International Journal of Computational Innovations, Intelligent Systems and AI
Julio 2026, Volumen: **2**, N.º **2** | Páginas 7–25
doi: 10.64439/cisai.v2i2.43
e-ISSN : 3084-7907

ARTÍCULO

Arquitectura moderna de ciberseguridad para prevención de fraudes en servicios administrativos

Modern cybersecurity architecture for fraud prevention in administrative services

Enrique Castellares Cuya* · **Jose Rengifo Espina**

Facultad de Ingeniería, Universidad Tecnológica del Perú, Lima, Perú

*Autor de Correspondencia: 1221978@utp.edu.pe

Enviado: 01 Junio 2026 Aceptado: 14 Julio 2026 Publicado en línea: 15 Julio 2026

Resumen

A medida que las organizaciones de servicios avanzan en su modernización digital, la seguridad no siempre crece al mismo ritmo, generando vulnerabilidades que los esquemas convencionales de protección no logran cubrir. Para abordar este problema, el presente trabajo propone una arquitectura de ciberseguridad orientada a la prevención de fraudes en una empresa del sector servicios de Lima, Perú, cuyo diseño se fundamenta en el análisis documental de 385 registros técnicos de incidentes. Dicho análisis evidenció que solo el 54,3% de las vulnerabilidades fueron atendidas con oscilaciones mensuales entre 44,4% y 62,5%. Mientras que el índice de respuesta efectiva no superó el 34%, con mínimos del 19% en los períodos de mayor carga operativa. Estos resultados, junto con la concentración de incidentes no resueltos en malware, amenazas desconocidas y ransomware, revelan fallas en la priorización y el cierre de eventos de seguridad. Ante este diagnóstico, la arquitectura propuesta articula gestión de identidades y accesos, protección avanzada contra amenazas, seguridad en entornos cloud, cumplimiento normativo y analítica extendida potenciada con inteligencia artificial, conformando una defensa en profundidad capaz de reducir la exposición residual y sostener una respuesta antifraude robusta en entornos administrativos digitalizados.

Palabras clave: Seguridad informática, gestión de riesgos, detección de amenazas, continuidad operacional, inteligencia artificial

Abstract

As service organizations advance in their digital modernization, security does not always keep pace, generating vulnerabilities that conventional protection schemes fail to address. To tackle this problem, the present work proposes a cybersecurity architecture oriented toward fraud prevention in a service sector company in Lima, Peru, whose design is grounded in the documentary analysis of 385 technical incident records. This analysis revealed that only 54.3% of vulnerabilities were addressed annually, with monthly fluctuations between 44.4% and 62.5%, while the effective response rate did not exceed 34%, reaching lows of 19% during periods of peak operational load. These findings, combined with the concentration of unresolved incidents in malware, unknown threats, and ransomware, expose structural failures in the prioritization and closure of security events. In response to this diagnosis, the proposed architecture integrates identity and access management, advanced threat protection, cloud security, regulatory compliance, and extended analytics powered by artificial intelligence, forming a defense-in-depth capable of reducing residual exposure and sustaining a robust anti-fraud response in digitalized administrative environments.

Keywords: Computer security, risk management, threat detection, operational continuity, artificial intelligence

Citar como: E. Castellares Cuya y J. Rengifo Espina, "Arquitectura moderna de ciberseguridad para prevención de fraudes en servicios administrativos," *CISAI*, vol. 2, no. 2, pp. 7–25, Julio 2026, doi: 10.64439/cisai.v2i2.43.

1. Introducción

En los últimos años, la digitalización ha pasado de ser una ventaja competitiva a convertirse en una condición operativa para la mayoría de las organizaciones. La incorporación masiva de servicios en nube, plataformas transaccionales y sistemas informáticos ha traído consigo una expansión considerable de la superficie de ataque, exponiendo a las organizaciones a amenazas cada vez más sofisticadas. En este escenario, el fraude informático ha dejado de ser un riesgo periférico para convertirse en una amenaza que compromete la disponibilidad, la continuidad operativa, la reputación institucional y la confianza de los usuarios, con consecuencias financieras de creciente magnitud. Algunos casos recientes permiten dimensionar el impacto real de estos ataques. Houghton [1] documenta el ataque a la Biblioteca Británica con pérdidas estimadas entre seis y siete millones de libras esterlinas, mientras que Wadesango et al. [2] reportan que los ciberataques contra instituciones de Zimbabue generan pérdidas anuales cercanas a 1.800 millones de dólares. En el contexto peruano, León [3] registra un crecimiento superior al 26% del cibercrimen económico y más de 30.000 casos de fraude electrónico en el primer trimestre de 2024, con un incremento interanual del 25%. Estas cifras revelan que ningún sector ni escala organizacional está exento de exposición, incluyendo empresas de servicios administrativos con infraestructura tecnológica de soporte.

Frente a este panorama, el avance del conocimiento en este campo muestra contribuciones valiosas, aunque parciales. Desde la perspectiva tecnológica, Wahid y Hassini [4] demuestran que la automatización y la inteligencia artificial no solo aportaría en la detección de amenazas, sino que también ayudaría en los patrones del fraude digital al dotarlo de mayor velocidad, adaptabilidad y capacidad de evasión. Sin embargo, este hallazgo cobra especial relevancia cuando se considera que los atacantes aprovechan las mismas tecnologías para sofisticar sus métodos, razón por la cual Abu-Khadrah [5] advierte que los ataques actuales combinan de forma coordinada la suplantación de identidad, el malware, la explotación de credenciales y el abuso de configuraciones, lo que dificulta su contención mediante controles aislados. Ambas contribuciones coinciden en señalar que la complejidad del fraude digital exige respuestas igualmente integradas, capaces de operar de manera simultánea en múltiples capas de la infraestructura organizacional. Desde una mirada ejecutiva y organizacional, la literatura complementa este enfoque con elementos igualmente relevantes. Por ejemplo, Agyepong et al. [6] subrayan que la medición del desempeño en centros de operaciones de seguridad es condición necesaria para sostener una respuesta efectiva a lo largo del tiempo, pues sin indicadores claros las brechas tienden a normalizarse y a volverse invisibles para la gestión. A esto se suma la advertencia de Yaseen [7], quien señala que la adopción tecnológica solo genera valor real cuando va acompañada de confianza por parte de los usuarios, lo que implica que cualquier arquitectura de seguridad debe considerar también la dimensión humana y cultural. En coherencia con ello, Cevallos et al. [8] asocian la resiliencia cibernética con capacidades organizacionales sostenidas, situando así la seguridad como un atributo institucional y no como una función técnica aislada. A pesar de estos avances, persiste un vacío aplicado sobre cómo traducir registros de ataques en una arquitectura antifraude medible, especialmente en organizaciones de servicios que disponen de herramientas dispersas pero carecen de indicadores unificados, dominios de control definidos y secuencias de implementación integradas.

Para atender este vacío, el presente artículo proporciona una arquitectura moderna de ciberseguridad orientada a la prevención de fraudes en servicios administrativos, sustentada en el análisis de registros técnicos de incidentes documentados durante el año 2025 en una empresa de Lima, Perú, que reportó un alto índice de ataques, interrupciones parciales en su operación y pérdidas estimadas en 120 mil soles. Este contexto justifica la necesidad de una respuesta estructurada y ofrece una base empírica para formular una solución verificable. En adelante, el artículo se estructura en seis secciones. La Sección 2, revisión de literatura, desarrolla los fundamentos conceptuales y empíricos que sustentan la propuesta. La Sección 3, caso de estudio, caracteriza el contexto organizacional que motiva la propuesta. La Sección 4, metodología, describe el enfoque del estudio, la fuente documental, las variables, los indicadores y el procedimiento analítico. La Sección 5, resultados, presenta el diagnóstico de vulnerabilidad, el desempeño del sistema de protección y la arquitectura propuesta frente a las brechas identificadas. La Sección 6, discusión, interpreta los hallazgos a la luz de la literatura y precisa sus implicancias para la gestión de la ciberseguridad. La Sección 7, conclusiones, sintetiza la contribución principal, las limitaciones del estudio y la ruta de validación futura.

2. Revisión de literatura

Esta sección examina los fundamentos teóricos que sustentan la propuesta arquitectónica del estudio. Se organiza en tres ejes: la detección de fraude y la adopción tecnológica, las operaciones de seguridad y sus indicadores de desempeño, y la resiliencia cibernética apoyada en inteligencia artificial y gobierno. A partir de esta revisión se identifican los avances consolidados en la disciplina, así como el vacío aplicado que justifica la contribución del presente trabajo.

2.1 Detección de fraude y adopción tecnológica

La comprensión del fraude digital ha experimentado una transformación significativa en los últimos años, pasando de enfoques basados en reglas estáticas hacia sistemas híbridos capaces de combinar aprendizaje supervisado, analítica de comportamiento y criterios de confianza institucional. En este sentido, Wahid y Hassini [4] destacan el valor de marcos híbridos con inteligencia artificial para plataformas transaccionales, demostrando que la combinación de múltiples capas de análisis mejora significativamente la capacidad de detección. En esa misma dirección, Afriyie et al. [9] muestran que los algoritmos supervisados alcanzan mayor precisión cuando se entrenan con patrones consistentes, lo que subraya la importancia de contar con registros históricos de calidad como base para cualquier solución. Sin embargo, la tecnología por sí sola no garantiza resultados si los usuarios no perciben transparencia, justicia y confiabilidad en los sistemas, aspecto que Yaseen [7] identifica como determinante para el éxito de cualquier implementación. En el contexto de servicios administrativos, esta tensión entre capacidad técnica y aceptación organizacional resulta especialmente relevante, porque la arquitectura debe proteger activos sensibles al tiempo que sostiene continuidad operativa, trazabilidad y decisiones auditables. Reconocer esta dualidad es el punto de partida, pero no es suficiente si no se cuenta con mecanismos que permitan medir, priorizar y responder de forma sistemática, aspecto que aborda el siguiente eje.

