Sesión de Formación 1 IDRC: Material de Pre-Lectura

24 abril 2025

En la primera sesión de este programa de formación se analizarán en detalle los siguientes temas.

Contenido:

1. Definición de términos:
    a. IA ética
    b. IA responsable
    c. IA segura
    d. IA de confianza
2. La Ética y la IA
    a. Importancia de la ética en el desarrollo de la IA
3. Marcos/directrices mundiales sobre ética de la IA en el ámbito de la salud
    a. Organización Mundial de la Salud (OMS)
    b. Organización de Cooperación y Desarrollo Económicos (OCDE)
    c. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO)

Nota: Se recomienda leer el siguiente material de lectura previa, ya que constituirá la base de la próxima sesión. La lectura de este material con antelación ayudará a garantizar una experiencia de aprendizaje más enriquecedora durante nuestra segunda sesión del 24 de abril de 2025.

Los 6 Principios Éticos Clave de la OMS para el Uso de la IA en la Salud

1. **Proteger la autonomía:** El uso de la IA en la salud no debe socavar la autonomía humana. En la salud, esto significa que los seres humanos deben seguir controlando los sistemas de salud y las decisiones médicas. Este principio también implica garantizar que los proveedores dispongan de la información necesaria para utilizar los sistemas de IA de forma segura y eficaz, y que las personas comprendan el papel de la IA en su atención. La protección de la privacidad y la confidencialidad y la obtención de un consentimiento informado válido mediante marcos jurídicos adecuados para la protección de datos también forman parte de este principio.

2. **Promover el bienestar humano, la seguridad de las personas y el interés público:** Las tecnologías de la IA no deben dañar a las personas. Deben cumplir los requisitos normativos de seguridad, precisión y eficacia antes de su despliegue, y deben aplicarse medidas continuas de control y mejora de la calidad. Los financiadores, desarrolladores y usuarios tienen el deber continuo de supervisar el rendimiento de los algoritmos de IA para garantizar que funcionan según lo previsto y no afectan negativamente a personas o grupos.

3. **Garantizar la transparencia, la explicabilidad y la inteligibilidad**: Los sistemas de IA deben diseñarse de forma que ofrezcan información clara y transparente sobre las tareas que pueden realizar y las condiciones en las que pueden lograr el desempeño deseado. La transparencia implica revelar el razonamiento que subyace al funcionamiento de los sistemas de IA, los datos utilizados como entrada, así como sus resultados, de modo que los sistemas de IA puedan auditarse y verificarse. La explicabilidad se refiere a la capacidad de ofrecer razones o justificaciones de los resultados de la IA de

forma comprensible para las partes interesadas. La inteligibilidad implica que la información proporcionada sea clara y comprensible.

4. **Fomentar la responsabilidad y la responsabilización**: Aunque las tecnologías de IA realicen tareas específicas, las partes interesadas tienen la responsabilidad de garantizar que estas tareas se realicen de forma adecuada, en condiciones apropiadas y por personas formadas. La responsabilidad puede reforzarse mediante la «garantía humana», que implica la evaluación por parte de pacientes y médicos en el desarrollo y despliegue de la IA. Deben existir mecanismos para cuestionar y exigir reparación a las personas o grupos afectados negativamente por decisiones basadas en algoritmos. Para evitar la difusión de la responsabilidad, un modelo de "responsabilidad colectiva" puede animar a todas las partes implicadas a actuar con integridad y minimizar los daños.

5. **Garantizar la inclusión y la equidad**: La IA utilizada en la atención de salud debe diseñarse para fomentar un uso y un acceso lo más adecuado y equitativo posible, independientemente de factores como la edad, el sexo, los ingresos o la capacidad. Para abordar y minimizar los prejuicios sociales arraigados, es crucial que las instituciones promuevan la diversidad entre quienes desarrollan, supervisan y despliegan la IA. Además, el diseño y la evaluación de las tecnologías de IA deben contar con la participación activa de quienes las utilizarán o se verán afectados por ellas, incluidos los proveedores, los pacientes y las comunidades que puedan sufrir impactos colaterales, independientemente de que sean usuarios del sistema.

