

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

Chat Robot: Robot social para la generación de una conversación
basada en inteligencia artificial

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Ingeniero en Computación

Presentado por:

Dayse Joselyne Maroto Lema

Ivonne Gabriela Cuadros Cruz

GUAYAQUIL - ECUADOR

Año: 2021

DEDICATORIA

Le dedico mi proyecto de tesis a mi mamá Carmen Cruz y a mi abuelita Luzmila Figueroa, ambas me alentaron a no rendirme a pesar de los obstáculos que se presentaron al comienzo del camino, estuvieron incondicionalmente conmigo para que pudiera alcanzar esta meta. Mi mamá me enseñó con su ejemplo a trabajar para poder cumplir mis sueños, sin ella no estaría el día de hoy aquí. También a mis tías que me apoyaron a lo largo de este viaje con sus consejos y palabras de aliento para poder alcanzar con éxito el resultado del día de hoy. Finalmente, a mi novio que siempre estuvo a mi lado apoyándome e impulsándome a no rendirme.

Ivonne Gabriela Cuadros Cruz

DEDICATORIA

El presente proyecto se lo dedico a todas las personas que diariamente luchan por seguir sus sueños, que no se rinden a pesar de las adversidades que se le presentan en el camino, que enseñan con el ejemplo y enorgullecen a esta carrera. Aquellos que a través de su conocimiento, perseverancia e innovación logran crear tecnologías que repercuten en el mundo entero y abren campos de trabajo para futuras generaciones.

Dayse Joselyne Maroto Lema

AGRADECIMIENTOS

Mi más sincero agradecimiento a la Dr. Marisol Villacres, ella estuvo acompañándonos en este último paso y gracias a sus consejos y guías es posible que nos encontremos culminando con éxito esta etapa en nuestras vidas.

Además, me gustaría agradecerle al Dr. Dennys Paillacho por apoyarnos como tutor de nuestro proyecto final, finalmente a todos los profesores de la ESPOL que nos guiaron a lo largo de estos años hasta convertirnos en los profesionales que somos hoy en día.

Ivonne Gabriela Cuadros Cruz

AGRADECIMIENTOS

Agradezco a mi madre, Gloria Lema, por apoyarme en este largo camino, por ser mi ejemplo, mi roca y mi guía. A la Ing. Margarita Filián por enseñarme conocimientos útiles para mi vida laboral. A mis amigos del DST que siempre fueron gentiles conmigo. Al Ing. Rodrigo Saraguro que me vio empezar y culminar la carrera y se convirtió en un colega y amigo. A todos mis maestros de la carrera que compartieron conmigo sus conocimientos muy valiosos que hoy en día aplico en mi vida profesional.

Dayse Joselyne Maroto Lema

DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; Dayse Joselyne Maroto Lema y Ivonne Gabriela Cuadros Cruz damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”



Dayse Joselyne Maroto



Ivonne Gabriela Cuadros

EVALUADORES

PhD.Lucia Marisol Villacres Falconi

PROFESOR DE LA MATERIA

PhD. Dennys Fabian Paillacho Chiluza

PROFESOR TUTOR

RESUMEN

Las tecnologías están evolucionando este es el caso de los robots sociales que han cobrado una mayor popularidad, provocando que las personas sientan un mayor interés en ellos, además de que ayudan a potenciar la vida de las personas.

Actualmente en la ESPOL, en el centro de investigación CIDIS se encuentra un robot social al que le vemos este potencial, Loly es un robot que cuenta con una aplicación móvil, esta aplicación ofrece videojuegos con historias animadas respecto a la prevención del COVID 19 y desafíos. Además de esto Loly cuenta con movimientos, gestos y diálogos pregrabados para utilizarlos según sea conveniente.

Sin embargo, Loly actualmente solo funciona a través de esta aplicación MOVIL, en la cual los usuarios pueden interactuar con el robot eligiendo una opción dentro del menú que presenta. Todo esto no resulta ser óptimo debido a que no tiene una interacción directa con el usuario, lo cual limita la comunicación entre ambos. El nuevo módulo que se implementó en el robot social será muy útil para solucionar estos problemas.

La propuesta entregada es un chatRobot que tiene 1 modelo de inteligencia artificial, en el cual el usuario va a hablar directamente con el robot diciéndole una frase o palabra, donde el robot escuchará y analizará lo que usuario le dice, después este procederá respondiendo con audios y movimientos dinámicos de acuerdo con el entrenamiento que se le dio sobre la clasificación de alimentos saludables y no saludables.

Palabras claves: ChatRobot, modelo de inteligencia artificial, aplicación móvil.

ABSTRACT

Technologies are evolving, this is the case of social robots that have become more popular, causing people to feel more interested in them, in addition to helping to enhance people's lives.

Currently at ESPOL, at the CIDIS research center there is a social robot that we see this potential for, Loly is a robot that has a mobile application, this application offers video games with animated stories regarding the prevention of COVID 19 and challenges. In addition to this Loly has pre-recorded movements, gestures and dialogues to use as needed.

However, Loly currently only works through this MOVIL application, in which users can interact with the robot by choosing an option within the menu it presents. All this does not turn out to be optimal because it does not have a direct interaction with the user, which limits the communication between both. The new module that was implemented in the social robot will be very useful to solve these problems.

The proposal delivered is a chatRobot that has 1 artificial intelligence model, in which the user will speak directly with the robot saying a phrase or word, where the robot will listen and analyze what the user says, then it will proceed responding with audio and dynamic movements according to the training given on the classification of healthy and unhealthy foods.

Keywords: *ChatRobot, artificial intelligence model, mobile application.*

ÍNDICE GENERAL

EVALUADORES	5
RESUMEN	I
<i>ABSTRACT</i>	II
ÍNDICE GENERAL	III
ABREVIATURAS	VI
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	VIII
CAPÍTULO 1	1
1. Introducción	1
1.1 Descripción del problema	2
1.2 Justificación del problema	3
1.3 Objetivos	4
1.3.1 Objetivo General	4
1.3.2 Objetivos Específicos	4
1.4 Marco teórico	4
1.4.1 Robots sociales	4
1.4.2 Inteligencia artificial	5
1.4.3 ChatBot	5
1.4.4 Chatbots con Inteligencia Artificial	6
1.4.5 Lematización	6
1.4.6 Sintetizador de Voz	6
1.4.7 Los chatbots y el procesamiento y de lenguajes	6
1.4.8 Los chatbot y las coincidencias de patrón	7
1.4.9 Redes neuronales artificiales	7
1.4.10 Arquitectura de chatbots	7

1.4.11	Modelos generativos	8
1.4.12	Modelos basados en la recuperación	8
1.4.13	Heurística basada en patrones	8
1.4.14	Reconocimiento de voz	8
1.4.15	Análisis del habla y respuesta	9
CAPÍTULO 2		10
2.	Metodología	10
2.1	Resumen del capítulo	10
2.2	Análisis Primario	10
2.3	Definición del producto	11
2.4	Usuarios del sistema	11
2.5	Evaluación y Selección de Arquitecturas	12
2.6	Esquema de trabajo	12
2.6.1	Fase 1: Síntesis de voz	12
2.6.2	Fase 2: Chatbot	15
2.6.3	Fase 3: Integración del módulo de síntesis de voz, chatbot y movimientos del robot.	16
2.7	Requerimientos de Hardware	17
CAPÍTULO 3		18
3.	RESULTADOS Y ANÁLISIS	18
3.1	Detalles del desarrollo de la solución	18
3.1.1	Proceso de entrenamiento del Chatbot	18
3.1.2	Proceso del sintetizador de voz	21
3.1.3	Proceso de los movimientos del robot	21
3.1.4	Proceso de integración del proyecto	22
3.2	Pruebas y resultados	23

