# **Project Leaders**

Yaofei Duan

# Unsupervised Active Learning for Efficient Annotation

In real-world applications, the performance of artificial intelligence models often declines when there are discrepancies between training data and the data encountered in deployment environments—a challenge commonly known as **domain shift**. This issue is particularly critical in data-sensitive fields such as medical imaging, where collecting and labeling new domain-specific data is costly and time-consuming.

To address this challenge, this project introduces an innovative framework that integrates **unsupervised domain adaptation** with **active learning**, aiming to improve model robustness under limited annotation budgets. The proposed approach enhances data efficiency by selectively identifying the most valuable samples from diverse data sources, reducing the need for extensive manual labeling. A key feature of this framework is the use of advanced **distribution alignment techniques**, which minimize domain discrepancies by harmonizing data characteristics across different environments. Furthermore, it incorporates **intelligent sample selection mechanisms** that consider both uncertainty and representativeness, ensuring that each labeled sample contributes meaningfully to model improvement.

This methodology has demonstrated strong adaptability and scalability, making it applicable not only to medical diagnostics but also to a wide range of cross-domain machine learning tasks. By combining data-efficient learning with enhanced generalization capabilities, the project provides a practical solution for real-world scenarios where data diversity and resource limitations are major concerns.

