



Sesión 3 de Formación IDRC: Material de Prelectura

22 mayo 2025

Los siguientes temas se explorarán en detalle en la tercera sesión de este programa de formación:

Contenido:

- I. Sesgos en la IA
 - a. Definición
 - b. Origen de los sesgos
- II. Partes interesadas en IA para la salud
 - a. Enfoques de gestión de partes interesadas
 - b. Cómo visualizar y clasificar a las partes interesadas en la gestión de proyectos
 - c. Panorama de los grupos de partes interesadas Stakeholder management approaches
- III. Marcos de identificación de partes interesadas
 - a. Perfilado de partes interesadas
 - b. Modelo de prominencia (Salience Model)
 - c. Matriz de poder e interés (Power-Interest Grid)
- IV. Participación de las partes interesadas
 - a. Escalera de participación (Ladder of Participation)
 - b. IA inclusiva (Inclusive AI)

Nota: Se recomienda consultar la documentación preparatoria adjunta, que servirá como base para la sesión del 22 de mayo de 2025. Leer este material con antelación ayudará a garantizar una experiencia de aprendizaje más enriquecedora y participativa.

¿Qué es el sesgo?

Según la Escuela de Salud Pública de Harvard, el sesgo se define como "juicios instantáneos que hacemos sobre personas y situaciones, basados en años de socialización subconsciente".

Sesgo implícito vs. sesgo explícito

- **Sesgo implícito**
 - o De acuerdo con Mahzarin Banaji, nuestro cerebro está naturalmente programado para emitir juicios rápidos usando suposiciones y experiencias previas, frecuentemente sin que seamos conscientes de ello. Estos sesgos inconscientes no son intencionales; surgen de cómo nuestro cerebro define lo "normal" o "positivo", y se moldean por historias personales, cultura, medios de comunicación e influencias sociales.
 - o Son inconscientes, involuntarios y a menudo sutiles.
 - o Todos los tenemos y afectan decisiones sin que nos demos cuenta.
- **Sesgo explícito**
 - o Se refiere a actitudes y creencias conscientes que mantenemos, especialmente sobre grupos específicos de personas.
 - o Estos sesgos son deliberados y somos plenamente conscientes de ellos.
 - o Generalmente se basan en rasgos observables o identidades sociales.

Fuente: [*Understanding Unconscious Bias*](#)

¿Qué es el sesgo en IA?

Según IBM, el sesgo en IA (también llamado sesgo de aprendizaje automático o sesgo algorítmico) ocurre cuando resultados parcializados surgen debido a prejuicios humanos que distorsionan los datos de entrenamiento originales o el algoritmo de IA, generando resultados erróneos y potencialmente dañinos.

Fuentes de sesgos:

- **Sesgo algorítmico:** Ocurre cuando el problema está mal definido o la retroalimentación dada al sistema de aprendizaje automático es inútil, lo que lleva a resultados engañosos o incorrectos.
- **Sesgo cognitivo:** Dado que los humanos construyen la IA, sus prejuicios inconscientes pueden influir tanto en los datos como en el comportamiento del modelo, sin que lo noten.
- **Sesgo de confirmación:** Cuando la IA refuerza suposiciones existentes en los datos en lugar de descubrir nuevos patrones, profundiza prejuicios previos y limita la innovación.
- **Sesgo de exclusión:** Surge cuando variables o grupos importantes se omiten involuntariamente del conjunto de datos, a menudo por descuido o enfoques limitados de los desarrolladores.
- **Sesgo de medición:** Causado por recopilación de datos incompleta o sesgada, donde se omiten segmentos clave de la población. Ejemplo: Estudiar solo a graduados para entender el éxito universitario ignora a los que dejaron sus estudios universitarios.
- **Sesgo de homogeneidad del exogrupo:** Los desarrolladores pueden entender mejor su propio grupo y generalizar sobre otros, lo que lleva a sistemas de IA que clasifican mal o marginan a grupos poco representados en los datos.
- **Sesgo por prejuicio:** Ocurre cuando estereotipos sociales se incrustan en los datos, generando resultados parcializados. Ejemplo: Asociar ciertos trabajos con un género específico.
- **Sesgo de memoria (recall bias):** Aparece durante el etiquetado de datos cuando juicios subjetivos generan etiquetas inconsistentes, distorsionando el aprendizaje.
- **Sesgo de muestra o selección:** Sucede cuando los datos de entrenamiento son demasiado pequeños, homogéneos o incompletos, haciendo que el modelo generalice mal. Ejemplo: Entrenar solo con profesores de un contexto puede excluir a otros con diferentes cualificaciones.
- **Sesgo por estereotipos:** Los sistemas de IA pueden reforzar estereotipos culturales o de género (ejemplo: traducir ciertos roles a grupos específicos). Eliminar atributos sensibles (como género o raza) no siempre resuelve el problema y puede perjudicar el rendimiento del modelo.

