

Universidad de Oriente

Sede “Julio Antonio Mella”

Facultad de Ingeniería en Telecomunicaciones, Informática y Biomédica

Trabajo de Diploma

En opción al título de Ingeniera en Informática

**Título:** “Plataforma de creación automática de asistentes virtuales con soporte multimedia”.

**Autora:** Claudia Queipo García.

**Tutor:** Dr. C. Dionis López Ramos.

Santiago de Cuba,

“Año de la Revolución”

Resumen

Los asistentes virtuales se han vuelto cada vez más populares y se espera que su uso continúe creciendo en el futuro. Estos sistemas de inteligencia artificial son capaces de comprender y responder a las preguntas y solicitudes del usuario de manera natural, lo que los convierte en una herramienta útil para mejorar la interacción hombre máquina y el acceso a la información en diversos ámbitos como la salud, el gobierno y la economía.

En esta investigación se aborda la creación de un sistema informático para la automatización de la creación y edición de asistentes virtuales capaces de procesar información multimedia (texto, audio e imágenes) utilizando el marco de desarrollo de software RASA, la integración de herramientas de procesamiento del lenguaje natural, como Spacy y un modelo computacional de tipo gran modelo de lenguaje o modelo fundacional. El objetivo es mejorar la capacidad de los asistentes de comprender y responder a las preguntas y solicitudes del usuario. Además, se implementaron tecnologías de reconocimiento de voz para convertir audio a texto y permitir al asistente virtual interactuar con el usuario a través de la voz.

Se evaluaron y entrenaron los asistentes, obteniendo resultados prometedores en precisión y comprensión del lenguaje natural en ámbitos como la asistencia de salud para la información hospitalaria.

Palabras clave: Asistente Virtual, RASA, Reconocimiento de voz, Whisper

*ABSTRACT*

Virtual assistants have become increasingly popular, and their usage is expected to continue growing in the future. These artificial intelligence systems can understand and respond to user questions and requests naturally, making them a valuable tool for automating tasks and enhancing efficiency in various domains.

In this thesis, we address the continuation of student Ernesto Duvalon's research on creating a system that allows the development of virtual assistants using the RASA software development framework and the integration of natural language processing tools, such as Spacy, to enhance the assistants' ability to understand and respond to user questions and requests. Additionally, voice recognition technologies were implemented to convert audio to text and enable the virtual assistant to interact with users through voice.

The thesis demonstrates the possibility of creating efficient and natural virtual assistants using RASA, natural language processing, and voice recognition with Whisper. The assistants were evaluated and trained, yielding promising results in accuracy and natural language understanding. In conclusion, the feasibility of this technology is confirmed.

**Keywords**: Virtual Assistant, RASA, Voice Recognition, Whisper

índice

[INTRODUCCIÓN 6](#_Toc129849672)

[CAPITULO 1 . MARCO TEÓRICO 11](#_Toc129849673)

[1.1 Procesamiento del Lenguaje Natural 11](#_Toc129849674)

[1.2 Asistente Personal Inteligente 11](#_Toc129849675)

[1.3 Reconocimiento Automático del Habla 12](#_Toc129849676)

[1.4 Estado del arte de las herramientas empleadas en el proceso de desarrollo 13](#_Toc129849677)

[1.4.1 DialogFlow 13](#_Toc129849678)

[1.4.2 Amazon Lex 14](#_Toc129849679)

[1.4.3 IBM Watson 15](#_Toc129849680)

[1.4.4 RASA Framework 16](#_Toc129849681)

[1.4.5 Conclusiones de las herramientas y marcos de trabajo analizados 17](#_Toc129849682)

[1.5 Herramientas y Lenguajes de Programación 18](#_Toc129849683)

[Python 18](#_Toc129849684)

[Visual Studio Code 19](#_Toc129849685)

[RASA 20](#_Toc129849686)

[Whisper 20](#_Toc129849687)

[Mongo DB 20](#_Toc129849688)

[FastAPI 21](#_Toc129849689)

