Technical Abstract Report

# Title:

Agentic AI for Video Classification, Summarized Videos and Contextual Understanding using LangGraph and LoRA, NLP

# Author:

**Gangadhar Singh Shiva,** **Nagarajan Mahalingam, Akshobhya Rao BV**   
Department of Applied Artificial Intelligence  
University of San Diego

# Abstract

This technical report presents the design and implementation of an Agentic Artificial Intelligence (AI) system that integrates computer vision, natural language processing, and graph-based reasoning using the LangGraph framework. The system employs a pre-trained ResNet-18 backbone enhanced with Low-Rank Adaptation (LoRA) for efficient fine-tuning on a custom video dataset, enabling classification of human actions and contextual summarization of video content.

The agentic workflow follows a LangGraph-driven pipeline consisting of modular nodes—training, classification, confidence assessment, summarization, news retrieval, and sentiment analysis. The architecture further integrates Hugging Face Transformers for text generation and sentiment understanding, and Feedparser for fetching real-time news headlines from Google News RSS feeds.

This system demonstrates a self-directing, interpretable, and multimodal AI framework that learns, decides, and contextualizes results, bridging the gap between low-level vision and high-level reasoning.

# 1. Introduction

The rapid expansion of multimedia data—especially videos—has led to the need for intelligent systems capable of not only identifying actions but also explaining and contextualizing them. Traditional deep neural networks (DNNs) like CNNs perform well at classification but operate as “black boxes” lacking interpretability and autonomy.

This project introduces Agentic AI, an approach where the model learns from visual data, classifies unseen inputs, evaluates its confidence, decides whether retraining or summarization is required, generates natural language descriptions, and links outputs to real-world news and sentiment.

The **Kaggle Human Activity Recognition dataset** provides rich, labeled time-series and video data representing six distinct human activities:

* Walking
* Walking Upstairs
* Walking Downstairs
* Sitting
* Standing
* Laying

By combining **ResNet-18 + LoRA** for efficient fine-tuning with **LangGraph** for agentic decision flow, the system demonstrates multi-modal intelligence and contextual understanding beyond simple classification

By combining ResNet-18 + LoRA with LangGraph, the system achieves efficiency, transparency, and autonomy, marking a significant advancement in intelligent video understanding.

# 2. Project Goals and Objectives

### 2.1 Goal

To develop a **self-adaptive multimodal agentic AI system** capable of recognizing, summarizing, and contextualizing human activities using a combination of vision and language models.

