ACADEMIA HUMANITARIA · HUMANITAIRES JOBS

Mini MBA IA y acción humanitaria: trabajar con Claude y Gemini

Material del curso · Semana 3 : Proteger los datos: lo que nunca se da a una IA y cómo anonimizar

Un programa de exactamente 3 meses (12 semanas, 144 horas) para utilizar los asistentes de IA Claude (Anthropic) y Gemini (Google) en el trabajo humanitario, de forma útil, segura y responsable: funcionamiento y límites, redacción de peticiones, protección de datos, ética y principios humanitarios, informes, propuestas, análisis de datos y MEAL, traducción y comunicación con las comunidades, investigación y verificación de fuentes, desinformación y seguridad, productividad y automatización, política de uso y despliegue en una organización. Ejercicios corregidos, tres exámenes mensuales con estudios de caso y un diploma verificable de la Academia Humanitaria. Programa independiente, sin asociación con Anthropic ni con Google.

https://humanitairejobs.org/es/academia/mini-mba-ia-claude-gemini/semaine-3/

Documento gratuito de la Academia Humanitaria. Puede imprimirse y compartirse libremente con fines de formación, sin modificaciones.

Semana 3 : Proteger los datos: lo que nunca se da a una IA y cómo anonimizar

Objetivos

- Aplicar los principios de protección de datos en la acción humanitaria al uso de los asistentes de IA.
- Clasificar una información (pública, interna, confidencial, sensible) y decidir si puede darse a un asistente, y con qué plan.
- Anonimizar o seudonimizar un documento antes de utilizarlo con la IA, y configurar los ajustes de privacidad.

Por qué es la semana más importante

Las organizaciones humanitarias poseen algunos de los datos más sensibles que existen: identidad de personas refugiadas o desplazadas, sobrevivientes de violencia, niños separados, datos de salud, posiciones de equipos, incidentes de seguridad, información sobre grupos armados. Una filtración puede exponer a personas a represalias, discriminación o violencia.

Cuando pega un texto en un asistente de IA, ese texto **sale de su ordenador** y se procesa en los servidores del proveedor, a menudo en otro país. Según el plan y los ajustes, puede ser **conservado**, a veces **leído por personas** encargadas de la calidad, o utilizado para **mejorar los modelos**.

> Principio: no dé nunca a un asistente de IA un dato que no tendría derecho a enviar por correo a un proveedor externo sin contrato de tratamiento de datos.

Los principios de referencia

El **Manual sobre protección de datos en la acción humanitaria** del CICR y del Brussels Privacy Hub (2.ª edición, 2020, que también trata de las nuevas tecnologías), las políticas de protección de datos de las agencias de la ONU y, para las organizaciones concernidas, el Reglamento General de Protección de Datos (RGPD) de la Unión Europea establecen principios comunes:

- **licitud y finalidad**: un dato se recoge para un fin preciso y no se reutiliza para otro fin incompatible;
- **minimización**: utilizar solo los datos estrictamente necesarios;
- **seguridad**: proteger los datos contra el acceso no autorizado;
- **transferencias**: regular toda transferencia a un tercero o a otro país;
- **derechos de las personas**: información, acceso, rectificación.

Dar los datos de una lista de beneficiarios a un asistente de IA de uso general es una **transferencia a un tercero**, para una **finalidad** que no se anunció a las personas. En la gran

mayoría de los casos, es contrario a estos principios.

En el sector se habla también de **datos demográficamente sensibles**: incluso sin nombres, datos agregados (por ejemplo la localización precisa de un grupo étnico o religioso minoritario) pueden exponer a una comunidad entera.

Lo que nunca se da a un asistente de uso general

- Nombres, números de teléfono, números de tarjeta o de registro, fotos, direcciones, coordenadas GPS de beneficiarios o de miembros del personal.
- Toda información sobre sobrevivientes de violencia de género, niños, casos de protección, salud mental o VIH, incluso «anonimizada» de forma aproximada.
- Informes detallados de incidentes de seguridad, itinerarios, posiciones y horarios de desplazamiento, planes de evacuación, información sobre negociaciones de acceso.
- Denuncias de explotación y abuso sexuales, de fraude o de corrupción en curso de investigación.
- Datos de salud individuales.
- Contraseñas, claves de acceso, datos bancarios.
- Documentos marcados como confidenciales por la organización, un donante o un socio.

Para estas categorías, ni siquiera un plan profesional con compromisos contractuales basta siempre: se necesita una decisión documentada de la organización (evaluación de impacto en la protección de datos) antes de cualquier uso.

Planes de uso general y planes profesionales

En noviembre de 2026, las grandes líneas son las siguientes (compruebe siempre las condiciones vigentes):

- **Planes de uso general (gratuitos o suscripción individual)**: ambos proveedores ofrecen ajustes para controlar la conservación del historial y el uso de las conversaciones para mejorar los modelos. En Google, la actividad de las aplicaciones de Gemini puede conservarse y algunas conversaciones pueden ser revisadas por personas; un ajuste permite desactivarla. En Anthropic, el usuario elige en los ajustes de privacidad si sus conversaciones pueden servir para el entrenamiento, y Claude ofrece un modo de conversación de incógnito.
- **Planes para organizaciones** (Claude Team y Enterprise; Gemini en Google Workspace para empresas y educación): los proveedores se comprometen contractualmente a no utilizar por defecto los datos de los clientes para entrenar sus modelos, con administración central, controles de acceso y condiciones de tratamiento de datos.

En concreto:

- una cuenta personal gratuita solo debe servir para información **pública o sin importancia**;
- la información **interna** (no sensible) corresponde a un plan profesional validado por la organización;

- los datos **personales y sensibles** quedan excluidos, salvo decisión formal y regulada.