3.2.1	Resultado del entrenamiento	23
3.2.2	Resultado de encuestas	23
3.3	Análisis de costos	26
CAPÍTULO 4		28
4.	CONCLUSIONES Y RECOMENDACIONES	28
4.1	Conclusiones	28
4.1.1	Técnicas	28
4.1.2	Usos	29
4.2	Recomendaciones	29
4.2.1	Técnicas	29
BIBLIOGRAFÍA		30
EVALUADORES		5

ABREVIATURAS

ESPOL	Escuela Superior Politécnica del Litoral
NLP	Procesamiento del Lenguaje Natural
MLP	Perceptrón Multicapa
TTS	Texto a Voz
IA	Inteligencia Artificial
CIDIS	Centro de Investigación, Desarrollo e Innovación de Sistemas Computacionales

ÍNDICE DE FIGURAS

Figura 2.1 Modelo conceptual de la solución	11
Figura 2.2 Modelo Conceptual de la Fase 1 – Síntesis de voz	13
Figura 2.3 Transcripciones en formato LJ Speech	14
Figura 2.4 Modelo Conceptual de la Fase 2 – Chatbot	15
Figura 2.5 Modelo Conceptual de la Fase 3 - Integración del módulo de síntesis de voz, chatbot y movimientos del robot	16
Figura 3.1 API de Loly	22
Figura 3.2 Precisión del modelo MLP	23

ÍNDICE DE TABLAS

Tabla 2.1 Requerimientos de Hardware	17
Tabla 3.1 Clasificación de alimentos	19
Tabla 3.2 Parámetros utilizados para el entrenamiento del modelo	20
Tabla 3.3 Valores de los recursos usados	27

CAPÍTULO 1

En el desarrollo de este capítulo se mostrará lo importante que el proyecto llegará a ser. También se mostrarán diferentes metodologías, es decir los modelos que se usarán para el desarrollo del proyecto y la importancia de estos, así como todos los conceptos que serán necesarios para entender de una mejor manera el mismo. Además, se mostrará como este nuevo módulo de implementación ayudará a desarrollar y potenciar las capacidades con las que ya cuenta el robot Loly.

1. INTRODUCCIÓN

En la actual era tecnológica, los robots sociales han cobrado una mayor popularidad, provocando que las diferentes personas se sientan con mayor interés a estos, además de ayudar a potenciar la vida de las personas. Este tipo de robot presenta una ventaja frente a los otros ya que no solamente realiza ciertas tareas repetitivas, sino que interactúan con los seres humanos

En el caso de los niños, estos ya poseen ciertas capacidades para poder comunicarse y desarrollarse, sin embargo, es importante crear una relación entre los niños y la tecnología ya que esta le ayudará a facilitar el aprendizaje de diferentes temas, además de que al relacionarse los enfrentará a nuevas experiencias y conocimientos (Abad et al., 2018).

Este proyecto está encaminado, al desarrollo de un chatbot en un robot social, que pueda reconocer las entradas que realice el usuario por medio del habla, y le pueda responder de acuerdo a la necesidad que este requiera tanto verbal como no verbal, se utilizarán las redes neuronales convolucionales para llevar los diálogos que tendrá gracias a su entrenamiento predefinidos de textos a audios y adicional a esto se pretende usar las redes neuronales recurrentes para que el chatbot pueda reconocer la voz del usuario y así darle la ayuda necesaria de acuerdo a su requerimiento. De acuerdo con el entrenamiento que se realice, se podrá medir el resultado final del este proyecto.

Loly es un robot social, en el cual el CIDIS ha decidido invertir para que pueda interactuar con las personas, especialmente con los niños. Actualmente el robot ya se encuentra funcional, sin embargo, hay ciertas funcionalidades que se le pretenden agregar con este nuevo módulo que lo beneficiaran a largo plazo.

El robot llamado Loly ofrece videojuegos con historias animadas, desafíos para niños de 4 a 7 años. Para los niños que reciben educación regular e inclusiva, principalmente niños con síndrome de trastorno del espectro autista (TEA), se puede utilizar como una herramienta complementaria para desarrollar habilidades sociales y cognitivas. Los niños interactúan con el robot Loly, atrayendo su atención reproduciendo movimientos mecánicos, audio y gestos en tiempo real, motivándolos a jugar. Esto se logra a través de juegos basados en MIDI (multimedia, interactivos, educativos, para niños) (*ESPOL - Proyecto educativo de ESPOL, premiado en concurso mundial del BID, 2020*).

1.1 Descripción del problema

Loly es un robot social que actualmente funciona a través de una aplicación web, en la cual los usuarios pueden interactuar con el robot eligiendo una opción dentro del menú que presenta. Pero esto no resulta ser tan óptimo debido a que la interacción entre el robot y el usuario es repetitiva. Esto se presenta como una desventaja para los creadores de Loly, ya que su comunicación con el usuario solo se puede dar a través de esta aplicación web que tiene integrada. Es por eso por lo que se está buscando que Loly posea una comunicación más dinámica a través del chatbot que se quiere implementar.

Este proyecto tiene como objetivo final la implementación de un chatbot por medio de técnicas de inteligencia artificial, este será responsable de interactuar de una manera más amena con el usuario, este nuevo módulo se basará en el reconocimiento de palabras o frases que los usuarios quieran decir relacionados a un tema en específico, actualmente el tema que se eligió para realizar este proyecto es la alimentación saludable, es decir los usuarios por ahora solo podrán realizar preguntas y obtener respuestas relacionadas a este tema.

Una vez que esta nueva fase esté completamente desarrollada existirá una mejor interacción entre los usuarios y el robot, ya que será tan sencillo como que el usuario diga una determinada frase relacionada al tema en específico y el robot pueda generar la respuesta automática para esta.

1.2 Justificación del problema

El nuevo módulo que se desea implementar en el robot social será muy útil para poder establecer una comunicación menos repetitiva y más dinámica con el usuario. Además de ayudar a tener una interacción más fluida, la cual se da por medio del

habla y del lenguaje no verbal que serán los movimientos que Loly proyecte cuando este interactuando con el niño.

Esta nueva integración podrá ser un punto de partida para Loly, para poder reconocer en un futuro cualquier tema que el usuario disponga y este mostrará una respuesta oportuna a cada una de ella.

La solución que se ha propuesto puede llegar a ser el principio de una mejor interacción entre los niños y el robot Loly, además de ser una mejora tecnológica, también ayudará a los niños a aprender a diferenciar los alimentos saludables y no saludables para que puedan mejorar su dieta alimenticia.

1.3 Objetivos

1.3.1 Objetivo General

Implementar un software que permita la interacción entre un usuario y un robot social a través del reconocimiento del habla utilizando técnicas de inteligencia artificial para lograr una comunicación verbal y no verbal.

1.3.2 Objetivos Específicos

Para lograr el objetivo general planteado en la sección previa, se fijaron los siguientes objetivos específicos.

1. Reconocer la voz humana a través de técnicas de inteligencia artificial para convertir la voz a texto.

2. Generar la voz del robot social usando técnicas de inteligencia artificial relacionadas a la síntesis de voz para que se pueda comunicar de forma verbal.
3. Generar los movimientos del robot social en base a una biblioteca de eventos clasificados de acuerdo a su tipo (respuestas positivas o negativas) para lograr que la comunicación sea dinámica y coherente.

1.4 Marco teórico

El marco teórico le brinda al lector la información necesaria para entender todos los conceptos que involucran el desarrollo de este proyecto. A continuación, se explicarán los componentes que forman parte de la solución, su funcionamiento y demás términos técnicos que permitan la total comprensión del tema.

1.4.1 Robots sociales

Los robots sociales son robots dedicados a actividades básicas de ocio tales como el aprendizaje, el juego, el entretenimiento; son juguetes interactivos que poseen un componente de software que los distingue de los demás mecanismos de baja tecnología (Pinel et al., 2018).

