

La _IA_ en el ejercicio de la auditoría

Ing. Nicolás Serrano, CISA, CISSP, CISM, CDPSE



Ministerio de Economía y Finanzas
Auditoría Interna de la Nación





Conceptos de IA 01

IA y Uruguay 02

Auditando la IA 03

AI + IA 04

Referencias 05

AGENDA



← // render.future >

01

Conceptos_ de IA_

Artificial Intelligence

Sistemas computacionales que realizan tareas asociadas con la inteligencia humana.

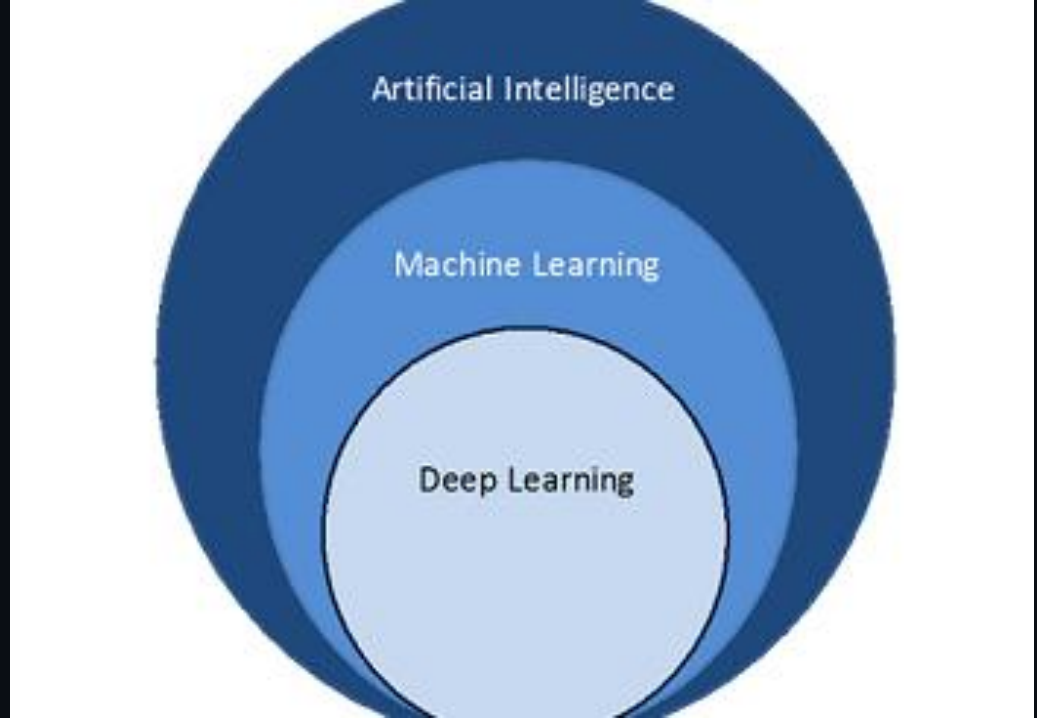
Machine Learning

Software que se autocorrigie tras sucesivas iteraciones usando resultados previamente obtenidos, sin necesidad de intervención humana. De un par a miles de características.

Deep Learning

Utiliza redes neuronales donde multiples capas de procesamiento extraen características de los datos. ChatGPT tiene cientos de billones de parámetros.

</create_with_intent>



<https://aws.amazon.com/what-is/deep-learning/>

Clasificación

APRENDIZAJE SUPERVISADO

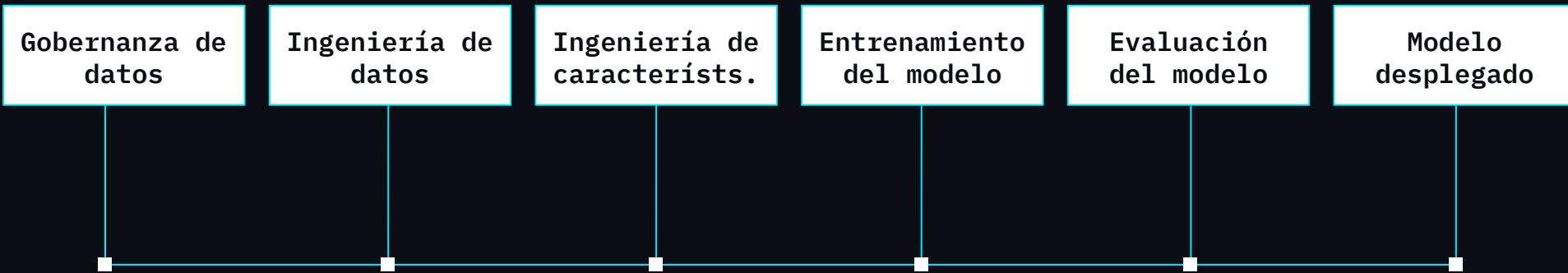
- Clasificación
- Regresión
- Híbridos

APRENDIZAJE NO SUPERVISADO

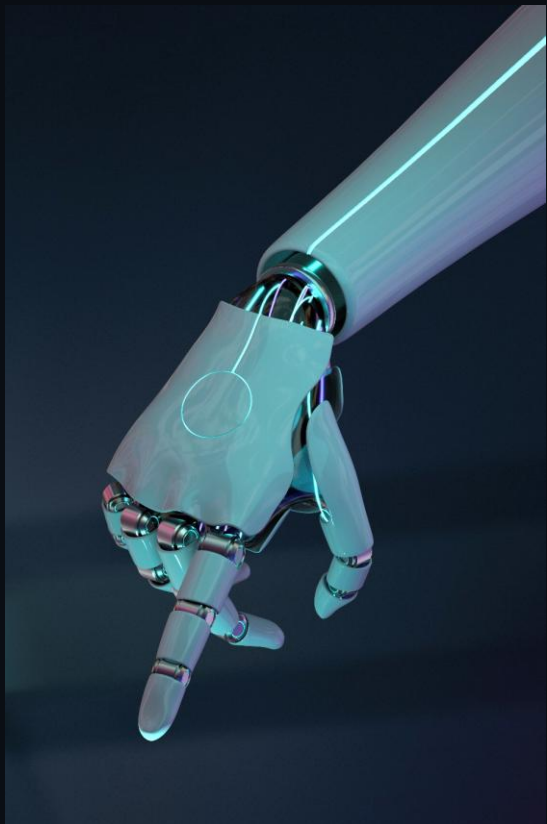
- Clustering
- Reducción de dimensiones
- Detección de anomalías
- Reglas de asociación

