AFD AGENCE FRANÇAISE DE DÉVELOPPEMENT

Funded by the European Union

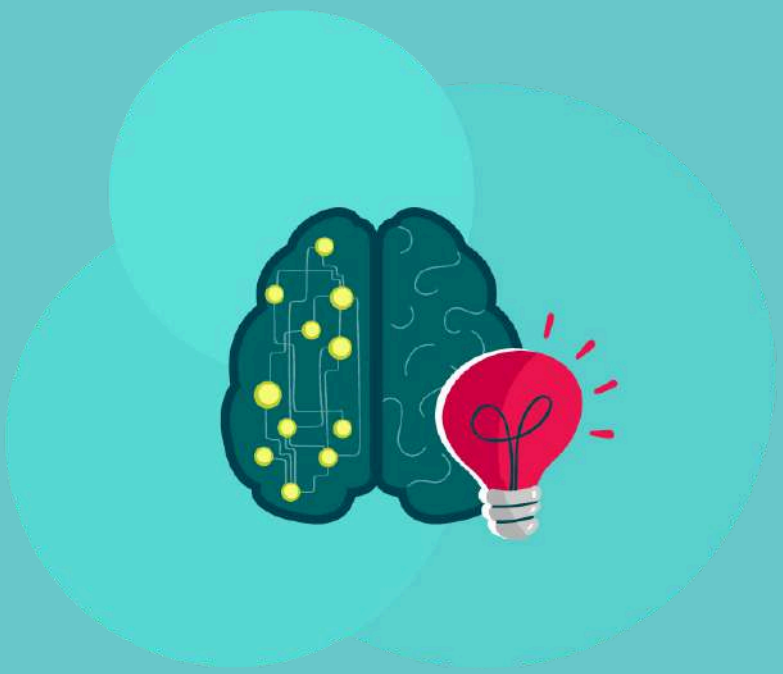
forus CONNECT SUPPORT INFLUENCE

INNOVUS

Gestión de ONG & Ética de la IA en Operación

Sesión #3

Ética y sesgo de la IA en la toma de decisiones



Actividad de Aprendizaje

1. Información general

Esta actividad te permitirá identificar procesos repetitivos en tu organización donde la IA podría generar eficiencia, reducir errores y liberar tiempo del equipo humano para tareas estratégicas o sensibles.

2. Guión de la sesión 3 (descrita dispositiva por dispositiva)

Minuto 1 - 2	**Acción del Facilitador**
	Da la bienvenida cordial al grupo. Pregunta en el chat de dónde se conectan.

Guión/instrucciones

Hola a todas y todos. Hoy finalizamos con el Módulo 3 de esta ruta formativa sobre Inteligencia Artificial para ONGs. Mi nombre es X y seré su facilitador/a.

Minuto 3	**Acción del Facilitador**
	Llamado a la reflexión.

Guión/instrucciones

Les invito a pensar por un momento en una situación real: una ONG que debe priorizar a quién brindar apoyo alimentario en una comunidad vulnerable. Han recolectado datos y deciden usar una herramienta de IA para segmentar beneficiarios por nivel de riesgo. Todo parece lógico, ¿verdad? Pero luego se dan cuenta que la IA está ignorando sistemáticamente a mujeres jefas de hogar porque en los datos históricos, los hombres fueron priorizados.

¿Qué harías tú? ¿Quién es responsable de esa decisión? ¿El algoritmo? ¿El programador? ¿La ONG?

Hoy vamos a hablar de esto: del poder y los peligros de delegar decisiones sensibles a la inteligencia artificial. Este módulo no es para técnicos: es para cualquier persona que lidere, diseñe o implemente programas sociales donde la IA pueda estar involucrada en algún nivel.

No se trata de decir “la IA es mala” o “la IA es buena”. Se trata de entender **cómo hacerla más ética, más justa y más humana**, especialmente cuando trabajamos con poblaciones en riesgo o con derechos vulnerados.

Así que empecemos cuestionandonos lo más básico: ¿confiaríamos en una decisión que afecta a alguien más si no sabemos cómo se tomó?”

Minuto 4	**Acción del Facilitador**

Guión/instrucciones

Minuto 5	**Acción del Facilitador**

Guión/instrucciones

En este módulo abordaremos cómo los algoritmos pueden heredar o amplificar prejuicios humanos, qué consecuencias tiene esto para organizaciones sin fines de lucro, y cómo podemos recurrir a marcos éticos y normativas vigentes para asegurarnos de usar la IA de manera responsable. Al final, desarrollaremos conjuntamente una serie de principios o pautas éticas adaptadas a nuestras organizaciones.

