Fundamentos algorítmicos y ética en IA: de la teoría a la práctica

Kit de herramientas para sesiones sincrónas

UC1 | Ética de la IA: un enfoque práctico

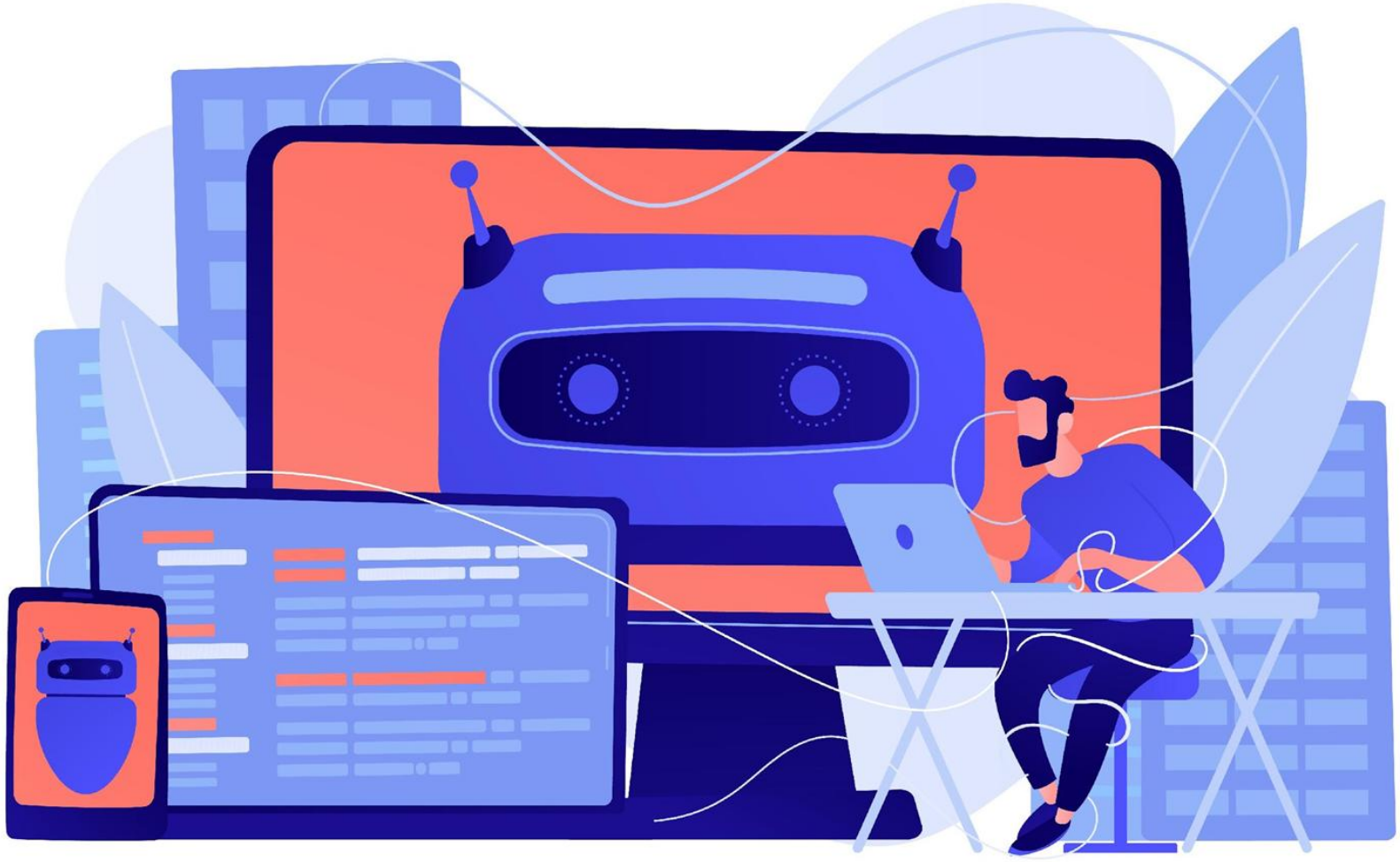
Soporte de diapositivas de PowerPoint

Número de proyecto| 2022-1-ES01-KA220-HED-000085257

ÍNDICE

- INTRODUCCIÓN - 3
- PRINCIPIOS ÉTICOS DE LA IA - 8
- MARCOS ÉTICOS, DIRECTRICES Y CONJUNTOS DE HERRAMIENTAS DE LA IA - 23
- LEY DE IA DE LA UE – 33
- CICLO DE VIDA DEL DESARROLLO DE IA: 40
- PARTICIPACIÓN DE LAS PARTES INTERESADAS EN EL DESARROLLO DE SOLUCIONES DE IA – 45
- PRINCIPIOS ÉTICOS Y PARTICIPACIÓN DE LAS PARTES INTERESADAS EN EL CICLO DE VIDA DEL DESARROLLO DE LA IA - 53

INTRODUCCIÓN



Fuente de la imagen| Freepik

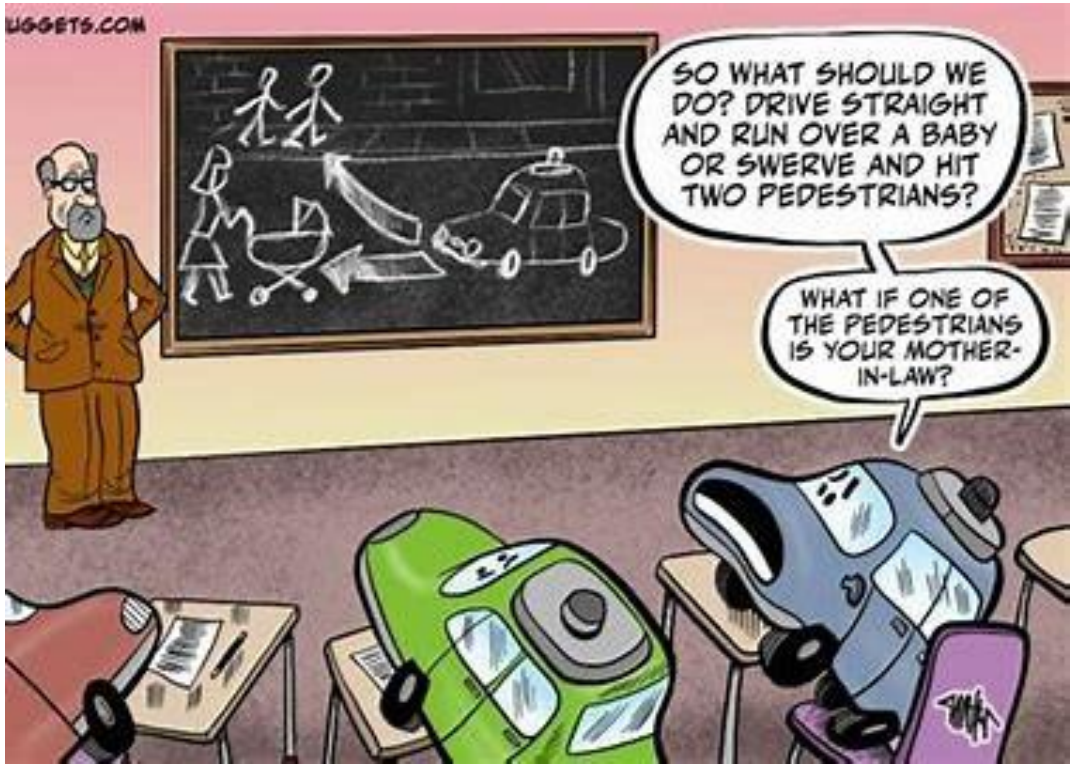
EN ESTA UNIDAD DE COMPETENCIA ENCONTRARÁS: LAS SIGUIENTES ASIGNATURAS:

- La ética de la IA y su importancia en el desarrollo de soluciones de IA
- Principios éticos de la IA
- Marcos éticos, directrices y conjuntos de herramientas
- Grupo de expertos de alto nivel; IA confiable – Marco
- Ley de IA de la UE
- Ciclo de vida del desarrollo de la IA
- Participación de las partes interesadas en el ciclo de vida del desarrollo de la IA
- Ciclo de vida del desarrollo de la IA con principios éticos y partes interesadas relacionados

AL FINAL DE LA UNIDAD DE COMPETENCIA, DEBERÍA SER CAPAZ DE:

- Conocer los principios éticos en el desarrollo y la implementación de la IA
- Reconocer diversos marcos y directrices éticas y su aplicación en escenarios del mundo real que involucran sistemas de IA
- Comprender la importancia de los principios éticos y la participación de las partes interesadas en el proceso de desarrollo ético de la IA

Preocupaciones éticas: ¿qué sucede si...



FUENTE DE LA IMAGEN| Dibujos animados: Enseñando ética a la IA - KDnuggets

- ¿La IA tendría el poder de decidir si obtienes un préstamo?
- ¿AI juzgaría quién será condenado a prisión?
- ¿Un coche autónomo tendría que decidir si salvar la vida del peatón o la de los que están dentro del coche?
- ¿Qué pasaría si un sistema de contratación basado en IA decidiera si contratar a una candidata femenina o masculina como capitán de vuelo?

La inteligencia artificial está revolucionando la vida

FUENTE DE LA IMAGEN | OCDE; IA: Riesgo y Oportunidad

Vídeo de: https://youtu.be/-CXkHs3cxa4

- La IA impacta en todas las facetas de la vida, influyendo en el trabajo, el ocio y brindando soluciones a problemas globales como el cambio climático y el acceso a la atención médica, al tiempo que plantea desafíos significativos para los gobiernos y las personas.
- La integración de la IA en las economías y las sociedades plantea interrogantes sobre los marcos políticos e institucionales adecuados necesarios para guiar su desarrollo y aplicación en beneficio de la sociedad.
- Las amplias consideraciones éticas han llevado a la creación y publicación de numerosos enfoques para garantizar la aplicación ética de la IA, que se tratan en este curso.

PRINCIPIOS ÉTICOS DE LA IA

FUENTE DE LA IMAGEN| Freepik

Principios éticos de la IA

¿Cuáles son los principios éticos de la IA?

- **Los principios éticos se definen como directrices generales que abordan las preocupaciones éticas de los sistemas de IA.** Estos principios son a menudo abstractos y no proporcionan instrucciones específicas sobre cómo aplicarlos en un contexto particular.
- Los principios éticos describen lo que se espera en términos de lo correcto y lo incorrecto y otras normas éticas. Los principios éticos de la IA se refieren a las restricciones normativas **sobre lo que se debe y no se debe hacer** en el uso de algoritmos en la sociedad.

(Prem 2023), (Zhou et al., 2020)

Principios éticos de la IA

- Se han publicado multitud de recomendaciones para guiar el uso ético de la inteligencia artificial.
- Las contribuciones provienen de varios sectores, por ejemplo:

  - Recomendación de la OCDE y esbozo de los principios éticos de la IA (publicado en mayo de 2019)
  - Principios de IA de Beijing para la investigación y el desarrollo (publicado en mayo de 2019)
  - Industria (por ejemplo, Google, IBM, Microsoft, Intel)
  - Gobierno (por ejemplo, Declaración de Montreal, Grupo de Expertos de la Comisión Europea)
  - Academias (por ejemplo, Future of Life Institute, IEEE, AI4People)

(Prem 2023), (Morley et al 2020)

Principios éticos de la IA

Ejemplos de principios éticos

El gráfico muestra cómo los diferentes marcos combinan principios éticos y cómo se superponen.
Las flechas horizontales vinculan los temas entre sí.

