

# UE Generative, Multimodal AI



Niveau d'étude  
Bac +5



ECTS  
6 crédits



Crédits ECTS  
Echange  
6.0



Composante  
UFR  
Informatique,  
mathématiques  
et  
mathématiques  
appliquées  
(IM2AG)



Période de  
l'année  
Automne (sept.  
à dec./janv.)

## Diplômes intégrant cet élément

- Master Informatique

- › **Langue(s) d'enseignement:** Anglais
- › **Ouvert aux étudiants en échange:** Oui
- › **Crédits ECTS Echange:** 6.0
- › **Code d'export Apogée:** GBX9AI03

## Présentation

### Description

The objective of this course is to provide the principles of advanced (supervised and unsupervised) machine learning algorithms, and explain their interest when applying them to address learning tasks using visual, auditory or textual data, as well as multi-modal combinations. Description The course is split into two parts. During the first part, a wide range of machine learning algorithms will be discussed. The second part will focus on deep learning, and presentations more applied to the three data modalities and their combinations. The following is a non-exhaustive list of topics discussed:

- Computing dot products in high dimension & Page Rank
- Matrix completion/factorization (Stochastic Gradient Descent, SVD)
- Monte-carlo, MCMC methods: Metropolis-Hastings and Gibbs Sampling
- Unsupervised classification: Partitionning, Hierarchical, Kernel and Spectral clustering
- Alignment and matching algorithms (local/global, pairwise/multiple), dynamic programming, Hungarian algorithm,...

- Introduction to Deep Learning concepts, including CNN, RNN, Metric learning
- Attention models: Self-attention, Transformers
- Auditory data: Representation, sound source localisation and separation.
- Natural language data: Representation, Seq2Seq, Word2Vec, Machine Translation, Pre-training strategies, Benchmarks and evaluation
- Visual data: image and video representation, recap of traditional features, state-of-the-art neural architectures for feature extraction
- Object detection and recognition, action recognition.
- Multimodal learning: audio-visual data representation, multimedia retrieval.
- Generative Adversarial Networks: Image-image translation, conditional generation

---

## Heures d'enseignement

|    |    |     |
|----|----|-----|
| CM | CM | 33h |
| TP | TP | 3h  |

## Infos pratiques

---

### Contacts

Responsable d'UE

Xavier Alameda-Pineda

✉ [xavier.alameda-pineda@inria.fr](mailto:xavier.alameda-pineda@inria.fr)

Responsable d'UE

Eric Gaussier

✉ [Eric.Gaussier@imag.fr](mailto:Eric.Gaussier@imag.fr)

---

### Campus

➤ Grenoble - Domaine universitaire