Minuto 6	**Acción del Facilitador**

Guión/instrucciones

Las organizaciones sin fines de lucro frecuentemente trabajan con poblaciones vulnerables y manejan información sensible. Si bien la inteligencia artificial puede potenciar nuestro impacto (automatizando tareas, analizando grandes datos, etc.), también puede causarnos problemas serios si la usamos sin precaución ética. Imaginemos que una ONG utiliza un algoritmo para decidir quién recibe un servicio: si este algoritmo está sesgado, podríamos sin querer **excluir o discriminar** a justo aquellos grupos a los que queremos ayudar. Recordemos que cuando los sesgos no se abordan, **obstaculizan la participación de las personas en la sociedad y reducen el potencial de la IA**, generando desconfianza especialmente entre grupos marginados. Para las ONG, cuya legitimidad se basa en la confianza de sus comunidades y donantes, **usar IA de forma ética es imprescindible**. No se trata sólo de evitar daños, sino de alinear la tecnología con nuestros valores misionales de justicia, inclusión y servicio.

Minuto 7

Acción del Facilitador

Guión/instrucciones

Definamos primero el concepto: hablamos de *sesgo algorítmico* cuando un sistema de IA produce resultados sistemáticamente parciales o discriminatorios. En otras palabras, la IA "prefiere" o beneficia injustamente a ciertos grupos sobre otros debido a cómo fue construida.

¿Cómo ocurre esto? Principalmente de tres formas:
(1) **Sesgo en datos de entrenamiento:** Si las bases de datos con que entrenamos al algoritmo están desbalanceadas o reflejan prejuicios (por ejemplo, datos demográficos incompletos, históricos de decisiones discriminatorias), la IA aprenderá esas tendencias.
(2) **Sesgo en el algoritmo:** Puede pasar que la forma en que se programó el modelo dé más peso a ciertos factores de manera injusta, quizá sin que el desarrollador se dé cuenta.
(3) **Sesgo humano o cognitivo:** Nuestras propias suposiciones y decisiones al elegir datos o ajustar el sistema pueden introducir sesgos sutiles.

El resultado final es que la IA puede **perpetuar estereotipos y desigualdades** de la sociedad. Por ejemplo, si siempre se le ha negado cierto servicio a un grupo, un algoritmo entrenado con esos registros podría seguir negándoselos en el futuro simplemente porque "así fue siempre". Sin intervención consciente, la IA termina **amplificando prejuicios** en vez de eliminarlos.

Minuto 8	Acción del Facilitador

Guión/instrucciones

Veamos algunos ejemplos concretos que han salido a la luz pública:

- En **salud**, los algoritmos usados para predecir riesgos médicos han fallado con poblaciones minoritarias. Un caso mostró que herramientas de diagnóstico asistido por computadora tenían peores resultados con pacientes negros que con blancos. ¿La razón? Los datos usados para “entrenar” al modelo provenían mayormente de ciertas poblaciones, dejando a otras subrepresentadas. Esto significó que a pacientes afrodescendientes se les evaluaba con menos precisión, una inequidad preocupante en algo tan crítico como la atención médica.

- En **contratación de personal**, una empresa de tecnología (Amazon) desarrolló una IA para filtrar currículums, buscando a los mejores candidatos automáticamente. ¿Qué sucedió? El sistema aprendió de datos históricos donde la industria tecnológica contrataba mayoritariamente hombres, y concluyó que los candidatos masculinos eran preferibles. Empezó a descartar solicitudes de mujeres o de universidades femeninas, replicando un sesgo de género evidente. La empresa tuvo que retirar esa herramienta al descubrir el sesgo.

- En el **sistema judicial**, algunas ciudades han usado IA para predecir dónde ocurrirán delitos o evaluar el riesgo de reincidencia de un acusado (por ejemplo, el polémico software COMPAS en EE.UU.). Estos algoritmos *parecen* objetivos, pero se alimentan de datos policiales históricos que están marcados por décadas de sesgo racial. El resultado: las comunidades minoritarias aparecían como de “mayor riesgo” simplemente porque eran objeto de más vigilancia y arrestos en el pasado. Esto refuerza perfiles raciales y puede guiar indebidamente las decisiones judiciales, perpetuando la injusticia que la IA debía ayudar a corregir.

- Otro ejemplo ocurrió en **publicidad en línea**: un estudio de Carnegie Mellon descubrió que la plataforma de anuncios de Google mostraba ofertas de empleo de alto salario con más frecuencia a usuarios identificados como hombres que a mujeres. Sin ninguna intervención humana directa, el algoritmo optimizó clics de una forma que terminó promoviendo un sesgo de género en oportunidades laborales.

Estos casos son **alertas tempranas**. Si esto pasa en empresas o gobiernos, **¿qué implicaciones tendría en una ONG?** Pensemos en ello: una IA sesgada podría, por ejemplo, dejar fuera de un programa social a cierto grupo étnico, o recomendar atención prioritaria solo a comunidades urbanas porque tiene menos datos de zonas rurales. Por eso es vital entender estos ejemplos, para reconocer sesgos potenciales en nuestras propias aplicaciones de IA y aprender cómo prevenirlos.