AI4People (publicado en noviembre de 2018) (Floridi et al. 2018)	Cinco principios clave para cualquier marco ético de la IA (L, Floridi y Clement-Jones, 2019)	Directrices éticas para una IA fiable (publicadas en abril de 2019) (Comisión Europea, 2019)	Recomendación del Consejo de Inteligencia Artificial (publicada en mayo de 2019) (OCDE 2019b)
Beneficencia	La IA debe ser beneficiosa para la humanidad	Respeto a la autonomía humana	Crecimiento inclusivo, desarrollo sostenible y bienestar
No maleficencia	La IA no debe infringir la privacidad ni socavar la seguridad	Prevención de daños	Robustez, seguridad y protección
Autonomía	La IA debe proteger y mejorar nuestra autonomía y capacidad para tomar decisiones y elegir entre alternativas		Valores centrados en el ser humano
Justicia	La IA debe promover la prosperidad	Equidad	Equidad
Explicabilidad	Los sistemas de IA deben ser comprensibles y explicables	Explicability	Transparencia y explicabilidad Responsabilidad

Modified from Morley et al 2020

Principios éticos de la IA

Principios compartidos en documentos éticos de IA

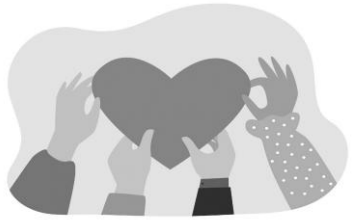
Fuente de la imagen| Image developed by VAMK

Los principios éticos de la IA muestran una superposición significativa, lo que sugiere una convergencia hacia un conjunto unificado de directrices éticas.

- Beneficioso y respetuoso con las personas y el medio ambiente (beneficencia).
- Robusto y seguro (no maleficencia).
- Respetuoso de los valores humanos (autonomía).
- Justo (justicia).
- Explicable, responsable y comprensible (explicabilidad).

Morley et al (2020), Jobin et al (2019)

Principios éticos de la IA



FUENTE DE LA IMAGEN| Freepik

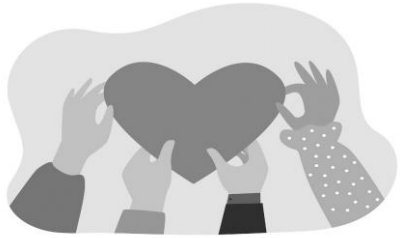
Beneficencia

La beneficencia en la IA abarca el fomento del bienestar de la humanidad

El principio de beneficencia hace hincapié en la creación de una IA que beneficie a las personas y al planeta, destacando la importancia de promover el bienestar y la sostenibilidad ambiental.

(Floridi et al 2018)

Principios éticos de la IA



FUENTE DE LA IMAGEN| Freepik

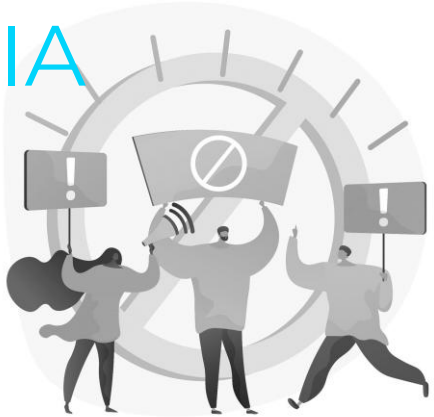
Ejemplo de principio de beneficencia

Recomendaciones sanitarias

- Considere un sistema de atención médica impulsado por IA que recomienda planes de tratamiento basados en los datos de los pacientes. Si la IA prioriza las opciones de tratamiento que son más rentables para el proveedor de atención médica sobre las que son más beneficiosas para la salud del paciente, no mantendrá la beneficencia. En este escenario, las recomendaciones de la IA priorizan las ganancias financieras en lugar de maximizar el bienestar del paciente.
- Watson for Oncology de IBM se enfrentó a desafíos éticos, ya que recomendaba tratamientos que no siempre se alineaban con las pautas médicas o los intereses de los pacientes, lo que generó preocupaciones sobre la priorización de las ganancias sobre los resultados de los pacientes. Para más información sobre el caso, visite Este enlace.

(Floridi et al 2018)

Principios éticos de la IA



Fuente de la imagen| Freepik

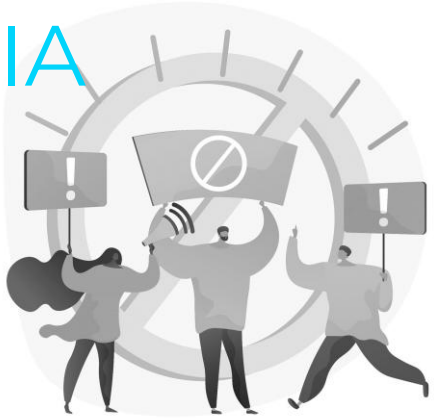
No maleficencia

La no maleficencia en la IA se centra en evitar el daño.

La no maleficencia se centra en prevenir el daño, ya sea por acciones humanas intencionales o comportamientos no intencionales de las máquinas, incluida la influencia no intencional en el comportamiento humano. Se trata de evitar resultados negativos, independientemente de la fuente.

(Floridi et al 2018)

Principios éticos de la IA



Fuente de la imagen| Freepik

Ejemplo de principio de no maleficencia

Vigilancia policial predictiva

- Imagínese un algoritmo de IA utilizado en la vigilancia policial predictiva que se dirige de manera desproporcionada a las comunidades minoritarias debido a los sesgos en los datos con los que se entrenó. Este sistema de IA podría perpetuar inadvertidamente el daño y la injusticia al someter a personas inocentes a una mayor vigilancia o escrutinio basado en suposiciones erróneas. A pesar de su intención de reducir los índices de criminalidad, las acciones de la IA pueden conducir al acoso injusto o a la persecución de ciertos grupos demográficos.

- El uso del software policial predictivo "Palantir" por parte del Departamento de Policía de Nueva Orleans se dirigió de manera desproporcionada a las comunidades minoritarias, lo que generó preocupaciones sobre la perpetuación de los sesgos sistémicos y la injusticia en las prácticas de aplicación de la ley. Para más información sobre el caso, visite Este enlace.

Bias in algorithms – Artificial intelligence and discrimination (europa.eu)

Principios éticos de la IA

FUENTE DE LA IMAGEN| Freepik

Autonomía

La autonomía en la IA implica un cuidadoso equilibrio entre la toma de decisiones humana y la de las máquinas, lo que hace hincapié en la necesidad de preservar la autodeterminación humana.

- La autonomía implica equilibrar el poder de decisión entre los humanos y la IA. Defiende el valor intrínseco de la elección humana para las decisiones importantes, abogando por un modelo de "decidir para delegar" en el que los humanos conservan la autoridad final para delegar y, si es necesario, anular las decisiones de la IA.

(Floridi et al 2018)

Principios éticos de la IA

Fuente de la imagen| Freepik

Ejemplo de principio de autonomía

Manipulación de las redes sociales

- Un ejemplo de cómo la IA socava la autonomía podría ser un **algoritmo de recomendación de redes sociales** que utiliza contenido personalizado para manipular el comportamiento y las opiniones de los usuarios sin su consentimiento explícito. Al alimentar continuamente a los usuarios con contenido que se alinea con sus creencias o sesgos existentes, la IA restringe la exposición de los usuarios a diversas perspectivas e influencias, limitando así su autonomía para tomar decisiones informadas basadas en una amplia gama de información.

- Ejemplo de la vida real: el algoritmo News Feedde Facebook, diseñado para personalizar el contenido en función de las interacciones de los usuarios, ha sido criticado por fomentar burbujas de filtro y cámaras de eco, amplificando potencialmente el contenido sensacionalista o polarizador e influyendo en el comportamiento de los usuarios sin consentimiento explícito. Para obtener más información sobre el tema, visite Este enlace.

Bias in algorithms – Artificial intelligence and discrimination (europa.eu) / Electronic Frontier Foundation (EFF), Pew Research Center, or Data & Society

Principios éticos de la IA

Fuente de la imagen| Freepik

Justicia

El principio de justicia en la IA se refiere a la equidad en la toma de decisiones y su distribución social, con el objetivo de abordar las disparidades y promover la equidad.

- El principio de justicia se refiere al uso de la IA para abordar injusticias históricas, garantizar una distribución equitativa de los beneficios de la IA y prevenir nuevos daños o interrupciones de la solidaridad social.

(Floridi et al 2018)

Principios éticos de la IA

Fuente de la imagen| Freepik

Ejemplo de justicia por principio

Algoritmos de sentencias penales

- Los algoritmos impulsados por IA utilizados en las sentencias penales pueden perpetuar involuntariamente los sesgos contra ciertos grupos demográficos. Si los datos históricos utilizados para entrenar estos algoritmos reflejan decisiones sesgadas tomadas por humanos, el sistema de IA puede recomendar sentencias más largas o castigos más severos para las personas de comunidades marginadas, exacerbando las desigualdades existentes en el sistema de justicia penal.
- Se descubrió que el sistema COMPAS, utilizado para la evaluación de riesgos en las sentencias penales, exhibía prejuicios raciales, etiquetando desproporcionadamente a los acusados negros como de mayor riesgo y contribuyendo a resultados injustos de las sentencias. Para más información sobre el caso, visite Este enlace.

Megan T. Stevenson and Christopher Slobogin, University of Pennsylvania Law Review

Principios éticos de la IA

Fuente de la imagen| Freepik

Explicabilidad

La explicabilidad en la IA significa asegurarnos de que entendemos cómo la IA toma sus decisiones y ser capaces de explicar esas decisiones con claridad. Garantiza que los sistemas de IA sean transparentes y responsables de sus acciones.

(Floridi et al 2018)

Principios éticos de la IA

Fuente de la imagen| Freepik

Ejemplo de explicabilidad de principio

Algoritmos de calificación crediticia

- Pensemos en un sistema de puntuación crediticia impulsado por IA que determine la solvencia de las personas, pero que carezca de transparencia en su proceso de toma de decisiones. Si a los solicitantes se les niegan préstamos o se les ofrecen términos desfavorables sin comprender los factores considerados por el algoritmo de IA, se socava su capacidad para impugnar o apelar las decisiones. Sin explicaciones claras de cómo la IA evalúa el riesgo crediticio, las personas pueden sentirse tratadas injustamente y perder la confianza en la integridad del sistema.

- La violación de datos de Equifax en 2017 expuso la información confidencial de millones de consumidores, lo que generó preocupaciones sobre la falta de transparencia y seguridad en los algoritmos de calificación crediticia, destacando problemas más amplios de la industria con respecto a la visibilidad limitada de los consumidores sobre los factores que afectan la solvencia. Para más información sobre el caso, visite Este enlace.