La diferencia entre un robot social y un juguete radica en la interacción que tienen con su entorno ya que están diseñados para interactuar y comunicarse con los seres humanos.

1.4.2 Inteligencia artificial

La inteligencia artificial es la habilidad de los ordenadores para hacer actividades que normalmente requieren inteligencia humana (Barrios et al., 2018), más detalladamente es la capacidad de las máquinas para utilizar algoritmos, aprender de los datos crudos y utilizarlos para la toma de decisiones de igual forma como lo haría un humano.

Una de las aplicaciones que se le puede dar a la inteligencia artificial, es en la creación y el entrenamiento de Chatbots, los chatbots son un software informático que tiene como finalidad automatizar ciertos requerimientos del usuario, por lo tanto, mediante estos se puede ofrecer un servicio o información sin necesidad de estar en contacto con el usuario.

1.4.3 ChatBot

Es un software que permite simular una conversación para poder cumplir con la tarea que ha dispuesto el usuario, estos chatbots pueden responder diferentes tipos de instrucciones ya sean escritas o verbales. Estos tipos de chatbots son la versión más simple ya que solo pueden relacionarse con los usuarios a través de las reglas previamente definidas por los desarrolladores. (*¿Qué son los chatbots? Por qué incorporarlos en la relación con nuestros clientes - Revista Digital de la Facultad de Ciencias Administrativas y RRHH, 2018*)

1.4.4 Chatbots con Inteligencia Artificial

Los chatbots con inteligencia artificial son la versión más evolucionada, porque permiten mantener una conversación más humana con los usuarios, además de que pueden aprender a partir de las conversaciones que intercambien. Los chatbots tienen que someterse a un entrenamiento para que logren ser capaces de reconocer los requerimientos del usuario, asegurando de esta manera brindarles a las personas una atención más personalizada y fluida (*Chatbots con inteligencia artificial: la solución que necesitas - Centribal, 2020*)

1.4.5 Lematización

La lematización es un método usado por los motores de búsqueda y chatbots para analizar el significado detrás de una palabra, utiliza el contexto en el que se usa la palabra.

1.4.6 Sintetizador de Voz

La síntesis de voz es el proceso de generar una voz artificial a partir de un texto de entrada. A la aplicación dedicada a este proceso se conocen como sintetizador de voz y su calidad se mide por diversos parámetros: la naturalidad, la inteligibilidad de la voz resultante y la complejidad del procesamiento.

El primer software que permitió la síntesis de voz fue desarrollado por Noriko Umeda en 1968. Con el paso del tiempo esos sistemas fueron perfeccionándose con la ayuda de la inteligencia artificial hasta el punto de alcanzar resultados sorprendentes, mejorando la calidad del audio y logrando una mayor similitud con las características de la voz humana (Paul et al., 2020).

1.4.7 Los chatbots y el procesamiento y de lenguajes

Para que los chatbots logren comprender el lenguaje se necesita realizar un entrenamiento con los datos con los que se va a trabajar o con los que se esperan obtener resultados. Gracias a que algunos chatbots manejan un sistema de aprendizaje autónomo pueden ir aprendiendo y mejorando constantemente sus habilidades de comunicación para poder entregar respuestas correctas a los seres humanos. (*Comprendiendo la Arquitectura de los Chatbots*, 2019).

1.4.8 Los chatbot y las coincidencias de patrón

Estos chatbots manejan un sistema de reconocimiento de patrones para poder clasificar la información que recibe y así poder ofrecer una respuesta óptima al usuario.

1.4.9 Redes neuronales artificiales

Las redes neuronales permiten calcular la interacción utilizando ciertas conexiones ponderadas que se toman a partir de repeticiones mientras se entrena con los datos. Con cada una de las repeticiones se logra mejorar la precisión del mensaje de salida. (*Comprendiendo la Arquitectura de los Chatbots*, 2019)

Las oraciones que van captando se van dividiendo en palabras, pero cada una de las palabras se usan como entradas para las redes neuronales. Las conexiones ponderadas se calculan gracias a las diferentes iteraciones a través de los datos de entrenamiento miles de veces. El algoritmo va mejorando con cada interacción hasta que la respuesta es más precisa.

1.4.10 Arquitectura de chatbots

Los chatbots poseen unos dos tipos de modelos; los que logran generar respuestas desde 0 mediante el uso del aprendizaje automático, y los que usan la heurística que logran seleccionar la mejor respuesta dentro de un conjunto que ya está predefinida.

1.4.11 Modelos generativos

Son modelos avanzados, por este motivo su uso no es muy común ya que requiere el uso de algoritmos complejos, estos presentan una mayor

dificultad de desarrollarse, además de esto se requiere un mayor tiempo y esfuerzo para generar los ejemplos que se utilizarán el entrenamiento. Se basa en un aprendizaje profundo para poder crear respuestas lógicas durante una conversación, sin embargo, no se tiene la seguridad de lo que este chatbot responderá

1.4.12 Modelos basados en la recuperación

A diferencia de los modelos generativos, este presenta una arquitectura más sencilla de elaborar, aunque sus respuestas no son 100% precisas, se puede estar seguro de que el chatbot no va a enviar una respuesta inapropiada. Hoy en día es el modelo que más se usa.

1.4.13 Heurística basada en patrones

La salida que es más fácil para poder generar respuestas es lograr definir un conjunto de reglas con patrones que ya se encuentran predefinidos para que actúen de una determinada manera gracias a las reglas enmarcadas.

Para el presente proyecto usaremos el modelo heurístico, debido a que será la solución óptima, porque se entrenará nuestro modelo con palabras predefinidas para que de acuerdo a estas tenga una respuesta ya establecida. El modelo se entrenará con alimentos saludables y no saludables para de acuerdo a lo que escuche del usuario este pueda darle una respuesta idónea.

1.4.14 Reconocimiento de voz

El reconocimiento de voz es un subcampo interdisciplinario de la PNL que desarrolla metodologías y tecnologías para permitir el reconocimiento y la traducción del lenguaje hablado a texto por computadoras. Los primeros sistemas de reconocimiento de voz (1950) podían comprender números, pero no palabras, IBM Shoebox (1960) fue el primero en comprender y responder a pocas palabras en inglés.

1.4.15 Análisis del habla y respuesta

El análisis del habla se puede dividir en tres etapas:

- Voz reconocimiento y conversión a texto
- Procesamiento de texto,
- Respuesta y adopción de medidas.

Estas etapas se explican como:

En primer lugar, el habla independiente del hablante pasa por un micrófono a un paquete de procesamiento de señal digital integrado en la computadora para convertirlo en un flujo de pulsos que contienen información de voz. Se pueden utilizar instrucciones específicas para leer ingrese voz y luego la convertida en texto. Esta etapa proporciona texto de voz para su procesamiento en la siguiente etapa.

En segundo lugar, el texto resultante se divide en palabras separadas para etiquetado con etiquetas de partes gramaticales de acuerdo con su posiciones y vecinos en la sentencia. Diferentes tipos de gramática se puede utilizar en esta etapa para dividir palabras etiquetadas para formar frases. Las palabras clave pueden ser extraído de estas frases eliminando palabras no deseadas en operaciones de tintineo. Estas palabras clave se pueden comprobar y corregido si no son correctos

Finalmente, se puede construir un Chatbot para brindar la respuesta inteligente a una conversación de habla en lenguaje natural. La entrada a este Chatbot son palabras clave liberadas del procesamiento de texto de voz; la salida es la respuesta programada, que será, por ejemplo, una aplicación que se ejecuta o cualquier otra respuesta de texto o voz.

CAPÍTULO 2

2. METODOLOGÍA

2.1 Resumen del capítulo

Este capítulo muestra una breve descripción de los usuarios que usarán esta implementación final, además de los parámetros técnicos utilizados para entrenar el modelo de IA que se emplearon en este proyecto; seguido de una descripción detallada de la obtención y preprocesamiento que se realizó a los datos crudos de audio para formar un dataset útil.