Minuto 9

Acción del Facilitador

Guión/instrucciones

Para una ONG, los sesgos algorítmicos no son un tema meramente técnico: **tienen consecuencias muy reales en nuestra labor social**. Explorémoslas:

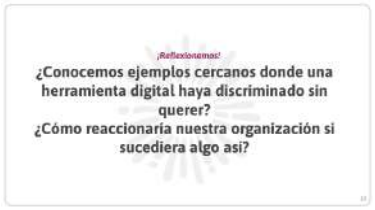
En primer lugar, existe el riesgo de **discriminación involuntaria**. Una decisión automatizada mal calibrada podría dejar fuera a personas que *deberían* ser beneficiarias. Imaginemos una ONG de educación que usa IA para seleccionar candidatos a becas y, sin darse cuenta, el sistema privilegia a ciertos barrios urbanos porque tiene más datos de ellos, dejando en desventaja a jóvenes de zonas rurales. Estaríamos, sin querer, **negando oportunidades** a quienes quizá más las necesitan. Esto socava la misión de equidad e inclusión que guían a la mayoría de organizaciones sociales.

Luego, pensemos en la **confianza**. Las ONG dependen de la confianza del público: de las comunidades a las que sirven, de los voluntarios y de los financiadores. Si se descubre que nuestras decisiones asistidas por IA son injustas o están sesgadas, la percepción pública puede volverse negativa rápidamente. Un ejemplo real: una asociación en EE.UU. implementó un chatbot para orientar a personas con trastornos alimenticios, pero el chatbot terminó dando consejos dañinos debido a datos defectuosos, lo que causó un escándalo y obligó a retirarlo. La lección fue clara: un *pequeño error ético* con IA puede derivar en un gran problema reputacional, erosionando años de buena imagen en la comunidad.

También están los **riesgos legales**. Si una ONG usa una plataforma de IA de terceros para, digamos, filtrar currículums de empleo o seleccionar beneficiarios, y esa herramienta tiene sesgos discriminatorios, podríamos enfrentar reclamos o incluso demandas por discriminación. Un abogado especializado en el sector sin ánimo de lucro advertía que si hay sesgo en un software de reclutamiento y una persona no es contratada por ese sesgo, **la organización podría afrontar un caso legal por**

discriminación. Es decir, la responsabilidad ética *y jurídica* de las decisiones automatizadas recae en nosotros como usuarios de la tecnología.

Finalmente, el permitir sesgos afecta nuestra **eficacia e impacto**. Las decisiones injustas llevan a asignar mal los escasos recursos: podríamos enfocarnos en poblaciones menos necesitadas dejando de lado a quienes intentamos empoderar, perpetuando la desigualdad que queríamos combatir. En suma, un sesgo algorítmico no corregido es contrario al *ADN* de las ONG (que buscan justicia social) y puede hacer que fallemos en nuestros objetivos fundamentales.

Minuto 10	**Acción del Facilitador**
	Invita al grupo a reflexionar.

Guión/instrucciones

¿Conocemos ejemplos cercanos (en nuestra región o sector) donde una herramienta digital haya discriminado sin querer? ¿Cómo reaccionaría nuestra organización si sucediera algo así? ¡Piensen un par de minutos!

Minuto 11	**Acción del Facilitador**

Guión/instrucciones

A veces creemos que los dilemas éticos son abstractos o técnicos, pero en el trabajo de las ONG, toman forma concreta todos los días. ¿Quién prioriza? ¿Cómo se decide quién recibe ayuda? ¿Estamos usando herramientas que realmente entienden nuestro contexto o simplemente replicamos desigualdades disfrazadas de eficiencia?

Con esta actividad buscamos activar esa reflexión. No hay respuestas correctas o incorrectas, pero sí hay decisiones que debemos tomar con conciencia, en equipo, y con criterios claros.

Por eso, mientras responden con emojis, piensen: ¿a quién podría afectar esta decisión?

¿qué voz falta en este proceso? ¿cómo podría hacerlo más justo?

Vamos a revisar juntos cada caso y debatir brevemente. El objetivo no es juzgar, sino aprender cómo mejorar nuestras prácticas con IA desde una mirada ética, humana y crítica.

Minuto 12	**Acción del Facilitador**
	Enuncia la pregunta y la situación y llama a que reacciones con los emojis en el chat.

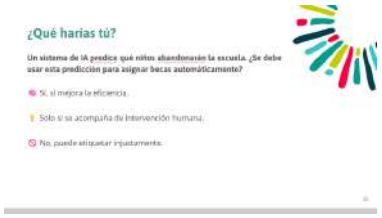
Guión/instrucciones
Veamos sus respuestas.

Minuto 13	**Acción del Facilitador**
	Enuncia la pregunta y la situación y llama a que reacciones con los emojis en el chat.

Guión/instrucciones
Veamos sus respuestas.

Minuto 14	**Acción del Facilitador**
	Enuncia la pregunta y la situación y llama a que reacciones con los emojis en el chat.

Guión/instrucciones
Veamos sus respuestas.