Keith Frankish & William M. Ramsey

MARCOS ÉTICOS, DIRECTRICES Y CONJUNTOS DE HERRAMIENTAS DE LA IA

Fuente de la imagen| Freepik

Marcos éticos, directrices y conjuntos de herramientas de la IA

Marcos éticos

- **Los marcos éticos van más allá de las declaraciones de principios de alto nivel. Proporcionan un enfoque estructurado para la toma de decisiones éticas.**
- Muchos marcos para la IA ética tienen como objetivo identificar posibles desafíos éticos y proponer algunos remedios para superar esos desafíos o mitigar los riesgos asociados.
- Un marco ético para la IA sirve como estructura fundamental para dar forma a leyes, reglas, normas técnicas y mejores prácticas en varios sectores y regiones.
- Los componentes típicos de los marcos éticos son::
    - Conceptos – Relevantes para el debate de los aspectos éticos
    - Principio – Principios éticos (por ejemplo, valores)
    - Preocupación: cómo se ven amenazados los principios por el uso y el desarrollo de sistemas de IA
    - Remedio: estrategias, reglas y pautas para abordar el problema.

(Floridi & Cowls, 2019), (Prem, 2023: 701) (Ayling & Chapman, 2022

Marcos éticos, directrices y conjuntos de herramientas de la IA

Herramientas y directrices éticas

- **Las directrices éticas ofrecen orientación práctica y reglas específicas para el comportamiento ético en diversos contextos.**
- Los conjuntos de herramientas y directrices éticas dan forma a la innovación basada en la IA y respaldan la aplicación práctica de los principios éticos de la IA.
- Los kits de herramientas y directrices explican cómo aplicar los principios éticos en el diseño, la implementación y el despliegue de la IA.
- Los marcos, las directrices y los conjuntos de herramientas se superponen con frecuencia en su uso.
- En las siguientes páginas, se presentan con más detalle dos ejemplos de frameworks; Marco de IA confiable del GEAN y Marco de clasificación de IA de la OCDE

(Zhou et al., 2020)

Marcos éticos, directrices y conjuntos de herramientas de la IA

Grupo de Expertos de Alto Nivel sobre Inteligencia Artificial; IA confiable

- A pesar del acuerdo de que la IA debe ser ética, existe un debate sobre qué constituye una "IA ética" y qué requisitos éticos y normas técnicas se necesitan para lograrla.
- La Comisión Europea presentó una estrategia para el desarrollo de la IA en Europa que hace hincapié en la IA ética, segura y avanzada.
- En abril de 2019, el Grupo de Alto Nivel de Inteligencia Artificial (GEAN) presentó las Directrices éticas para una IA fiable con el fin de promover la IA fiable.
- **Se reconoce que una IA fiable es el objetivo esencial para fomentar la confianza en el desarrollo y la aplicación de la IA, sobre la base de un marco sólido que garantice su fiabilidad**.
- Ver más en: https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai

(Leijnen et al., 2020), HLEG (2019). High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Brussels: European Commission

Marcos éticos, directrices y conjuntos de herramientas de la IA

IA fiable - Marco

Una IA fiable tiene tres componentes, que deben cumplirse a lo largo de todo el ciclo de vida del sistema:

- Legal: respetuoso de todas las leyes y regulaciones aplicables.
- Ético: respeto de los principios y valores éticos.
- Robusto, tanto desde el punto de vista técnico como teniendo en cuenta su entorno social

Ethics guidelines for trustworthy AI. (2019). Publications Office of the European Union.

Marcos éticos, directrices y conjuntos de herramientas de la IA

Grupo de expertos de alto nivel sobre inteligencia artificial: marco para una IA fiable

Fundamentos de una IA fiable

Cuatro principios éticos basados en los derechos fundamentales

Realización de una IA fiable

- Implemente **siete requisitos clave**

Evaluación de la IA fiable

- Para poner en práctica los requisitos clave, adapte la lista de evaluación a la aplicación de IA específica

HLEG (2019). High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Brussels: European Commission.

Marcos éticos, directrices y conjuntos de herramientas de la IA

Fundamentos de una IA fiable: principios éticos

Respeto a la autonomía humana

- La IA debe mejorar la autodeterminación, evitar la manipulación y apoyar las habilidade cognitivas, sociales y culturales humanas. Debe permitir la supervisión humana y la elección significativa, y apoyar a los seres humanos en el lugar de trabajo.

Prevención de daños

- La IA no debe causar daño, salvaguardar la dignidad humana, ser segura y robusta, y prestar especial atención a los grupos vulnerables y al medio ambiente. Debería preven efectos adversos debidos a desequilibrios de poder o de información.

Equidad

- La IA debe garantizar una distribución equitativa de los beneficios y los costos, evitar los sesgos y la discriminación, y promover la igualdad de oportunidades. Requiere un equilibrio en los medios y los fines y permite la impugnación y la reparación contra las decisiones de la IA.

Explicability

- Los procesos de IA deben ser transparentes, con una comunicación clara de las capacidades y los propósitos. Cuando no es posible una explicación completa, otras medidas como la trazabilidad y la auditabilidad deben garantizar el respeto de los derechos.

- El Grupo de Expertos de Alto Nivel sobre IA enumera cuatro principios éticos, arraigados en los derechos fundamentales, que deben respetarse para garantizar que los sistemas de IA se desarrollen, desplieguen y utilicen de manera fiable.
- Los principios éticos se especifican como imperativos éticos, de modo que los profesionales de la IA siempre deben esforzarse por cumplirlos.
- El Grupo de Expertos de Alto Nivel de AI también reconoce y aborda las posibles tensiones entre estos principios.

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Marcos éticos, directrices y conjuntos de herramientas de la IA

Realización de una IA fiable: requisitos clave

- Los principios éticos establecidos deben convertirse en **requisitos definitivos** para la realización de una IA confiable. Estos requisitos se aplican a todos los participantes en el ciclo de vida del sistema de IA:
- Desarrolladores: personas o equipos que realizan investigación, diseño o desarrollo de IA.
- Implementadores: organizaciones que implementan IA en sus operaciones y ofrecen productos o servicios impulsados por IA.
- Usuarios finales: personas que interactúan con los sistemas de IA, ya sea directa o indirectamente.
- Sociedad en general: aquellos que se ven afectados por la IA, tanto directa como indirectamente.

Agencia humana y supervisión

- Derechos fundamentales, agencia humana y supervisión humana

Robustez técnica y seguridad

- Resistencia a los ataques y a la seguridad, plan de reserva y seguridad general, precisión, fiabilidad y reproducibilidad

Privacidad y gobernanza de datos

- Respeto de la privacidad, la calidad y la integridad de los datos, y el acceso a los datos

Transparencia

- Trazabilidad, explicabilidad y comunicación

Diversidad, no discriminación y equidad

- la prevención de sesgos injustos, la accesibilidad y el diseño universal, y la participación de las partes interesadas

Bienestar social y medioambiental

- sostenibilidad y respeto al medio ambiente, impacto social, sociedad y democracia

Responsabilidad

- auditabilidad, minimización y notificación de impactos negativos, compensaciones y compensaciones.

HLEG (2019). High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. Brussels: European Commission

Marcos éticos, directrices y conjuntos de herramientas de la IA

Evaluación de la IA fiable: lista de evaluación de ALTAI

- Para ayudar a desarrollar una IA fiable, se ha desarrollado una herramienta que transforma estos principios de la IA en medidas prácticas.
- Conocida como Assessment List for Trustworthy AI (ALTAI), esta herramienta proporciona una lista de verificación completa y dinámica que sirve como hoja de ruta para que los desarrolladores e implementadores integren estos principios en sus aplicaciones de IA.
- ALTAI facilita este proceso al ofrecer un conjunto de pasos concretos para la autoevaluación, asegurando así que las tecnologías de IA se desarrollen de una manera que maximice los beneficios para el usuario y minimice la exposición a riesgos innecesarios.
- Acceso a herramienta ALTAI

Fuente de la imagen| ALTAI

(High-Level Expert Group on AI (AI HLEG), 2020)

Marcos éticos, directrices y conjuntos de herramientas de la IA

Marco de clasificación de la IA de la OCDE

- El marco de clasificación de IA de la OCDE, creado por la OCDE. La Red de Expertos en IA está diseñada para ayudar a diversas partes interesadas a evaluar el potencial y los riesgos de diversos sistemas de IA.
- Esta herramienta facilita el desarrollo de estrategias de IA y tiene como objetivo mantener la coherencia de las políticas a nivel internacional.
- El Marco se basa en los Principios de IA de la OCDE, respaldando valores como la equidad, la transparencia, la seguridad y la rendición de cuentas, así como abogando por el desarrollo de la capacidad humana y la colaboración internacional.

Fuente de la imagen| OECD Framework

https://www.oecd.org/digital/artificial-intelligence/

Ley de Inteligencia Artificial de la UE

Fuente de la imagen| Freepik

Ley de Inteligencia Artificial de la UE

- The AI Act is a draft EU law on AI Artificial Intelligence Act | AI Act - the first of its kind in the world. It is forecasted that the Act will be enforced in 2026.
- It applies to the development, deployment, and use of AI in the EU or when it will affect people in the EU.
- The European Commission released the proposed Regulation on Artificial Intelligence (AI Act) on 21 April 2021. The Act is the first-ever legal framework that sets out harmonised rules for the development, placing on the market, and use of artificial intelligence in the European Union.

Fuente de la imagen| The AI Act Explorer | EU Artificial Intelligence Act

https://artificialintelligenceact.com
https://www.euaiact.com

Ley de Inteligencia Artificial de la UE

Marco basado en el riesgo

Los principios éticos establecidos deben convertirse en requisitos definitivos para la realización de una IA confiable. Estos requisitos se aplican a todos los participantes en el ciclo de vida del sistema de IA:

- **Desarrolladores: personas o equipos que realizan investigación, diseño o desarrollo de IA.**
- **Implementadores: organizaciones que implementan IA en sus operaciones y ofrecen productos o servicios impulsados por IA.**
- **Usuarios finales: personas que interactúan con los sistemas de IA, ya sea directa o indirectamente.**
- **Sociedad en general: aquellos que se ven afectados por la IA, tanto directa como indirectamente.**

Ley de Inteligencia Artificial de la UE 'Pyramid of Risks' (Fuente: European Parliament)

Ley de Inteligencia Artificial de la UE

Marco basado en riesgos | Riesgo mínimo o nulo

Aplicaciones de IA que se considere que presentan un riesgo mínimo o nulo para los derechos o la seguridad de los ciudadanos. La gran mayoría de los sistemas de IA entran en la categoría de riesgo mínimo.

Ejemplos de aplicaciones de IA de riesgo mínimo:

- Sistemas de recomendación habilitados para IA
- Filtros de spam

De acuerdo con la normativa actual, estos sistemas no están sujetos a ninguna nueva obligación específica, sino que deben adherirse a leyes preexistentes.

Ley de Inteligencia Artificial de la UE

Marco basado en riesgos | Riesgo limitado

Los sistemas de IA de riesgo limitado requieren transparencia.

En el caso de estos sistemas de IA, los requisitos se refieren a la transparencia: los usuarios deben ser informados de que están interactuando con la IA y disponer de la información necesaria para decidir sobre su uso continuado.

Ejemplos de aplicaciones de IA de riesgo limitado:

- IA que genera o manipula contenido de imagen o audio.
- IA que crea o altera contenido de video, como deepfakes.

Ley de Inteligencia Artificial de la UE

Marco basado en riesgos | Alto riesgo

Los sistemas de IA de alto riesgo pueden tener un impacto significativo en las oportunidades de vida de un usuario.

Estos sistemas deben cumplirse antes de su implantación en el mercado de la UE, incluidas las obligaciones de gestión de riesgos y gobernanza de datos.

