Imperial College London

University of the Year 2022 by The Times and Sunday Times' Good University Guide presents

Imperial's virtual masterclass

Advanced Machine Learning and Artificial Intelligence

- Techniques to solve real-world data-driven problems

Learn from Imperial's Data Science Institute experts live online!

Online via MS Teams

MASTERCLASS OVERVIEW

Everyone knows data is essential, but society still needs to gain the skills and tools to understand large datasets. Humans and AI applications are producing more data than ever, so it becomes more important to process the data and draw the right conclusion by understanding the limitations of your models.

This master class will give participants an understanding of these technologies and apply the knowledge and learning experience to design and develop machine learning techniques specific to real-world datasets. The course also focuses on advanced Machine Learning techniques drawn from the hands-on research experience of the presenters.

Topics covered include:

- **Introduction to machine learning**
  This will be a hands-on guide on dealing with data for a typical machine learning pipeline, plus advanced skills required to process data effectively.
- **Computer vision and applications**
  Focus on the latest advancements in computer vision and understand how all these techniques can be applied to various applications, including healthcare.
- **Text data**
  To process text data efficiently and understand the basics of text data analysis.
- **Data learning**
  To approach using data assimilation techniques integrated with machine learning to solve complex data-driven problems.

Team based learning via group project:

As part of this masterclass, students will have the opportunity to work in small teams on a group project. Students will be asked first to select a dataset and produce a machine learning pipeline to process the data. A list of sample datasets will be provided in the first week lecture, but students are also encouraged to bring their own data to work on. The group should be focusing on developing a project that uses some of the data processing techniques presented in the course work. Students will present their project on the last day of the programme.

On completion of this masterclass, students will be able to:

- Process and produce an advanced Machine Learning pipeline.
- Understand structure of an efficient Machine Learning algorithm that can be used for text and image processing, as well as Data Learning.
- Apply the knowledge and experience gained to develop a Machine Learning project and understand the performance metrics.
- Evaluate uncertainty and noise in big data systems and propose methods to reduce the errors propagation.
- Evaluate, select, and apply models and technologies to perform data learning.
- Design strategies to manage big data for real world forecasting models using data compressions and data decompressions.

PROGRAMME STRUCTURE AND TEACHING METHODS

Live lectures and tutorials spread over 8 sessions covering the following:

- 7 live lectures delivering via Microsoft Teams with group discussions and questions.
- Web based exercises and quizzes will be provided for formative feedback.
- Group projects for assessing the learning outcomes, supported by 5 times tutorials.
- 1 session on final day for project presentation.

Live classes will be delivered on weekdays between 08:00 and 12:00 UK time / 16:00 to 20:00 China time.

Project work will be done through team-based learning and tutorials. Final projects will be presented in groups on the last day of the programme. A prize will be awarded to the team with the best project.

The programme will be delivered over Microsoft Teams. Online project channels will be allocated to each team for project work. Students will be able to use the channel at any time to work on their project.

The entire programme will be taught in English.

CERTIFICATION

Students will receive a verified Imperial College London digital certificate on successful completion of this masterclass and a prize will be awarded to the best project team. Each student will also receive a transcript for their project marks.

ENTRY REQUIREMENTS

This masterclass is designed for undergraduate or postgraduate students studying in a **technical subject**, e.g. Engineering, Computing, Software Engineering, Math, Physics or related disciplines from a well-recognised university in China.

English requirements:
All students are required to have a good command of English, and if it is not their first language, they will need to satisfy the College requirement as follows:

- a minimum score of IELTS (Academic Test) 6.5 overall (with no less than 6.0 in any element) or equivalent.
- TOEFL (iBT) 92 overall (minimum 20 in all elements)
- CET- 4 (China) minimum score of 550
- CET- 6 (China) minimum score of 520

Technical requirements:
All students are expected to ***already have basic Artificial Intelligence and Machine learning knowledge*** and have ***a good level of Python, with a good understanding of statistics and numerical processing*****.**

Students will need to have access to a computer with a webcam, microphone and good internet connection to attend the live classes.