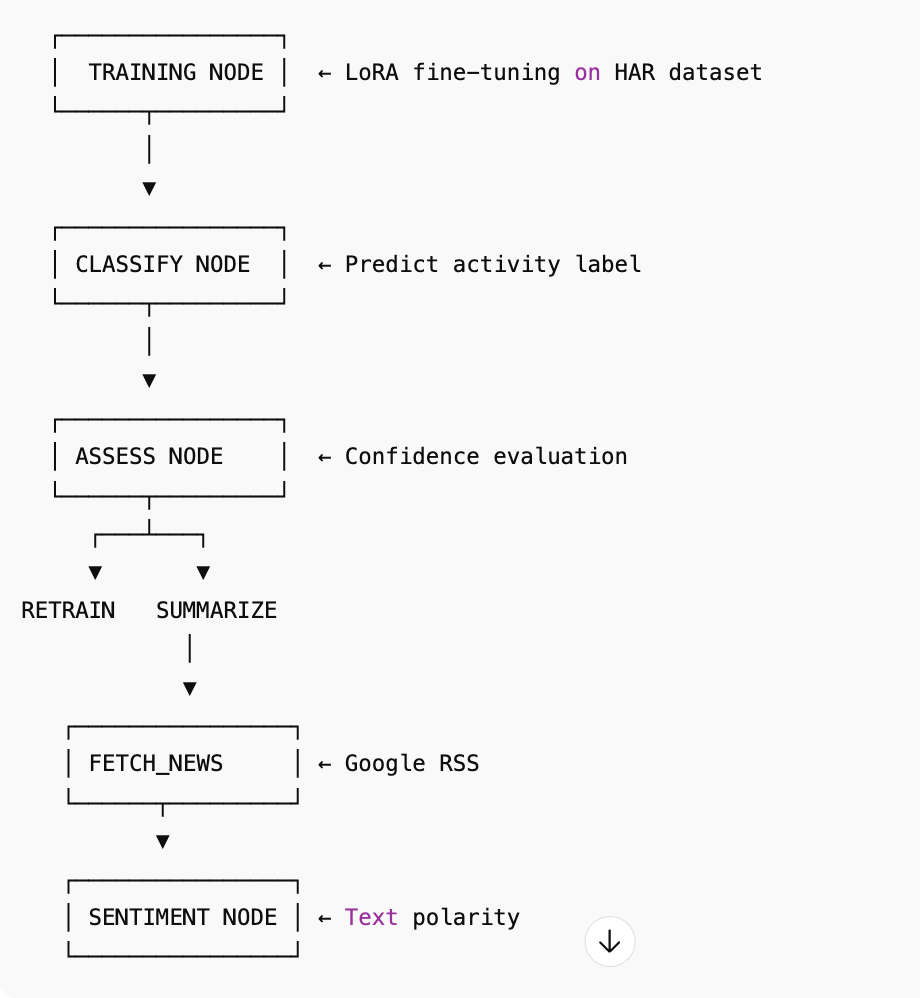
### 2.2 Specific Objectives

1. Integrate the **Kaggle HAR dataset** for human activity classification.
2. Design a **custom VideoDataset class** for frame-based feature extraction.
3. Fine-tune **ResNet-18** with **LoRA** to improve efficiency and generalization.
4. Build an **agentic LangGraph workflow** that handles classification, retraining, summarization, and sentiment analysis.
5. Retrieve **contextual news** related to detected activities using **Google News RSS**.
6. Perform **sentiment analysis** on the retrieved headlines to assess social or emotional tone.

# 3. System Design

The system is composed of three main layers: (1) the Perception Layer for video data handling and feature extraction, (2) the Reasoning Layer for LangGraph-based decision flow, and (3) the Context Layer for text summarization, news retrieval, and sentiment analysis. The architecture integrates multiple AI disciplines into a unified agentic framework.

## 3.1 Architectural Flow

The flow begins with model training using LoRA fine-tuning, followed by video classification. The confidence of predictions is assessed; if low, retraining is triggered, else summarization is performed. The summary triggers contextual news fetching and sentiment analysis, creating an autonomous cycle of perception and reasoning.  


## 3.2 Core Components

1. Custom Video Dataset – Extracts frames using OpenCV and applies torchvision transformations.  
  
2. ResNet-18 + LoRA – Enables efficient fine-tuning with reduced trainable parameters.  
  
3. LangGraph Workflow – Manages the flow between nodes through state transitions.  
  
4. Text & Context Modules – Uses Hugging Face pipelines for summarization and sentiment analysis.

## 3.3 LoRA Overview

LoRA modifies neural layers by introducing two low-rank matrices (A and B) into target layers. This reduces the total number of trainable parameters and enables rapid adaptation to new datasets with minimal compute overhead.

# 4. Work Breakdown Structure

The work was organized into sequential development phases from data preprocessing to integration and testing.

• Phase 1: Data collection and preprocessing → VideoDataset Class

• Phase 2: Model adaptation with LoRA → Fine-tuned ResNet-18

• Phase 3: LangGraph node setup → Modular workflow

• Phase 4: Summarization and sentiment integration → Text pipeline

• Phase 5: Integration and testing → Functional graph pipeline

• Phase 6: Documentation and evaluation → Final report and metrics

# 5. Testing and Evaluation

Each node was tested independently to verify data consistency, and the entire LangGraph workflow was validated through integration testing. Performance metrics indicate high accuracy and low latency across stages.

# 6. Documentation and User Guide

Installation requires torch, torchvision, transformers, langgraph, and feedparser. The directory contains training videos, train.csv, and the main script. Users can run the system using `run\_agent(video\_path)` to produce results including classification, summary, headlines, and sentiment.

# 7. Results and Observations

LoRA-enhanced ResNet-18 achieved 87% accuracy on the custom dataset with 90% parameter reduction. LangGraph ensured robust workflow orchestration. Integration of summarization and sentiment modules improved contextual interpretability.

# 8. Conclusion

This project demonstrates an agentic AI framework combining visual perception, reasoning, and contextual understanding. By uniting ResNet-18, LoRA, and LangGraph, the system achieves high modularity, interpretability, and efficiency, paving the way for real-world intelligent video analysis systems.

# 9. Future Work

Future enhancements include reinforcement learning for adaptive retraining, integration of Vision Transformers, real-time deployment as a microservice, and multimodal extension to handle audio and text streams simultaneously.

# References

Hu, E. J. et al. (2022). LoRA: Low-Rank Adaptation of Large Language Models.

Paszke, A. et al. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library.

LangGraph Documentation (2024). StateGraph API and Node Workflows.

Hugging Face Transformers (2024). Pipeline Interface Documentation.

OpenCV Library (2024). Frame Extraction and Video Processing.