Plano de Ensino – 2024/2

Código	DCC851
Disciplina	TECC: Seminários Avançados em Grandes Modelos de Linguagem — Eficiência
Turma	PG
Professor	Rodrygo Luis Teodoro Santos
Público	Pós-graduação
Carga horária	30h

Descrição do curso: Grandes modelos de linguagem (LLMs, na sigla em inglês) são uma nova classe de modelos de aprendizado de máquina que podem compreender e gerar texto com fluência e precisão sem precedentes. No entanto, tais capacidades vêm acompanhadas de uma alta demanda por recursos computacionais. Esta disciplina visa proporcionar um estudo dirigido sobre inovações recentes para tornar os LLMs mais eficientes computacionalmente, incluindo direções de pesquisa relacionadas à compressão de modelos, pré-treinamento eficiente, fine-tuning eficiente, inferência eficiente e projeto de arquiteturas eficientes. A disciplina terá o formato de seminários, com uma visão geral da área apresentada pelo professor e as demais aulas apresentadas pelos discentes matriculados, cobrindo artigos recentes relacionados aos vários tópicos da disciplina, seguidos de discussões em sala.

Pré-requisitos (informais): Conhecimentos sobre aprendizado de máquina ou aprendizado profundo; conhecimentos específicos sobre LLMs são desejáveis, porém não obrigatórios; em caso de dúvida, contate com o professor: rodrygo@dcc.ufmg.br.

Ementa: Fundamentos teóricos; custos e oportunidades; compressão de modelos; pré-treinamento eficiente; fine-tuning eficiente; inferência eficiente; arquiteturas eficientes

Programa (tentativo)

Class	Content
1	Overview of LLMs
2	Costs and Opportunities
3	Research Seminars
4	Research Seminars
5	Research Seminars
6	Research Seminars
7	Research Seminars
8	Research Seminars
9	Research Seminars

10	Research Seminars
11	Research Seminars
12	Research Seminars
13	Research Seminars
14	Research Seminars
15	Research Seminars

Bibliografia

1. Efficient Large Language Models: A Survey, by Wan et al. (2024)
2. Efficient Transformers: A Survey, by Tay et al. (2020)
3. Required and recommended readings, by several authors

Avaliações (tentativo)

1	Apresentação de seminários	60 pontos
2	Avaliação de seminários	20 pontos
3	Presença e participação	20 pontos