STRATEGOS
REVISTA INTERNACIONAL DE ESTUDIOS EN CIENCIAS ADMINISTRATIVAS

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Análisis de riesgos en la seguridad de la información con la implementación de IA en los sistemas de atención al cliente

Risk analysis in information security with the implementation of AI in customer service systems

Karen Lissette Estacio Corozo[1] Y William Lenin Chenche Jácome [2]
[1]*Instituto Superior Tecnológico ARGOS, Guayaquil, Ecuador, k_estacio@tecnologicoargos.edu.ec.*
[2]*Universidad de Guayaquil, Guayaquil, Ecuador, william.chenchej@ug.edu.ec.*
(2024). Análisis de riesgos en la seguridad de la información con la implementación de IA en los sistemas de atención al cliente. *STRATEGOS Research Journal*, *4*(2), 1-19.

Recibido: 01 octubre 2024. Aceptación: 01 noviembre 2024. Publicado: 01 diciembre 2024.

Resumen
Esta investigación es un estudio documental descriptivo que realiza una revisión exhaustiva de la literatura, incluyendo artículos científicos, informes de la industria como los de ESET y estudios sobre la implementación de la inteligencia artificial (IA) en sistemas de atención al cliente. El objetivo principal es evaluar los riesgos y vulnerabilidades en la seguridad de la información que surgen con la implementación de la IA, identificar las principales amenazas y proponer medidas preventivas y correctivas para proteger los datos sensibles y garantizar la integridad de los sistemas.
Los hallazgos indican que, aunque la IA mejora significativamente la eficiencia y automatización en la atención al cliente, también introduce riesgos importantes como errores en los algoritmos, vulnerabilidades en la privacidad de los datos y un aumento en incidentes de seguridad como el robo de información y ataques de ransomware. Para mitigar estos riesgos, la literatura recomienda implementar medidas como el cifrado robusto de datos, la privacidad diferencial, la autenticación multifactor y el control de acceso basado en roles. Además, se enfatiza la importancia del

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

monitoreo continuo, la gestión de parches, auditorías de seguridad periódicas y procedimientos sólidos de copia de seguridad y recuperación, asegurando así una respuesta efectiva ante incidentes de seguridad y la protección de los datos sensibles.
Palabras clave: Riesgo, Seguridad de la Información, IA, Atención al Cliente.

Abstract
This research is a descriptive documentary study that conducts an exhaustive literature review, including scientific articles, industry reports such as those from ESET, and studies on the implementation of artificial intelligence (AI) in customer service systems. The main objective is to evaluate the risks and vulnerabilities in information security that arise with the implementation of AI, identify the primary threats, and propose preventive and corrective measures to protect sensitive data and ensure the integrity of the systems.
The findings indicate that although AI significantly improves efficiency and automation in customer service, it also introduces important risks such as errors in algorithms, vulnerabilities in data privacy, and an increase in security incidents like data theft and ransomware attacks. To mitigate these risks, the literature recommends implementing measures such as robust data encryption, differential privacy, multifactor authentication, and role-based access control. Additionally, the importance of continuous monitoring, patch management, periodic security audits, and robust backup and recovery procedures is emphasized, thereby ensuring an effective response to security incidents and the protection of sensitive data.
Keywords: Risk, Information Security, AI, Customer Service.

Introducción

La implementación de la Inteligencia Artificial (IA) en los sistemas de atención al cliente ha revolucionado la forma en que las empresas interactúan con los clientes, ofreciendo importantes ventajas en términos de automatización, eficiencia y personalización. Los sistemas impulsados por IA, como los chatbots, asistentes virtuales y algoritmos de aprendizaje automático, mejoran las experiencias de los clientes al proporcionar respuestas más rápidas y precisas, reducir los costos operativos y permitir la disponibilidad de servicio las 24 horas del día (Davenport & Ronanki, 2018). Sin embargo, junto con estos beneficios surgen varias vulnerabilidades, particularmente en la seguridad de la información, que plantean riesgos significativos durante la fase de implementación (Lawlor, 2021).