2.2 Operaciones de seguridad e indicadores de desempeño

Identificar las amenazas es solo una parte del problema, pues la otra, igualmente crítica, es la capacidad de gestionarlas con eficiencia y consistencia a lo largo del tiempo. En esta línea, Agyepong et al. [6] proponen evaluar el desempeño de analistas y centros de operaciones mediante indicadores que superen el simple conteo de alertas, reconociendo que la efectividad real de un equipo de seguridad no se refleja en el volumen de eventos procesados sino en la calidad y oportunidad de su respuesta. Complementando esta perspectiva, Forsberg y Frantti [10] resaltan la importancia de evaluar vulnerabilidades considerando detección, priorización y remediación como etapas interdependientes, de modo que una falla en cualquiera de ellas compromete el resultado final. Por su parte, Umer et al. [11] vinculan la eficiencia de mitigación con mecanismos de disuasión y confianza tecnológica, añadiendo una dimensión institucional a la gestión operativa. Estos aportes revelan que una organización puede disponer de herramientas avanzadas y aun así fallar si no las integra y gobierna con indicadores comparables y procesos definidos. El desafío, por tanto, no consiste en acumular soluciones tecnológicas, sino en construir una arquitectura donde eventos, identidades, activos, controles y respuestas fluyan con baja fricción y alta visibilidad. Este nivel de integración, sin embargo, requiere algo más que procesos bien medidos, pues exige también que la organización sea capaz de anticipar, adaptarse y recuperarse, lo que sitúa la discusión en el terreno de la resiliencia cibernética que se aborda en la próxima subsección.

2.3 Resiliencia cibernética, inteligencia artificial y gobierno

La gestión operativa eficiente abre paso a una dimensión más estratégica del problema, que es la capacidad de las organizaciones para anticipar riesgos, absorber impactos y aprender de cada incidente. En este marco, Nandy et al. [12] discuten el uso de modelos predictivos para anticipar riesgos antes de que se materialicen, lo que representa un cambio paradigmático respecto a los enfoques reactivos tradicionales. En coherencia con ello, Moura et al. [13] analizan el papel de la inteligencia artificial en la detección y respuesta a incidentes, mientras que Raj [14] vincula las arquitecturas modernas con controles adaptativos y monitoreo extendido capaces de ajustarse dinámicamente al panorama de amenazas. Esta orientación converge con enfoques contemporáneos de seguridad que integran protección en la nube, gestión de identidades, monitoreo continuo, seguridad de datos y cumplimiento normativo en un marco cohesionado. Sin

embargo, la automatización sin criterios claros de priorización puede producir fatiga de alertas o decisiones opacas, lo que exige un gobierno explícito de los procesos de seguridad. A pesar de estos avances, la literatura revisada aborda cada dimensión de forma independiente, sin ofrecer una propuesta que articule detección, operación y resiliencia en una arquitectura integrada, medible y replicable para organizaciones de servicios. Es precisamente ese vacío el que el presente estudio busca atender, proponiendo una arquitectura sociotécnica que articule herramientas avanzadas, procesos medibles, roles claros, capacitación, auditoría periódica y mejora continua como respuesta verificable al problema del fraude en entornos administrativos digitalizados.

Por tanto, lo expuesto hasta aquí muestra que la ciberseguridad organizacional ha madurado desde la detección puntual hacia una visión integrada que combina tecnología, operación y gobierno. No obstante, esta madurez teórica contrasta con la realidad de muchas organizaciones de servicios, donde las capacidades siguen siendo fragmentadas y los indicadores inexistentes. Es desde esa brecha que el presente estudio construye su propuesta, trasladando los fundamentos revisados hacia una arquitectura, medible y adaptada al contexto operativo analizado, cuya base se describe a continuación en la sección de metodología.

3. Caso de estudio

La organización analizada es una empresa de Lima que presta servicios de soporte logístico, financiero, contable, tecnológico y administrativo a un grupo de empresas del sector automotriz peruano. Todo su funcionamiento depende de plataformas digitales, sistemas de gestión empresarial y servicios en la nube, lo que significa que cualquier fallo en su infraestructura tecnológica no se limita a una sola área, sino que se propaga de forma inmediata a todas las empresas del grupo, paralizando procesos, bloqueando operaciones y generando pérdidas que se acumulan con cada hora de inactividad. Esta dependencia convierte a la organización en un blanco especialmente atractivo para los atacantes, quienes pueden explotar cualquier brecha con un efecto en cadena de gran alcance.

A pesar de esa exposición, el nivel de protección no está a la altura del riesgo que enfrenta. Los registros disponibles documentaron más de 600 intentos de ataque, 20 interrupciones parciales en los servicios y pérdidas económicas estimadas en 120 mil soles, lo que puso de manifiesto que la situación exigía una respuesta más profunda y estructurada. Con ese panorama como punto de partida, el presente estudio se propuso ir más allá del registro de lo ocurrido para comprender las causas subyacentes de esas brechas y traducirlas en una arquitectura de ciberseguridad que no solo corrija lo que falló, sino que ofrezca una base sólida para reducir de manera progresiva la exposición al riesgo.

4. Metodología

Esta sección explica cómo se llevó a cabo el estudio, qué información se utilizó y cómo se analizó para llegar a la propuesta arquitectónica. Se organiza en cuatro apartados que describen el enfoque del estudio, la forma en que se seleccionaron los datos, los indicadores que se calcularon y el procedimiento seguido para obtener los resultados.

4.1 Enfoque y unidad de análisis

Para poder proponer una solución de ciberseguridad que funcione en la práctica, primero es necesario entender qué está fallando. Por esa razón, el punto de partida del estudio fue revisar los registros que dejó el sistema de seguridad de la organización a lo largo de 2025. Cada vez que el sistema detectaba un intento de ataque, una alerta o un problema de seguridad, generaba un registro con información sobre lo que ocurrió, cuándo ocurrió y si fue resuelto o no. A partir de ese análisis fue posible identificar con precisión los puntos más débiles del sistema y determinar los elementos que debía contemplar la propuesta, la cual responde a lo que realmente está ocurriendo y no a escenarios hipotéticos.

4.2 Muestra, criterios e instrumentos

En este apartado se describe el origen de los datos utilizados en el estudio, los criterios que debía cumplir un registro para ser incluido en el análisis y el instrumento empleado para su extracción. Dado que los sistemas de seguridad generan continuamente registros de todo lo que ocurre en la infraestructura, no

todos contienen información suficiente para ser analizados con rigor, pues algunos están incompletos, otros son duplicados y otros simplemente no aportan datos útiles. Por esa razón se definieron criterios claros de selección que permitieran trabajar únicamente con los registros más completos y confiables, incluyendo aquellos que contaban con fecha y hora del evento, tipo de ataque, nivel de gravedad, estado de resolución y otros campos clave para el análisis. Tras ese proceso de revisión y depuración, quedaron 385 registros válidos distribuidos a lo largo de los doce meses del año 2025. Para extraer la información de forma ordenada y consistente se utilizó una ficha de extracción estructurada, diseñada a partir del formato que ya usaba el propio equipo de seguridad de la organización, lo que facilitó la recolección y aseguró que todos los datos fueran comparables entre sí. La Tabla 1 ilustra un ejemplo concreto de cómo lucía un registro válido una vez extraído mediante dicha ficha.

Tabla 1. Ejemplo de registro incluido en el análisis

Campo	Ejemplo de registro
Fecha y hora	14/03/2025 – 02:47
Tipo de ataque	Malware
Nivel de gravedad	Alta
Nivel de confianza	94%
Versión del agente	6.2.1
Categoría de análisis	Endpoint
Estado de mitigación	No atendido

4.3 Variables e indicadores

Con los registros depurados, el siguiente paso fue definir cómo medir el estado real de la ciberseguridad. Para ello se calcularon dos indicadores. El primero es la tasa de vulnerabilidad atendida, que responde a una pregunta concreta: de todos los problemas de seguridad que se detectaron, cuántos fueron realmente resueltos. Si este porcentaje es bajo, significa que muchas amenazas quedan sin atender, dejando a la organización expuesta. Este indicador se calcula como se muestra en la Ecuación 1.

$$\text{TVA}(\%) = \frac{\text{Vulnerabilidades atendidas}}{\text{Vulnerabilidades totales}} \times 100 \tag{1}$$

El segundo indicador es el índice de respuesta, que mide cuántas de las alertas generadas por el sistema fueron gestionadas correctamente, es decir, investigadas, clasificadas y resueltas de manera adecuada. Este indicador muestra si el equipo de seguridad no solo detecta los problemas, sino que también es capaz de resolverlos, y se expresa mediante la Ecuación 2.

$$\text{IR}(\%) = \frac{\text{Alertas procesadas correctamente}}{\text{Alertas totales}} \times 100 \tag{2}$$

Ambos indicadores se calcularon mes a mes y luego se resumieron en un promedio anual y un rango de variación mensual. Considerados de forma conjunta, permiten construir una imagen diagnóstica del estado real de la ciberseguridad en la organización, superando la limitación de las métricas de detección tradicionales, que registran lo que ocurre pero no dan cuenta de si fue resuelto ni con qué eficacia. En ese sentido, ambos indicadores se articulan de manera que uno acota la magnitud de la exposición no resuelta mientras el otro revela si el equipo de seguridad tiene capacidad para cerrar el ciclo completo de gestión de un incidente. Por tanto, esta dualidad es la que permite trazar una distinción relevante entre una organización que detecta amenazas pero no las neutraliza y una que responde de forma reactiva pero sin la sistematicidad necesaria para sostener una postura de seguridad robusta, distinción que resulta determinante para orientar con precisión los componentes de la arquitectura propuesta. Bajo esta premisa, la proxima subsección aborda el procedimiento de análisis y los aspectos éticos que han guiado la investigación.

4.4 Procedimiento de análisis y aspectos éticos

Una vez definidos los indicadores y seleccionados los registros, el análisis se desarrolló en cuatro pasos secuenciales que permitieron ir desde los datos en bruto hasta las conclusiones diagnósticas. En el primer paso se revisaron todos los registros para descartar los que tenían información incompleta o repetida, asegurando así que el análisis posterior se apoyara en datos confiables. En el segundo paso los eventos válidos se agruparon por mes, por tipo de amenaza, por nivel de gravedad y por estado de resolución, lo que permitió identificar con claridad qué tipos de ataque se repetían con mayor frecuencia y cuáles quedaban sistemáticamente sin resolver. En el tercer paso se calcularon los dos indicadores mes a mes, obteniendo una visión completa de cómo evolucionó la situación de seguridad a lo largo del año. En el cuarto y último paso, cada problema identificado se relacionó con un área específica de la arquitectura de ciberseguridad, estableciendo así la conexión directa entre lo que falló y lo que la arquitectura propuesta busca corregir. En cuanto a los aspectos éticos, el análisis se realizó sobre registros agregados que no contienen datos personales identificables, por lo que no implicó ningún tipo de intervención sobre personas. Los hallazgos obtenidos a partir de este procedimiento se presentan en la siguiente sección.

