

DI 725– Transformers and Attention-Based Deep Networks

S Y L L A B U S

Year, Semester:	2023-2024 Spring
Course Conduct:	Face-to-face Lectures II-06 Wednesday@13:40 Lecture notes and reading materials on ODTUClass Lab sessions on Google Colab
Lecturers:	Prof. Dr. Tuğba Taşkaya Temizel, ttemizel@metu.edu.tr Prof. Dr. Alptekin Temizel, atemizel@metu.edu.tr
Teaching Assistant:	Mert Çağlar, mert.caglar@metu.edu.tr

Course Objective

This course explores advanced concepts and applications of transformers and attention-based models in various domains, focusing particularly on natural language processing (NLP), time series and computer vision as well as unified vision and language understanding. It covers topics such as attention, vanilla transformer, large language models (LLM), LLM frameworks, NLP applications with LLM, Unified Vision-Language Understanding and Multi-modal Transformers, Distillation and data-efficient transformers, explainability, flash attention, in-context learning, prompting, and ethical concerns. The course aims to give both theoretical and practical aspects of the topics and present real-world use cases.

Reference Material:

There is no textbook for the course. The following publications will be used as part of in-class materials:

- Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. Advances in neural information processing systems, 30.
- Chang, Y., Wang, X., Wang, J., Wu, Y., Zhu, K., Chen, H., ... & Xie, X. (2023). A survey on evaluation of large language models. arXiv preprint arXiv:2307.03109.
- Min, S., Lyu, X., Holtzman, A., Artetxe, M., Lewis, M., Hajishirzi, H., & Zettlemoyer, L. (2022). Rethinking the role of demonstrations: What makes in-context learning work?. arXiv preprint arXiv:2202.12837.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.
- Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).
- Carion, Nicolas, et al. "End-to-end object detection with transformers." European conference on computer vision. Cham: Springer International Publishing, 2020.
- Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, 2021.
- Kirillov, Alexander, et al. "Segment anything." arXiv preprint arXiv:2304.02643 (2023).
- Huang, Xinyu, et al. "Tag2text: Guiding vision-language model via image tagging." arXiv preprint arXiv:2303.05657 (2023).
- Touvron et al., Training data-efficient image transformers & distillation through attention. ICML, 2021

Grade Distribution:

4x Quizzes (Scheduled)	30%
3x Assignments	40%
Final project	30%

Deliverables

Documents and necessary files of the assignments must be uploaded to ODTUClass by students before the specified due dates.

Late Submissions

Late submissions are accepted with a penalty of 10% grade decrease for each day delayed.

University Policies

All students are **expected to obey** the university code of integrity and avoid academic dishonesty or plagiarism.

No	Date	
1	21 February	Introduction to the Course Introduction to NLP, NLP Processing Techniques
2	28 February	Foundations of Transformers Embeddings Sequence models Encoder-decoder models Attention mechanism
3	6 March	Transformer Architectures Introductory Lab Building blocks of transformer architecture
4	13 March	Lab Session – I Large Language Models (LLM) Encoder only, decoder only Encoder-decoder models Applications in natural language processing (NLP) LLM frameworks
5	20 March	Transformers in Computer Vision Quiz-1 Vision Transformers Detection Transformers
6	27 March	Computer Vision Applications Segment Anything Track Anything
7	3 April	Multi-modal Transformers Unified Vision-Language Understanding Image Tagging Guided Vision-Language Models YoloWorld
	10 April	Public holiday
8	17 April	Lab Session -II & Quiz-2 Vision Applications
9	24 April	Dynamic Learning In-context learning, Retrieval Augmented Generation (RAG) Prompting Zero and few shot learning
	1 May	Public holiday
10	8 May	Data Efficient Transformers Distillation and data-efficient transformers Self-distillation Sparsity, scaling
11	15 May	Transformers for time series Applications and architectures in time series domain Lab Session -III & Quiz-3
12	22 May	Hybrid Models, Performance Optimization CNNs and Transformers, hybrid models Flash Attention Performance Optimized Models
13	29 May	Explainable AI and Social Aspects Quiz-4 Security and misuse, harm and ethical concerns Expressivity, interpretability and visualization