certiprof®

AI RISK MANAGER

PROFESSIONAL CERTIFICATION

AIRMPC™ Versión 032024

AI Risk Manager Professional Certification (AIRMPC)

Basada en NIST Artificial Intelligence Risk Management Framework

AIRMPC™ Versión 032024

¿Quién es Certiprof®?

Certiprof® es una entidad certificadora fundada en los Estados Unidos en 2015, ubicada actualmente en Sunrise, Florida.

Nuestra filosofía se basa en la creación de conocimiento en comunidad y para ello su red colaborativa está conformada por:

- **Nuestros Lifelong Learners (LLL)** se identifican como Aprendices Continuos, lo que demuestra su compromiso inquebrantable con el aprendizaje permanente, que es de vital importancia en el mundo digital en constante cambio y expansión de hoy. Independientemente de si ganan o no el examen.
- Las universidades, centros de formación, y facilitadores en todo el mundo forman parte de nuestra red de aliados **ATPs (Authorized Training Partners.)**
- **Los autores (co-creadores)** son expertos de la industria o practicantes que, con su conocimiento, desarrollan contenidos para la creación de nuevas certificaciones que respondan a las necesidades de la industria.
- **Personal Interno:** Nuestro equipo distribuido con operaciones en India, Brasil, Colombia y Estados Unidos está a cargo de superar obstáculos, encontrar soluciones y entregar resultados excepcionales.

Nuestras Afiliaciones

Memberships

Digital badges issued by

IT Certification Council – ITCC

Certiprof® es un miembro activo de ITCC.
Una de las ventajas de hacer parte del ITCC es como líderes del sector colaboran entre sí en un formato abierto para explorar maneras nuevas o diferentes formas de hacer negocios que inspiran y fomentan la innovación, estableciendo y compartiendo buenas prácticas que nos permiten extender ese conocimiento a nuestra comunidad.

Certiprof ha contribuido a la elaboración de documentos blancos en el Career Path Ways Taskforce, un grupo de trabajo que se implementó internamente para ofrecer a los estudiantes la oportunidad de saber qué camino tomar después de una certificación.

Algunos de los miembros del ITCC

- **IBM**
- **CISCO**
- **ADOBE**
- **AWS**
- **SAP**
- **GOOGLE**
- **ISACA**

itcc
IT CERTIFICATION COUNCIL

Agile Alliance

Certiprof® es un miembro corporativo de Agile Alliance.

Al unirnos al programa corporativo Agile Alliance, continuamos empoderando a las personas ayudándolas a alcanzar su potencial a través de la educación. Cada día, brindamos más herramientas y recursos que permiten a nuestros socios formar profesionales que buscan mejorar su desarrollo profesional y sus habilidades.

https://www.agilealliance.org/organizations/certiprof/

Credly

Esta alianza permite que las personas y empresas certificadas con Certiprof® cuenten con una distinción a nivel mundial a través de un distintivo digital.

Credly es el emisor de insignias más importante del mundo y empresas líderes en tecnología como IBM, Microsoft, PMI,Nokia, la Universidad de Stanford, entre otras, emiten sus insignias con Credly.

Empresas que emiten insignias de validación de conocimiento con Credly:

- **IBM**
- **Microsoft**
- **PMI**
- **Universidad de Stanford**
- **Certiprof**

Insignias Digitales

Según el estudio del IT Certification Council (ITCC), años atrás, la gente sabía muy poco sobre las insignias digitales. Hoy, grandes empresas e instituciones educativas de todo el mundo expiden insignias.

Las insignias digitales contienen metadatos detallados sobre quién las ha obtenido, las competencias requeridas y la organización que las ha expedido. Algunas insignias incluso están vinculadas a las actividades necesarias para obtenerlas.

Para las empresas e instituciones educativas, las insignias y la información que proporcionan son tan importantes que muchas decisiones, como las de contratación o admisión, se basan en los datos que aportan.

¿Por qué son importantes?

- **Facilidad de Compartir y Verificar Logros:**

  Las insignias digitales permiten a los profesionales mostrar y verificar sus logros de manera instantánea y global. Según un informe de Credly, **los perfiles de LinkedIn con insignias digitales reciben un 40% más de atención por parte de reclutadores y empleadores.**

- **Visibilidad en Plataformas Digitales:**

  En una encuesta realizada por Pearson y Credly, el **85%** de los usuarios que obtuvieron insignias digitales **las compartieron en LinkedIn,** y el **75%** reportó que esto mejoró su **credibilidad profesional en sus redes.** Además, el **76%** de los empleadores encuestados afirmó que las insignias digitales les ayudan a identificar rápidamente habilidades específicas.

¿Por qué son importantes?

- **Impacto en la Contratación:**

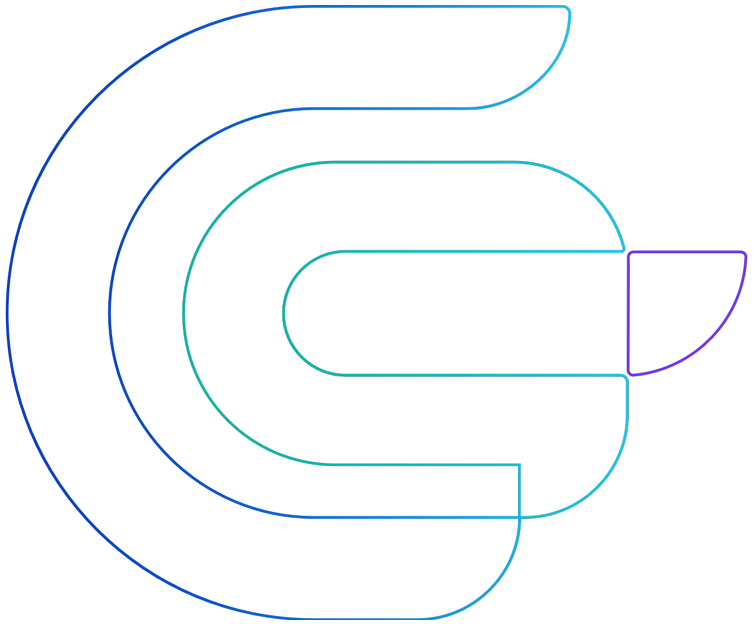
  Un estudio de la **Asociación Internacional de Gestión de Proyectos (PMI)** encontró que los candidatos que muestran insignias digitales de gestión de proyectos tienen **un 60%** más de probabilidades de ser contratados en comparación con aquellos que solo mencionan sus habilidades sin verificación digital.

¿Por qué son importantes?

- **Empoderamiento de la Marca Personal:**

La visibilidad y verificación instantánea proporcionada por las insignias digitales permiten a los profesionales no solo demostrar sus habilidades, sino también construir una marca personal fuerte. Según un estudio de LinkedIn, los profesionales que utilizan insignias digitales tienen un 24% más de probabilidades de avanzar en sus carreras.

La certificación y las insignias digitales no son solo una validación del conocimiento, sino también una herramienta poderosa para la mejora continua y la empleabilidad. En un mundo donde el aprendizaje permanente se ha convertido en la norma, estas credenciales son clave para el desarrollo profesional y la competitividad en el mercado laboral global.



No todas las insignias son iguales, y en **Certiprof,** estamos comprometidos con ofrecerte más que un simple reconocimiento digital. Al obtener una insignia emitida por certiprof, estarás recibiendo una validación de tu conocimiento respaldada por una de las entidades líderes en certificación profesional a nivel mundial.

Da el siguiente paso y obtén la insignia que te abrirá puertas y te posicionará como un experto en tu campo.

¿Por qué es importante obtener su certificado?

- **Prueba de experiencia:** Su certificado es un reconocimiento formal de las habilidades y conocimientos que ha adquirido. Sirve como prueba verificable de sus cualificaciones y demuestra su compromiso con la excelencia en su campo.

- **Credibilidad y reconocimiento:** En el competitivo mercado laboral actual, las empresas y los compañeros valoran las credenciales que le distinguen de los demás. Un certificado de una institución reconocida, como Certiprof, proporciona credibilidad instantánea e impulsa su reputación profesional.

- **Avance profesional:** Tener tu certificado puede abrirte las puertas a nuevas oportunidades. Ya se trate de un ascenso, un aumento de sueldo o un nuevo puesto de trabajo, las certificaciones son un factor diferenciador clave que los empleadores tienen en cuenta a la hora de evaluar a los candidatos.

¿Por qué es importante obtener su certificado?

- **Oportunidades de establecer contactos:** Poseer un certificado le conecta con una red de profesionales certificados. Muchas organizaciones cuentan con grupos de antiguos alumnos o de trabajo en red en los que puede compartir experiencias, intercambiar ideas y ampliar su círculo profesional.

- **Logro personal:** Obtener una certificación es un logro importante, y su certificado es un recordatorio tangible del trabajo duro, la dedicación y el progreso que ha realizado. Es algo de lo que puede sentirse orgulloso y mostrar a los demás.

Insignia

AI Risk Management Professional Certification™

Issued by Certiprof

Holders of the Professional Certificate in AI Risk Management, based on the NIST AI Risk Management Framework, develop skills in identifying, assessing, and mitigating the risks associated with AI technologies, ensuring that they align with ethical and social values for professionals involved in the development, deployment, or management of AI systems.

Certification | $ Paid

Skills

AI Risk | Continuous Risk Management | Critical Thinking | Ethical and Social Principles | NIST Framework | Problem-Solving | Risk Assessment | Risk Identification | Teamwork

Earning Criteria

Badge holders have passed a multiple choice exam, scoring a minimum of 32 out of 40 (80%). The exam is a closed-book exam lasting 60 minutes.

https://www.credly.com/org/certiprof/badge/ai-risk-management-professional-certification.1

Aprendizaje Permanente

- Certiprof ha creado una insignia especial para reconocer a los aprendices constantes.
- Para el 2024, se han emitido más de 1,000,000 de estas insignias en más de 11 idiomas.

Propósito y Filosofía

- Esta insignia está destinada a personas que creen firmemente en que la educación puede cambiar vidas y transformar el mundo.
- La filosofía detrás de la insignia es promover el compromiso con el aprendizaje continuo a lo largo de la vida.

Acceso y Obtención de la Insignia

- La insignia de Lifelong Learning se entrega sin costo a aquellos que se identifican con este enfoque de aprendizaje.
- Cualquier persona que se considere un aprendiz constante puede reclamar su insignia visitando:

https://certiprof.com/pages/certiprof-lifelong-learning

Introducción

- Programa de Certificación CertiProf para el Marco NIST de Gestión de Riesgos de IA (AI RMF)

- Desarrollado por CertiProf, este programa integral lo prepara para identificar y mitigar los riesgos asociados a la Inteligencia Artificial (IA).

- Componentes Clave:

  - **Framing Risk:** Aprende a comprender y abordar los riesgos, impactos y daños potenciales de la IA.
  - **Audiencia y Ciclo de Vida de la IA**: Identifica a los actores involucrados en el desarrollo y uso de la IA, y las etapas de este ciclo.
  - **Riesgos de la IA y Confiabilidad:** Descubre cómo la confiabilidad y la gestión de riesgos están interrelacionadas para minimizar los impactos negativos de la IA.
  - **Efectividad del AI RMF**: Conoce los beneficios que este marco ofrece a los usuarios.
  - **Núcleo del AI RMF** (Govern, Map, Measure, Manage): Aprende las 4 funciones clave para gestionar el riesgo y desarrollar sistemas de IA confiables.
  - **Perfiles del AI RMF**: Descubre cómo se implementan las funciones del marco en escenarios específicos.

- *Fuente: https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF*

Objetivo del Programa de Certificación del NIST Artificial Intelligence Risk Manager Framework (AIRMF)

- Comprender los Principios Fundamentales del AI RMF
- Implementar la Gobernanza de IA
- Entender Métodos de Evaluación y Mitigación de Riesgos
- Promover la Transparencia, Justicia y Seguridad de IA
- Mejora Continua y Aprendizaje

Contenido

- **Parte 1: Información Fundamental**
  - Atributos clave del Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF)
  - Enmarcando el Riesgo
  - Audiencia
  - Riesgos de IA y Confiabilidad
  - Efectividad del AI RMF

- **Parte 2: Núcleo y Perfiles**
  - Núcleo del AI RMF
  - Perfiles del AI RMF

- **Parte 3: Informacion Anexa al AI RFM**
  - Descripciones de Tareas de Actores de IA
  - Cómo los Riesgos de IA Difieren de los Riesgos de Software Tradicionales
  - Gestión de Riesgos de IA e Interacción Humano-IA

Bibliografía

- National Institute of Standards and Technology (NIST). (2021). Artificial Intelligence Risk Management Framework (AI RMF). https://www.nist.gov/AI/RMF
- Doe, J. (2019). Managing AI Risks: A New Framework for the Age of AI. 2nd ed. TechPress.

"CertiProf adopta el Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF) desarrollado por el NIST como fundamento de este programa. El objetivo es ampliar la divulgación de las prácticas recomendadas por el gobierno de los Estados Unidos a la comunidad global, fomentando así la adopción de estrategias de gestión de riesgos en IA que sean sólidas, eficaces y pertinentes a nivel mundial".

Ismael Ramirez,

Certiprof®

Audiencia

- El público objetivo de la certificación AI Risk Management Professional Certification (AIRMPC), incluye:
  - **Profesionales de TI y Ciberseguridad:** Que buscan comprender los riesgos asociados con los sistemas de IA y cómo gestionarlos.
  - **Gerentes de Proyecto de IA y Líderes de Equipo**: Que necesitan incorporar prácticas de gestión de riesgos en sus proyectos de IA.
  - **Desarrolladores de IA y Científicos de Datos**: Que deseen aplicar prácticas de desarrollo seguras y éticas en la creación de modelos de IA.
  - **Ejecutivos y Directores de Organizaciones:** Interesados en supervisar la integración de la IA en sus estrategias de negocio asegurando la conformidad y mitigación de riesgos.
  - **Auditores y Consultores de Riesgos de IA:** Profesionales que evalúan sistemas de IA y asesoran sobre la mejora de la gestión de riesgos de IA.
  - **Responsables de Cumplimiento y Regulación:** Que deben asegurar que los sistemas de IA cumplan con las regulaciones actuales y futuras.

Certificación Prompt Engineering Foundation

FREE

Beneficios

- Mejora en la interacción con herramientas de IA generativa, capacidad para escribir prompts efectivos, maximización del potencial de herramientas como ChatGPT, Gemini y Copilot.

Habilidades Desarrolladas

- Comprensión de la IA generativa, habilidades en generación de prompts, desarrollo y uso de chatbots, aplicación de técnicas de PE en contextos reales, optimización del uso de herramientas de IA.

https://certiprof.com/collections/agile/products/design-thinking-professional-certificate-dtpc

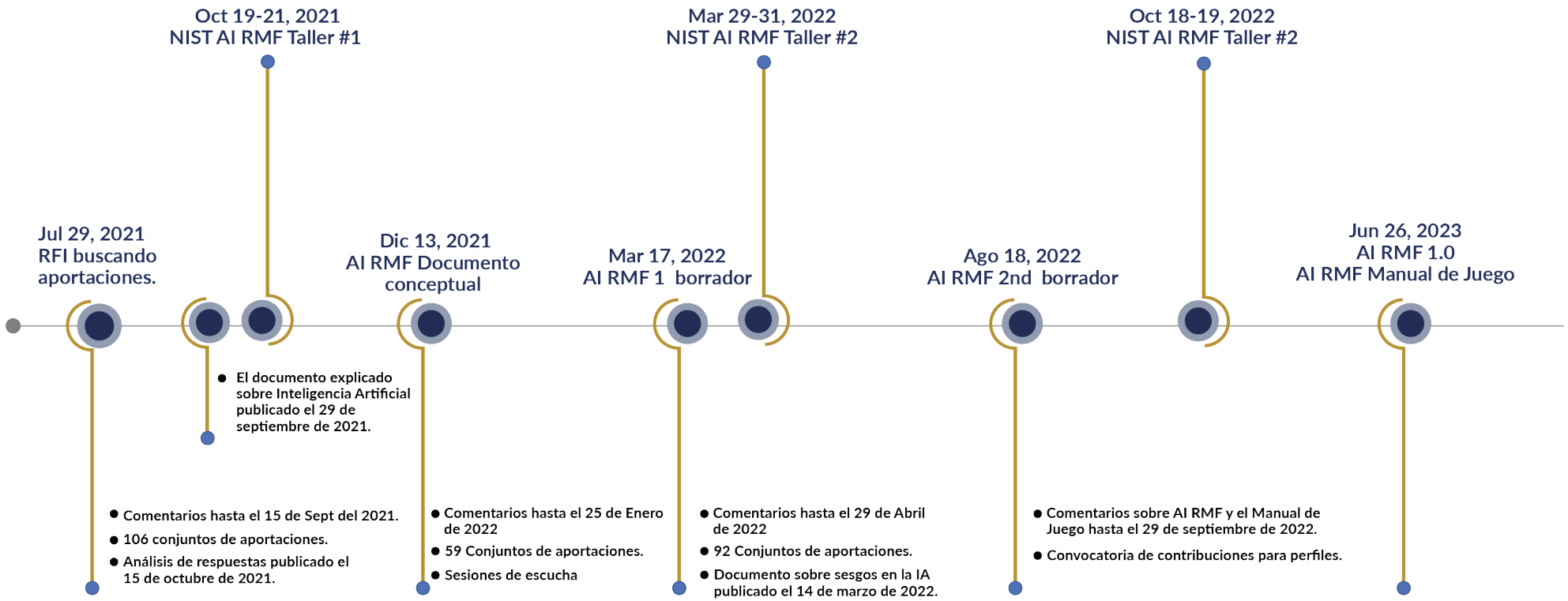
COMPARTE Y VERIFICA TUS LOGROS DE APRENDIZAJE FÁCILMENTE

#AIRMPC #certiprof

certiprof®

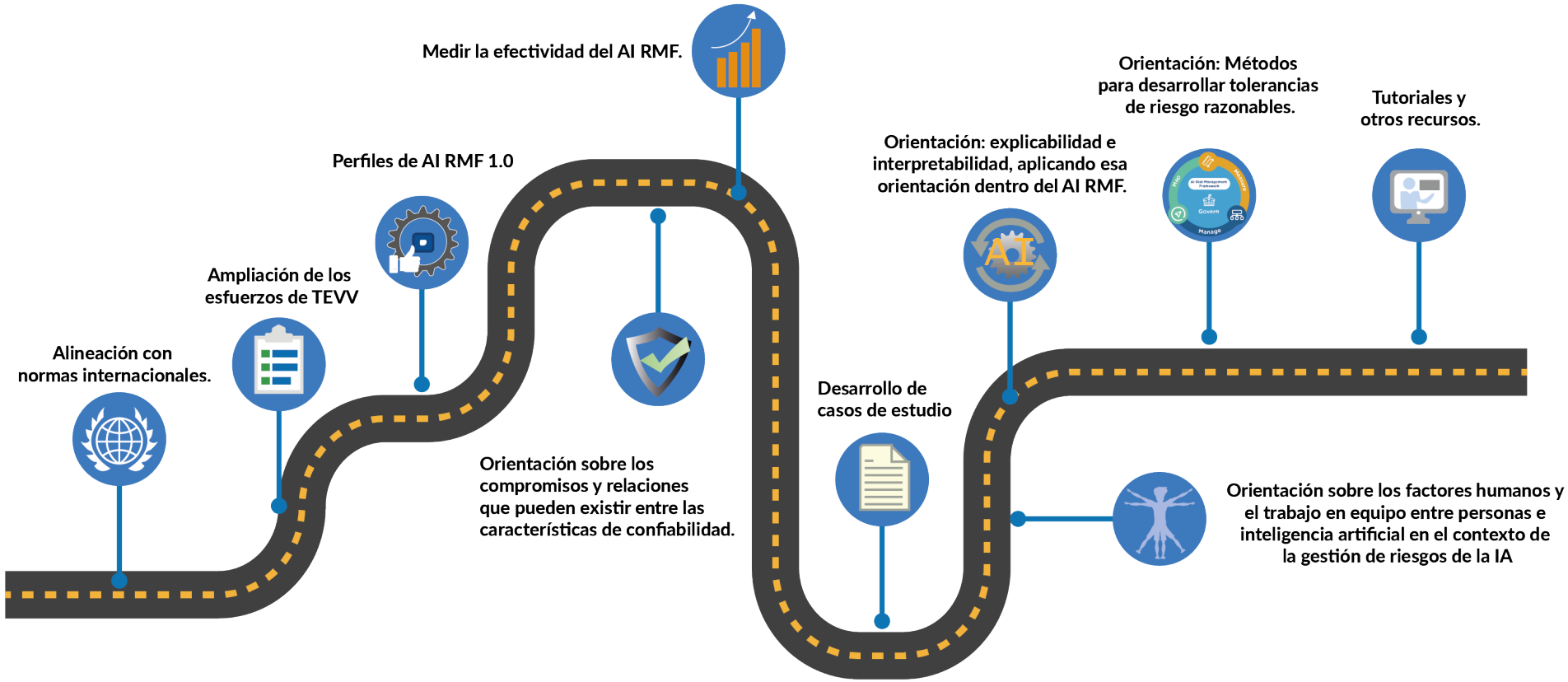

Map
Measure
AI Risk Management Framework
Govern
Manage

Línea de tiempo y participaciones del AI RMF.



Fuente: https://www.nist.gov/itl/ai-risk-management-framework

Desarrollo AI RMF



Fuente: https://www.nist.gov/itl/ai-risk-management-framework/roadmap-nist-artificial-intelligence-risk-management-framework-ai

¿QUÉ ES EL AI RMF?