Algunos ejemplos de IA de alto riesgo son los siguientes::

- Gestión y operación de infraestructuras críticas
- Educación y formación profesional
- Empleo, gestión de trabajadores y acceso al trabajo por cuenta propia
- Acceso y disfrute de los servicios privados esenciales y de los servicios y prestaciones públicos
- Aplicación de la ley
- Gestión de la migración, el asilo y el control de fronteras
- Asistencia en la interpretación jurídica y aplicación de la ley.

Ley de Inteligencia Artificial de la UE

Marco basado en riesgos | Riesgo inaceptable

Se prohíben los sistemas de IA **con riesgo inaceptable** que se considere que suponen una clara amenaza para los derechos fundamentales de las personas. Los sistemas de riesgo inaceptable son aquellos que son explotadores, manipuladores o utilizan técnicas subliminales.

La venta de estos sistemas de IA en el mercado de la UE.

Algunos ejemplos de usos prohibidos de la IA son los siguientes::

- Manipulación cognitivo-conductual de personas o grupos vulnerables específicos: por ejemplo, juguetes activados por voz que fomentan comportamientos peligrosos en los niños
- Puntuación social: clasificar a las personas en función del comportamiento, el estatus socioeconómico o las características personales
- Identificación biométrica y categorización de personas
- Sistemas de identificación biométrica en tiempo real y remotos, como el reconocimiento facial

CICLO DE VIDA DEL DESARROLLO DE LA IA

Fuente de la imagen| Freepik

Ciclo de vida de desarrollo de la IA

Principios éticos, marcos y directrices en acciones concretas

La importancia de los principios, marcos y directrices en el desarrollo de la IA es bien reconocida, pero su aplicación práctica durante la creación de soluciones de IA presenta varios **obstáculos:**

- Los principios de la IA tienden a ser abstractos, diversos e intrincados.
- Existe una complejidad en la conversión del comportamiento humano matizado en herramientas sencillas y universales.
- Los principios a menudo están sujetos a una interpretación personal, influenciada por los antecedentes individuales.
- Hay una ausencia de estandarización universal en las regulaciones, marcos y directrices (incluso si se han hecho algunas generalizaciones como se ha presentado anteriormente)
- Si bien los marcos son útiles para la identificación de problemas, es posible que no proporcionen una guía práctica sobre los pasos específicos que se deben tomar

Morley (2019) / Prem (2023)

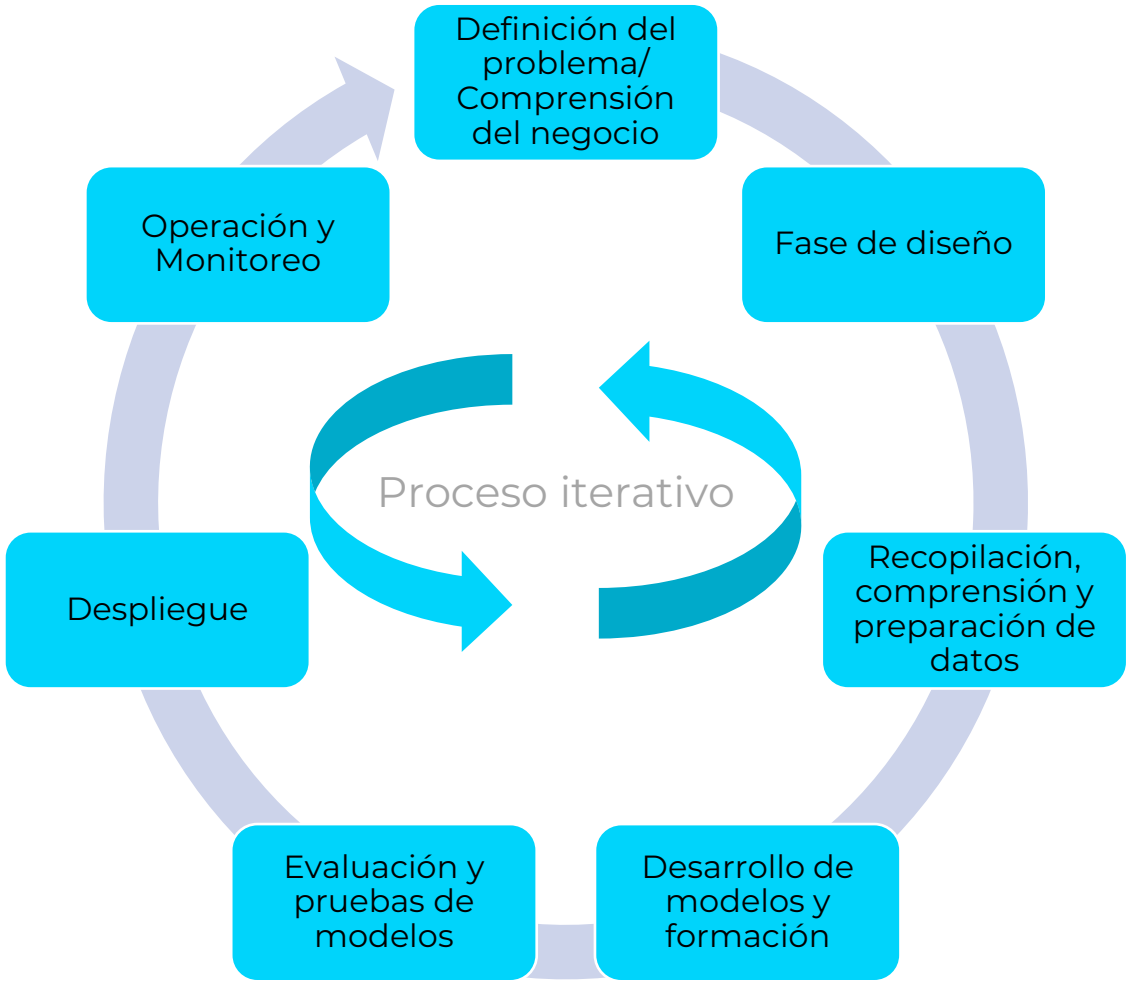
Ciclo de vida de desarrollo de la IA

Principios éticos, marcos y directrices en acciones concretas: Ciclo de vida de desarrollo de la IA

- Disponible Las herramientas diseñadas para abordar los desafíos éticos en las soluciones de IA son accesibles, pero a menudo cubren solo un número selecto de principios o etapas de desarrollo de soluciones de IA
- Para transformar los principios éticos en acciones concretas, se recomienda ver el proceso de desarrollo de la IA como una secuencia de pasos que comienzan con un caso de negocio y terminan con el modo operativo de la solución. (Ciclo de vida de desarrollo de la IA)
- Ciclo de vida de desarrollo de la IA, adopta una visión holística de las consideraciones éticas a lo largo de todas las etapas del desarrollo

Morley (2019) / Prem (2023)

Ciclo de vida de desarrollo de la IA

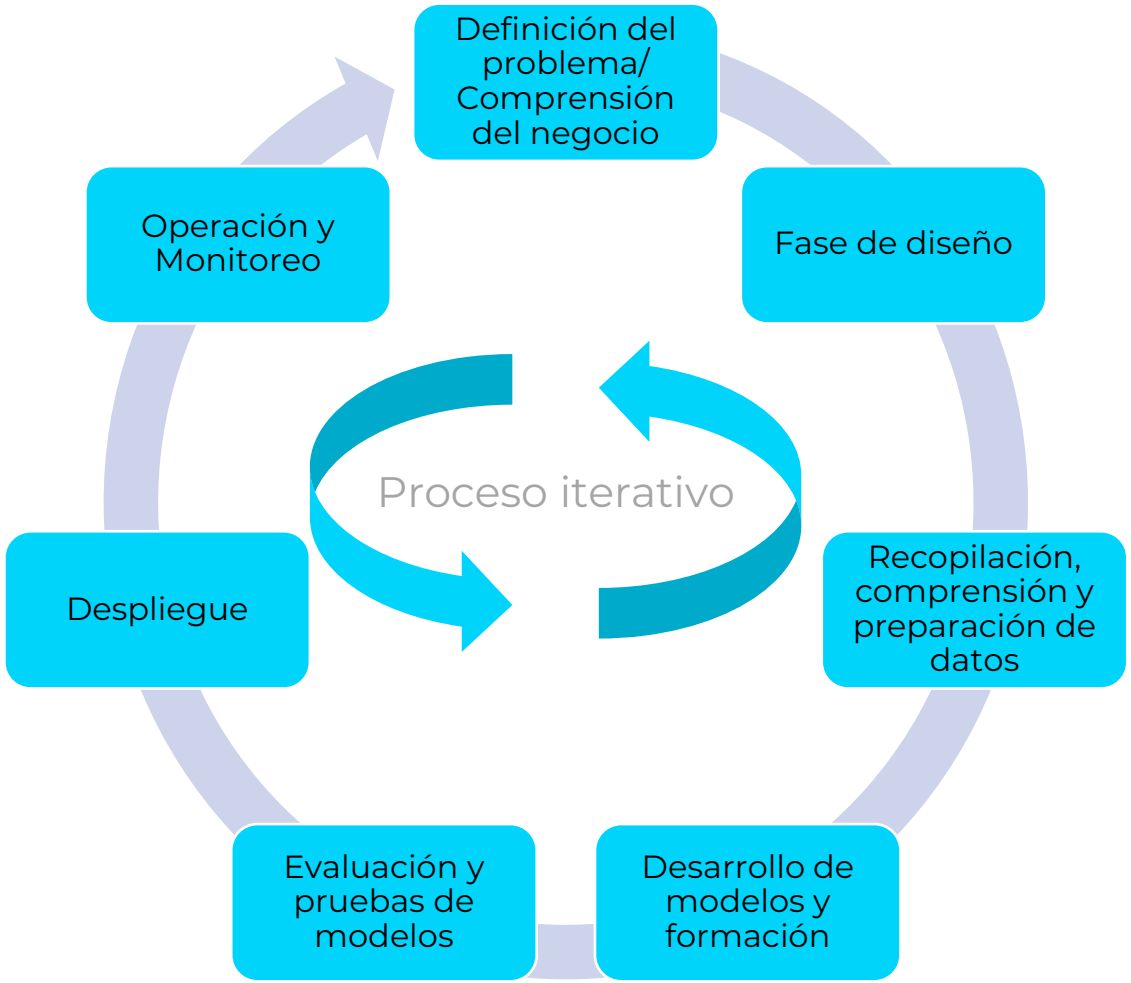


- Como cualquier proyecto, se recomienda seguir los pasos de diseño para transformar las ideas innovadoras en soluciones prácticas y bien definidas
- Es importante entender que cada paso juega un papel vital en el desarrollo de la solución, formando un Ciclo de vida de desarrollo de la IA
- El proceso es ITERATIVO, lo que significa que podría ser necesario volver a la etapa anterior en caso de que sea necesario y modificar la solución

Las etapas de diseño de desarrollo de IA que se presentan aquí son una combinación de los diferentes modelos descritos por los artículos de Alianza de Procesos de Ciencia de Datos (2023), Morley et.al' (2019), Prehm (2022) y Rochel et.al'a (2020).