Minuto 15	**Acción del Facilitador**
	Enuncia la pregunta y la situación y llama a que reacciones con los emojis en el chat.

Guión/instrucciones
Veamos sus respuestas.

Minuto 16	**Acción del Facilitador**
	Lee el título y continua

Guión/instrucciones
Ahora, detectemos sesgos ocultos en las decisiones de la IA

Minuto 17	**Acción del Facilitador**
	Presenta la actividad, divide los grupos aleatoriamente y asigna 10 minutos para la discusión. Una vez regresa hace el informe de la actividad.

Guión/instrucciones

Vamos a hacer una pequeña dinámica para aplicar lo que acabamos de ver. Les presento un escenario basado en hechos reales: *Una organización no lucrativa decide agilizar su proceso de contratación usando una herramienta de IA que rankea candidatos según su CV. Tras un tiempo, observan un patrón preocupante: la mayoría de CVs que la IA recomienda pertenecen a hombres, y muchas candidatas calificadas quedan fuera de las ternas.* Esto choca con el compromiso de la ONG con la equidad de género.

Dividamos la sala en pequeños grupos. Cada grupo discutirá durante unos 5 minutos las preguntas en pantalla: **¿Qué tipo de sesgo creen que está presente en este sistema? ¿Cómo podrían detectarlo o confirmarlo? Y lo más importante: si ustedes fueran el equipo de la ONG, ¿qué harían para corregir o mitigar este sesgo?**

Piensen en las posibles causas: ¿Será que los datos de entrenamiento tenían mayormente currículums de hombres en posiciones previas? ¿Quizá la IA aprendió a valorar ciertas palabras o experiencias que son más comunes en hombres por factores culturales? – justo como pasó con el caso de Amazon. También reflexionen sobre las soluciones: por ejemplo, revisar el algoritmo, incorporar criterios de diversidad en la selección, o incluso quitar ciertas variables (como el género o proxies del género) para evitar ese sesgo.

Debrief luego de la actividad: Era un caso de sesgo de género en IA. Posibles soluciones que habrán surgido incluyen: entrenar el modelo con datos más balanceados, hacer pruebas ciegas donde se oculta el género al algoritmo, o establecer una supervisión humana que revise las recomendaciones de la IA antes de decidir. La idea central es que debemos auditar y ajustar nuestras herramientas de IA para asegurarnos de que no traicionen nuestros valores. Gracias por sus ideas; con esto en mente, pasemos a ver cómo los marcos éticos nos ayudan a prevenir situaciones así.

Minuto 18	Acción del Facilitador

Guión/instrucciones

Afortunadamente, no partimos de cero para responder a estos retos: en los últimos años han surgido **marcos éticos internacionales** y **regulaciones** que nos sirven de guía. Destacaremos algunos:

En 2021 la **UNESCO** (Organización de la ONU para la Educación, la Ciencia y la Cultura) logró que 193 Estados miembros –prácticamente todo el mundo– adoptaran un acuerdo histórico: la *Recomendación sobre la Ética de la Inteligencia Artificial*. Es el primer estándar global en la materia. ¿Qué dice este documento? Su *piedra angular* es **proteger la dignidad humana y los derechos humanos** en todo desarrollo de IA. Plantea principios como la **transparencia y equidad** en los algoritmos, la importancia de la **supervisión humana** (que no deleguemos decisiones totalmente a máquinas), la promoción de la diversidad y la inclusión, y garantizar que la IA no cause daño ni profundice desigualdades. Gabriela Ramos, subdirectora de UNESCO, ha dicho claramente que los gobiernos y organizaciones deben desarrollar la IA de modo que *no reproduzca los sesgos del mundo real que no nos gustan*, construyendo así sociedades más **inclusivas y libres de discriminación**. La Recomendación UNESCO es un **faro ético**: no es una ley obligatoria, pero orienta a países (y por ende a las instituciones dentro de ellos) sobre cómo establecer buenas prácticas y políticas en IA, desde la

gestión de datos hasta la educación y la investigación en este campo.

Por otro lado, la **OCDE (Organización para la Cooperación y el Desarrollo Económicos)** publicó en 2019 unos *Principios para la IA* que rápidamente obtuvieron respaldo internacional (G20, UE, EE.UU., etc.). Estos principios se centran en valores compartidos: promover un **crecimiento inclusivo** y el bienestar general mediante la IA, **respetar los derechos humanos, la democracia y la equidad**, asegurar la **transparencia** de los sistemas, garantizar la **robustez, seguridad y ausencia de sesgos** en los algoritmos, y asegurar la **rendición de cuentas** (accountability) de quienes desarrollan y usan IA. En 2024, la OCDE actualizó estos principios para reforzar temas como la **integridad de la información** (evitar desinformación generada por IA) y la **seguridad** con mecanismos de control sobre sistemas potencialmente dañinos. Para las ONG, estos valores de la OCDE ofrecen un lenguaje común para nuestras políticas: por ejemplo, nos alineamos con el principio de transparencia si explicamos a nuestros beneficiarios cómo una IA tomó cierta decisión, o con el principio de inclusión si nos aseguramos de que la tecnología beneficie a todos los grupos y no solo a unos pocos.