APRENDIZAJE POR REFUERZO

Categorías de Machine Learning



Fases en el desarrollo de un aplicaciones de ML



← // render.future >

02

IA_ y
Uruguay_

Puesto 47, 3ro en LATAM

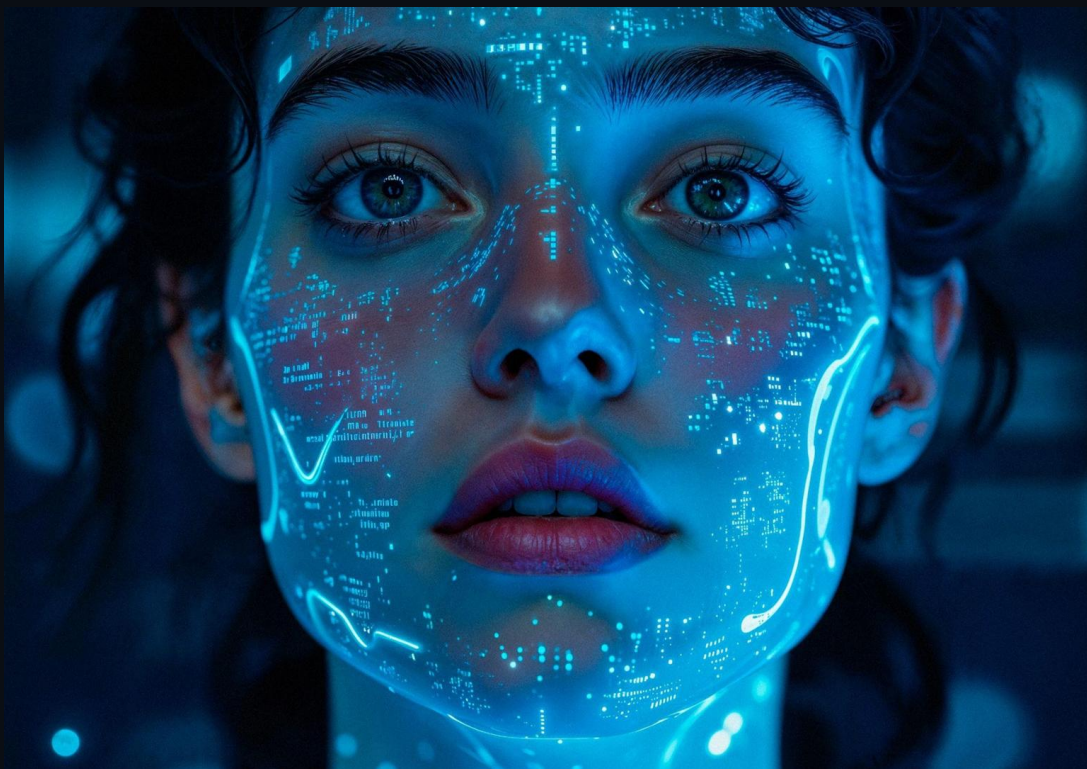
Índice de preparación de los gobiernos
para la IA, 2024, e ILIA 2025.

Estrategia Nacional 2024-2030

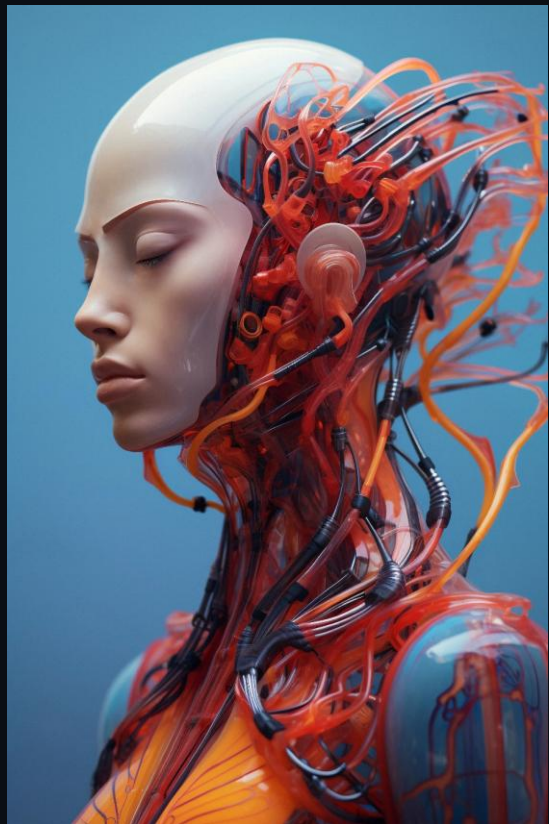
27 proyectos de IA

Sector público, fines del 2024.

</create_with_intent>



Algunas cifras...