5. Resultados

Esta sección presenta los hallazgos obtenidos a partir del análisis de los 385 registros técnicos de seguridad documentados durante 2025. Se organiza en tres apartados que corresponden a los tres ejes de análisis del estudio: la vulnerabilidad de los servicios administrativos frente a intentos de ataque, el desempeño del sistema de protección existente, y la arquitectura de ciberseguridad propuesta como respuesta a las brechas identificadas. Los resultados se expresan mediante estadística descriptiva, recuentos, porcentajes y rangos, dado que el estudio no formuló hipótesis ni aplicó contrastes inferenciales.

5.1 Vulnerabilidad de los servicios administrativos

En este apartado se caracteriza la exposición de la organización frente a amenazas informáticas durante 2025, a partir de los indicadores calculados sobre los registros técnicos de incidentes. A lo largo del año se documentaron 385 eventos de seguridad, con una frecuencia mensual que osciló entre 14 y 101 ataques. El mes con mayor volumen fue noviembre, con 101 eventos registrados, mientras que los meses con menor actividad fueron setiembre y octubre, con 14 y 18 ataques respectivamente. La Figura 1 presenta la distribución mensual de estos eventos.

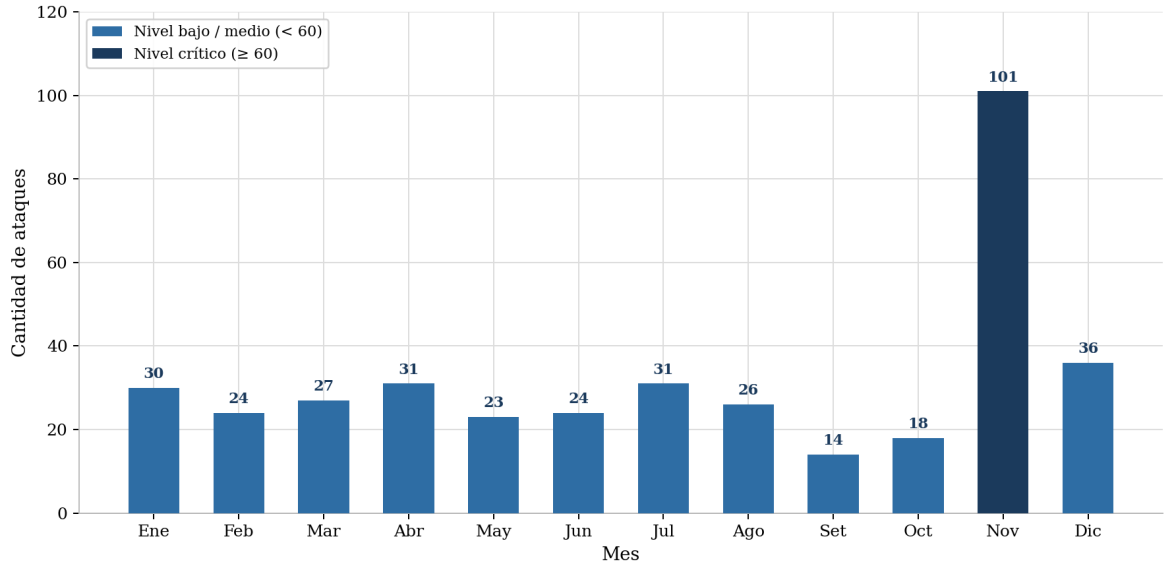

Figura 1. Cantidad mensual de ataques registrados en 2025

La clasificación por rangos de frecuencia mensual permitió identificar tres niveles de criticidad. Ocho meses se ubicaron en un rango de 10 a 30 ataques y fueron clasificados como nivel bajo, tres meses se ubicaron entre 30 y 60 ataques y correspondieron al nivel medio, y un mes superó los 60 ataques alcanzando el nivel crítico, asociado al incremento de actividad propio del cierre anual. Esta distribución muestra que la

exposición extrema se concentró en un período específico, aunque la presencia de ataques fue constante a lo largo de todo el año.

En cuanto al tipo de amenaza, los ataques desconocidos y de tipo packed predominaron en los niveles de menor criticidad, el malware se presentó de forma recurrente en los niveles medio y alto con 54 eventos registrados en criticidad alta, y el ransomware se concentró en el nivel crítico con 30 ataques documentados. La Figura 2 resume la distribución por tipo y nivel de criticidad.

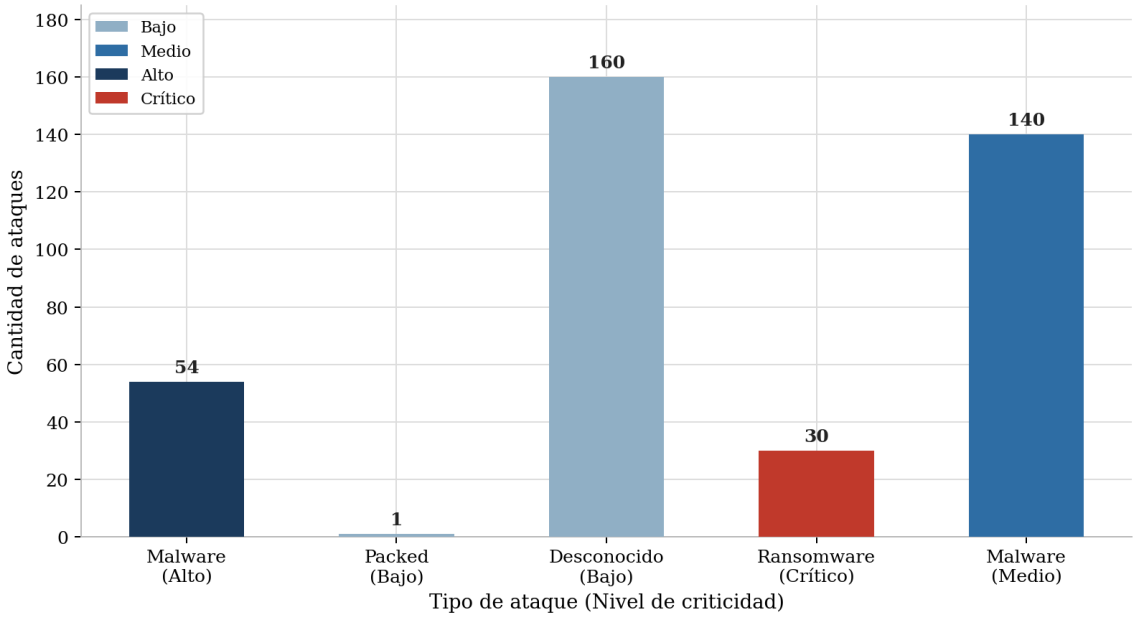

Figura 2. Análisis por tipo de ataque y nivel de criticidad

El análisis trimestral mostró que los ataques de criticidad alta o crítica oscilaron entre 3 y 5 eventos en cada uno de los tres primeros trimestres, mientras que en el cuarto trimestre ese rango se amplió significativamente, entre 12 y 32 eventos, lo que evidencia una mayor carga de amenazas graves en la etapa final del año. La Figura 3 ilustra la evolución de cada nivel de criticidad a lo largo de los cuatro trimestres.

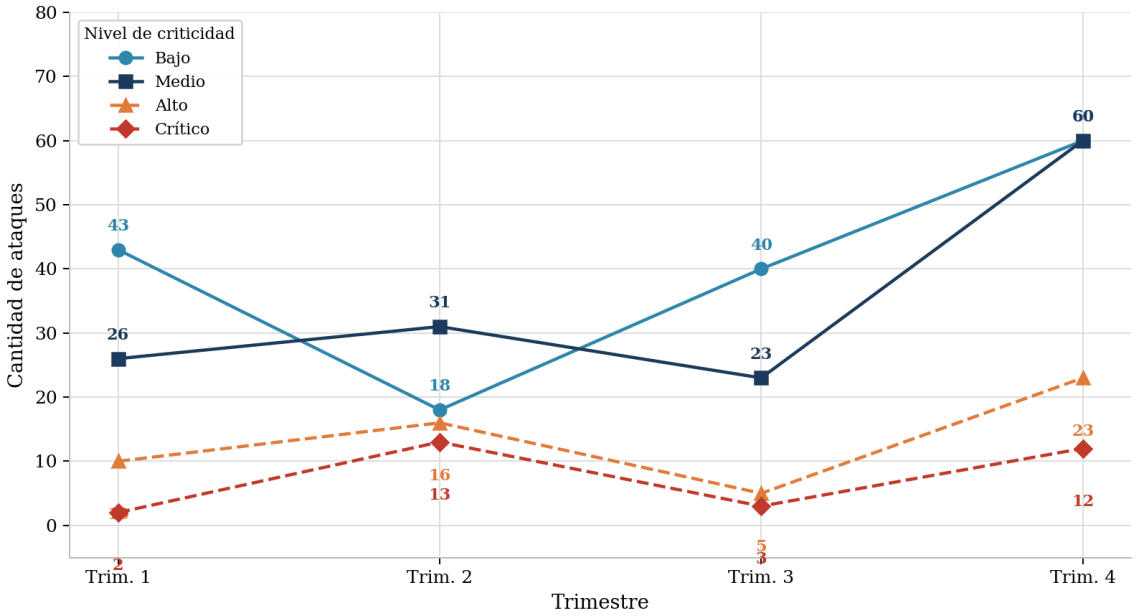

Figura 3. Análisis trimestral de ataques por nivel de criticidad

La tasa de vulnerabilidad atendida, calculada con 209 vulnerabilidades atendidas sobre 385 registros totales, alcanzó un valor anual de 54,3%, con una variación mensual entre 44,4% y 62,5%. Esto significa que prácticamente la mitad de los eventos registrados no fueron resueltos, dejando a la organización con una

exposición residual sostenida a lo largo de todo el período. La Figura 4 muestra la evolución mensual de este indicador.