En la sección de requerimientos de Hardware, se menciona las características técnicas del equipo con el que se entrenó al modelo de IA.

Después, en la sección Evaluación y Selección de Arquitecturas se detalla la arquitectura que se implementó para el modelo de inteligencia artificial y el proceso que utilizo para el reconocimiento de voz y para reproducir la respuesta de voz al usuario.

Finalmente, en la sección esquema de trabajo se puede observar un modelo conceptual de cómo funciona todo el proyecto, y también se menciona las fases que se siguieron con su respectiva explicación para tener la alineación final del proyecto

2.2 Análisis Primario

Se logró establecer 10 reuniones con el cliente al menos una vez por semana, en las cuales se validaron los avances y se plantearon soluciones a las fallas que se presentaron en estos avances.

Gracias a las reuniones se tuvo un mejor enfoque de la orientación del proyecto y sobre todo permitió definir a los usuarios principales que se beneficiarán de esta implementación.

2.3 Definición del producto

El producto por entregar es un robot social que consta de un modelo de IA entrenados, el primer modelo de síntesis de voz entrenado con 124 audios de la voz de loly proporcionados por el cliente, el segundo modelo de chatbot entrenado con un dataset de texto relacionado con la clasificación de alimentos para que el robot social sea capaz de clasificar los alimentos como saludables y no saludable; y que genere al mismo tiempo una respuesta tanto en audio como con expresiones relacionadas a lo que se está respondiendo.

A continuación, se muestra un modelo conceptual que se definió para llevar a cabo la solución del proyecto.

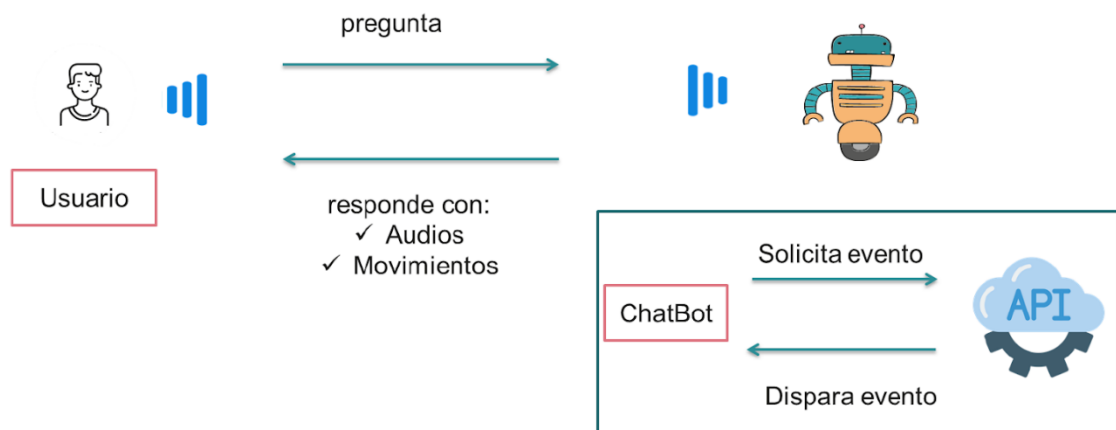


Figura 2.1 Modelo conceptual de la solución

2.4 Usuarios del sistema

Los usuarios o clientes del sistema son principalmente los niños, debido a que esta nueva implementación busca que los niños mejoren su alimentación reconociendo cuáles son los alimentos que les favorecen nutricionalmente, pero por motivos de la pandemia mundial por el COVID 19 no se pudieron realizar las pruebas en los usuarios mencionados.

2.5 Evaluación y Selección de Arquitecturas

El primer modelo de IA para la síntesis de voz utiliza una arquitectura de una red neuronal que ayuda a la síntesis para pasar de texto a voz, esta arquitectura tiene 2 componentes principales:

1. Tiene una versión cambiada de wavenet que ayuda para formar ondas en el dominio del tiempo enlazadas a los marcos del espectrograma de mel.
2. Tiene una red recurrente que le ayuda para la reproducción de features secuenciales además de ayudarlo a predecir secuencias de cuadros de espectrograma de mel y todo esto se da gracias a la información de entrada.

El segundo modelo de IA para el chatbot utiliza la arquitectura Encoder-Decoder que se base en redes neuronales recurrentes la cual es útil para resolver problemas complejos del lenguaje como lo es la traducción automática, creación de chatbot, respuesta a preguntas, etc.

2.6 Esquema de trabajo

Nuestro modelo conceptual mostrado en la Figura 1 se divide en 3 fases que se mencionan a continuación:

Fase 1: Síntesis de voz

Fase 2: Chatbot

Fase 3: Integración del módulo de síntesis de voz, chatbot y movimientos del robot.

2.6.1 Fase 1: Síntesis de voz

En la fase 1, para la síntesis de voz se necesita generar un dataset óptimo para el entrenamiento con el modelo Tacotron 2, y probar el modelo hasta lograr una voz parecida a la del robot Loly, este procedimiento se detalla en la Figura 2.

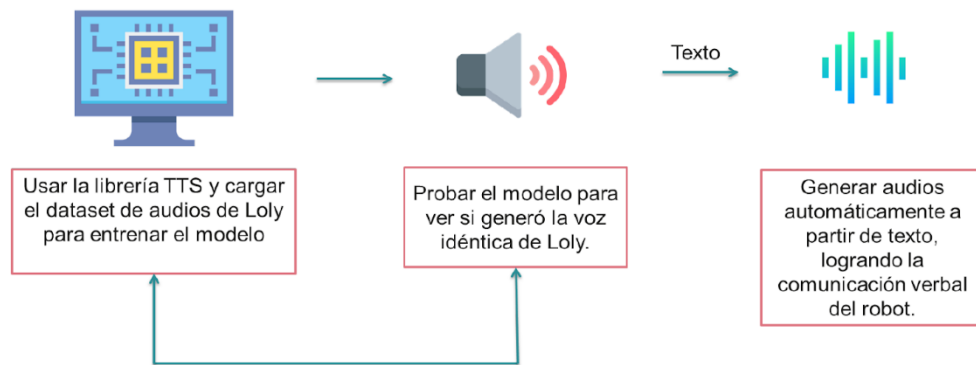


Figura 2.2 Modelo Conceptual de la Fase 1 – Síntesis de voz

2.6.1.1 Dataset

Los audios con la voz de Loly que se recibieron por parte del cliente estaban en formato .mp3 por lo que tuvieron que ser sometidos a un preprocesamiento que se detalla a continuación:

1. Se transformó los audios de formato .mp3 a .wav
2. Se modificó el sample rate de los audios de 48000 Hz a 22050 Hz.
3. Se cambió el channel de los audios de “stereo” a “mono”.
4. Se generó las transcripciones de los audios en formato LJ Speech con ayuda de la librería `speech_recognition` que usa el reconocimiento de voz de Google.