← // render.future >

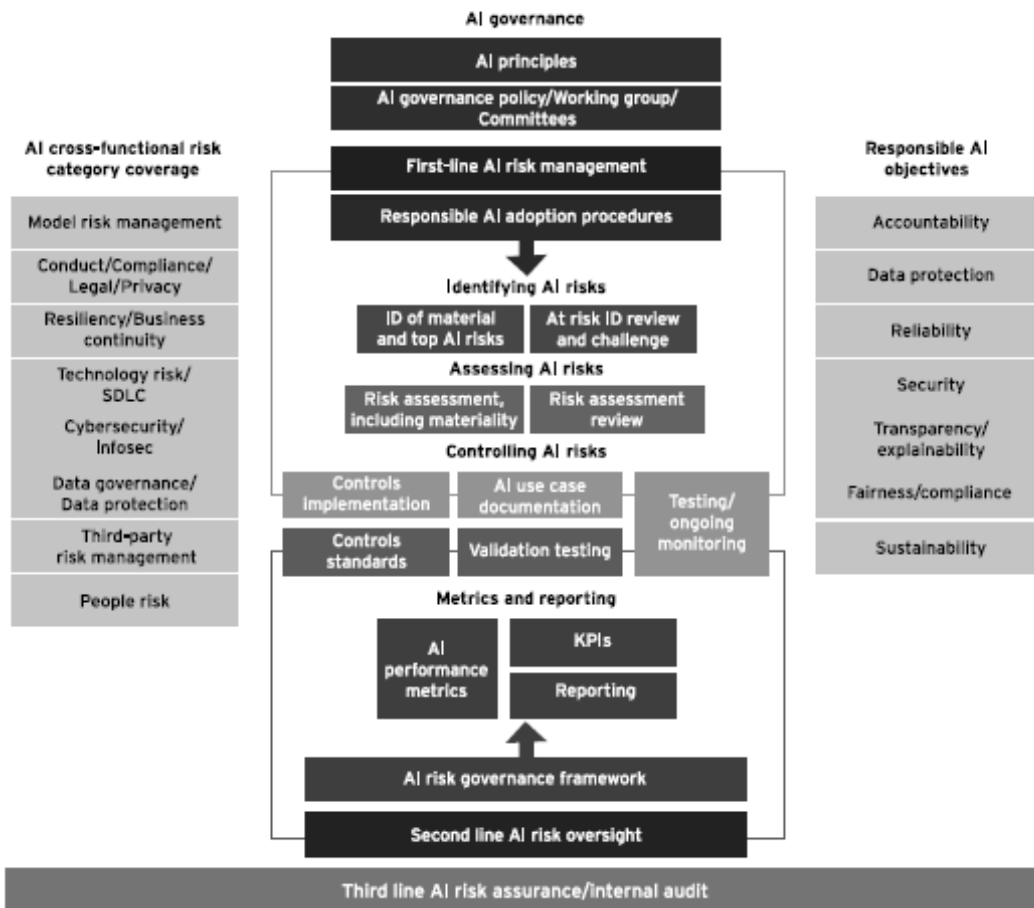
03

Auditando_ la IA_

¿Gobernanza de IA?

Board of Directors/Executive Leadership

Program sponsorship, guiding principles, industry standards/alliances



CHECKLIST FOR AI AUDITING – EDPB 1/23

Foco de la auditoría

Hacerlo sólo en aspectos técnicos dejaría muchos elementos sin auditar.

Marco de trabajo

Modelo – Sistema – Proceso.

Alcance del documento

Modelos de machine learning.



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

IDENTIFICATION AND TRANSPARENCY OF THE AI-BASED COMPONENT:

Inventory of the audited AI-based component

Is the AI-based component identified in the documentation by means of a name or code, identification of version and date of creation?

Identification of responsibilities

Is there an identification about the person(s) or institution(s) who manage the life cycle stages of the AI-based component?

Transparency

Are data sources documented?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

PURPOSE OF THE AI-BASED COMPONENT:

Identification of intended purposes and uses

Is the intended purpose of the AI-based component documented both in quantitative and qualitative terms?

Definition of the intended context of the AI-based component

Are there any legal, social, economic, organizational, technical, scientific or other contexts identified related to the inclusion of the AI-based component in the processing? Are they documented?

Analysis of proportionality and necessity

Has the use of the AI component been assessed against other possible options from an approach focusing on the rights and freedoms of data subjects?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

PURPOSE OF THE AI-BASED COMPONENT:

Definition of the potential recipients of data

Is the information obligation to data subjects identified regarding data processing arising from the inclusion of the AI-based component?

Limitation of data storage

Have appropriate technical and organizational measures and criteria been defined to storage personal data?

Analysis of categories of data subjects

Are the categories of data subjects affected by the development of the AI component and its use in the framework of the intended processing identified?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

BASES OF THE AI COMPONENT:

Identification of the AI-based component development policy

Do the documents with development policies of products and systems consider the data protection policy?

Involvement of the Data Protection Officer

Does the DPO have the necessary professional qualifications and, particularly, the legal and technical expertise, as well as data protection practice appropriate to the project?

Adjustment of basic theoretical models

Has an analysis been carried out regarding the theoretical framework and previous similar experience on which the development of the AI component is based?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

BASES OF THE AI COMPONENT:

Appropriateness of the methodological framework

Are the metrics for measuring the behaviour of the AI component duly selected and measured?

Identification of the basic architecture of the AI-based component

Is there documentation which assure that, when programming AI-based components, the coding principles, codes and coding, best practices applied are followed in order to guarantee that the code is readable, secure, low-maintenance and robust?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

DATA MANAGEMENT:

Data quality assurance

Is there a documented procedure to manage and ensure proper data governance, which allows to verify and provide guarantees of the accuracy, integrity, accuracy, veracity, update and adequacy...?

Definition of the origin of the data sources

Are legal grounds to used personal data in the different stages of the AI-based component life cycle identified?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

DATA MANAGEMENT:

Preprocessing of personal data

Are data cleaning techniques and best practices used in the data cleansing process properly selected and documented?

Bias control

Is there a procedure to assess the need to have additional data for improving precision or removing any possible bias?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

VERIFICATION AND VALIDATION:

Adapting the verification and validation process of the AI based component

Is there documentation that duly describe the verification and validation process, ...?

Verification and validation of the AI-based component

Is white-box testing at code and implementation levels included in the testing plan?

Performance

Are metrics or sets of aggregated metrics used to determine the precision, accuracy, sensitivity or other performance parameters of the relevant component in consideration of the principle of data accuracy established?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

VERIFICATION AND VALIDATION:

Consistency

Has a threshold been established to determine when an obtained result deviates from the expected result based on identical or similar data inputs?

Stability and robustness

...are the factors, whose variation may impact the properties of the AI component and may establish the need to manage its readjustment, identified?



← </redefine_logic>

CHECKLIST FOR AI AUDITING – EDPB 1/23

VERIFICATION AND VALIDATION:

Traceability

Is there a version control system in place for all elements of the AI-based component: used datasets, AI-based component code, libraries used and any other element associated with the component?

Security

Has a risk analysis developed with regard to the risks for rights and freedoms of persons? Have the results of this risk analysis been used to determine the security and privacy requirements...?



← </redefine_logic>

ICO AI AUDITS

POSIBLES ALCANCES:

Gobernanza

Contratos y tercerizaciones

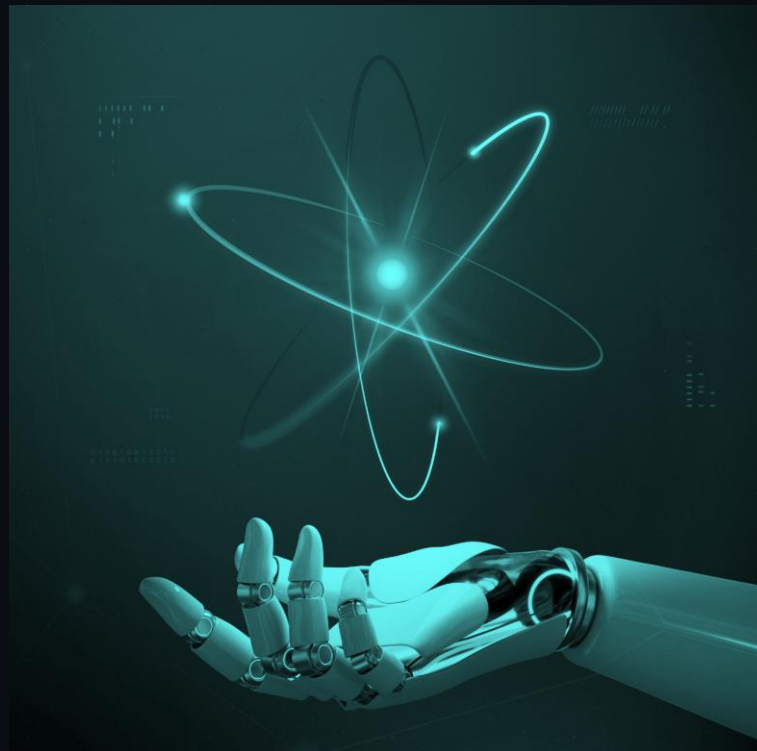
Entrenamiento

Protección de datos

Base legal

Costo – beneficio / oportunidad - riesgos

Exactitud estadística



← </redefine_logic>

ICO AI AUDITS

POSIBLES ALCANCES:

Discriminación y sesgos

Seguridad e integridad

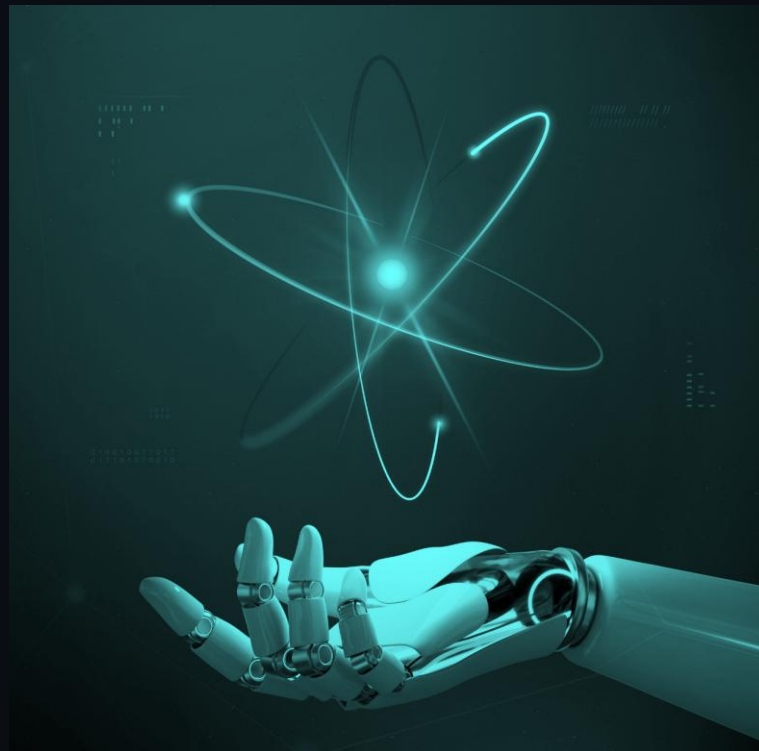
Transparencia

Minimización de datos

Derechos de las personas

Revisión humana

← </redefine_logic>



Si bien la IA promete transformar servicios, decisiones y procesos en el sector público, su adopción es desigual y plagada de barreras.

Taxonomía centrada en los “desafíos datos-IA” estructurada en tres dimensiones principales: tecnológica, organizativa y ambiental.

Pasar de “¿esta IA es responsable?” a “¿está la organización preparada para adoptar IA responsablemente?”.

Muchos fracasos no se deben sólo a la tecnología, sino a ausencias en datos, infraestructura, cultura organizativa o gobernanza.

Nikiforova, A., Lnenicka, M., Melin, U., Valle-Cruz, D., Gill, A., Casiano Flores, C., Sirait, E., Luterek, M., Dreyling, R. M., and Tesarova, B. (2025).