Fuente: [¿Qué es el sesgo de la IA?](#)

¿Quién es una parte interesada?

Según la ISO 15288, una parte interesada (o stakeholder) es un individuo u organización con un derecho, participación, reclamo o interés en un sistema, especialmente en cómo este satisface sus necesidades y expectativas.

Enfoques de Gestión de Partes Interesadas

Eskerod & Huemann (2013) distinguen dos enfoques de gestión de partes interesadas:

1. Gestión **-de**-partes-interesadas (Management-**of**-stakeholders): Las partes interesadas se consideran como recursos, donde la atención se basa en su potencial para ayudar o perjudicar (visión instrumental).
2. Gestión **-para**-partes-interesadas (Management-**for**-stakeholders): Todas las partes interesadas merecen atención de gestión, independientemente de su capacidad para contribuir (visión normativa).

3. Los autores sugieren que ambos enfoques son esenciales para el desarrollo sostenible en la gestión de partes interesadas de proyectos.

Source: [*Stakeholder roles in artificial intelligence projects*](#) (only available in English)

Cómo Visualizar y Clasificar a las Partes Interesadas en la Gestión de Proyectos

- **Visión Instrumental (PMBOK, 2021):** Las partes interesadas se identifican según su capacidad para apoyar o perjudicar los objetivos del proyecto.
- **Visión Normativa:** Hace hincapié en la responsabilidad ética de los gestores para gobernar las organizaciones y responsabilizar a todas las partes relevantes, incluidas aquellas pasivas que suelen pasarse por alto.
- **Conclusión Clave:** La categorización de partes interesadas depende de los objetivos del proyecto y de cómo la organización o el gestor definen el éxito.

Source: [*Stakeholder roles in artificial intelligence projects*](#) (only available in English)

Panorama de los Grupos de Partes Interesadas

1. **Partes Interesadas en el Desarrollo**
 - a. Tienen una relación formal con el proyecto durante la fase de desarrollo de la IA.
 - b. Pueden ser internas o externas según la gobernanza del proyecto.
2. **Partes Interesadas en el Uso**
 - a. Mantienen una relación formal durante la fase de utilización de la IA.
 - b. Su participación depende de la estructura organizacional y el contexto.
 - c. Relación formal durante la fase de utilización de la IA.
 - d. La participación depende de la estructura organizativa y del contexto.
3. **Partes Interesadas Externas**
 - a. No tienen un rol directo en el proyecto, pero pueden influir o verse afectadas.
4. **No Partes Interesadas**
 - a. Forman parte del ecosistema de IA pero no están asociadas al proyecto (ej.: académicos, socios industriales).

Source: [*Stakeholder roles in artificial intelligence projects*](#) (only available in English)

Marco de Identificación de Partes Interesadas:

1. **Perfilado de Partes Interesadas**
 - Aunque no es una técnica de "mapeo" en el sentido visual, la creación de perfiles (también conocidos como personas) para cada parte interesada o grupo de interés puede ser extremadamente valiosa. Cada perfil incluirá detalles como:
 - Sus intereses y necesidades
 - Nivel de influencia
 - Preferencias de comunicación
 - Historial de interacciones con el proyecto u organización
 - **¿Cuándo utilizar el perfilado de partes interesadas?**
 - Cuando se requiere un entendimiento detallado de cada parte interesada o grupo
 - Especialmente útil en:
 - Proyectos a largo plazo
 - Escenarios de gestión continua de partes interesadas
 - Resulta valioso para personalizar estrategias de comunicación y participación para cada parte interesada o grupo
2. **Modelo de Prominencia (Salience Model)**
 - Esta técnica considera tres atributos:
 - **Poder** (la capacidad de influir en el proyecto)
 - **Legitimidad** (lo apropiado de su participación)