Además de los marcos éticos, existen ya **leyes y reglamentos** en marcha. El más conocido es el **GDPR** (Reglamento General de Protección de Datos de la Unión Europea), vigente desde 2018, que aunque se centra en privacidad, tiene un artículo (el 22) muy relevante para IA: **otorga a las personas el derecho a no ser sometidas a decisiones puramente automatizadas que tengan efectos significativos en ellas, sin intervención humana**. Es decir, en Europa (y de facto para cualquier entidad que trate datos de europeos, incluso ONG internacionales) no se podría usar un algoritmo para, digamos, aceptar o rechazar a alguien en un programa **sin** un mecanismo de revisión humana. Esta protección legal busca frenar los “desequilibrios de poder” entre individuos y algoritmos opacos, y esencialmente obliga a introducir **supervisión humana y transparencia**, algo clave para mitigar sesgos.

En la Unión Europea también se está discutiendo la **Ley Europea de IA (AI Act)**, que será la primera ley integral específica sobre inteligencia artificial a nivel mundial. Aún no está en vigor (se espera su aprobación final pronto), pero ya se sabe que clasificará los sistemas de IA por nivel de riesgo (alto, medio, bajo) y exigirá requisitos estrictos –como evaluaciones de conformidad, documentación, eliminación de sesgos– a aquellos usos considerados de *alto riesgo* (por ejemplo en salud, empleo, seguridad pública). Esto significa que proveedores de IA y también organizaciones usuarias tendrán obligaciones para garantizar la **ética y seguridad** de sus sistemas. Para ilustrar la relevancia: el Parlamento Europeo ya avanzó un borrador y se le conoce como la primera regulación comprensiva en la materia.

Fuera de Europa, otros países están siguiendo camino: por ejemplo, Estados Unidos publicó un **Blueprint for an AI Bill of Rights** con principios para desarrolladores (como prevenir discriminación algorítmica y asegurar explicabilidad), y varios países de Latinoamérica están adoptando o discutiendo estrategias nacionales de IA con componentes éticos. Incluso a nivel de Naciones Unidas, se habla de un posible

instrumento global en el futuro.

Minuto 19	Acción del Facilitador

Guión/instrucciones

En resumen, las ONG no están solas en esta conversación ética: podemos apoyar nuestras políticas internas en referentes sólidos como UNESCO y OCDE, y debemos estar al tanto de las leyes vigentes (p. ej. de protección de datos) que nos apliquen localmente. Estas directrices nos ayudan a traducir valores en acciones concretas al usar IA.

Minuto 20	Acción del Facilitador

Guión/instrucciones

Pasemos de la teoría a la práctica: ¿Cómo aterrizamos estos principios éticos en nuestras operaciones diarias y proyectos con IA? Les propongo algunas **acciones concretas** que una ONG puede tomar:

- **Evaluación de impacto y riesgos previos:** Antes de lanzar una herramienta de IA, hagamos una pausa y evaluemos: *¿Qué puede salir mal?* Por ejemplo, si vamos a usar un algoritmo para decidir la asignación de microcréditos en una comunidad, debemos analizar si podría, por su diseño, **excluir** a algún grupo (mujeres, minorías étnicas, mayores de cierta edad, etc.). Un buen hábito es realizar una **Evaluación de Impacto Ética** (inspirada en la propuesta de UNESCO) donde identifiquemos riesgos de sesgo, invasión de privacidad u otros daños, y planifiquemos cómo mitigarlos. Existen listas de verificación o incluso metodologías formalizadas para esto – la idea es **no lanzarnos a usar IA a ciegas**, sino anticipar las consecuencias sociales y éticas.

- **Uso de datos y diseño inclusivo:** Mucho del sesgo viene de los datos, así que ahí tenemos que actuar primero. Asegurémonos de que los datos con que alimentamos la IA **representen la diversidad** de las personas a las que servimos. Si vamos a crear un modelo para una iniciativa regional, por ejemplo, incorporamos datos de todas las zonas, no solo de la capital. Y si notamos que nos faltan datos de cierto grupo (pongamos que tenemos menos casos de mujeres rurales en el historial), seamos conscientes de ese vacío. A veces, la solución es complementar con datos externos o incluso **ajustar manualmente** ciertos parámetros para corregir el sesgo. Además, involucremos a partes interesadas en el diseño: ¿por qué no invitar a miembros de la comunidad o a expertos en derechos humanos a probar el sistema piloto y dar feedback? Esa retroalimentación diversa puede revelar sesgos que los técnicos pasamos por alto.