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

El aprendizaje automático (ML) junto con la IA pueden generar tanto valor como destruirlo en el contexto empresarial. Estas tecnologías prometen mejorar la eficiencia de los procesos, reducir costos y optimizar la toma de decisiones. No obstante, el valor generado depende de la adecuada gestión de los elementos involucrados y de las decisiones resultantes de los algoritmos. Si los algoritmos arrojan resultados incorrectos, estos errores pueden replicarse rápidamente, afectando la reputación de la empresa. Para cualquier sistema de IA, se identifican tres componentes clave iniciando por los datos de entrada, seguido de los algoritmos de procesamiento, incluyendo aprendizaje supervisado y no supervisado, y las decisiones de salida. La calidad y fiabilidad de cada uno de estos componentes es esencial para evitar la destrucción de valor (Canhoto y Clear, 2019).

Los sistemas de IA dependen en gran medida de grandes conjuntos de datos que a menudo contienen información confidencial de los clientes. El aumento de la sofisticación de los ciberataques, como el phishing, el ransomware y el malware basado en troyanos, convierte a estos sistemas en objetivos principales de explotación. Según el Informe de Ciberseguridad de ESET (2024), el 69% de las organizaciones en América Latina reportaron al menos un incidente de seguridad en el último año, siendo el robo de datos y el acceso no autorizado las principales preocupaciones. Dada esta realidad, es sumamente importante para las organizaciones evaluar los riesgos y vulnerabilidades que la IA introduce en los sistemas de atención al cliente, centrándose en las posibles amenazas a la integridad de los datos y proponiendo medidas para mitigar estos riesgos.

El objetivo de esta investigación es evaluar los riesgos y vulnerabilidades en la seguridad de la información que surgen durante la implementación de la IA en los sistemas de atención al cliente, identificar las principales amenazas y proponer medidas preventivas y correctivas para proteger los datos sensibles y garantizar la integridad de los sistemas en base a la revisión de la literatura.

METODOLOGÍA

Esta investigación es un estudio documental descriptivo basado en una revisión exhaustiva de la literatura. Se recopilaron datos de artículos científicos, informes de la industria y tendencias de ciberseguridad, incluidos informes de ESET y estudios sobre la implementación de IA. Se iniciará abordando conceptos teóricos de IA y en la sección resultados se expondrán los riesgos y

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

vulnerabilidades de la integración de IA en sistemas de atención al cliente finalizando con las recomendaciones para garantizar la seguridad de la información.

Historia y evolución de la inteligencia artificial

La IA, como campo de estudio, comenzó oficialmente en la década de 1950, con pioneros como John McCarthy, quien acuñó el término "inteligencia artificial" en 1956. Desde entonces, la IA ha evolucionado desde simples reglas programadas hacia modelos avanzados basados en el aprendizaje automático y el procesamiento de lenguaje natural. Los primeros sistemas de IA estaban limitados a tareas sencillas, pero con el tiempo, el aumento en la capacidad computacional y el acceso a grandes volúmenes de datos han permitido el desarrollo de sistemas complejos capaces de interactuar de manera efectiva con los clientes (Russell & Norvig, 2021).

IA en atención al cliente.

Durante los años 2000, las empresas comenzaron a integrar sistemas de IA en la atención al cliente mediante chatbots básicos. Sin embargo, fue en la última década cuando se produjo un cambio significativo, con la incorporación de algoritmos de aprendizaje profundo y modelos de lenguaje natural como GPT-3, que han revolucionado la capacidad de las máquinas para mantener conversaciones complejas (Goodfellow, Bengio, & Courville, 2016).

Conceptos básicos de inteligencia artificial

La IA se refiere a la capacidad de los sistemas informáticos para realizar tareas que normalmente requieren inteligencia humana, como el reconocimiento de voz, la toma de decisiones y la resolución de problemas.