Alianza de Procesos de Ciencia de Datos/Morley (2019) / Prem (2023) / Rochel & Evequoz (2020)

Ciclo de vida de desarrollo de la IA



- Como cualquier proyecto, se recomienda seguir los pasos de diseño para transformar las ideas innovadoras en soluciones prácticas y bien definidas
- Es importante entender que cada paso juega un papel vital en el desarrollo de la solución, formando un Ciclo de vida de desarrollo de la IA
- El proceso es ITERATIVO, lo que significa que podría ser necesario volver a la etapa anterior en caso de que sea necesario y modificar la solución

Las etapas de diseño del desarrollo de IA que se presentan aquí son una combinación de los diferentes modelos descritos por los artículos de Data Science Process Alliance (2023), Morley et.al' (2019), Prehm (2022) y Rochel et.al'a (2020).

Alianza de Procesos de Ciencia de Datos/Morley (2019) / Prem (2023) / Rochel & Evequoz (2020)

Ciclo de vida de desarrollo de la IA

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO

La definición del problema es crucial para comprender las necesidades del cliente y el papel de la solución de IA.

FASE DE DISEÑO

La fase de diseño convierte los requisitos de negocio en requisitos de diseño para los ingenieros.

RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN

La recopilación, la comprensión y la preparación de datos son la base de la solución

DESARROLLO Y FORMACIÓN DE MODELOS

El desarrollo y la evaluación de modelos tienen como objetivo convertir los datos en una solución factible que aborde el problema inicial

EVALUACIÓN Y PRUEBAS DE MODELOS

El modelo debe probarse exhaustivamente y la funcionalidad se refleja en las necesidades del cliente

DESPLIEGUE

La implementación es llevar la solución a sus usuarios

OPERACIÓN Y MONITOREO

La solución debe ser monitoreada durante el tiempo y mantenida de acuerdo con las necesidades cambiantes

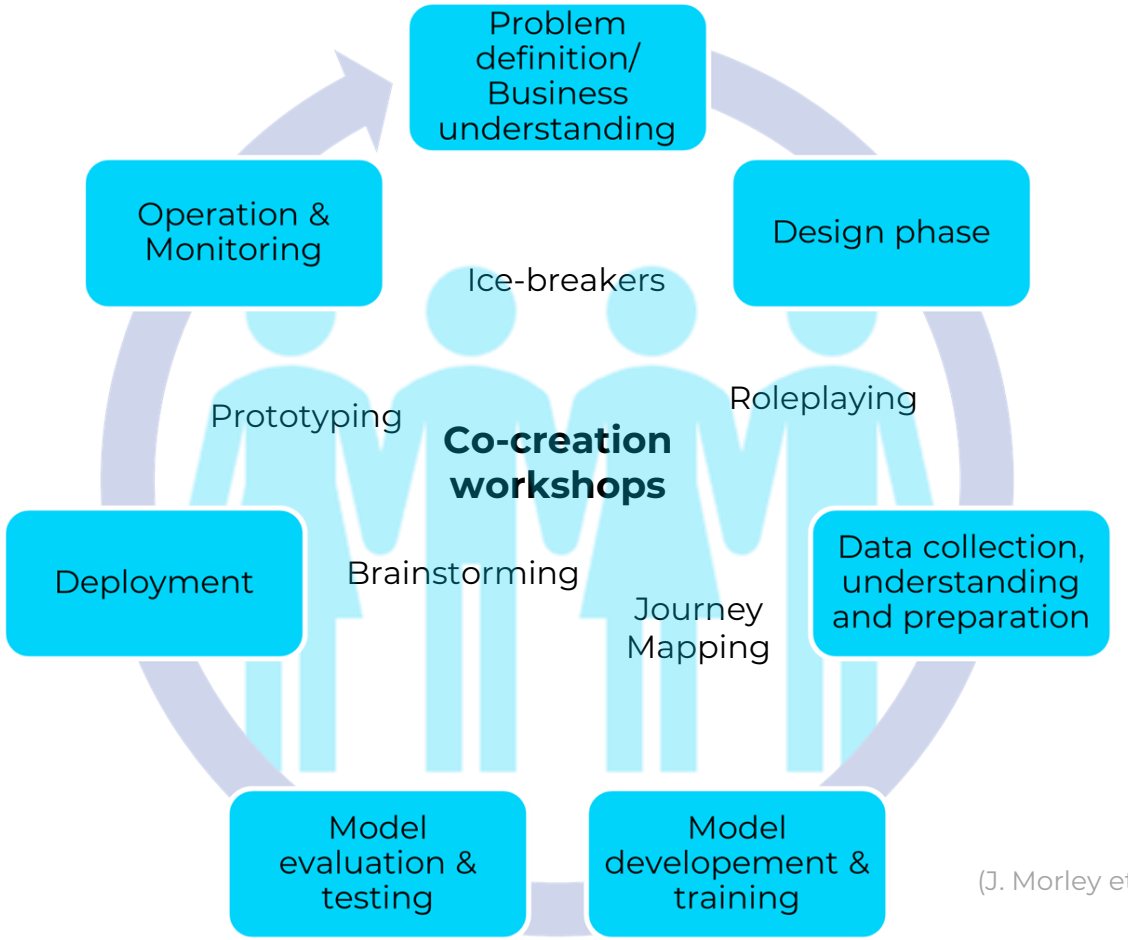
Alianza de Procesos de Ciencia de Datos/Morley (2019) / Prem (2023) / Rochel & Evequoz (2020)

PARTICIPACIÓN DE LAS PARTES INTERESADAS EN EL DESARROLLO DE SOLUCIONES DE IA

Fuente de la imagen| Freepik

Participación de las partes interesadas en el desarrollo de soluciones de IA

Why is it important to engage various stakeholders in AI solution development?



- Para garantizar que los impactos en el desarrollo de la IA se consideren desde diferentes puntos de vista y para desarrollar sistemas fiables, es esencial involucrar y consultar a múltiples partes interesadas en el proceso de desarrollo, incluidas aquellas que se verán afectadas por la solución de IA.
- La co-creación es una práctica utilizada para colaborar con las partes interesadas, incluidos los usuarios finales o clientes, durante un proceso de diseño para compartir ideas entre las partes interesadas con diferentes roles y experiencias, y permite comprender mejor las necesidades de los usuarios finales y, por lo tanto, diseñar soluciones que tengan en cuenta a los diferentes usuarios.

(J. Morley et al.,2019) / (UNESCO, 2023) / (Interaction Design Foundation, Co-creation) / (Anika Sanin, 2020)

Participación de las partes interesadas en el desarrollo de soluciones de IA

Partes interesadas en el desarrollo

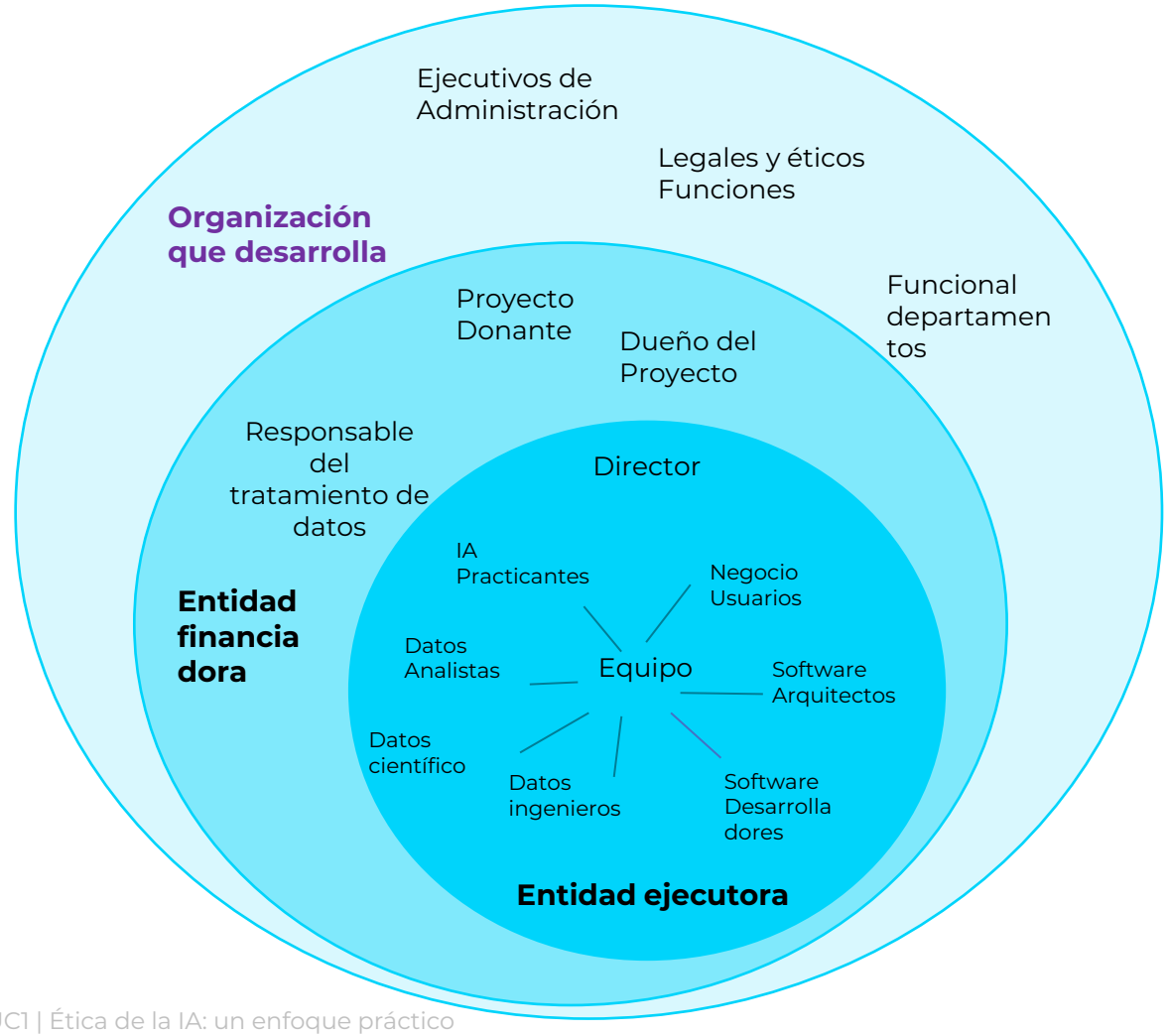
Organización en desarrollo

- Patrocina el proyecto de IA y determina el alcance del proyecto.
- Los roles del proyecto son la gestión ejecutiva, los departamentos funcionales y las funciones legales y éticas:
  - La dirección ejecutiva es responsable de las decisiones estratégicas, financieras y de desarrollo de la empresa.
  - El departamento funcional es una unidad organizativa especializada dentro de una empresa que se centra en un conjunto específico de tareas o actividades, a menudo agrupadas en función de funciones comunes como finanzas, marketing o recursos humanos.
  - Las funciones legales y éticas dentro de una organización implican varias funciones y responsabilidades para garantizar el cumplimiento de las leyes, regulaciones y estándares éticos.

