



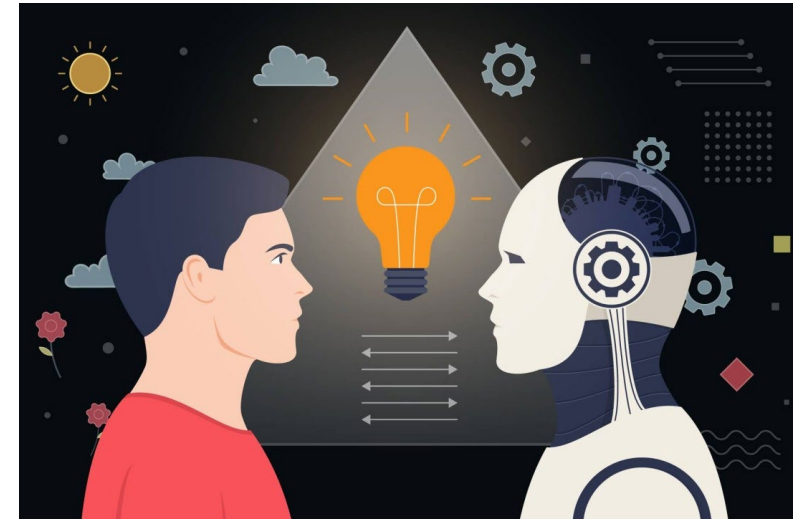
# Ética en la Inteligencia Artificial

Jonnatan Arias Garcia

# Introducción

La ética en la IA pretende abordar diversos desafíos como:

- Sesgos con poblaciones
- Información falsa e Información parcial
- Privacidad
- Transparencia
- Implicaciones
- Confianza
- Uso cético



En casos mas específicos como en la medicina

- Cuidados en el rendimiento, fiabilidad y las implicaciones de los desenlaces

# Enfoques

El manejo de la IA debe ir hacia la **validación de Impactos y riesgos** tanto Sociales como económico

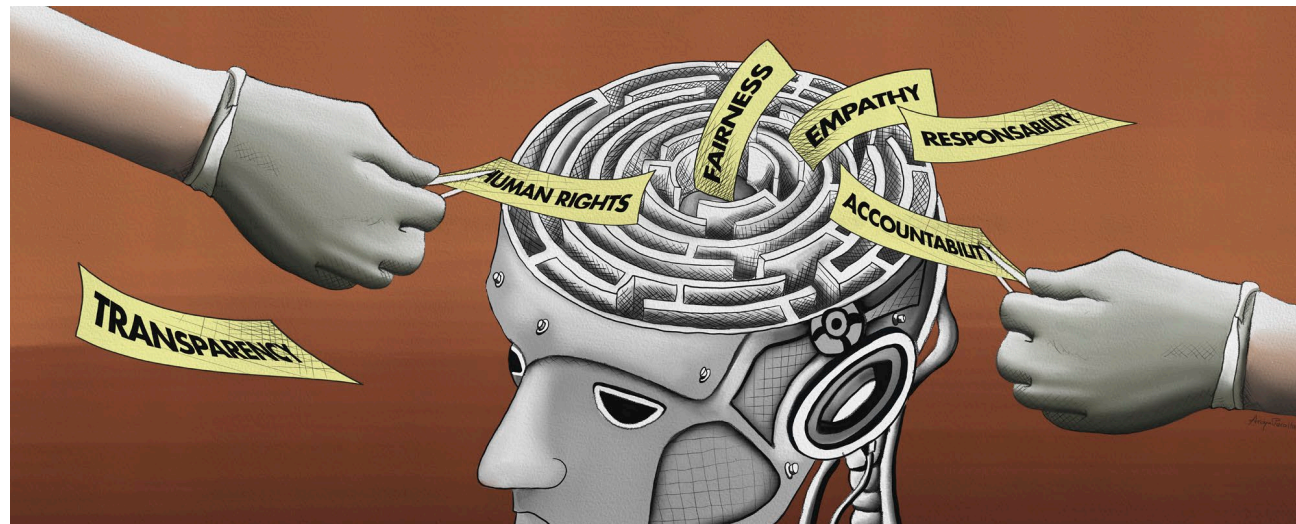
Por ende es necesario un tipo de **gobernanza** que se centre en validar todo con valores y leyes, primando el análisis del Diseño de la IA, su desarrollo y despliegue.



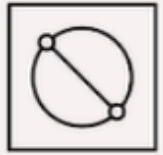
Para finalmente dar Licencias de operación confiables y transparentes

# Enfoques Necesarios

1. Construir inteligencias Artificiales basadas en normas éticas con entendimiento y conciencia de los impactos y peligros.
2. Estructura de gobernanza que funcione a escala (Necesaria para dar apoyo a la AI-ética)
3. Operacionalización que pueda manejar claridad de los principios y pilares de la ética en la IA



# Definiendo la ética



Data responsibility  
and privacy



Moral agency



Accountability



Inclusion

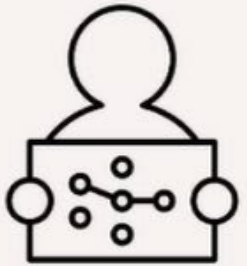


Value  
alignment



Technology  
misuse

# Pilares de la ética en IA



Explainability



Fairness



Robustness



Transparency



Privacy



- **Explainability / explicabilidad**: Como y Porque se llego a un desenlace particular o recomendación
- **Fairness / Justicia**: Tratar individuos o grupos equitativamente
- **Robustness / Robustez** : Manejar efectivamente condiciones excepcionales como datos anormales o ataques adversarios
- **Transparency / Transparencia**: Compartir información apropiada con humanos sobre como se diseño y desarrollo la IA.
- **Privacy / Privacidad**: Dpara priorizar y asegurar los datos y privacidad humana

# Entendiendo el Sesgo en AI

- El sesgo tiene que ver con el comportamiento no deseado en el sistema de IA
- El sesgo da desventajas sistemáticas a ciertos grupos o individuos.