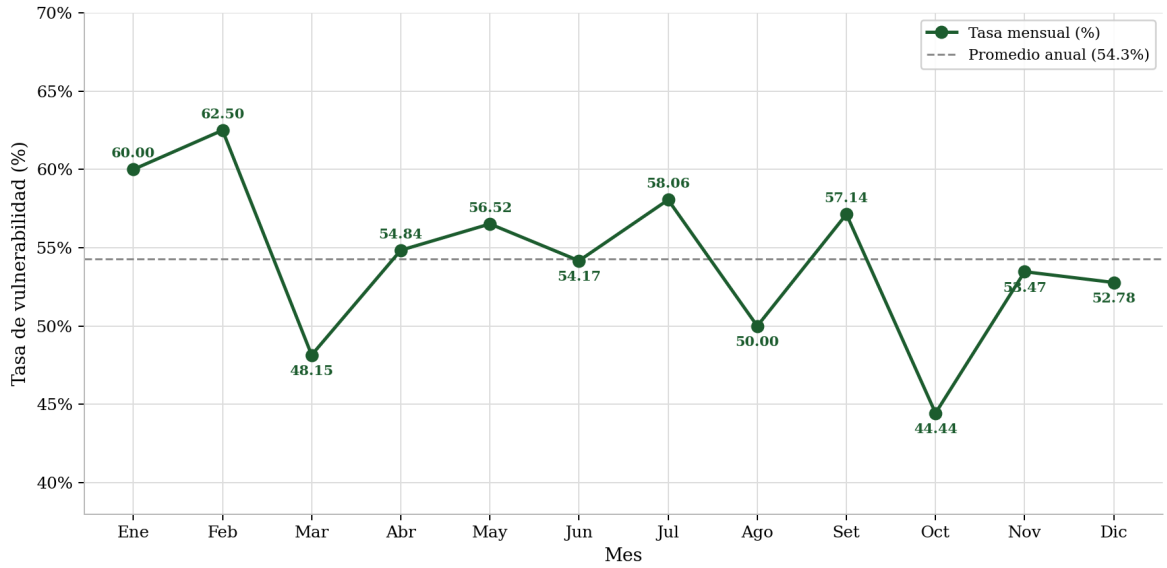

Figura 4. Evolución mensual de la tasa de vulnerabilidad atendida (%) en 2025

5.2 Desempeño del sistema de protección

Una vez caracterizada la exposición, se procedio a evaluar la capacidad del sistema de protección existente para hacer frente a esos eventos, tomando como referencia los estados de mitigación y la oportunidad de las atenciones registradas. Los ataques no mitigados oscilaron entre 6 y 47 casos mensuales, los ataques con atención deficiente entre 1 y 18 casos mensuales, y los ataques con atención óptima entre 0 y 41 casos mensuales. Noviembre concentró el mayor número de ataques sin mitigar (47), mientras que los meses de marzo y julio no registraron ninguna atención óptima. La Figura 5 presenta la evolución mensual de estos tres estados de mitigación.

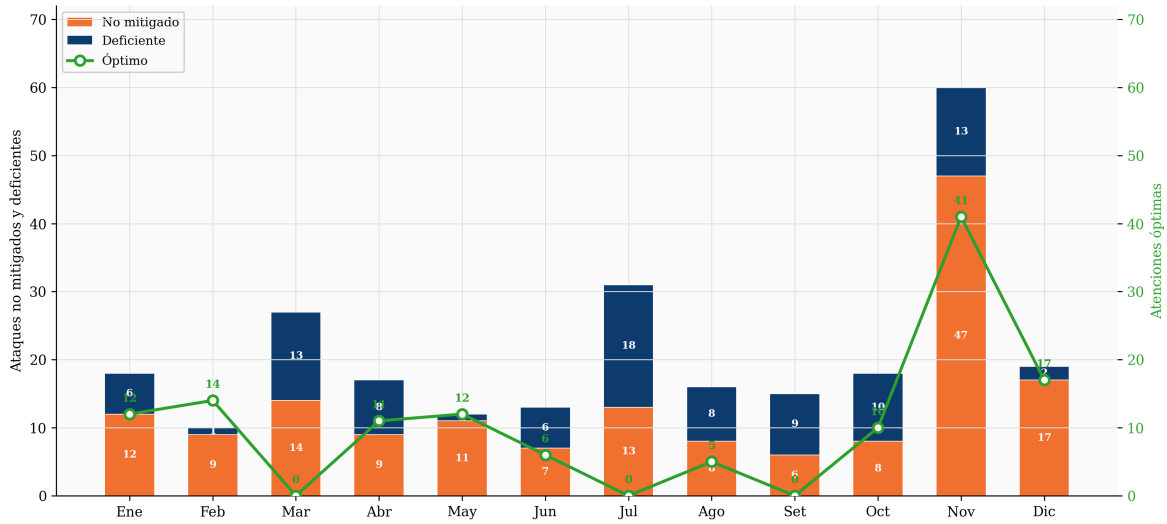

Figura 5. Cantidad de ataques por nivel de mitigación

De igual manera, al analizar las atenciones óptimas según el nivel de criticidad, se revela no solo cuántos incidentes fueron resueltos de manera adecuada sino qué tipos de amenazas concentraron la mayor capacidad de respuesta efectiva del equipo durante el período, tal como ilustra la Figura 6. En ese marco, los casos críticos atendidos dentro del plazo óptimo oscilaron entre 0 y 5 por mes, rango que no refleja estabilidad operativa sino la escasa frecuencia con que los incidentes de mayor severidad lograron gestionarse adecuadamente, confirmando que precisamente los eventos de mayor riesgo fueron los que con mayor frecuencia escaparon a la capacidad de respuesta oportuna. A ello se suma que los casos de riesgo

alto presentaron un rango ligeramente mayor, entre 0 y 8 atenciones óptimas mensuales, manteniéndose igualmente en valores bajos en relación con el volumen total registrado en esa categoría. Por su parte, los casos de criticidad media mostraron la mayor dispersión, con valores entre 0 y 17 atenciones óptimas y un pico notable en noviembre que coincidió con el mayor volumen de ataques del año, lo que sugiere una movilización reactiva de recursos ante la presión del volumen y no una mejora estructural sostenida. Por tanto, este patrón evidencia una asignación de recursos que priorizó el volumen sobre la severidad, limitación que la arquitectura propuesta en la sección siguiente busca corregir mediante mecanismos de priorización basados en criticidad real.

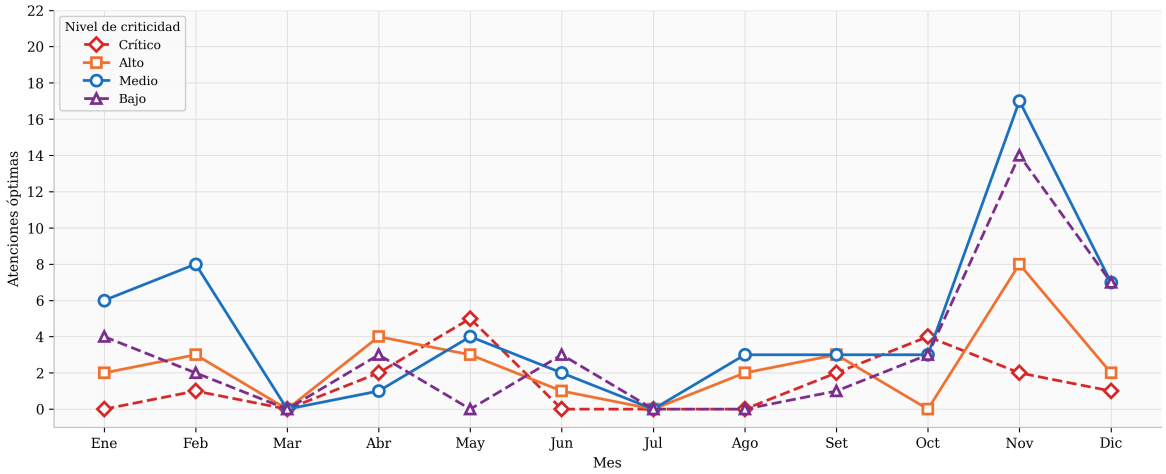

Figura 6. Atenciones óptimas por nivel de criticidad

Ahora bien, al examinar la composición por tipo de ataque entre los eventos mitigados de forma óptima, se observa una variabilidad considerable que refleja tanto la naturaleza cambiante de las amenazas recibidas como la capacidad desigual del sistema para responder a cada categoría en distintos momentos del año, tal como se observa en la Figura 7. En ese contexto, el malware representó entre el 7% y el 33% de las atenciones óptimas mensuales, el ransomware exhibió la mayor amplitud de variación con un rango de 6% a 78%, y los ataques de origen desconocido oscilaron entre el 10% y el 54%, lo que indica que ninguna categoría dominó de forma estable la composición de las atenciones efectivas a lo largo del período. A ello se añade que marzo y julio no registraron ninguna atención óptima, lo que significa que la totalidad de los eventos recibidos en esos meses quedó fuera del umbral de resolución adecuada. Ante ese escenario, las brechas absolutas de respuesta observadas en dichos períodos no pueden atribuirse a un tipo específico de amenaza sino a limitaciones generales de la arquitectura de seguridad, lo que refuerza la necesidad de una intervención estructural y no meramente reactiva.