- **Urgencia** (la presión de sus demandas)
- Estos tres atributos se combinan en un diagrama de Venn para clasificar a las partes interesadas.
- ¿Cuándo usar el modelo de prominencia? En proyectos complejos donde se requiere un entendimiento profundo de las partes interesadas, especialmente cuando se necesita considerar simultáneamente:
 - Urgencia
 - Legitimidad
 - Poder
- Ejemplo práctico:
Una universidad que lanza un nuevo programa académico podría utilizar una matriz de prominencia para:
 - Identificar y priorizar:
 - Miembros de la facultad
 - Estudiantes potenciales
 - Agencias de acreditación
 - Egresados
 - Socios de la industria local
- Dado que el lanzamiento de un nuevo programa suele tener plazos definidos, este modelo ayuda a determinar qué partes interesadas tienen mayor prominencia según su poder, legitimidad y urgencia. Esta información guiaría a la universidad en:
 - Personalizar sus estrategias de comunicación
 - Diseñar planes de participación efectiva
 - Alinear expectativas durante el lanzamiento y desarrollo continuo del programa

Salience Model in AI in Health

- **Poder (Nguyen et al., 2019): Capacidad de una parte interesada para influir en el proyecto:**
 - i. Coercitivo – Puede sancionar el incumplimiento.
 - ii. De recompensa – Puede ofrecer incentivos.
 - iii. Legítimo – Posee autoridad formal.
 - iv. Experto – Tiene conocimiento crítico.
 - v. Referente – Influye como modelo a seguir.
- **Legitimidad (Mitchell et al., 1997): ¿Está justificada su participación?**
 - i. Según contratos (términos, SOW), leyes, roles laborales o normas organizacionales.
- **Urgencia (Jones, 1991): ¿Requiere su demanda acción inmediata?**
 - i. Inmediatez temporal – Plazo para actuar.
 - ii. Proximidad – Grado de afectación directa (etapas de Desarrollo, Uso o Consecuencias).
- **Daño potencial (Someh et al., 2019): ¿Podría resultar perjudicada?**
 - i. Físico/Psicológico – Lesiones, vigilancia, estrés.
 - ii. Pérdidas – Privacidad, financieras, seguridad laboral.
 - iii. Perjuicios – Reputación, confianza, medio ambiente.

Fuente:

- [Stakeholder Mapping: When, Why, and How to Map Stakeholders](#)
- [Stakeholder roles in artificial intelligence projects \(only available in English\)](#)

3. Matriz de Poder e Interés (Power-Interest Grid)

- **Definición:**
Herramienta común donde las partes interesadas se ubican en una matriz según:
 - **Poder** (capacidad de influencia)

- **Interés** (nivel de preocupación)
...respecto a un proyecto o decisión.
- **¿Cuándo usarla?**
En proyectos *sencillos* donde se requiere priorizar rápidamente a las partes interesadas según su poder e interés.
- **Ejemplo:**
Una ONG de preservación de biodiversidad podría usar esta matriz para:
 - Identificar y priorizar actores como:
 - Agencias gubernamentales
 - Activistas ambientales influyentes
 - Grandes donantes
 - Científicos
 - Comunidades locales
 - Optimizar la colaboración en iniciativas de conservación.
 - **Alto poder - Alto interés:**
Estos actores clave son tomadores de decisiones y tienen el mayor impacto en el éxito del proyecto. Por lo tanto, debes gestionar sus expectativas muy de cerca mediante comunicación frecuente y atención personalizada.
 - **Alto poder - Bajo interés:**
Estas partes interesadas deben mantenerse informadas y satisfechas, aunque muestren poco interés, dado su poder de influencia. Se recomienda manejar estas relaciones con cautela, ya que podrían ejercer su poder de manera negativa si se sienten desatendidas.
 - **Bajo poder - Alto interés:**
Mantén a estos grupos adecuadamente informados y en diálogo constante para identificar posibles problemas. Suelen ser valiosos para aportar detalles operativos y perspectivas especializadas del proyecto.
 - **Bajo poder - Bajo interés:**
Monitorea a estas partes con comunicaciones mínimas pero estratégicas, evitando saturarlas con información excesiva. Un seguimiento básico es suficiente.