## #BreakTheBias





# Como surge el sesgo en AI

La IA es entrenada usando decisiones que los humanos han hecho en el pasado

Estas decisiones humanas podrían tener implícito o explícitamente un sesgo, y estos podrían reflejarse en el entrenamiento

- Las muestras de datos podrían introducir sesgo (por ejemplo, conducir sobre representaciones o sub representaciones)
- El sesgo también puede introducirse a través de la fase de procesamiento o de la introducción de datos (por ejemplo si la población donde se toma es sesgada)



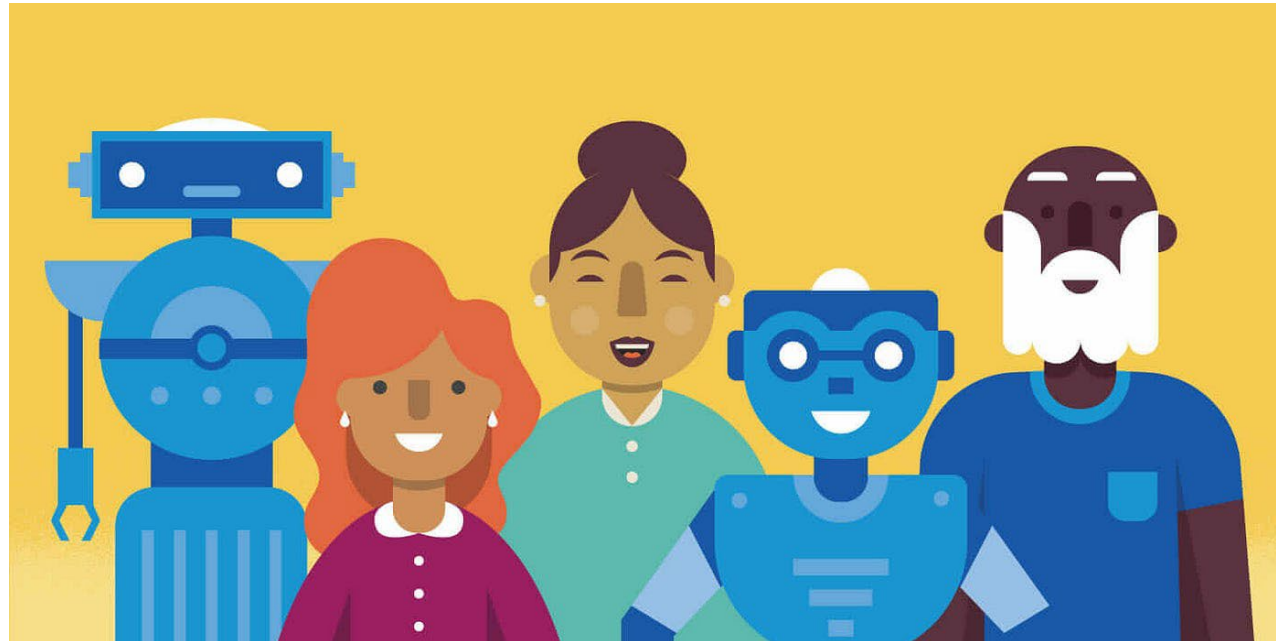
# Mitigación del sesgo

- Usando equipos de trabajo diversos con diferentes perspectivas
- Búsqueda de data de alta calidad
- Enfoques técnicos



# Consideraciones frente al sesgo

- Debemos asegurarnos que el impacto sea siempre positivo
- Eliminar sesgo según estándares de justicia y equidad
- Validando la inclusión



# Ética + Regulaciones

Una Regulación es una regla gubernamental sobre el desarrollo y despliegue usando la legalidad y ética

Tenemos 5 políticas imperativas para sistemas de IA:

- Diseñar un estándar líder oficial para la ética en IA
- Desarrollar diferentes reglas para diferentes riesgos
- No esconderla, hacer la IA transparente
- Explicar la IA y su uso
- Probar tus IA para sesgos



# Gobernanza de la AI

Organización que Actúe como gobierno IA a través de la incorporación de instrucciones, personal, procesos y sistemas

- El objetivo es dar confianza en la IA estableciendo requerimientos para rendición de cuentas, responsabilidad y supervisión

Sus principales beneficios serian

- Confianza
- Eficiencia
- Cumplimiento



# ESG

Debemos tener especial cuidado con el modelo ESG (Enviromental, Social, Governance)

- Impacto medio ambiental
- Impacto equitativo
- Impacto ético





# Como generar Confianza en la IA

- Alcance y plan
  - ¿Cuáles son las expectativas empresariales en cuanto a equidad o transparencia?
  - ¿Qué regulación debemos cumplir?
  - ¿Cómo obtenemos acceso a atributos de datos confidenciales?
  - ¿Cuál es la frecuencia con la que se deben proporcionar las explicaciones?
- Recolección y organización
- Construcción y pruebas
- Validación y entrega
- Monitorear y gestionar



