ACADÉMIE HUMANITAIRE · HUMANITAIRES JOBS

Mini MBA IA et action humanitaire : travailler avec Claude et Gemini

Support de cours · Semaine 3 : Protéger les données : ce qu'on ne donne jamais à une IA, et comment anonymiser

Programme de 3 mois exactement (12 semaines, 144 heures) pour utiliser les assistants d'IA Claude (Anthropic) et Gemini (Google) dans le travail humanitaire, de façon utile, sûre et responsable : fonctionnement et limites, formulation des demandes, protection des données, éthique et principes humanitaires, rapports, propositions, analyse de données et MEAL, traduction et communication avec les communautés, recherche et vérification des sources, désinformation et sécurité, productivité et automatisations, politique d'usage et déploiement dans une organisation. Exercices corrigés, trois examens mensuels avec études de cas, et un diplôme vérifiable de l'Académie Humanitaire. Programme indépendant, sans partenariat avec Anthropic ni avec Google.

https://humanitairejobs.org/fr/academie/mini-mba-ia-claude-gemini/semaine-3/

Document gratuit de l'Académie Humanitaire. Il peut être imprimé et partagé librement à des fins de formation, sans modification.

Semaine 3 : Protéger les données : ce qu'on ne donne jamais à une IA, et comment anonymiser

Objectifs

- Appliquer les principes de protection des données dans l'action humanitaire à l'usage des assistants d'IA.
- Classer une information (publique, interne, confidentielle, sensible) et décider si elle peut être donnée à un assistant, et avec quelle offre.
- Anonymiser ou pseudonymiser un document avant de l'utiliser avec l'IA, et régler les paramètres de confidentialité.

Pourquoi c'est la semaine la plus importante

Les organisations humanitaires détiennent des données parmi les plus sensibles qui soient : identité de personnes réfugiées ou déplacées, survivantes de violences, enfants séparés, données de santé, positions d'équipes, incidents de sécurité, informations sur des groupes armés. Une fuite peut exposer des personnes à des représailles, à des discriminations ou à des violences.

Quand vous collez un texte dans un assistant d'IA, ce texte **quitte votre ordinateur** et est traité sur les serveurs du fournisseur, souvent dans un autre pays. Selon l'offre et les réglages, il peut être **conservé**, parfois **relu par des personnes** chargées de la qualité, ou utilisé pour **améliorer les modèles**.

> Principe : ne donnez jamais à un assistant d'IA une donnée que vous n'auriez pas le droit d'envoyer par courriel à un prestataire extérieur sans contrat de traitement des données.

Les principes de référence

Le **Manuel sur la protection des données dans l'action humanitaire** du CICR et du Brussels Privacy Hub (2e édition, 2020, qui traite aussi des nouvelles technologies), les politiques de protection des données des agences de l'ONU et, pour les organisations concernées, le Règlement général sur la protection des données (RGPD) de l'Union européenne posent des principes communs :

- **licéité et finalité** : une donnée est collectée pour un but précis et n'est pas réutilisée pour un autre but incompatible ;
- **minimisation** : n'utiliser que les données strictement nécessaires ;
- **sécurité** : protéger les données contre l'accès non autorisé ;
- **transferts** : encadrer tout transfert vers un tiers ou un autre pays ;
- **droits des personnes** : information, accès, rectification.

Donner les données d'une liste de bénéficiaires à un assistant d'IA grand public est un **transfert à un tiers**, pour une **finalité** qui n'a pas été annoncée aux personnes. C'est, dans la grande majorité des cas, contraire à ces principes.

Dans le secteur, on parle aussi de **données démographiquement sensibles** : même sans noms, des données agrégées (par exemple la localisation précise d'un groupe ethnique ou religieux minoritaire) peuvent exposer une communauté entière.