**Responsible AI Adoption in the Public Sector:
A Data-Centric Taxonomy of AI Adoption Challenges.**

In Proceedings of the 59th Hawaii International Conference on System Sciences

TOE dimension: Technology

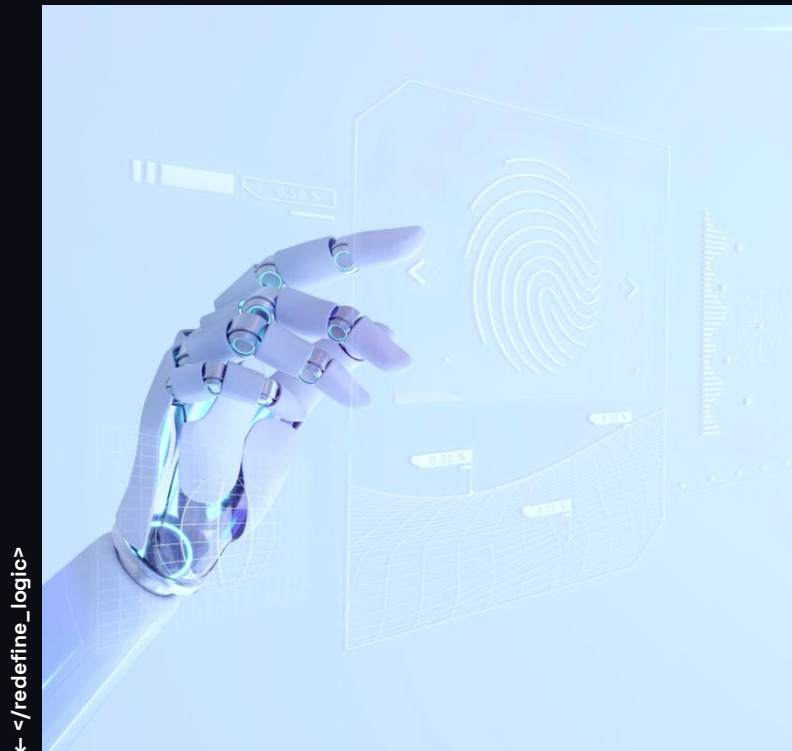
Availability of data sources

Computing and data storage resources

Data and AI systems, platforms, tools, and services

AI-ready data

Algorithm transparency and standardization



← </redefine_logic>

TOE dimension: Organization

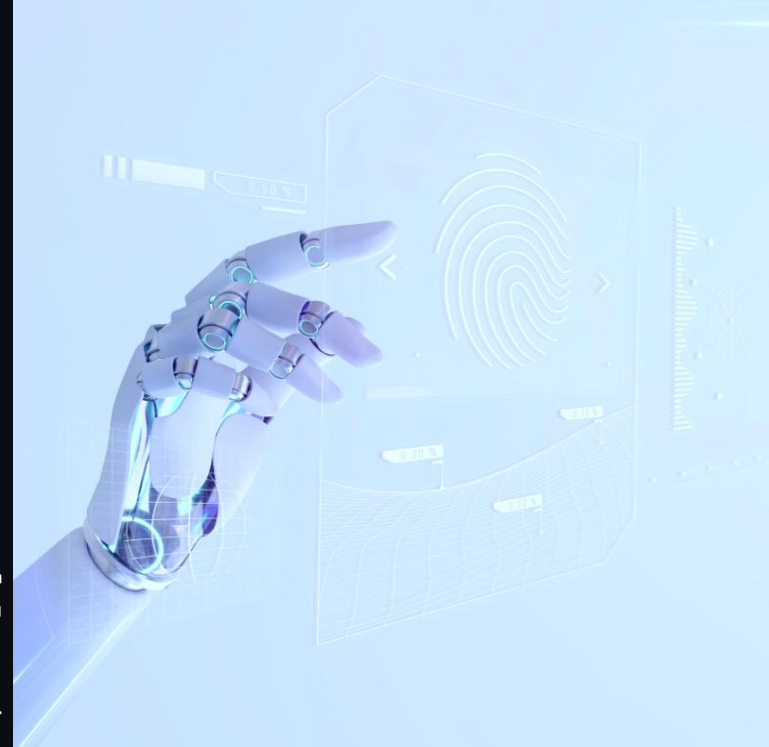
Governance and management of data, AI systems, and PDEs

Actions and Interactions between humans, data, and AI

Impacts on decision-making and use

Stakeholders, data, and AI literacy

← </redefine_logic>



TOE dimension: Environment

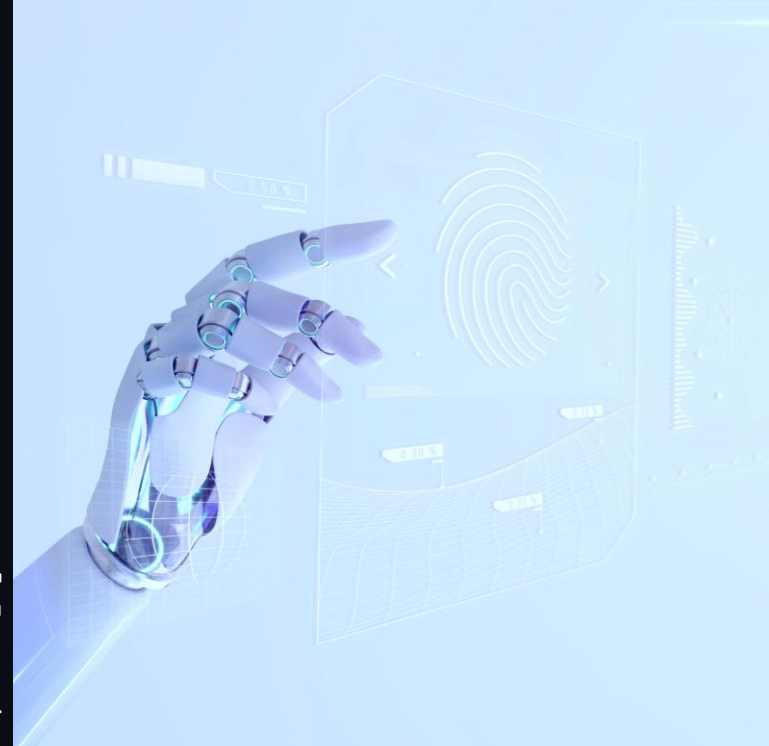
Data collaboration and exchange

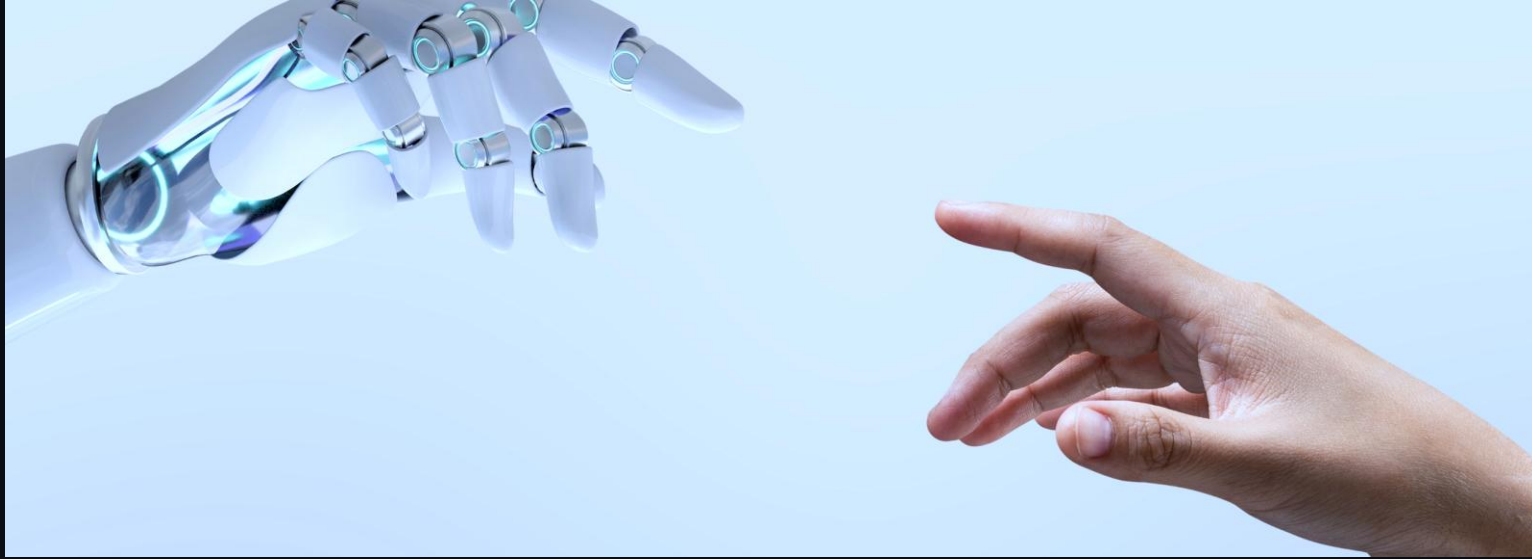
Data and AI security, privacy, and ethics