- **Transparencia y explicación:** Para ganar la confianza de nuestros beneficiarios y staff, debemos ser transparentes sobre el uso de IA. Esto implica dos cosas: primero, **divulgar** que se usó una IA en la decisión (por ejemplo, “las puntuaciones fueron generadas con un modelo predictivo basado en X criterios”), y segundo, **explicar en términos sencillos** los motivos de la decisión. Por muy complejo que sea el algoritmo, la persona afectada merece saber las razones básicas: “Te ofrecemos esta ayuda porque el sistema identificó que cumples con ciertos criterios prioritarios (A, B, C)”. La explicabilidad también es interna: el equipo técnico debe ser capaz de interpretar qué factores está usando la IA para decidir. Si es una “caja negra” incomprensible, es difícil confiar en ella o detectar sesgos. La transparencia crea un ambiente de rendición de cuentas y permite que cualquiera pueda cuestionar y apelar decisiones si algo parece incorrecto.

Minuto 21	Acción del Facilitador

Guión/instrucciones

- **Supervisión humana continua:** Un principio fundamental es **no ceder el control absoluto**. Ninguna decisión de alto impacto en nuestra ONG debería tomarse 100% por una máquina sin alguien que la valide. Por ejemplo, si un algoritmo

sugiere excluir a una familia de un programa por considerar que “no califican”, un trabajador humano debe revisar ese caso antes de la decisión final, pudiendo revertirla si ve algo injusto. La IA debería ser *como un asistente*, no el director. Además, hay que **monitorear** su desempeño en el tiempo: establecer indicadores de posibles sesgos (¿estamos llegando equilibradamente a distintos grupos demográficos? ¿ha habido quejas de decisiones injustas?) y auditar periódicamente los resultados. Si detectamos desviaciones, ajustar o re-entrenar el sistema. La ética de la IA no es “configurar y olvidar”; requiere vigilancia permanente, igual que cualquier programa de la ONG.

- **Capacitación y cultura ética:** Finalmente, nada de esto funciona si nuestro equipo no está sensibilizado. Es clave capacitar a nuestro personal – tanto a los técnicos como a quienes implementan los programas – en conceptos de IA básica, en entender qué es un sesgo algorítmico, qué dice la legislación (por ejemplo, GDPR sobre protección de datos personales), cómo manejar datos de manera responsable, etc. Pero más allá del conocimiento, hay que cultivar una **cultura organizacional** donde se tomen en serio los dilemas éticos. Animar a que, si alguien en el equipo nota un posible sesgo o un uso inapropiado de datos, lo diga sin temor. Tal vez eso signifique crear un pequeño comité de ética digital o incluir este tema en las reuniones de proyecto. Lo importante es que la ética de la IA se vuelva una conversación continua en la ONG, no un check de una sola vez.

Aplicando estas prácticas, las ONG pueden **aprovechar la IA minimizando sus riesgos**. Como menciona un reporte de Nonprofit Quarterly, tener políticas y “guardrails” claros no solo evita daños, sino que permite que usemos la IA **de forma ética y consciente para potenciar nuestra misión social** en lugar de socavarla.

Minuto 22	**Acción del Facilitador**
	Lee el título y continua

Guión/instrucciones
Principios rectores para la adopción ética de la IA en las ONG

Minuto 23	Acción del Facilitador
	Explica uno a uno los principios propuestos.

Guión/instrucciones

Proponemos ahora una lista de **Principios Rectores** que podrían conformar el "código de ética" para el uso de IA en una ONG. Estos principios actúan como columnas vertebrales para cualquier política o decisión tecnológica que tomemos:

1. **Equidad e Inclusión:** Toda solución de IA que adoptemos debe ser examinada bajo la lupa de la equidad. ¿Está accesible para distintos grupos? ¿Sus resultados son justos para mujeres, hombres, minorías étnicas, personas con discapacidad, diferentes edades, zonas geográficas? Nos comprometemos a identificar y eliminar cualquier sesgo que lleve a tratos desiguales. La UNESCO en su recomendación destaca este punto: la IA debe promover la justicia social y la **no discriminación**, asegurando que sus beneficios alcancen a todos. Por ejemplo, si desarrollamos un bot de atención, verificaremos que entienda e incluya la diversidad cultural y lingüística de nuestros usuarios, evitando respuestas estereotipadas.

2. **Transparencia y Explicabilidad:** Como principio, no usaremos "cajas negras" de forma irresponsable. Seremos transparentes sobre cuándo y cómo utilizamos IA. Si un donante recibe un correo generado por IA o un beneficiario es evaluado por un algoritmo, lo comunicaremos abiertamente. Y si alguien pregunta "¿por qué este algoritmo decidió así?", podremos explicarlo en lenguaje sencillo. La transparencia genera confianza y permite a terceros escrutar y señalar posibles fallos. Imaginemos que un beneficiario cuestiona por qué no fue seleccionado en un programa; bajo este principio, debemos poder revisar la decisión, explicarla ("priorizamos casos con tales características porque...") y corregir si hallamos un error. Nada de "el computador lo dijo y es infalible".