(C.J. Miller, 2022)

Participación de las partes interesadas en el desarrollo de soluciones de IA

Development stakeholders



Entidad financiadora

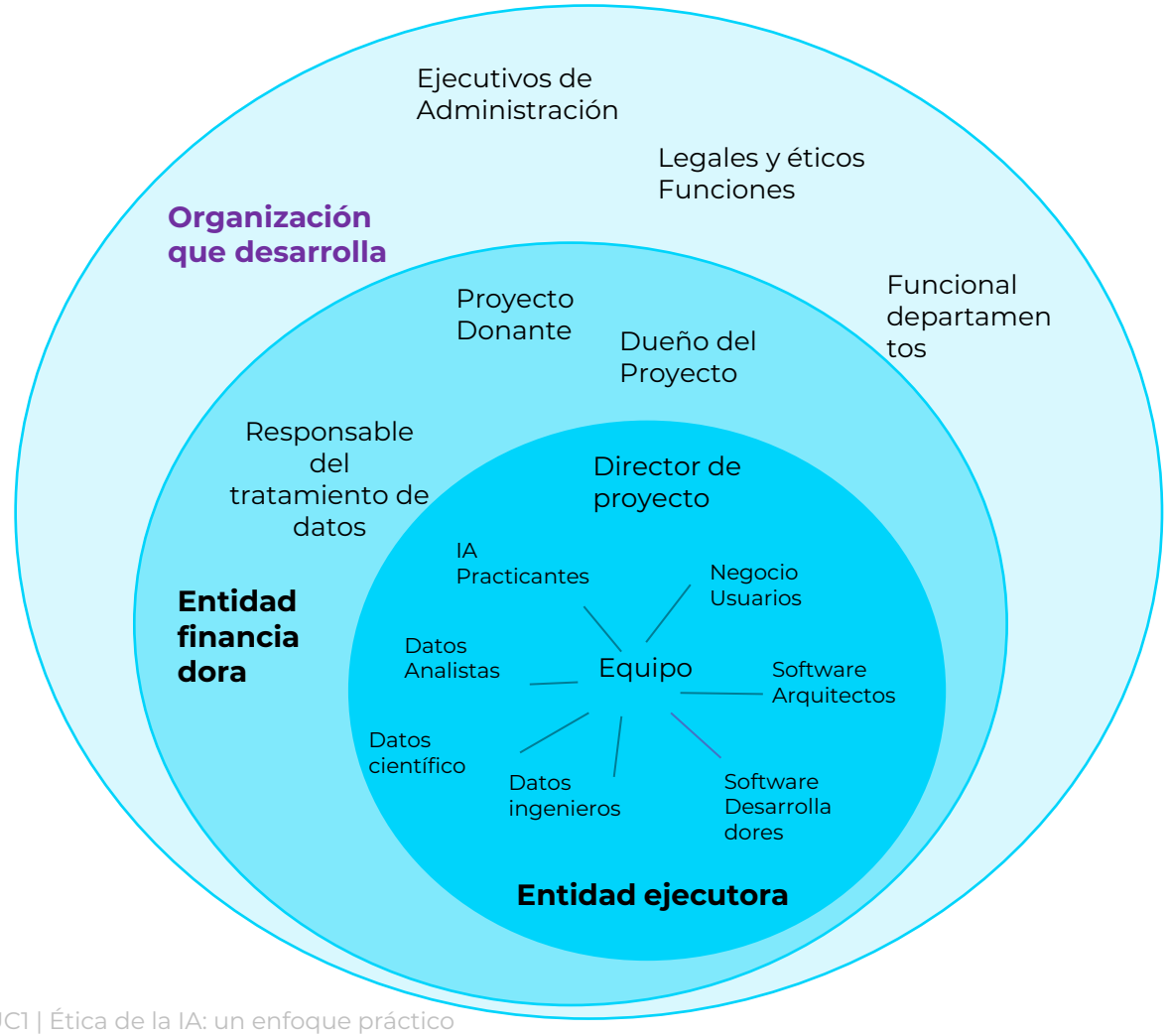
Los roles del proyecto son financiador del proyecto, propietario del proyecto y controlador de datos**.**

- **El financiador del proyecto financia el proyecto y asigna los recursos para el proyecto.**
- **El propietario del proyecto ofrece dirección estratégica.**
- **El responsable del tratamiento de datos es responsable de la gestión de los datos y de su uso en el desarrollo del sistema.**

(G.J. Miller, 2022)

Participación de las partes interesadas en el desarrollo de soluciones de IA

Development stakeholders



Entidad ejecutora

Los roles del proyecto son gerentes de proyecto y equipo de proyecto.

- **El director del proyecto** es responsable de los resultados del proyecto y se asegura de que se respeten las normas y expectativas éticas, de privacidad y seguridad, y de que se implique a las partes interesadas pertinentes.
- **El equipo del proyecto de IA** está formado por arquitectos de software, desarrolladores de software, ingenieros de datos, científicos de datos, analistas de datos, profesionales de la IA.

(G.J. Miller, 2022)

Participación de las partes interesadas en el desarrollo de soluciones de IA

Partes interesadas en el uso

Organización operativa

- Una organización operativa es una parte interesada del cliente.
- Al igual que en la organización de desarrollo, las funciones del proyecto de la organización operativa incluyen la gestión ejecutiva y las funciones legales y éticas.
- En el nivel de operaciones, los roles del proyecto son usuarios finales, mantenedores de modelos, controladores de datos y propietarios de plataformas.
    - Los **usuarios finales** son individuos o grupos que interactúan con el sistema de IA a través de contratos de trabajo o servicio con la organización operadora o como consumidores.
    - Los **mantenedores de modelos** son responsables de la gestión, las actualizaciones y la optimización continuas de los modelos de aprendizaje automático utilizados en el sistema de IA.
    - Los **responsables del tratamiento de datos** determinan los fines y los medios del tratamiento de datos personales dentro del sistema de IA.
    - Los **propietarios de las plataformas** son responsables de la infraestructura general y del despliegue del sistema de IA.

(G.J. Miller, 2022)

Participación de las partes interesadas en el desarrollo de soluciones de IA

Grupos de interés externos

Individuos y sociedad

- **Los sujetos individuales son personas que tienen acuerdos con la organización de desarrollo u operación, como contratos de servicio o de trabajo. Este grupo incluye a los sujetos de datos, los sujetos de decisiones y los trabajadores, y se consideran partes interesadas pasivas**.
- **La sociedad está formada por los individuos y el público. Podrían verse afectados negativamente por el desarrollo o el uso del sistema de IA**.
- Este grupo comprende comunidades locales, entornos y grupos vulnerables como personas con discapacidades, grupos minoritarios, menores y el público en general.

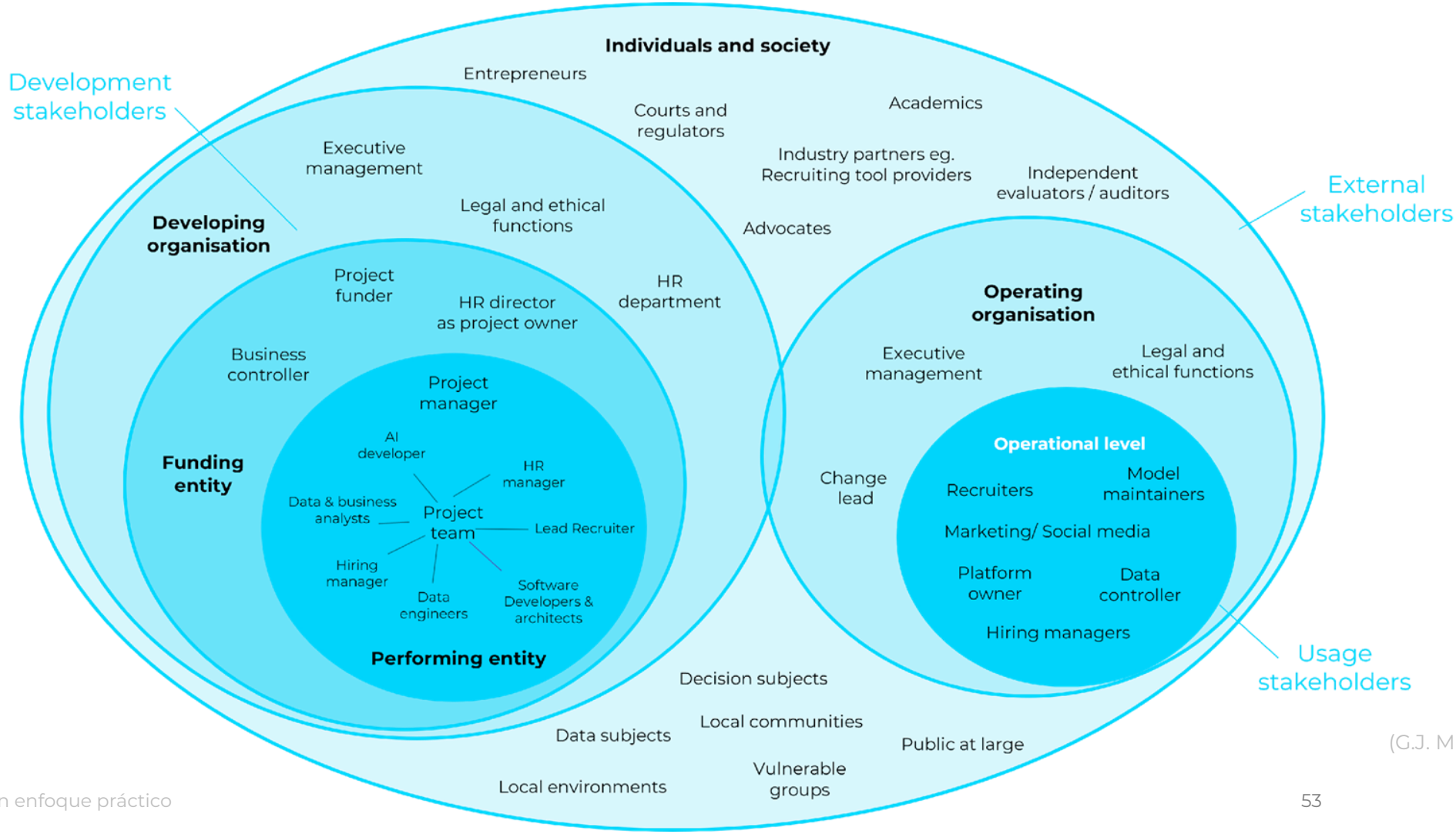
  Los ingenieros de IA forman parte de un debate social más amplio, en el que deben desempeñar un papel activo en la integración de otras perspectivas.

  - Los representantes son grupos y organizaciones que no se ven afectados por el sistema de IA, pero que podrían representar a otros en el proyecto.
  - Los representantes formales incluyen tribunales, leyes, reguladores y sindicatos.
  - Los emprendedores pueden participar en el desarrollo del sistema de IA.
  - Los evaluadores independientes son personas o grupos externos, por ejemplo, reporteros, auditores o revisores.

(C.J. Miller, 2021), (Rodrigues, 2020)

Participación de las partes interesadas en el desarrollo de soluciones de IA

Partes interesadas en el desarrollo de la IA



(G.J. Miller, 2022)

PRINCIPIOS ÉTICOS Y PARTICIPACIÓN DE LAS PARTES INTERESADAS EN EL

Ciclo de vida de desarrollo de la IA

Fuente de la imagen| Freepik

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

A lo largo del ciclo de vida de la IA, es importante mantener un enfoque en las consideraciones éticas y la participación de las partes interesadas.

En las siguientes páginas se analiza el ciclo de vida desde tres perspectivas:

1. **Descripción de la etapa**
2. **Principios éticos fundamentales**
3. **Grupos de interés implicados**

No todos los principios éticos tienen el mismo peso en todos los contextos o en todas las etapas del desarrollo y, por lo tanto, solo se presentan las principales consideraciones y puntos de vista a continuación.