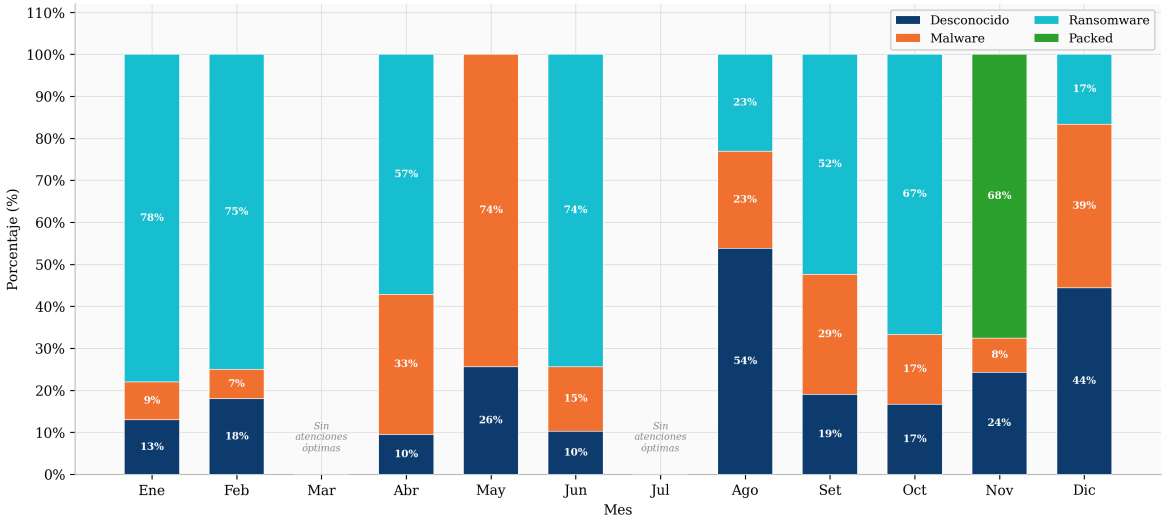

Figura 7. Distribución porcentual mensual del tipo de ataque en atenciones óptimas

Siguiendo esa misma línea, al examinar el total de ataques que no recibieron atención efectiva según el

tipo de amenaza, se observa una concentración preocupante en las categorías de mayor impacto potencial, donde 90 correspondieron a malware clasificado con criticidad alta por el riesgo de inoperatividad que representa para los servicios administrativos, 73 a amenazas de origen desconocido clasificadas con criticidad media dado su elevado potencial de generar pérdida o exposición de información, y 13 a ransomware clasificado con criticidad baja por tratarse de una amenaza que pudo ser contenida parcialmente antes de causar daño irreversible. Ante esa distribución, el hecho de que el mayor volumen de eventos no atendidos corresponda precisamente a malware de criticidad alta indica que la arquitectura actual no solo falla en cantidad sino en selectividad, dejando sin resolución los incidentes de mayor consecuencia operativa. Este déficit se agudiza aún más cuando se examina la distribución geográfica interna de los ataques no atendidos, pues el sector comercial concentró 70 eventos de origen desconocido y 38 de malware, acumulando el mayor volumen de la organización por su exposición constante a múltiples canales digitales de comunicación, lo que lo posiciona como el frente más vulnerable y prioritario para la intervención de la arquitectura propuesta, tal como se evidencia en la Figura 8.

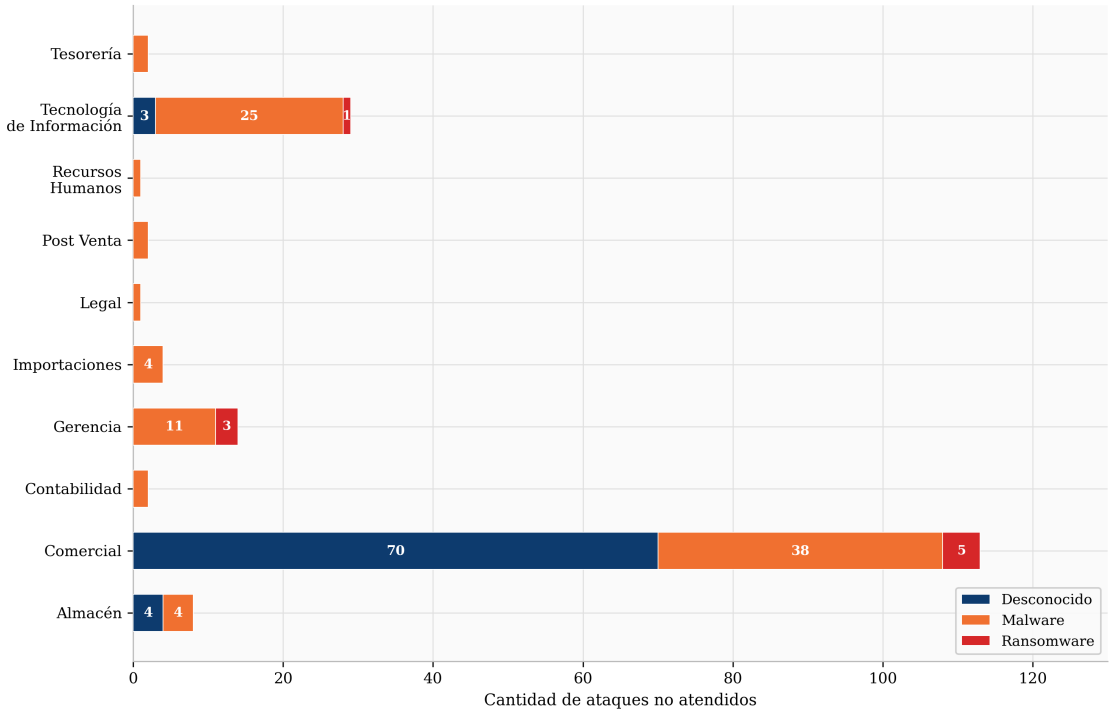

Figura 8. Ataques no atendidos por área y tipo de amenaza

Para complementar lo anterior la Figura 9 presenta la evolución mensual de este indicador.

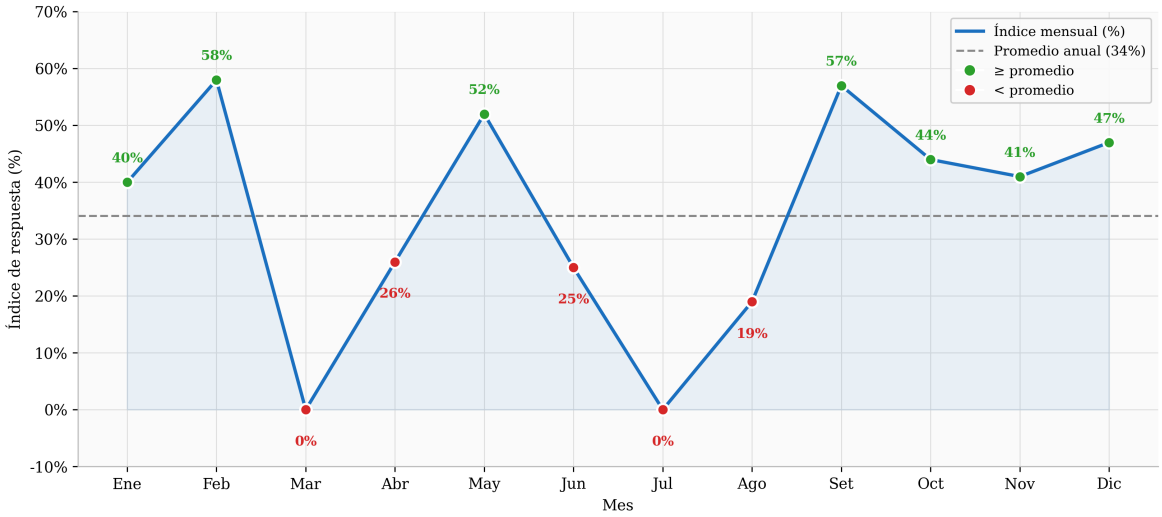

Figura 9. Evolución mensual del índice de respuesta (%) en 2025

5.3 Arquitectura moderna de ciberseguridad propuesta

Los hallazgos presentados en las subsecciones anteriores revelan una tasa de vulnerabilidad atendida del 54,3 % y un índice de respuesta del 34 %, cifras que no son aisladas sino el reflejo de una arquitectura de seguridad que carece de integración, priorización y capacidad de cierre efectivo. Frente a ese diagnóstico, la solución no puede limitarse a reforzar controles existentes sino que exige una respuesta estructural. Con ese propósito, se diseñó una arquitectura moderna de ciberseguridad orientada a reducir la exposición residual, mejorar el desempeño del sistema de protección y fortalecer la prevención del fraude informático en entornos administrativos digitalizados.

La arquitectura se organiza en seis dominios funcionales que operan de manera interconectada y se articulan alrededor de un núcleo central de detección avanzada e inteligencia artificial, tal como se ilustra en la Figura 10. El primer dominio, gestión de identidades y accesos, controla quién puede ingresar a cada sistema con qué privilegios y bajo qué condiciones, implementando autenticación multifactor, inicio de sesión único y control de acceso basado en roles para reducir el riesgo de accesos no autorizados que fueron una fuente recurrente de incidentes en el diagnóstico. El segundo dominio, protección contra amenazas, detecta, contiene y neutraliza eventos maliciosos como el malware y el ransomware antes de que puedan causar daño, apoyándose en detección y respuesta en endpoints, filtrado de tráfico de red y motores antimalware en tiempo real. El tercer dominio, seguridad en la nube, protege los servicios ERP y las plataformas digitales que sostienen la operación, asegurando configuraciones correctas y visibilidad sobre las aplicaciones desplegadas en entornos cloud mediante brokers de seguridad de acceso y verificación de cumplimiento de configuración.

Figura 10. Arquitectura moderna de ciberseguridad propuesta

En cuanto al cuarto dominio, la protección de información clasifica los datos según su nivel de sensibilidad y establece controles de prevención de fuga y cifrado tanto en reposo como en tránsito, un riesgo especialmente crítico en organizaciones que manejan información financiera y contable. Por su parte, el quinto dominio, cumplimiento y auditoría, verifica periódicamente que los controles estén funcionando, documenta evidencias mediante logs de monitoreo y genera las acciones correctivas necesarias alineadas con estándares como ISO 27001. A ello se suma el sexto dominio, respuesta a incidentes, que garantiza que cada evento sea atendido bajo un protocolo definido y trazable mediante orquestación, automatización y playbooks estandarizados, cerrando el ciclo que la arquitectura actual dejaba abierto. De forma transversal, una capa de gobierno de seguridad, cumplimiento normativo, capacitación continua y gestión del riesgo sustenta la operación de todos los dominios, reconociendo que la seguridad es tanto un problema técnico como organizacional.

El valor de esta arquitectura no reside únicamente en la existencia de los seis dominios sino en la manera en que estos interactúan a través del núcleo de detección avanzada, convirtiendo lo que en la arquitectura actual era una colección de controles aislados en un sistema de respuesta cohesionado y autoajustable. En ese sentido, la Figura 11 ilustra el flujo operativo que sigue un incidente desde su detección hasta su resolución y aprendizaje, evidenciando que cada etapa no opera de forma independiente sino que alimenta y condiciona a la siguiente. Cuando el sistema identifica un evento mediante sensores, logs o agentes EDR, este es normalizado y filtrado mediante parseo, triage y deduplicación antes de ser enviado al núcleo de inteligencia artificial, donde se correlaciona con otros eventos para determinar si forma parte de un patrón de ataque más amplio y no un incidente aislado. A partir de esa correlación, el evento se prioriza según su nivel de criticidad real y se activa una respuesta automatizada mediante SOAR y playbooks si el caso lo permite, o se escala al analista correspondiente cuando la complejidad del incidente requiere criterio humano, garantizando así que los recursos de atención se asignen en función de la severidad y no del orden de llegada.