Las transcripciones se almacenaron en un archivo .csv que sigue la siguiente estructura:

1. Debe tener una etiqueta que haga referencia al archivo .wav
2. Seguido de “[”
3. Finaliza con una descripción del audio repetida 2 veces tal como se muestra en la Figura 3.

loly_fo01|Bravo lo hiciste muy bien este es un animal.|Bravo lo hiciste muy bien este es un animal.
loly_fo02|Sí. Bravo. Lo hiciste muy bien.|Sí. Bravo. Lo hiciste muy bien.
loly_fo03|Deshacerlo.|Deshacerlo.
loly_fo04|Arrastra la planta hacia el cuadrado.|Arrastra la planta hacia el cuadrado.
loly_fo05|Miren aquí hay muchas plantas. Todas las plantas son seres vivos. Las plantas necesitan el calor del :
loly_fo06|Le brindaste venir la señora de tus senos tu frutos mejores y esforzarse como los mayores un sende
loly_fo07|Bienvenidos.|Bienvenidos.
loly_fo08|Vemos juego.|Vemos juego.
loly_fo09|Himno.|Himno.
loly_fo10|Regresar.|Regresar.
loly_fo11|Salir.|Salir.
loly_fo12|Servicio reemplazada en la costa son tus hijos coraje y pujanza rebeldía orgullo y templanza fortale:
loly_fo13|Estupendo. Me ha gustado lo que acabas de hacer.|Estupendo. Me ha gustado lo que acabas de hac
loly_fo14|Vibrante lago sierra cantando aprendo hablar. Qué vibrante helados tierra mía canta un panel sobre
loly_fo15|Un pasado glorioso es ejemplo de quevedo semilla latente recuerdo de honor en la mente en tu hist
loly_fo16|Un puñado de hombres valientes de recónditos lares llegaron y al poblado.|Un puñado de hombres
loly_fo17|Hola soy loli una lorita de la especie farinosa originaria de los bosques tropicales del ecuador nació er
torear sus avances a través de una plataforma online evaluación por medio de la interacción que realice con e
loly_fo18|Qué vibrante lados tierra canta un pasado como prendas.|Qué vibrante lados tierra canta un pasadc
loly_fo19|Amiguito me llamo loli. Y yo voy a ser tu compañera en esta gran aventura mira la pantalla y presior
s. Hoy iremos al bosque. Y yo seré tu guía. Elige con qué historia quieres comenzar presionando uno de estos
loly_fo20|Bien te felicito. Completado el rompecabezas de las rocas.|Bien te felicito. Completado el rompeca

Figura 2.3 Transcripciones en formato LJ Speech

Una vez que se obtuvo el dataset adecuado para entrenar el modelo Tacotron 2 se pasó a configurar los parámetros del entrenamiento.

2.6.1.2 Preparando el config.json

Para poder entrenar nuestro modelo se necesitó contar con un archivo de configuración conocido como config.json en los cuales se encontró los hiperparámetros y distintas configuraciones de entrenamiento del modelo.

Los cambios que se realizó al archivo para que el modelo funcione son:

- Reducir el batch size
- Reducir el group batch size
- Reducir el numloaderwork

Modificar estos parámetros nos permitió entrenar más lento el modelo, pero evitó que haya falla con una tarjeta de video de poca potencia.

2.6.2 Fase 2: Chatbot

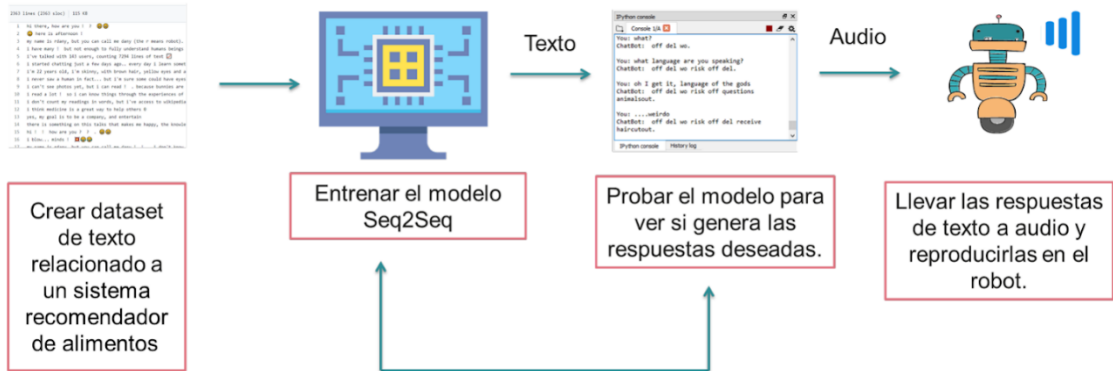


Figura 2.4 Modelo Conceptual de la Fase 2 – Chatbot

2.6.2.1 Dataset

Para la elaboración del chatbot, se creó un dataset basado en la clasificación de distintos alimentos; saludables y no saludables que nos sirvió para entrenar al modelo Seq2Seq, este modelo tiene una estructura Encoder-Decoder, lo que nos quiere decir que la entrada y la salida se dan en forma de secuencia.

Existieron datos que redujo la precisión del modelo por lo que se decidió evitar la mezcla de caracteres en mayúscula y minúscula, dígitos, caracteres especiales, etc.

Gracias al codificador se pudo convertir la entrada que es una secuencia que puede llegar a tener una longitud variable en un vector que posee longitud fija. Por otro lado, el decodificador convirtió a este vector que tiene longitud fija en una salida con longitud variable.

2.6.2.2 Lectura de datos

Una vez que se logró crear el dataset, se usó un script de Python para que los datos previamente creados se puedan leer para poderlos utilizar en el modelo.

2.6.3 Fase 3: Integración del módulo de síntesis de voz, chatbot y movimientos del robot.



Figura 2.5 Modelo Conceptual de la Fase 3 - Integración del módulo de síntesis de voz, chatbot y movimientos del robot

2.6.3.1 Biblioteca local de eventos

Se elaboró una copia de los eventos que ya tiene creado Loly para poder usarlo en una biblioteca local donde cada evento esta mapeado con una respuesta previamente definida de acuerdo con una clasificación de alimentos saludables y no saludables; dependiendo de la salida que ejecute el robot los movimientos son positivos o negativos.

2.6.3.2 Generación de movimientos del robot

Se creó distintos eventos en el robot por medio de su API, estos eventos corresponden a los diferentes tipos de respuestas que puede dar el robot.

2.6.3.3 Integración

Se elaboró un script en Python que reconoce la voz del usuario y utiliza los modelos generados anteriormente para dar una respuesta a la pregunta del usuario, y que utiliza el API del robot para generar movimientos mientras está

dando la respuesta.

2.7 Requerimientos de Hardware

Al momento de empezar a entrenar el modelo de síntesis de voz se utilizó una computadora con las siguientes características:

Tabla 2.1 Requerimientos de Hardware

Equipo	Características	Motivos
Computadora	<ul style="list-style-type: none">• 16 GB de RAM• Tarjeta gráfica de Nvidia de la serie 1050 Ti• 500 GB de almacenamiento	<ul style="list-style-type: none">• Se debe contar con una tarjeta de video, ya que, si solo se utiliza el procesador de CPU, la computadora llegar a crashear y además el entrenamiento será muy lento.• Tener espacio en el disco duro para poder almacenar los diferentes checkpoints que genere el modelo.

CAPÍTULO 3

El propósito de este capítulo es mostrar los resultados del proyecto que se obtuvieron a medida que se fue desarrollando, incluyendo las pruebas de percepción. Al finalizar el capítulo se encuentra una sección de análisis de costos, que indica los costos que se dieron a lo largo del proyecto.

3. RESULTADOS Y ANÁLISIS

3.1 Detalles del desarrollo de la solución

Luego de diseñar la solución, en el proceso de desarrollo de esta se dieron particularidades que consideramos importantes de recalcar para quienes deseen replicar nuestra experiencia. Entre los aspectos de la implementación que detallamos a continuación se encuentran:

1. El proceso de entrenamiento del Chatbot, el cual es el más importante para lograr crear el recomendador de alimentos.
2. El sintetizador de voz, el cual ayudo a que el usuario pueda obtener una respuesta clara a su interacción

3.1.1 Proceso de entrenamiento del Chatbot

El Chatbot comenzó con la creación de un dataset, esta base de datos constaba con nombres de alimentos saludables y no saludables, cada uno de estos alimentos ya se los había categorizado según la naturaleza a la que pertenecían. A pesar de que existe una infinidad de alimentos para entrenar nuestro chatbot, solo se hizo uso de 87 alimentos saludables y 24 de alimentos no saludables. Sin embargo, el modelo que usamos queda abierto a la posibilidad de poder agregar más alimentos en el futuro y así realizar una mejor clasificación de los alimentos.

