SYLLABUS

1. Information regarding the programme

1.1 Higher education institution	**Babeş-Bolyai University**
1.2 Faculty	**Faculty of Mathematics and Computer Science**
1.3 Department	**Department of Computer Science**
1.4 Field of study	**Computer Science**
1.5 Study cycle	**Bachelor**
1.6 Study programme / Qualification	**Computer Science**

2. Information regarding the discipline

2.1 Name of the discipline (en) (ro)			**Understanding and Developing Large Language Models (LLMs)** **Înțelegerea și Implementarea de Modele Lingvistice Mari (LLMs)**				
2.2 Course coordinator			**Assist. Drd. Bogdan Mursa**				
2.3 Seminar coordinator			**Assist. Drd. Bogdan Mursa**				
2.4. Year of study	**3**	2.5 Semester	**6**	2.6. Type of evaluation	**E**	2.7 Type of discipline	**Optional**
2.8 Code of the discipline		**MLE5247**					

3. Total estimated time (hours/semester of didactic activities)

3.1 Hours per week	5	Of which: 3.2 course	2	3.3 seminar/laboratory	1lab + 2proj
3.4 Total hours in the curriculum	60	Of which: 3.5 course	24	3.6 seminar/laboratory	36
Time allotment:					hours
Learning using manual, course support, bibliography, course notes					12
Additional documentation (in libraries, on electronic platforms, field documentation)					16
Preparation for seminars/labs, homework, papers, portfolios, and essays					25
Tutorship					6
Evaluations					6
Other activities:					

3.7 Total individual study hours	65
3.8 Total hours per semester	125
3.9 Number of ECTS credits	5

4. Prerequisites (if necessary)

4.1. curriculum	• Python programming • Linear Algebra • Statistics • Data Structures and Algorithms
4.2. competencies	• Average programming skills in a high-level programming language and very good knowledge on data structures and algorithms.

5. Conditions (if necessary)

5.1. for the course	• Classroom with a video project device
5.2. for the seminar /lab activities	• Lab equipped with high-performance computers and Python installed.

6. Specific competencies acquired

Professional competencies	• CE1.3 Using the methods, techniques, and algorithms from AI in order to model several classes of problems • CE1.4 Identify and explain specific AI techniques and algorithms and using them to solve specific problems • CE1.5 Integrating the models and the specific solutions from AI in dedicated applications • C4.2 Interpretation of mathematical models and computer science (formal) • C4.3 Identifying appropriate models and methods to solve real problems • C4.5 Incorporation of formal models in specific applications in various fields
Transversal competencies	• CT1 Ability to conform to the requirements of organized and efficient work, to develop a responsible approach towards the academic and scientific fields, in order to make the most of one's own creative potential, while obeying the rules and principles of professional ethic. • CT3 Using efficient methods and techniques for learning, information, research and developing capabilities for using knowledge, for adapting to a dynamic society and for communicating in Romanian and in a worldwide spoken language.

7. Objectives of the discipline (outcome of the acquired competencies)

7.1 General objective of the discipline	• The goal of this course is to familiarize students with the field of natural language processing, focusing particularly on the latest advancements brought by transformer architecture. Students will be taught how to analyze, design, implement, and evaluate various NLP problems. This course aims to elucidate how NLP serves as a bridge between human language and machine understanding, enabling tasks

	like text classification, entity extraction, text summarization, text generation, chatbots, among others. Specifically, all these will be accomplished by leveraging the latest technical breakthroughs in Large Language Models (LLMs)
7.2 Specific objective of the discipline	• Understand various architectures of Large Language Models (LLMs) with a focus on transformer architectures for tasks such as text classification, entity extraction, text summarization, text generation, and many others. • Solve and analyze a natural language processing problem using specific theoretical frameworks and methodologies inherent to LLMs. • Understand and develop effective strategies for prompt engineering, including techniques for eliciting desired responses from LLMs through well-crafted prompts. • Learn techniques for fine-tuning and retraining Large Language Models to enhance performance and adaptability to specific NLP tasks. • Understand the metrics used to evaluate the performance of LLMs and the principles behind deploying these models in real-world applications, including bot creation.