Source: [Stakeholder Analysis using the Power Interest Grid](#)

Escalera de Participación

La escalera de Arnstein es un modelo que ilustra los distintos niveles de participación ciudadana en los procesos de toma de decisiones, especialmente en la planificación urbana, la administración pública y la formulación de políticas. Este modelo muestra que no toda participación es igual: algunas formas son meramente simbólicas (tokenismo), mientras que otras implican una verdadera distribución del poder.

1. **No Participación**
 - a. **Manipulación:** Los funcionarios “educan” o “curan” a los participantes en lugar de escuchar genuinamente sus opiniones. Los ciudadanos no tienen ninguna influencia real.
 - b. **Terapia:** La participación se utiliza como una forma de terapia en lugar de generar un cambio real; nuevamente, los ciudadanos no tienen poder de decisión.
2. **Grados de Tokenismo (Participación Simbólica)**
 - a. **Información:** Se informa a los ciudadanos sobre sus derechos, responsabilidades y opciones, pero no se les da oportunidad de dar retroalimentación ni negociar.
 - b. **Consulta:** Se solicita la opinión de los ciudadanos (por ejemplo, mediante encuestas o audiencias públicas), pero no hay garantía de que sus aportes influirán en las decisiones.

- c. **Aplacamiento:** Los ciudadanos pueden asesorar o participar en la planificación, pero quienes detentan el poder se reservan el derecho de juzgar la legitimidad o viabilidad de sus propuestas.
- 3. Grados de Poder Ciudadano**
- a. **Asociación (Partnership):**
El poder se redistribuye mediante la negociación entre ciudadanos y autoridades; se comparten las responsabilidades de planificación y toma de decisiones.
 - b. **Poder Delegado:**
Los ciudadanos tienen autoridad predominante en la toma de decisiones sobre un programa o política específica.
 - c. **Control Ciudadano:**
Los ciudadanos gestionan todo el proceso, con plena autoridad para tomar decisiones.

Source: [*Ladder of Citizen Participation*](#)

IA Inclusiva:

- La participación pública no es intrínsecamente “buena” y puede ser coercitiva.
- La participación pública no puede solucionar problemas relacionados con productos y sistemas que son perjudiciales por naturaleza.
- La participación pública no garantiza automáticamente la inclusión de miembros de comunidades socialmente marginadas ni de perspectivas diversas.
- El “público” –y los muchos grupos y comunidades que lo componen– no debe tratarse como un bloque homogéneo ni como entidades completamente aisladas.
- Se debe “encontrar a los participantes donde están” para apoyar su participación plena.
- No se debe obligar al público general a participar en procesos de consulta para proteger sus derechos o bienestar.
- La confianza no es un requisito previo ni un resultado garantizado de la participación pública.
- Llevar a cabo una participación pública efectiva requiere formación y experiencia.

Source: [*Guidance for Inclusive AI*](#)

Pregunta de Reflexión:

1. ¿Es posible eliminar completamente los sesgos en los sistemas de IA?

Herramientas para las Actividades de la Sesión

Para la tercera sesión, se utilizarán las siguientes herramientas interactivas. Es importante familiarizarse con cada una de ellas con antelación para asegurar una experiencia participativa y dinámica:

- **Mentimeter** – <https://www.mentimeter.com/es-ES>
Se recomienda tener el teléfono móvil a mano para escanear los códigos QR que se mostrarán durante la sesión del 22 de mayo. Si tu teléfono no permite escanear códigos QR, no te preocupes: los facilitadores también compartirán los enlaces directos a las actividades a través del chat.
- **Miro Board** – <https://miro.com/es/>
Durante la sesión se compartirá un enlace a la pizarra colaborativa de Miro. Este espacio permitirá a todos los participantes aportar sus ideas y reflexiones de forma visual y escrita.

Recomendaciones de Lectura Adicional:

1. [*¿Qué es el sesgo de la IA?*](#)
2. [*Inteligencia artificial y sesgos algorítmicos*](#)
3. [*¿Salud para quién? Interseccionalidad y sesgos de la inteligencia artificial para el diagnóstico clínico*](#)
4. [*Guía de planificación de participación de las partes interesadas*](#)
5. [*Pensamiento de diseño y marcos éticos para la Inteligencia Artificial: una mirada a la participación de las múltiples partes interesadas*](#)