

Supervisory Team	Dr Qian Zhang
Short introduction & description of research project	<p>Title: Context-Based Learning with Large Language Models for Video Understanding in Activity Recognition and Quality Assessment</p> <p>This research aims to develop a novel context-based learning framework that enhances video understanding for activity recognition and quality assessment by integrating Large Language Models (LLMs), Vision-Language Models (VLMs), and motion capture data. Traditional models for video-based action recognition rely heavily on large-scale annotated datasets and struggle with capturing contextual and temporal dependencies. As a result, they perform poorly in evaluating action correctness, sequence adherence, and execution quality—critical in fields such as rehabilitation, sports analytics, and industrial training.</p> <p>The proposed project explores three key directions: (1) using context-aware knowledge distillation to train action recognition models with auxiliary textual data, reducing dependency on paired video-text datasets; (2) integrating visual, textual, and skeletal inputs through LLMs to enhance spatiotemporal reasoning in activity quality assessment; and (3) employing motion capture data to benchmark model performance and support fine-grained evaluation of movement quality. By incorporating domain knowledge, sequential understanding, and multimodal fusion, this research addresses core limitations in current models, enabling improved generalization, explainability, and feedback generation. The outcome will be a scalable and robust AI system capable of understanding complex human activities in real-world contexts with reduced annotation cost and enhanced interpretability.</p>
Contact points	Qian Zhang: qian.zhang@nottingham.edu.cn