



UNIVERSITÀ DI PISA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
Dottorato di Ricerca in Ingegneria dell'Informazione

Doctoral Course

**“Advanced Topics in Deep Learning:
the rise of Transformers in sequence and image modeling”**

Federico A. Galatolo, Ph.D.

University of Pisa - Italy

federico.galatolo@unipi.it

Short Abstract:

Transformers have brought a significant shift in the AI field through the introduction of a novel learning and information processing paradigm. Their attention mechanisms, including self-attention, empower them to grasp intricate patterns and relationships in data, rendering them highly adaptable in addressing a variety of complex tasks. Transformers are capable of executing an extensive spectrum of AI tasks: Machine Translation, Text Generation, Sentiment Analysis, Named Entity Recognition, Text Classification, Image Generation and Processing, Image Captioning, Image Generation, Style Transfer, Object Detection, Multimodal AI, Multimodal Translation, Visual Question Answering, Text-to-Image Synthesis, Recommendation Systems, Time Series Analysis and Prediction, Speech Recognition and Synthesis, Graph-based Tasks, Molecular Structure Prediction, Conversational AI, Summarization, Question Answering, Formulation of Robot Instructions .

First introduced by Google in 2017, Transformers are today the core of revolutionary technologies, such as ChatGPT, Google Search, Dall-E, and Microsoft Copilot, overtaking the most commonly employed Deep Learning neural network architectures across various applications, such as Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs).

In this series of lectures, participants will achieve professional experience with the industrial frameworks and languages for building transformer-based architectures in the domains of sequence and image modeling. The lectures focus on decoder/encoder architectures, fine-tuning techniques, alignment problems, vision models, multimodal models, text-to-image models, and large-scale inference.

Course Contents in brief:

- Introduction

- The transformer architecture and its applications (BERT, GPT, ViT, DeTR, etc.) **[1] [2] [3] [4]**
- Python and the huggingface framework (transformers, tokenizers, datasets, diffusers, etc.)
- Decoder-only architectures
 - GPT family **[3]**
 - Few Shot Learning **[1] [5]**
- Encoder-only architectures
 - BERT family **[4]**
 - Clustering and Classification
- Encoder-decoder architectures
 - T5 family **[6]**
 - Summarization, translation, paraphrasing
- LLM Fine Tuning
 - Parameter-efficient fine-tuning (PEFT, LoRA family, etc.) **[7]**
 - Quantization methods (GPTQ, GGML, etc.)
- Alignment problem
 - ChatGPT **[8]**
 - Instruction Following fine-tuning
 - Reinforcement Learning from Human Feedback (RLHF)
- Vision Models
 - Vision Transformers (ViT) **[9]**
 - Detection Transformers (DeTR) **[10]**
 - Bootstrapping Language-Image Pre-training (BLIP)
- Multimodal Models
 - Contrastive Language-Image Pretraining (CLIP) **[11]**
 - Large Language-and-Vision Assistant (LLaVA)
 - Visual Question Answering (VQA)
 - Document Question Answering (DQA)
- Text to Image Models
 - Image Synthesis history (GANs)
 - Diffusion family **[12]**
- Large Scale Inference
 - FastAPI
 - vLLM

- Text Generation Inference

Total # of hours of lecture: 20 hours

References:

- [1] Galatolo, F.; Cimino, M. and Vaglini, G. (2022). Zero-shot Mathematical Problem Solving via Generative Pre-trained Transformers. In *Proceedings of the 24th International Conference on Enterprise Information Systems - Volume 1*, (pp. 479-483).
- [2] Orlandi, D., Galatolo, F. A., Cimino, M. G., La Rosa, A., Pagli, C., & Perilli, N. (2022). Enhancing land subsidence awareness via InSAR data and Deep Transformers. In *2022 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* (pp. 98-103). IEEE.
- [3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.
- [4] Kenton, J. D. M. W. C., & Toutanova, L. K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
- [5] 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.
- [6] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1), 5485-5551.
- [7] Hu, E., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., & Chen, W. (2021). LoRA: Low-Rank Adaptation of Large Language Models.
- [8] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4.
- [9] Kolesnikov, A., Dosovitskiy, A., Weissenborn, D., Heigold, G., Uszkoreit, J., Beyer, L., ... & Zhai, X. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.
- [10] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In *European conference on computer vision* (pp. 213-229). Cham: Springer International Publishing.
- [11] Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International conference on machine learning* (pp. 8748-8763). PMLR.
- [12] Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34, 8780-8794.
-

CV of the Teacher

Dr. Federico A. Galatolo is a researcher at the Department of Information Engineering of the University of Pisa. He got the Ph.D. in Information Engineering with a thesis work on "Efficient Information Representation and Propagation in Artificial Neural Networks". He teaches "Computational Intelligence and Deep Learning" at the MSc Artificial Intelligence and Data

Engineering. He is a founding member of the Machine Learning and Process Intelligence (MLPI) research initiative at the University of Pisa.

Dr. Galatolo is member of the working group “WP 1.1 Explainable AI for synergistic Human-AI collaboration”, of the Extended Partnership PNRR PE1 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGenerationEU program.

He collaborated to the following research projects, as principal designer of innovative Deep Learning architectures: MUR-FISR 2019 SERICA “Interactive Texts and intelligent Networks”, MUR-PRIN 2020 REASONING “foRmal mEthods for computAtional analysis for diagnOsis and progNosis in imaging”, SecureB2C “A distributed platform for cybersecurity and payments for business-to-consumer companies”, DoctOraAI “Diagnosis of Oral Cancer via Explainable Artificial Intelligence”, PRA 2022 “Decision Support Systems for territorial networks for managing ecosystem services”. Dr. Galatolo co-authored 20+ publications in international conferences and journals. He is an open source contributor, with over 80 public open source projects.

Final Exam: project

Room and Schedule

Room: *Aula Riunioni del Dipartimento di Ingegneria dell'Informazione, Via G. Caruso 16, Pisa – Ground Floor*

Aula Riunioni del Piano 6 del Dipartimento di Ingegneria dell'Informazione, Largo Lucio Lazzarino 1, Pisa

Schedule:

11/03/2024: 10.00 – 13.00, Lecture 1 (3h) - **Via Caruso**

12/03/2024: 15.00 – 18.00, Lecture 2 (3h) - **Via Caruso**

13/03/2024: 15.00 – 18.00, Lecture 3 (3h) - **Largo L. Lazzarino**

14/03/2024: 15.00 – 18.00, Lecture 4 (3h) - **Largo L. Lazzarino**

15/03/2024: 15.00 – 18.00, Lecture 5 (3h) - **Via Caruso**

18/03/2024: 15.00 – 17.00, Lecture 6 (2h) - **Via Caruso**

22/03/2024: 15.00 – 18.00, Lecture 7 (3h) - **Via Caruso**