Advanced Artificial Intelligence and Machine Learning: Natural Language Processing

LMH Summer Programmes are provided by Lady Margaret Hall, a college in the University of Oxford

Course:	Advanced Artificial Intelligence and Machine Learning: Natural Language Processing
Available:	Programme Session 2: 15th July 2024 to 2nd August 2024
Lectures:	18 Hours
Seminars:	12 Hours
Tutorials:	3 Hours
Independent Study:	Approximately 120 Hours
Recommended Credit:	15 CATS / 7.5 ECTS / 4 US Credits

About this Course:	Through predictive text, translation tools, and smart devices natural language processing (NLP) is increasingly a part of our day-to-day lives, and in large language models like Chat-GPT we see the enormous future potential of this exciting area of research. This advanced course examines the theoretical concepts of NLP and its current and potential future application in diverse domains. The course begins with an introduction to attention mechanisms, examining self-attention, transformers, and byte pair encoding, before turning to large language models (LLMs) and natural language generation, exploring how they use prompting and reinforcement learning with human feedback. You will look closely at the varied applications of NLP and LLMs in particular, such as question answering, translation, and code generation. In the final part of the course you will discover how language and vision can interact in applications such as video captioning or text to image generation, before looking to the future of NLP research and considering the limitations, biases, ethical concerns, and potential misuses of NLP. This intensive course offers students theoretical understanding and practical experience in a range of natural language processing concepts and techniques, offering career skills as well as excellent foundations for future research.

Course Overview:	Week 1
	• Attention Mechanism
	  ○ Introduction to the Attention Mechanism and to Recurrent Neural Networks and their shortcomings.
	• Self-Attention and Transformers
	  ○ Self-attention, transformer architecture, and Byte Pair Encoding.
	• Large Language Models and Natural Language Generation
	  ○ The legacy of LLMs and the latest LLMs.

Lady Margaret Hall, University of Oxford
Norham Gardens, Oxford, OX2 6QA, United Kingdom
vacation.programmes@lmh.ox.ac.uk
+44 (0)1865 274295

	○ Masked training mechanisms. ○ Bert, GPT, T5, and ChatGPT. • Prompting and Reinforcement Learning with Human Feedback ○ Prompt Engineering and RLHF in ChatGPT. Week 2 • Llama 2 [Falcon] and its workings ○ Hugging Face and the Transformers Library • Fine-Tuning LLMS ○ Methods for fine-tuning LLMs, including Low-Rank Adaptation and Quantized Low-Rank Adaptation. • Applications of NLP ○ Translation tools and question-answering. Week 3 • Applications of NLP ○ NLP for Code-Generation. • Vision and Text ○ The connections between language and computer vision. ○ Image and Video Captioning. ○ CLIP. ○ Text-to-image generation. • Model Analysis and Interpretability ○ How do we interpret what LLMs learn? • Advanced Considerations of NLP ○ Limitations and Biases. ○ Ethical Concerns and Potential Misuses of LLMs.
Key Texts:	Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, Cambridge MA, 2016. Olah, C., and Carter, S., *Attention and Augmented Recurrent Neural Networks*, Distill, 2016. Indurkhya, N., and Damerau, F.J. (eds.), *Handbook of Natural Language Processing* (2nd Ed), Boca Raton FL, 2010.
Learning Outcomes:	By the end of this course, you will: • Be able to demonstrate understanding of the algorithms and methods used to process textual data. • Understand the functionality of large language models and their training through finetuning, low-rank adaptation, and quantized low-rank adaptation. • Demonstrate understanding of the practical applications of natural language processing. • Be able to discuss the potential limitations, biases, ethical concerns, and misuses of NLP.
Admissions Requirements:	LMH Summer Programmes are designed for students who want to gain and develop knowledge in their chosen subject area. LMH Summer Programmes are intensive courses of study aimed at undergraduates who have completed one, two, or three years of their degree, or entry level postgraduate students. We will consider each applicant's academic ability and expect successful applicants to have a minimum grade point average equivalent to 2:1 level on the British grading scale. For example, this would mean at least a 3.2 GPA on the 4.0 grading scale in the United States, and 80% in China.

	This course would suit STEM students with intermediate level experience in artificial intelligence, machine learning, and natural language processing concepts and techniques, including those undertaking, or looking ahead to, graduate level study or research. **Specifically, students on this course must have experience of the following topics:** • **Knowledge of the deep learning libraries.** • **Understanding of deep learning, recurrent neural networks, GRU, and LSTMs.** • **Strong background in optimization and probability.** • **Familiarity with the Python programming language.** To participate fully in the programme all students will need to have proficiency in English. English language requirements for students who are not native English speakers: • Overall TOEFL score of 85; • or IELTS score of 6.5 (no less than 6.0 in each component); • or CET-4 at 550 or CET-6 at 520. If the language of instruction in your home institution is English you do not need to provide evidence of your English proficiency.
Teaching Methods:	Core syllabus material will be covered in lectures. Students attend four lectures each week and each lecture lasts 90 minutes. Seminars in smaller groups offer students space to discuss and debate, to dig deeper into difficult concepts, and to explore their own ideas. Student contribution to seminars is vital, and tutors will ensure everyone takes part in discussions. Seminars last 1 hour and students will take part in four seminars each week. Independent study is a crucial part of an LMH Summer Programme and of the Oxford teaching model. Tutors will recommend important reading to do between lectures and seminars that will enable students to come to class equipped to understand the information presented and prepared to take part in discussion and debate. Each week students will have an assignment of independent work to complete and submit in advance of the tutorial. There is an appropriate amount of space in the timetable to complete the necessary reading, preparation, and assignments. Students should expect to do around 40 hours of independent study each week. The final class each week is a tutorial, a very small class typically including only 2-4 students and central to the teaching methods used by the University of Oxford and on LMH Summer Programmes. Guided by their tutor, students will receive feedback on their assignments and be challenged to defend, justify, or even rethink their work and ideas. These rigorous academic discussions help develop and facilitate learning in a way that cannot be done with lectures and seminars alone.
Assessment:	On a three-week LMH Summer Programme students produce one piece of assessed work every week, which is submitted to the tutor and then discussed in a tutorial. At the end of each week students will receive a percentage grade for their submitted work. Each week's work counts for a third of the final percentage grade, so the final grade is an average of the mark received for each piece of work. Students who stay for six or nine weeks will receive a separate grade for each 3-week course.
Academic Credit:	Lady Margaret Hall will provide a transcript of students' assessed work, and can send this directly to your home institution if required. LMH Summer Programmes are designed to be eligible for academic credit, and we will communicate with home

	institution to facilitate this as needed. As a guide, we recommend the award of 15 CATS / 7.5 ECTS / 4 US Credits for each 3-week course.

Lady Margaret Hall, University of Oxford
Norham Gardens, Oxford, OX2 6QA, United Kingdom
vacation.programmes@lmh.ox.ac.uk
+44 (0)1865 274295