Tipos de inteligencia artificial

IA débil o específica.

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Se centra en realizar una tarea específica, como la atención al cliente. Los sistemas de IA débil no poseen verdadera "inteligencia" en el sentido humano, sino que operan dentro de parámetros predefinidos (Russell & Norvig, 2021).

IA fuerte o general.

Aspira a tener una capacidad cognitiva similar a la humana, permitiendo la resolución de problemas en una variedad de contextos. Aunque este tipo de IA aún no se ha logrado completamente, representa un área activa de investigación.

Inteligencia artificial especializada en atención al cliente

La IA débil es la que ha sido más ampliamente implementada en la atención al cliente. Los sistemas especializados, como los chatbots, están diseñados para manejar un conjunto limitado de interacciones, como responder preguntas frecuentes o procesar solicitudes comunes (Sheehan, 2020). La IA en la atención al cliente utiliza varios algoritmos y modelos, como el procesamiento del lenguaje natural (PLN) y redes neuronales profundas. Estos modelos permiten a los sistemas comprender y generar lenguaje de manera más eficiente.

Procesamiento del lenguaje natural (PLN)

El PLN permite a las computadoras entender y generar lenguaje humano. Los chatbots y asistentes virtuales que emplean PLN pueden procesar consultas en lenguaje natural y responder de manera coherente. Modelos avanzados como GPT y BERT se han destacado por su capacidad para generar respuestas con un alto grado de precisión (Devlin et al., 2019).

Redes neuronales profundas

Las redes neuronales profundas, un subconjunto del aprendizaje automático, se utilizan para mejorar la comprensión del lenguaje y prever comportamientos de los clientes. Estas redes imitan la estructura del cerebro humano para procesar información y tomar decisiones en base a patrones previamente identificados (LeCun, Bengio, & Hinton, 2015).

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Métodos de aprendizaje en inteligencia artificial

Los sistemas de IA pueden aprender a través de varios métodos, que determinan su capacidad para mejorar con el tiempo. En la atención al cliente, los más utilizados son:

Aprendizaje supervisado

En este método, el sistema de IA es entrenado con un conjunto de datos etiquetados, donde el resultado deseado está previamente especificado. Esto permite que el modelo aprenda patrones y los aplique a nuevos datos (Goodfellow et al., 2016). Un ejemplo sería entrenar un chatbot con transcripciones de interacciones pasadas para mejorar su capacidad de respuesta.

Aprendizaje no supervisado

Aquí, la IA no tiene acceso a resultados etiquetados. En su lugar, el sistema debe identificar patrones ocultos en los datos sin la intervención humana. Este tipo de aprendizaje es útil para el análisis de sentimientos en tiempo real, donde las respuestas de los clientes se agrupan según patrones emocionales (Russell & Norvig, 2021).

Aprendizaje por refuerzo

Este método enseña a la IA a tomar decisiones en un entorno dinámico mediante recompensas o penalizaciones. En la atención al cliente, el aprendizaje por refuerzo puede ayudar a mejorar las respuestas de un chatbot al evaluar la satisfacción del cliente en tiempo real (Sutton & Barto, 2018).

Métodos de optimización en programas informáticos de IA

La optimización de los programas informáticos de IA es clave para mejorar su rendimiento, los métodos más comunes son optimización mediante descenso de gradiente y Algoritmos genéticos.

Optimización mediante descenso de gradiente

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Es un método iterativo utilizado para ajustar los parámetros del modelo de IA y minimizar el error. En los sistemas de atención al cliente, este enfoque ayuda a mejorar la precisión de las respuestas al entrenar modelos con grandes cantidades de datos (Goodfellow et al., 2016).