Economic sustainability

Environmental sustainability

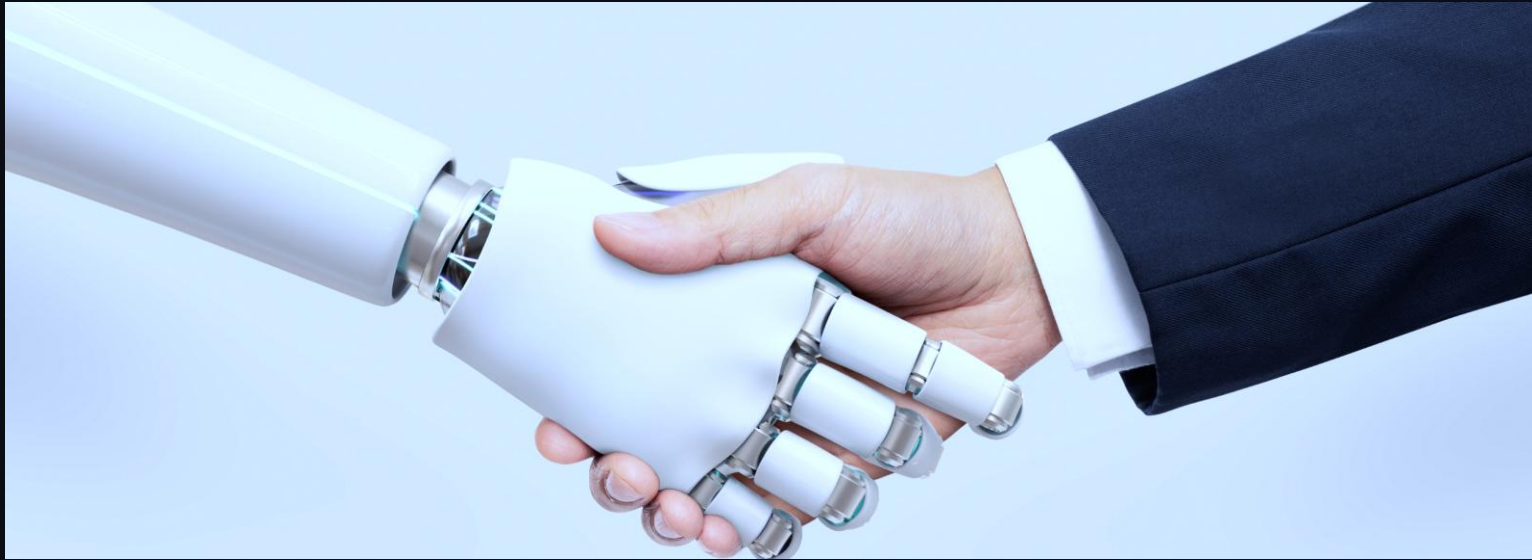
← </redefine_logic>





IA Generativa = Democratización de IA

Esto genera más presión para su uso dentro de las organizaciones, apurando consecuentemente a los auditores.



</innovation_pipeline>

IA Act - EU

Paraguas y catalizador para estándares/guías de auditorías.

1

Artificial Intelligence Systems with Unacceptable Risk (Art. 5)

Prohibited

- Manipulation of behavior, opinions, and human decisions.
- Classification of people based on their social behavior.
- Mass biometric identification remotely and in real-time, with certain exceptions.

EXAMPLE: Social scoring

2

High-Risk Artificial Intelligence Systems (HRAIS, Art. 6)

Allowed if the requirements of the AI for the ex-ante conformity assessment are met.

- Key Aspects of the Regulation (Annex III).
- Common regimes already subject to harmonized EU standard.
- Additional list to be reviewed annually by the EAIB (Art. 84).

Example: Hiring

3

Artificial Intelligence Systems with Specific Transparency Obligations (Art. 52)

Allowed but subject to information/transparency obligations.

- Human interaction.
- Use to detect emotions or determine categories based on biometric data.
- Generation of manipulated content.

Example: Personification (bots)

4

Artificial Intelligence Systems with No or Minimal Risk

Allowed without restrictions.

Example: Predictive maintenance

OWASP Machine Learning Security Top Ten

[Main](#)[Charter](#)[Related](#)[Glossary](#)[owasp](#) [incubator](#) [License](#) [CC BY-SA 4.0](#)

Important Information

The current version of this work is in draft and is being modified frequently. Please refer to the [project wiki](#) for information on how to contribute and project release timelines.

Overview

Welcome to the repository for the OWASP Machine Learning Security Top 10 project! The primary aim of the OWASP Machine Learning Security Top 10 project is to deliver an overview of the top 10 security issues of machine learning systems. More information on the project scope and target audience is available in our [project working group charter](#)

Top 10 Machine Learning Security Risks

- ML01:2023 Input Manipulation Attack
- ML02:2023 Data Poisoning Attack
- ML03:2023 Model Inversion Attack
- ML04:2023 Membership Inference Attack
- ML05:2023 Model Theft
- ML06:2023 AI Supply Chain Attacks
- ML07:2023 Transfer Learning Attack
- ML08:2023 Model Skewing
- ML09:2023 Output Integrity Attack
- ML10:2023 Model Poisoning

OWASP Top 10 for Large Language Model Applications

[Main](#)[Example](#)

About This Repository

This is the repository for the **OWASP Top 10 for Large Language Model Applications**. However, this project has now grown into the comprehensive **OWASP GenAI Security Project** - a global initiative that encompasses multiple security initiatives beyond just the Top 10 list.

OWASP GenAI Security Project

The OWASP GenAI Security Project is a global, open-source initiative dedicated to identifying, mitigating, and documenting security and safety risks associated with generative AI technologies, including large language models (LLMs), agentic AI systems, and AI-driven applications. Our mission is to empower organizations, security professionals, AI practitioners, and policymakers with comprehensive, actionable guidance and tools to ensure the secure development, deployment, and governance of generative AI systems.

Learn more about our mission and charter: [Project Mission and Charter](#)

Visit our main project site: genai.owasp.org

LLM01: Prompt Injection

Manipulating LLMs via crafted inputs can lead to unauthorized access, data breaches, and compromised decision-making.

LLM02: Insecure Output Handling

Neglecting to validate LLM outputs may lead to downstream security exploits, including code execution that compromises systems and exposes data.

LLM03: Training Data Poisoning

Tampered training data can impair LLM models leading to responses that may compromise security, accuracy, or ethical behavior.

LLM04: Model Denial of Service

Overloading LLMs with resource-heavy operations can cause service disruptions and increased costs.

LLM05: Supply Chain Vulnerabilities

Depending upon compromised components, services or datasets undermine system integrity, causing data breaches and system failures.

LLM06: Sensitive Information Disclosure

Failure to protect against disclosure of sensitive information in LLM outputs can result in legal consequences or a loss of competitive advantage.

LLM07: Insecure Plugin Design

LLM plugins processing untrusted inputs and having insufficient access control risk severe exploits like remote code execution.

LLM08: Excessive Agency

Granting LLMs unchecked autonomy to take action can lead to unintended consequences, jeopardizing reliability, privacy, and trust.

LLM09: Overreliance

Failing to critically assess LLM outputs can lead to compromised decision making, security vulnerabilities, and legal liabilities.

LLM10: Model Theft

Unauthorized access to proprietary large language models risks theft, competitive advantage, and dissemination of sensitive information.