6. **Promover una inteligencia artificial sensible y sostenible**: Las tecnologías de IA para la salud deben diseñarse y utilizarse de forma que respondan a las necesidades de los sistemas de salud y sean sostenibles desde el punto de vista medioambiental. Esto incluye considerar los recursos necesarios para su desarrollo, despliegue y mantenimiento, así como su impacto más amplio en el medio ambiente. La viabilidad y adaptabilidad a largo plazo de la IA en el sector sanitario son aspectos clave de este principio.

Estos seis principios pretenden guiar a gobiernos, desarrolladores de tecnología, empresas, sociedad civil y organizaciones intergubernamentales en la adopción de enfoques éticos para el uso apropiado de la IA para la salud.

Fuente: https://www.who.int/publications/i/item/9789240029200

Recomendación de la UNESCO Sobre la Ética de la Inteligencia Artificial

Basándose en la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial, los principios éticos que deben guiar el desarrollo y el uso de los sistemas de la IA son:

1. **Proporcionalidad y no hacer daño**: Se reconoce que las tecnologías de IA no garantizan intrínsecamente el florecimiento humano y medioambiental. Los procesos relacionados con el ciclo de vida de los sistemas de IA no deben exceder lo necesario para alcanzar objetivos legítimos y deben ser adecuados al contexto. En caso de daño potencial a los seres humanos, los derechos humanos, las comunidades, la sociedad o el medio ambiente, deben aplicarse procedimientos de evaluación de riesgos y medidas preventivas. La elección de los sistemas y métodos de IA debe estar justificada por su adecuación y proporcionalidad al objetivo legítimo, la no infracción de los valores fundamentales (especialmente los derechos humanos) y unas bases científicas rigurosas. En escenarios irreversibles o de vida o muerte, debe prevalecer la determinación humana final. Los sistemas de IA no deben utilizarse para el escrutinio social o la vigilancia masiva.

2. **Seguridad y protección**: Los daños no deseados (riesgos de seguridad) y las vulnerabilidades a los ataques (riesgos de seguridad) deben evitarse, abordarse, prevenirse y eliminarse a lo largo del ciclo de vida del sistema de IA para garantizar la seguridad humana, medioambiental y del ecosistema. Una IA segura y protegida es posible gracias a marcos de acceso a los datos sostenibles y protectores de la privacidad que fomenten una mejor formación y validación de los modelos de IA utilizando datos de calidad.

3. **Equidad y no discriminación**: Los actores de la IA deben promover la justicia social, proteger la equidad, y prevenir la discriminación de cualquier tipo, en conformidad con el derecho internacional. Esto incluye garantizar que los beneficios de la IA estén disponibles y sean accesibles para todos, teniendo en cuenta las necesidades específicas de los diferentes grupos y trabajando para abordar las brechas digitales. Deben hacerse esfuerzos razonables para minimizar y evitar reforzar o perpetuar aplicaciones y resultados discriminatorios o sesgados, y debe disponerse de recursos efectivos contra la discriminación. Las brechas digitales y de conocimiento dentro de los países y entre ellos deben abordarse a lo largo de todo el ciclo de vida del sistema de IA para garantizar un trato equitativo.

4. **Sostenibilidad**: El impacto de las tecnologías de la IA en las dimensiones humana, social, cultural, económica y medioambiental debe evaluarse continuamente teniendo en cuenta las implicaciones para la sostenibilidad como un conjunto de objetivos en evolución, como los Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas.

5. **Derecho a la intimidad y a la protección de datos**: La privacidad, esencial para la dignidad humana, la autonomía y la agencia, debe ser respetada, protegida y promovida a lo largo de todo el ciclo de vida del sistema de IA, en consonancia con el derecho internacional y los valores y principios de la Recomendación. Deben establecerse marcos adecuados de protección de datos y mecanismos de gobernanza a nivel nacional o internacional, protegidos por sistemas judiciales, y garantizados durante todo el ciclo de vida del sistema de IA, haciendo referencia a los principios internacionales de protección de datos. Los sistemas algorítmicos requieren evaluaciones del impacto sobre la privacidad y la aplicación de la privacidad desde el diseño.