[Git 22](#_Toc129849690)

[1.6 Metodología de desarrollo de Software 22](#_Toc129849691)

[SCRUM 22](#_Toc129849692)

[Conclusiones del capítulo 23](#_Toc129849693)

[CAPITULO 2 . ORGANIZACIÓN Y DISEÑO 24](#_Toc129849694)

[2.1 Sprint Backlog 24](#_Toc129849695)

[2.2 Arquitectura del Sistema 25](#_Toc129849696)

[2.3 Requisitos no Funcionales 25](#_Toc129849697)

[2.3.1 Requisitos de Software 25](#_Toc129849698)

[2.3.2 Requisitos de Hardware 25](#_Toc129849699)

[2.4 Diagrama de Secuencia 26](#_Toc129849700)

[2.5 Diagrama de Actividades del Asistente 26](#_Toc129849701)

[2.6 Diseño de la interfaz 27](#_Toc129849702)

[CAPITULO 3 . IMPLEMENTACIÓN Y PRUEBA 28](#_Toc129849703)

[3.1 Instalación de las herramientas empleadas 28](#_Toc129849704)

[CONCLUSIONES Y RECOMENDACIONES 30](#_Toc129849705)

[REFERENCIAS BIBLIOGRÁFICAS 31](#_Toc129849706)

[GLOSARIO DE TÉRMINOS 35](#_Toc129849707)

[ANEXOS 36](#_Toc129849708)

INTRODUCCIÓN

Los asistentes virtuales se han convertido en una parte integral de la vida cotidiana en el siglo XXI. Estas herramientas de innovación han desempeñado un papel clave en la transformación de la interacción entre las empresas y los consumidores. Están diseñados para ser una forma de ayuda interactiva, capaz de proporcionar apoyo tanto a los usuarios como a los clientes.

Los chatbot tuvieron sus inicios en los años 60, simultáneamente con el procesamiento del lenguaje natural (NLP, Natural Language Processing), en un sistema de preguntas y respuestas (QA systems, question answering systems). A su vez, se denominan como un modelo característico de sistemas de conversaciones que ejecutan tareas específicas de interacción textual [1].

Hay tres partes principales en un sistema de chatbot: una parte de comprensión del lenguaje natural (NLU) que obtiene las intenciones del usuario, una parte de administración de diálogo que monitorea el sistema actual y el estado de la conversación, y una parte de generación de lenguaje natural que responde al usuario. A medida que pasaron los años y la tecnología se hizo más eficiente, los chatbots se volvieron más versátiles y utilizados, aunque el principal catalizador podría haber sido el impulso de Facebook por “Messenger Chatbots” [2], donde se popularizó en gran medida su uso. Cabe destacar el trabajo de varias plataformas líderes en la nube para la creación de chatbots: DialogFlow de Google, wit.ai de Facebook, Microsoft LUIS, IBM Watson Conversation, Amazon Lex, SAP Conversation AI, y otras plataformas conocidas como RASA, Botsify, Chatfuel, Manychat, Flow XO, Chatterbot, Pandorabots, Botkit y Botlytics.

A partir del 2020 con la crisis desatada por el COVID19 los chatbots han emergido como una herramienta clave para la lucha global contra la pandemia, ya que permite la identificación de personas con síntomas y muchas otras enfermedades con un alto grado de precisión, además de proporcionar respuestas precisas a la mayoría de las preguntas, por medio de fuentes creíbles de información ante la gran cantidad de personas necesitadas de consejos [3].

Actualmente han evolucionado y pueden proporcionar información significativa que ayude a las empresas a mejorar su servicio al cliente. Estos utilizan un lenguaje natural para que sea más fácil para los usuarios interactuar con ellos y obtener respuestas inmediatas. A través de la Inteligencia Artificial, los chatbots pueden seleccionar la mejor respuesta entre varias posibilidades, lo que les permite aprender y mejorar con cada conversación.