Recurso voluntario para organizaciones que diseñan, desarrollan, implementan o utilizan sistemas de IA para gestionar riesgos de IA y promover una IA confiable y responsable.

Introducción a AI RMF

- Las tecnologías de inteligencia artificial (IA) tienen el potencial significativo de transformar la sociedad y la vida de las personas en múltiples sectores como el comercio, la salud, el transporte, la ciberseguridad, el medio ambiente y nuestro planeta.

Potencial de la IA:

- La IA puede impulsar el crecimiento económico inclusivo y apoyar avances científicos que mejoren las condiciones de nuestro mundo, promoviendo soluciones innovadoras a desafíos globales.

Riesgos Asociados con la IA:

- Sin embargo, las tecnologías de IA también presentan riesgos que pueden impactar negativamente a individuos, grupos, organizaciones, comunidades, la sociedad, el medio ambiente y el planeta.
- Estos riesgos, al igual que los riesgos de otros tipos de tecnología, pueden surgir de diversas maneras y caracterizarse como de largo o corto plazo, de alta o baja probabilidad, sistémicos o localizados, y de alto o bajo impacto.

Introducción al Marco de Gestión de Riesgos de IA (AI RMF)

- El AI RMF está diseñado para ser práctico, adaptarse al panorama cambiante de la IA a medida que las tecnologías continúan desarrollándose y ser operacionalizado por organizaciones en diversos grados y capacidades.

Atributos clave del Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF)

- **Comunicación Efectiva de Riesgos de IA**
  - Uso de lenguaje claro y comprensible.
  - Facilita la comunicación de riesgos de IA a través de diferentes niveles organizacionales y con el público.
- **Fomento de un Lenguaje Común**
  - Proporciona taxonomía, terminología, definiciones, métricas, y caracterizaciones para el riesgo de IA.
- **Usabilidad y Compatibilidad**
  - Fácilmente utilizable y coherente con otros aspectos de la gestión de riesgos.
  - Adaptable a estrategias y procesos más amplios de gestión de riesgos.
- **Aplicabilidad Universal**
  - Útil para una amplia gama de perspectivas, sectores, y dominios tecnológicos.
- **Enfoque en Resultados**
  - Ofrece un catálogo de resultados y enfoques sin prescribir requisitos únicos.

Atributos clave del Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF)

- **Mejora y Conciencia de Estándares**
  - Aprovecha y aumenta la conciencia de estándares, mejores prácticas y herramientas existentes.
- **Neutralidad Legal y Regulatoria**
  - Apoya la capacidad de las organizaciones para operar bajo regímenes legales o regulatorios aplicables.
- **Documento (Marco) Vivo**
  - Se actualiza regularmente para reflejar avances tecnológicos y cambios en la comprensión y enfoques hacia la confiabilidad de la IA.

Información Fundamental

- Enmarcando el Riesgo
- Gestión del Riesgo de IA: Camino hacia la minimización de impactos negativos potenciales de los sistemas de IA.
- Impactos Negativos: Amenazas a libertades civiles y derechos como foco de atención.
- Maximización de Impactos Positivos: Exploración de oportunidades para mejorar los resultados de IA.
- Documentación y Gestión: Pasos clave para abordar eficazmente los riesgos y potenciales impactos negativos.
- Sistemas de IA Confiables: Resultado de una gestión de riesgos eficaz.

Introducción al Manejo de Riesgos de IA

- **Definición de Riesgo**: Medida compuesta por la probabilidad de ocurrencia de un evento y la magnitud de sus consecuencias.

- **Impactos de los Sistemas de IA**: Positivos, negativos o ambos; pueden resultar en oportunidades o amenazas.

Comprendiendo el Impacto Negativo y el Riesgo

- **Función del Riesgo**: Impacto negativo o magnitud del daño vs. probabilidad de ocurrencia.

- **Afectados por el Impacto Negativo**: Individuos, grupos, comunidades, organizaciones, sociedad, medio ambiente y el planeta.

Gestión de Riesgos

- **Definición de Gestión de Riesgos**: Actividades coordinadas para dirigir y controlar una organización con respecto al riesgo.

- **Objetivo de la Gestión de Riesgos**: Minimizar impactos negativos anticipados y maximizar impactos positivos de los sistemas de IA.

Beneficios de una Gestión de Riesgos Efectiva

- **Resultados de una Gestión de Riesgos Efectiva**: Sistemas de IA más confiables y beneficios potenciales para personas, organizaciones y sistemas/ecosistemas.

- **Importancia de la Gestión de Riesgos**: Permite a los desarrolladores y usuarios de IA comprender impactos, considerar limitaciones y mejorar el rendimiento y la confiabilidad del sistema.

Flexibilidad y Evolución del AI RMF

- **Diseño del AI RMF**: Para abordar nuevos riesgos a medida que surgen, adaptándose a impactos no fácilmente previsibles y aplicaciones en evolución.

- **Desafíos en la Evaluación de Impactos Negativos**: Dificultades para evaluar impactos negativos y el grado de daños

Percepciones y Expectativas Humanas sobre la IA

- **Percepciones Humanas**: Los sistemas de IA a menudo se perciben como más objetivos o capaces que los humanos.

- **Importancia de Considerar Percepciones**: Gestión de riesgos de IA debe tener en cuenta las expectativas humanas sobre el funcionamiento de los sistemas de IA.

Fig. 1. Examples of potential harms related to AI systems. Trustworthy AI systems and their responsible use can mitigate negative risks and contribute to benefits for people, organizations, and ecosystems.

Fuente: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

Challenges for AI Risk Management

Introducción a la Medición del Riesgo de IA

- Importancia de definir y entender adecuadamente los riesgos de IA para su medición.
- La medición inapropiada no indica necesariamente un riesgo alto o bajo.

Challenges for AI Risk Management

Desafíos en la Medición del Riesgo

- Influencia de software, hardware y datos de terceros en la medición de riesgos.
- Divergencias en métricas y metodologías de riesgo entre organizaciones desarrolladoras y usuarias.
- La complejidad de la gestión de riesgos en el uso o integración de componentes de terceros

Desafíos en la gestión de riesgos

- Medición del Riesgo
- Tolerancia al Riesgo
- Priorización del riesgo
- Integración Organizacional y Gestión del Riesgo

Challenges for AI Risk Management

Rastreo de Riesgos Emergentes

- Mejora de la gestión de riesgos mediante la identificación y seguimiento de riesgos emergentes.
- Enfoques para evaluar impactos de sistemas de IA en contextos específicos.

Challenges for AI Risk Management

Disponibilidad de Métricas Confiables

- Falta de consenso sobre métodos de medición robustos y verificables para el riesgo y la confiabilidad de la IA.
- Desafíos y trampas potenciales en la medición de riesgos y daños negativos.

Medición de Impactos en Poblaciones

- Importancia del contexto en la medición de impactos.
- Diferencias en cómo los daños pueden afectar a diversos grupos o subgrupos.

Challenges for AI Risk Management

Riesgo en Diferentes Etapas del Ciclo de Vida de la IA

- Variabilidad en la medición de riesgos a lo largo de las etapas del ciclo de vida de la IA.
- Perspectivas de riesgo diferenciadas entre desarrolladores de IA y usuarios finales.

Riesgo en Entornos Reales

- Diferencias entre riesgos medidos en entornos controlados versus entornos operativos reales.

Challenges for AI Risk Management

Inescrutabilidad de Sistemas de IA

- Cómo la opacidad y la falta de transparencia de los sistemas de IA complican la medición de riesgos.

Comparación con Baselines Humanos

- Desafíos en la gestión de riesgos para sistemas de IA que buscan aumentar o reemplazar actividades humanas.
- Dificultades en establecer métricas de comparación base.

Challenges for AI Risk Management

Tolerancia al Riesgo en la Gestión de Riesgos de IA

- Definición de tolerancia al riesgo: Disposición de una organización o actor de IA para asumir riesgos con el fin de alcanzar sus objetivos.
- Influencia de requisitos legales y regulatorios en la tolerancia al riesgo.

Contextualidad de la Tolerancia al Riesgo

- La tolerancia al riesgo y el nivel de riesgo aceptable son altamente contextuales y específicos según la aplicación y el caso de uso.
- Influencia de políticas y normas establecidas por propietarios de sistemas de IA, organizaciones, industrias, comunidades o formuladores de políticas.

Challenges for AI Risk Management

Evolución de la Tolerancia al Riesgo

- Cambios en las tolerancias al riesgo a lo largo del tiempo con la evolución de sistemas de IA, políticas y normas.
- Variación de tolerancias al riesgo entre organizaciones debido a prioridades organizacionales y consideraciones de recursos.

Desafíos en la Especificación de Tolerancias al Riesgo de IA

- Desarrollo y debate continuo de conocimientos y métodos para informar mejor los intercambios entre daños/beneficios.
- Contextos en los que el marco de gestión de riesgos puede no ser aplicable aún para mitigar riesgos negativos de IA.

Challenges for AI Risk Management

Flexibilidad y Mejora de las Prácticas de Riesgo Existentes

- Intención del marco de ser flexible y complementar las prácticas de riesgo existentes, alineándose con leyes, regulaciones y normas aplicables.
- Seguimiento de regulaciones y directrices existentes sobre criterios de riesgo, tolerancia y respuesta establecidos por requerimientos organizacionales o sectoriales.

Definición y Gestión de la Tolerancia al Riesgo

- Importancia de definir una tolerancia al riesgo razonable en ausencia de directrices establecidas.
- Uso del AI RMF para gestionar riesgos y documentar procesos de gestión de riesgos una vez definida la tolerancia.

Challenges for AI Risk Management

Introducción a la Priorización de Riesgos

- La eliminación total del riesgo negativo es impráctica; no todos los incidentes y fallos pueden ser eliminados.
- Las expectativas poco realistas sobre el riesgo pueden conducir a una asignación ineficiente de recursos.

Challenges for AI Risk Management

Cultura de Gestión de Riesgos

- Importancia de reconocer que no todos los riesgos de IA son iguales.
- La asignación de recursos debe ser intencionada, basada en la evaluación del riesgo.

Gestión de Riesgos Accionable

- Establecimiento de pautas claras para evaluar la confiabilidad de cada sistema de IA.
- Priorización de políticas y recursos basada en el nivel de riesgo y el impacto potencial.

Challenges for AI Risk Management

Factores Contribuyentes

- La personalización del sistema de IA según el contexto específico de uso influye en la gestión del riesgo.

Aplicación del AI RMF

- Los riesgos más altos dentro de un contexto dado deben recibir la mayor urgencia y un proceso de gestión de riesgos más exhaustivo.
- En casos de riesgo negativo inaceptable, se debe cesar el desarrollo y despliegue de forma segura hasta que los riesgos puedan gestionarse adecuadamente.

Challenges for AI Risk Management

Priorización Según el Contexto

- Diferenciación en la priorización de riesgos entre sistemas de IA que interactúan directamente con humanos frente a aquellos que no lo hacen.
- Mayor priorización inicial para sistemas de IA entrenados con datos sensibles o que impactan directa o indirectamente en los humanos.

Challenges for AI Risk Management

Riesgo Residual

- Definido como el riesgo que permanece después del tratamiento del riesgo.
- La documentación de riesgos residuales es crucial para informar a los usuarios finales sobre los impactos negativos potenciales.

Challenges for AI Risk Management

Integración Organizacional y Gestión del Riesgo

- Gestión de Riesgos de IA: No debe considerarse en aislamiento.
- Roles Diversos: Responsabilidades y conciencia varían a lo largo del ciclo de vida de la IA.

Integración con la Gestión de Riesgos Empresariales

- Incorporación en Estrategias más Amplias: La gestión del riesgo de IA como parte de la gestión de riesgos empresariales.
- Beneficios de la Integración: Mejora en la eficiencia organizacional y resultados integrados.

Challenges for AI Risk Management

Riesgos Comunes en el Desarrollo de Software

- Privacidad y Uso de Datos: Preocupaciones al entrenar sistemas de IA.
- Implicaciones Energéticas y Ambientales: Demandas computacionales intensivas.
- Seguridad de la Información: Confidencialidad, integridad y disponibilidad del sistema y sus datos.

Mecanismos de Responsabilidad y Cultura Organizacional

- Establecimiento de Responsabilidades: Mecanismos de responsabilidad y estructuras de incentivo adecuadas.
- Compromiso Organizacional: Importancia del compromiso a niveles superiores y cambio cultural.

Challenges for AI Risk Management

Desafíos para Organizaciones Pequeñas y Medianas

- **Diferencias en Capacidad y Recursos**: Desafíos únicos frente a las grandes organizaciones.
- **Implementación del AI RMF**: Adaptación según las capacidades y recursos disponibles.

Audiencia

Audiencia del AI RMF

- Gestión de Riesgos de IA: Necesidad de perspectivas y actores diversos a lo largo del ciclo de vida de la IA.
- Diversidad en IA: Importancia de equipos demográfica y disciplinariamente diversos.

Ciclo de Vida y Dimensiones de la IA

- Marco de la OCDE: Clasificación de actividades de IA en cinco dimensiones socio-técnicas.
- Modificaciones de NIST: Enfatizando procesos de prueba, evaluación, verificación y validación (TEVV).

Audiencia

Dimensiones Clave de la IA

- Contexto de Aplicación: Cómo se utiliza la IA.
- Datos y Entrada: La información que alimenta los sistemas de IA.
- Modelo de IA: Las técnicas y algoritmos utilizados.
- Tarea y Salida: Lo que hace el sistema de IA.

Actores de IA en el AI RMF

- Diseño y Desarrollo: Quienes construyen y configuran sistemas de IA.
- Despliegue y Uso: Quienes implementan y operan sistemas de IA.
- Evaluación TEVV: Expertos que integran evaluaciones a lo largo del ciclo de vida de la IA.

Audiencia

Contribuciones de los Actores de IA

- Gestión de Riesgos: Esfuerzos conjuntos para manejar riesgos y lograr IA confiable y responsable.
- Percepciones de TEVV: Aportes en estándares técnicos, sociales, legales y éticos.

Dimensión de Personas y Planeta

- Derechos Humanos y Bienestar: Centralidad de los derechos humanos y el bienestar general.
- Actores Informativos: Grupos que informan y orientan a la audiencia principal del AI RMF.

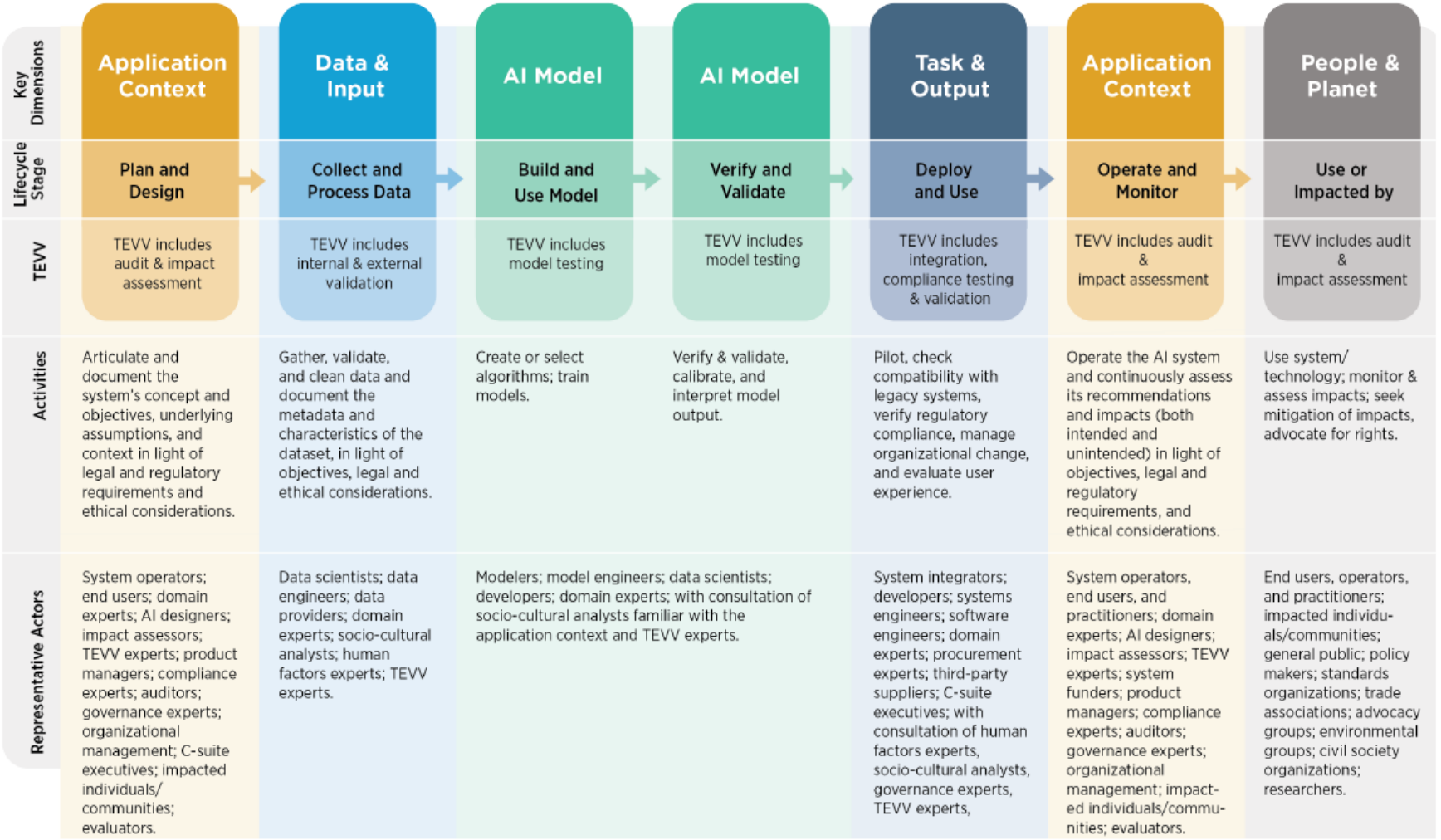
Audiencia

Gestión de Riesgos Exitosa

- Responsabilidad Colectiva: La importancia de la responsabilidad compartida entre los actores de la IA.

- Perspectivas Diversas: El valor de equipos diversos para una gestión de riesgos efectiva.

Audiencia



Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

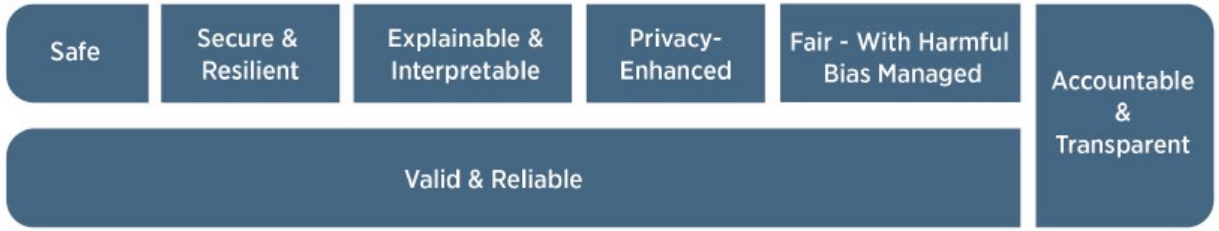
AI Risks and Trustworthiness

Introducción a la Confiabilidad en Sistemas de IA

- Los sistemas de IA confiables deben responder a múltiples criterios de valor para las partes interesadas.
- Mejorar la confiabilidad de la IA puede reducir los riesgos negativos asociados.

Características de la IA Confiable

- Validación y fiabilidad
- Seguridad
- Seguridad y resiliencia
- Rendición de cuentas y transparencia



AI Risks and Trustworthiness

Características Continuadas

- Explicabilidad e interpretabilidad
- Mejora de la privacidad
- Justicia y gestión del sesgo perjudicial

Balance de Características

- Necesidad de equilibrar estas características basándose en el contexto de uso del sistema de IA.
- La rendición de cuentas y la transparencia se relacionan con todos los demás aspectos.

AI Risks and Trustworthiness

Influencias Sociales y Organizacionales

- Características de la confiabilidad de IA vinculadas al comportamiento social y organizacional, datos, selección de modelos y algoritmos de IA, y las decisiones de quienes los construyen.

Juicio Humano y Confiabilidad

- El juicio humano es crucial al decidir métricas específicas y valores umbral para características de confiabilidad de IA.

AI Risks and Trustworthiness

Navegando Compromisos

- Abordar individualmente las características de confiabilidad no garantiza la confiabilidad del sistema de IA.
- La gestión de riesgos de IA puede requerir decisiones difíciles para equilibrar estas características.

Ejemplos de Compromisos

- Ejemplos de compromisos entre interpretabilidad y privacidad, precisión predictiva y privacidad, o precisión y justicia en ciertos dominios.

AI Risks and Trustworthiness

Mejorando la Conciencia Contextual en el Ciclo de Vida de la IA

- Abordajes para aumentar la conciencia contextual a lo largo del ciclo de vida de la IA, incluyendo la evaluación por expertos y la alineación de parámetros.

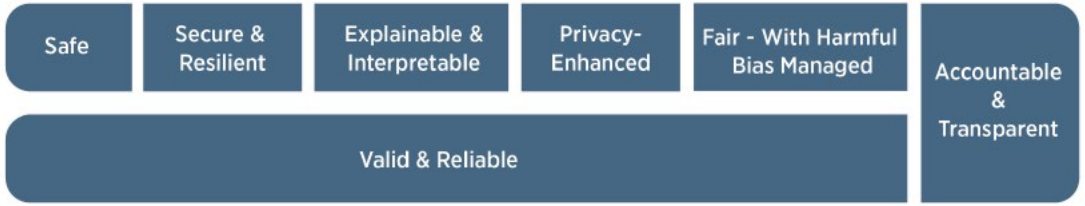
Responsabilidad Compartida y Evaluación Contextual

- Es responsabilidad conjunta de todos los actores de IA determinar el uso adecuado y responsable de la tecnología de IA.
- La decisión de comisionar o desplegar un sistema de IA debe basarse en una evaluación contextual de las características de confiabilidad y los riesgos relativos.

