

Propostas de Iniciação Científica em Inteligência Artificial

Este documento apresenta três propostas de temas de Iniciação Científica em Inteligência Artificial, desenvolvidas para alinhar a pesquisa acadêmica na UFSCar com as competências mais valorizadas em grandes empresas de tecnologia (Meta, Microsoft, Netflix, Snap).

1. LLMs & Prompt Engineering

Objetivo: Investigar técnicas de prompt optimization e instruction fine-tuning para melhorar a performance de LLMs em tarefas específicas (ex.: sumarização de textos científicos ou análise de documentos jurídicos).

Metodologia:

- Implementar prompts manuais e automatizados via prompt tuning.
- Testar modelos open-source (LLaMA, Falcon, Mistral).
- Avaliar métricas: precisão, coerência, custo computacional.

Entregáveis:

- Artigo comparativo de técnicas de prompting.
- Repositório GitHub com scripts de fine-tuning.
- Dataset curado para experimentos.

2. Generative AI Multimodal

Objetivo: Desenvolver protótipo multimodal que integre texto + imagem (e vídeo opcional) para geração de recomendações personalizadas.

Metodologia:

- Usar CLIP embeddings para alinhamento imagem-texto.
- Testar diffusion models para síntese de imagens.
- Integrar transformers multimodais para contexto.

Entregáveis:

- Protótipo de geração de conteúdo multimodal.
- Benchmark em datasets (COCO, LAION).
- Notebook demonstrativo (Colab/Jupyter).

3. Optimization for Large-Scale Training

Objetivo: Avaliar estratégias de paralelização e otimização em GPU clusters para acelerar o treinamento de modelos de deep learning.

Metodologia:

- Implementar baseline em PyTorch.
- Aplicar data parallelism e model parallelism.
- Comparar eficiência em tempo de treinamento e consumo de GPU.

Entregáveis:

- Relatório técnico comparativo.
- Scripts CUDA/PyTorch otimizados.
- Submissão potencial em simpósio HPC/IA.

As três propostas foram elaboradas para garantir relevância acadêmica e, ao mesmo tempo, alta aderência às competências exigidas por grandes empresas de tecnologia. Cada tema pode ser ajustado de acordo com a disponibilidade de datasets, infraestrutura e orientadores na UFSCar.