# INTRODUCTION TO MULTIMODAL ARTIFICIAL INTELLIGENCE (5 ECTS, S2 93001387):

An introduction To Machine learning and deep learning working with multimodal data and related application fields (image, video, speech, music, audio, text, …).

Processing and extracting knowledge from acoustics: reviewing traditional machine learning-based approaches.

Deep learning architectures for modelling acoustic signals as: time series (1D CNNs), time-frequency images (2D CNNs), sequences of tokens (LSTM, Transformers, etc.), differentiable digital signal processing, physics-informed neural networks, etc.

Building applications combining signal processing and deep learning: emotion recognition, pathological voices, speech recognition, music genre recognition, timbre- -transformation, sound source separation, virtual analogue modelling, room acoustics.

Introducing generative models and large multimodal models for applications in music, speech, sound generation from text, images, etc.

Ethical issues in generative AI