AI Risks and Trustworthiness

Introducción a la Validez y Fiabilidad en IA

- Definición de validez: Confirmación de que se han cumplido los requisitos para un uso o aplicación específicos.
- Importancia de la validez: Evitar el despliegue de sistemas de IA inexactos o poco fiables.
- Definición de fiabilidad: Capacidad de un sistema par condiciones dadas.



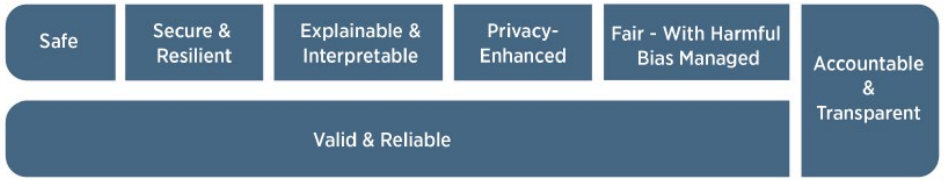
Validez en Sistemas de IA

- Definición de Precisión: Cercanía entre los resultados de observaciones o cálculos y los valores verdaderos o aceptados como tales.
- Importancia de la Precisión: Contribuye a la validez y confiabilidad de los sistemas de IA.
- Medidas de Precisión: Incluir medidas centradas en la computación y la colaboración humano-IA.

AI Risks and Trustworthiness

Fiabilidad y Operación Correcta de IA

- Meta de la Fiabilidad: Correctitud en la operación del sistema de IA bajo condiciones esperadas.
- Incluye: Toda la vida útil del sistema.
- Contribución a la Confiabilidad: Precisión y robustez.



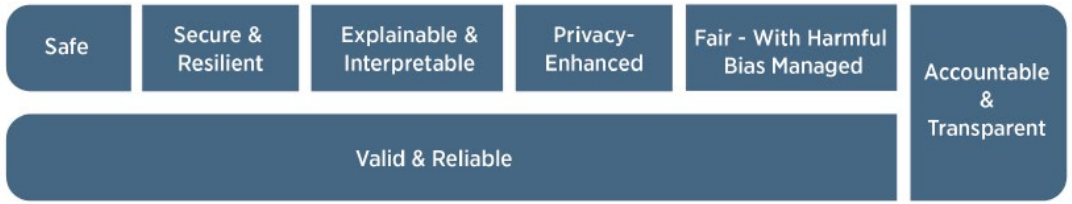
Robustez y Generalización

- Definición de Robustez: Capacidad de mantener el nivel de rendimiento bajo diversas circunstancias.
- Meta de la Robustez: Funcionalidad adecuada en un conjunto amplio de condiciones.
- Importancia de la Robustez: Minimizar daños potenciales en escenarios no anticipados.

AI Risks and Trustworthiness

Medición y Monitoreo de la Validez y Fiabilidad

- **Evaluación Continua**: Testing y monitoreo en curso para confirmar el rendimiento según lo previsto.
- **Medición de Validez y Fiabilidad**: Contribuye a la confiabilidad y debe considerar los daños potenciales de fallos específicos.
- **Gestión de Riesgos de IA**: Priorizar la minimización de impactos negativos y la intervención humana en la detección y corrección de errores.



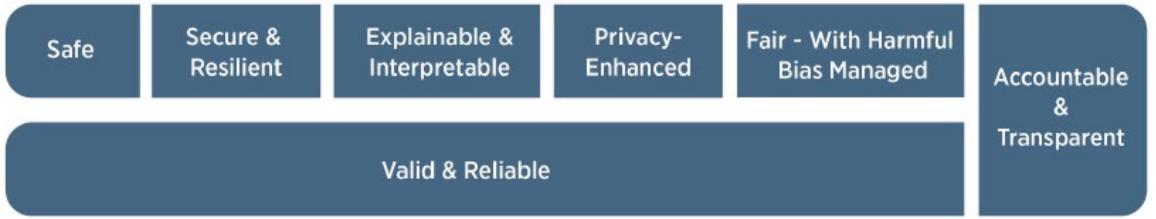
SAFE

Definición de Seguridad

- Los sistemas de IA no deben, bajo condiciones definidas, llevar a un estado que ponga en peligro la vida, la salud, la propiedad o el medio ambiente. (Fuente: ISO/IEC TS 5723:2022).

Mejora de la Operación Segura

- Prácticas responsables de diseño, desarrollo e implementación.
- Información clara sobre el uso responsable del sistema a los implementadores.
- Toma de decisiones responsable por parte de los implementadores y usuarios finales.
- Documentación y explicaciones de riesgos basadas en evidencias empíricas de incidentes.



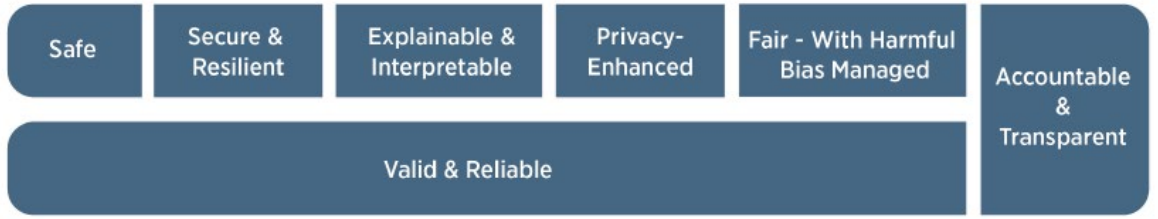
SAFE

Riesgos de Seguridad y Enfoques de Gestión

- Necesidad de enfoques de gestión de riesgos de IA personalizados según el contexto y la gravedad de los riesgos potenciales.
- Los riesgos de seguridad que presentan un riesgo potencial de lesiones graves o muerte requieren una priorización urgente y un proceso de gestión de riesgos exhaustivo.

Consideraciones de Seguridad durante el Ciclo de Vida de la IA

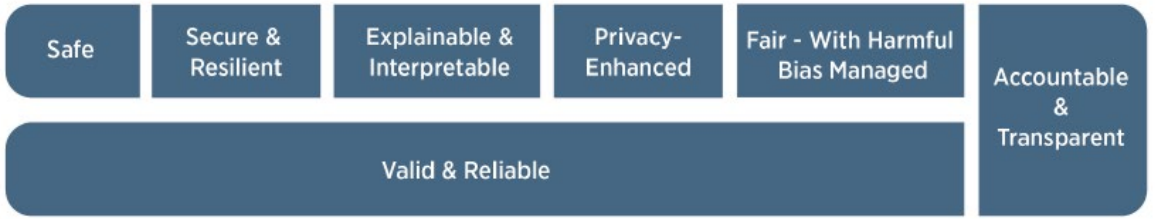
- Iniciar tan pronto como sea posible con la planificación y el diseño para prevenir fallos o condiciones peligrosas.
- Enfoques prácticos para la seguridad de la IA incluyen simulaciones rigurosas, pruebas en el dominio específico, monitoreo en tiempo real, y la capacidad de apagar, modificar o intervenir humanamente en sistemas que se desvían de la funcionalidad intencionada o esperada.



SAFE

Alineación con Esfuerzos y Directrices de Seguridad Existentes

- Los enfoques de gestión de riesgos de seguridad de IA deben tomar ejemplo de esfuerzos y directrices de seguridad en campos como el transporte y la salud.
- Alinear con guías o estándares específicos del sector o de la aplicación existentes.



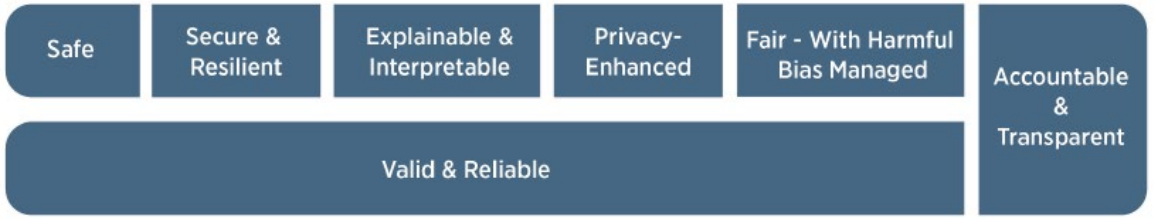
Seguridad y Resiliencia en Sistemas de IA

Introducción a la Seguridad y Resiliencia en Sistemas de IA

- Definición de Resiliencia: Capacidad de los sistemas de IA y sus ecosistemas para soportar eventos adversos inesperados o cambios en su entorno o uso.
- Mantener Funciones y Estructura: Importancia de la adaptabilidad frente a cambios internos y externos, y degradación segura cuando es necesario.

Desafíos Comunes en la Seguridad de la IA

- Ejemplos Adversariales: Manipulación de sistemas de IA para inducir errores.
- Envenenamiento de Datos: Inserción de datos falsos o engañosos durante el entrenamiento.
- Exfiltración de Información: Robo de modelos de IA, datos de entrenamiento o propiedad intelectual a través de puntos finales del sistema de IA.



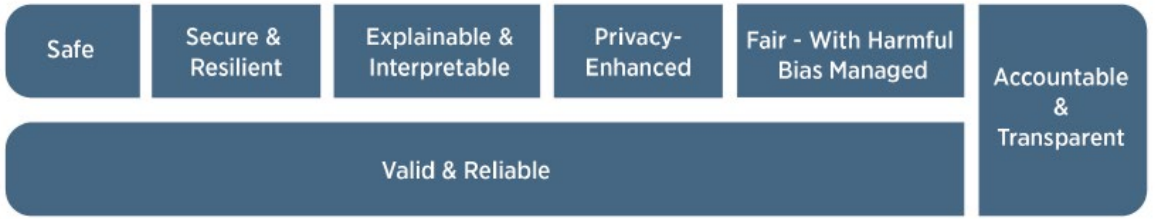
Introducción a la Transparencia y Responsabilidad en IA

Introducción a la Transparencia y Responsabilidad en IA

- Dependencia de la IA Confiable: Esencialidad de la responsabilidad y la transparencia.
- Premisa de Responsabilidad: Necesidad de transparencia para la rendición de cuentas.

Definición de Transparencia

- Disponibilidad de Información: Importancia del acceso a información sobre el sistema de IA y sus resultados.
- Transparencia Significativa: Adaptación de los niveles de información según la etapa del ciclo de vida de la IA y el rol del actor.



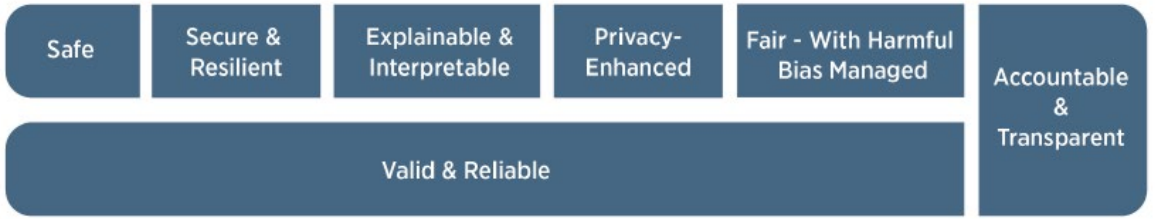
Introducción a la Transparencia y Responsabilidad en IA

Beneficios de la Transparencia

- Promoción de la Comprensión: Cómo la transparencia aumenta la confianza en los sistemas de IA.
- Alcance de la Transparencia: Desde decisiones de diseño hasta uso final e interacciones.

Transparencia y Redress

- Importancia para el Redress Accionable: Transparencia frente a resultados incorrectos o impactos negativos.
- Interacción Humano-IA: Notificación de resultados adversos potenciales o reales.



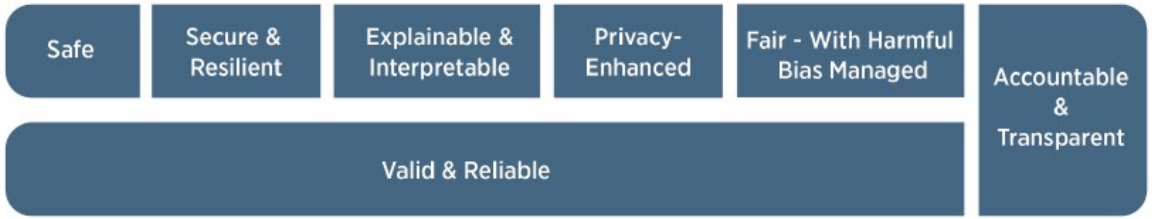
Introducción a la Transparencia y Responsabilidad en IA

Transparencia ≠ Precisión/Aseguramiento

- Limitaciones de la Transparencia: No garantiza precisión, privacidad, seguridad o equidad por sí sola.
- Desafíos de Sistemas Opacos: Dificultad para determinar características en sistemas complejos y evolutivos.

Responsabilidad y Actores de IA

- Relación entre Riesgo y Responsabilidad: Variabilidad según contextos culturales, legales y sectoriales.
- Prácticas de Transparencia y Responsabilidad: Ajustes proporcionales en contextos de alto riesgo.



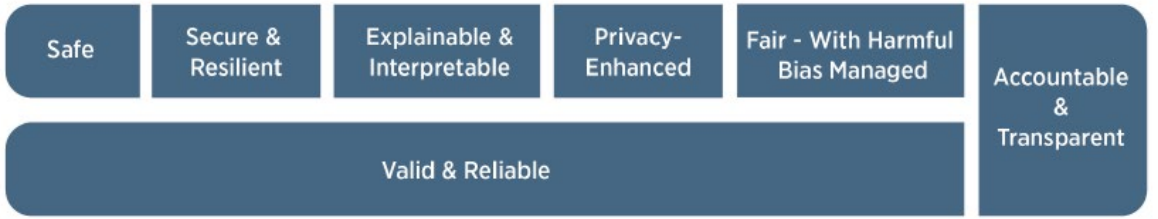
Introducción a la Transparencia y Responsabilidad en IA

Estrategias para Mejorar la Responsabilidad

- Estructuras Organizativas: Para reducción de daños y gestión de riesgos.
- Consideraciones de Implementación: Impacto en entidades implementadoras, recursos necesarios y protección de información propietaria.

Transparencia, Proveniencia de Datos y Herramientas

- Mantenimiento de la Proveniencia de Datos: Asistencia en transparencia y responsabilidad.
- Evolución de Herramientas de Transparencia: Fomento de la experimentación y colaboración en el desarrollo de sistemas de IA.



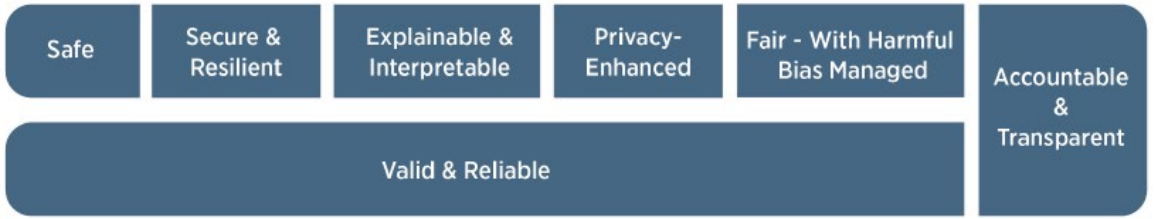
Introducción a la Explicabilidad e Interpretabilidad

Introducción a la Explicabilidad e Interpretabilidad

- Explicabilidad: Representación de los mecanismos que subyacen en la operación de los sistemas de IA.
- Interpretabilidad: Significado de la salida de los sistemas de IA en el contexto de sus propósitos funcionales diseñados.

Importancia de la Explicabilidad e Interpretabilidad

- Facilitan a quienes operan o supervisan un sistema de IA, así como a los usuarios, obtener una comprensión más profunda de la funcionalidad y confiabilidad del sistema, incluyendo sus salidas.
- Ayudan a manejar percepciones de riesgo negativo derivadas de la incapacidad para entender o contextualizar adecuadamente la salida del sistema.



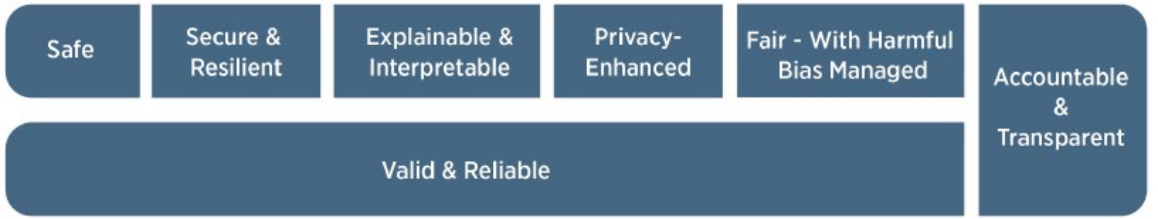
Introducción a la Explicabilidad e Interpretabilidad

Beneficios de la IA Explicable e Interpretable

- Ofrecen información que ayuda a los usuarios finales a entender los propósitos y el impacto potencial de un sistema de IA.
- Permiten un mejor diagnóstico, monitoreo, documentación, auditoría y gobernanza de los sistemas.

Manejo de Riesgos por Falta de Explicabilidad

- La gestión del riesgo puede involucrar describir cómo funcionan los sistemas de IA, adaptando las descripciones a las diferencias individuales como el rol, conocimiento, y nivel de habilidad del usuario.



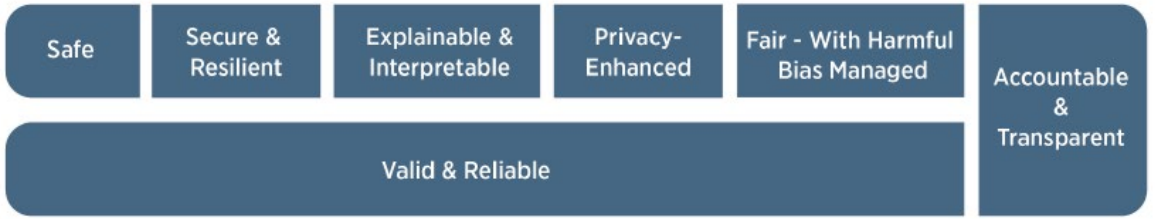
Introducción a la Explicabilidad e Interpretabilidad

Abordando Riesgos para la Interpretabilidad

- Los riesgos a la interpretabilidad a menudo pueden ser abordados comunicando una descripción de por qué un sistema de IA hizo una predicción o recomendación particular.

Transparencia, Explicabilidad e Interpretabilidad

- **Transparencia**: Responde a la pregunta de "qué sucedió" en el sistema.
- **Explicabilidad**: Responde a la pregunta de "cómo" se tomó una decisión en el sistema.
- **Interpretabilidad**: Responde a la pregunta de "por qué" se tomó una decisión por el sistema y su significado o contexto para el usuario.



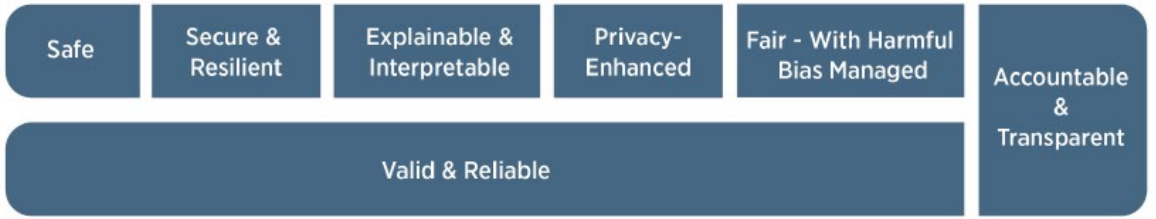
Privacidad Mejorada

Introducción a la Privacidad

- Definición: La privacidad protege la autonomía, identidad y dignidad humanas.
- Normas y Prácticas: Enfocadas en evitar intrusiones, limitar observación y permitir el control individual sobre la identidad.

Valores de la Privacidad

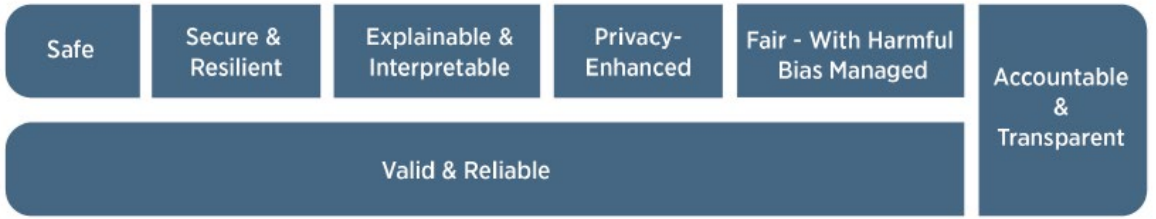
- **Anonimato, Confidencialidad, Control**: Pilares en el diseño de sistemas de IA.
- **Guía para el Diseño de Sistemas de IA**: Estos valores deben orientar las decisiones de diseño, desarrollo y despliegue



Privacidad Mejorada

Riesgos Relacionados con la Privacidad

- Influencia en Seguridad, Sesgo y Transparencia: Los riesgos de privacidad pueden afectar otras características de los sistemas de IA.
- Trade-offs: Equilibrios necesarios entre privacidad, seguridad, sesgo y transparencia.



