

Título do curso: Auditoria de Algoritmos

Resumo:

Essa disciplina tem o objetivo de prover um entendimento dos princípios e práticas de auditoria de algoritmos. Ela aborda os principais conceitos de IA Responsável, como justiça (fairness), transparência, responsabilidade (accountability), privacidade, e segurança. Os alunos vão analisar artigos científicos relevantes e participar de um projeto prático, no qual irão realizar a auditoria de um algoritmo real, escolhido de um sistema real. Ao final do curso, os alunos devem ser capazes de identificar, medir e mitigar problemas em algoritmos, além de propor recomendações. O trabalho final do curso é um artigo de um experimento real para ser submetido a uma conferência de qualidade, preferencialmente internacional.

Estrutura do curso

- **Semanas 1-3:** Introdução e Fundamentos
- **Semanas 4-8:** Apresentação e Discussão de Artigos
- **Semanas 9-12:** Propostas para o Projeto prático de Auditoria.
- **Semanas: 16:** Apresentação Final dos Projetos

Semanas 1- 3: Introdução e Fundamentos

Aula 1: Introdução à Auditoria de Algoritmos

- O que é auditoria de algoritmos?
- A importância dos princípios de justiça, transparência, responsabilidade, privacidade e segurança
- Conceitos básicos: viés, justiça e discriminação.
- O papel da auditoria em aplicações reais: algoritmos em recrutamento, saúde e justiça.

Aula 2: Viés e Justiça

- Tipos de viés.
- Definições de justiça: paridade demográfica, igualdade de oportunidades.
- Métricas para avaliação de justiça.
- Exemplos e frameworks de auditoria de algoritmos (Aequitas, Fairlearn).

Aula 3: Transparência Algorítmica

- A necessidade de transparência.
- Conceitos de black box, grey box, white box
- Auditoria em algoritmos black box: métodos e desafios

- Explicabilidade (XAI): técnicas e aplicações.(SHAP, LIME, etc)

Aula 4: Auditoria Legal e Ética

- "Frameworks" legais como GDPR, LGPD, AI act entre outros.
- Aspectos éticos e a responsabilidade dos desenvolvedores de algoritmos.
- Governança e o papel dos stakeholders.

Semanas 4–8: Apresentação e Discussão de Artigos

Durante essas semanas, os alunos apresentarão artigos científicos sobre auditoria de algoritmos, seguidos de discussões em sala. O objetivo é analisar criticamente as vantagens e limitações das técnicas de auditoria existentes. Paralelamente, os alunos começarão a preparar seus próprios projetos de auditoria, definindo o escopo, métodos e ferramentas a serem utilizados.

Semanas 9–12: Propostas para o Projeto Prático

- Apresentação das propostas de auditoria de algoritmos.
- A proposta deve incluir:
 - **Seleção de um algoritmo:** escolha de um algoritmo para auditar em áreas como saúde, redes sociais, recrutamento, etc.
 - **Definição do escopo da auditoria:** identificação dos aspectos que serão auditados (justiça, transparência, etc.).
 - **Pesquisa inicial:** coleta de dados e formulação de hipóteses sobre o algoritmo.
 - Proposta inicial da organização (título, resumo e cronograma) do artigo do projeto final.

Relatório Final e Apresentação (Semana 16)

- **Condução da auditoria:** Uso das ferramentas e técnicas propostas na literatura para auditar o algoritmo selecionado.
- **Conteúdo do relatório:** Métodos utilizados, resultados obtidos e recomendações.
- **Apresentação final:** Discussão dos resultados do projeto em sala de aula.