3. **Responsabilidad y Rendición de Cuentas:** Aunque usemos IA de terceros o automatizaciones, *la responsabilidad final es nuestra*. Este principio establece que habrá **rendición de cuentas interna** por el funcionamiento ético de la IA. ¿Cómo se ve esto? Por un lado, asignando claramente roles: quién supervisa el algoritmo, a quién reportar incidencias, quién decide ajustes. Por otro, teniendo *auditabilidad*: registro de cómo se toman las decisiones automatizadas para poder reconstruir y entender un fallo. Si, por ejemplo, una familia fue excluida por error de una ayuda, debemos investigar qué pasó en el proceso algorítmico y enmendarlo. También implica crear canales para que usuarios o empleados puedan reportar sesgos o resultados problemáticos sin burocracia. En consonancia con los marcos

internacionales, fomentaremos **algoritmos auditables, trazables y sujetos a debido control**. La ONG responderá por sus herramientas así como responderá por las acciones de su personal.

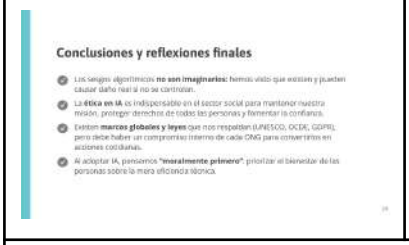
4. **Privacidad y Seguridad de datos:** Este principio refuerza nuestro deber de proteger los datos personales al máximo nivel. Sólo utilizaremos datos de personas si tenemos autorización clara y un propósito legítimo. Si empleamos IA que requiere datos sensibles (salud, situación socioeconómica), nos aseguraremos de cumplir normativas como la GDPR y las locales, minimizando la recolección al mínimo necesario. Además, implementaremos medidas de **ciberseguridad** para prevenir brechas o accesos no autorizados a los datos o modelos. La ética también es no exponer a quienes confiaron sus datos a nosotros. Por ejemplo, si alimentamos un algoritmo con información de nuestros beneficiarios, jamás lo haremos en plataformas públicas o sin revisar términos y condiciones (evitando que esos datos terminen mal usados por terceros). La confidencialidad y la integridad de la información son sagradas.
5. **Supervisión Humana:** Esto es clave: **ninguna IA operará en piloto automático absoluto**, especialmente en decisiones que afectan a personas. Siempre habrá un ser humano en el circuito de decisión para validar o tener la palabra final. Este principio evita muchos riesgos. Supongamos que un sistema recomienda rechazar a alguien de un programa; con supervisión humana, ese alguien tiene implícitamente derecho a una segunda revisión o apelación humana. Además, mantenemos así el control: la IA nos asiste pero *no sustituye nuestra responsabilidad humana*. Como políticas concretas, podríamos decir: “si un algoritmo marca un caso como fraudulento, un auditor humano lo revisará antes de tomar acción” o “las respuestas del chatbot de la ONG serán monitoreadas regularmente por un supervisor humano para corregir cualquier respuesta inadecuada”. Esto también se alinea con legislación como el artículo 22 de GDPR que mencionamos: en nuestras ONG, cumpliremos dando siempre opción de intervención humana.

6. **Proporcionalidad y No Hacer Daño:** Inspirado en la bioética y recogido por UNESCO, este principio nos invita a preguntarnos: *¿Necesitamos realmente usar IA para esto?* y si la usamos, ¿hemos calibrado bien su alcance para no causar daño accidentalmente? La **proporcionalidad** significa que la complejidad o intrusividad de la IA debe ser acorde al problema que resuelve. Por ejemplo, no justificaríamos implementar reconocimiento facial (con todos sus riesgos de privacidad y sesgo racial) solo para pasar lista en una capacitación, ¿cierto? Pero quizá sí para un proyecto crítico de reunificación familiar donde no hay otra forma de identificar personas. “No hacer daño” es un recordatorio de nuestro mandato humanitario: si la IA presenta una posibilidad alta de perjudicar a nuestros beneficiarios (p. ej., etiquetándolos con puntajes que puedan estigmatizarlos), preferiremos no usarla. Siempre ponderaremos beneficios vs. posibles daños antes de adoptar la tecnología.

7. **Capacitación y Mejora Continua:** Finalmente, un principio hacia adentro: nos

comprometemos a **aprender y mejorar continuamente** en este tema. La ética de IA evoluciona, surgen nuevos desafíos (piensen en IA generativa, los deepfakes, etc.), y las regulaciones también se actualizan. Nuestras directrices no pueden ser estáticas. Entonces, incluiremos acciones de capacitación periódica al personal en datos/IA, actualizaremos nuestras políticas tecnológicas regularmente y compartiremos lecciones aprendidas con la comunidad. También es válido sumarse a redes o iniciativas de "AI for Good" donde las ONG comparten prácticas éticas. La humildad para ajustar el curso es clave: puede que implementemos un sistema con las mejores intenciones y luego descubramos un sesgo; debemos estar listos para reconocerlo, informar a las partes interesadas y corregirlo.