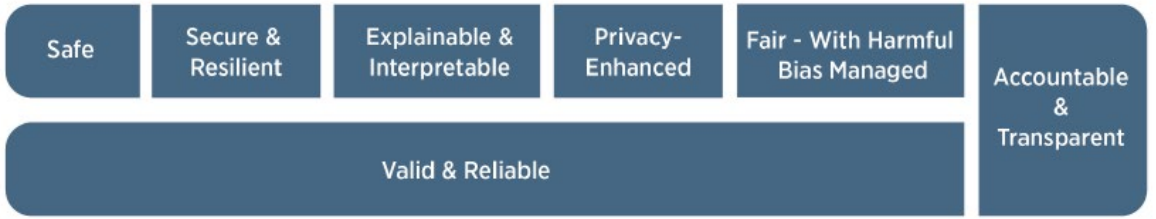
Privacidad Mejorada

Tecnologías para Mejorar la Privacidad (PETs)

- **Definición y Aplicación**: Tecnologías diseñadas para fortalecer la privacidad en sistemas de IA.
- **Ejemplos**: De-identificación, agregación y otras técnicas de minimización de datos.

Desafíos y Compromisos

- **Equilibrio entre Privacidad y Exactitud**: Cómo las técnicas de mejora de privacidad pueden impactar la precisión de los sistemas de IA.
- **Decisión sobre la Equidad**: Consideraciones de equidad y otros valores en el contexto de la privacidad mejorada.



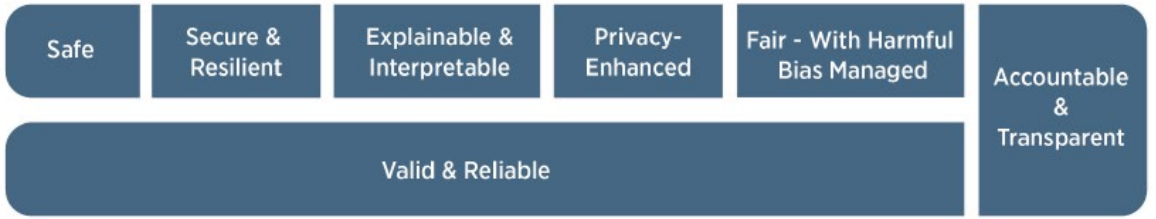
Justo – con el Sesgo Nocivo Gestionado

Introducción a la Equidad en IA

- Preocupación por la Equidad: Incluye igualdad y equidad, abordando sesgos nocivos y discriminación.
- Complejidad de la Equidad: Estándares difíciles de definir, variando entre culturas y aplicaciones.

Reconocimiento de Diferencias

- Gestión de Riesgos Mejorada: Por el reconocimiento y consideración de diferencias en percepciones de equidad.
- Desafíos de Equidad: Sistemas sin sesgos nocivos pueden seguir siendo injustos.



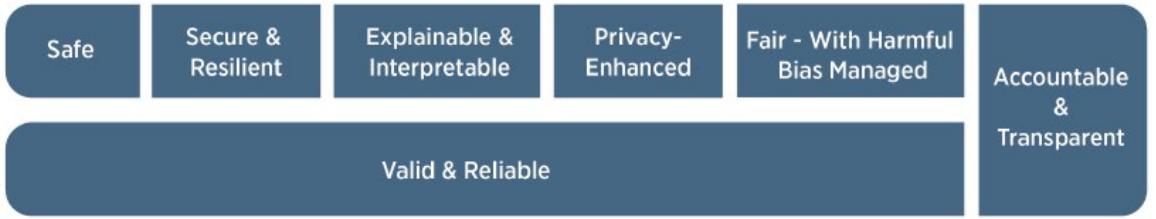
Justo – con el Sesgo Nocivo Gestionado

Más Allá del Equilibrio Demográfico

- Inaccesibilidad y Brecha Digital: Ejemplos de cómo la equidad va más allá del balance demográfico.
- Exacerbación de Disparidades: Potencial de aumentar disparidades existentes o sesgos sistémicos.

Categorías de Sesgo en IA

- Sesgo Sistémico: Presente en datasets de IA, normas organizacionales y la sociedad.
- Sesgo Computacional y Estadístico: Errores sistemáticos en datos y algoritmos.
- Sesgo Cognitivo Humano: Percepciones y decisiones influenciadas en el ciclo de vida de la IA.



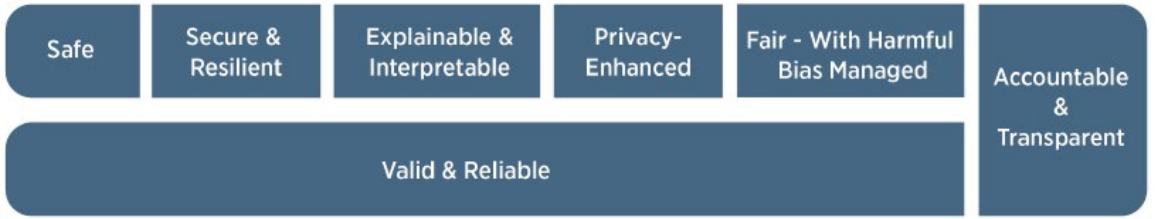
Justo – con el Sesgo Nocivo Gestionado

Impactos del Sesgo

- Formas y Consecuencias del Sesgo: Cómo se incrusta y amplifica en sistemas automatizados.
- Velocidad y Escala: Potencial de IA para aumentar y perpetuar daños.
- Transparencia y Equidad: Asociación estrecha con conceptos de transparencia en la sociedad.

Gestionando el Sesgo en IA

- Enfoque en Tres Categorías Principales: Sesgo sistémico, computacional/estadístico, y cognitivo humano.
- Publicación Especial de NIST 1270: Hacia un Estándar para Identificar y Gestionar el Sesgo en Inteligencia Artificial.



Efectividad del AI RMF

Efectividad del AI RMF

- Introducción: Evaluaciones de efectividad del AI RMF, incluyendo métodos para medir mejoras en la confiabilidad de sistemas de IA.
- Futuras Actividades de NIST: Colaboración con la comunidad de IA para desarrollar métricas y metodologías.

Evaluación Continua

- Evaluación por Parte de las Organizaciones: Importancia de evaluar periódicamente el impacto del AI RMF en la gestión de riesgos de IA.
- Aspectos Clave: Políticas, procesos, prácticas, planes de implementación, indicadores, mediciones y resultados esperados.

Efectividad del AI RMF

Beneficios del Uso del Marco

- Mejora de Procesos: Gobernanza, mapeo, medición y gestión de riesgos de IA documentados claramente.
- Conciencia Mejorada: Relación y compensaciones entre características de confiabilidad, enfoques socio-técnicos y riesgos de IA.

Decisiones y Responsabilidad

- Procesos Explícitos: Para decisiones de condicionamiento y despliegue de sistemas.
- Mejora de la Responsabilidad Organizacional: Políticas y procedimientos enfocados en riesgos de sistemas de IA.

Efectividad del AI RMF

Cultura Organizacional y Conocimiento Contextual

- Cultura Organizacional Mejorada: Prioriza la identificación y gestión de riesgos de IA.
- Conocimiento Contextual: Para una mayor conciencia de los riesgos aguas abajo.

Comunicación y Colaboración

- Compartir Información: Dentro y entre organizaciones sobre riesgos y prácticas de decisión.
- Engagement con Partes Interesadas: Fortalecimiento de la participación con partes interesadas relevantes.

Efectividad del AI RMF

Capacidad Aumentada para TEVV

- **TEVV de Sistemas de IA**: Capacidad aumentada para Pruebas, Evaluación, Verificación y Validación de sistemas de IA y riesgos asociados.

Core and Profiles

Core (Núcleo)

Introducción al Núcleo del AI RMF

- Objetivo
  - Facilitar el diálogo, la comprensión y las actividades para gestionar riesgos de IA y desarrollar sistemas de IA confiables.

- Componentes Clave:
  - **GOVERN, MAP, MEASURE, MANAGE.**

Core (Núcleo)

GOVERN

- Descripción:
  - Establecer marcos de gobernanza para informar y ser informado por otras funciones del AI RMF.
- Importancia:
  - Función transversal esencial para una gestión efectiva de riesgos de IA.

Core (Núcleo)

MAP

- **Descripción**: Mapear relaciones, dependencias, y el entorno de IA para identificar riesgos y oportunidades.
- **Objetivo**: Crear una comprensión clara del ecosistema de IA y sus interacciones.

Core (Núcleo)

MEASURE

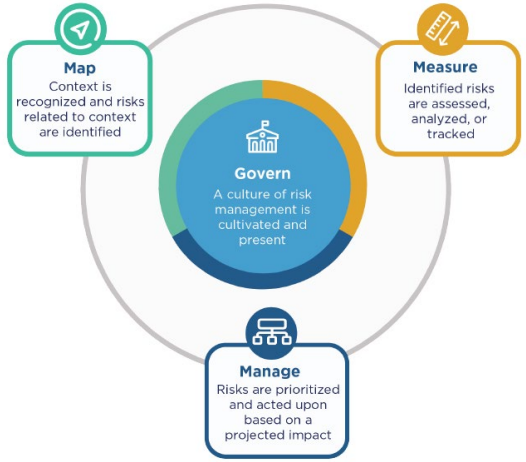
- **Descripción**: Medir y evaluar riesgos, rendimientos y conformidad dentro del contexto de IA.
- **Enfoque**: Uso de métricas y evaluaciones para informar decisiones de gestión de riesgos.

Core (Núcleo)

MANAGE

- **Descripción**: Administrar y mitigar riesgos de IA a través de acciones estratégicas y operativas.
- **Estrategias**: Implementación de medidas de control y seguimiento de la efectividad.

Core (Núcleo)

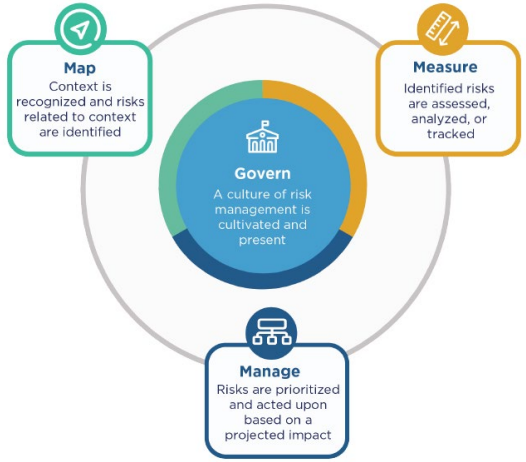


Gestión Continua de Riesgos

- **Proceso Continuo**: La gestión de riesgos debe ser oportuna y realizarse a lo largo de todo el ciclo de vida del sistema de IA.

- **Perspectivas Diversas**: Reflejar puntos de vista diversos y multidisciplinarios para una gestión de riesgos más efectiva.

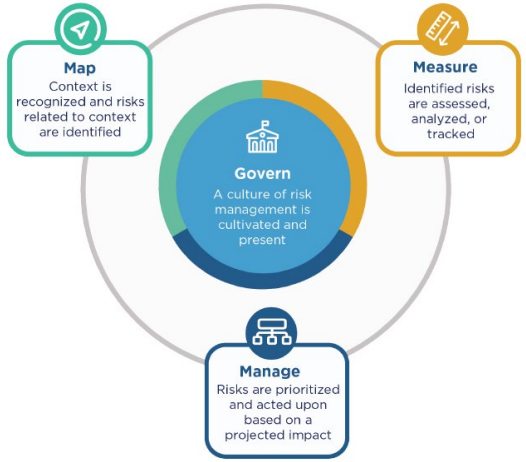
Core (Núcleo)



Contribución de un Equipo Diverso

- **Beneficios**: Mayor intercambio de ideas y detección de problemas y riesgos emergentes.
- **Meta**: Fomentar la inclusión de perspectivas de actores de IA tanto internos como externos a la organización.

Core (Núcleo)

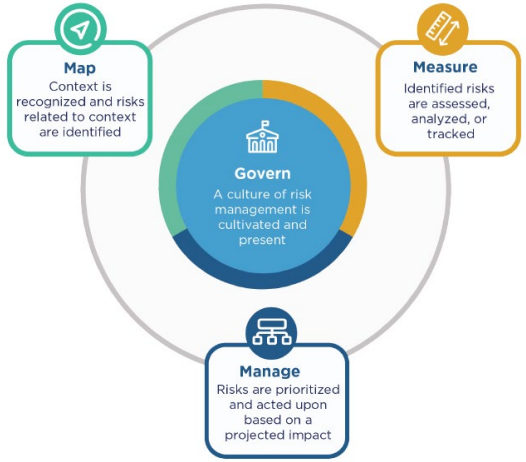


Voluntariedad y Personalización

- **Uso Voluntario**: Las organizaciones pueden utilizar las sugerencias del Playbook según sus necesidades e intereses.

- **Guía Personalizada**: Posibilidad de crear orientación adaptada a partir del material sugerido.

Core (Núcleo)

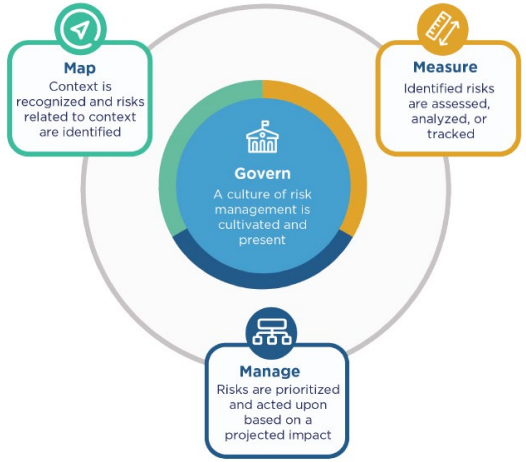


Contribuciones Comunitarias

- **Compartir Sugerencias**: Usuarios pueden contribuir con sus propias sugerencias para enriquecer la comunidad.

- **Centro de Recursos de IA Confiable y Responsable de NIST**: El Playbook es parte de esta iniciativa más amplia.

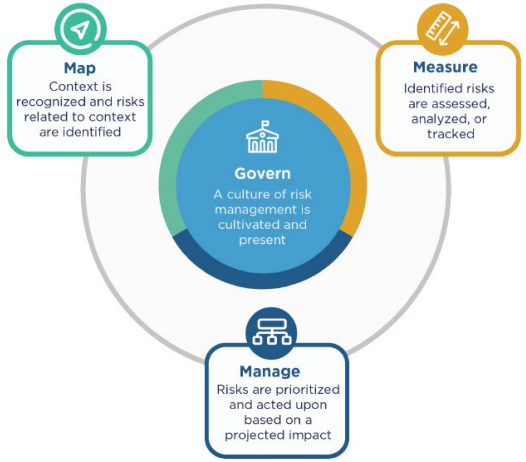
Core (Núcleo)



Aplicación Flexible de Funciones

- **Adaptación a Necesidades y Capacidad**: Elección entre categorías y subcategorías basadas en recursos y capacidades.
- **Integración Flexible**: Las funciones pueden ser aplicadas en cualquier orden a lo largo del ciclo de vida de la IA.

Core (Núcleo)

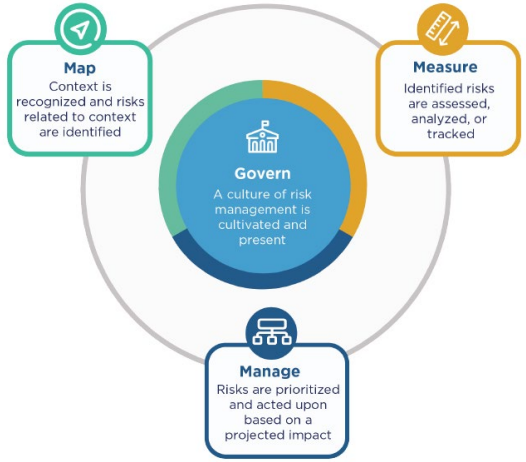


Proceso Iterativo y Referencia Cruzada

- **Iteración y Referencias Cruzadas**: Proceso iterativo con referencias cruzadas entre funciones según sea necesario.

- **Elementos Aplicables a Múltiples Funciones**: Categorías y subcategorías que se aplican de manera transversal.

Introducción a GOVERN



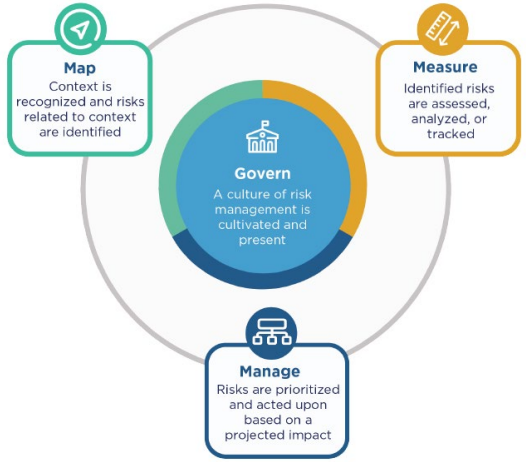
Introducción a la Función GOVERN

- Objetivo: Fomentar una cultura de gestión de riesgos en organizaciones que diseñan, desarrollan, despliegan, evalúan o adquieren sistemas de IA.

Cultura de Gestión de Riesgos

- Cultivar e Implementar: Estrategias y prácticas organizacionales centradas en la gestión proactiva de riesgos de IA.
- Componentes Clave: Procesos, documentos, y esquemas organizacionales para anticipar, identificar y gestionar riesgos.

Introducción a GOVERN



Procesos y Documentación

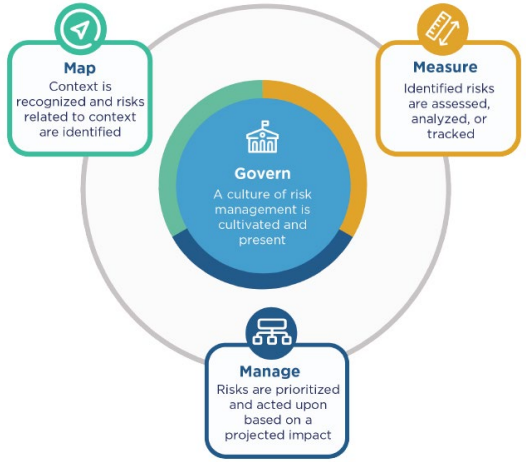
- Esquemas Organizacionales: Diseñar procesos que abordan directamente los riesgos que los sistemas pueden plantear a usuarios y a la sociedad.

- Alcanzando Resultados: Procedimientos claros para lograr una gestión efectiva de riesgos.

Evaluación de Impactos

- Incorporación de Procesos: Evaluación de los impactos potenciales de los sistemas de IA en el entorno y en los individuos.

Introducción a GOVERN



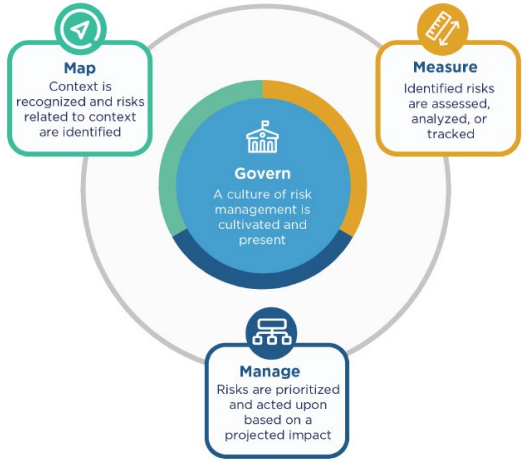
Alineación con Principios Organizacionales

- Estructura y Prioridades: Cómo la gestión de riesgos de IA se alinea con los principios, políticas y prioridades estratégicas de la organización.

Conexión entre Técnico y Organizacional

- Valores y Principios: Conectar aspectos técnicos del diseño y desarrollo de sistemas de IA con los valores y principios organizacionales.
- Competencias para el Personal: Facilitar prácticas organizacionales y competencias para individuos involucrados en la adquisición, entrenamiento, despliegue y monitoreo de sistemas de IA.

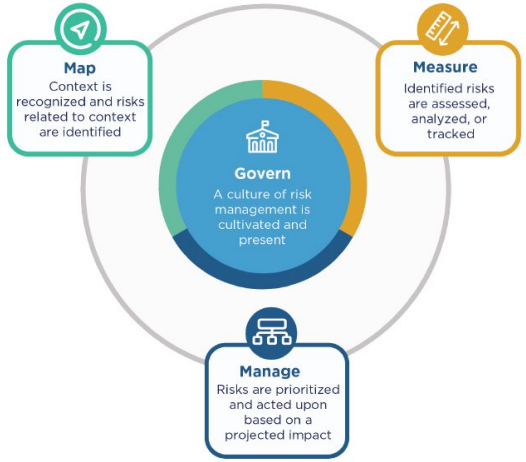
Introducción a GOVERN



Ciclo de Vida del Producto y Consideraciones Legales

- Ciclo Completo del Producto: Abordar todos los aspectos del ciclo de vida del producto y los procesos asociados, incluyendo el uso de software o hardware de terceros y cuestiones de datos.

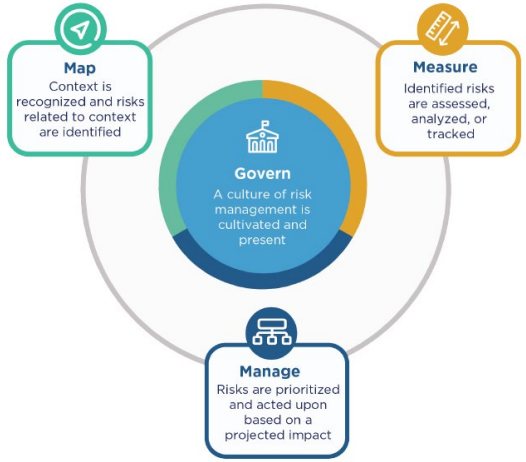
Introducción a GOVERN



- **Función Transversal**: GOVERN es esencial en la gestión de riesgos de IA, impactando y mejorando todas las otras funciones.