Se creó un archivo `intense.json`, el cual es la parte central del chatbot ya que cuenta con diferentes etiquetas, estas ayudarán a clasificar los diferentes patrones que puede recibir el chatbot del usuario, así mismo contiene las

posibles respuestas que recibirá el usuario a las peticiones que realice y finalmente los movimientos que pueda dar el chatbot dependiendo de la situación en la que se encuentre.

Para tener una mejor clasificación, se etiquetó cada alimento dependiendo del grupo al que pertenece.

Tabla 3.1 Clasificación de alimentos

Saludables	No Saludables
Frutas	Snacks
Vegetales	Chatarra
Carbohidratos	Procesados
Proteínas	Dulces
Lácteos	
Frutos Secos	

Antes de entrenar el modelo, se usó una técnica para el preprocesamiento de datos (NLP), debido a que en este proyecto usamos tokenización para poder extraer lo que el usuario diga como una lista palabras, adicional a esto se usó la lematización para poder dejar una raíz común de las posibles expresiones que ingrese el usuario. Finalmente, también se trató de estandarizar las palabras que se iban a utilizar quitando signos de puntuación o tildes que podrían existir dentro del grupo de palabras.

Una vez que obtuvimos los datos limpios para poder realizar el entrenamiento, se armó una red MLP la cual es necesaria para el aprendizaje cognitivo del Chatbot. Se utilizó esta red debido a que el desarrollo del chatbot en esta red usa clasificaciones simples ya que nuestro chatbot lo que busca específicamente es clasificar las expresiones y agruparlas a una clase en específico y esto se solucionó por medio de este modelo.

La Tabla 3.1 nos muestra la clasificación de los alimentos, el modelo MultiLayer Perceptrón (MLP) se logró configurar de esta manera para que

cuando el usuario diga el nombre de un alimento, este lo clasifique de acuerdo con el tipo al que pertenece para poder clasificarlo como saludable y no saludable.

El entrenamiento del modelo se realizó bajo los siguientes parámetros.

Tabla 3.2 Parámetros utilizados para el entrenamiento del modelo

Tabla de Parámetros	
Epoch	100
Batch Size	4
Entrenamiento	90%
o	
Test	10%
Optimizer	SGD

La Tabla 3.2 nos muestra los parámetros que se usaron para el entrenamiento del modelo MLP, se escogieron estos valores finales ya que durante los entrenamientos que se realizaron, estos arrojaron los mejores resultados, las épocas tienen ese valor porque entre más lo aumentábamos mejor era su resultado, el batch size dependió de la capacidad que tuvo la computadora en donde se realizó el entrenamiento.

Del total de los datos que tenemos para realizar el entrenamiento solo se usa el 90% de los datos y se utiliza el otro 10% para validar si el modelo nos está ofreciendo una respuesta adecuada.

3.1.2 Proceso del sintetizador de voz

Se trató de usar la voz original de Loly, se trabajó con los 124 audios que estaban disponibles para realizar el entrenamiento, sin embargo, no se obtuvo el resultado esperado ya que no se lograba entender el audio debido a que se necesitaban más audios para realizar el entrenamiento respectivo.

Por lo tanto, para el sintetizador de voz se utilizó una librería de Python llamada pyttsx3, la cual permitió usar una voz que tenga instalado en el sistema operativo que se esté trabajando, estas voces podrían presentar ciertos cambios dependiendo de donde las estén usando.

Además de crear audios por medio de un texto ingresado, esto fue de ayuda porque se pudo llevar a audios las respuestas que nos daba el modelo previamente mencionado del chatbot. Una ventaja que da la librería pyttsx3 es que se puede modificar la voz fácilmente.

3.1.3 Proceso de los movimientos del robot

Para lograr los movimientos del robot, se crearon varios eventos en el servidor de Loly a través de su interfaz gráfica como se muestra en la Figura 3.1.

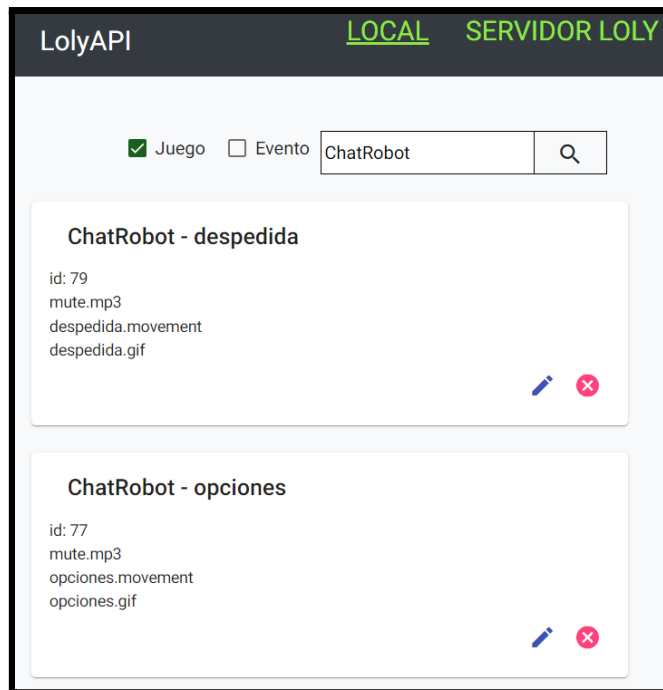


Figura 3.1 API de Loly

Los movimientos del robot se almacenaron en un archivo .csv que contenía las etiquetas previamente mencionadas en el archivo intense.json con su respectivo ID que corresponden a un movimiento en específico que ya fue previamente creado en el servidor de Loly.

3.1.4 Proceso de integración del proyecto

Cuando se obtuvo el modelo correctamente entrenado del chatbot, se creó un script en Python para integrar todas las partes del proyecto. El script consistía en generar un token de acceso para usar el API de Loly, este script logró escuchar todo lo que el usuario decía a través de un micrófono integrado con el que ya se contaba previamente.

Para lograr el reconocimiento de voz se utilizó la librería de Python `speech_recognition`, esta permitió llevar a texto lo que el usuario decía, este texto se enviaba al modelo del chatbot para que retorne una respuesta coherente con su respectiva etiqueta, esta se buscaba en la biblioteca local para obtener el movimiento coherente que debía ejecutar el robot de acuerdo con

cada respuesta. Una vez obtenido el movimiento se envía una petición http al API de Loly para que ejecute el movimiento.

3.2 Pruebas y resultados

Para comprobar que el robot haga lo que tiene que hacer, hicimos una serie de pruebas que se detallan a continuación:

3.2.1 Resultado del entrenamiento

Después del aprendizaje del modelo MLP se obtuvo la siguiente gráfica:

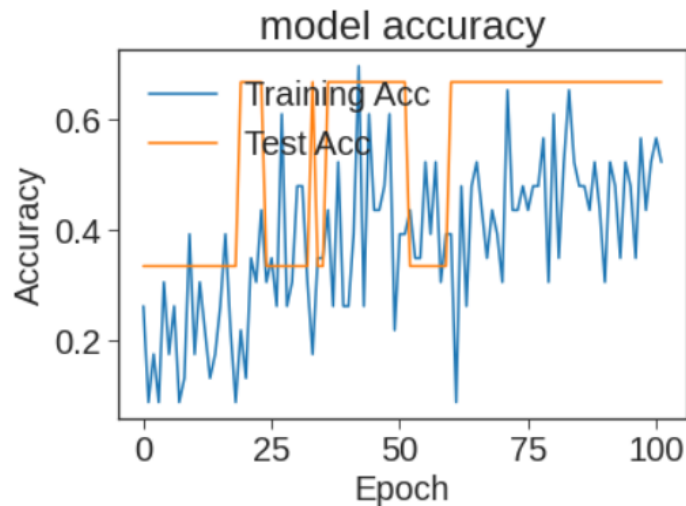


Figura 3.2 Precisión del modelo MLP

La Figura 3.2 es el resultado del último modelo de entrenamiento, e indica como ha ido aprendiendo por épocas el modelo, comenzó con una precisión de 0.2 finalizando con una precisión mayor a 0.6 para los datos de entrenamiento.