> Desactivar el historial o el uso para el entrenamiento reduce los riesgos, pero no convierte una cuenta personal en una herramienta autorizada para datos de beneficiarios.

Clasificar y luego anonimizar

Antes de cada uso, clasifique la información según la política de la organización, por ejemplo:

- **Pública**: informes publicados, convocatorias de propuestas, textos legales. Utilizable con cualquier asistente.
- **Interna**: borradores de propuestas, procedimientos, actas sin datos personales. Plan profesional autorizado.
- **Confidencial**: presupuestos detallados, evaluaciones de socios, negociaciones. Plan profesional y acuerdo del responsable, tras retirar los elementos innecesarios.
- **Sensible**: datos personales, protección, seguridad, investigaciones. Excluida.

Anonimizar significa hacer imposible la identificación de una persona, incluso por cruce de datos. **Seudonimizar** significa sustituir los identificadores por códigos («Beneficiario A»); la persona sigue siendo identificable por quien tiene la tabla de correspondencia: siguen siendo datos personales.

Trampas frecuentes: una aldea de 300 habitantes más «mujer de 34 años, viuda, cuatro hijos» basta para identificar a alguien; una cita de un grupo focal puede contener un nombre o un apodo; los metadatos de un archivo (autor, comentarios, revisiones) contienen nombres. Elimine los detalles innecesarios, agrupe las edades y los lugares, y suprima los comentarios y el historial de los documentos.

Los buenos reflejos del día a día

- Utilice la **cuenta proporcionada por la organización** para el trabajo, nunca una cuenta personal para documentos internos.
- Verifique los **ajustes de privacidad** en la primera conexión, y luego cada trimestre: historial, uso para el entrenamiento, memoria, conectores activados.
- Limite los **conectores** (acceso al correo, a Drive) a lo estrictamente necesario: un asistente que puede leer todo su buzón también puede leer en él datos sensibles.
- No comparta **enlaces públicos** a conversaciones que contengan información interna.
- En el terreno, cuidado con las **pantallas visibles** y los dispositivos compartidos; cierre la sesión.
- En caso de error (datos sensibles pegados por descuido): elimine la conversación y luego **notifique el incidente** al referente de protección de datos sin esperar. Un error notificado rápido limita los daños; un error ocultado los agrava.

Ejercicio: ¿se puede utilizar la IA para este documento?

Situación. Un equipo de protección quiere ahorrar tiempo en cinco tareas: A) resumir el informe público de un clúster de protección; B) traducir al inglés el acta de una reunión interna sin nombres; C) analizar una hoja de cálculo de 2 300 hogares con nombres, números de teléfono y vulnerabilidades para detectar duplicados; D) reformular el relato de una sobreviviente de violencia para un informe de caso; E) proponer un plan de formación del personal en PEAS. La organización dispone de un plan profesional validado por el responsable de protección de datos, con compromiso de no utilización para el entrenamiento.

Trabajo solicitado. Para cada tarea: ¿autorizada con una cuenta personal, autorizada solo con el plan profesional, o excluida? Justifique.

Corrección. A) Autorizada con cualquier cuenta: documento público. B) Solo con el plan profesional: documento interno no sensible, tras verificar que realmente no contiene ningún nombre ni elemento de seguridad. C) Excluida en esta forma: datos personales y de vulnerabilidad de 2 300 hogares; la eliminación de duplicados se hace en la herramienta de gestión de beneficiarios de la organización, o se puede pedir a la IA que redacte una fórmula o un script de eliminación de duplicados probado con datos ficticios, y luego ejecutarlo localmente. D) Excluida: datos de violencia de género, extremadamente sensibles, incluso con un plan profesional. E) Autorizada con cualquier cuenta si la petición no contiene información interna sensible; preferiblemente con el plan profesional para adjuntar los procedimientos internos.

Para recordar

- Un texto pegado en un asistente sale de la organización: el proveedor lo procesa y a veces lo conserva.
- Minimización, finalidad y regulación de las transferencias se aplican al uso de la IA.
- Nunca datos personales de beneficiarios, casos de protección, salud o seguridad en un asistente de uso general.
- Los planes profesionales ofrecen compromisos contractuales, pero no sustituyen una decisión de la organización para los datos sensibles.
- Seudonimizar no es anonimizar; un error se notifica inmediatamente.

Cuestionario

1. Pegar una lista de beneficiarios en un asistente de uso general es:
   A. Una transferencia a un tercero para una finalidad no anunciada, que hay que evitar
   B. Autorizado si el historial está desactivado
   C. Autorizado si la lista está en inglés
   D. Sin consecuencias

2. Sustituir los nombres por «Beneficiario A, B, C» conservando una tabla de correspondencia es:
   A. Anonimizar
   B. Cifrar
   C. Suprimir
   D. Seudonimizar: siguen siendo datos personales

3. Un informe detallado de incidente de seguridad con itinerarios y horarios:
   A. Puede resumirse con una cuenta gratuita
   B. No debe darse a un asistente sin decisión formal de la organización
   C. Puede darse si está traducido
   D. Debe publicarse

4. Ha pegado por error datos sensibles. ¿Qué hacer?
   A. Nada
   B. Continuar la conversación
   C. Eliminar la conversación y notificar el incidente al referente sin esperar
   D. Cambiar de navegador

5. El principio de minimización significa:
   A. Escribir peticiones cortas
   B. Utilizar el menor número de tokens posible
   C. Pagar el plan más barato
   D. Utilizar solo los datos estrictamente necesarios

Respuestas del cuestionario : 1. A · 2. D · 3. B · 4. C · 5. D