



Proyecto Final 3

Análisis de emociones

Presentado a: Ing. Diego Alejandro Barragán Vargas
Julian Sebastian Alvarado Monroy , Cód. 2308042

Resumen— Este informe presenta el desarrollo e integración de un sistema de chatbot basado en DeepSeek AI en el robot social Pepper. El objetivo del proyecto fue habilitar a Pepper para mantener una conversación fluida con humanos mediante el uso de un modelo de lenguaje avanzado (DeepSeek) alojado en un servidor Flask, comunicándose a través de una red local. Se utilizó Docker para contenerizar el servidor de DeepSeek, y el cliente del robot se programó en Python 2.7 con la plataforma NAOqi de SoftBank. Como resultado, Pepper pudo escuchar al usuario, convertir el habla a texto, enviar la consulta al servidor DeepSeek y pronunciar la respuesta generada, mostrando incluso un estilo de comunicación personalizado (por ejemplo, con un tono humorístico al estilo Bart Simpson). Las pruebas realizadas confirmaron la correcta interacción, reproduciendo respuestas coherentes y apropiadas al contexto conversacional.

I. Introducción

Los robots sociales están diseñados para interactuar con las personas de forma natural y emocional. Pepper, un robot humanoide de Aldebaran Robotics (SoftBank), es el primer robot social que puede reconocer rostros y emociones humanas. Pepper está optimizado para la interacción a través de conversación y cuenta con sensores y módulos de reconocimiento de voz integrados. Debido a esto, Pepper se ha empleado en entornos comerciales y educativos para recibir, informar y guiar a usuarios, habiendo más de 17,000 unidades desplegadas hasta 2021.

Por otro lado, los modelos conversacionales avanzados basados en IA han experimentado un rápido avance. Los modelos de lenguaje grande (LLMs) como GPT-4 han demostrado gran habilidad para generar texto coherente y mantener diálogos naturales. En particular, DeepSeek AI es un nuevo LLM desarrollado en China como alternativa a GPT-4 o Gemini, que ofrece modelos de código abierto (como DeepSeek-R1) y acceso por API. Este modelo está optimizado para eficiencia y coste, y su API es compatible con el formato de OpenAI, permitiendo reutilizar SDKs existentes.

En este contexto, el proyecto buscó integrar Pepper con el modelo DeepSeek. Inspirados en proyectos previos que conectan Pepper con chatbots basados en IA, se planteó una arquitectura cliente-servidor donde Pepper funciona como cliente local que envía preguntas a un servidor remoto. Dicho servidor ejecuta una aplicación Flask que maneja las peticiones de chat y devuelve las respuestas de DeepSeek. A

continuación se desarrolla la teoría necesaria para comprender cada componente de este sistema.

El informe se organiza en cinco etapas: construcción del dataset, ejecución de código de captura, implementación del análisis de voz, revisión tecnológica del estado del arte, y publicación técnica en repositorios colaborativos.

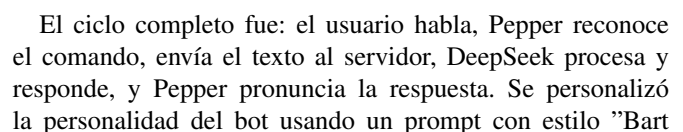
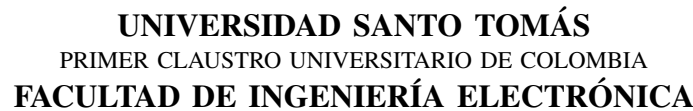
II. Marco Teórico

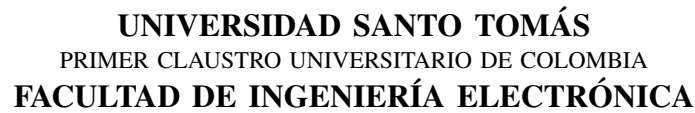
A. Robótica social (Pepper)

Los robots sociales constituyen una subclase de robots programables diseñados para interactuar socialmente con los humanos. Estos robots realizan tareas dentro de contextos sociales, comunicándose o provocando respuestas sociales en las personas. Pepper es un ejemplo destacado: es un robot humanoide social capaz de percibir emociones humanas y optimizado para el diálogo con usuarios. Pepper incluye micrófonos y un procesador de lenguaje para reconocer palabras, así como un sintetizador de voz para conversar con fluidez. Además, Pepper dispone de módulos de percepción (cámaras 2D/3D, sensores táctiles) y reconocimiento de voz en 15 idiomas distintos, lo que le permite comprender comandos verbales básicos y responder de forma multilingüe.

B. Reconocimiento de voz (ASR)

El reconocimiento automático del habla (ASR) es la tecnología que convierte señales de audio en texto legible por la computadora. Es un campo interdisciplinario que involucra aprendizaje automático y lingüística, y es fundamental para los sistemas de interacción oral con máquinas. En nuestro sistema, Pepper utiliza su motor ASR interno (ALSpeechRecognition de NAOqi) para transcribir la voz del usuario. Una vez transformado en texto, ese mensaje puede procesarse con un chatbot. El proceso inverso (texto a voz) se realiza mediante síntesis de voz (text-to-speech), que en Pepper está a cargo del módulo ALTextToSpeech. El uso combinado de ASR y TTS permite una conversación bidireccional hablada.





<https://github.com/JulianAlva24>
<https://www.overleaf.com/project/68465a8445555eeea03a75ab8>