3.2.2 Resultado de encuestas

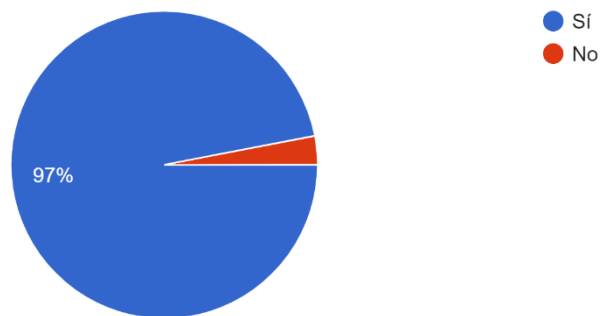
Por motivos de la pandemia mundial COVID-19 se optó por este método para evaluar la usabilidad del chatbot para distintos usuarios.

Para que los usuarios tuvieran conocimiento sobre el proyecto y lo que tenían que evaluar se preparó un video usando el chatbot en los laboratorios del CIDIS y se procedió a subirlo a la plataforma de YouTube.

Luego se preparó una encuesta que contenía el enlace al video de YouTube y además 7 preguntas que se mencionan a continuación:

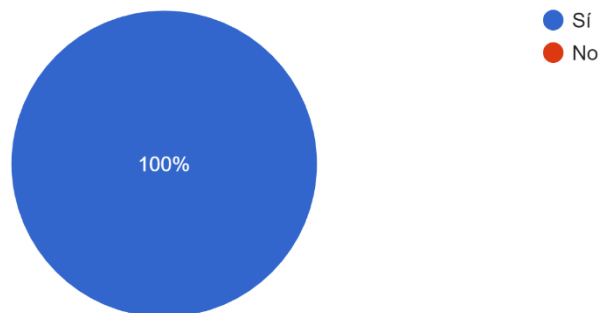
¿Entiende usted lo que el robot dice?

33 respuestas



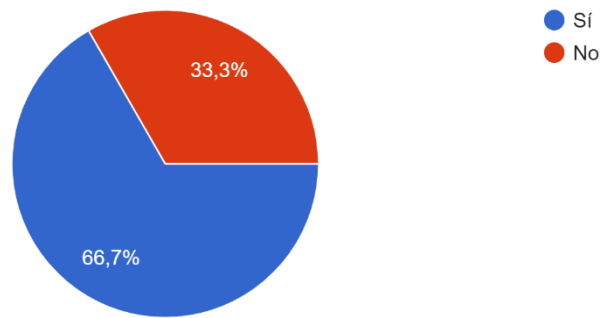
¿Cree usted que el robot cumple la función de clasificar los alimentos saludables y no saludables?

33 respuestas



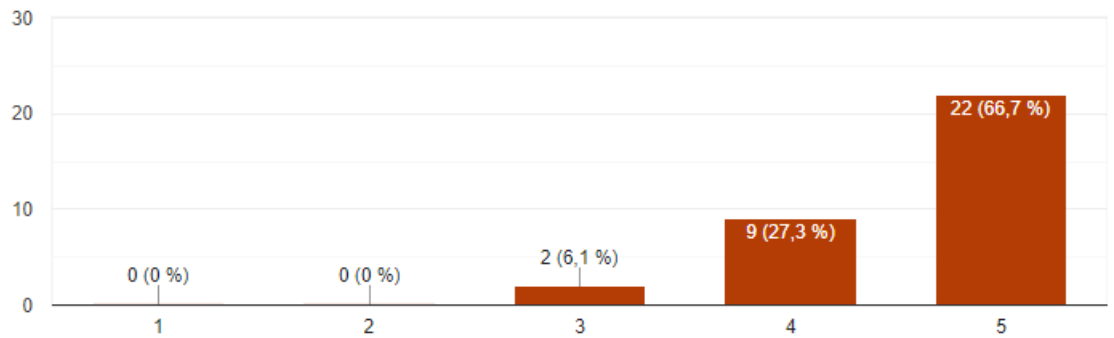
¿Cree usted que los movimientos del robot van acorde a lo que se le comunica al usuario?

33 respuestas



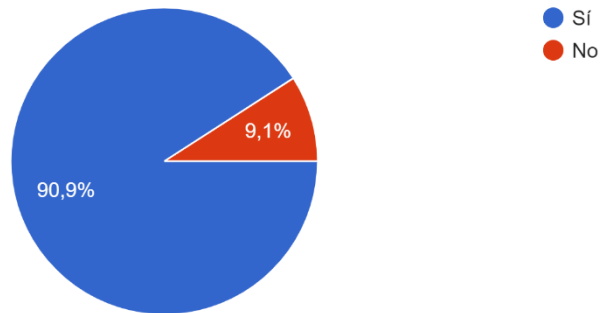
En una escala del 1 al 5, siendo 5 la más alta y 1 la más baja. ¿Cree usted que Loly da una clasificación correcta de acuerdo al alimento que se le dice?

33 respuestas



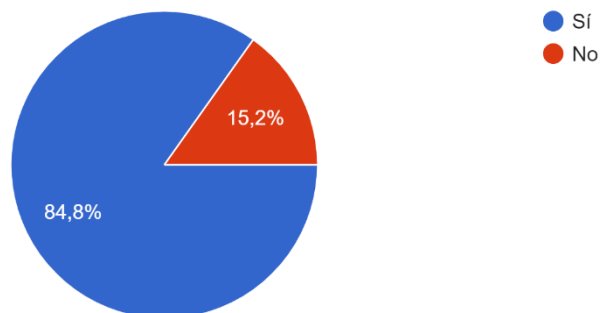
¿Cree usted que los movimientos del robot aportan una comunicación más interactiva con el usuario?

33 respuestas



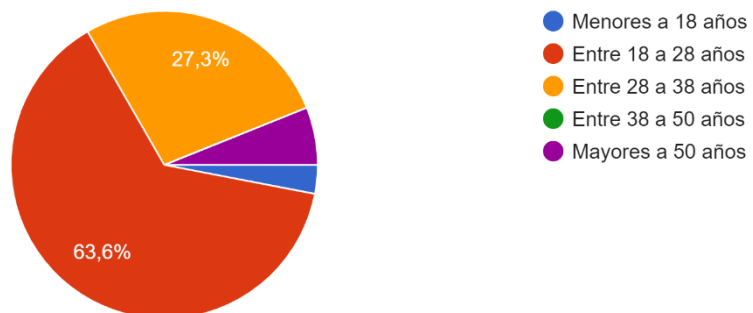
¿Usaría usted este recomendador de alimentos?

33 respuestas



¿En qué rango de edad se encuentra usted?

33 respuestas



Después de realizar las encuestas de percepción se encontró una respuesta positiva de los usuarios ante el prototipo presentado como lo muestran los gráficos, dando como resultado una viabilidad positiva del proyecto.

3.3 Análisis de costos

El entrenamiento del chatbot se dio gracias a una computadora con estas características:

- Tarjeta gráfica superior a 1050
- Procesador mínimo Intel Core i7
- 16 Gb de RAM
- Almacenamiento mínimo 500 Gb

Pero se encontraron muchas limitaciones al usar una computadora física es por eso por lo que adicional a esta se decidió usar una máquina virtual, esta computadora tuvo un costo de \$10 mensuales, esta nos ofrecía mayor poder de procesamiento gracias a su GPU. Este servicio ayuda a alcanzar los resultados esperados en un menor tiempo.