Estos también han ayudado a las empresas a reducir los costos, ya que pueden manejar diferentes tareas a la vez, como la atención al cliente, el seguimiento de pedidos, el procesamiento de pagos y la recopilación de datos. Esto permite a las empresas ahorrar tiempo y dinero al reducir los costos laborales y en general se han convertido en una parte invaluable del mundo empresarial del siglo XXI.

Es importante mencionar que existen dos tipos principales de chatbots: los de dominios específicos de conocimiento y los de dominios generales.

Los chatbots de dominios específicos de conocimiento son aquellos que han sido diseñados para trabajar en un área particular de conocimiento o campo de especialización. Estos chatbots están diseñados para comprender el lenguaje y la terminología específica de ese campo y para responder preguntas y proporcionar información relacionada con ese tema en particular. Por ejemplo, un chatbot de dominio específico de conocimiento en medicina podría responder preguntas sobre síntomas, enfermedades y tratamientos médicos.

Por otro lado, los chatbots de dominios generales son aquellos que han sido diseñados para comprender el lenguaje natural en general y responder preguntas sobre una amplia variedad de temas. Estos chatbots no están especializados en un campo particular de conocimiento y pueden responder preguntas generales sobre temas como el clima, las noticias y el entretenimiento [4].

Los chatbots de dominios específicos de conocimiento ofrecen información precisa y detallada en un campo particular, pero pueden tener dificultades para comprender preguntas fuera de su especialización. En comparación, los asistentes virtuales como ChatGpt y otros LLM son más flexibles pero presentan sesgos y errores, lo que puede afectar la precisión de los resultados y limitar su capacidad para comprender ciertos aspectos del lenguaje, además pueden ser utilizados de manera inapropiada o malintencionada, lo que puede tener graves consecuencias sociales y éticas.

Por otro lado, los chatbots de dominios generales son muy útiles cuando se trata de proporcionar respuestas rápidas y generales a una variedad de preguntas. Estos chatbots son flexibles y pueden comprender una amplia variedad de preguntas y temas. Sin embargo, debido a que no están especializados en ningún campo en particular, pueden proporcionar respuestas menos precisas y detalladas a preguntas específicas en un área de conocimiento determinada.

**Problema de investigación:**

El problema de investigación de esta tesis se centra en el desarrollo de un sistema web que permita crear asistentes virtuales que utilicen canales de texto, audio y video, con el objetivo de abordar las dificultades de atención y comunicación que enfrentan las personas ancianas, discapacitadas, con problemas visuales y niños. Se busca proporcionar respuestas precisas y adaptadas a su estado emocional, brindando una experiencia más efectiva y personalizada para estos grupos.

**Objeto de estudio:**

Desarrollo de un sistema web para la creación de asistentes virtuales que utilizan canales de texto, audio y video

**Campo de Estudio:**

Asistentes Virtuales

**Campo de Acción:**

Abarca la investigación y desarrollo de algoritmos de procesamiento de lenguaje natural, reconocimiento de voz, análisis de video y sistemas de inteligencia artificial que permitan la interacción efectiva entre las personas vulnerables y los asistentes virtuales. Además, se busca diseñar interfaces de usuario accesibles y amigables que faciliten la interacción para aquellos con limitaciones visuales o motoras.

**Objetivo General:**

Integrar la tesis de Ernesto Duvalón cambiando el algoritmo de generación de preguntas y respuestas para implementar un modelo largo del lenguaje y mejorar la experiencia del usuario al agregar funciones de audio (comunicación en lenguaje natural) y reconocimiento de emociones faciales al asistente virtual, con el objetivo de brindar respuestas adaptadas y comprender las necesidades de los usuarios .

**Objetivos Específicos:**

Análisis del estado del arte.