Algoritmos genéticos

Se inspiran en la evolución natural y son utilizados para resolver problemas de optimización. Se emplean para mejorar la eficiencia de los sistemas automatizados, ajustando continuamente las respuestas para maximizar la satisfacción del cliente.

Implementación de IA en atención al cliente

La implementación de IA en la atención al cliente ha sido exitosa en muchas empresas globales, como Amazon y Google. Estos estudios de caso destacan cómo la IA ha mejorado la eficiencia operativa y la experiencia del cliente.

Amazon ha implementado IA en su servicio al cliente mediante Alexa, un asistente virtual que interactúa con los clientes para responder preguntas y procesar solicitudes. Según McKinsey & Company (2021), Alexa ha logrado reducir en un 30% los tiempos de espera en atención al cliente, mejorando significativamente la satisfacción.

Google utiliza IA para predecir comportamientos futuros de los clientes y optimizar las interacciones. La empresa ha integrado algoritmos de aprendizaje automático en su servicio de atención, lo que le permite personalizar las respuestas y anticiparse a las necesidades de los clientes (Deloitte, 2022).

Ventajas del uso de IA en atención al cliente

- **Automatización de tareas repetitivas**: Los chatbots pueden manejar tareas simples y frecuentes, liberando tiempo para que los agentes humanos se concentren en problemas más complejos (Sheehan, 2020).

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

- **Disponibilidad 24/7**: La IA permite a las empresas ofrecer soporte continuo, algo que es particularmente beneficioso en mercados globales con diferentes zonas horarias (Gartner, 2021).

- **Personalización**: Mediante análisis predictivos, la IA puede adaptar las interacciones a las preferencias del cliente, mejorando la satisfacción (McKinsey & Company, 2021).

Desventajas del uso de IA en atención al cliente

- **Falta de empatía**: A pesar de que la IA ha avanzado en el procesamiento del lenguaje natural, todavía carece de la capacidad para comprender las emociones humanas con la precisión de un agente humano (Deloitte, 2022).

- **Problemas de seguridad**: Al manejar grandes volúmenes de datos personales, los sistemas de IA pueden ser vulnerables a brechas de seguridad si no se implementan correctamente (Salesforce, 2021).

Vulnerabilidades en la Implementación de la IA

La implementación de la IA en los sistemas de atención al cliente presenta varias vulnerabilidades, principalmente debido a la integración de procesamiento de datos a gran escala y capacidades de toma de decisiones autónomas (Davenport & Ronanki, 2018). Estas vulnerabilidades incluyen:

1. **Violaciones de la privacidad de los datos**: Los sistemas de IA a menudo acceden a datos personales, lo que aumenta el riesgo de violaciones de datos. Según Lawlor (2021), la creciente dependencia de la IA la ha convertido en un objetivo principal de los ciberataques.

2. **Acceso no autorizado**: Muchos sistemas de atención al cliente basados en IA operan en tiempo real y son vulnerables al acceso no autorizado. El informe de ESET

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

2024 mencionó que el acceso no autorizado a los sistemas fue la segunda preocupación más reportada entre las organizaciones latinoamericanas.

3. **Phishing y Ransomware**: Los sistemas de IA pueden ser explotados mediante ataques de spear-phishing y ransomware. En 2022, los ataques de ransomware en América Latina aumentaron un 26%, lo que destaca la importancia de defensas robustas contra estas amenazas (ESET, 2024).

En la Fig. 1 se analizan los desafíos que enfrentan las organizaciones al adoptar tecnologías de IA, con un enfoque en equilibrar proyectos ambiciosos y aplicaciones prácticas e incrementales.

