Executive Summary

Project Title: NewsBot Intelligence System 2.0
Student: Rubén Darío Valenzuela
Course: ITAI 2373 – Natural Language Processing
Date: August 7, 2024

The NewsBot Intelligence System 2.0 is a production-ready natural language processing (NLP) platform that classifies, summarizes, and analyzes news articles using advanced AI techniques. Originally based on a basic classification model developed for the midterm project, this final version integrates eight core modules of NLP and machine learning to create a robust and versatile text analysis system.

This project supports multilingual input with automatic language detection and translation, performs sentiment and emotion analysis, extracts named entities and syntactic dependencies, and generates abstractive summaries using transformer models. It demonstrates mastery of NLP concepts such as tokenization, lemmatization, TF-IDF vectorization, text classification, sentiment scoring, named entity recognition (NER), dependency parsing, and text summarization.

The system was built using Python, scikit-learn, pandas, NLTK, spaCy, VADER, TextBlob, Hugging Face Transformers, and Google Translate API. It achieved 97% classification accuracy on a real-world dataset (BBC News corpus) across multiple categories such as politics, tech, entertainment, and business.

NewsBot’s pipeline is modular and scalable, designed for real-world applications in journalism, social media monitoring, or content moderation. The system includes a command-line interface and a future-ready architecture for deployment as a web or app service. This project has not only deepened my understanding of NLP workflows but also enhanced my ability to translate academic knowledge into practical AI solutions.

This summary reflects a commitment to developing intelligent, ethical, and scalable AI solutions, and showcases a professional-grade implementation of a modern NLP system.