8. Content

8.1 Course	Teaching methods	Remarks
1. **Introduction to LLMs and the Landscape of Generative AI.** Overview of the history of Natural Language Processing with a focus on Large Language Models (LLMs) and their significance in the field of generative artificial intelligence. Examination of various applications and tasks LLMs are employed for, highlighting their versatility.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
2. **The Evolution of Text Generation Technologies.** Tracing the development of text generation from pre-transformer models to current methodologies.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
3. **Deep Dive into Transformer Architecture.** Techniques and strategies for utilizing transformers in text generation tasks. Exploration of transformer architecture, the backbone of modern LLMs.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
4. **The Principle of Attention in Transformers.** Understanding the "Attention is all you need" concept and its revolutionary impact on LLMs.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
5. **Mastering Prompt Engineering.** Learning how to effectively design prompts to guide LLMs in generating desired outputs.	• Interactive exposure • Explanation • Conversation	

	• Didactical demonstration	
6. **Pre-Training Large Language Models and Scaling Laws.** Insights into the pre-training process, computational challenges, and the principles of scaling laws for LLMs.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
7. **Fine-Tuning LLMs for Specific Tasks.** Strategies for instruction-based fine-tuning, including single and multi-task adaptations.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
8. **Advanced Fine-Tuning Techniques.** Introduction to Parameter Efficient Fine-Tuning (PEFT) methods such as LoRA and Soft Prompts.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
9. **Reinforcement Learning from Human Feedback (RLHF).** Fundamentals of aligning LLMs with human values through RLHF, including feedback collection and reward models.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
10. **Enhancing LLM output using Reasoning and Act.** Explore the landscape of advanced fine-tuning and prompting strategies through method like Chain-of-thought (CoT, Reason Only), Act-only and ReAct across different domains, highlighting their task-solving trajectories and the distinct advantages of the ReAct approach.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
11. **Implementing LLMs in Real-World Applications & Introduction to LangChain.** Combining the exploration of deploying LLMs in real-world applications with an introduction to LangChain, covering document loading, vector stores, embeddings, and the fundamentals of Retrieval Augmented Generation (RAG).	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
12. **Ethics of AI.** Discover the evolving field of generative AI, emphasizing the need for responsible use and continuous innovation in LLM-powered applications.	• Interactive exposure • Explanation • Conversation • Didactical demonstration	
13. **Presentation of the student projects.**	• Interactive exposure • Explanation • Conversation • Dialogue, debate	
14. **Presentation of the student projects.**	• Interactive exposure • Explanation • Conversation • Dialogue, debate	

Bibliography
1. Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Pellat, M., Robinson, K., Valter, D., . . . Wei, J. (2022). **Scaling Instruction-Finetuned Language Models.**
2. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). ReAct: Synergizing **Reasoning and Acting in Language Models.**
3. Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). **BloombergGPT: A Large Language Model for Finance.**
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). **Attention Is All You Need**
5. Alammar, J,m Grootendorst, M. (2024). **Hands-On Large Language Models.**
6. Auffarth, B. (2023). **Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT and other LLMs**

8.2 Seminar / laboratory	Teaching methods	Remarks
1. **Introduction to LLMs and Text Generation.** Get hands-on experience with basic LLM operations, focusing on generating text using pre-trained models.	• Interactive exposure • Explanation • Conversation • Individual and group work • Dialogue, debate	
2. **Exploring Transformer Architectures.** Dive into transformer models, understanding attention mechanisms and their implementation in text generation tasks.	• Interactive exposure • Explanation • Conversation • Individual and group work • Dialogue, debate	
3. **Advanced Text Generation and Prompt Engineering.** Experiment with advanced text generation techniques and learn the art of prompt engineering to guide LLM outputs.	• Interactive exposure • Explanation • Conversation • Individual and group work • Dialogue, debate	
4. **Pre-Training and Fine-Tuning Strategies.** Hands-on session on the basics of pre-training LLMs and strategies for fine-tuning them on specific tasks.	• Interactive exposure • Explanation • Conversation • Individual and group work • Dialogue, debate	
5. **Reinforcement Learning from Human Feedback (RLHF).** Implement RLHF techniques, setting up feedback loops and understanding reward models to align LLM outputs with human values.	• Interactive exposure • Explanation • Conversation • Individual and group work • Dialogue, debate	
6. **Introduction to LangChain and Retrieval Augmented Generation (RAG)**. Begin working with LangChain, focusing on document loading, vector stores, and embeddings. Explore the implementation of RAG for enhancing LLM applications.	• Interactive exposure • Explanation • Conversation • Individual and group work • Dialogue, debate	
7. **Building a Chatbot.** Students will apply the	• Interactive exposure	

