ACADEMIA HUMANITÁRIA · HUMANITAIRES JOBS

Mini MBA IA e ação humanitária: trabalhar com o Claude e o Gemini

Material do curso · Semana 3 : Proteger os dados: o que nunca se dá a uma IA e como anonimizar

Um programa de exatamente 3 meses (12 semanas, 144 horas) para utilizar os assistentes de IA Claude (Anthropic) e Gemini (Google) no trabalho humanitário, de forma útil, segura e responsável: funcionamento e limites, redação de pedidos, proteção de dados, ética e princípios humanitários, relatórios, propostas, análise de dados e MEAL, tradução e comunicação com as comunidades, pesquisa e verificação de fontes, desinformação e segurança, produtividade e automatização, política de utilização e implementação numa organização. Exercícios corrigidos, três exames mensais com estudos de caso e um diploma verificável da Academia Humanitária. Programa independente, sem parceria com a Anthropic nem com a Google.

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

Documento gratuito da Academia Humanitária. Pode ser impresso e partilhado livremente para fins de formação, sem modificações.

Semana 3 : Proteger os dados: o que nunca se dá a uma IA e como anonimizar

Objetivos

- Aplicar os princípios de proteção de dados na ação humanitária à utilização dos assistentes de IA.
- Classificar uma informação (pública, interna, confidencial, sensível) e decidir se pode ser dada a um assistente, e com que plano.
- Anonimizar ou pseudonimizar um documento antes de o utilizar com a IA, e configurar as definições de privacidade.

Porque é a semana mais importante

As organizações humanitárias detêm alguns dos dados mais sensíveis que existem: identidade de pessoas refugiadas ou deslocadas, sobreviventes de violência, crianças separadas, dados de saúde, posições de equipas, incidentes de segurança, informações sobre grupos armados. Uma fuga pode expor pessoas a represálias, discriminação ou violência.

Quando cola um texto num assistente de IA, esse texto **sai do seu computador** e é processado nos servidores do fornecedor, muitas vezes noutro país. Consoante o plano e as definições, pode ser **conservado**, por vezes **lido por pessoas** encarregadas da qualidade, ou utilizado para **melhorar os modelos**.

> Princípio: nunca dê a um assistente de IA um dado que não teria o direito de enviar por e-mail a um fornecedor externo sem contrato de tratamento de dados.

Os princípios de referência

O **Manual sobre proteção de dados na ação humanitária** do CICV e do Brussels Privacy Hub (2.ª edição, 2020, que trata também das novas tecnologias), as políticas de proteção de dados das agências da ONU e, para as organizações abrangidas, o Regulamento Geral sobre a Proteção de Dados (RGPD) da União Europeia estabelecem princípios comuns:

- **licitude e finalidade**: um dado é recolhido para um fim preciso e não é reutilizado para outro fim incompatível;
- **minimização**: utilizar apenas os dados estritamente necessários;
- **segurança**: proteger os dados contra o acesso não autorizado;
- **transferências**: enquadrar toda a transferência para um terceiro ou para outro país;
- **direitos das pessoas**: informação, acesso, retificação.

Dar os dados de uma lista de beneficiários a um assistente de IA de uso geral é uma **transferência para um terceiro**, para uma **finalidade** que não foi anunciada às pessoas. Na grande maioria dos casos, é contrário a estes princípios.

No setor fala-se também de **dados demograficamente sensíveis**: mesmo sem nomes, dados agregados (por exemplo a localização precisa de um grupo étnico ou religioso minoritário) podem expor uma comunidade inteira.

O que nunca se dá a um assistente de uso geral

- Nomes, números de telefone, números de cartão ou de registo, fotografias, moradas, coordenadas GPS de beneficiários ou de membros do pessoal.
- Qualquer informação sobre sobreviventes de violência baseada no género, crianças, casos de proteção, saúde mental ou VIH, mesmo «anonimizada» de forma aproximada.
- Relatórios detalhados de incidentes de segurança, itinerários, posições e horários de deslocação, planos de evacuação, informações sobre negociações de acesso.
- Denúncias de exploração e abuso sexuais, de fraude ou de corrupção em curso de investigação.
- Dados de saúde individuais.
- Palavras-passe, chaves de acesso, dados bancários.
- Documentos marcados como confidenciais pela organização, por um doador ou por um parceiro.

Para estas categorias, nem sequer um plano profissional com compromissos contratuais basta sempre: é necessária uma decisão documentada da organização (avaliação de impacto sobre a proteção de dados) antes de qualquer utilização.

Planos de uso geral e planos profissionais

Em novembro de 2026, as grandes linhas são as seguintes (verifique sempre as condições em vigor):

- **Planos de uso geral (gratuitos ou subscrição individual)**: ambos os fornecedores oferecem definições para controlar a conservação do histórico e a utilização das conversas para melhorar os modelos. Na Google, a atividade das aplicações Gemini pode ser conservada e algumas conversas podem ser revistas por pessoas; uma definição permite desativá-la. Na Anthropic, o utilizador escolhe nas definições de privacidade se as suas conversas podem servir para o treino, e o Claude oferece um modo de conversa anónima.
- **Planos para organizações** (Claude Team e Enterprise; Gemini no Google Workspace para empresas e educação): os fornecedores comprometem-se contratualmente a não utilizar por predefinição os dados dos clientes para treinar os seus modelos, com administração central, controlos de acesso e condições de tratamento de dados.

Em concreto:

- uma conta pessoal gratuita só deve servir para informações **públicas ou sem importância**;
- as informações **internas** (não sensíveis) correspondem a um plano profissional validado pela organização;
- os dados **pessoais e sensíveis** ficam excluídos, salvo decisão formal e enquadrada.