Figura 11. Flujo operativo de detección, priorización y respuesta a incidentes

Una vez contenido y remediado el incidente mediante aislamiento, parchado o rollback, el sistema no concluye el ciclo sino que registra las lecciones aprendidas y actualiza las métricas de desempeño para mejorar la detección de eventos similares en el futuro, cerrando un bucle de mejora continua que la arquitectura actual no contempla. Esta capacidad de aprendizaje acumulativo es precisamente lo que distingue una arquitectura reactiva de una adaptativa, pues cada incidente gestionado se convierte en insumo para afinar los umbrales de detección, ajustar las reglas de correlación y reducir progresivamente los tiempos de respuesta. Este diseño aborda directamente la debilidad central identificada en el diagnóstico, donde la tasa de vulnerabilidad atendida del 54,3 % y el índice de respuesta del 34 % no eran el resultado de falta de recursos sino de la ausencia de un proceso sistemático que conectara la detección con la respuesta efectiva, la priorización por criticidad y el registro estructurado de aprendizajes, tres dimensiones que este flujo integra de manera explícita y medible.

Ahora bien, para llevar esta arquitectura a la práctica, su implementación se organiza en tres fases progresivas a lo largo de doce meses, tal como se muestra en la Figura 12. La primera fase, denominada fundamentos y protección y comprendida entre los meses 1 y 4, instala los controles de identidad, acceso y protección básica, establece las políticas de seguridad y realiza el inventario de activos y evaluación de riesgos que sirven de base para las fases siguientes, con la meta de elevar la tasa de vulnerabilidad atendida por encima del 65 % y el índice de respuesta por encima del 40 %. Sobre esa base, la segunda fase, detección y automatización, desarrollada entre los meses 5 y 8, conecta los dominios entre sí mediante el núcleo de inteligencia artificial, activa la correlación de eventos, despliega la protección en la nube y la respuesta automatizada, persiguiendo una tasa de vulnerabilidad atendida superior al 75 % y un índice de respuesta superior al 55 %. A partir de los avances consolidados en esa etapa, la tercera fase, optimización y mejora continua, comprendida entre los meses 9 y 12, refuerza el proceso mediante auditorías periódicas, capacitación del personal, ajuste dinámico de reglas y revisión de la arquitectura, con el objetivo de que la tasa de vulnerabilidad atendida supere el 85 % y el índice de respuesta alcance el 70 %. Este despliegue escalonado garantiza que cada fase construya sobre los logros de la anterior y que los indicadores sirvan de señal objetiva de avance a lo largo de todo el proceso de transformación.

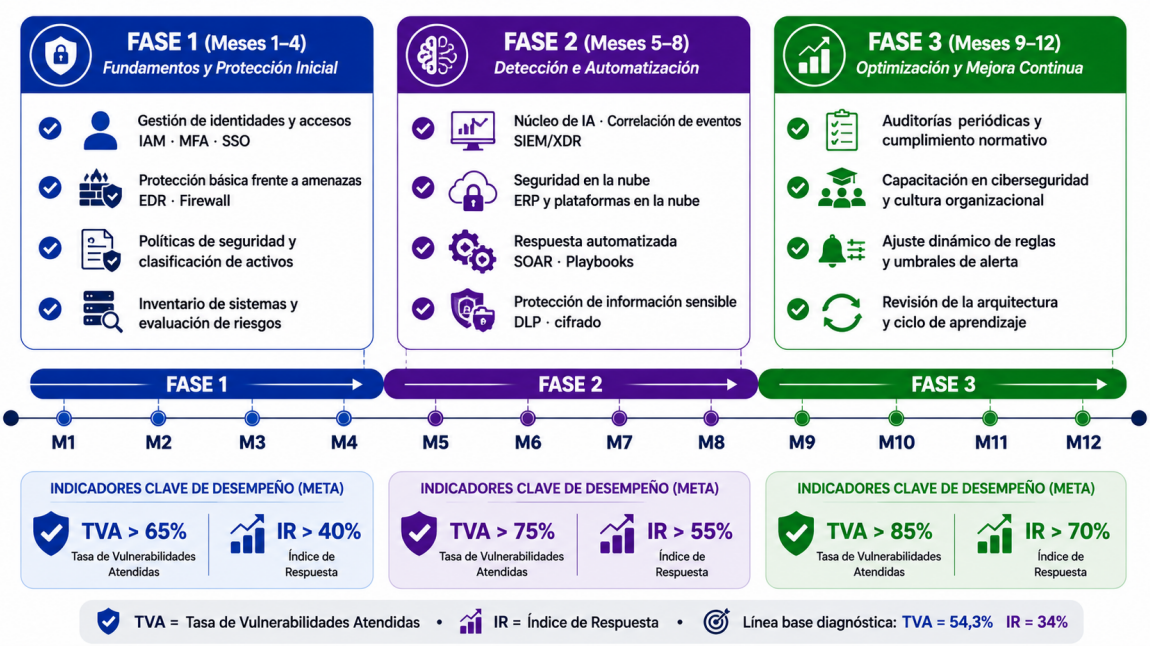

Figura 12. Hoja de ruta para la implementación de la arquitectura propuesta

Junto a los fundamentos conceptuales ya descritos, la solidez de una arquitectura de ciberseguridad no depende únicamente de la coherencia entre sus componentes sino del respaldo empírico que justifica cada decisión de diseño. En ese sentido, el conjunto de estudios que sustenta esta propuesta cubre un espectro amplio de dimensiones, desde la detección inteligente de anomalías y el análisis estadístico de patrones hasta la ingeniería social, la concientización del usuario y la vinculación entre riesgo de seguridad y procesos de negocio. Cada referente fue seleccionado por su correspondencia directa con alguna de las brechas identificadas en el diagnóstico, lo que garantiza que la propuesta no constituya una solución genérica sino una respuesta calibrada a las limitaciones específicas de la organización.

De esta manera, la arquitectura no solo se apoya en evidencia consolidada sino que establece un puente explícito entre los hallazgos del análisis y las decisiones de diseño adoptadas. Algunos estudios respaldan la selección de indicadores y la lectura cuantitativa de patrones anómalos, otros justifican la integración de mecanismos de concientización y cultura de seguridad, y otros conectan directamente el riesgo operativo con los procesos administrativos digitalizados. En conjunto, estos referentes confirman que cada componente de la arquitectura propuesta tiene sustento en investigación previa y no responde a criterios arbitrarios de diseño, relación que se sintetiza en la Tabla 2.

Tabla 2. Estudios complementarios que sustentan la propuesta arquitectónica

Autor	Eje de aporte	Uso en el manuscrito
Zioviris et al. [15]	Detección inteligente	Soporta el uso de modelos secuenciales para fraude.
Zogaj et al. [16]	Análisis estadístico	Refuerza la lectura cuantitativa de patrones anómalos.
Vanini et al. [17]	Gestión de riesgo	Conecta anomalías de pago con gestión antifraude.
Ukwandu et al. [18]	Infraestructura crítica	Muestra impactos operativos en sectores digitalizados.
Torres Gamarra [19]	Brechas estructurales	Contextualiza amenazas, legislación y brechas regionales.
Szczepaniuk et al. [20]	Sistemas inteligentes	Aporta criterios para seguridad en entornos conectados.
Santillán et al. [21]	Revisión integral	Sustenta la visión multidominio de ciberseguridad.
Rodrigues et al. [22]	Comercio electrónico	Relaciona fraude digital con prevención operativa.
Pittoli et al. [23]	Arquitecturas	Apoya la necesidad de diseños adaptados al entorno.
Pauta Ayabaca et al. [24]	Tecnologías emergentes	Vincula adopción tecnológica con capacidades preventivas.
Orosco-Fabian et al. [25]	Bibliometría	Sitúa la investigación regional en ciberseguridad.
Schroeder et al. [26]	Medición	Respalda la selección de indicadores de seguridad.
Awobelem George [27]	Gobierno digital	Refuerza la seguridad como capacidad institucional.
Khan et al. [28]	Detección de anomalías	Apoya la analítica para eventos no conocidos.
Hernández Aros et al. [29]	Fraude financiero	Conecta aprendizaje automático con detección de fraude.
Fueres Quilumbango et al. [30]	Ingeniería social	Integra usuarios y capacitación en la arquitectura.
Chaudhary et al. [31]	Concientización	Justifica medir programas de cultura de seguridad.
Rosado et al. [32]	Procesos de negocio	Conecta riesgo de seguridad con procesos administrativos.
Bennouk et al. [33]	Vulnerabilidades	Respalda detección y gestión automatizada de fallas.
Alshomrani et al. [34]	IA explicable	Aporta transparencia para detección temprana.

A partir de ese marco teórico, cada dominio de la arquitectura fue asociado a una referencia central que fundamenta tanto su diseño como su función operativa dentro del sistema integrado de protección. Esta correspondencia entre evidencia académica y componente arquitectónico garantiza que las decisiones de diseño respondan a criterios validados en la literatura y no únicamente a consideraciones tecnológicas, como se detalla en la Tabla 3.

Al examinar los dominios, se aprecia que su diseño no obedece a una lógica aditiva sino a una arquitectura de interdependencias donde cada componente refuerza y condiciona al siguiente. Por ejemplo, el dominio de identidad y acceso constituye el punto de entrada de toda la cadena de protección, pues sin un control riguroso de privilegios y autenticación, cualquier mecanismo posterior opera sobre una base comprometida. A partir de ese control, el dominio de protección contra amenazas extiende la cobertura hacia los endpoints y la red, asumiendo que los incidentes de malware y actividad desconocida identificados en el diagnóstico no pueden contenerse únicamente en el perímetro. En ese mismo nivel operativo, la seguridad en la nube complementa esa cobertura al proteger los entornos ERP y los servicios digitales sobre los que descansa la operación administrativa de la organización. Sobre esa base técnica, el dominio de protección de información añade una capa de custodia orientada específicamente a los datos sensibles, cuya exposición o pérdida representa el mayor riesgo de fraude en contextos financieros y contables. A ello se suma el dominio de cumplimiento y auditoría, que cierra el ciclo de control verificando periódicamente

que los mecanismos anteriores funcionen según lo previsto y generando las evidencias necesarias para la rendición de cuentas. De esta manera, el dominio de detección avanzada e inteligencia artificial opera de forma transversal sobre todos los anteriores, correlacionando eventos, priorizando alertas y aprendiendo de patrones históricos para que el sistema mejore su capacidad de respuesta con cada incidente gestionado.

