Course Syllabus: Natural Language Processing

Objectives

THE COURSE PROVIDES THEORETICAL, METHODOLOGICAL, TECHNOLOGICAL, AND DESIGN KNOWLEDGE RELATED TO AUTOMATIC LANGUAGE AND TEXT UNDERSTANDING, FRAMING THE INNOVATIVE PARADIGMS INTRODUCED BY LARGE LANGUAGE MODELS AND THE NUMEROUS MODERN APPLICATIONS OF SUCH TECHNOLOGIES WITHIN THE GENERAL FRAMEWORK FOR THE DEVELOPMENT OF NATURAL LANGUAGE PROCESSING SYSTEMS.

Knowledge and Understanding

BASIC CONCEPTS OF NATURAL LANGUAGE PROCESSING SYSTEMS. STANDARD LANGUAGE MODELS. LARGE LANGUAGE MODELS BASED ON TRANSFORMERS. NATURAL LANGUAGE PROCESSING APPLICATIONS WITH LARGE LANGUAGE MODELS. PROMPT ENGINEERING. FINE-TUNING OF LARGE LANGUAGE MODELS.

Applying Knowledge and Understanding

DESIGN AND IMPLEMENTATION OF A NATURAL LANGUAGE PROCESSING SYSTEM BASED ON LARGE LANGUAGE MODELS, EFFECTIVELY INTEGRATING EXISTING TECHNOLOGIES AND TOOLS AND OPTIMALLY CONFIGURING OPERATING PARAMETERS.

Prerequisites

REQUIRED EXAM: MACHINE LEARNING

Course Content

The course is structured into four teaching units:

UNIT 1: FUNDAMENTALS OF NATURAL LANGUAGE PROCESSING

- Basic concepts, tasks, evolution, and applications of Natural Language Processing.

- Text representation, tokenization, stemming, lemmatization, bag of words, n-grams, similarity measures, word embeddings.

- TF-IDF vectors, text classification and clustering, word embeddings.

- Neural networks and text analysis, application of CNNs, recurrent networks, and LSTMs.

- Implementation of a text classifier.

- Information extraction from text, named entity recognition, and question answering.

- Development of a simple chatbot using Python and SpaCy/Rasa.

UNIT 2: TRANSFORMERS
- Self-attention, multi-head attention, positional encoding, masking.

- Encoder and decoder structure of a Transformer.

- Introduction to Hugging Face.

- Encoder-decoder or seq2seq models (translation and summarization).

- Encoder-only models (sentence classification and named entity recognition).

- Decoder-only models (text generation).

- Definition and training of a large language model.

- Training a large language model.

UNIT 3: PROMPT ENGINEERING
- Zero-shot, few-shot, and chain-of-thought prompting, self-consistency, generated knowledge, prompt chaining, ReAct, Retrieval-Augmented Generation (RAG).

- Application of basic prompting techniques.

- Hands-on practice on RAG and LangChain.

UNIT 4: FINE-TUNING LARGE LANGUAGE MODELS
- Feature-based fine-tuning, updating the output layers, updating all layers, parameter-efficient fine-tuning (PEFT), and Low-Rank Adaptation (LoRA).

- Fine-tuning large language models.

- Reinforcement Learning with Human Feedback (RLHF).

- Hands-on practice on RLHF.

- Final project.

Teaching Methods

THE COURSE INCLUDES LECTURES AND CLASSROOM EXERCISES. THE LECTURES WILL PROVIDE STUDENTS WITH FUNDAMENTAL KNOWLEDGE ON THE MAIN BASIC AND ADVANCED TECHNIQUES FOR REPRESENTING, ANALYZING, AND CLASSIFYING TEXT IN NATURAL LANGUAGE USING LARGE LANGUAGE MODELS. THE EXERCISES WILL DEVELOP THE ABILITY TO APPLY THESE TECHNIQUES TO THE IMPLEMENTATION OF TEXT CLASSIFICATION AND ANALYSIS TOOLS AND QUESTION ANSWERING SYSTEMS. ATTENDANCE AT LECTURES IS MANDATORY, AND A MINIMUM ATTENDANCE OF 70% IS REQUIRED TO TAKE THE FINAL EXAM. ATTENDANCE WILL BE MONITORED THROUGH THE UNIVERSITY'S AUTOMATIC SYSTEM EASYBADGE.

Assessment

THE FINAL EXAM CONSISTS OF A GROUP PROJECT AND AN ORAL EXAM. THE PROJECT REQUIRES STUDENTS TO CRITICALLY APPLY THE METHODOLOGIES LEARNED DURING THE COURSE TO A PRACTICAL CASE. THE ORAL EXAM WILL ASSESS THE THEORETICAL KNOWLEDGE ACQUIRED DURING THE COURSE, THE ABILITY TO ARGUE DESIGN CHOICES MADE IN THE PROJECT WORK, AND THE ABILITY TO ANSWER QUESTIONS ON SPECIFIC TOPICS COVERED IN THE LECTURES. THE FINAL GRADE WILL BE DETERMINED BY THE AVERAGE OF THE TWO ASSESSMENTS.

References

REFERENCE TEXTBOOK:
H. LANE, C. HOWARD, H. M. HAPKE: 'NATURAL LANGUAGE PROCESSING IN ACTION - UNDERSTANDING, ANALYZING, AND GENERATING TEXT WITH PYTHON,' MANNING.

ADDITIONAL TEACHING MATERIAL WILL BE AVAILABLE IN THE DEDICATED SECTION OF THE COURSE ON THE UNIVERSITY'S E-LEARNING PLATFORM (HTTPS://ELEARNING.UNISA.IT), ACCESSIBLE TO STUDENTS USING THEIR UNIVERSITY CREDENTIALS.