knowledge gained in LangChain and RAG to to build a functional chatbot. **PROJECT** **Phase 1 (Weeks 1 and 2): Introduction and Topic Selection** Presentation of a list of project topics that incorporate LLMs, focusing on the requirements from the standpoint of real-world clients. Students choose or propose their own project topics, working in groups. Discussion about the chosen projects to ensure feasibility and relevance by using the methodology of Generative AI project lifecycle. Initial state-of-the-art analysis, focusing on how similar challenges are approached using LLMs. **Phase 2 (Weeks 3 and 4): Preparation and Planning** Following their selected topic, each team is tasked with identifying and defining a list of NLP applications, then conducting a literature review to determine the highest performing pretrained models for those specified use cases. **Phase 3 (Weeks 5 and 6): Adapt and Align model I.** Apply prompt engineering techniques to refine the model's output without undergoing retraining, followed by an evaluation of the model's performance. **Phase 4 (Weeks 7 and 8): Adapt and Align model II.** Implement fine-tuning methods to retrain the models, enhancing their performance for the particularities of the selected topic, then proceed to evaluate the model. **Phase 5 (Weeks 9 and 10): Adapt and Align model III.** Incorporate Reinforcement Learning from Human Feedback (RLHF) and reward models to tailor the LLM output more closely with human values. **Phase 6 (Weeks 11 and 12): LangChain and** Retrieval Augmented Generation (RAG) Utilizing LangChain and RAG, students are	• Explanation • Conversation • Individual and group work • Dialogue, debate • Interactive exposure • Explanation • Conversation • Individual and group work	

required to integrate the LLM they developed into an actual application workflow. This integration should ensure the LLM's output is in harmony with topic-specific requirements, accomplished through the employment of document loading, vector stores, and embeddings. **Phase 7 (Weeks 13 and 14):** Oral presentations		

Bibliography

1. Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S. S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Pellat, M., Robinson, K., Valter, D., . . . Wei, J. (2022). **Scaling Instruction-Finetuned Language Models.**
2. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). ReAct: Synergizing **Reasoning and Acting in Language Models.**
3. Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). **BloombergGPT: A Large Language Model for Finance.**
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). **Attention Is All You Need**
5. Alammar, J,m Grootendorst, M. (2024). **Hands-On Large Language Models.**
6. Auffarth, B. (2023). **Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT and other LLMs**

9. Corroborating the content of the discipline with the expectations of the epistemic community, professional associations and representative employers within the field of the program

- Similar courses exist in the studying program of major universities in Europe and abroad.
- The course respects the IEEE and ACM Curricula Recommendations for Computer Science studies.
- The knowledge and skills gained from this course not only provide students with a foundation for embarking on a career in scientific research but also position them as sought-after LLM engineers in the industry, where there is a high demand for experts.

10. Evaluation

Type of activity	10.1 Evaluation criteria	10.2 Evaluation methods	10.3 Share in the grade (%)
10.4 Course	• The capability to utilize the knowledge acquired from the course and practiced in the labs to address practical problems and real-world requirements with applications in natural language processing and generative AI.	Oral examination (project)	60%
10.5 Seminar/lab activities	• The student possesses a thorough comprehension of Large Language Model	Practical Examination under continuous observation (solving lab tasks)	40%

	(LLM) concepts, including transformer architectures, prompt engineering, and LangChain applications.		

10.6 Minimum performance standards

- Students must prove that they acquired an acceptable level of knowledge and understanding of the core concepts taught in the class, that they are capable of using this knowledge in a coherent form, that they have the ability to establish certain connections and to use the knowledge in solving various computer vision problems.
- The final grade (weighted average between the two presented evaluation methods) should be at least 5 (no rounding, from a scale from 1 to 10).

Date	Signature of course coordinator	Signature of seminar coordinator
12.03.2024	**Assist. Drd. Bogdan Mursa**	**Assist. Drd. Bogdan Mursa**

Date of approval ..... Signature of the head of department .....