeña como docente investigador. Sus intereses de investigación se centran en las metodologías de desarrollo de *software*, la inteligencia artificial y los sistemas de información de apoyo a los negocios.

EDGAR SERRANO PÉREZ

Es doctor y maestro en Tecnología Avanzada y egresado de Ingeniería en Control y Automatización por el Instituto Politécnico Nacional. Actualmente, realiza una estancia posdoctoral académica en la UAEMéx. Es integrante del Sistema Nacional de Investigadoras e Investigadores (SNII) de la SECIHTI. Sus intereses actuales de investigación incluyen las tecnologías de inteligencia artificial, el control de procesos mediante microcontroladores y la realidad virtual y aumentada destinada al desarrollo de objetos de aprendizaje y recursos educativos interactivos.

ANABELEM SOBERANES MARTIN

Es doctora en Ciencias de la Educación por el Colegio de Estudios de Posgrado de la Ciudad de México, maestra en Educación por la Universidad de las Américas y licenciada en Sistemas de Computación Administrativa por la Universidad del Valle de México. Actualmente, se desempeña como profesora de tiempo completo. Cuenta con el reconocimiento del PRODEP y es integrante del Sistema Nacional de Investigadoras e Investigadores (SNII) de la SECIHTI. Sus intereses de investigación se centran en el desarrollo de soluciones computacionales con impacto social, que incorporen el uso de tecnologías emergentes, enfocándose principalmente en tecnología educativa.

ANGELO TACO-JIMENEZ

Holds a bachelor's degree in Systems Engineering from the University of Lima, graduating summa cum laude. Specializes in software development for mobile applications and currently serves as an Android Engineer at ITLAB and Android Project Manager at White Pencil Consulting. In addition, serves as a teaching assistant in the Systems Engineering program at the University of Lima and is pursuing a master's degree in Innovation Management at the University of Lima's Graduate School. Research interests include self-adaptive software and human–computer interaction.

ANGEL RICARDO VERA SANTOS

Es un estudiante de pregrado de Tecnologías de la Información en Civil de la Universidad Técnica de Machala, Ecuador. Su pertenencia a una facultad enfocada en la ingeniería y la tecnología refleja un claro interés por las ciencias aplicadas y por la búsqueda de soluciones en el ámbito del ecosistema laboral y tecnológico del país.

SERGIO ZABALA-VARGAS

Es magíster en Administración de Proyectos, magíster en E-learning y doctor en Tecnología Educativa, con mención *cum laude*. Es ingeniero electrónico. Cuenta con más de veinte años de experiencia en docencia universitaria, investigación y gestión de proyectos multidisciplinares. Es investigador sénior reconocido por el Ministerio de Ciencias, Tecnología e Innovación de Colombia, y ha liderado grupos de investigación clasificados en categoría A, así como múltiples proyectos con financiación nacional e internacional. Su producción científica incluye más de 25 artículos en revistas indexadas, seis libros de investigación y diversos desarrollos tecnológicos registrados. Actualmente, se desempeña como docente investigador en programas de posgrado y doctorado en instituciones de Colombia y el exterior, con énfasis en la integración de tecnologías emergentes, analítica de datos y transformación digital en contextos educativos y organizacionales.


FONDO
EDITORIAL
ULIMA