Relatório Final

Introdução

Este projeto tem como objetivo o desenvolvimento de um sistema de chat baseado em modelos de linguagem (LLM) capaz de interagir com os usuários sobre conteúdos extraídos de arquivos CSV fornecidos. O sistema permite a extração e organização de informações de maneira eficiente, garantindo transparência, segurança e conformidade com as leis de proteção de dados pessoais.

A escolha deste tema justifica-se pela necessidade de análise de dados em arquivos extensos, como grandes tabelas de dados. Na aplicação desenvolvida, o usuário poderá interrogar agentes inteligentes sobre assuntos gerais e, em especial, sobre arquivos CSV, que podem ser analisados pela tecnologia, fazendo com que ela seja útil em casos reais que requerem análise de tabelas.

Desenvolvimento

O sistema foi desenvolvido com uma arquitetura modular composta por três principais componentes:

1. **Agentes:** Responsáveis pelo processamento das interações e pela lógica de negócios do chat. O agente principal utiliza um modelo de linguagem treinado para interpretar comandos do usuário e gerar respostas com base nos dados do arquivo CSV fornecido. A lógica de processamento envolve a extração de dados relevantes, a formatação adequada das respostas e a gestão do histórico de conversa para fornecer um diálogo coeso.
2. **Backend:** Implementado em Python, o backend atua como intermediário entre o frontend e os agentes, gerenciando as requisições dos usuários e garantindo a segurança dos dados processados. O backend recebe os comandos do frontend e os encaminha ao agente responsável, e, em seguida, retorna com a resposta que será exibida ao usuário via chat.
3. **Frontend:** A interface do usuário foi desenvolvida com HTML, CSS e JavaScript, proporcionando uma experiência interativa e intuitiva. O frontend permite que o usuário envie arquivos CSV, visualize os dados processados e interaja com o chatbot de maneira fluida, além de contar com um histórico de conversações para facilitar o acompanhamento das interações.

Durante o desenvolvimento, foi realizada uma análise de riscos para garantir a proteção dos dados pessoais inseridos nos arquivos CSV. Esta análise foi baseada na metodologia STRIDE e na abordagem de modelagem de ameaças de Torr, permitindo uma avaliação sistemática dos possíveis vetores de ataque e suas consequências.

As principais ameaças identificadas incluem:

- **Spoofing (Falsificação de identidade):** O risco de um agente malicioso se passar por um usuário autorizado.
- **Tampering (Manipulação de dados):** A possibilidade de alteração não autorizada dos dados armazenados ou em trânsito.
- **Repudiation (Repúdio):** Falta de mecanismos para evitar que um usuário negue uma ação realizada.
- **Information Disclosure (Divulgação de informações):** O risco de exposição indevida de dados sensíveis.
- **Denial of Service (Negativa de serviço):** A possibilidade de sobrecarga do sistema por ataques maliciosos.
- **Elevation of Privilege (Elevação de privilégio):** A obtenção de acessos indevidos dentro do sistema.

Para mitigar esses riscos, foram consideradas diversas medidas de segurança, tais como:

- Implementar autenticação multifator para verificar a identidade dos usuários.
- Utilizar assinaturas digitais e mecanismos de verificação de integridade para detectar alterações não autorizadas.
- Manter logs de auditoria detalhados e protegidos para rastrear as ações dos usuários.
- Classificar os dados e aplicar controles de acesso baseados em papéis, além de criptografar dados sensíveis.
- Implementar mecanismos de detecção e prevenção de intrusões, além de limitar as requisições por usuário.
- Aplicar o princípio do menor privilégio e revisar regularmente as permissões dos usuários.

A implementação dessas medidas assegura que os dados processados pelo sistema estejam protegidos contra ameaças comuns, alinhando-se às melhores práticas do setor.

O código-fonte completo do projeto está disponível no repositório do GitHub: github.com/LuizSoriano/TrabalhoSistemasDistruibuidos

Considerações Finais

O desenvolvimento deste sistema proporcionou um amplo aprendizado sobre a importância da proteção de dados pessoais e dos desafios envolvidos na implementação de soluções seguras. A análise de riscos realizada permitiu identificar potenciais vulnerabilidades e propor soluções eficazes para mitigação desses problemas.

O projeto demonstrou a viabilidade de construir um sistema de chat interativo baseado em LLMs, garantindo a conformidade com normas de proteção de dados.

Além disso, a aplicação de metodologias consagradas na segurança da informação reforça a relevância e confiabilidade da solução desenvolvida.

Por fim, espera-se que os conhecimentos adquiridos neste projeto possam ser aplicados em futuras implementações, contribuindo para a construção de sistemas cada vez mais seguros e eficientes.