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

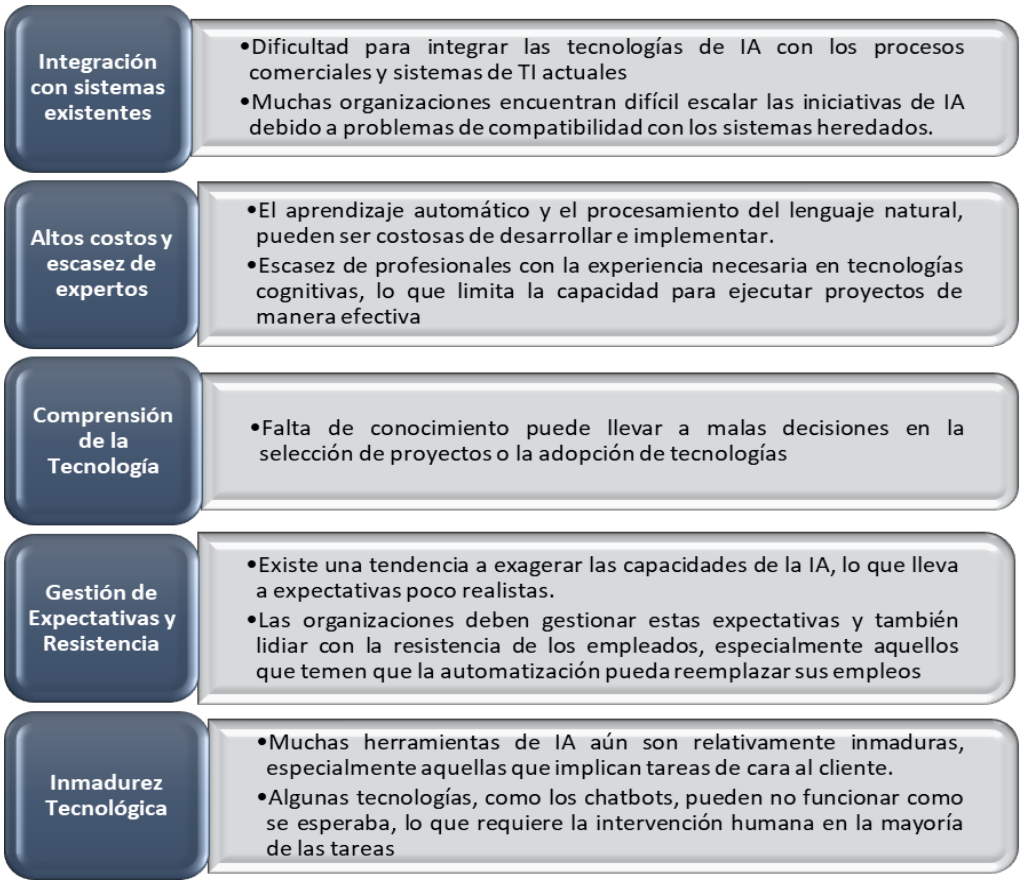

Fig. 1. Desafíos identificados en la adopción de IA

Fuente: Elaboración propia a partir de (Davenport & Ronanki, 2018)

RESULTADOS

Análisis Comparativo de la Literatura y Hallazgos

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

La comparación de la literatura (Tabla 1) muestra que, si bien la IA aporta importantes avances en el servicio al cliente, no se pueden ignorar los riesgos asociados a la seguridad de los datos. La dependencia de la IA hace que los sistemas sean más susceptibles a la explotación, y el aumento del número de ataques resalta la necesidad de medidas de seguridad más robustas.

Tabla 1. Resumen de Hallazgos Clave y Riesgos Identificados en la Implementación de IA en Atención al Cliente y Otros Sectores

Autor	Principales Hallazgos	Riesgo Identificado
Canhoto y Clear (2019)	La IA y ML mejoran la eficiencia, pero los errores pueden destruir valor.	Errores en la IA que dañan la calidad del servicio y la reputación
Davenport & Ronanki (2018)	Destacaron la eficiencia de la IA en la automatización del servicio al cliente, pero advirtieron sobre los riesgos de privacidad de los datos.	Privacidad de los datos y Acceso no autorizado
Informe de Ciberseguridad de ESET (2024)	Informó un aumento del 69% en incidentes de seguridad, particularmente relacionados con el robo de datos y el ransomware.	Robo de datos y Ransomware
Lawlor (2021)	Identificó el uso creciente de la IA en varios sectores, destacando su vulnerabilidad a las ciberamenazas.	Ciberataques a sistemas de IA
Pothumsetty (2020)	La IA está transformando las funciones financieras, como el asesoramiento financiero y la detección de fraudes.	Riesgos en la toma de decisiones incorrectas en trading automatizado y préstamos