- **Importancia**: Establecimiento de una cultura de gestión de riesgos dentro de la organización a lo largo del ciclo de vida de los sistemas de IA.

Introducción a GOVERN

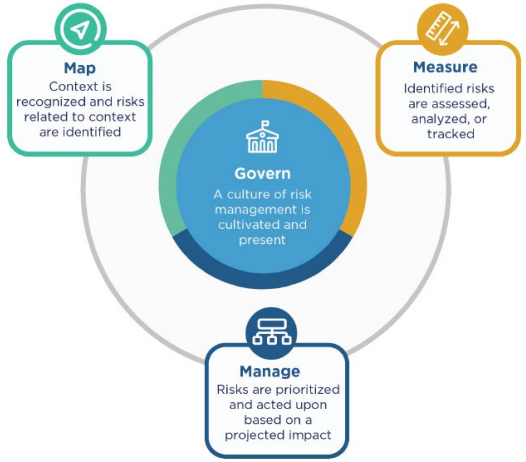


Cultura y Estructura de Gestión de Riesgos

- **Cultivo de una Cultura de Riesgo**: Implementación de prácticas y normas internas para facilitar una cultura organizacional enfocada en el riesgo.

- **Estructura Organizacional**: Conexión de aspectos técnicos del diseño y desarrollo de sistemas de IA con valores y principios organizacionales.

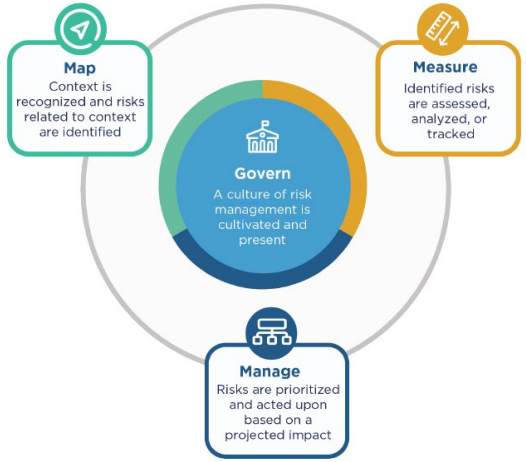
Introducción a GOVERN



Procesos y Documentación

- **Procesos**: Esquemas organizativos para anticipar, identificar y gestionar riesgos, incluidos impactos potenciales en usuarios y la sociedad.
- **Documentación**: Mejora de la transparencia y la rendición de cuentas a través de procesos de revisión humana y documentación adecuada.

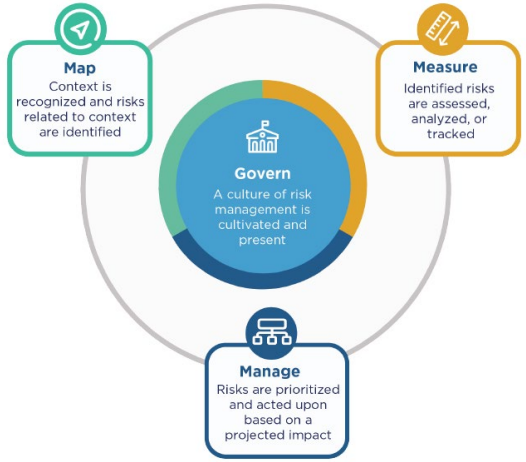
Introducción a GOVERN



Liderazgo y Gobernanza

- **Autoridades de Gobernanza**: Determinación de políticas generales que dirigen la misión, metas, valores, cultura y tolerancia al riesgo de la organización.
- **Liderazgo**: El liderazgo senior establece el tono para la gestión de riesgos dentro de la organización, influyendo en la cultura organizacional.

Introducción a GOVERN

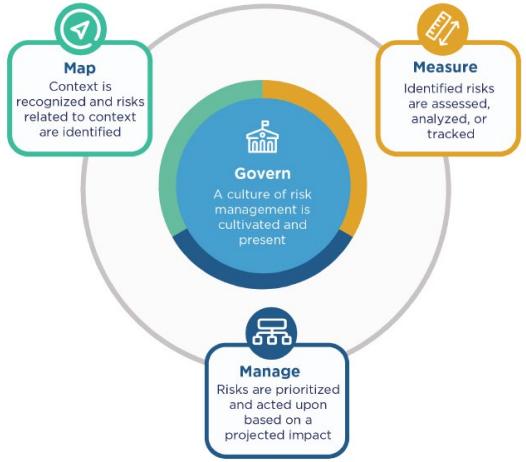


Alineación Técnica y Operativa

- **Alineación con Políticas**: Integración de aspectos técnicos de la gestión de riesgos de IA con políticas y operaciones organizacionales.

- **Beneficios Organizacionales**: Cultura orientada al propósito enfocada en el entendimiento y la gestión de riesgos.

Introducción a GOVERN

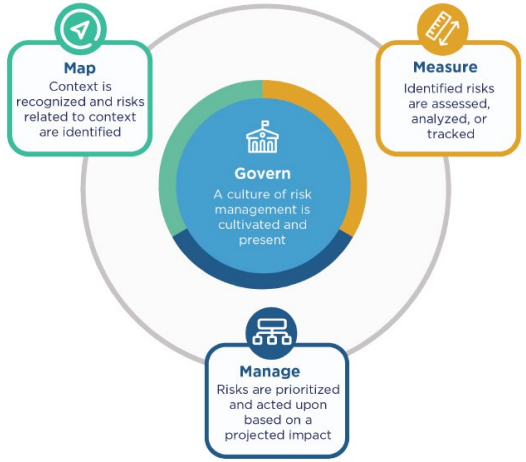


Continuidad y Evolución

- **Ejecución Continua**: Importancia de continuar ejecutando la función GOVERN conforme evolucionan los conocimientos, culturas y necesidades.

- **Adaptación**: Ajuste a las expectativas cambiantes de los actores de IA a lo largo del tiempo.

Introducción a GOVERN



NIST AI RMF Playbook

- **Prácticas de Gobernanza**: Descripción de prácticas relacionadas con la gobernanza de riesgos de IA en el Playbook de AI RMF de NIST.
- **Categorías y Subcategorías**: Referencia a la Tabla 1 para detalles específicos de la función GOVERN.

Categorías y subcategorías para la función de GOVERNANZA (GOVERN).

Categoría	Subcategoría	Descripción
GOVERNANZA 1		**Políticas, procesos, procedimientos y prácticas en toda la organización** relacionadas con el mapeo, medición y gestión de los riesgos de IA están establecidos, son transparentes y se implementan de manera efectiva.
	GOVERNANZA 1.1	Los requisitos legales y reglamentarios que involucran a la IA se comprenden, gestionan y documentan.
	GOVERNANZA 1.2	Las características de una IA confiable se integran en las políticas, procesos, procedimientos y prácticas organizacionales.
	GOVERNANZA 1.3	Se establecen procesos, procedimientos y prácticas para determinar el nivel necesario de actividades de gestión de riesgos basado en la tolerancia al riesgo de la organización.
	GOVERNANZA 1.4	El proceso de gestión de riesgos y sus resultados se establecen a través de políticas transparentes, procedimientos y otros controles basados en las prioridades de riesgo organizacionales.

Categorías y subcategorías para la función de GOVERNANZA (GOVERN).

Categoría	Subcategoría	Descripción
GOVERNANZA 1		Políticas, procesos, procedimientos y prácticas en toda la organización relacionadas con el mapeo, medición y gestión de los riesgos de IA están establecidos, son transparentes y se implementan de manera efectiva.
	GOVERNANZA 1.5	Se planifica el monitoreo continuo y la revisión periódica del proceso de gestión de riesgos y sus resultados, y se definen claramente los roles y responsabilidades organizacionales, incluyendo la determinación de la frecuencia de la revisión periódica.
	GOVERNANZA 1.6	Se disponen mecanismos para inventariar los sistemas de IA y se asignan recursos de acuerdo con las prioridades de riesgo organizacionales.
	GOVERNANZA 1.7	Se establecen procesos y procedimientos para el desmantelamiento y eliminación gradual de sistemas de IA de manera segura y que no aumente los riesgos ni disminuya la confiabilidad de la organización.

Categorías y subcategorías para la función de GOVERNANZA (GOVERN).

Categoría	Subcategoría	Descripción
GOVERNANZA 2		Estructuras de responsabilidad están establecidas para que **los equipos y personas apropiados estén facultados**, sean responsables y estén capacitados para el mapeo, medición y gestión de riesgos de IA.
	GOVERNANZA 2.1	Los roles y responsabilidades, así como las líneas de comunicación relacionadas con el mapeo, medición y gestión de riesgos de IA están documentados y son claros para los individuos y equipos a lo largo de la organización.
	GOVERNANZA 2.2	El personal y los socios de la organización reciben capacitación en gestión de riesgos de IA para habilitarlos a desempeñar sus deberes y responsabilidades de manera consistente con las políticas, procedimientos y acuerdos relacionados.
	GOVERNANZA 2.3	El liderazgo ejecutivo de la organización asume la responsabilidad de las decisiones sobre los riesgos asociados con el desarrollo e implementación de sistemas de IA.

Categorías y subcategorías para la función de GOVERNANZA (GOVERN).

GOVERNANZA 3		Los procesos de **diversidad, equidad, inclusión y accesibilidad de la fuerza laboral** se priorizan en el mapeo, medición y gestión de los riesgos de IA a lo largo del ciclo de vida.
	GOVERNANZA 3.1	La toma de decisiones relacionada con el mapeo, medición y gestión de los riesgos de IA a lo largo del ciclo de vida está informada por un equipo diverso (por ejemplo, diversidad de demografía, disciplinas, experiencia, experticia y antecedentes).
	GOVERNANZA 3.2	Existen políticas y procedimientos para definir y diferenciar los roles y responsabilidades para las configuraciones humano-IA y la supervisión de los sistemas de IA.

Categorías y subcategorías para la función de GOVERNANZA (GOVERN).

GOVERNANZA 4		Los equipos organizacionales están **comprometidos con una cultura**
	GOVERNANZA 4.1	Las políticas y prácticas organizacionales están establecidas para fomentar un pensamiento crítico y una mentalidad de seguridad primero en el diseño, desarrollo, despliegue y uso de los sistemas de IA para minimizar impactos negativos potenciales.
	GOVERNANZA 4.2	Los equipos organizacionales documentan los riesgos e impactos potenciales de la tecnología de IA que diseñan, desarrollan, despliegan, evalúan y usan, y comunican sobre los impactos más ampliamente.
	GOVERNANZA 4.3	Existen prácticas organizacionales para permitir las pruebas de IA, la identificación de incidentes y el intercambio de información.

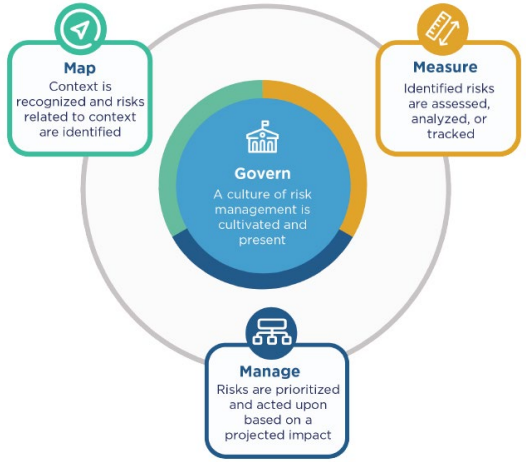
Categorías y subcategorías para la función de GOVERNANZA (GOVERN).

GOVERNANZA 5		Existen procesos para un **compromiso sólido con actores relevantes de la IA**.
	GOVERNANZA 5.1	Las políticas y prácticas organizacionales están establecidas para recopilar, considerar, priorizar e integrar la retroalimentación de aquellos externos al equipo que desarrolló o desplegó el sistema de IA respecto a los posibles impactos individuales y sociales relacionados con los riesgos de IA.
	GOVERNANZA 5.2	Se establecen mecanismos para permitir que el equipo que desarrolló o desplegó sistemas de IA incorpore regularmente la retroalimentación adjurada de actores relevantes de la IA en el diseño e implementación del sistema.

Categorías y subcategorías para la función de GOVERNANZA (GOVERN).

GOVERNANZA 6		Políticas y procedimientos están establecidos para abordar los riesgos y beneficios de IA que surgen de software de terceros y datos y otros problemas de la cadena de suministro.
	GOVERNANZA 6.1	Existen políticas y procedimientos que abordan los riesgos de IA asociados con entidades de terceros, incluidos los riesgos de infracción de la propiedad intelectual u otros derechos de terceros.
	GOVERNANZA 6.2	Se encuentran establecidos procesos de contingencia para manejar fallos o incidentes en datos de terceros o sistemas de IA considerados de alto riesgo.

Map

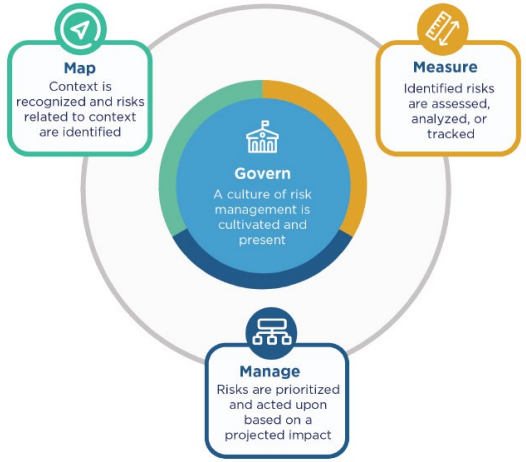


Introducción a MAP

- **Establecimiento de Contexto**: La función MAP define el contexto para enmarcar riesgos relacionados con sistemas de IA.

- **Ciclo de Vida de IA**: Consiste en actividades interdependientes que involucran a un conjunto diverso de actores.

Map

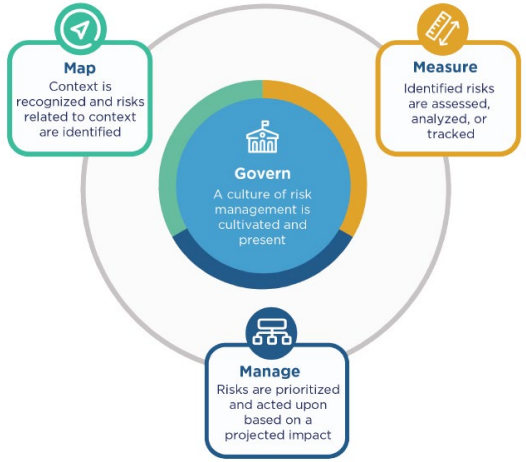


Visibilidad y Control

- **Desafíos de Visibilidad**: Los actores de IA a cargo de una parte del proceso a menudo no tienen visibilidad o control total sobre otras partes.

- **Interdependencias**: Las relaciones entre actividades y actores de IA complican la anticipación fiable de impactos.

Map

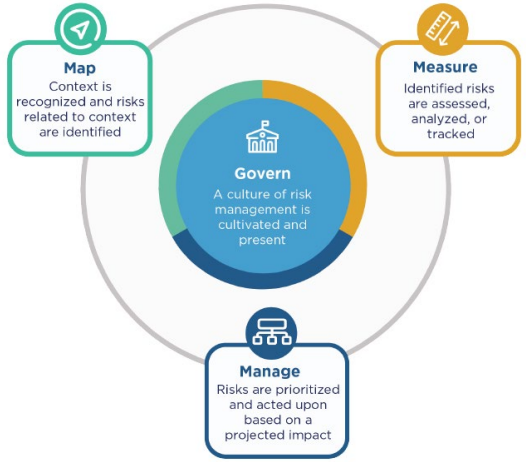


Anticipación de Impactos

- **Decisiones Tempranas**: Identificación de propósitos y objetivos de un sistema de IA puede alterar su comportamiento y capacidades.

- **Configuración de Despliegue**: Cómo las dinámicas del entorno de despliegue pueden modelar los impactos de las decisiones del sistema de IA.

Map

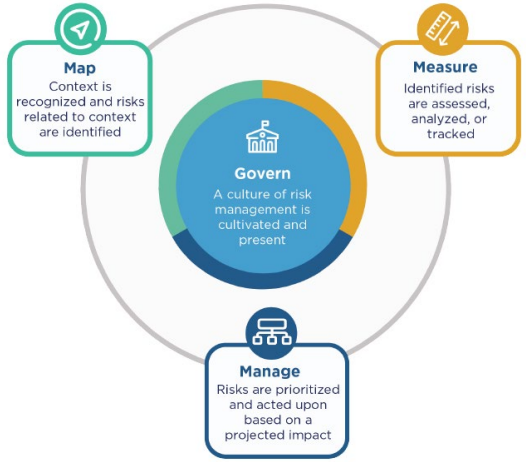


Interacciones y Condiciones

- **Interacciones Complejas**: Las mejores intenciones en una dimensión del ciclo de vida de IA pueden verse socavadas por interacciones con decisiones y condiciones en otras actividades posteriores.

- **Ejemplo Ilustrativo**: Ejemplos de cómo las decisiones tempranas pueden influir en las etapas posteriores y alterar los resultados esperados

Map

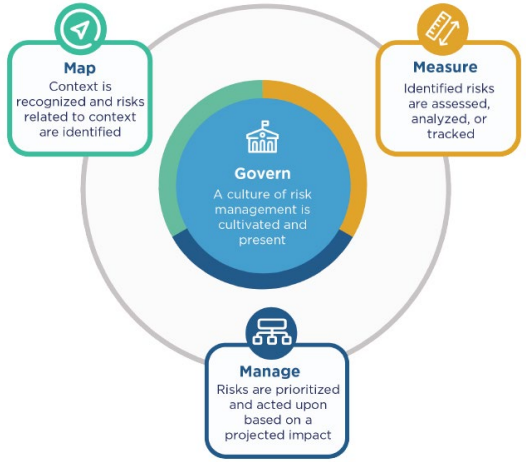


Gestión Efectiva de Riesgos

- **Enfoque Integrado**: Necesidad de un enfoque integrado y consciente de las interdependencias para gestionar eficazmente los riesgos de IA.

- **Colaboración entre Actores**: Fomento de la colaboración y la comunicación entre los actores de IA a lo largo del ciclo de vida de la IA.

Map

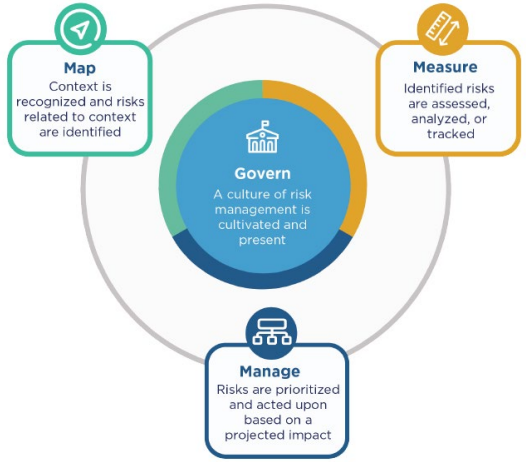


Complejidad y Visibilidad en la Gestión de Riesgos

- **Desafíos**: La complejidad y los niveles variables de visibilidad introducen incertidumbre en la gestión de riesgos.

- **Mitigación de la Incertidumbre**: Anticipación, evaluación y dirección de fuentes potenciales de riesgo negativo.

Map

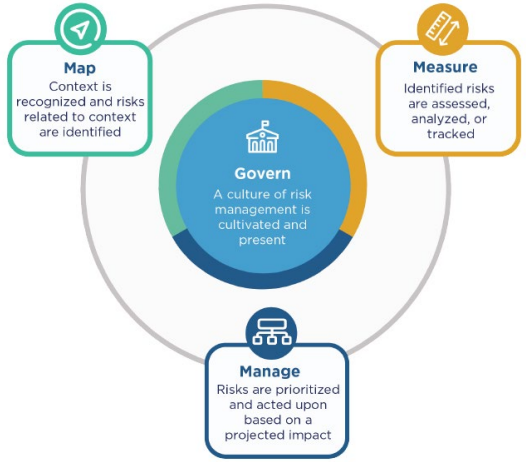


Función y Objetivos de MAP

- **Prevención de Riesgos Negativos**: Utilización de la información recopilada para prevenir riesgos negativos e informar decisiones.

- **Base para MEASURE y MANAGE**: Los resultados en MAP son esenciales para las funciones de medición y gestión de riesgos.

Map

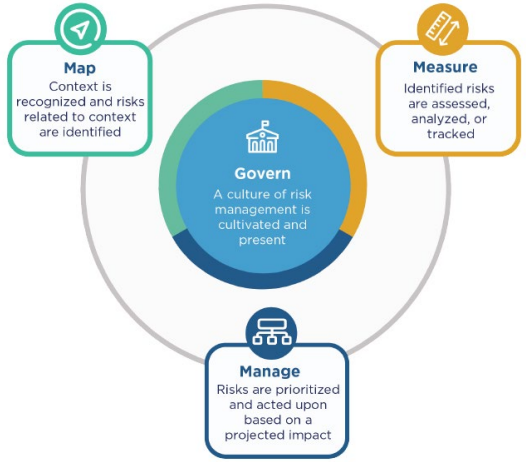


Mejora de la Capacidad Organizativa

- **Identificación de Riesgos**: La función MAP busca mejorar la habilidad de las organizaciones para identificar riesgos y factores contribuyentes.