Tabla 3. Dominios y componentes de la arquitectura propuesta

Referencia	Dominio	Función principal
Raj [14]	Identidad y acceso	Controlar privilegios, autenticación multifactor, accesos condicionales y cuentas de alto riesgo.
Umer et al. [11]	Protección contra amenazas	Detectar, contener y responder ante malware, eventos desconocidos y actividad maliciosa en endpoints y red.
Cevallos Ramos et al. [8]	Seguridad en la nube	Proteger servicios ERP, configuraciones, cargas de trabajo y políticas desplegadas en entornos cloud.
Yaseen [7]	Protección de información	Clasificar, custodiar y monitorear datos sensibles para reducir fuga, manipulación o pérdida.
Agyepong et al. [6]	Cumplimiento y auditoría	Verificar controles, evidencias, responsables, brechas y acciones correctivas de forma periódica.
Moura et al. [13]	Detección avanzada e IA	Correlacionar eventos, priorizar alertas, automatizar respuesta y aprender de patrones históricos.

6. Discusión

Después de haber expuesto los resultados y la propuesta arquitectónica, corresponde ahora interpretar su alcance, contrastarlos con la literatura disponible y delimitar las implicaciones para la gestión organizacional. Los servicios administrativos digitalizados concentran procesos financieros, logísticos y documentales que, por su interdependencia y volumen, representan un objetivo persistente para el fraude organizacional. En ese contexto, la capacidad de una organización para detectar amenazas y convertirlas en respuestas oportunas no depende únicamente de la tecnología disponible, sino de la coherencia entre sus procesos de detección, priorización y cierre operativo. Los resultados de este estudio revelan precisamente una desconexión entre ambas capacidades. La tasa de vulnerabilidad atendida de 54,3% indica que el sistema identifica más de la mitad de los eventos de riesgo, pero el índice de respuesta de 34% evidencia que solo una fracción de esas alertas se convierte en acciones trazables, oportunas y verificables. Esta brecha no constituye un déficit técnico puntual, sino que refleja una insuficiencia en la arquitectura organizacional de seguridad, donde la detección y la gestión operativa operan como procesos desacoplados, lo que amplía la exposición a suplantación de identidad, manipulación de datos e interrupción de servicios.

Esta interpretación se alinea con la advertencia de Agyepong et al. [6], quienes señalan que la madurez en operaciones de seguridad no se mide por el volumen de alertas generadas, sino por la calidad de las decisiones que estas producen. De manera complementaria, Forsberg y Frantti [10] sostienen que una gestión efectiva de vulnerabilidades exige que detección, priorización y remediación formen un ciclo medible y cerrado. Bajo estos criterios, los indicadores empleados capturan las etapas iniciales del ciclo pero no su cierre, dado que no se dispone de datos sobre tiempo medio de contención, severidad residual, reincidencia o costo por incidente. Esta acotación no invalida los hallazgos, sino que precisa su alcance. Los resultados son diagnósticos de actividad operativa y no de madurez sistémica, por lo que deben leerse como punto de partida para una evaluación más amplia.

La dimensión del fraude digital introduce una capa adicional de complejidad que merece atención particular. Afriyie et al. [9] demuestran que los modelos de detección basados en aprendizaje automático dependen críticamente de la consistencia de los datos de entrenamiento y de la estabilidad de los patrones opera-

tivos. Abu-Khadrah [5] subraya que las modalidades de fraude digital evolucionan con suficiente rapidez como para dejar obsoletos los sistemas de reglas estáticas. Proyectados sobre el contexto estudiado, ambos planteamientos indican que una arquitectura de respuesta efectiva no puede limitarse a automatizar controles preexistentes ni a incrementar el volumen de alertas. Su valor reside en integrar datos, contexto, priorización y explicabilidad para reducir falsos negativos, mitigar la fatiga de alertas y eliminar decisiones opacas. La ausencia de estas capacidades en el sistema actual es coherente con la brecha observada entre detección y cierre operativo.

Sobre esta base diagnóstica, la arquitectura propuesta responde a brechas que atraviesan múltiples dominios de control organizacional. Raj [14] promueve marcos modernos con monitoreo extendido y respuesta adaptativa. Moura et al. [13] evidencian que la inteligencia artificial fortalece la detección proactiva cuando se integra con procesos gobernados. Cevallos Ramos et al. [8] vinculan estos enfoques con ciberresiliencia organizacional. En ese marco, identidad, nube, protección de datos, cumplimiento, auditoría, XDR e inteligencia artificial deben operar como capas interdependientes. Cuando una falla, las demás aportan contención, trazabilidad y aprendizaje. Esta redundancia persigue un objetivo concreto, que es impedir que un evento aislado escale hacia fraude operativo o afecte la continuidad administrativa en entornos con usuarios distribuidos y datos compartidos en nube. La propuesta es arquitectónica y no experimental, por lo que su validez práctica permanece condicionada a una implementación controlada con indicadores comparables antes y después de la intervención.

La traducción a decisiones de gestión requiere reconocer que la seguridad es una capacidad operativa central y no una función de soporte técnico. Los costos organizacionales asociados a la brecha detectada incluyen interrupciones de servicio, restauración de información, sobrecarga del equipo de infraestructura y erosión de la confianza institucional. Por ello, la implementación de la arquitectura propuesta demanda responsables definidos, cronograma, presupuesto, capacitación continua y auditorías periódicas. Requiere además tableros que conecten indicadores técnicos con continuidad del negocio, riesgo residual y tiempos de respuesta verificables. El impacto deberá medirse con una línea base preimplementación que emplee los mismos indicadores utilizados en este diagnóstico, condición que permitirá comparar resultados con rigor y evitar atribuir mejoras que no hayan sido verificadas empíricamente.

Por último, las limitaciones del estudio deben leerse como delimitadores de su contribución y no como objeciones a su pertinencia. El caso analizado corresponde a una sola organización, por lo que la transferibilidad de los hallazgos depende del grado de similitud en contexto, tamaño, madurez digital y disponibilidad de recursos. El diseño es descriptivo y no experimental, de modo que identifica brechas y justifica una arquitectura pero no establece causalidad ni prueba efectividad. Los registros disponibles cubren un periodo operativo específico y su calidad está condicionada por la plataforma corporativa de clasificación. Estas condiciones definen el paso siguiente con precisión, que consiste en validar la arquitectura mediante un diseño longitudinal pre y postimplementación con indicadores comparables, incluyendo riesgo residual, continuidad operativa y capacidad real de respuesta antifraude.

7. Conclusiones

Este estudio proporciona evidencia que las brechas de seguridad en entornos administrativos digitalizados no responden a carencias tecnológicas aisladas sino a la ausencia de integración entre detección y respuesta efectiva. Sobre 385 registros técnicos documentados durante 2025, la tasa de vulnerabilidad atendida no superó el 54,3% y el índice de respuesta efectiva se situó en el 34%, lo que significa que casi la mitad de los eventos detectados circularon por el sistema sin generar una acción trazable ni verificable. Estos valores no son el reflejo de una capacidad técnica insuficiente sino de una arquitectura que carece de la integración necesaria para cerrar el ciclo completo entre la detección de una amenaza y su resolución operativa. Frente a esa condición, la arquitectura diseñada introduce una lógica de interdependencia deliberada en la que gestión de identidad y acceso, protección contra amenazas avanzadas, seguridad en entornos cloud, custodia de información sensible, cumplimiento normativo, auditoría continua, analítica extendida, detección proactiva mediante inteligencia artificial, capacitación del personal y mejora continua institucionalizada conforman un sistema donde el fallo de un dominio no produce un colapso sino una contención, porque los dominios restantes sostienen la trazabilidad y la respuesta. Esta propiedad, ausente en la arquitectura

actual, es la que determina la diferencia entre un sistema que detecta y uno que protege.

Conviene precisar que la arquitectura propuesta en este trabajo constituye un diseño fundamentado en evidencia diagnóstica y respaldo teórico, pero cuya validación no forma parte del alcance de este estudio. Esa validación representa la línea de trabajo más inmediata y necesaria, y debería articularse mediante estudios de seguimiento que midan la evolución de los mismos indicadores tras la implementación, incorporando métricas complementarias de tiempo de contención, severidad residual y tasa de reincidencia. Extender posteriormente este análisis a organizaciones de distinto tamaño y sector permitiría examinar la transferibilidad de los hallazgos y contribuir al desarrollo de marcos de referencia comparables para la ciberseguridad en entornos administrativos digitalizados.

Contribución de los autores. Los autores confirman su contribución al artículo de la siguiente manera: conceptualización, E.C.C.; metodología, J.R.E. y E.C.C.; investigación, E.C.C. y J.R.E.; análisis formal, J.R.E.; redacción – preparación del borrador original, E.C.C.; redacción – revisión y edición, E.C.C.; visualización, J.R.E. y E.C.C. Todos los autores revisaron los resultados y aprobaron la versión final del manuscrito.

Agradecimientos. Los autores agradecen a la Universidad Tecnológica del Perú y a los especialistas que facilitaron la comprensión operativa del caso, manteniendo la confidencialidad de datos sensibles.

Financiamiento. El estudio ha sido autofinanciado por los autores.

Declaración de consentimiento informado. No aplica. El estudio utilizó registros técnicos organizacionales anonimizados y no involucró reclutamiento directo de participantes humanos.

Declaración de disponibilidad de datos. Los datos operativos pertenecen a la empresa estudiada. Las matrices agregadas, criterios de codificación e instrumentos pueden solicitarse razonablemente a los autores, sujetos a restricciones de confidencialidad.

Conflictos de intereses. Los autores declaran no tener conflictos de intereses.

Declaración de uso de inteligencia artificial. Los autores emplearon herramientas de inteligencia artificial generativa exclusivamente para mejorar la claridad y el estilo del texto. El análisis, la interpretación y las conclusiones son responsabilidad intelectual exclusiva de los autores, quienes validaron y aprobaron la versión final.

Referencias

[1] F. Houghton, M. Winterburn, and K. Oakley, "Rhysida Ransomware Attack on the British Library: Prioritisation, Expertise, and Funding Issues," *Information Technology and Libraries*, vol. 44, no. 1, Mar. 2025. doi: 10.5860/ital.v44i1.17112.