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Raimundo & Rosário (2021)	La IA es clave para la seguridad de los sistemas de datos, pero presenta desafíos en la gestión de datos masivos y vigilancia.	Vulnerabilidades en ciberseguridad, como ataques basados en visión computacional y supervisión de redes.
Vanneschi et al. (2018)	La IA puede optimizar la atención al cliente, pero la dependencia excesiva puede generar respuestas ineficaces.	Fallos en la interacción humano-tecnología y pérdida de confianza del cliente
Zerfass et al. (2020)	Impacto significativo esperado en la gestión de la comunicación, desafíos en competencias y responsabilidades.	Desvalorización de habilidades humanas y responsabilidades confusas.

Fuente: Elaboración propia

Medidas preventivas y correctivas para la protección de datos sensibles en IA aplicada a la atención al cliente

La protección de los datos sensibles en los sistemas de inteligencia artificial (IA) integrados en la atención al cliente es esencial para mitigar riesgos asociados a brechas de seguridad y garantizar la integridad de los sistemas. Diversos estudios y normativas de seguridad han propuesto un conjunto de medidas que se dividen en acciones preventivas y correctivas. A continuación, se describen las principales medidas recomendadas por la literatura especializada.

Cifrado de datos

El uso de cifrado robusto para proteger los datos sensibles tanto en tránsito como en reposo. Los algoritmos de cifrado como AES (Advanced Encryption Standard) y RSA (Rivest-Shamir-Adleman) son ampliamente recomendados por organismos como el National Institute of Standards and Technology (NIST, 2020) y la International Organization for Standardization (ISO/IEC 27002, 2013). En caso de una vulneración de datos, la medida correctiva debe incluir la revocación de las

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

claves de cifrado comprometidas y la regeneración de nuevas claves para asegurar la continuidad del sistema sin exposición de datos.

Privacidad diferencial

Esta metodología añade "ruido" a los datos antes de utilizarlos en modelos de IA, lo que impide la identificación directa de los usuarios mientras se mantienen los patrones útiles para el entrenamiento. Google AI (2020) y Dwork et al. (2014) proponen la privacidad diferencial como una de las principales técnicas para salvaguardar la información sensible. Si se detecta una fuga de información, la medida correctiva incluiría reentrenar los modelos utilizando datos anonimizados.

Autenticación multifactor (MFA)

El uso de autenticación multifactor es otra medida preventiva clave para proteger los sistemas de IA que manejan datos sensibles. La combinación de contraseñas, tokens físicos o aplicaciones de autenticación añade una capa adicional de seguridad frente a accesos no autorizados (Microsoft, 2019; NIST, 2017). Si se compromete una cuenta o se detecta actividad sospechosa, la medida correctiva inmediata debe ser la deshabilitación del acceso a la cuenta comprometida y una investigación exhaustiva para identificar posibles fallas de seguridad.

Control de acceso basado en roles (RBAC)

Esta medida limita el acceso a los sistemas y modelos de IA únicamente a usuarios autorizados según su función dentro de la organización (Ferraiolo & Kuhn, 1992; NIST, 2014). Esto reduce las posibilidades de exposición de datos sensibles. En caso de un incidente de seguridad, se recomienda revisar y ajustar los permisos de acceso para prevenir futuros incidentes, retirando accesos innecesarios.