Tabla 3.3 Valores de los recursos usados

Detalle	Costo (USD)
Equipo Computacional	\$1.400,00
Desarrolladores de Software (2 personas)	\$2.000,00
Horas de entrenamiento del modelo	\$0,00
Librerías de Python	\$0,00
Total	\$3.400,00

Se tomó la decisión de usar los equipos previamente mencionados, debido a la disponibilidad que se tenía en ese momento, el costo no influyó debido a que ya se contaba con el equipo previamente descrito. Sin embargo, se realizó una valoración total del valor mínimo que se debe tener disponible para invertir en un equipo que tenga

las mismas características a las previamente citadas que es de aproximadamente \$1.400 USD.

Adicional a esto se hizo un análisis en el mercado actual del Ecuador para los programadores, los cuales están cobrando alrededor de \$1.000, debido a que este proyecto se realizó entre 2 personas se llegó a la conclusión de que la inversión total del recurso humano sería de \$2.000 dando como resultado final un proyecto con el valor de \$3.400,00 USD.

CAPÍTULO 4

Siendo este el capítulo final del documento se presentarán las conclusiones y recomendaciones que serían factibles para lograr mejorar el proyecto.

4. CONCLUSIONES Y RECOMENDACIONES

Se presenta en esta sección la importancia del proyecto realizado, así mismo se tratan de mencionar todas las dificultades que se presentaron a lo largo del desarrollo y los resultados obtenidos al momento de resolver cada dificultad.

4.1 Conclusiones

4.1.1 Técnicas

- ✓ A pesar de que se intentó usar el modelo Tacotron 2 inicialmente para la síntesis de voz, se demostró que usando este modelo de inteligencia artificial con pocos audios para el entrenamiento no se obtendrá los resultados esperados, ya que no tendremos una buena calidad en los audios.
- ✓ Se puede concluir que el dataset utilizado por el chatbot pudo haber sido mayor, sin embargo, se obtuvieron buenos resultados con los datos con los que se trabajaron, y se deja la posibilidad abierta a que este dataset continúe creciendo con el tiempo.
- ✓ El modelo de inteligencia artificial utilizado para el chatbot a diferencia de otros, tiene una arquitectura sencilla, lo que permitió que fuera manipulable ciertos parámetros consiguiendo los resultados esperados.
- ✓ Las redes MLP que se usaron como modelo de inteligencia artificial fueron suficientemente eficaces para lograr la correcta clasificación de los alimentos saludables y no saludables, debido a que estas son de gran utilidad en clasificaciones simples.

4.1.2 Usos

- ✓ El chatbot clasifica correctamente los alimentos saludables y no saludables que menciona el usuario de manera dinámica y coherente a sus movimientos.

4.2 Recomendaciones

4.2.1 Técnicas

- ✓ Se observó que en el modelo actual se puede mejorar y aumentar la calidad de la clonación de voz implementándole técnicas extras al modelo principal que es Tacotron 2, uno de estos puede ser Data augmentation, esta ayuda a la calidad de la clonación de voz cuando se tiene dataset pequeño para realizar el entrenamiento.
- ✓ Mejorar el modelo de inteligencia artificial del sintetizador de voz, para poder usar audios propios y aumentar la experiencia de usuario, utilizando el clonador de voz que se desarrolló en un módulo anterior de Loly.
- ✓ Para el chatbot se recomienda clasificar correctamente los alimentos, y tratar de evitar caracteres especiales, ya que estos podrían presentar inconvenientes al momento de realizar el entrenamiento.
- ✓ Finalmente, se podría agregar una variedad de idiomas para el chatbot para lograr crear una versatilidad y que no se limite solamente al idioma español, sino que pueda entender otros idiomas y así lograr un mayor alcance.

BIBLIOGRAFÍA

Asociación Estadounidense de Psiquiatría. (2014). Manual de diagnóstico y estadístico de los trastornos mentales (DSM-5) (5ª ed.). Madrid: Panamericana.

Organización Mundial de la Salud. (Abril de 2017). Trastornos del espectro autista.

Obtenido de Centro de prensa:

<http://www.who.int/mediacentre/factsheets/autismspectrum-disorders/es/>

ESPOL - Proyecto educativo de ESPOL, premiado en concurso mundial del BID.

(2020). Espol.edu.ec.

<https://www.espol.edu.ec/es/noticias/proyecto-educativo-espol-premiado-concurso-mundial-bid>

Meachin, J., Smjth, O., & Lvar. (n.d.). Long-Term Outcome for Children With Autism Who Received Early Intensive Behavioral Treatment.

<https://earlyautismservices.com/wp-content/uploads/2018/01/McEachin-1993.pdf>

Robots sociales que ayudan al niño autista a abrirse al mundo | Telos Fundación

Teléfonoica. (2021). Telos Fundación Teléfonoica.

<https://telos.fundaciontelefonica.com/la-cofa/robots-sociales-que-ayudan-a-abrirse-al-mundo-al-nino-autista/>

Home. (2019). Leka.io. <https://leka.io/>

PARO Therapeutic Robot. (2014). Parorobots.com. <http://www.parorobots.com/>

¿Qué son los chatbots? Por qué incorporarlos en la relación con nuestros clientes - Revista Digital de la Facultad de Ciencias Administrativas y RRHH. (2018, November 2). Revista Digital de La Facultad de Ciencias Administrativas Y RRHH.

<https://www.administracion.usmp.edu.pe/revista-digital/numero-1/que-son-los-chat-bots/>

Chatbots con inteligencia artificial: la solución que necesitas - Centribal. (2020, November 18). Centribal.

<https://centribal.com/es/chatbots-con-inteligencia-artificial-la-solucion-tecnologica-necesaria-para-el-entorno-digital/>

Pinel, V., RendónL. A., & Adrover-Roig, D. (2018). Los robots sociales como promotores de la comunicación en los Trastornos del Espectro Autista (TEA). *Letras de Hoje*, 53(1), 39. <https://doi.org/10.15448/1984-7726.2018.1.28920>

Barrios, K., López, J., Mendieta, S., Benavides, R., & Sáez, Y. (2018). Sistema de reconocimiento de voz: un enlace en la comunicación hombre-máquina. *Revista De Iniciación Científica*, 4, 92-95. <https://doi.org/10.33412/rev-ric.v4.0.1827>

mlearnere. (2021, February 16). *Learning from Audio: The Mel Scale, Mel Spectrograms, and Mel Frequency Cepstral Coefficients*. Medium; Towards Data Science.

<https://towardsdatascience.com/learning-from-audio-the-mel-scale-mel-spectrograms-and-mel-frequency-cepstral-coefficients-f5752b6324a8>

González, A. (2020, May 19). *Reconocimiento de voz: Que es, cómo funciona y programas que existen*. Ayuda Ley Protección Datos; AyudaLeyProteccionDatos. <https://ayudaleyprotecciondatos.es/2020/05/19/reconocimiento-voz/>

Paul, D., Pantazis, Y., & Stylianou, Y. (2020). Speaker Conditional WaveRNN: Towards Universal Neural Vocoder for Unseen Speaker and Recording Conditions. ArXiv.org. <https://arxiv.org/abs/2008.05289>

Elias, I., Zen, H., Shen, J., Zhang, Y., Jia, Y., Skerry-Ryan, R., & Wu, Y. (n.d.). *Parallel Tacotron 2: A Non-Autoregressive Neural TTS Model with Differentiable Duration Modeling*. Retrieved November 8, 2021, from <https://arxiv.org/pdf/2103.14574.pdf>

Abad, Balfour y Vilanova (2018). Modelo pedagógico de la escuela de innovación. Azogues, Ecuador. Editorial Universidad Nacional de Educación.