[2] N. Wadesango and E. Maveneka, "Cyberthreats and their impact on financial integrity: Evaluating the effectiveness of local authorities cybersecurity policies in preventing and detecting fraud," *Corporate Law and Governance Review*, vol. 7, no. 2, p. 32, Apr. 2025. doi: 10.22495/clgrv7i2p3.

[3] V. H. Che Leon Antinori, "Influence of Economic Cybercrime in Peru," *SCIENDO*, vol. 27, no. 4, pp. 415–418, Oct. 2024. doi: 10.17268/sciendo.2024.071.

[4] D. F. Wahid and E. Hassini, "An augmented AI-based hybrid fraud detection framework for invoicing platforms," *Applied Intelligence*, vol. 54, no. 2, pp. 1297–1310, Jan. 2024. doi: 10.1007/s10489-023-05223-x.

[5] A. Abu-Khadrah, S. Al-Washmi, A. Mohd Ali, and M. Jarrah, "Cyber-fraud detection methodology by using machine learning algorithms," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 15, no. 4, p. 3949, Aug. 2025. doi: 10.11591/ijece.v15i4.pp3949-3956.

[6] E. Agyepong, Y. Cherdantseva, P. Reinecke, and P. Burnap, "A systematic method for measuring the performance of a cyber security operations centre analyst," *Computers & Security*, vol. 124, p. 102959, Jan. 2023. doi: 10.1016/j.cose.2022.102959.

[7] H. Yaseen and A. Al-Amarneh, "Adoption of Artificial Intelligence-Driven Fraud Detection in Banking: The Role of Trust, Transparency, and Fairness Perception in Financial Institutions in the United Arab Emirates and Qatar," *Journal of Risk and Financial Management*, vol. 18, no. 4, p. 217, Apr. 2025. doi: 10.3390/jrfm18040217.

[8] C. D. R. Cevallos Ramos, F. F. Navarrete Chavez, F. R. Marquez Sanay, and M. P. Andrade Romero, "Strengthening cyber resilience in universities using artificial intelligence for proactive threat detectio," *Data and Metadata*, vol. 4, p. 1109, Jul. 2025. doi: 10.56294/dm20251109.

[9] J. K. Afriyie, K. Tawiah, W. A. Pels, S. Addai-Henne, H. A. Dwamena, E. O. Owiredu, S. A. Ayeh, and J. Eshun, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decision Analytics Journal*, vol. 6, p. 100163, Mar. 2023. doi: 10.1016/j.dajour.2023.100163.

[10] J. Forsberg and T. Frantti, "Technical performance metrics of a security operations center," *Computers & Security*, vol. 135, p. 103529, Dec. 2023. doi: 10.1016/j.cose.2023.103529.

[11] M. A. Umer, E. G. Belay, and L. B. Gouveia, "Fortifying Industry 4.0: Internet of Things Security in Cloud Manufacturing through Artificial Intelligence and Provenance Blockchain—A Thematic Literature Review," *Sci*, vol. 6, no. 3, p. 51, Sep. 2024. doi: 10.3390/sci6030051.

[12] M. Nandy and A. Dubey, "Applications of artificial intelligence to strengthening cyber security for threat detection and response," in *AIP Conference Proceedings*, Bangkok, Thailand, 2026, p. 020029. doi: 10.1063/5.0298670.

[13] L. Moura, A. Barcaui, and R. Payer, "AI and Financial Fraud Prevention: Mapping the Trends and Challenges Through a Bibliometric Lens," *Journal of Risk and Financial Management*, vol. 18, no. 6, p. 323, Jun. 2025. doi: 10.3390/jrfm18060323.

[14] A. Raj, V. Narayan, V. Muskan, A. Sani, P. Sharma, and S. S. Sarma, "Modern ransomware: Evolution, methodology, attack model, prevention and mitigation using multitiered approach," *SECURITY AND PRIVACY*, vol. 7, no. 6, p. e436, Nov. 2024. doi: 10.1002/spy2.436.

[15] G. Zioviris, K. Kolomvatsos, and G. Stamoulis, "An intelligent sequential fraud detection model based on deep learning," *The Journal of Supercomputing*, vol. 80, no. 10, pp. 14 824–14 847, Jul. 2024. doi: 10.1007/s11227-024-06030-y.

[16] G. Zogaj, F. Ismaili, E. Idrizi, and A. Luma, "Statistical Analysis of Unique Web Application Vulnerabilities: A Quantitative Assessment of Scanning Tool Efficiency," *SEEU Review*, vol. 20, no. 1, pp. 136–152, Jun. 2025. doi: 10.2478/seeur-2025-0021.

[17] P. Vanini, S. Rossi, E. Zvizdic, and T. Domenig, "Online payment fraud: from anomaly detection to risk management," *Financial Innovation*, vol. 9, no. 1, p. 66, Mar. 2023. doi: 10.1186/s40854-023-00470-w.

[18] E. Ukwandu, M. A. Ben-Farah, H. Hindy, M. Bures, R. Atkinson, C. Tachtatzis, I. Andonovic, and X. Bellekens, "Cyber-Security Challenges in Aviation Industry: A Review of Current and Future Trends," *Information*, vol. 13, no. 3, p. 146, Mar. 2022. doi: 10.3390/info13030146.

[19] N. Torres Gamarra and M. Zuniga Carnero, "Panorama actual de la ciberseguridad: amenazas, legislacion y brechas estructurales desde una revision sistematica," *Zenodo*, Jun. 2025. doi: 10.5281/ZENODO.15605545.

[20] E. K. Szczepaniuk and H. Szczepaniuk, "Cybersecurity of Smart Grids: Requirements, Threats, and Countermeasures," *Energies*, vol. 18, no. 18, p. 5017, Sep. 2025. doi: 10.3390/en18185017.

[21] H. Santillan, J. A. Arevalo Satan, and P. Wong, "A Comprehensive Analysis of Cybersecurity Infrastructure in Academic Environments," *Ingenieria*, vol. 35, no. 1, pp. 11–23, Oct. 2024. doi: 10.15517/ri.v35i1.60075.

[22] V. F. Rodrigues, L. M. Policarpo, D. E. Da Silveira, R. Da Rosa Righi, C. A. Da Costa, J. L. V. Barbosa, R. S. Antunes, R. Scorsatto, and T. Arcot, "Fraud detection and prevention in e-commerce: A systematic

literature review," *Electronic Commerce Research and Applications*, vol. 56, p. 101207, Nov. 2022. doi: 10.1016/j.elerap.2022.101207.

[23] P. Pittoli, P. David, and T. Noel, "Security architectures in constrained environments: A survey," *Ad Hoc Networks*, vol. 79, pp. 112–132, Oct. 2018. doi: 10.1016/j.adhoc.2018.06.001.

[24] S. L. Pauta Ayabaca and J. E. Alvarez Gavilanes, "Uso y necesidad de Tecnologias Emergentes en las empresas cuencanas para el fortalecimiento academico," *Pacha. Revista de Estudios Contemporaneos del Sur Global*, vol. 3, no. 9, p. e210123, Sep. 2022. doi: 10.46652/pacha.v3i9.123.

[25] J. R. Orosco-Fabian, "Ciberseguridad en educacion superior: una revision bibliometrica," *Revista Digital de Investigacion en Docencia Universitaria*, vol. 18, no. 2, p. e1933, Jul. 2024. doi: 10.19083/ridu.2024.1933.

[26] K. Schroeder, H. Trinh, and V. Y. Pillitteri, "Measurement guide for information security volume 1 : identifying and selecting measures," National Institute of Standards and Technology (U.S.), Gaithersburg, MD, Tech. Rep. NIST SP 800-55v1, Dec. 2024. doi: 10.6028/NIST.SP.800-55v1.

[27] Awobelem A. George, Akeem Olakunle Ogundipe, and Aminat Bolaji Bello, "Cybersecurity In healthcare systems: safeguarding electronic health records (EHRs) and medical devices against emerging cyber threats," *World Journal of Advanced Research and Reviews*, vol. 25, no. 2, pp. 2249–2262, Feb. 2025. doi: 10.30574/wjarr.2025.25.2.0592.

[28] W. Khan and N. Ebrahim, "ANOGAT-Sparse-TL: A hybrid framework combining sparsification and graph attention for anomaly detection in attributed networks using the optimized loss function incorporating the Twersky loss for improved robustness," *Knowledge-Based Systems*, vol. 311, p. 113144, Feb. 2025. doi: 10.1016/j.knosys.2025.113144.

[29] L. Hernandez Aros, L. X. Bustamante Molano, F. Gutierrez-Portela, J. J. Moreno Hernandez, and M. S. Rodriguez Barrero, "Financial fraud detection through the application of machine learning techniques: a literature review," *Humanities and Social Sciences Communications*, vol. 11, no. 1, p. 1130, Sep. 2024. doi: 10.1057/s41599-024-03606-0.

[30] E. A. Fueres Quilumbango, R. O. Andrade Paredes, and M. V. Yamba Yugsi, "Proteccion contra ataques de ingenieria social: analisis cualitativo de estrategias, tecnologias y patrones especificos," *Religacion*, vol. 10, no. 44, p. e2501310, Oct. 2024. doi: 10.46652/rgn.v10i44.1310.

[31] S. Chaudhary, V. Gkioulos, and S. Katsikas, "Developing metrics to assess the effectiveness of cybersecurity awareness program," *Journal of Cybersecurity*, vol. 8, no. 1, p. tyac006, Jan. 2022. doi: 10.1093/cybsec/tyac006.

[32] D. G. Rosado, L. E. Sanchez, A. J. Varela-Vaca, A. Santos-Olmo, M. T. Gomez-Lopez, R. M. Gasca, and E. Fernandez-Medina, "Enabling security risk assessment and management for business process models," *Journal of Information Security and Applications*, vol. 84, p. 103829, Aug. 2024. doi: 10.1016/j.jisa.2024.103829.

[33] K. Bennouk, N. Ait Aali, Y. El Bouzekri El Idrissi, B. Sebai, A. Z. Faroukhi, and D. Mahouachi, "A Comprehensive Review and Assessment of Cybersecurity Vulnerability Detection Methodologies," *Journal of Cybersecurity and Privacy*, vol. 4, no. 4, pp. 853–908, Oct. 2024. doi: 10.3390/jcp4040040.

[34] M. Alshomrani, A. Albeshri, A. A. Alsulami, and B. Alturki, "An Explainable Hybrid CNNTransformer Architecture for Visual Malware Classification," *Sensors*, vol. 25, no. 15, p. 4581, Jul. 2025. doi: 10.3390/s25154581.