¿Dónde poner el foco como auditores?

En las estructuras de control y gobernanza que se han definido en la organización **para los desarrollos o uso de IA**, y determinar que se encuentran operando efectivamente.

¿Cuáles son algunos de los desafíos para el auditor de IA?

- Regulaciones o marcos de auditoría específicos para IA inmaduros.
 - Precedentes limitados para casos de uso de IA.
 - Definiciones y taxonomías poco claras para IA.
 - Gran variedad de sistemas y soluciones de IA.
 - IA por naturaleza es una tecnología emergente.
 - Falta de guías explícitas para auditar IA.
 - Falta de prioridades estratégicas.
 - Curva de aprendizaje compleja para el auditor de IA.
- Riesgo generado por la tercerizaciones que son necesarias en IA.

**INVESTIGAR + APRENDER +
INVOLUCRAR INTERESADOS +
ASESORARLOS + ADAPTAR O
CREAR MARCOS DE
REFERENCIA**

Seguimiento a la auditoría es difícil para IA inmaduros.
Precedentes limitados para casos de uso de IA.
Definiciones y taxonomías poco claras para IA.
Gran variedad de sistemas y soluciones de IA.
IA por naturaleza es una tecnología emergente.
Falta de guías explícitas para auditar IA.
Falta de prioridades estratégicas.
Curva de aprendizaje compleja para el auditor de IA.
Riesgo generado por la tercerizaciones que son necesarias en IA.

¿Dónde podemos agregar valor al auditar la IA?

Protección de datos

Concientizar y velar por el cumplimiento normativo en lo relativo a protección de datos y demás datos sensibles.

Debido uso de recursos económicos

Observar que los desarrollos de IA no impliquen un mal uso de los fondos públicos, ya sea por proyectos inviables o por malas ejecuciones de los mismos.

Resultados confiables

Contribuir a que la toma de decisiones sea a partir de información confiable, sin perjuicios por sesgos o datos de mala calidad.

Independencia

Mirada independiente de aspectos normativos, tecnológicos, de datos, financieros, de ciberseguridad, etc.

Postura de la organización

Demostrar que la organización es madura y que aplica la innovación de manera responsable con la debida diligencia.

Adaptabilidad de los auditores

Dejar en claro que como auditores estamos en formación constante y preparados para las nuevas tecnologías.



¿Dónde podemos agregar valor al auditar la IA?

Protección de datos

Concientizar y velar por el cumplimiento normativo en lo relativo a protección de datos y demás datos sensibles.

Debido uso de recursos económicos

Observar que los desarrollos de IA no impliquen un mal uso de los fondos públicos, ya sea por proyectos inviables o por malas ejecuciones de los mismos.

Resultados confiables

Contribuir a que la toma de decisiones sea a partir de información confiable, sin perjuicios por sesgos o datos de mala calidad.

Independencia

Mirada independiente de aspectos normativos, tecnológicos, de datos, financieros, de ciberseguridad, etc.

Postura de la organización

Demostrar que la organización es madura y que aplica la innovación de manera responsable con la debida diligencia.

Adaptabilidad de los auditores

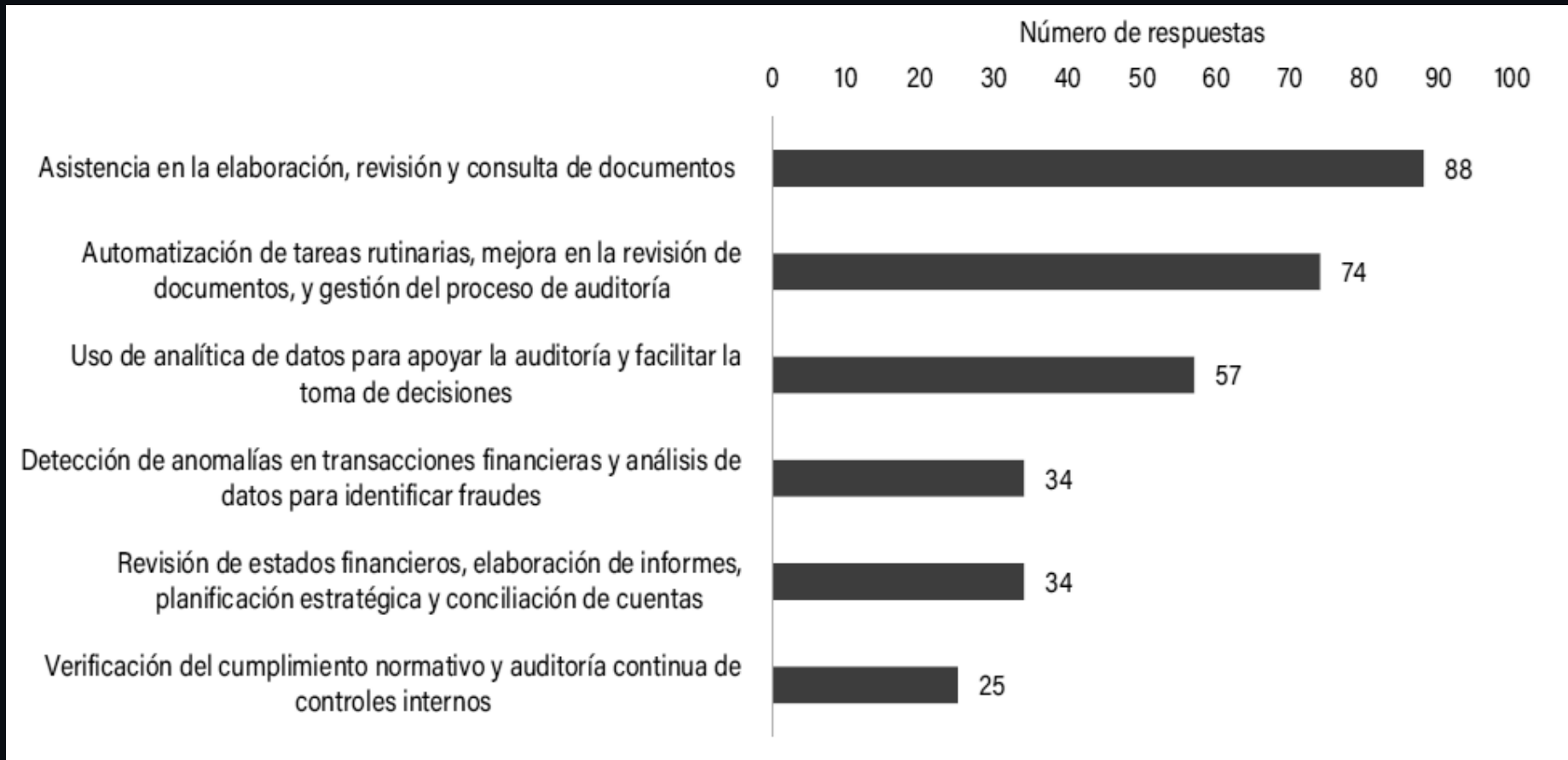
Dejar en claro que como auditores estamos en formación constante y preparados para las nuevas tecnologías.



