# Master's Department of Engineering Sciences
## MSc in Automation Engineering and Robotics

**Name of the unit:** Advanced Deep Learning and Multimodal Learning

**Lecturer:** Nicolas Obin (nicolas.obin@sorbonne-universite.fr)
**Course code:** UM5RBI30
**Student workload:** 16 h of lectures, 24 h of labs
**Credits:** 6 ECTS
**Specialization tracks:** ☐ ROBOT ☒ SI ☐ MeDH
**Semester offered:** ☐ S1 ☐ S2 ☒ S3 ☐ S4
**Language of instruction:** ☐ French ☒ English
**Targeted audience:** ☒ Engineering Sciences students ☒ Other (specify): Poltech Sorbonne
**Localization :** ☒ PMC Campus ☐ Other (specify):

### Course Overview
This course unit provides advanced training in deep learning (neural networks and reinforcement learning), with a focus on learning from multimodal data such as text, sound and images. The course examines current neural architectures in depth (transformers, large language models, foundation models and generative models) and introduces emerging or complementary topics for intelligent systems. It also addresses the environmental footprint of the methods used, the interpretation and explainability of network outputs, and practical applications of embedded AI and AI for robotic systems.
*Keywords:* Deep learning. Generative modeling. Multimodal learning. Interpretability and explainability. Hardware optimization. Energy and environmental cost.

### Prerequisites
Students should have previously acquired the following prerequisites to follow this course:

- Mathematics: linear algebra, vector and matrix calculus, numerical optimization of continuous multivariable functions, differential calculus, probability
- Computer science: install and use a programming environment (language, IDE), use a virtual environment and version-control software, and use programming environments dedicated to deep learning (TensorFlow, PyTorch)
- Machine learning: master the fundamental concepts and methods of machine learning and deep learning; master data preparation for learning; master model-evaluation methodologies.

### Intended Learning Outcomes
By the end of this course, students will be able to:

1. C1.2. Analyze a problem: choose an algorithm suited to the task, data and available resources
2. C1.4. Analyze a problem: incorporate hardware, energy-consumption and environmental-footprint constraints when designing a solution to a learning problem
3. C2.1. Design: algorithmic foundations
4. C2.3. Design: specify hyperparameters to optimize the learning of model parameters
5. C2.4. Design: choose the learning process
6. C3.3. Implementation: prepare data for learning
7. C3.4. Implementation: carry out the learning process defined during the design phase to optimize the algorithm's parameters and hyperparameters
8. C3.5. Implementation: reproduce training processes and experimental results

9. C4.3. Evaluate: diagnose the causes of suboptimal network behavior and propose corrective solutions
10. C4.4. Evaluate: interpret and explain a model's decisions using attention maps and bias checks

### Indicative Teaching Sequence and Methods

| Week | C/TD/TP* | Content | Preparation | Learn. outc. |
|---|---|---|---|---|
| S1 | C1 (2h) | C1 - Transformeurs | | AAV 1, 3, 4, 5 |
| S2 | C2 (2h) - TP1 (4h) | C2 - Large language models. Self-supervised learning. Masking. Multimodality TP1 - Transformers | | AAV 1, 3, 4, 5, 6, 7 |
| S3 | C3 (2h) - TP2 (4h) P1 (4h) | C3 - Foundation models. Text, sound and images. Learning multimodal representations: CLIP, CLAP TP2 - Large language models P1 - Project session 1 | | AAV 1, 3, 4, 5, 6, 7 |
| S4 | C4 (2h) - TP3 (4h) | C4 - Generative models: GANs, diffusion, language models TP3 - Foundation models | | AAV 1, 3, 4, 5, 6, 7 |
| S5 | TP4 (4h) | TP4 - Generative models | | AAV 1, 3, 4, 5, 6, 7, 10 |
| S6 | C6 / TP6 (2h+2h) Quiz 1 P2 (4h) | C6 / TP6 - Environmental footprint of deep learning P2 - Project session 2 | | AAV 2 |
| S7 | C7 / TP7 (2h+2h) Quiz 2 | C7 / TP7 - Compression and acceleration | | AAV 2 |
| S8 | C8 / TP8 (2h+2h) Quiz 3 | C8 / TP8 - Interpretability and explainability | | AAV 2 |
| S9 | C9 / TP9 (2h+2h) Quiz 4 P3 (4h) | C9 / TP9 - Foundation models for robotics P3 - Project session 3 | | AAV 1, 3, 4, 5 |
| S10 | | | | |
| S11 | P4 (4h) | P3 - Project session 4 | | |
| S14 | Project defense | | | |
| S15 | ER (2h) | | | |

* C/TD/TP respectively corresponds to lectures, tutorials and lab sessions.

**Sequence of the unit:**
The course unit is organized into three sequences:
- Sequence 1: four advanced lectures on the fundamentals of current deep-learning architectures (C1-5, TP1-5, practical exam)
- Sequence 2: four introductory lectures on emerging or complementary topics related to intelligent systems (C6-9, ER)
- Sequence 3: a project in which students reproduce the results of a research paper (P1-4, project)

The first two sequences are consecutive. The project sequence begins in October.

### Indicative Assessment of Intended Learning Outcomes (1st session)

| Week | Individ./group | In-person/remote | Type of exam | Evaluated outcomes | Scale % |
|---|---|---|---|---|---|
| S7-10 | Individual | In-person | Quiz | AAV 6-10 | 20 % |
| S14 | Group | In-person | Other | AAV 1-10 | 40 % |
| S15 | Individual | In-person | Written exam | AAV 1-5 | 40 % |

### 2nde session of exam

| Session | Individ./group | In-person/remote | Type of exam | Evaluated outcomes | Scale % |
|---|---|---|---|---|---|

### Bibliographic references


- I. Goodfellow, Y. Bengio, A. Courvielle, Deep Learning, MIT Press, 2016
- Richard S. Sutton et Andrew G. Barto, Reinforcement Learning : An Introduction, second edition , 2018
- François Chollet, Deep Learning with Python. Mannan, 2017-2021

- Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2006

**Date of generation of this unit description:** October 6, 2026