



IA y human-in-the-loop en el sector humanitario

¿Te resultan confusos los términos de IA y no sabes cómo aplicarlos en el trabajo humanitario? Esta guía práctica, informada con la experiencia de Data Friendly Space, ofrece una introducción accesible al human-in-the-loop.

Esta guía ha sido creada como continuación de la iniciativa de investigación conjunta 2025-26 realizada por Humanitarian Leadership Academy y Data Friendly Space: ¿Cómo utilizan la inteligencia artificial los trabajadores humanitarios en 2025? Mapeo de la práctica actual y el potencial futuro.

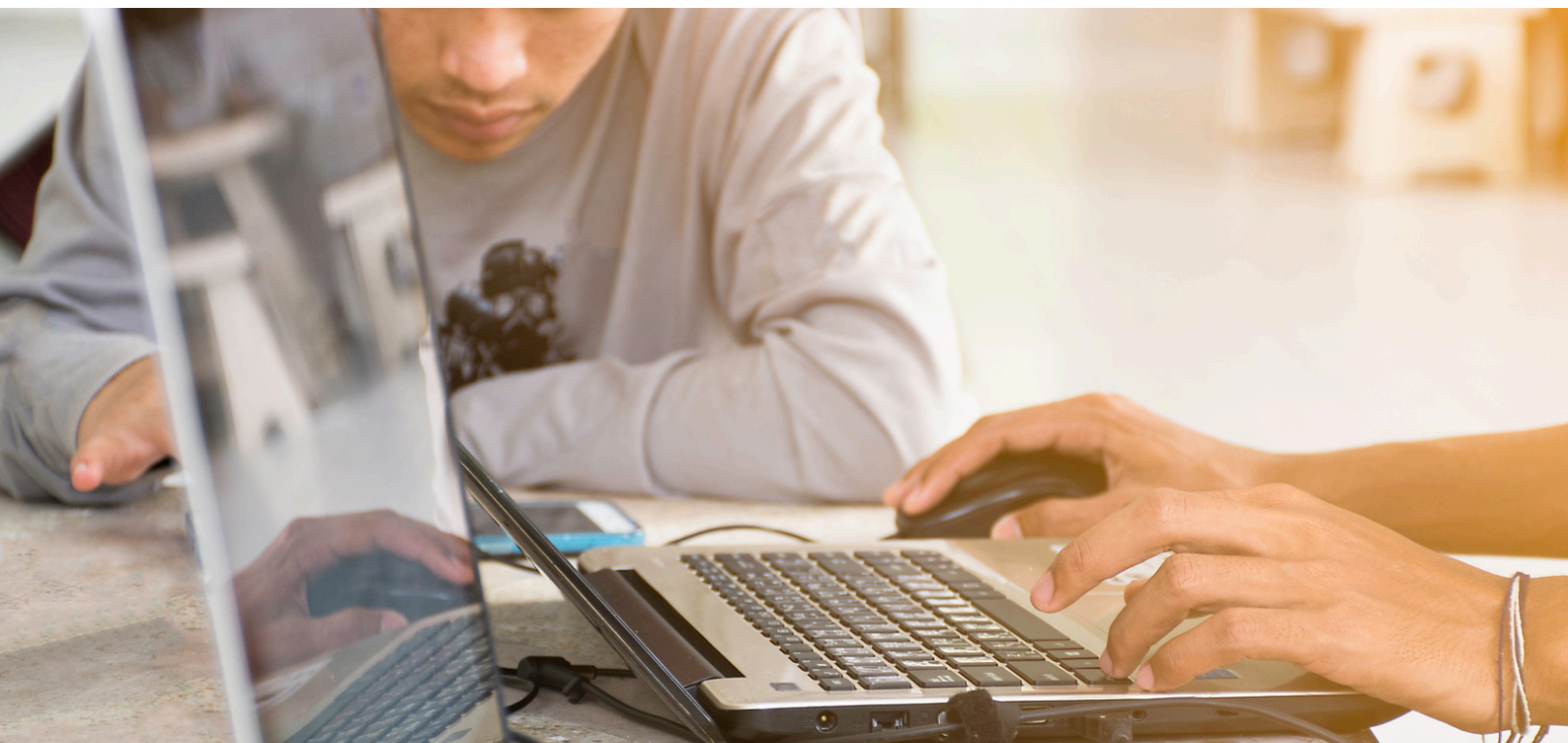
La IA se está utilizando en todo el trabajo humanitario, pero no siempre de forma segura

En 2026, el 75% de los profesionales humanitarios utilizan herramientas de IA regularmente en su trabajo. Sin embargo, menos de una cuarta parte de las organizaciones cuenta con una política formal de IA.

