

# Voice Assistant with Emotion Recognition\*

1<sup>st</sup> Bernardo Quindimil Micó  
*Universidad Intercontinental de la Empresa*  
A Coruña, España  
bernardo.quindimil.01@uie.edu

**Abstract**—This paper presents an AI voice assistant capable of detecting and responding to human emotions using a deep learning-based facial emotion recognition system. Inspired by the YOLO (You Only Look Once) framework, our approach enables real-time emotion analysis, facilitating context-aware interactions. This system has potential applications in psychological support, customer service, and personalized AI-driven experiences.

## I. INTRODUCTION

Recent advances in artificial intelligence (AI) have paved the way for more sophisticated human-computer interactions. Emotion recognition plays a critical role in understanding user needs and enhancing communication. This project aims to develop a voice assistant that can detect and analyze users' emotions through facial expression recognition and provide relevant responses. By combining deep learning-based emotion detection with NLP-based dialogue systems, this assistant can serve as an interactive and empathetic AI companion.

## II. OBJECTIVES

- Develop a deep learning model for real-time facial emotion recognition, inspired by YOLO.
- Construct a labeled dataset containing facial images annotated with emotional states.
- Implement a voice-based assistant capable of processing and responding to spoken language.
- Integrate emotional analysis with NLP techniques to generate context-aware responses.
- Explore potential applications in psychological support and mental well-being.

## III. METHODOLOGY

- **Dataset Collection:** Acquire and preprocess a dataset of facial images labeled with emotions such as happiness, sadness, anger, and surprise.
- **Deep Learning Model:** Develop a convolutional neural network (CNN) architecture inspired by YOLO to detect and classify facial emotions in real time.
- **Speech Processing:** Implement NLP techniques for speech recognition and response generation.
- **System Integration:** Combine vision and NLP components to enable the assistant to detect emotions and provide context-aware responses.
- **Evaluation:** Assess system performance using metrics such as f1-score, response relevance and user satisfaction.

Identify applicable funding agency here. If none, delete this.