- **Conocimiento Contextual**: La importancia del conocimiento contextual y la conciencia de riesgos para una gestión efectiva de riesgos.

Map

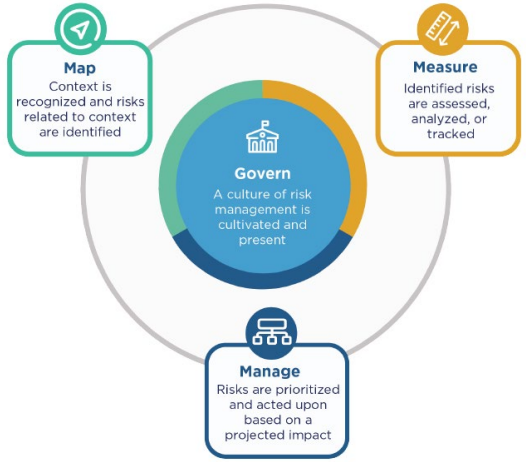


Incorporación de Perspectivas Diversas

- **Equipos Internos Diversos**: La implementación de la función MAP se enriquece con la incorporación de perspectivas de un equipo interno diverso.

- **Compromiso con Externos**: Colaboración con colaboradores externos, usuarios finales, y comunidades potencialmente impactadas.

Map

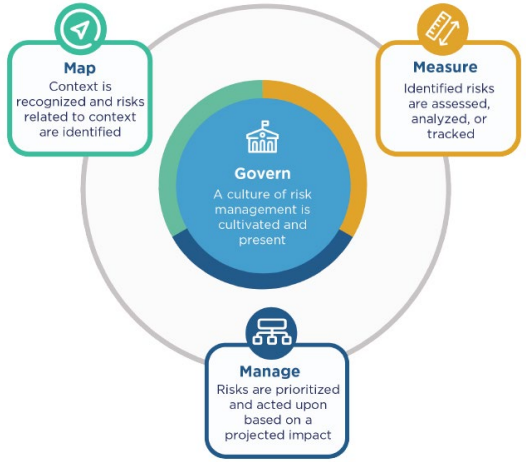


Beneficios de Perspectivas Amplias

- **Prevención Proactiva de Riesgos**: Cómo la recopilación de perspectivas amplias puede ayudar a prevenir riesgos negativos.

- **Desarrollo de Sistemas de IA Confiables**: Mejora de la capacidad de las organizaciones para desarrollar sistemas de IA más confiables

Map

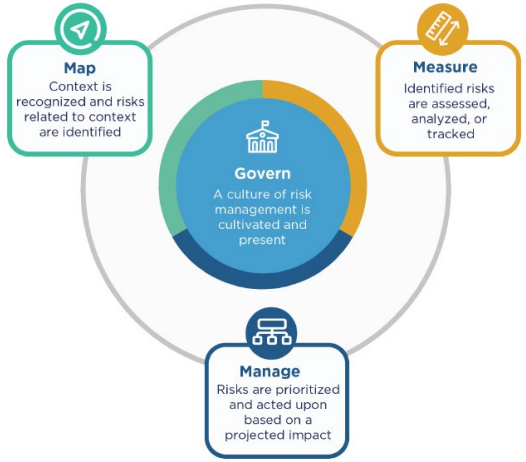


Mejoras en Comprensión y Aplicación

- **Comprensión de Contextos y Suposiciones**: Mejorar la capacidad de entender contextos y verificar suposiciones sobre el uso.

- **Reconocimiento de Funcionalidad**: Habilitación del reconocimiento de cuándo los sistemas no son funcionales dentro o fuera de su contexto intencionado.

Map

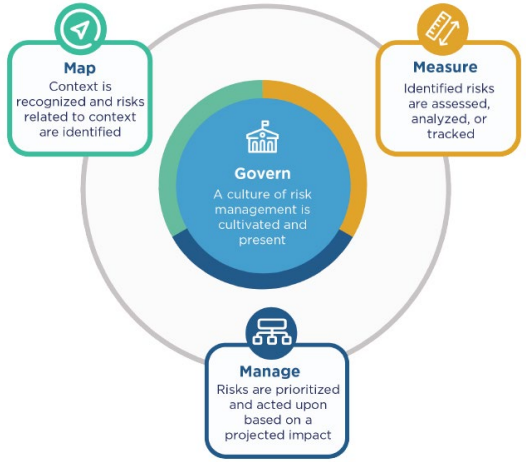


Identificación y Anticipación de Riesgos

- **Usos Positivos y Limitaciones**: Identificar usos beneficiosos y entender las limitaciones en procesos de IA y ML.

- **Impactos Negativos Previsibles**: Anticipar riesgos del uso de sistemas de IA más allá del uso intencionado.

Map

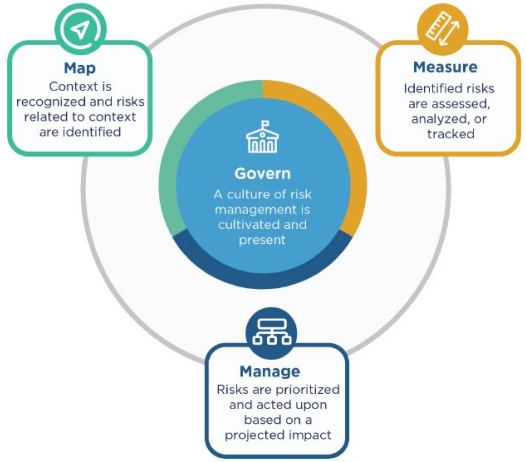


Completando la Función MAP

- **Conocimiento Contextual**: Los usuarios del marco deberían obtener suficiente conocimiento contextual sobre los impactos de los sistemas de IA.

- **Decisión Inicial Go/No-Go**: Informar sobre si proceder con el diseño, desarrollo o despliegue de un sistema de IA.

Map

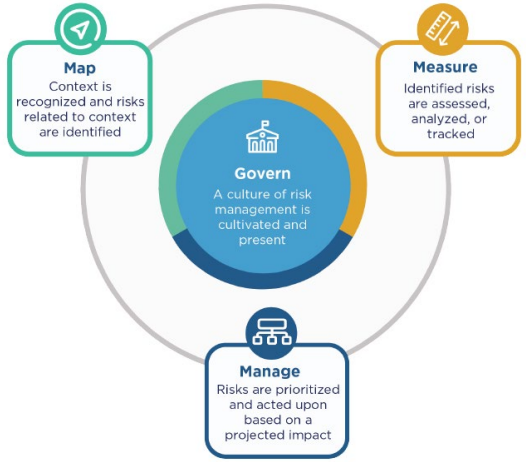


Proceder con Cautela

- **Uso de Funciones MEASURE y MANAGE**: Utilizar estas funciones junto con las políticas y procedimientos establecidos en GOVERN para gestionar riesgos de IA.

- **Importancia de la Continuidad**: Necesidad de continuar aplicando la función MAP a medida que evolucionan el contexto, las capacidades, los riesgos, los beneficios y los impactos potenciales.

Map

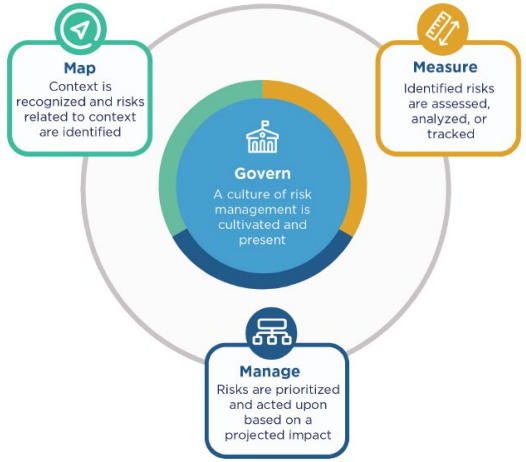


Evolución del Contexto y Capacidades

- **Adaptabilidad**: La función MAP debe ser adaptativa para reflejar cambios en el contexto y capacidades de los sistemas de IA.

- **Evaluación Continua**: Importancia de la reevaluación continua para garantizar que las decisiones sigan siendo informadas y adecuadas.

Map

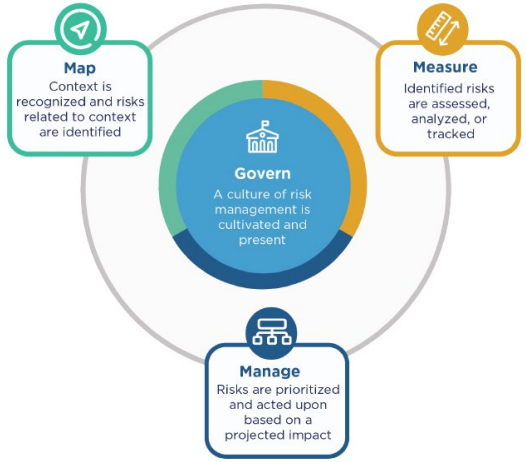


Gestión de Riesgos con MAP

- **Aplicación Continuada**: La necesidad de aplicar continuamente la función MAP a lo largo del ciclo de vida del sistema de IA.

- **Identificación de Riesgos y Beneficios**: Cómo la función MAP ayuda en la identificación y gestión de riesgos y beneficios a lo largo del tiempo.

Map



Prácticas de Mapeo en el NIST AI RMF Playbook

- **Guía Detallada**: Descripción de prácticas relacionadas con el mapeo de riesgos de IA en el NIST AI RMF Playbook.

- **Aplicación de Mejores Prácticas**: Fomento del uso de mejores prácticas y directrices para el mapeo efectivo de riesgos de IA.

Categorías y subcategorías para la función de Map.

Categoría	Subcategoría	Descripción
MAP 1		Se establece y comprende el contexto.
	MAP 1.1	Los propósitos previstos, usos potencialmente beneficiosos, leyes específicas del contexto, normas y expectativas, y los entornos prospectivos en los que se desplegará el sistema de IA se comprenden y documentan. Las consideraciones incluyen: el conjunto específico o tipos de usuarios junto con sus expectativas; impactos positivos y negativos potenciales de los usos del sistema para individuos, comunidades, organizaciones, sociedad y el planeta; suposiciones y limitaciones relacionadas sobre los propósitos, usos y riesgos del sistema de IA a lo largo del ciclo de vida del desarrollo o producto de IA; y métricas de TEVV y del sistema relacionadas.
	MAP 1.2	Los actores de IA interdisciplinarios, competencias, habilidades y capacidades para establecer el contexto reflejan diversidad demográfica y una amplia experiencia en dominios y experiencia del usuario, y su participación se documenta. Se priorizan las oportunidades de colaboración interdisciplinaria.

Categorías y subcategorías para la función de Map.

Categoría	Subcategoría	Descripción
MAP 1	**MAP 1.3**	La misión de la organización y los objetivos pertinentes para la tecnología de IA se comprenden y se documentan.
	MAP 1.4	Se ha definido claramente el valor comercial o el contexto de uso comercial o, en el caso de la evaluación de sistemas de IA existentes, se ha reevaluado.
	MAP 1.5	Se determinan y documentan las tolerancias al riesgo organizacional.
	MAP 1.6	Los requisitos del sistema (por ejemplo, "el sistema respetará la privacidad de sus usuarios") se solicitan y comprenden por los actores relevantes de IA. Las decisiones de diseño tienen en cuenta las implicaciones socio-técnicas para abordar los riesgos de IA.

Categorías y subcategorías para la función de Map.

MAP 2		Se realiza la categorización del sistema de IA.
	MAP 2.1	Se definen las tareas específicas y los métodos utilizados para implementar las tareas que el sistema de IA soportará (por ejemplo, clasificadores, modelos generativos, sistemas de recomendación).
	MAP 2.2	Se documenta información sobre los límites del conocimiento del sistema de IA y cómo los humanos pueden utilizar y supervisar la salida del sistema. La documentación proporciona información suficiente para asistir a los actores relevantes de IA al tomar decisiones y realizar acciones subsiguientes.
	MAP 2.3	Se identifican y documentan la integridad científica y las consideraciones de TEVV, incluidas las relacionadas con el diseño experimental, la recolección y selección de datos (por ejemplo, disponibilidad, representatividad, idoneidad), la confiabilidad del sistema y la validación de constructo.

Categorías y subcategorías para la función de Map.

Categoría	Subcategoría	Descripción
MAP 3		Las capacidades de IA, el uso previsto, los objetivos y los beneficios y costos esperados en comparación con referencias apropiadas se comprenden.
	MAP 3.1	Se examinan y documentan los beneficios potenciales de la funcionalidad y rendimiento del sistema de IA previsto.
	MAP 3.2	Se examinan y documentan los costos potenciales, incluidos los costos no monetarios, que resultan de errores esperados o realizados de IA o de la funcionalidad y confiabilidad del sistema – en conexión con la tolerancia al riesgo organizacional.
	MAP 3.3	Se especifica y documenta el ámbito de aplicación objetivo basado en la capacidad del sistema, el contexto establecido y la categorización del sistema de IA.
	MAP 3.4	Se definen, evalúan y documentan los procesos para la competencia de operadores y practicantes con el rendimiento y la confiabilidad del sistema de IA – y las normas técnicas y certificaciones relevantes.
	MAP 3.5	Se definen, evalúan y documentan los procesos para la supervisión humana de acuerdo con las políticas organizacionales de la función GOBIERNO.

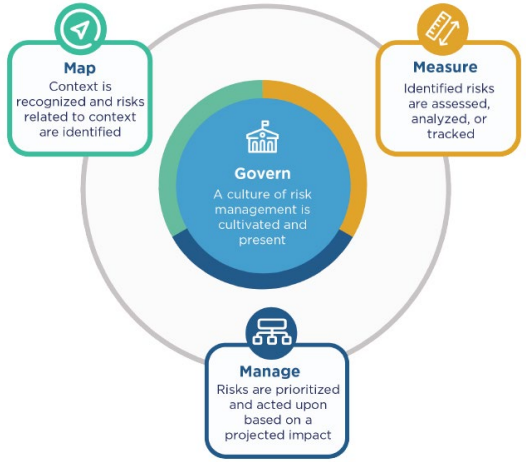
Categorías y subcategorías para la función de Map.

Categoría	Subcategoría	Descripción
MAP 4		Los riesgos y beneficios se mapean para todos los componentes del sistema de IA, incluido el software y los datos de terceros.
	MAP 4.1	Los enfoques para mapear la tecnología de IA y los riesgos legales de sus componentes – incluido el uso de datos o software de terceros – están establecidos, se siguen y documentan, al igual que los riesgos de infracción de la propiedad intelectual de un tercero u otros derechos.
	MAP 4.2	Se identifican y documentan los controles de riesgo internos para los componentes del sistema de IA, incluidas las tecnologías de IA de terceros.

Categorías y subcategorías para la función de Map.

Categoría	Subcategoría	Descripción
MAP 5		Se caracterizan los impactos a individuos, grupos, comunidades, organizaciones y la sociedad.
	MAP 5.1	Se identifican y documentan la probabilidad y magnitud de cada impacto identificado (tanto potencialmente beneficioso como perjudicial) basado en el uso esperado, usos pasados de sistemas de IA en contextos similares, informes de incidentes públicos, retroalimentación de aquellos externos al equipo que desarrolló o desplegó el sistema de IA, u otros datos.
	MAP 5.2	Se establecen y documentan prácticas y personal para apoyar el compromiso regular con actores relevantes de IA e integrar la retroalimentación sobre impactos positivos, negativos y no anticipados.

Measure

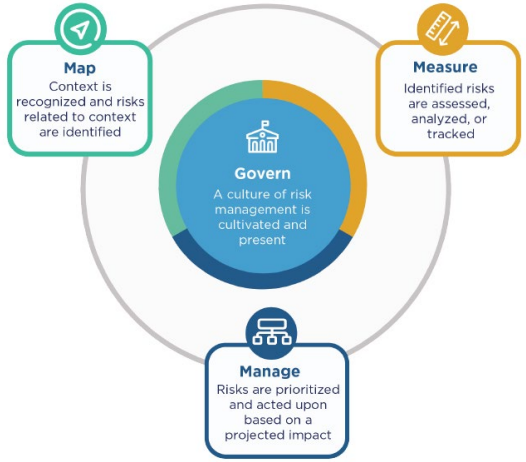


Introducción a la Función MEASURE

- **Propósito**: Utilizar herramientas y metodologías cuantitativas, cualitativas o mixtas para analizar, evaluar, y monitorear riesgos de IA y sus impactos relacionados.

- **Base de Conocimiento**: Emplea el conocimiento relevante identificado en la función MAP e informa la función MANAGE.

Measure

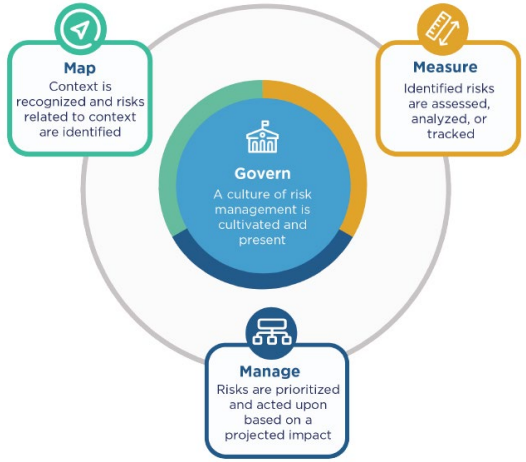


Evaluación de Riesgos de IA

- **Pruebas de Sistemas de IA**: Importancia de probar los sistemas de IA antes de su despliegue y de manera regular durante su operación.

- **Mediciones de Riesgo de IA**: Documentación de aspectos de la funcionalidad y confiabilidad de los sistemas.

Measure

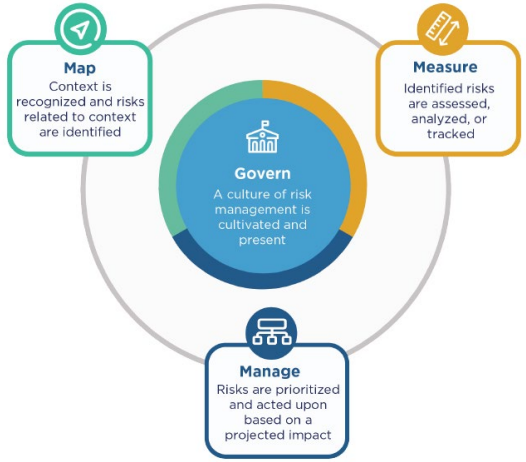


Métricas y Características Confiables

- **Seguimiento de Métricas**: Incluye métricas para características confiables, impacto social y configuraciones humano-IA.

- **Procesos de Evaluación**: Desarrollo o adopción de procesos que incluyan pruebas rigurosas de software y metodologías de evaluación de rendimiento.

Measure

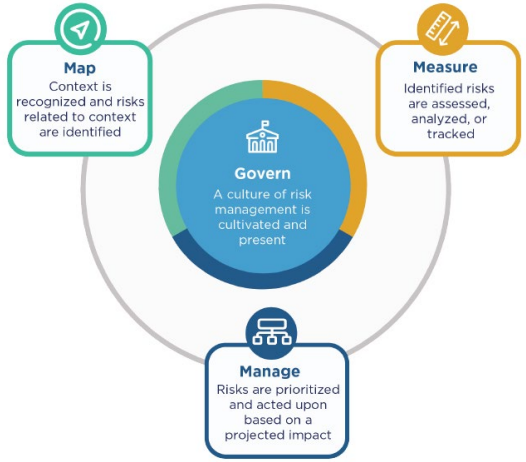


Metodologías y Documentación

- **Metodologías Rigurosas**: Pruebas de software y evaluación de rendimiento con medidas de incertidumbre y comparaciones con referencias de rendimiento.

- **Reportes Formalizados**: Documentación y reporte formal de resultados para mejorar la transparencia y la rendición de cuentas.

Measure

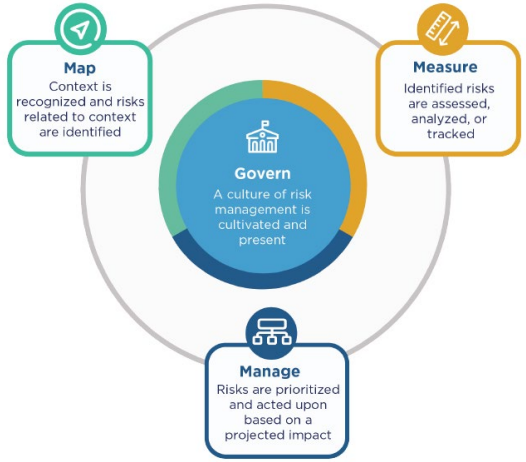


Revisión Independiente y Mitigación de Sesgos

- **Importancia de la Revisión Independiente**: Puede mejorar la efectividad de las pruebas y mitigar sesgos internos y potenciales conflictos de interés.

- **Procesos de Evaluación Independiente**: Contribuyen a una evaluación más objetiva y confiable de los riesgos y rendimiento de IA.

Measure

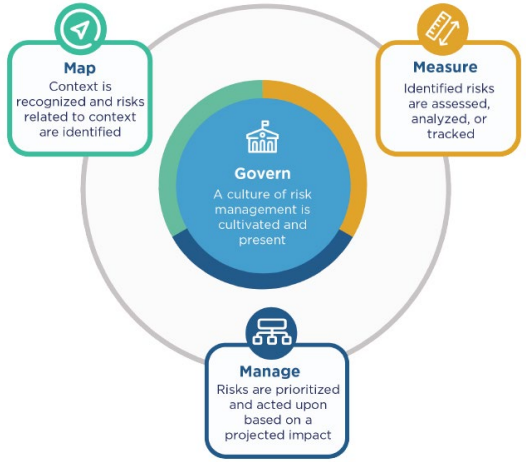


Manejo de Compensaciones

- **Compensaciones entre Características Confiables**: Cómo la medición proporciona una base trazable para informar decisiones de gestión ante compensaciones.