Cuando la IA opera sin intervención humana, puede:

- Generar información plausible pero incorrecta
- Pasar por alto contexto crítico sobre el terreno
- Reflejar y amplificar sesgos existentes en los datos

En contextos humanitarios, actuar a partir de resultados de IA defectuosos no solo desperdicia recursos: puede desviar la ayuda de las personas que más la necesitan. Por eso, todo resultado de IA debe combinarse con supervisión humana.



¿Qué significa human-in-the-loop (HITL)?

Human-in-the-Loop (HITL) es un enfoque en el que las personas guían, validan, mejoran y/o anulan activamente los resultados de la IA antes de que se tomen decisiones.

Este enfoque combina las fortalezas de los seres humanos y las máquinas, garantizando que las tareas complejas o de alto riesgo se beneficien del juicio y la supervisión humanos.

HITL no se trata de ralentizar la IA, sino de garantizar que la IA trabaje para las personas, no solo sobre las personas.

HITL no es un único punto de control. Es un ciclo continuo integrado en el uso de la IA, que garantiza que las personas adecuadas sigan siendo responsables ante su organización y ante las personas a las que sirven.



¿Por qué es importante el human-in-the-loop en contextos humanitarios?

El principio de no causar daño también se aplica a la IA.

La IA puede pasar por alto el contexto. Puede reflejar sesgos en los datos de entrenamiento. La IA no puede comprender plenamente la experiencia vivida de las personas afectadas por crisis.

Si un algoritmo marca a la familia equivocada para su exclusión o clasifica erróneamente un riesgo de protección, el daño es real.

El daño no es solo físico. Una IA mal gestionada puede exponer datos sensibles, reidentificar a personas o facilitar la vigilancia masiva. La supervisión protege a las personas en todos los niveles.

La revisión humana es la forma en la que detectamos estos posibles riesgos antes de que afecten a las personas a las que intentamos ayudar.



El human-in-the-loop no debe ser un marco abstracto

El human-in-the-loop debe implementarse en todos los niveles del uso de la IA. ¿Cómo se ve esto en la práctica?

- **Define los puntos de decisión con antelación:** Antes de implementar cualquier herramienta de IA, identifica exactamente dónde debe aplicarse la revisión humana. No lo dejes al azar ni supongas que alguien más lo está verificando.
- **Asigna el revisor adecuado a cada decisión:** La persona que revisa un resultado de IA debe tener el conocimiento y la autoridad para actuar en consecuencia. Un oficial de WASH, un especialista en protección y un analista de datos detectarán cosas diferentes.
- **Facilita la posibilidad de anular:** Si anular una recomendación de IA parece arriesgado, burocrático o desalentador, las personas no lo harán. Construye culturas y sistemas donde cuestionar los resultados de la IA sea la norma.
- **Documenta cada anulación:** Cuando una persona modifica o rechaza un resultado de IA, regístralo. Estas decisiones son valiosas. Revelan dónde la IA tiene bajo rendimiento, mejoran los modelos futuros y ayudan a otros a saber qué tener en cuenta.

Los seres humanos aportan habilidades especializadas que la IA no puede replicar

Tu experiencia no es un complemento de la IA: es lo que hace que la IA sea segura de usar.

- **Conocimiento local:** Los datos significan cosas diferentes en diferentes lugares. Un número que parece normal en un contexto puede indicar una crisis en otro. Solo las personas que conocen un lugar pueden hacer ese juicio.
- **Juicio ético:** Decidir qué publicar, qué proteger y a quién podría afectar es un acto de responsabilidad. Eso no puede delegarse a un algoritmo.
- **Sensibilidad cultural:** La IA no puede reconocer de forma fiable enfoques inapropiados, suposiciones dañinas o puntos ciegos específicos de cada cultura. Los revisores humanos detectan lo que los modelos pasan por alto.
- **Interpretar los silencios:** El hallazgo más importante es a veces lo que los datos no dicen. Quién no está contado, qué está ausente, qué significa el silencio. La IA no tiene forma de saber lo que no sabe.