6. **Supervisión y determinación humanas**: Los Estados miembros deben garantizar que la responsabilidad ética y jurídica de todas las etapas del ciclo de vida del sistema de IA, así como de las soluciones, pueda atribuirse siempre a personas físicas o jurídicas. La supervisión humana incluye la supervisión pública individual e inclusiva. Aunque los seres humanos pueden optar por confiar en la IA para lograr eficacia, la decisión de ceder el control en contextos limitados sigue siendo humana, y los sistemas de IA no pueden sustituir a la responsabilidad y la responsabilidad humana en última instancia. Las decisiones de vida o muerte no deben cederse a los sistemas de IA.

7. **Transparencia y explicabilidad**: La transparencia y la explicabilidad de los sistemas de IA suelen ser esenciales para garantizar el respeto de los derechos humanos, las libertades fundamentales y los principios éticos, así como para que los regímenes de responsabilidad sean eficaces. La transparencia es necesaria para impugnar las decisiones basadas en los resultados de la IA y evitar que se infrinja el derecho a un juicio justo y a un recurso efectivo. El nivel de transparencia y explicabilidad debe ser adecuado al contexto y al impacto, equilibrándolo con otros principios como la privacidad y la seguridad. Las personas deben ser informadas cuando se tomen decisiones basadas en la IA, especialmente cuando afecten a la seguridad o a los derechos humanos, y deben tener la oportunidad de pedir información explicativa y solicitar su revisión. Una mayor transparencia contribuye a sociedades más pacíficas, justas, democráticas e inclusivas al permitir el escrutinio público. La explicabilidad se refiere a hacer inteligibles los resultados de los sistemas de IA y a proporcionar información sobre la entrada, la salida y el funcionamiento de los componentes algorítmicos.

8. **Responsabilidad y responsabilización**: Los actores de la IA y los Estados miembros deben respetar, proteger y promover los derechos humanos, las libertades fundamentales y el medio ambiente,

asumiendo su respectiva responsabilidad ética y legal a lo largo del ciclo de vida del sistema de IA, de acuerdo con la legislación nacional e internacional. La responsabilidad ética y la responsabilidad por las decisiones y acciones basadas en la IA deben atribuirse en última instancia a los actores de la IA en función de su papel. Deben desarrollarse mecanismos adecuados de supervisión, evaluación de impacto, auditoría y diligencia debida para garantizar la responsabilización.

9. **Concienciación y conocimiento**: La concienciación pública y la comprensión de las tecnologías de IA y el valor de los datos deben promoverse a través de la educación accesible, el compromiso cívico, las competencias digitales, la formación ética en IA y la alfabetización mediática e informacional, teniendo en cuenta la diversidad lingüística, social y cultural. El aprendizaje sobre el impacto de la IA debe incluir el aprendizaje sobre, a través de y para los derechos humanos, las libertades fundamentales y el medio ambiente.

10. **Gobernanza y colaboración multisectorial y adaptativa**: Deben respetarse el derecho internacional y la soberanía nacional en el uso de los datos. La participación de las diferentes partes interesadas a lo largo del ciclo de vida del sistema de IA es necesaria para una gobernanza inclusiva de la IA, beneficios compartidos y desarrollo sostenible. Esto incluye a los gobiernos, las organizaciones internacionales, la comunidad técnica, la sociedad civil, los investigadores, los medios de comunicación, la educación, los responsables políticos, el sector privado, las instituciones de derechos humanos y los grupos de jóvenes y niños. Deben establecerse normas abiertas e interoperables para facilitar la colaboración, y las medidas deben tener en cuenta los cambios tecnológicos y permitir una participación significativa de los grupos marginados.

Fuente: https://unesdoc.unesco.org/ark:/48223/pf0000381137

Principios de IA de la OCDE

Los Principios de IA de la OCDE, establecidos en 2019 y actualizados en 2024, proporcionan un marco para el desarrollo y despliegue responsables de los sistemas de inteligencia artificial (IA). Constan de cinco principios basados en valores y cinco recomendaciones para los responsables políticos, con el objetivo de promover una IA digna de confianza que respete los derechos humanos y los valores democráticos.

Principios basados en valores:

1. **Crecimiento inclusivo, desarrollo sostenible y bienestar**: La IA debe contribuir positivamente a la sociedad mejorando las capacidades humanas, promoviendo la inclusión, reduciendo las desigualdades y protegiendo el medio ambiente.