- **Opciones de Gestión**: Recalibración, mitigación de impacto, remoción del sistema, y controles compensatorios, detectivos, disuasorios, directivos y de recuperación.

Measure

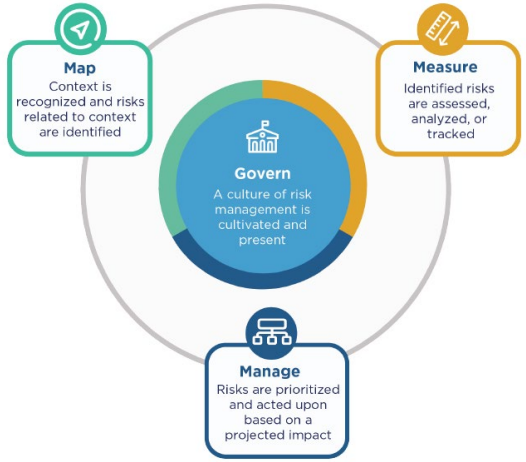


Informando Decisiones de Gestión

- **Base para Decisiones de MANAGE**: La función MEASURE informa y apoya decisiones estratégicas en la gestión de riesgos de IA.

- **Enfoque Integral**: Importancia de un enfoque integral que incorpore evaluación continua, revisión independiente, y ajuste basado en mediciones confiables.

Measure

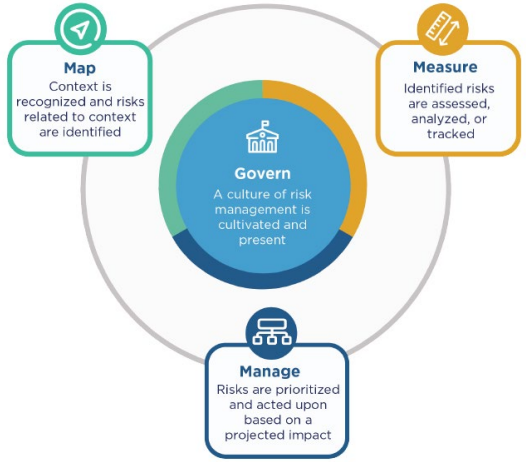


Completando la Función MEASURE

- **Implementación de TEVV**: Procesos objetivos, repetibles y escalables de prueba, evaluación, verificación y validación (TEVV) están establecidos y documentados.

- **Enfoque en Métricas y Metodologías**: Uso de métricas, métodos y metodologías siguiendo normas científicas, legales y éticas.

Measure

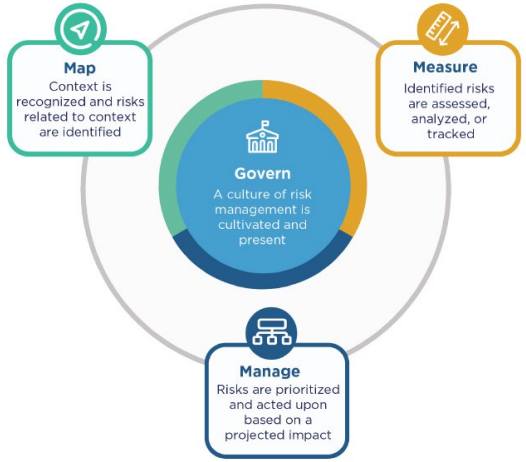


Transparencia y Desarrollo de Medidas

- **Proceso Abierto y Transparente**: Las mediciones deben realizarse de manera abierta y transparente para fomentar la confianza.

- **Desarrollo de Nuevas Medidas**: La necesidad de desarrollar tipos nuevos de medición, tanto cualitativos como cuantitativos.

Measure

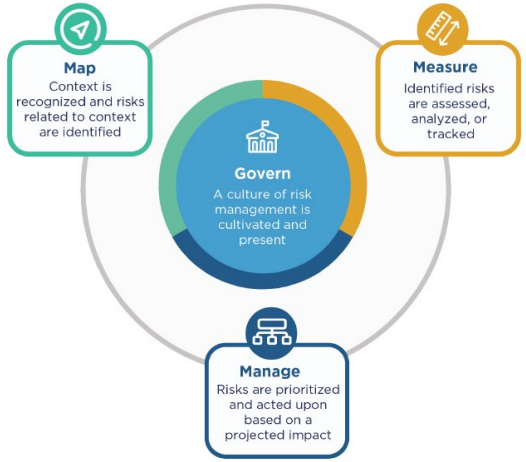


Evaluación de Tipos de Medición

- **Unicidad y Significado**: Consideración del grado en que cada tipo de medición aporta información única y significativa a la evaluación de riesgos de IA.

- **Evaluación Integral**: Mejora de la capacidad para evaluar de manera comprensiva la confiabilidad del sistema.

Measure

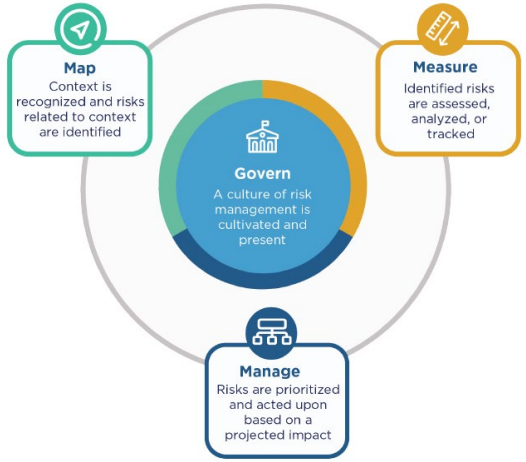


Mejora de la Capacidad de Evaluación

- **Identificación de Riesgos**: Capacidad mejorada para identificar y seguir riesgos existentes y emergentes.

- **Verificación de Eficacia de Métricas**: Verificación de la eficacia de las métricas utilizadas en la evaluación de riesgos.

Measure

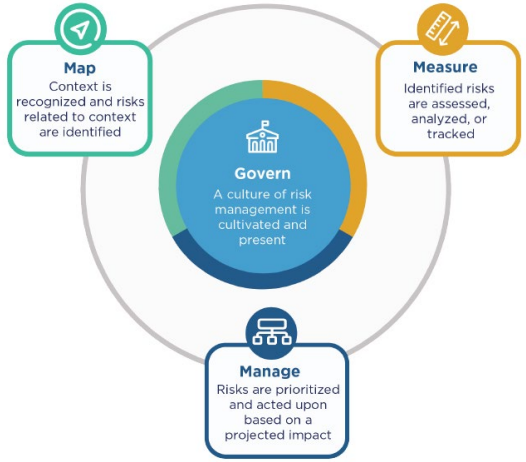


Utilización de Resultados de Medición

- **Soporte a la Función MANAGE**: Los resultados de la medición se utilizan para asistir en los esfuerzos de monitoreo de riesgos y respuesta.

- **Monitoreo y Respuesta de Riesgos**: Importancia de una gestión informada por mediciones objetivas y repetibles.

Measure



Aplicación Continua de MEASURE

- **Evolución del Conocimiento y Metodologías**: Necesidad de continuar aplicando la función MEASURE a medida que evolucionan el conocimiento, las metodologías, los riesgos y los impactos.

- **Adaptabilidad y Mejora Continua**: Compromiso con la adaptabilidad y mejora continua en la evaluación de riesgos de IA.

Measure

Categoría	Subcategoría	Descripción
MEDIR 1		Se identifican y aplican métodos y métricas apropiados.
	MEDIR 1.1	Se seleccionan para implementación los enfoques y métricas para la medición de riesgos de IA enumerados durante la función MAP, comenzando con los riesgos de IA más significativos. Los riesgos o características de confiabilidad que no se medirán – o no pueden ser medidos – se documentan adecuadamente.
	MEDIR 1.2	La idoneidad de las métricas de IA y la efectividad de los controles existentes se evalúan y actualizan regularmente, incluyendo informes de errores y posibles impactos en las comunidades afectadas.
	MEDIR 1.3	Expertos internos que no sirvieron como desarrolladores principales del sistema y/o evaluadores independientes están involucrados en evaluaciones y actualizaciones regulares. Se consulta a expertos en el dominio, usuarios, actores de IA externos al equipo que desarrolló o desplegó el sistema de IA, y a las comunidades afectadas en apoyo de las evaluaciones según sea necesario de acuerdo con la tolerancia al riesgo organizacional.

Measure

Categoría	Subcategoría	Descripción
MEDIR 2		Los sistemas de IA son evaluados por características de confiabilidad.
	MEDIR 2.1	Se documentan los conjuntos de pruebas, métricas y detalles sobre las herramientas utilizadas durante la TEVV.
	MEDIR 2.2	Las evaluaciones que involucran sujetos humanos cumplen con los requisitos aplicables (incluida la protección del sujeto humano) y son representativas de la población relevante.
	MEDIR 2.3	El rendimiento del sistema de IA o los criterios de aseguramiento se miden cualitativa o cuantitativamente y se demuestran para condiciones similares a los entornos de despliegue. Las medidas se documentan.
	MEDIR 2.4	La funcionalidad y el comportamiento del sistema de IA y sus componentes – como se identificaron en la función MAP – se monitorean cuando están en producción.
	MEDIR 2.5	Se demuestra que el sistema de IA a desplegar es válido y confiable. Se documentan las limitaciones de la generalización más allá de las condiciones bajo las cuales se desarrolló la tecnología.

Measure

	MEDIR 2.6	El sistema de IA se evalúa regularmente por riesgos de seguridad – como se identificó en la función MAP. Se demuestra que el sistema de IA a desplegar es seguro, su riesgo negativo residual no excede la tolerancia al riesgo y puede fallar de manera segura, particularmente si se hace operar más allá de sus límites de conocimiento. Las métricas de seguridad reflejan la confiabilidad y robustez del sistema, el monitoreo en tiempo real y los tiempos de respuesta para fallos del sistema de IA.
	MEDIR 2.7	La seguridad y resiliencia del sistema de IA – como se identificó en la función MAP – se evalúan y documentan.
	MEDIR 2.8	Se examinan y documentan los riesgos asociados con la transparencia y responsabilidad – como se identificó en la función MAP.
	MEDIR 2.9	El modelo de IA se explica, valida y documenta, y la salida del sistema de IA se interpreta dentro de su contexto – como se identificó en la función MAP – para informar el uso responsable y la gobernanza.
	MEDIR 2.10	Se examina y documenta el riesgo de privacidad del sistema de IA – como se identificó en la función MAP.
	MEDIR 2.11	La equidad y el sesgo – como se identificó en la función MAP – se evalúan y los resultados se documentan.
	MEDIR 2.12	El impacto ambiental y la sostenibilidad de las actividades de entrenamiento y gestión del modelo de IA – como se identificó en la función MAP – se evalúan y documentan.
	MEDIR 2.13	La efectividad de las métricas y procesos TEVV empleados en la función MEDIR se evalúan y documentan.

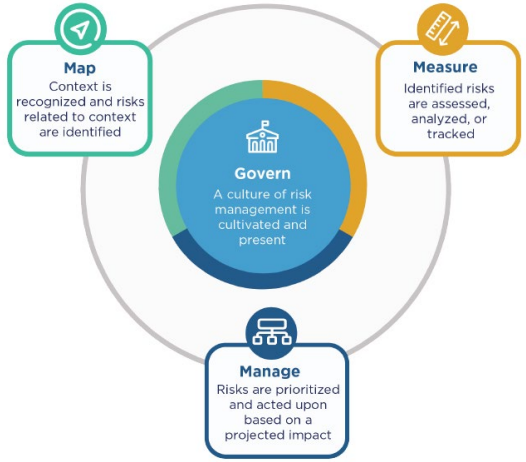
Measure

Categoría	Subcategoría	Descripción
MEDIR 3		Mecanismos para rastrear los riesgos de IA identificados a lo largo del tiempo están establecidos.
	MEDIR 3.1	Se disponen de enfoques, personal y documentación para identificar y rastrear regularmente los riesgos de IA existentes, no anticipados y emergentes basados en factores como el rendimiento previsto y real en contextos desplegados.
	MEDIR 3.2	Los enfoques de rastreo de riesgos se consideran para entornos donde los riesgos de IA son difíciles de evaluar utilizando las técnicas de medición actualmente disponibles o donde aún no hay métricas disponibles.
	MEDIR 3.3	Se establecen y integran en las métricas de evaluación del sistema de IA procesos de retroalimentación para usuarios finales y comunidades impactadas para informar problemas y apelar los resultados del sistema.

Measure

Categoría	Subcategoría	Descripción
MEDIR 4		La retroalimentación sobre la eficacia de la medición es recopilada y evaluada.
	MEDIR 4.1	Los enfoques de medición para identificar riesgos de IA están conectados a los contextos de despliegue e informados mediante la consulta con expertos en el dominio y otros usuarios finales. Los enfoques están documentados.
	MEDIR 4.2	Los resultados de medición respecto a la confiabilidad del sistema de IA en contextos de despliegue y a lo largo del ciclo de vida de la IA están informados por la entrada de expertos en el dominio y actores relevantes de la IA para validar si el sistema está funcionando de manera consistente según lo previsto. Los resultados están documentados.
	MEDIR 4.3	Mejoras o declives en el rendimiento medible basados en consultas con actores relevantes de la IA, incluyendo comunidades afectadas, y datos de campo sobre riesgos relevantes al contexto y características de confiabilidad son identificados y documentados.

Manage

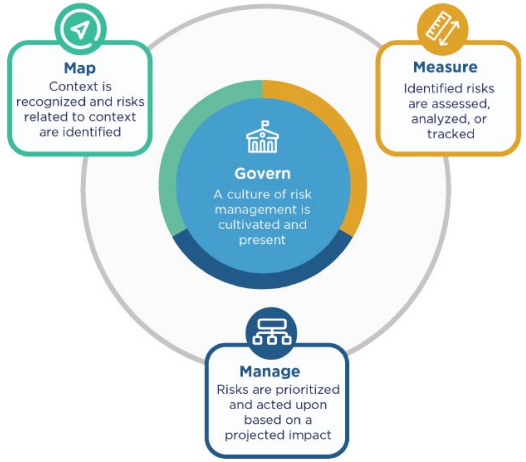


Introducción a la Función MANAGE

- **Asignación de Recursos de Riesgo**: Gestión y asignación de recursos a los riesgos identificados y medidos, siguiendo las directrices de GOVERN.

- **Tratamiento de Riesgos**: Implementación de planes para responder, recuperar y comunicar incidentes o eventos.

Manage

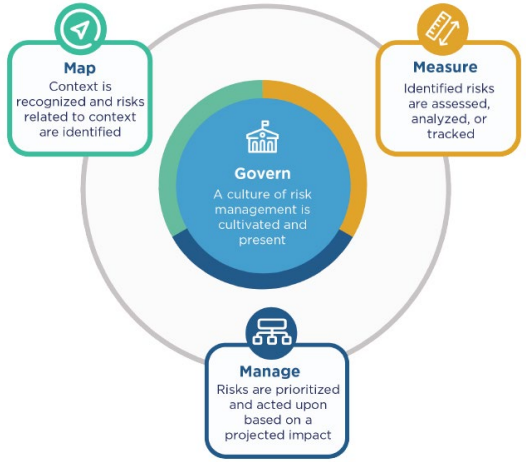


Uso de Información Contextual

- **Consultas de Expertos y Entrada de Actores de IA**: La información contextual obtenida es crucial para reducir la probabilidad de fallos del sistema e impactos negativos.

- **Fundamentos en GOVERN y MAP**: La información establecida y ejecutada a través de estas funciones es vital para la función MANAGE.

Manage

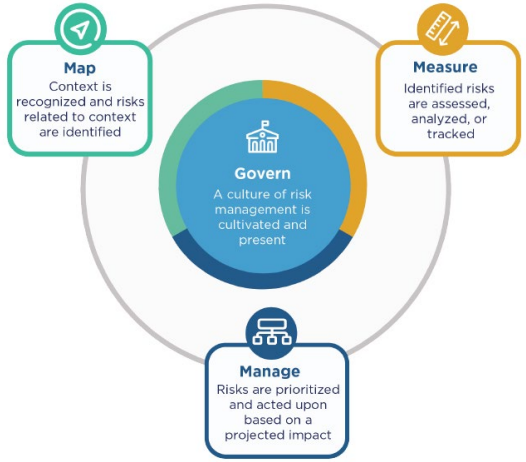


Documentación Sistemática y Transparencia

- **Prácticas de Documentación**: Establecidas en GOVERN y empleadas en MAP y MEASURE para reforzar la gestión de riesgos de IA.

- **Aumento de la Transparencia y Responsabilidad**: A través de una documentación sistemática y prácticas de gestión de riesgos.

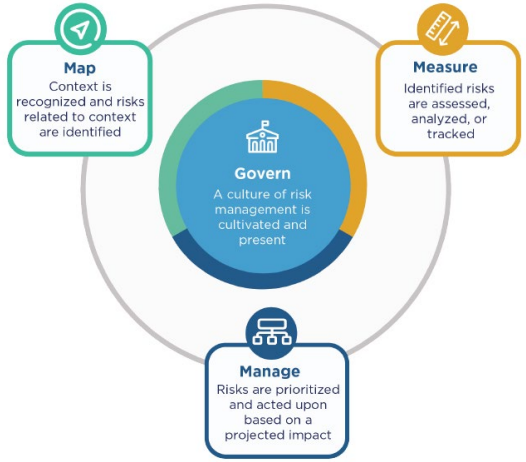
Manage



Evaluación de Riesgos Emergentes

- **Mecanismos de Mejora Continua**: Procesos para evaluar riesgos emergentes y mecanismos establecidos para la mejora continua en la gestión de riesgos.

Manage

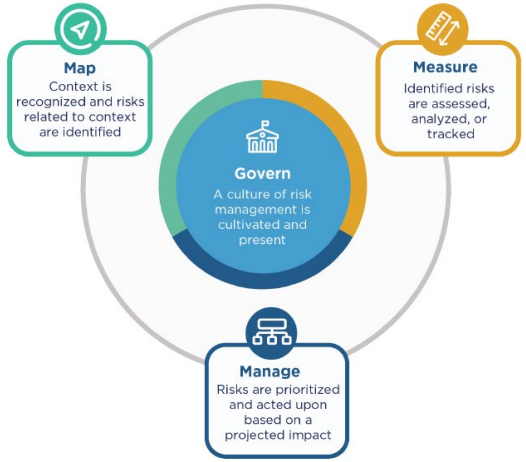


Planificación y Monitoreo Post-MANAGE

- **Priorización y Monitoreo de Riesgos**: Desarrollo de planes para la priorización de riesgos y el monitoreo regular y la mejora de procesos.

- **Capacidad Mejorada para Gestionar Riesgos**: Usuarios del marco tendrán una capacidad mejorada para gestionar los riesgos de sistemas de IA desplegados.

Manage



Aplicación Continua de MANAGE

- **Adaptabilidad ante Evolución**: Necesidad de continuar aplicando la función MANAGE conforme evolucionan métodos, contextos, riesgos, y las necesidades o expectativas de los actores de IA relevantes.

- **Compromiso con la Gestión de Riesgos**: Es fundamental para los usuarios del marco adaptarse y responder a los cambios de manera efectiva.

Manage

Categoría	Subcategoría	Descripción
GESTIONAR 1		Los riesgos de IA basados en evaluaciones y otros resultados analíticos de las funciones MAPA y MEDIR son priorizados, respondidos y gestionados.
	GESTIONAR 1.1	Se determina si el sistema de IA alcanza sus propósitos previstos y objetivos declarados y si su desarrollo o despliegue debe continuar.
	GESTIONAR 1.2	El tratamiento de riesgos de IA documentados se prioriza basado en el impacto, la probabilidad y los recursos o métodos disponibles.
	GESTIONAR 1.3	Se desarrollan, planifican y documentan respuestas a los riesgos de IA considerados de alta prioridad, según lo identificado por la función MAPA. Las opciones de respuesta al riesgo pueden incluir mitigar, transferir, evitar o aceptar.
	GESTIONAR 1.4	Se documentan los riesgos residuales negativos (definidos como la suma de todos los riesgos no mitigados) tanto para los adquirientes de sistemas de IA como para los usuarios finales.

Manage

Categoría	Subcategoría	Descripción
GESTIONAR 2		Estrategias para maximizar los beneficios de la IA y minimizar los impactos negativos son planificadas, preparadas, implementadas, documentadas e informadas por la entrada de actores relevantes de la IA.
	GESTIONAR 2.1	Se tienen en cuenta los recursos necesarios para gestionar los riesgos de la IA, junto con sistemas, enfoques o métodos alternativos viables no basados en IA, para reducir la magnitud o la probabilidad de impactos potenciales.
	GESTIONAR 2.2	Se establecen y aplican mecanismos para sostener el valor de los sistemas de IA desplegados.
	GESTIONAR 2.3	Se siguen procedimientos para responder y recuperarse de un riesgo previamente desconocido cuando se identifica.
	GESTIONAR 2.4	Se establecen y aplican mecanismos, y se asignan y comprenden responsabilidades, para reemplazar, desactivar o desactivar sistemas de IA que demuestren un rendimiento o resultados inconsistentes con el uso previsto.