PROVISIONAL SCHEDULE

Time (MS Teams)	**Session 1: Introduction** *Led by: Dr Ovidiu Șerban*
08:00 – 09:30	**Content** • Introduction to Machine Learning and AI • Data management, curation, and quality measurements • A guide to good practices in Machine Learning • Q&A
10:00 – 11:00	**Tutorial 1**

Time (MS Teams)	**Session 2: Training your model: pre-processing, sampling, and the loss function** *Led by: Dr Ovidiu Șerban*
08:00 – 09:30	**Content** • A discussion about various data types and feature representation • Sampling strategies • Definition of the loss function and impact on the overall training process • Q&A
10:00 – 11:00	**Tutorial 2**

Time (MS Teams)	**Session 3: Computer vision and applications (1)** *Led by: Dr Wenjia Bai*
08:00 – 09:30	**Content** • Supervised and unsupervised image classification • Image segmentation • Latest advances in computer vision • Q&A

Time (MS Teams)	**Session 4: Computer vision and applications (2)** *Led by: Dr Wenjia Bai*
08:00 – 09:30	**Content** • Applications to healthcare • Large-scale medical imaging data analysis • Q&A
10:00 – 11:00	**Tutorial 3**

Time (MS Teams)	**Session 5: Text processing** *Led by: Dr Ovidiu Șerban*
08:00 – 09:30	**Content** • Document representation • Transformer-based models for document representation • Task definition for various text processing problems • Data visualisation techniques for model performance • Q&A
10:00 – 11:00	**Tutorial 4**

Time (MS Teams)	**Session 6: Data Learning (1)** *Led by: Dr Rossella Arcucci*
08:00 – 09:30	**Content** • Uncertainty quantification, backward error analysis and propagation of errors in data-driven models • Data Assimilation for uncertainty minimization and data learning for effectively merging heterogeneous data • Q&A

Time (MS Teams)	**Session 7: Data Learning (2)** *Led by: Dr Rossella Arcucci*
08:00 – 09:30	**Content** • Data Reduction, data compression and optimal data selection. • Data Learning for big-data problems from real-world applications. • Q&A
10:00 – 11:00	**Tutorial 5**

Time (MS Teams)	**Session 8: Project Presentations** *Led by: Dr Ovidiu Șerban*
08:00 – 10:00	**Content** • Group project presentations • Q&A and feedback • Announcement of winning project group.

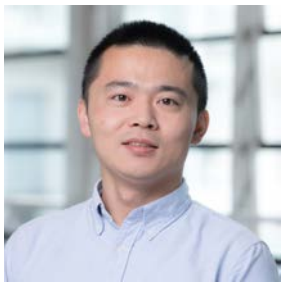
THE PRESENTERS

Dr Ovidiu Serban (Programme Director)
Research Fellow in Intelligent Data Processing and Curation, Data Science Institute,
Imperial College London.
https://www.imperial.ac.uk/people/o.serban

Ovidiu Șerban is a Research Fellow at the Data Science Institute (DSI), Imperial College London. His current work includes real-time Natural Language Processing, Data Curation and Large Scale Visualisation Systems.

Ovidiu's research topics are Natural Language Processing, Machine Learning, Affective Computing and Interactive System Design. He holds a joint PhD from INSA de Rouen Normandy (France) and "Babeș-Bolyai" University (Romania), while working at LITIS Laboratory in France.

In his youth, Ovidiu worked at the Institute for Security Science and Technology (ISST), Imperial College London; Computer Lab, University of Cambridge, UK and ISR Laboratory, University of Reading, UK.



Dr Wenjia Bai
Senior Lecturer at the Department of Computing & Department of Brain Sciences,

Imperial College London.

https://www.doc.ic.ac.uk/~wbai

Wenjia is Senior Lecturer (Associate Professor) jointly at Department of Computing and Department of Brain Sciences, Imperial College London. I am affiliated with Biomedical Image Analysis Group and Data Science Institute.

His research is at the interface between machine learning and medical imaging. He is interested in developing computational and machine learning algorithms to understand the structure, motion and function of anatomical organs from medical images. He works with colleagues with a wide spectrum of knowledge from computing to medicine.