Data Science Process Alliance /Morley (2019) / Prem (2023) / Rochel & Evequoz (2020)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Definición de problemas y comprensión del negocio

Primera etapa crucial para el éxito de un proyecto

- Comprender el problema a resolver o la oportunidad a investigar.
- Llevar a cabo investigaciones y entrevistas con las partes interesadas para formular un planteamiento claro del problema.
- Comprenda las necesidades de los clientes y cómo se traducen en una solución empresarial viable.
- Definir los objetivos y criterios de éxito de la solución.
- Realizar un estudio de viabilidad para determinar si la IA es la solución más adecuada.
- Evaluar la viabilidad técnica y ética del proyecto.

Data Science Process Alliance /Morley (2019) / DeSilva & Alahakoon (2022)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Definición de problemas y comprensión del negocio

Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

Sentar las bases: abordar cuestiones éticas básicas como los derechos humanos, el impacto ambiental y las preocupaciones sociales..

Beneficencia: garantizar que la solución de IA mejore el bienestar individual y el bien social. Áreas clave a tener en cuenta:

- **Participación de las partes interesadas**
  - Construir sistemas confiables que apoyen el florecimiento humano.
  - Involucrar a las personas impactadas en el proceso de desarrollo.
- **Protección de los derechos fundamentales**
  - Salvaguardar los derechos individuales, reforzando la integridad ética.
- **Sostenibilidad Ambiental**
  - Minimizar el impacto ambiental a través de prácticas sostenibles.
- **Justificación**
  - Definir claramente y vincular el propósito del sistema con beneficios sociales tangibles.

Morley (2019) / Prem (2023)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Definición de problemas y comprensión del negocio

Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

No maleficencia

- Asegurar que la identificación del problema y la solución propuesta no perjudiquen a las personas o a las comunidades.
- Considere y monitoree cuidadosamente el impacto del sistema en el bienestar físico y mental.
- Evaluar los posibles riesgos de seguridad y las estrategias de mitigación.

Justicia

- Evaluar el impacto del sistema en las instituciones, la democracia y la sociedad.
- Considere si ciertos grupos están favorecidos o en desventaja.
- Examinar si el proyecto podría infringir la autonomía individual.

Morley (2019)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Definición de problemas y comprensión del negocio

Participantes

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

- **Usuarios finales: usuarios principales de la solución de IA**
- **IA/científicos de datos: formular el problema y conceptualizar una solución**
- **Comunidades afectadas**
- **Expertos en la materia: conocimiento profundo de la industria en la que se aplica la IA**
- **Desarrolladores de IA: la participación temprana les permite comprender profundamente el problema**
- **Gestores de proyectos: supervisión de todo el progreso del proyecto**
- **Grupos de interés del negocio: garantizar la alineación con los objetivos y recursos del negocio**
- **Especialistas en ética y ciencias sociales: destacando los impactos potenciales**

(D. De Silva & D. Alahakoon, 2022)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Fase de diseño

Esta etapa transforma el trabajo preliminar establecido durante la etapa de definición del problema en planes detallados para construir e implementar el sistema de IA

- Especificar lo que el proyecto pretende alcanzar (objetivos y resultados)
- Evalúe el costo financiero frente a los beneficios anticipados.
- Esbozar las funcionalidades que debe poseer el sistema de IA.
- Decida el enfoque para el uso de datos, como el uso de modelos previamente entrenados.
- Establecer métricas para medir el éxito del proyecto.
- Integrar las directrices éticas que rigen el proceso de desarrollo.
- Desarrollar un cronograma y gestionar la participación de las partes interesadas.

Rochel (2020)/ Morley (2019) / & DeSilva&Alahakoon (2022)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Fase de diseño

Justicia | Evitar sesgos injustos
Requisitos de diseño

- Implementar mecanismos para identificar y mitigar sesgos en datos y algoritmos.
- Apuntar a resultados equitativos entre diferentes grupos.

- **Concienciación de los desarrolladores**
  Asegúrese de que los desarrolladores sean conscientes de los posibles sesgos, especialmente los relacionados, por ejemplo, con la raza, el color, la ascendencia, el género, la edad, el idioma, la religión, la opinión política, el origen nacional, étnico, social, la situación económica, la condición de nacimiento, la discapacidad
- **Comunicación y respuesta**
  Establecer protocolos claros para comunicarse, informar y responder a sesgos, impactos negativos o compensaciones.
- **Gestión de datos**
  Considere la posibilidad de utilizar sistemas de almacenamiento de datos para evitar sesgos de datos aislados.
- **Gestión de riesgos**
  Desarrollar medidas de "Qué sucede si" para manejar resultados inesperados.

Rochel (2020)/ Morley (2019) / & DeSilva&Alahakoon (2022)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Fase de diseño
Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

Explicabilidad / Transparencia
- Garantice una documentación clara y explicaciones fáciles de usar de las decisiones de la IA.
- Garantizar la transparencia de las fuentes de datos y los procesos algorítmicos.

Beneficencia
Involucrar a las partes interesadas en el proceso de desarrollo para que consideren diversos aspectos éticos..

Autonomía
Implementar mecanismos de supervisión humana para mantener los valores humanos en las soluciones de IA.

HLEG/ Unesco / Morley / Prem

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Design phase
Stakeholders

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

- **Arquitectos de software: sientan las bases técnicas de la solución de IA, teniendo en cuenta tanto las necesidades actuales como las futuras.**
- **Diseñadores de experiencia de usuario (UX): fundamentales para crear interfaces que mejoren la adopción y la satisfacción del usuario y para garantizar que los usuarios puedan interactuar eficazmente con el sistema de IA.**
- **Expertos en ética: abordan las preocupaciones éticas.**
- **Profesionales de la IA: colaboran con los diseñadores de UX para integrar los componentes de la IA sin problemas.**
- **Funciones legales y éticas: mitigar los riesgos legales abordando los requisitos de cumplimiento en las primeras etapas de la fase de diseño y ayudar a diseñar soluciones que respeten las regulaciones de privacidad y protección de datos.**
- **Personas y comunidades afectadas por el modelo de IA: enfoque de diseño participativo.**

(D. De Silva & D. Alahakoon, 2022) / (Peters, 2024) / (White, 2024)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Recopilación, comprensión y preparación de datos

Una vez establecido el objetivo y el plan para alcanzarlo, el siguiente paso consiste en trabajar con datos relevantes.

- Recopile todos los datos necesarios para el proyecto.
- Describa, explore, verifique, seleccione, limpie, construya, integre y formatee los datos en categorías y conceptos adecuados para el propósito previsto.
- Aborde y resuelva los problemas relacionados con los datos faltantes para garantizar la integridad.
- Reconozca y comprenda las suposiciones subyacentes a los datos existentes.
- Aclarar cómo se utilizarán los conjuntos de datos para la toma de decisiones.
- Mantener una alta calidad de los datos, especialmente en casos de información poco clara o faltante.

Esta es una etapa que requiere mucho tiempo y en la que se construyen los cimientos de la solución de trabajo. La fiabilidad y el rendimiento de la solución están directamente relacionados con la calidad de los datos subyacentes.

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Recopilación, comprensión y preparación de datos

Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OFUNCIONAMIENTO Y SEGUIMIENTO

No maleficencia1/2

- **La integridad y la calidad de los datos causan sesgos**
  - Aborda los posibles sesgos que favorecen a ciertos grupos demográficos. Garantizar la precisión al combinar conjuntos de datos (por ejemplo, no se deben mezclar personas con el mismo nombre)
  - Llevar a cabo una debida diligencia exhaustiva para evaluar y verificar la exactitud de los datos, asegurándose de que no perpetúen sesgos injustos.
  - Evite el uso de datos aislados basados en datos demográficos específicos, ya que pueden aumentar los sesgos.
  - Implemente un repositorio de datos unificado, como un almacén de datos, para minimizar el sesgo (nota: costoso).

¿Puede el sesgo ser un factor relevante en la solución de IA? Por ejemplo, el seguro de automóvil influenciado por los tipos de accidentes y los datos demográficos involucrados. Limitaciones legales: El género no puede determinar las tarifas de los seguros de automóviles, según el Tribunal de Justicia de la Unión Europea, 2012.

HLEG (2019) / De Silva & Alahakoon (2022) / Rochel et. Al (2023)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Recopilación, comprensión y preparación de datos

Principios

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OFUNCIONAMIENTO Y SEGUIMIENTO

No maleficencia 2/2

- **Privacidad y confidencialidad de los datos:**
  - Aborde los riesgos típicos en torno a la privacidad y confidencialidad de los datos durante la recopilación de datos.
  - Asegúrese de que los sistemas de IA mantengan la privacidad de los datos durante todo el ciclo de vida, cumpliendo con regulaciones como el RGPD.

- **Resistencia a los ataques y seguridad:**
  - Proteja los conjuntos de datos de los ataques y aborde los posibles riesgos de seguridad para mantener una seguridad de datos sólida.

HLEG (2019) / De Silva & Alahakoon (2022) / Rochel et. Al (2023)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Recopilación, comprensión y preparación de datos
Principios

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OFUNCIONAMIENTO Y SEGUIMIENTO

- **Los científicos de datos encuentran patrones y tendencias en los datos para convertirlos en conocimiento.**
- **Ingenieros de IA: responsables de evaluar y verificar los datos.**
- **Los analistas de datos contribuyen a comprender las características de los datos, lo que ayuda a informar las estrategias de preparación de datos.**
- **Ingenieros de datos: responsables de preparar los datos y asegurar su calidad.**
- **Oficial de protección de datos (DPO): responsabilidad de asegurarse de que su organización siga las reglas al manejar datos personales de empleados, clientes, proveedores u otras personas (denominadas sujetos de datos).**
- **Éticos: abordan las consideraciones éticas en el manejo y selección de datos.**

(Miller, 2022), (J. Rochel & F. Evéquoz, 2020), (European Data Protection Supervisor, 2024)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Desarrollo y formación de modelos

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

En esta fase, se desarrolla un modelo de IA para abordar el problema definido. Esto incluye varios pasos críticos:

- Seleccione y configure el modelo algorítmico y las herramientas, definiendo su uso por adelantado.
- Enumere y evalúe las implicaciones de las herramientas elegidas, especialmente teniendo en cuenta los posibles inconvenientes y compensaciones.
- Decida qué elementos del modelo deben enfatizarse y cuáles pueden convertirse en secundarios o incluso ocultos.
- Participe en múltiples rondas de refinamiento durante el proceso de entrenamiento, mejorando continuamente el modelo.
- La selección de herramientas a menudo implica equilibrar intereses contrapuestos. Por ejemplo, decidir si es preferible arriesgarse a tener correos electrónicos no deseados en la bandeja de entrada real o correos electrónicos válidos que terminan en la carpeta de correo no deseado.

Esta etapa es crucial, ya que influye directamente en la eficacia y eficiencia de la solución de IA.

Rochel & Ezequiz (2022), Data Science Process Alliance

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Desarrollo y formación de modelos
Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	**DESARROLLO Y FORMACIÓN DE MODELOS**	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

Explicabilidad

- Al centrarse en los aspectos técnicos, es crucial comprobar la explicabilidad y si existen sesgos dentro del sistema de IA, lo que ofrece una excelente oportunidad para la intervención.
- Mantenga la transparencia durante todo el proceso de desarrollo para garantizar que las partes interesadas puedan entenderlo. Esto implica una documentación clara de las fuentes de datos, las opciones de modelos y la justificación de estas decisiones.

