



Plano de Desenvolvimento da Disciplina

MC934 - Tópicos em Engenharia de Computação
MO436 - Tópicos em Aprendizado de Máquina - Turma A
Oferecimento: Impactos sociais de inteligência artificial
Docentes: Prof. Jaques Wainer & Marcos M. Raimundo

Descrição

Este documento descreve, de forma sucinta, o plano de desenvolvimento da disciplina (PDD) de MC934 - Tópicos em Engenharia de Computação – abordando o impacto de inteligência artificial na sociedade, mais especificamente da turma A do primeiro semestre de 2024. Em particular, são destacados, de acordo com os requisitos do Regimento Geral de Graduação, o cronograma de atividades, os critérios de avaliação, punição para fraudes e plágios e a bibliografia a ser utilizada ao longo do semestre.

1. Programa da disciplina

O programa desta turma da disciplina cobrirá, em diferentes níveis de profundidade, tópicos relacionados à impactos sociais de inteligência artificial. O objetivo terminal da disciplina é que o aluno seja capaz de: a) diagnosticar os potenciais de uso de ferramentas de aprendizado de máquina e suas limitações legais, b) identificar os riscos reais das ferramentas de aprendizado de máquina que gerem danos a indivíduos e grupos vulneráveis e, c) julgar quais ferramentas devem ser criadas dentro de contexto de ética e quando criada diagnosticar caminhos de redução de impactos danosos. Para isso, serão abordados os seguintes assuntos (com possíveis alterações):

- **Bases conceituais da Inteligência Artificial, Ética e Direito:**
 - O que é IA e seus principais usos.
 - Ambiente social e jurídico que possibilitou a expansão (inadvertida) de IA.
 - Ética e modernidade.
 - O boom de IA, suas razões e desafios.
- **Impactos sociais da Inteligência Artificial:**
 - O papel de IA na reprodução e amplificação de vieses discriminatórios.

- Capitalismo de vigilância.
- Controle/indução de comportamento.
- Vigilância governamental e policiamento.
- Veículos autônomos (carros, e robos autônomas) e seus riscos.
- IA seu impacto em perda de empregos, desigualdade, e produtividade.
- IA, Ética e Industria
- Ameaça existencial.
- **Abordagens Éticas para a Inteligência Artificial:**
 - Leis regulatórias de Dados/IA.
 - Abordagens técnicas controle de vieses discriminatórios.
 - O direito a explicação e IA explicável.

2. Cronograma de atividades

O cronograma de atividades está descrito acima, consistindo numa série de palestras sobre diversos tópicos. A avaliação ocorrerá por 3 ensaios (de até 3 paginas) de cada bloco de palestras (Bases Conceituais E_1 , Impactos Sociais E_2 , Abordagens Éticas E_3).

3. Critérios de avaliação

A avaliação ocorrerá por ensaios (de até 3 paginas) das palestras, entregues semanalmente antes da aula posterior. A nota final NF consiste na media ponderada das notas:

$$NF = \frac{E_1 + 2 \times E_2 + E_3}{4} \quad (1)$$

A nota irá se basear na qualidade do resumo. Seguindo os seguintes nortes:

- 9/10 - Execução exemplar (raros equívocos), com pesquisa adicional com referências.
- 7/8 - Execução exemplar (raros equívocos), sem pesquisa adicional com referências. Execução com falhas, mas com pesquisa adicional com referências.
- 5/6 - Execução com falhas e sem pesquisa adicional referências.

- 3/4 - Execução com falhas excessivas.
- 0 - Não execução.

4. Punição para fraudes e plágios

Detecção de fraude ou plágio em um resumo implica em $E_i = 0$ para todos os envolvidos (quem recebeu ajuda e também quem ajudou). Reincidência implica em $NF = 0$ para todos os envolvidos.

5. Bibliografia

A bibliografia será apresentada durante o decorrer do semestre.