Estudio de la investigación de José Ernesto Duvalón

Modificación del algoritmo de generación de preguntas y respuestas

Integración de un modelo largo del lenguaje

Análisis y diseño del proceso de integración y creación de la plataforma

Estudio de tecnologías para el análisis de texto y conversión de audio a texto (p.ej., Wispher)

Implementación de la plataforma

* Pruebas y despliegue de la plataforma

**Hipótesis:**

Se desarrollará un software que permita la integración de RASA, Whisper y FaceAPI para la inclusión del canal de voz y reconocimiento de emociones en un sistema que permita crear asistentes virtuales, con el objetivo de mejorar la interacción con los usuarios y ofrecer una experiencia más personalizada y satisfactoria en la resolución de dudas y consultas. Esta herramienta permitirá una mayor facilidad a la hora de la creación de asistentes, mayor precisión en el reconocimiento de voz y una mejor interpretación de las emociones de los usuarios, lo que se traducirá en una mayor eficiencia en las respuestas ofrecidas por los asistentes virtuales creados.

**Métodos de investigación:**

* Método histórico: se aplicó al realizar el análisis de otros sistemas, tecnologías o herramientas que le pudieran dar solución a la problemática en cuestión.
* Método de análisis y síntesis: se aplicó al realizar el análisis de todo el

proceso llevado a cabo en el desarrollo del proyecto y sintetizar las ideas que fueron

surgiendo; extrayendo los elementos comunes al objeto de estudio.

**Estructura del informe:**

Capítulo 1: Muestra el marco teórico, en el cual se analizan los

diferentes aspectos teóricos y su fundamentación.

Capítulo 2: Plantea la propuesta del sistema, entre otros elementos que proporciona la metodología utilizada.

Capítulo 3: Muestra la implementación del sistema, las pruebas realizadas, los

posibles resultados a obtener.

Conclusiones, Recomendaciones, Referencias Bibliográficas y Anexos.

**Aportes de la investigación:**

* Desarrollo de asistentes virtuales con los cuales se pueda establecer una comunicación mediante canales de audio y video.
* Desarrollo de una API para el sistema de creación de asistentes virtuales lo cual permitiría su integración en plataformas web.

# MARCO TEÓRICO

La creación de asistentes virtuales se ha convertido en una tendencia cada vez más frecuente para aquellos que buscan aprovechar la tecnología para ofrecer una mejor experiencia de usuario. Estas herramientas permiten a los usuarios interactuar con la maquina a través de una conversación lo que les permite obtener información de forma sencilla y eficaz. En esta sección se examinarán algunas de las principales herramientas de construcción de asistentes virtuales, conceptos importantes del contexto que se está trabajando, por último, se explicará cual es la mejor opción para utilizarse en este proyecto y el porqué.

## Procesamiento del Lenguaje Natural

Procesamiento de Lenguaje Natural (PLN), una disciplina que fusiona elementos de la Inteligencia Artificial y la Lingüística para permitir que las computadoras comprendan y se relacionen con los idiomas humanos. El propósito fundamental del PLN es cerrar la brecha entre la comunicación humana y la interacción con computadoras, proporcionando a los usuarios la capacidad de interactuar con estas últimas mediante el uso de lenguaje natural.

El PLN se desglosa en dos componentes esenciales: la Comprensión del Lenguaje Natural y la Generación del Lenguaje Natural. La Comprensión del Lenguaje Natural implica la habilidad de la computadora para comprender el lenguaje humano, abarcando aspectos como la fonología, morfología, sintaxis, semántica y pragmática. Cabe destacar que figuras destacadas como el lingüista Noah Chomsky han realizado contribuciones significativas al estudio de la sintaxis en la lingüística.

Por otro lado, la Generación del Lenguaje Natural se enfoca en la capacidad de la computadora para producir texto coherente y significativo a partir de representaciones internas. Este componente implica la creación de frases, oraciones y párrafos en lenguaje humano [5].

## Asistente Personal Inteligente

Un Asistente Personal Inteligente es una herramienta tecnológica que utiliza la inteligencia artificial para interactuar con los usuarios de manera natural y realizar diversas tareas diarias. Estos asistentes pueden entender el contexto de las preguntas y peticiones del usuario, y realizar acciones cada vez más complejas y útiles, como organizar la agenda del usuario, controlar dispositivos inteligentes, o informar sobre noticias diarias.