Previously, he completed my D.Phil in Engineering Science at University of Oxford and my M.Eng and B.Eng in Automation at Tsinghua University.

Dr Rossella Arcucci
Lecturer at the Department of Earth Science & Engineering,

Imperial College London.

https://www.imperial.ac.uk/people/r.arcucci

Elected member of the WMO (World Meteorological Organization), Rossella contributes to the development of AI models for Climate and Environmental impact as part of the data assimilation and observing systems working group.

Rossella is a lecturer in Data Science and Machine Learning at Imperial College London where she leads the Data Assimilation and Machine Learning (Data Learning) Group. Rossella has been with the Data Science Institute at Imperial College since 2017, where she has created the Data Learning Group which is now a focal point for researchers and students of several departments at Imperial and other Universities in UK and Europe.

Since February 2022, she is the elected speaker of the Artificial Intelligence Network of Excellence at Imperial College, where she represents ~250 academics working on AI.

She collaborates with the Leonardo Centre at Imperial College Business School, where she contributes to the development of integrative, just and sustainable models of economic and social development by discovering, testing and diffusing new logics of business enterprise.
The models Rossella has developed have produced impact in many applications such as finance (to estimate optimal parameters of economic models), social science (to merge twitter and pooling data to better estimate the sentiment of people), engineering (to optimise the placement of sensors and reduce the costs), geoscience (to improve accuracy of forecasting), climate changes and others. She has developed accurate and efficient models with data analysis, fusion and data assimilation for incomplete, noisy or Big Data problems, always including uncertainty quantifications and minimizations.

She works on numerical and parallel techniques for accurate and efficient **Data Assimilation** and **Machine Learning** models. Efficiency is achieved by virtue of designing models specifically to take full advantage of massively parallel computers.

She finished her PhD in Computational and Computer Science in February 2012. She received the acknowledgement of Marie Sklodowska-Curie fellow from European Commission Research Executive Agency in Brussels in February 2017.

IMPERIAL COLLEGE LONDON AND THE DATA SCIENCE INSTITUTE

Consistently rated amongst the world's best universities (3rd in Europe and 7th in the World, QS World University Rankings 2022), Imperial College London is a science-based institution with an international reputation for excellence in teaching and research. Imperial attracts over 19,000 students and 8,000 staff of the highest international quality from over 136 different countries. Imperial has recently been named University of the Year 2022 by The Times and Sunday Times' Good University Guide.

Since its foundation in 1907, Imperial's contributions to society have included the discovery of penicillin, the development of holography and the foundations of fibre optics. This commitment to the application of research for the benefit of all continues today, with current areas of focus including interdisciplinary collaborations to improve global health, tackle climate change, develop sustainable sources of energy, address security challenges, develop data management and analysis technologies for supporting data driven research, and tackling problems at molecular scale.

Imperial's Centre for Continuing Professional Development had extensive experience in developing and running a range of online summer schools and masterclasses for undergraduate and postgraduate students. We draw on Imperial's education pedagogy in online learning to design and deliver programmes that provide an engaging learning experience for students. Various interactive applications are used to support live teaching and online group projects are designed to assess students' learning outcomes.

The Data Science Institute (DSI) is a major Imperial College London initiative that brings together Imperial's existing data science activities and expertise and provides a focus and a catalyst for new partnerships.

The DSI supports multidisciplinary collaborations between the College's academic experts in many disciplines such as healthcare, financial services, climate science, and city infrastructure to create solutions to complex problems. Alongside research, the Institute fosters the next generation of data scientists and engineers by developing a range of postgraduate and executive courses.

The DSI includes 7 Academic Labs, has attracted over £50m in funding for data science research, technology and infrastructure and has published over 300 papers.

The Institute's Data Observatory (DO) was one of the first and largest visualisation suites in Europe. It provides a multi-dimensional and immersive environment to analyse large and complex data sets and to work collaboratively. Thanks to its many research collaborations both across college and with a variety of external academic and industrial partners, the DSI is establishing its role as an international hub in data science.

Organsied by Imperial's Centre for Continuing Professional Development
www.imperial.ac.uk/cpd