

PhD Position

Generating Synthetic Videos for Automated Visual Learning

Context

Recent advances in computer vision and generative AI have enabled major breakthroughs in image and video understanding. However, modern deep learning models remain critically dependent on massive amounts of annotated data, whose collection, annotation, and curation are costly, time-consuming, and increasingly constrained by privacy regulations such as GDPR. These limitations now constitute one of the main bottlenecks in the development of robust and generalizable visual learning systems.

This PhD project aims to explore the foundations of **synthetic video generation for automated visual learning**, with the objective of designing new learning paradigms where synthetic data is generated dynamically to enrich and optimize training datasets. The project focuses on human-centered computer vision tasks, including facial expressions, body motion, interaction modeling, and deepfake analysis.

The research will investigate how synthetic data can be leveraged not only for data augmentation, but also for creating entirely new forms of continual, interactive, and multimodal learning systems inspired by human cognition.

Research Objectives

The PhD will address three major research directions:

1. Controllable and Generalizable Video Generation

This direction will aim to develop generative models capable of producing realistic and controllable human-centric videos. Particular attention will be given to:

- controllable latent-space manipulation,
 - generation of facial expressions, gaze, motion, and interaction,
 - adaptation of image diffusion models to video generation.
- The objective is to design models that generalize across settings while requiring limited training data and computational resources.

2. Learning from Synthetic Data

The second direction will investigate how synthetic data can be integrated directly into the training process of computer vision systems. Topics include:

- self-supervised learning from generated videos,
- continual and lifelong learning,

- mitigation of catastrophic forgetting and dataset imbalance.
- The overarching goal is to develop visual learning frameworks where synthetic generation and model training are tightly interleaved.

3. Detection of Generated Videos and Deepfakes

As synthetic media becomes increasingly realistic, robust detection methods are essential. Here the objective will be to investigate generation and detection jointly in a co-evolutionary “cat-and-mouse” setting.

Research Environment

The project will be conducted within the interdisciplinary environment of **3IA Côte d’Azur** and specifically at the Inria Center at Université Côte d’Azur. Inria, the French National Institute for computer science and applied mathematics, promotes “scientific excellence for technology transfer and society”.

The STARS research team ([STARS – Spatio-Temporal Activity Recognition of Social interactions](#)) combines advanced theory with cutting edge practice focusing on cognitive vision systems.

The research for this Ph.D. lies at the intersection of Computer Vision, Generative AI, Representation Learning, and Continual Learning.

Candidate Profile

We are looking for highly motivated candidates with a Master’s degree in Computer Science, Artificial Intelligence, Applied Mathematics, or related fields. In addition, a strong background in machine learning and deep learning, experience in Python and deep learning frameworks (PyTorch preferred) is required. Experience in diffusion models, transformers, or video understanding is a plus.

Application

Applicants should send:

- CV,
- motivation letter,
- academic transcripts,
- and contact information for 2 references.

To apply, please email a full application to Antitza Dantcheva (antitza.dantcheva@inria.fr), indicating “3IA Ph.D. position” in the e-mail subject line.