Monitoreo de sistemas en tiempo real

Es una medida preventiva fundamental para detectar actividades inusuales o no autorizadas en tiempo real. Herramientas de monitoreo de seguridad ayudan a identificar intentos de intrusión

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

o uso indebido de los sistemas (ISO/IEC 27001, 2013; ENISA, 2018). Si se detecta una violación, se deben revisar los registros de monitoreo para identificar el origen del ataque y aplicar las correcciones necesarias, como la desconexión de sistemas comprometidos.

Gestión de parches y actualizaciones

Mantener actualizados los sistemas de IA con parches y actualizaciones de seguridad regulares es una práctica preventiva crítica para evitar que vulnerabilidades conocidas sean explotadas por atacantes (CIS, 2021). Si se detecta un ataque que explote una vulnerabilidad ya identificada, se deben aplicar de inmediato los parches necesarios y realizar una auditoría completa del sistema para asegurar que no haya más puntos vulnerables.

Copia de seguridad y recuperación

Realizar copias de seguridad periódicas de los datos y modelos de IA garantiza que, en caso de un fallo del sistema o ataque, como ransomware, los datos puedan recuperarse sin pérdida significativa (ISO/IEC 27002, 2013; NIST, 2020). Ante una crisis de este tipo, la medida correctiva es restaurar el sistema desde las copias de seguridad más recientes y ajustar las políticas de respaldo para mejorar la frecuencia y la seguridad de las mismas.

Análisis forense en IA

Las recomendaciones de instituciones como el SANS Institute (2020) sugieren que, tras la identificación de la brecha, se deben aplicar correcciones en los modelos afectados y reentrenarlos si es necesario, para garantizar la integridad de los resultados futuros.

Auditorías de seguridad en sistemas de IA

La realización de auditorías de seguridad en los sistemas de IA que manejan datos sensibles es otra medida preventiva importante. Estas auditorías permiten identificar y corregir posibles vulnerabilidades antes de que sean explotadas por actores malintencionados (NIST, 2020; ENISA, 2018). En caso de un incidente, una auditoría detallada del sistema debe realizarse para ajustar tanto los modelos como las políticas de seguridad en función de los hallazgos.

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

CONCLUSIONES

La revisión de la literatura ha evidenciado que, aunque la implementación de la IA en los sistemas de atención al cliente conlleva significativos avances en eficiencia y automatización, también introduce una serie de riesgos y vulnerabilidades en la seguridad de la información que no deben ser subestimados. Los estudios analizados, destacan amenazas como errores en los algoritmos de IA, vulnerabilidades en la privacidad de los datos, y un aumento sustancial en incidentes de seguridad como el robo de información y ataques de ransomware. Además, la dependencia creciente de la IA incrementa la exposición a ciberataques específicos dirigidos a sistemas automatizados, lo que resalta la necesidad imperante de fortalecer las medidas de seguridad existentes.

Para mitigar estos riesgos, la literatura propone una serie de medidas preventivas y correctivas que son esenciales para proteger los datos sensibles y garantizar la integridad de los sistemas de IA en atención al cliente. Entre las estrategias más destacadas se encuentran el cifrado robusto de datos, la implementación de privacidad diferencial, la autenticación multifactor, y el control de acceso basado en roles. Estas acciones, combinadas con un monitoreo continuo de los sistemas, una gestión efectiva de parches y actualizaciones, y la realización de auditorías de seguridad periódicas, constituyen un marco integral para reducir la vulnerabilidad frente a amenazas cibernéticas. Además, la capacidad de realizar análisis forenses y contar con procedimientos de copia de seguridad y recuperación robustos permite una respuesta rápida y efectiva ante incidentes de seguridad, minimizando el impacto de posibles brechas.

REFERENCIAS

Canhoto, A. I., & Clear, F. (2019). *Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential.* Business Horizons, 63(2), 183-194.