Manage

Categoría	Subcategoría	Descripción
GESTIONAR 3		Los riesgos y beneficios de IA provenientes de entidades de terceros se gestionan.
	GESTIONAR 3.1	Los riesgos y beneficios de IA provenientes de recursos de terceros se monitorean regularmente, y se aplican y documentan controles de riesgo.
	GESTIONAR 3.2	Los modelos preentrenados que se utilizan para el desarrollo se monitorean como parte del monitoreo regular y mantenimiento del sistema de IA.

Manage

Categoría	Subcategoría	Descripción
GESTIONAR 4		Los tratamientos de riesgo, incluyendo respuesta y recuperación, y planes de comunicación para los riesgos de IA identificados y medidos están documentados y se monitorean regularmente.
	GESTIONAR 4.1	Se implementan planes de monitoreo de sistemas de IA post-despliegue, incluyendo mecanismos para capturar y evaluar la entrada de usuarios y otros actores relevantes de la IA, apelación y anulación, desmantelamiento, respuesta a incidentes, recuperación y gestión de cambios.
	GESTIONAR 4.2	Actividades medibles para mejoras continuas se integran en las actualizaciones del sistema de IA e incluyen compromiso regular con partes interesadas, incluyendo actores relevantes de la IA.
	GESTIONAR 4.3	Incidentes y errores se comunican a actores relevantes de la IA, incluidas las comunidades afectadas. Se siguen y documentan procesos para rastrear, responder y recuperarse de incidentes y errores.

Perfiles

Introducción a los Perfiles del AI RMF

- **Definición**: Los perfiles del AI RMF son implementaciones específicas del marco adaptadas a entornos o aplicaciones particulares.

- **Objetivo**: Ayudar a las organizaciones a gestionar los riesgos de IA alineados con sus objetivos, considerando requisitos legales/regulatorios y mejores prácticas.

Perfiles

Tipos de Perfiles del AI RMF

- **Perfiles de Caso de Uso**: Ejemplos incluyen perfiles de contratación o de vivienda justa, adaptados a aplicaciones específicas.

- **Perfiles Temporales**: Descripciones del estado actual o del estado objetivo deseado de la gestión de riesgos de IA en un contexto dado.

Perfiles

Comparación de Perfiles Actuales y Objetivos

- **Identificación de Brechas**: La comparación revela brechas que deben abordarse para cumplir con los objetivos de gestión de riesgos de IA.

- **Planes de Acción**: Desarrollo de planes para abordar estas brechas y cumplir con los resultados deseados.

Perfiles

Perfiles Transversales

- **Cobertura de Riesgos**: Abordan riesgos de modelos o aplicaciones usables en múltiples casos o sectores.

- **Gestión de Riesgos Comunes**: Incluye cómo gobernar, mapear, medir y gestionar riesgos en procesos de negocio comunes a varios sectores.

Perfiles

Flexibilidad en la Implementación

- **Sin Plantillas Prescriptivas**: El marco permite flexibilidad, no prescribiendo plantillas de perfil para permitir una implementación adaptativa.

- **Enfoque Basado en Riesgos**: Permite a los usuarios del marco comparar sus enfoques y evaluar los recursos necesarios de manera coste-efectiva.

Certificación Generative AI Professional

Beneficios

- Ampliación de conocimientos en IA, preparación para el manejo de proyectos de IA, capacidad para evaluar el impacto de la IA en el lugar de trabajo, mejora de la competitividad y eficiencia operativa.

Habilidades Desarrolladas

- Comprensión de modelos de IA, fundamentos del aprendizaje automático y profundo, identificación y aplicación de IA en diversos contextos, gestión de proyectos de IA, evaluación de preocupaciones de seguridad en IA.

https://certiprof.com/collections/new-technologies-certifications/products/generative-ai-professional-certification-gaipc

Descripciones de tareas de Actores de IA

AIRMPC™ Versión 032024

Diseño de IA: Ubicación y Actividades Clave

•**1. Ubicación en el Ciclo de Vida de la IA**

•**Fases de Contexto de Aplicación y Datos e Insumos**: El Diseño de IA se lleva a cabo durante estas etapas esenciales.

•**2. Creación de Conceptos y Objetivos**

•**Desarrollo del Sistema**: Los actores de IA definen el concepto y los objetivos del sistema de IA, asegurando su legalidad y adecuación para el propósito previsto.

•**3. Principales Actividades de Diseño de IA**

•**Planificación y Diseño**: Establecimiento de la base para el sistema de IA, incluyendo la arquitectura y los objetivos funcionales.

•**Recolección y Procesamiento de Datos**: Identificación, limpieza y documentación de datos necesarios para el entrenamiento y operación del sistema.

•**Documentación**: Articulación de conceptos, objetivos, suposiciones subyacentes, contexto y requisitos del sistema.

•**4. Actores de IA Involucrados**

•Incluyen, entre otros, científicos de datos, expertos en dominios específicos, analistas socio-culturales, y expertos en diversidad, equidad, inclusión y accesibilidad.

•**5. Resultados Esperados**

•Un sistema de IA diseñado de manera integral que refleje una comprensión profunda de su aplicación, con todos los datos y metadatos necesarios bien documentados y preparados.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Desarrollo de IA: Ubicación y Actividades Principales

•**1. Ubicación en el Ciclo de Vida de la IA**

•**Fase de Modelo de IA**: Las tareas de Desarrollo de IA se llevan a cabo durante esta fase clave del ciclo de vida.

•**2. Infraestructura Inicial y Construcción de Modelos**

•**Establecimiento de Infraestructura**: Provisión de la infraestructura inicial necesaria para sistemas de IA.

•**Creación y Selección de Modelos**: Involucra la creación, selección, calibración, entrenamiento y/o prueba de modelos o algoritmos.

•**3. Actores de IA en el Desarrollo**

•Incluyen expertos en aprendizaje automático, científicos de datos, desarrolladores, entidades de terceros, y expertos en gobernanza legal y de privacidad.

•Expertos en factores socio-culturales y contextuales asociados con el entorno de despliegue.

•**4. Objetivos del Desarrollo de IA**

•**Interpretación de Modelos**: Responsables de las tareas de construcción e interpretación de modelos.

•**Asegurar Conformidad**: Garantizar que los modelos sean efectivos, éticos y cumplan con regulaciones legales y de privacidad.

•**5. Resultados Esperados**

•Desarrollo de modelos de IA robustos, confiables y adecuados para su propósito, reflejando una comprensión profunda de su aplicación y contexto.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Despliegue de IA: Ubicación y Actividades Clave

•**1.Tarea y Fase de Salida**: Las tareas de Despliegue de IA se realizan durante esta fase del ciclo de vida.

•**2. Actividades Principales**

•**Piloto y Compatibilidad**: Piloteo del sistema y verificación de compatibilidad con sistemas existentes.

•**Cumplimiento Regulatorio**: Aseguramiento del cumplimiento con las regulaciones aplicables.

•**Gestión del Cambio Organizacional**: Adaptación de la organización al nuevo sistema de IA.

•**Evaluación de la Experiencia del Usuario**: Verificación de la usabilidad y aceptación del sistema por parte de los usuarios.

•**3. Actores de IA Involucrados**

•Incluyen integradores de sistemas, desarrolladores de software, usuarios finales, operadores, evaluadores y expertos en dominios específicos.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Operación y Monitoreo de IA: Fases y Actividades Clave

- **1. Ubicación en el Ciclo de Vida**

- **Fase de Contexto de Aplicación/Operar y Monitorear**: Las tareas se realizan durante esta etapa específica.

- **2. Actividades Principales**

- **Operación del Sistema**: Responsables de la operación continua del sistema de IA.

- **Evaluación Regular**: Colaboración para evaluar regularmente la salida y los impactos del sistema.

- **3. Actores de IA Involucrados**

- Incluyen operadores de sistemas, expertos en dominios, diseñadores de IA, usuarios, desarrolladores de productos, evaluadores y auditores, expertos en cumplimiento, gestión organizacional y miembros de la comunidad de investigación.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Test, Evaluation, Verification, and Validation (TEVV) - Tareas TEVV en el Ciclo de Vida de la IA

•**1. Cobertura en el Ciclo de Vida**

•**A lo Largo del Ciclo de Vida de la IA**: Las tareas TEVV se realizan en todas las fases, desde el diseño hasta la operación.

•**2. Principales Actividades TEVV**

•**Diseño y Planificación**: Validación interna y externa de suposiciones para el diseño del sistema, recolección de datos y mediciones.

•**Desarrollo (Construcción de Modelos)**: Incluye la validación y evaluación de modelos.

•**Despliegue**: Validación de sistema e integración en producción, pruebas y recalibración para integración de sistemas y procesos, experiencia de usuario y cumplimiento de especificaciones legales, regulatorias y éticas.

•**Operaciones**: Monitoreo continuo para actualizaciones periódicas, pruebas y recalibración por expertos en la materia, seguimiento y gestión de incidentes o errores reportados.

•**3. Actores de IA en TEVV**

•Examinan el sistema de IA o sus componentes y detectan y remedian problemas. Idealmente, los actores que realizan verificación y validación son distintos de aquellos que llevan a cabo pruebas y evaluaciones.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Factores Humanos en el Ciclo de Vida de la IA

•**1. Ubicación en el Ciclo de Vida**

•**A lo Largo de Todas las Dimensiones**: Las tareas de Factores Humanos se encuentran integradas en cada fase del ciclo de vida de la IA.

•**2. Actividades Clave**

•**Diseño Centrado en el Humano**: Aplicación de prácticas y metodologías que ponen a las personas al centro del proceso de diseño.

•**Involucramiento Activo de Usuarios Finales**: Promoción de la participación activa de usuarios finales y otras partes interesadas.

•**Incorporación de Normas y Valores Específicos**: Integración de normas y valores contextuales en el diseño del sistema.

•**Evaluación y Adaptación de Experiencias de Usuario**: Ajuste continuo de la experiencia del usuario para mejorar la interacción y satisfacción.

•**Integración de Dinámicas Humanas**: Inclusión amplia de humanos y dinámicas humanas en todas las fases.

•**3. Contribución de Profesionales de Factores Humanos**

•Ofrecen habilidades multidisciplinarias para entender el contexto de uso, promover la diversidad, diseñar y evaluar la experiencia del usuario, y realizar evaluaciones centradas en el humano.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Rol de los Expertos de Dominio en IA

•1. Aportación de Expertos

•Entrada Multidisciplinaria: Los expertos de dominio ofrecen conocimientos especializados sobre un sector industrial, económico, contexto o área de aplicación específica.

•2. Guía Esencial en Diseño y Desarrollo

•Diseño de Sistemas de IA: Proporcionan orientación crucial para el diseño y desarrollo de sistemas de IA, asegurando relevancia y aplicabilidad.

•Interpretación de Salidas: Apoyan la interpretación de los resultados del sistema de IA, mejorando la utilidad y precisión de estas.

•3. Apoyo a Equipos TEVV e Impacto de IA

•Colaboración con TEVV: Contribuyen al proceso de Prueba, Evaluación, Verificación y Validación mediante su expertise.

•Evaluación de Impacto de IA: Ofrecen perspectivas valiosas para la evaluación de impacto de IA, enriqueciendo el análisis con su conocimiento del dominio.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Evaluación de Impacto de IA: Actividades Clave

•**1. Objetivo de la Evaluación**

•**Análisis Integral**: Evaluación de los requerimientos para la rendición de cuentas de sistemas de IA, combate contra sesgos perjudiciales, seguridad del producto, responsabilidad y seguridad.

•**2. Áreas de Enfoque**

•**Responsabilidad de Sistemas de IA**: Asegurar que los sistemas sean transparentes y responsables.

•**Combate contra Sesgos Nocivos**: Identificación y mitigación de sesgos para promover la equidad.

•**Seguridad del Producto y Responsabilidad**: Evaluación de la seguridad de los productos de IA y cuestiones relacionadas con la responsabilidad.

•**Seguridad**: Análisis de las medidas de seguridad implementadas para proteger los sistemas de IA.

•**3. Actores de IA Involucrados**

•Incluyen evaluadores de impacto y expertos técnicos, humanos, socio-culturales y legales

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Adquisición de IA: Proceso y Responsabilidades

•**1. Objetivo de la Adquisición**

•**Adquisición Responsable**: Realización de tareas por actores de IA con autoridad financiera, legal o de gestión de políticas para la adquisición de modelos, productos o servicios de IA.

•**2. Enfoque de las Tareas**

•**Selección de Desarrolladores**: Elección de desarrolladores de terceros, vendedores o contratistas adecuados para proporcionar soluciones de IA.

•**Evaluación Legal y de Políticas**: Asegurar que todas las adquisiciones cumplan con las regulaciones legales aplicables y las políticas internas.

•**Gestión Financiera**: Administración de los recursos financieros asignados para la adquisición de tecnologías de IA.

•**3. Actores de IA Involucrados**

•Incluyen aquellos con autoridad en finanzas, legalidad y gestión de políticas implicadas en el proceso de adquisición.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Gobernanza y Supervisión de IA: Funciones y Actores Clave

•**1. Objetivo de Gobernanza y Supervisión**

•**Aseguramiento de Responsabilidad**: Ejecución de tareas por actores de IA con autoridad de gestión, fiduciaria y legal para supervisar el diseño, desarrollo y despliegue de sistemas de IA.

•**2. Principales Responsabilidades**

•**Establecimiento de Directrices**: Desarrollo y aplicación de políticas y procedimientos para el uso ético y responsable de la IA.

•**Monitoreo de Impacto y Sostenibilidad**: Evaluación continua del impacto de los sistemas de IA en la organización y su sostenibilidad a largo plazo.

•**Cumplimiento Legal y Ético**: Asegurar que las actividades de IA cumplan con todas las regulaciones legales y principios éticos.

•**3. Actores Clave en la Gobernanza de IA**

•Incluyen la gestión organizacional, liderazgo senior y la Junta Directiva, quienes tienen un interés directo en el impacto y la sostenibilidad de la organización.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

El Papel de las Entidades de Terceros en IA

•**1. Quiénes son**

•**Proveedores de Componentes Clave**: Incluyen proveedores, desarrolladores, vendedores y evaluadores de datos, algoritmos, modelos, sistemas y servicios relacionados.

•**2. Responsabilidades**

•**Contribución al Diseño y Desarrollo de IA**: Responsables de tareas de diseño y desarrollo de IA, total o parcialmente, para otras organizaciones o sus clientes.

•**3. Características Específicas**

•**Externos al Equipo Interno**: Definidos por su posición externa al equipo de diseño, desarrollo o despliegue de la organización que adquiere sus tecnologías o servicios.

•**Complejidad y Opacidad**: Las tecnologías adquiridas pueden ser complejas o poco claras, y las tolerancias al riesgo pueden no alinearse con la organización que despliega u opera el sistema de IA.

Key Dimensions	Application Context	Data & Input	AI Model	AI Model	Task & Output	Application Context	People & Planet
Lifecycle Stage	Plan and Design	Collect and Process Data	Build and Use Model	Verify and Validate	Deploy and Use	Operate and Monitor	Use or Impacted by
TEVV	TEVV includes audit & impact assessment	TEVV includes internal & external validation	TEVV includes model testing	TEVV includes model testing	TEVV includes integration, compliance testing & validation	TEVV includes audit & impact assessment	TEVV includes audit & impact assessment
Activities	Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations.	Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations.	Create or select algorithms; train models.	Verify & validate, calibrate, and interpret model output.	Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience.	Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations.	Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights.
Representative Actors	System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators.	Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts.	Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts.		System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts,	System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators.	End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers.

Actores Adicionales de IA

1. **Usuarios Finales**
   - •Individuos o grupos que utilizan el sistema de IA para propósitos específicos.
2. **Individuos/Comunidades Afectadas**
   - •Todos aquellos directa o indirectamente afectados por sistemas de IA o decisiones basadas en la salida de estos sistemas.
3. **Otros Actores de IA**
   - •Incluyen asociaciones comerciales, organizaciones de desarrollo de estándares, grupos de defensa, investigadores, grupos ambientales y organizaciones de la sociedad civil.
4. **Público General**
   - •Quienes probablemente experimenten los impactos positivos y negativos de las tecnologías de IA. Incluye individuos, comunidades y consumidores relacionados con el contexto de desarrollo o despliegue de sistemas de IA.

Cómo los Riesgos de IA Difieren de los Riesgos de Software Tradicionales

Listado de Riesgos Específicos de IA

1. **Representación Inadecuada de Datos**: Posibles sesgos y problemas de calidad.
2. **Dependencia y Complejidad de Datos**: Volumen y complejidad crecientes.
3. **Cambios Durante el Entrenamiento**: Pueden alterar el rendimiento del sistema.
4. **Desactualización de Datos**: Pérdida de relevancia respecto al contexto de despliegue.
5. **Escala y Complejidad del Sistema**: Integración en aplicaciones de software más tradicionales.
6. **Modelos Preentrenados**: Aumentan incertidumbre estadística y problemas de sesgo.
7. **Predicción de Modos de Falla**: Dificultades con propiedades emergentes de modelos a gran escala.
8. **Riesgo de Privacidad**: Potenciado por capacidades de agregación de datos.
9. **Mantenimiento Frecuente**: Requerido por desviaciones en datos, modelos o conceptos.
10. **Opacidad y Reproducibilidad**: Aumento de preocupaciones.
11. **Estándares de Prueba Subdesarrollados**: Para prácticas basadas en IA.
12. **Costos Computacionales y Ambientales**: Impacto del desarrollo de sistemas de IA.
13. **Efectos Secundarios Inesperados**: Dificultades en la predicción más allá de medidas estadísticas.

Gestión de Riesgos de IA e Interacción Humano-IA

AIRMPC™ Versión 032024

Gestión de Riesgos de IA e Interacción Humano-IA

Aspecto	Descripción
Roles y responsabilidades humanas	Se necesita definir y diferenciar claramente los roles y responsabilidades humanos en la toma de decisiones y supervisión de sistemas de IA. Las configuraciones humano-IA pueden variar desde totalmente autónomas hasta completamente manuales, con sistemas que pueden tomar decisiones de forma autónoma, delegar la toma de decisiones a un experto humano o ser utilizados como una opinión adicional por un tomador de decisiones humano.
Sesgos cognitivos sistémicos y humanos	Las decisiones que se toman en el diseño, desarrollo, despliegue, evaluación y uso de sistemas de IA reflejan sesgos cognitivos sistémicos y humanos. Estos sesgos pueden introducirse en cualquier etapa del ciclo de vida de la IA a través de suposiciones, expectativas y decisiones humanas, y pueden ser exacerbados por la opacidad de los sistemas de IA y la falta de transparencia.

Gestión de Riesgos de IA e Interacción Humano-IA

Aspecto	Descripción
Variabilidad en la interacción humano-IA	Los resultados de la interacción humano-IA varían y, bajo ciertas condiciones, la parte de IA puede amplificar los sesgos humanos, llevando a decisiones más sesgadas que las de la IA o el humano por separado. Sin embargo, al tener en cuenta estas variaciones de manera juiciosa al organizar equipos humano-IA, se puede lograr complementariedad y mejorar el rendimiento general.
Presentación de la información del sistema de IA a humanos	Presentar la información de los sistemas de IA a los humanos es complejo. Los humanos perciben y derivan significado de la salida y explicaciones del sistema de IA de diferentes maneras, reflejando preferencias, rasgos y habilidades individuales diversas.

Certificación Artificial Intelligence Professional

Beneficios

- Ampliación de conocimientos en IA y Aprendizaje Automático, capacidad para aplicar técnicas de análisis de datos en la toma de decisiones, mejora en la competitividad y en la eficiencia operativa.

Habilidades Desarrolladas

- Fundamentos de IA y aprendizaje automático, métodos de aprendizaje supervisado y no supervisado, análisis de datos, programación en Python, comprensión de algoritmos y sus limitaciones, aplicación de técnicas de clustering y regresión.

https://certiprof.com/collections/new-technologies-certifications/products/artificial-intelligence-professional-certificate-caipc

Atributos clave del Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF)

AIRMPC™ Versión 032024

Atributos clave del Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF)

- **Comunicación Efectiva de Riesgos de IA**
  - Uso de lenguaje claro y comprensible.
  - Facilita la comunicación de riesgos de IA a través de diferentes niveles organizacionales y con el público.
- **Fomento de un Lenguaje Común**
  - Proporciona taxonomía, terminología, definiciones, métricas, y caracterizaciones para el riesgo de IA.
- **Usabilidad y Compatibilidad**
  - Fácilmente utilizable y coherente con otros aspectos de la gestión de riesgos.
  - Adaptable a estrategias y procesos más amplios de gestión de riesgos.
- **Aplicabilidad Universal**
  - Útil para una amplia gama de perspectivas, sectores, y dominios tecnológicos.
- **Enfoque en Resultados**
  - Ofrece un catálogo de resultados y enfoques sin prescribir requisitos únicos.

Certificación Artificial Intelligence Expert

Beneficios

- Mejora en la comprensión y aplicación de técnicas de Inteligencia Artificial y Aprendizaje Automático. Capacidad para utilizar análisis de datos en la toma de decisiones. Reducción de errores mediante la comprensión de los límites de los algoritmos. Preparación para roles avanzados en proyectos de IA. Incremento de la eficiencia operativa y competitividad en el mercado laboral.

Habilidades Desarrolladas

- Fundamentos de IA y aprendizaje automático. Métodos de aprendizaje supervisado y no supervisado. Análisis de datos para la toma de decisiones. Programación en Python y conocimientos matemáticos esenciales en IA. Métodos básicos de programación. Aplicación de técnicas de clustering y regresión.

https://certiprof.com/collections/new-technologies-certifications/products/artificial-intelligence-expert-certificate-caiec

Conoce nuestro
plan carrera en

New Technologies

¡Certifícate hoy!

certiprof®

¡Síguenos, ponte en contacto!

www.certiprof.com