Un ejemplo de asistente de voz, que es un tipo de asistente personal inteligente, es Google Assistant. Este asistente puede interactuar con el teléfono móvil y sus aplicaciones solo con comandos de voz, convirtiendo el dispositivo en el mando de un hogar inteligente. Google Assistant puede ajustar la temperatura y la iluminación, controlar los dispositivos inteligentes del usuario incluso cuando no está en casa, ofrecer información antes de que la pida y programar recordatorios para ayudarle en el momento oportuno.

Además de los asistentes de voz, existen otros tipos de asistentes virtuales que pueden ayudar al usuario en tareas cotidianas relacionadas con el manejo de dispositivos electrónicos. Estos asistentes pueden entablar una conversación natural con el usuario, entender el contexto de sus preguntas y peticiones, y llevar a cabo acciones cada vez más complejas y útiles para el usuario [6].

## Reconocimiento Automático del habla

Desde tiempo atrás ha sido de gran interés para el ser humano la posibilidad de entablar comunicación verbal con las máquinas. La ciencia ficción nos ha dado muchos ejemplos y la realidad poco a poco nos acerca a ese futuro. Para que esto sea posible es necesario que las máquinas escuchen y entiendan nuestro lenguaje y la primera parte de ese proceso es el reconocimiento automático del habla.

Los sistemas de reconocimiento automático del habla (automatic speech recognition o speech-to-text) son aquellos que traducen expresiones y enunciados de su forma hablada a texto. Este reconocimiento no conlleva la interpretación y entendimiento del significado del texto, ya que esta tarea es realizada posteriormente, en caso de que la aplicación lo requiera, mediante técnicas de Procesamiento de Lenguaje Natural.

El reconocimiento automático del habla (ASR, Automatic Speech Recognition por sus siglas en inglés) es un proceso en el que un sistema convierte la señal acústica de una expresión hablada en texto escrito. El proceso utiliza un enfoque probabilístico para asignar una puntuación a cada secuencia de palabras candidatas, evaluando su propiedad fonética, conocimiento lingüístico, contigüidad de las palabras y gramática para seleccionar la mejor opción. El esquema típico del proceso de ASR incluye preprocesamiento, extracción de rasgos característicos, decodificación y posprocesamiento del resultado. El preprocesamiento incluye la segmentación de la señal, la reducción del ruido de fondo y la identificación de las partes de silencio entre frases. La extracción de rasgos característicos implica la división de la señal acústica en ventanas de tiempo y la aplicación de transformaciones matemáticas, filtros y procesos de normalización para obtener un vector de coeficientes representativo de la señal. En la etapa de decodificación, se calcula la secuencia de palabras más probable de corresponder a la señal representada por los vectores de rasgos característicos [7].

## Estado del arte de las herramientas

Actualmente, existen diversas herramientas y plataformas para la creación de chatbots y asistentes virtuales. En este estado del arte se presentan cuatro de las más utilizadas: Rasa Framework, DialogFlow, IBM Watson y Amazon Lex.

Rasa Framework es una herramienta de código abierto basada en Python que permite la creación de chatbots y asistentes virtuales con inteligencia artificial. Esta herramienta es altamente personalizable y ofrece una gran flexibilidad en cuanto a la implementación de algoritmos de procesamiento de lenguaje natural y la integración con otros sistemas. Además, Rasa Framework cuenta con una gran comunidad de desarrolladores que contribuyen constantemente a su mejora y actualización.

### DialogFlow



*Figura 1. Logotipo de DialogFlow*

DialogFlow, propiedad de Google, es otra herramienta popular para la creación de chatbots y asistentes virtuales. Ofrece una interfaz visual amigable para la creación de flujos de conversación y una gran cantidad de integraciones con otras herramientas y plataformas. DialogFlow también cuenta con la capacidad de integrar soluciones de inteligencia artificial y aprendizaje automático de Google, como Google Cloud Speech-to-Text y Google Cloud Translate [8].