CIS (2021). *CIS Controls v8*. Center for Internet Security. https://www.cisecurity.org/controls/

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*. Recuperado de https://store.hbr.org/product/artificial-intelligence-for-the-real-world/r1801h?sku=R1801H-PDF-ENG

Deloitte. (2022). *AI in customer service: Security challenges*. Deloitte Insights. https://www2.deloitte.com/

Deloitte. (2022). *Global AI adoption trends in customer service*. Deloitte Insights. https://www.deloitte.com/global/adoption-ai

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.04805

Dwork, C., Roth, A., Abowd, J., et al. (2014). *The Algorithmic Foundations of Differential Privacy*. Foundations and Trends® in Theoretical Computer Science, 9(3–4), 211-407.

ENISA. (2018). *Cybersecurity training programs: Development and delivery*. European Union Agency for Cybersecurity.

Ferraiolo, D. F., & Kuhn, D. R. (1992). Role-Based Access Controls. *National Computer Security Conference*.

Gartner. (2021). *AI in customer service: The state of implementation*. Gartner Research. https://www.gartner.com/en/research

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Google AI. (2020). *Privacy-preserving machine learning with differential privacy*. Google Research. https://ai.googleblog.com/

Grand View Research. (2023). *Customer service AI market forecast 2023-2028*. Grand View Research. https://www.grandviewresearch.com

Informe de Ciberseguridad de ESET. (2024). *Tendencias de ciberseguridad en América Latina*. Recuperado de https://web-assets.esetstatic.com/wls/es/articulos/reportes/cybersecurity-trends-2024-es.pdf

ISO/IEC 27001. (2013). *Information security management systems – Requirements*. International Organization for Standardization.

ISO/IEC 27002. (2013). *Code of practice for information security controls*. International Organization for Standardization.

Lawlor, B. (2021). Artificial intelligence and machine learning: Forging a new world for scholarly communication. *Chemistry International*, 8(1), 8-13. Recuperado de https://doi.org/10.3233/ISU-190068

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. https://doi.org/10.1038/nature14539

McKinsey & Company. (2021). *The power of AI in customer service: Case studies and insights*. McKinsey. https://www.mckinsey.com

Microsoft. (2019). *Best practices for securing AI systems*. Microsoft Security Documentation. https://docs.microsoft.com/

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

NIST. (2014). *Special Publication 800-53: Security and Privacy Controls for Federal Information Systems and Organizations*. National Institute of Standards and Technology.

NIST. (2017). *Special Publication 800-63: Digital Identity Guidelines*. National Institute of Standards and Technology.

NIST. (2020). *Framework for Improving Critical Infrastructure Cybersecurity*. National Institute of Standards and Technology.

OWASP. (2021). *Top Ten Vulnerabilities 2021*. Open Web Application Security Project.

Pothumsetty, R. (2020). *Implementation of artificial intelligence and machine learning in financial services.* International Research Journal of Engineering and Technology (IRJET), 7(3), 3186-3193.

Raimundo, R., & Rosário, A. (2021). *The impact of artificial intelligence on data system security: A literature review.* Sensors, 21(21), 7029.

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

Salesforce. (2021). *State of Service report 2021*. Salesforce Research. https://www.salesforce.com/research

SANS Institute. (2020). *Incident Handling and Response*. SANS Information Security Training.

Sheehan, M. (2020). *Chatbots and the future of customer service*. *Customer Experience Journal*, 15(2), 34-42. https://doi.org/10.1002/cxj.v15.2

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.

ISSN:1390-9568
e-ISSN:3073-1291
Diciembre 2024

Vanneschi, L., Bush, S., & Castelli, M. (2018). *Artificial intelligence and its impacts on industries.* Journal of AI and Society, 23(3), 145-162.

Zerfass, A., Hagelstein, J., & Tench, R. (2020). *Artificial intelligence in communication management: A cross-national study on adoption and knowledge, impact, challenges, and risks.* Journal of Communication Management, 24(3), 207-227.