Ce qu'on ne donne jamais à un assistant grand public

- Noms, numéros de téléphone, numéros de carte ou d'enregistrement, photos, adresses, coordonnées GPS de bénéficiaires ou de membres du personnel.
- Toute information sur des survivantes et survivants de violences basées sur le genre, des enfants, des cas de protection, de santé mentale ou de VIH, même « anonymisée » de façon approximative.
- Rapports d'incidents de sécurité détaillés, itinéraires, positions et horaires de déplacement, plans d'évacuation, informations sur des négociations d'accès.
- Allégations d'exploitation et d'abus sexuels, de fraude ou de corruption en cours d'enquête.
- Données de santé individuelles.
- Mots de passe, clés d'accès, coordonnées bancaires.
- Documents marqués confidentiels par l'organisation, un bailleur ou un partenaire.

Pour ces catégories, même une offre professionnelle avec engagements contractuels ne suffit pas toujours : il faut une décision documentée de l'organisation (analyse d'impact sur la protection des données) avant tout usage.

Offres grand public et offres professionnelles

En novembre 2026, les grandes lignes sont les suivantes (vérifiez toujours les conditions en vigueur) :

- **Offres grand public (gratuites ou abonnement individuel)** : les deux fournisseurs proposent des réglages pour contrôler la conservation de l'historique et l'utilisation des conversations pour améliorer les modèles. Chez Google, l'activité des applications Gemini peut être conservée et des conversations peuvent être relues par des personnes ; un réglage permet de la désactiver. Chez Anthropic, l'utilisateur choisit dans les réglages de confidentialité si ses conversations peuvent servir à l'entraînement, et Claude propose un mode de conversation incognito.
- **Offres pour organisations** (Claude Team et Enterprise ; Gemini dans Google Workspace pour les entreprises et l'éducation) : les fournisseurs s'engagent contractuellement à ne pas utiliser les données des clients pour entraîner leurs modèles par défaut, avec une administration centrale, des contrôles d'accès et des conditions de traitement des données.

Concrètement :

- un compte personnel gratuit ne doit servir qu'à des informations **publiques ou sans enjeu** ;

- les informations **internes** (non sensibles) relèvent d'une offre professionnelle validée par l'organisation ;
- les données **personnelles et sensibles** restent exclues, sauf décision formelle et encadrée.

> Désactiver l'historique ou l'utilisation pour l'entraînement réduit les risques, mais ne transforme pas un compte personnel en outil autorisé pour des données de bénéficiaires.

Classer, puis anonymiser

Avant chaque usage, classez l'information selon la politique de l'organisation, par exemple :

- **Publique** : rapports publiés, appels à propositions, textes de loi. Utilisable avec tout assistant.
- **Interne** : brouillons de propositions, procédures, comptes rendus sans données personnelles. Offre professionnelle autorisée.
- **Confidentielle** : budgets détaillés, évaluations de partenaires, négociations. Offre professionnelle et accord du responsable, après retrait des éléments inutiles.
- **Sensible** : données personnelles, protection, sécurité, enquêtes. Exclue.

Anonymiser signifie rendre impossible l'identification d'une personne, y compris par recoupement. **Pseudonymiser** signifie remplacer les identifiants par des codes (« Bénéficiaire A ») ; la personne reste identifiable par qui détient la table de correspondance : ce sont encore des données personnelles.

Pièges courants : un village de 300 habitants plus « femme de 34 ans, veuve, quatre enfants » suffit à identifier quelqu'un ; une citation de groupe de discussion peut contenir un nom ou un surnom ; les métadonnées d'un fichier (auteur, commentaires, révisions) contiennent des noms. Retirez les détails inutiles, regroupez les âges et les lieux, et supprimez les commentaires et l'historique des documents.