¿Cómo podemos mantener a las personas verdaderamente en control?

El human-in-the-loop es tan sólido como la cultura y los procesos que lo respaldan.

- **Incluye a las comunidades afectadas:** El human-in-the-loop no debe referirse únicamente al personal. Siempre que sea posible, crea canales para que las personas afectadas por crisis puedan señalar errores, hacer preguntas o impugnar las decisiones que se tomen sobre ellas.
- **Presta atención al sesgo de automatización:** Las personas tienden a confiar en los resultados de la IA, especialmente bajo presión. Capacita a tus equipos para que cuestionen activamente los resultados, no solo los aprueben sin más.
- **Revisa la supervisión a medida que las herramientas evolucionan:** Los sistemas de IA cambian. El proceso de supervisión que diseñaste hace seis meses puede que ya no sea suficiente. Establece puntos de control regulares para reevaluar si la revisión humana sigue siendo efectiva.
- **Asigna responsabilidades claras:** Cada decisión informada por IA debe tener una persona responsable del resultado. Si nadie se hace cargo, nadie lo detectará cuando algo salga mal.
- **Actúa cuando las cosas salen mal:** Si un resultado de IA causa daño o se comete un error grave, detente, infórmalo y asegúrate de que las personas adecuadas estén informadas. Rendir cuentas significa actuar ante lo que salió mal.

No todos los resultados de la IA implican el mismo riesgo: ante la duda, verifica

Saber cuándo examinar con más detenimiento es una habilidad fundamental del HITL. El tiempo que lleva verificar siempre será menor que el coste de actuar a partir de un hallazgo defectuoso. Esto puede incluir situaciones en las que:

- **El resultado involucra a poblaciones afectadas:** Los hallazgos sobre comunidades desplazadas, sobrevivientes de violencia o grupos marginados requieren una interpretación sensible a la protección que la IA no puede proporcionar.
- **El contexto es cambiante o políticamente sensible:** En crisis activas, la IA puede quedarse atrás frente a las realidades sobre el terreno. La verificación humana ancla los resultados en lo que realmente está ocurriendo sobre el terreno.
- **El hallazgo parece sorprendente o sospechosamente ordenado:** Tanto las anomalías como los resultados demasiado limpios pueden indicar errores o lagunas en la cobertura de fuentes. Confía en tu instinto para profundizar.
- **El resultado informará una decisión importante:** La asignación de recursos, la respuesta a crisis y las estrategias de intervención conllevan riesgos elevados. Estos resultados exigen validación experta antes de su uso.

Preguntas que toda persona del sector humanitario debería hacerse

Antes de actuar a partir de un resultado de IA: una guía de verificación HITL.

- ¿Este resultado fue producido dentro de un marco definido por humanos?
- ¿Ha sido verificado con fuentes de confianza?
- ¿Refleja el contexto local, incluyendo lo que podría estar ausente?
- ¿Podría este resultado o decisión causar daño a las personas a las que pretende ayudar?
- ¿Es seguro y apropiado compartirlo con quienes toman decisiones?
- ¿Quién entrenó este modelo y con qué datos?
- ¿Refleja este resultado una necesidad genuina o una laguna en nuestros datos?
- ¿Puede una persona afectada cuestionar o impugnar esta decisión?
- ¿Compromete esta herramienta de IA nuestra neutralidad o independencia?
- ¿Existe una persona responsable del resultado?
- ¿Qué ocurre cuando la IA se equivoca?

Nuestros principios no se detienen cuando usamos herramientas de IA

Los valores que guían la acción humanitaria — no causar daño, imparcialidad, neutralidad, independencia — no dejan de aplicarse cuando usamos IA.

El marco human-in-the-loop es la forma en que llevamos esos principios a la era de la IA.

Recursos

La HLA, en colaboración con Data Friendly Space, lidera un estudio global sobre cómo las personas del sector humanitario utilizan la IA. Accede a los productos de investigación, artículos, seminarios, y podcasts en nuestro [Centro de Recursos](#).



Aprende gratis en línea en kayaconnect.org. Obtén insignias digitales HPass y compártelas en redes sociales para para mostrar tu aprendizaje en IA humanitaria.