> Desativar o histórico ou a utilização para o treino reduz os riscos, mas não transforma uma conta pessoal numa ferramenta autorizada para dados de beneficiários.

Classificar e depois anonimizar

Antes de cada utilização, classifique a informação segundo a política da organização, por exemplo:

- **Pública**: relatórios publicados, convites à apresentação de propostas, textos legais. Utilizável com qualquer assistente.
- **Interna**: rascunhos de propostas, procedimentos, atas sem dados pessoais. Plano profissional autorizado.
- **Confidencial**: orçamentos detalhados, avaliações de parceiros, negociações. Plano profissional e acordo do responsável, depois de retirados os elementos desnecessários.
- **Sensível**: dados pessoais, proteção, segurança, investigações. Excluída.

Anonimizar significa tornar impossível a identificação de uma pessoa, mesmo por cruzamento de dados. **Pseudonimizar** significa substituir os identificadores por códigos («Beneficiário A»); a pessoa continua identificável por quem tem a tabela de correspondência: continuam a ser dados pessoais.

Armadilhas frequentes: uma aldeia de 300 habitantes mais «mulher de 34 anos, viúva, quatro filhos» basta para identificar alguém; uma citação de um grupo focal pode conter um nome ou uma alcunha; os metadados de um ficheiro (autor, comentários, revisões) contêm nomes. Elimine os pormenores desnecessários, agrupe as idades e os lugares, e suprima os comentários e o histórico dos documentos.

Os bons reflexos do dia a dia

- Utilize a **conta fornecida pela organização** para o trabalho, nunca uma conta pessoal para documentos internos.
- Verifique as **definições de privacidade** na primeira ligação, e depois a cada trimestre: histórico, utilização para o treino, memória, conectores ativados.
- Limite os **conectores** (acesso ao e-mail, ao Drive) ao estritamente necessário: um assistente que pode ler toda a sua caixa de correio também pode ler nela dados sensíveis.
- Não partilhe **ligações públicas** para conversas que contenham informações internas.
- No terreno, cuidado com os **ecrãs visíveis** e os equipamentos partilhados; termine a sessão.
- Em caso de erro (dados sensíveis colados por descuido): elimine a conversa e depois **comunique o incidente** ao ponto focal de proteção de dados sem esperar. Um erro comunicado depressa limita os danos; um erro escondido agrava-os.

Exercício: pode-se utilizar a IA para este documento?

Situação. Uma equipa de proteção quer poupar tempo em cinco tarefas: A) resumir o relatório público de um cluster de proteção; B) traduzir para inglês a ata de uma reunião interna sem nomes; C) analisar uma folha de cálculo de 2 300 agregados familiares com nomes, números de telefone e vulnerabilidades para detetar duplicados; D) reformular o relato de uma sobrevivente de violência para um relatório de caso; E) propor um plano de formação do pessoal em PEAS. A organização dispõe de um plano profissional validado pelo responsável de proteção de dados, com compromisso de não utilização para o treino.

Trabalho pedido. Para cada tarefa: autorizada com uma conta pessoal, autorizada apenas com o plano profissional, ou excluída? Justifique.

Correção. A) Autorizada com qualquer conta: documento público. B) Apenas com o plano profissional: documento interno não sensível, depois de verificar que não contém realmente nenhum nome nem elemento de segurança. C) Excluída nesta forma: dados pessoais e de vulnerabilidade de 2 300 agregados; a eliminação de duplicados faz-se na ferramenta de gestão de beneficiários da organização, ou pode-se pedir à IA que redija uma fórmula ou um script de eliminação de duplicados testado com dados fictícios, e depois executá-lo localmente. D) Excluída: dados de violência baseada no género, extremamente sensíveis, mesmo com um plano profissional. E) Autorizada com qualquer conta se o pedido não contiver informações internas sensíveis; de preferência com o plano profissional para juntar os procedimentos internos.

A reter

- Um texto colado num assistente sai da organização: o fornecedor processa-o e por vezes conserva-o.
- Minimização, finalidade e enquadramento das transferências aplicam-se à utilização da IA.
- Nunca dados pessoais de beneficiários, casos de proteção, saúde ou segurança num assistente de uso geral.
- Os planos profissionais oferecem compromissos contratuais, mas não substituem uma decisão da organização para os dados sensíveis.
- Pseudonimizar não é anonimizar; um erro comunica-se imediatamente.

Quiz

1. Colar uma lista de beneficiários num assistente de uso geral é:
   A. Uma transferência para um terceiro para uma finalidade não anunciada, a evitar
   B. Autorizado se o histórico estiver desativado
   C. Autorizado se a lista estiver em inglês
   D. Sem consequências
2. Substituir os nomes por «Beneficiário A, B, C» conservando uma tabela de correspondência é:
   A. Anonimizar
   B. Cifrar
   C. Suprimir
   D. Pseudonimizar: continuam a ser dados pessoais
3. Um relatório detalhado de incidente de segurança com itinerários e horários:
   A. Pode ser resumido com uma conta gratuita
   B. Não deve ser dado a um assistente sem decisão formal da organização
   C. Pode ser dado se estiver traduzido
   D. Deve ser publicado
4. Colou por erro dados sensíveis. O que fazer?
   A. Nada
   B. Continuar a conversa
   C. Eliminar a conversa e comunicar o incidente ao ponto focal sem esperar
   D. Mudar de navegador
5. O princípio de minimização significa:
   A. Escrever pedidos curtos
   B. Utilizar o menor número de tokens possível
   C. Pagar o plano mais barato
   D. Utilizar apenas os dados estritamente necessários

Respostas do quiz : 1. A · 2. D · 3. B · 4. C · 5. D