Estos principios rectores sirven de brújula. Cada vez que consideremos un nuevo proyecto de IA o evaluemos uno en marcha, podemos contrastarlo con esta lista: ¿estamos siendo justos? ¿transparentes? ¿Quién se hace responsable? etc. Si alguna respuesta nos incomoda, es señal de replantear la herramienta o las condiciones de su uso. La adopción ética de la IA no es un estado final, sino **un proceso continuo de toma de decisiones conscientes** alineadas con nuestros valores como organización social.

Minuto 24	Acción del Facilitador
	

Guión/instrucciones

Para cerrar este módulo, recapitulemos los puntos clave que espero se lleven consigo:

- **Reconocer el problema:** El sesgo algorítmico es real y documentado. No es un temor teórico ni algo que “solo le pasa a otros”. Como vimos, desde la salud hasta la seguridad pública, pasando por recursos humanos, los algoritmos pueden agravar desigualdades pre-existentes si no somos cuidadosos. Como ONG, debemos partir de la premisa de que *sí, los algoritmos pueden equivocarse e incluso ser injustos*, y por eso requieren nuestra vigilancia ética constante. No podemos delegar ciegamente decisiones en la IA.

- **Ética = Esencial en nuestras organizaciones:** Hemos discutido cómo nuestros valores fundamentales (igualdad, justicia, inclusión, no discriminación) deben reflejarse en cualquier uso de tecnología avanzada. La ética de la IA no es un “adorno” o un cumplido; es tan importante como nuestros códigos de conducta o

nuestras salvaguardas de protección al beneficiario. Si descuidamos la ética, arriesgamos causar daños a quienes queremos ayudar y perder la confianza que tanto nos ha costado ganar en las comunidades y aliados. Por el contrario, si incorporamos la ética desde el diseño, la IA puede ser una aliada poderosa para el bien social.

- **Apoyarse en guías y normas existentes:** No estamos solos en este camino. Contamos con orientaciones valiosas a nivel internacional – por ejemplo, podemos alinear nuestras políticas con los **Principios de la OCDE**, que hablan de transparencia, equidad, rendición de cuentas, etc., o con la **Recomendación UNESCO**, que, recordemos, enfatiza la primacía de los derechos humanos y la necesidad de no repetir con IA los vicios de nuestra sociedad. Asimismo, la legislación vigente como GDPR nos da un marco legal que seguir, especialmente en protección de datos y decisiones automatizadas. Estar al día con estas referencias nos da legitimidad y solidez en nuestras prácticas.

- **Compromiso interno y diario:** Nada cambia si estos temas quedan solo en el papel. La adopción ética de IA requiere un compromiso de *toda la organización*, desde la alta dirección –que debe promover y dotar de recursos para implementar estas políticas– hasta cada técnico o usuario que interactúa con un sistema inteligente. Es un esfuerzo transversal: por ejemplo, el área de Recursos Humanos evaluando la equidad de una herramienta de selección, el área de programas verificando que un sistema de focalización no esté sesgado, el área de comunicaciones siendo transparente sobre cómo usamos chatbots, etc. En otras palabras, debe haber una cultura de "**pensar primero en lo humano y ético, luego en lo digital**" en cada decisión que tomemos con IA.

- **El lado positivo:** Quiero dejar muy en claro que este módulo no busca asustarlos y que renuncien a la IA. Al contrario, bien usada, la IA puede **potenciar enormemente el impacto social**. Puede ayudarnos a analizar datos de necesidades más rápido, a personalizar intervenciones, a liberar tiempo administrativo para enfocarlo en trabajo comunitario, y hasta a innovar soluciones a problemas complejos (vimos ejemplos de ONG que usan IA para monitorear el cambio climático, mejorar agricultura, etc.). La clave está en "tomar los recaudos éticos". Como señalaba un artículo reciente, si primero nos ocupamos de abordar los posibles riesgos éticos y legales, podremos usar la IA para *maximizar los beneficios sociales*, aportando un contrapeso positivo frente al uso de la IA en otros sectores orientados solo a lucro.

Minuto 25	Acción del Facilitador

Guión/instrucciones

En síntesis, **ética y sesgo en la IA** no es un tema pasajero; llegó para quedarse. Pero con conocimiento, principios claros y acción proactiva, nuestras ONG pueden liderar con el ejemplo en la adopción responsable de la IA. Así demostraremos que la tecnología más avanzada **sí puede** estar al servicio de la humanidad y la justicia, si se implementa con conciencia. Los invito a que lleven estas conversaciones a sus equipos, revisen sus proyectos a la luz de lo aprendido y, sobre todo, sigan curiosos y críticos frente a la IA. Solo así podremos navegar juntos este nuevo terreno, aprovechando sus ventajas sin perder nuestro norte ético.

¡Gracias a todos por su participación activa! Quedo atento/a a preguntas o comentarios finales que quieran compartir.