### Amazon Lex



*Figura 2. Logotipo de Amazon Lex*

Amazon Lex es un servicio de Amazon Web Services que permite a los desarrolladores construir interfaces de conversación para aplicaciones y bots. Con Amazon Lex, los desarrolladores pueden construir aplicaciones que pueden entender y responder a las solicitudes de los usuarios en lenguaje natural. Los bots de Amazon Lex pueden entender las entradas de los usuarios proporcionadas con texto o voz y conversar en lenguaje natural.

Amazon Lex permite a los desarrolladores crear bots que pueden realizar tareas automatizadas, como pedir una pizza, reservar un hotel, pedir flores, etc. Un bot de Amazon Lex cuenta con capacidades de reconocimiento automático de voz (ASR) y comprensión del lenguaje natural (NLU). Cada bot debe tener un nombre único en su cuenta.

Los bots de Amazon Lex están diseñados para una interacción de solicitud y respuesta o una conversación en streaming continua. Mediante la interacción de solicitud y respuesta, cada entrada (voz o texto) del usuario se procesa como una llamada API independiente. En una conversación en streaming, todas las entradas del usuario en diferentes turnos se procesan como una sola llamada a la API.

Amazon Lex admite varios idiomas, lo que permite a los desarrolladores crear bots que pueden interactuar con usuarios de todo el mundo.

En términos de costos, con Amazon Lex, no hay costos por anticipado ni cuotas mínimas. Solo se le cobrará por las solicitudes de texto o voz que se realicen. Los precios del pago por uso y el bajo costo de cada solicitud hacen que el servicio sea un método rentable para crear interfaces de conversación [9]

### IBM Watson



*Figura 3. Logotipo de IBM Watson*

IBM Watson Assistant es una plataforma líder en el mercado, de inteligencia artificial conversacional diseñada para ayudarte a superar la fricción del soporte tradicional y ofrecer experiencias excepcionales a prospectos, clientes y empleados.

Impulsado por modelos de lenguaje de gran envergadura (LLMs) en los que puedes confiar y una interfaz de usuario intuitiva, Watson Assistant permite a tus equipos crear agentes de voz y chatbots impulsados por inteligencia artificial que brindan soporte automatizado de autoservicio a través de todos los canales y puntos de contacto, con una integración perfecta con las herramientas que impulsan tu negocio [10].

### RASA Framework



*Figura 4. Logotipo de RASA*

Rasa es un marco de aprendizaje automático de código abierto utilizado para crear chatbots conversacionales. Se utiliza para automatizar asistentes de texto y voz.

Es capaz de mejorar las interacciones con los clientes y, por lo tanto, ayuda a las empresas. Se puede utilizar para abordar aspectos muy cruciales de las empresas, como la contratación, donde la inteligencia artificial se puede utilizar para construir chatbots que ofrezcan un excelente proceso de entrevista al proporcionar una gran experiencia al candidato. Para las empresas, los clientes que visitan su sitio web tienen una gran experiencia al conversar con una IA. La lista de implicaciones podría ser extensa.

Se compone de 2 partes:

• RASA NLU: Esta parte se encarga del lenguaje natural, para ello se nutre de los idiomas disponibles en la librería spaCy. Su misión es realizar el procesado del mensaje y su posteriormente transformarlo en datos con estructura concreta.

• RASA Core: Esta parte se encarga de la gestión del diálogo, se conecta con RASA NLU. El componente decide qué acciones tomar en cada momento [11] .

### BotPress



*Figura 5. Logotipo de Botpress*

 Botpress es una herramienta que simplifica el proceso de creación de chatbots para los desarrolladores. La plataforma reúne el código y la infraestructura necesarios para poner en marcha un chatbot y ofrece una plataforma completa y fácil de usar que incluye todas las herramientas necesarias para crea, desplegar y gestionar chatbots en tiempo récord [12].

