📌 Segurança – Julia-IA (POC Felicitômetro)

1️⃣ Fluxo da POC

1. Usuário preenche **formulário** estilo Felicitômetro:
    a. Humor do dia
    b. Mensagem livre
2. Form envia os dados via **request** para a API (.NET 8 MVC).
3. Back-end envia esses dados + contexto básico do cadastro para a **OpenAI API**.
4. OpenAI retorna a resposta, que é enviada de volta ao front-end para exibição.

2️⃣ Identificação de riscos

Risco	Descrição	Impacto
Dados sensíveis	O formulário pode conter informações delicadas (saúde, conflitos internos, demissões, assédio, etc.)	Vazamento de dados ou exposição indevida de informações pessoais
Prompt injection / manipulação da IA	Usuário pode escrever mensagens tentando manipular o modelo para gerar respostas impróprias	A IA pode retornar conselhos inseguros ou informações não desejadas
Uso inadequado da IA	Dependência excessiva da IA para situações críticas	A IA não substitui RH/gestão; risco de dar conselhos incorretos ou criar falsas expectativas
Logs e rastreabilidade	Logs podem armazenar mensagens completas do usuário	Possível exposição de dados sensíveis caso o log seja acessado indevidamente

3 Medidas de mitigação

Risco	Medida adotada
Dados sensíveis	- Não armazenar o texto completo do usuário em banco - Apenas armazenar métricas agregadas para estatísticas (ex.: número de interações por humor)
Prompt injection	- Sanitizar inputs básicos (remover caracteres suspeitos) - Limitar tamanho máximo da mensagem - Modelar prompts de forma a não permitir instruções arbitrárias para a IA
Uso inadequado da IA	- Mensagem clara no front: *"A IA não substitui RH ou gestão. Para casos críticos, procure atendimento humano."* - Respostas críticas detectadas automaticamente por palavras-chave e redirecionadas para RH
Logs e rastreabilidade	- Armazenar apenas logs de eventos (ex.: "usuário enviou mensagem X com humor Y") sem conteúdo sensível - Criptografia de logs temporários se necessário

4 Observações

- A POC **não armazenará histórico completo do usuário**, apenas dados temporários para gerar resposta da IA.
- Qualquer situação crítica (**assédio, depressão, suicídio**) será **identificada automaticamente** e **não respondida pela IA**, apenas encaminhada ao RH.
- O foco da POC é **demonstrar integração IA → resposta ao usuário** com **medidas mínimas de segurança**, sem comprometer dados sensíveis.