Justicia

- Reconozca e informe de forma transparente cualquier compensación entre la eficiencia y otros requisitos.
- Establecer un proceso para reconocer y evaluar el impacto de estas compensaciones.
- Los ingenieros de IA deben justificar sus decisiones, teniendo en cuenta los posibles impactos negativos en las personas afectadas.

Morley (2019) Prem (2023)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Desarrollo de modelos y formación

Participantes

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

- **Los científicos de IA/ML transforman la definición del problema en un modelo de IA prototípico.**
- **Los ingenieros de IA tienen experiencia para seleccionar las herramientas algorítmicas más adecuadas para determinados fines.**
- **Los científicos de datos son fundamentales para el desarrollo de modelos de aprendizaje automático y su optimización.**
- **Los expertos en ética son esenciales para integrar los principios éticos en el proceso de desarrollo y evitar consecuencias no deseadas.**
- **Los usuarios empresariales se aseguran de que los modelos cumplan con los requisitos y expectativas empresariales**.

(D. De Silva & D. Alahakoon, 2022), (J. Rochel & F. Évequoz, 2020)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Evaluación y pruebas de modelos

Una vez que el modelo ha sido desarrollado y entrenado, su rendimiento debe ser probado y evaluado:

- Evalúe el modelo en función de objetivos y criterios de éxito predefinidos.
- Articular claramente los supuestos y estándares subyacentes en los que se basan los conjuntos de datos.
- Realice pruebas utilizando datos o pasos no vistos en el proceso para garantizar la solidez.
- Si el rendimiento no es satisfactorio, considere la posibilidad de modificar los parámetros del modelo, la arquitectura o incluso los conjuntos de datos que hay detrás del modelo de datos.
- Cualquier deficiencia debe discutirse abierta y honestamente con el líder del proyecto y otras partes interesadas, ya que podría afectar significativamente a las personas.
- Acordar si continuar con más iteraciones o pasar a la fase de implementación.

Esta etapa es crucial para garantizar la fiabilidad y eficacia del modelo antes de que entre en producción.

Rochel (2023)/ Data Science Process Alliance

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Model evaluation & testing

Ethical principles to be considered

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓNN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

Justicia / Beneficencia / No maleficencia

- **Auditorías éticas y métricas:**
  - Asegurar el cumplimiento de principios éticos como la justicia, la beneficencia y la no maleficencia a través de auditorías y métricas diseñadas para la auditabilidad..
- **Seguridad y resiliencia:**
  - Pruebe la resistencia a los ataques y establezca un plan alternativo para garantizar la seguridad del sistema.
- **Abordar los impactos negativos y los posibles riesgos:**
  - Implemente estrategias para minimizar y reportar problemas negativos.
  - Establecer un mecanismo de reparación para abordar los posibles daños si las cosas salen mal.
  - Formular factores de evaluación de riesgos centrados en la privacidad, las implicaciones sociales y los sesgos..
- **Explicabilidad:**
  - Explique claramente tanto los procesos técnicos como las decisiones humanas involucradas en el desarrollo del modelo..

Tenga en cuenta que los ingenieros bajo presión para finalizar las soluciones para la implementación pueden enfrentar un mayor riesgo de pasar por alto problemas críticos.

Rochel (2023) / Prem (2019) / De Silva & Alahakoon (2022)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Evaluación y pruebas de modelos
Participantes

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓNN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

- **Los ingenieros de IA interpretan los resultados después del modelado.**
- **Los equipos de aseguramiento de la calidad (QA) garantizan que los modelos cumplan con los estándares de calidad y estén libres de errores o consecuencias no deseadas.**
- **Los expertos en ética garantizan el uso ético de los modelos de IA.**
- **Los defensores/representantes de los usuarios se aseguran de que los modelos de IA se alineen con las expectativas y necesidades de los usuarios.**
- **Los equipos legales y de cumplimiento mitigan los riesgos legales al garantizar el cumplimiento de las regulaciones relevantes.**

(J. Rochel & F. Évequoz, 2020) / (Moore, 2023)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Despliegue

Una vez que la solución de IA se ha finalizado a satisfacción, se despliega en un entorno de producción para abordar problemas del mundo real. La complejidad de esta fase depende de la propia solución.

- Inicialmente, la solución suele presentarse a un grupo más pequeño de expertos y usuarios.
- Integre el modelo con los sistemas existentes, cree una aplicación o servicio que utilice el modelo o aproveche la información en un contexto fuera de línea, como un informe para la administración.
- Establezca un bucle de retroalimentación para supervisar el rendimiento y los resultados, con iteraciones posteriores si es necesario.
- Implemente estrategias de administración de riesgos para abordar posibles problemas durante la implementación.
- Plan de seguimiento y mantenimiento continuos en la fase de explotación.
- Revisar todo el proyecto y elaborar un informe final que documente el proceso y los resultados..

Este enfoque estructurado garantiza que la solución de IA esté integrada de forma eficaz y sea capaz de funcionar según lo previsto en escenarios del mundo real.

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Despliegue
Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

Autonomía

- **Agencia humana**
  - Asegúrese de que los usuarios estén al tanto de cuándo una IA genera decisiones, contenido, consejos o resultados. Destaque la interacción con la solución de IA para evitar la dependencia excesiva.
- **Supervisión humana**
  - Garantizar que exista supervisión humana para intervenir cuando sea necesario, asegurando que las decisiones puedan ser revisadas o modificadas por humanos..

Justicia

- Implementar medidas de auditoría sólidas para evaluar regularmente el rendimiento del sistema y el cumplimiento ético.
- Establecer mecanismos para que los usuarios puedan apelar decisiones o presentar quejas, garantizando que existan protecciones para los denunciantes..

Unesco (2023) / Morley (2019)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Despliegue
Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	**DESPLIEGUE**	OPERACIÓN Y MONITOREO

Explicabilidad

- Resalte la importancia de explicar a los usuarios tanto los procesos técnicos como las decisiones humanas relacionadas.
- Enfréntate al reto de simplificar comportamientos humanos complejos en herramientas que sean fácilmente comprensibles, evitando la jerga técnica, > esto aumenta la confianza y la aceptación del sistema.
- Mantener la transparencia sobre las funcionalidades y limitaciones de la IA.

Evaluating long-term impacts

- Evaluar los impactos a largo plazo en la sociedad y el medio ambiente en esta etapa.
- Considere las mejoras necesarias para garantizar que el sistema de IA contribuya positivamente a lo largo del tiempo.

Morley / UNESCO / Prem

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Despliegue
Participantes

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

- **Los ingenieros de IA/ML transforman el modelo de IA prototípico en un servicio o solución desplegada que es accesible para todas las partes interesadas y los usuarios finales.**
- **Los expertos en seguridad protegen el sistema implementado contra los riesgos de ciberseguridad.**
- **Los usuarios finales deben identificar cualquier problema o limitación que pueda no haber sido evidente durante las pruebas en entornos controlados.**
- **Los defensores/representantes de los usuarios recopilan comentarios de los usuarios sobre la usabilidad y el rendimiento de la solución de IA implementada y se aseguran de que la solución implementada cumpla con las expectativas y necesidades de los usuarios.**

(D. De Silva & D. Alahakoon, 2022) / (Junklewitz et al., 2023)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Operación y monitoreo

Después de la implementación, la solución de IA requiere mantenimiento continuo, actualizaciones y evaluación continua.

- Asegúrese de que el modelo siga funcionando como se espera mediante la supervisión periódica de su eficacia.
- Supervise continuamente la actividad del usuario final para recopilar información operativa y detectar posibles problemas.
- Actualice periódicamente el modelo con los datos más recientes para mantener su precisión y relevancia.
- Perfeccione el modelo en función de los comentarios de los usuarios para mejorar la funcionalidad y el rendimiento.
- Las actualizaciones iterativas suelen ser necesarias y deben verse como una parte normal del ciclo de vida de la IA, no como un fracaso.
- Supervise el retorno de la inversión y otras métricas relevantes en función de los objetivos establecidos inicialmente.

Esta fase es fundamental para garantizar que la solución de IA siga siendo eficaz y siga satisfaciendo las necesidades de sus usuarios.

De Silva & Alahakoon (2022) / Data Science Process Alliance

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Operación y monitoreo
Principios éticos a tener en cuenta

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

Seguimiento de los principios éticos

- Revisar y evaluar periódicamente el cumplimiento de los principios éticos del sistema de IA, teniendo en cuenta los posibles cambios en los conjuntos de datos y las estructuras de los modelos..

Participación continua de las partes interesadas

- Mantenga un compromiso activo con los usuarios, las organizaciones y la sociedad en general para recopilar comentarios continuos.
- Establecer mecanismos estructurados para garantizar la recopilación regular y sistemática de información y respuestas..

Unesco (2023)/ HLEG (2019)

Principios Éticos y partes interesadas en el ciclo de vida de desarrollo de la IA

Operación y monitoreo
Participantes

DEFINICIÓN DE PROBLEMAS Y COMPRENSIÓN DEL NEGOCIO	FASE DE DISEÑO	RECOPILACIÓN DE DATOS, COMPRENSIÓN Y PREPARACIÓN	DESARROLLO Y FORMACIÓN DE MODELOS	EVALUACIÓN Y PRUEBAS DE MODELOS	DESPLIEGUE	OPERACIÓN Y MONITOREO

- **Los equipos de DevOps y operaciones de TI supervisan el rendimiento del sistema y garantizan la disponibilidad continua.**
- **Los certificadores de seguridad evalúan la seguridad de la solución de IA desarrollada.**
- **Paneles de expertos, comités directivos u organismos reguladores revisan el proyecto en aspectos técnicos y éticos.**

(D. De Silva & D. Alahakoon, 2022) / (Immersant Data Solutions, 2023) / (Miller, 2022)

Ciclo de vida de desarrollo de la IA

Herramientas

Las herramientas disponibles por consideraciones éticas varían mucho en función de la etapa de desarrollo de la IA. Muchos enfoques pueden ser válidos en todas las etapas, pero algunos solo en ciertas etapas de desarrollo. A continuación, la categorización muestra una imagen aproximada de cómo se pueden dividir las herramientas disponibles entre las etapas de diseño.

Enfoque ex ante, que incluye:

- Marcos: amplios y pueden abarcar varias etapas (por ejemplo, el Marco de la UNESCO que abarca las etapas ex ante y ex post)
- Bibliotecas
- Algoritmos

- Auditorías; Examen formal de los componentes éticos
- Listas
- Métricas
- Posibilidad de iteraciones

Enfoque ex post, que incluye:

- Declaraciones
- Etiquetas
- Licencias
- En la etapa de monitoreo, solo se dispone de unas pocas herramientas

Modified from Prem (2023) / UNESCO (2023)

GRACIAS

Cofinanciado por la Unión Europea

Financiado por la Unión Europea. Las opiniones y puntos de vista expresados solo comprometen a su(s) autor(es) y no reflejan necesariamente los de la Unión Europea o los de la Agencia Ejecutiva Europea de Educación y Cultura (EACEA). Ni la Unión Europea ni la EACEA pueden ser considerados responsables de ellos.

2022-1-ES01-KA220-HED-000085257