2. **Derechos humanos y valores democráticos, incluidas la imparcialidad y la privacidad**: Los actores de la IA deben defender el estado de derecho, los derechos humanos y los valores democráticos. Esto incluye garantizar la no discriminación, la igualdad, la libertad, la dignidad, la autonomía, la privacidad, la protección de datos y la lucha contra la desinformación mientras respetando la libertad de expresión.

3. **Transparencia y explicabilidad**: Los sistemas de IA deben ser transparentes, proporcionando información significativa para fomentar la comprensión de sus capacidades y limitaciones. Esto incluye informar a las partes interesadas sobre sus interacciones con la IA y, cuando sea posible, ofrecer explicaciones de las decisiones a los afectados.

4. **Robustez, seguridad y protección**: Los sistemas de IA deben funcionar de forma fiable y segura durante todo su ciclo de vida, con medidas para gestionar los riesgos derivados de un uso indebido o de

condiciones adversas. Los mecanismos deben permitir la anulación, reparación o desmantelamiento seguros de los sistemas de IA cuando sea necesario.

5. **Responsabilidad:** Los actores de la IA tienen la responsabilidad de garantizar que los sistemas de IA funcionen según lo previsto y respetan estos principios. Esto implica mantener la trazabilidad de los datos y los procesos, aplicar una gestión sistemática de los riesgos y abordar cuestiones como la parcialidad y la privacidad.

Recomendaciones para los responsables políticos:

1. **Invertir en investigación y desarrollo de IA**: Los gobiernos deben apoyar la inversión pública y privada en I+D de IA para promover la innovación en IA fiable, abordando los retos técnicos y las implicaciones sociales.

2. **Fomentar un ecosistema inclusivo que propicie la IA**: Desarrollar un entorno digital dinámico y sostenible que incluya datos, tecnologías, infraestructuras y mecanismos de intercambio de conocimientos, promoviendo un intercambio de datos seguro y ético.

3. **Creación de un entorno político y de gobernanza interoperable para la IA**: Promover políticas ágiles que faciliten la transición de la IA del desarrollo a la implantación, incluido el uso de entornos de prueba controlados y enfoques basados en los resultados, garantizando al mismo tiempo la interoperabilidad entre jurisdicciones.

4. **Creación de capacidades humanas y preparación para la transformación del mercado laboral**: Colaborar con las partes interesadas para preparar a la sociedad para los cambios en la mano de obra debidos a la IA, incluyendo iniciativas de educación y formación, un apoyo justo a la transición de los trabajadores y la promoción del uso responsable de la IA en el lugar de trabajo.

5. **Cooperación internacional para una IA confiable:** Implicarse en colaboración internacional para promover estos principios, compartir conocimientos sobre IA, desarrollar normas técnicas mundiales y crear indicadores comparables para evaluar los avances en la implantación de la IA.

Estos principios y recomendaciones pretenden orientar a los agentes de la IA y a los responsables políticos en el fomento de sistemas de IA que sean innovadores, fiables y acordes con los valores de la sociedad.

Fuente: https://oecd.ai/en/ai-principles

Preguntas de reflexión:

- ¿Cómo imagina poner en práctica estos principios en sus propios proyectos? ¿Qué medidas o estrategias concretas podría adoptar para aplicarlos eficazmente en su contexto específico?
- ¿Qué principio es el más difícil de implementar?

Otras recomendaciones de lectura :

1. **Ethics guidelines for trustworthy AI, High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence**
https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

2. **The Cambridge Handbook of The Law, Ethics And Policy Of Artificial Intelligence**
https://doi.org/10.1017/9781009367783

3. **Insights into suggested Responsible AI (RAI) practices in real-world settings: a systematic literature review**
https://link.springer.com/article/10.1007/s43681-024-00648-7?fromPaywallRec=false

4. **Ethics in AI through the practitioner's view: a grounded theory literature review**
https://link.springer.com/article/10.1007/s10664-024-10465-5?fromPaywallRec=false

5. **Empowering responsible AI practices**
https://www.microsoft.com/en-us/ai/principles-and-approach