Les bons réflexes au quotidien

- Utilisez le **compte fourni par l'organisation** pour le travail, jamais un compte personnel pour des documents internes.
- Vérifiez les **réglages de confidentialité** à la première connexion, puis tous les trimestres : historique, utilisation pour l'entraînement, mémoire, connecteurs activés.
- Limitez les **connecteurs** (accès à la messagerie, au Drive) au strict nécessaire : un assistant qui peut lire toute votre boîte mail peut aussi y lire des données sensibles.
- Ne partagez pas de **liens publics** vers des conversations contenant des informations internes.
- Sur le terrain, attention aux **écrans visibles** et aux appareils partagés ; déconnectez-vous.
- En cas d'erreur (données sensibles collées par mégarde) : supprimez la conversation, puis **signalez l'incident** au référent protection des données sans attendre. Une erreur signalée vite limite les dégâts ; une erreur cachée les aggrave.

Exercice d'application : peut-on utiliser l'IA pour ce document ?

Situation. Une équipe protection veut gagner du temps sur cinq tâches : A) résumer le rapport public d'un cluster protection ; B) traduire en anglais un compte rendu de réunion interne sans noms ; C) analyser un tableur de 2 300 ménages avec noms, numéros de téléphone et vulnérabilités pour repérer les doublons ; D) reformuler le récit d'une survivante de violence pour un rapport de cas ; E) proposer un plan de formation du personnel à la PSEA. L'organisation dispose d'une offre professionnelle validée par le responsable de la protection des données, avec engagement de non-utilisation pour l'entraînement.

Travail demandé. Pour chaque tâche : autorisé avec un compte personnel, autorisé avec l'offre professionnelle seulement, ou exclu ? Justifiez.

Corrigé. A) Autorisé avec tout compte : document public. B) Offre professionnelle seulement : document interne non sensible, après vérification qu'il ne contient vraiment aucun nom ni élément de sécurité. C) Exclu sous cette forme : données personnelles et de vulnérabilité de 2 300 ménages ; le dédoublonnage se fait dans l'outil de gestion des bénéficiaires de l'organisation, ou on peut demander à l'IA de rédiger une formule ou un script de dédoublonnage testé sur des données fictives, puis l'exécuter localement. D) Exclu : données de violence basée sur le genre, extrêmement sensibles, même avec une offre professionnelle. E) Autorisé avec tout compte si la demande ne contient aucune information interne sensible ; de préférence avec l'offre professionnelle pour y joindre les procédures internes.

À retenir

- Un texte collé dans un assistant quitte l'organisation : il est traité, parfois conservé, chez le fournisseur.
- Minimisation, finalité et encadrement des transferts s'appliquent à l'usage de l'IA.
- Jamais de données personnelles de bénéficiaires, de cas de protection, de santé ou de sécurité dans un assistant grand public.
- Les offres professionnelles offrent des engagements contractuels, mais ne remplacent pas une décision de l'organisation pour les données sensibles.
- Pseudonymiser n'est pas anonymiser ; une erreur se signale immédiatement.

Quiz

1. Coller une liste de bénéficiaires dans un assistant grand public est :
   A. Un transfert à un tiers pour une finalité non annoncée, à proscrire
   B. Autorisé si l'historique est désactivé
   C. Autorisé si la liste est en anglais
   D. Sans conséquence

2. Remplacer les noms par « Bénéficiaire A, B, C » en gardant une table de correspondance, c'est :
   A. Anonymiser
   B. Chiffrer
   C. Supprimer
   D. Pseudonymiser : ce sont encore des données personnelles

3. Un rapport d'incident de sécurité détaillé avec itinéraires et horaires :
   A. Peut être résumé par un compte gratuit
   B. Ne doit pas être donné à un assistant sans décision formelle de l'organisation
   C. Peut être donné s'il est traduit
   D. Doit être publié

4. Vous avez collé par erreur des données sensibles. Que faire ?
   A. Rien
   B. Continuer la conversation
   C. Supprimer la conversation et signaler l'incident au référent sans attendre
   D. Changer de navigateur

5. Le principe de minimisation signifie :
   A. Écrire des demandes courtes
   B. Utiliser le moins de jetons possible
   C. Payer l'offre la moins chère
   D. N'utiliser que les données strictement nécessaires

Corrigé du quiz : 1. A · 2. D · 3. B · 4. C · 5. D