← // render.future >

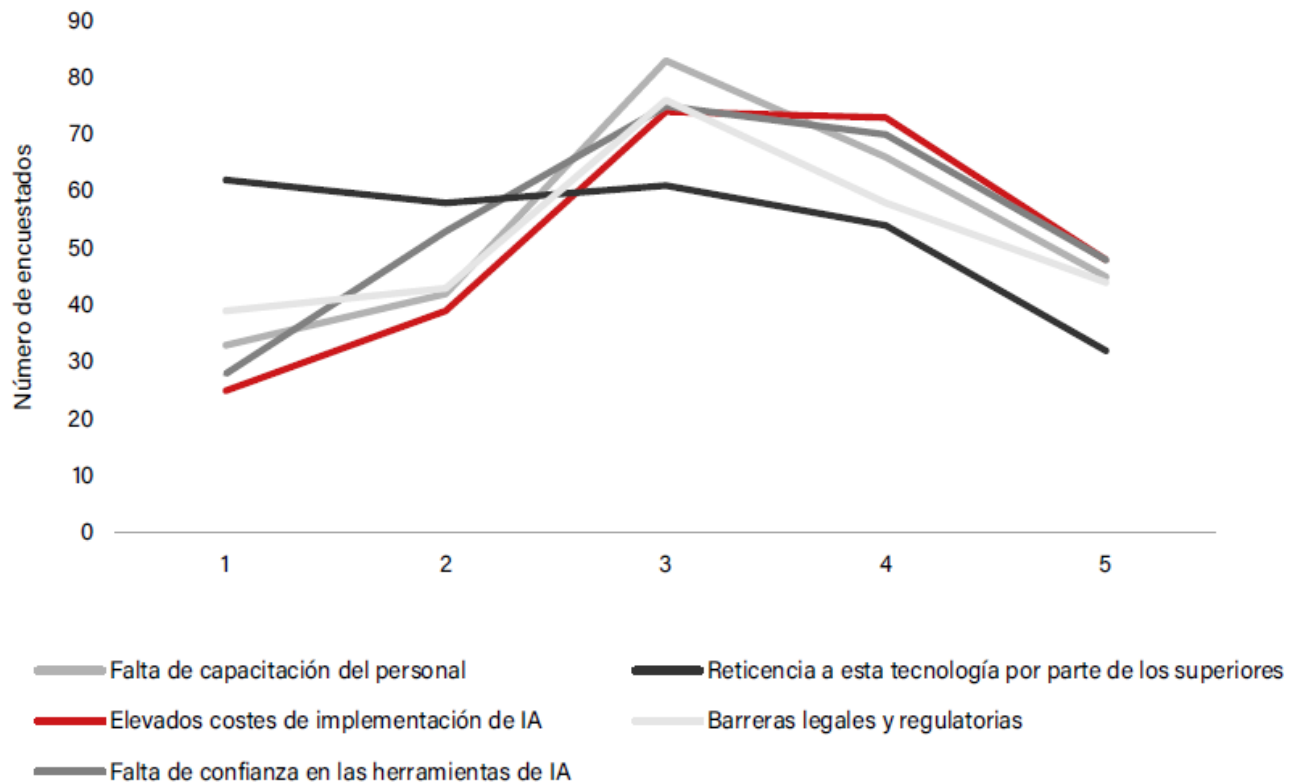
04

AI + IA_



USOS PARA EQUIPOS DE AUDITORÍA

//// GRÁFICO 25 Obstáculos para implementar la IA en la firma



USOS PARA EQUIPOS DE AUDITORÍA

Consideraciones críticas

Confiar, pero verificar

Resultados semanticamente correctos, pero con errores conceptuales.

Considerar los requisitos de cumplimiento existentes

Velar por que toda normativa aplicable se esté cumpliendo.

Revisar las políticas existentes

O crearlas si no existen, sobre los usos aceptables de la IA en nuestras organizaciones.

Adaptar las políticas y programas de ciberseguridad y privacidad existentes

Utilizar enfoques de seguridad y privacidad por diseño.

Promover la alfabetización en IA

Capacitar y educar a los empleados sobre tecnologías de IA y sus riesgos.

Consideraciones críticas

Realizar un análisis de costos

Evaluar el costo de implementar sistemas de IA junto con los posibles ahorros que estos sistemas puedan generar.

Considerar los impactos sociales

Pérdida y desplazamiento de empleos. Uso masivo. ¿Este documento es auténtico o lo generaron con chatGPT?

Establecer auditorías y trazabilidad

Evitar en casa de herrero, cuchillo de palo...

Designar un responsable de IA

Alguien a cargo del enfoque y seguimiento de las herramientas de IA en uso en la organización.

Desarrollar directrices éticas de IA

Documentadas y compartidas en toda la organización.

UNDERSTANDING THE EU AI
ACT: REQUIREMENTS AND
NEXT STEPS. ISACA, 2024



← // render.future >

05

Referencias_

Referencias

AUDITING ARTIFICIAL INTELLIGENCE

ISACA, 2018

ÍNDICE LATINOAMERICANO DE IA 2025

CENIA, CEPAL

GOVERNMENT AI READINESS INDEX 2024

OXFORD INSIGHTS

A GUIDE TO ICO AUDIT ARTIFICIAL INTELLIGENCE (AI) AUDITS

Information Commissioner's Office, ICO, UK

AI AUDITING CHECKLIST FOR AI AUDITING

European Data Protection Board, 2023

AUDIT PRACTITIONER'S GUIDE TO MACHINE LEARNING

ISACA, 2022

Referencias

EL RETO DE LA INTELIGENCIA ARTIFICIAL PARA LA AUDITORÍA

ICAC-ASEPUC 2024

INTERNAL AUDIT OF AI APPLIED TO BUSINESS PROCESSES

Thought Factory, The IIA–Spain. 2024

UNDERSTANDING THE EU AI ACT: REQUIREMENTS AND NEXT STEPS

ISACA, 2024

> Muchas gracias

Ing. Nicolás Serrano,
CISA, CISSP, CISM, CDPSE



Ministerio de Economía y Finanzas
Auditoría Interna de la Nación

CREDITS: This presentation template was created by **Slidesgo**, and includes icons, infographics & images by **Freepik**