La plataforma incluye:

* Tareas integradas de procesamiento del lenguaje natural, como reconocimiento de intenciones, corrección ortográfica, extracción de entidades y etiquetado de ranuras (y muchas otras).
* Un visual conversation studio para diseñar conversaciones y flujos de trabajo multiturno
* Un emulador y un depurador para simular conversaciones y depurar tu chatbot
* Compatibilidad con canales de mensajería populares como Slack, Telegram, MS Teams, Facebook Messenger y un chat web integrable.
* Un SDK y un editor de código para ampliar las funciones
* Herramientas posteriores a la implementación, como paneles de [análisis de chatbot](https://botpress.com/es/blog/chatbot-analytics), traspaso humano y [más](https://botpress.com/es/pricing)...

Si se realiza un análisis crítico sobre todas estas plataformas y cual de estas se ajusta a los requerimientos del proyecto en cuestión, RASA sería la mejor opción a elegir, dado que DialogFlow pese a que tiene una capa gratuita, limitaría el proyecto a la hora de que el usuario quisiera hacer peticiones sobre este, además la administración del asistente debe ser en la propia plataforma. En el caso de Amazon Lex se desecha la idea dado que los asistentes que se creen tienen que ser desplegados obligatoriamente en Amazon Web Services, lo cual estaría bien en su capa gratuita, pero esta solamente dura un año y pide tarjeta de crédito extranjera para poder registrarse en la plataforma AWS. Por parte de IBM Watson presenta desventajas en el proyecto dado que es una herramienta muy cara, la documentación no es tan abundante y este solo permite su interacción en inglés. Por último Botpress pese a no tener una gran curva de aprendizaje y tiene integración con ChatGPT, los bots creados con la plataforma se mantienen en la plataforma, es decir, no se podría hacer un host personalizado para el asistente, además solo permite la integración con ChatGPT y el precio de Botpress se basa en la cantidad de solicitudes mensuales, luego de exceder las 1000 gratuitas se debe pagar 0.005 dólares por cada una.

## Herramienta de Reconocimiento automático del Habla y Transcripción de Voz a Texto

Whisper es una herramienta de inteligencia artificial diseñada para convertir el habla en texto. Es una solución innovadora que ha revolucionado el campo de la transcripción, ya que puede manejar una amplia variedad de idiomas y ofrece un alto nivel de precisión en el reconocimiento de voz [15] ADDI Espagueti","URL":"https://codigoespagueti.com/noticias/tecnologia/que-es-como-funciona-whisper/","accessed":{"date-parts":[[2023,10,7]]},"userID":12642822,"index":5,"suppress-author":f.

Este se define como un "sistema de reconocimiento automático de voz (ASR)" y ha sido entrenado con más de 680.000 horas de datos. Este modelo puede accederse a través de la web o de forma local y es abierto a todos los usuarios [13]. Además, Whisper puede trabajar con más de 100 idiomas y cuenta con varios modelos de diferentes tamaños para su uso en servidores o computadoras con menos recursos, es clave destacar que mientras más grande sea el modelo, más consumo de recursos tendrá pero tendrá una mayor precisión.

En comparación con otras herramientas de transcripción, Whisper destaca por su capacidad para manejar acentos, ruido de fondo y lenguaje técnico. Además, ha demostrado ser especialmente efectivo para la traducción de voz a texto, superando incluso a los modelos que se especializan en el rendimiento de LibriSpeech [13].

Whisper es una herramienta valiosa para realizar transcripciones con un alto nivel de precisión. Su capacidad para manejar acentos y ruido de fondo, así como su precisión en la transcripción y su uso de forma local, lo hacen especialmente útil para este proyecto.

## Lenguajes de Pogramación utilizados

### Python

Python es un lenguaje de programación de alto nivel, interpretado y de propósito general que se destaca por su simplicidad y legibilidad, lo que lo hace especialmente adecuado para principiantes en programación. Fue creado por Guido van Rossum y se lanzó por primera vez en 1991 [14].

Python es versátil y se utiliza en una variedad de aplicaciones. Algunas de las principales aplicaciones de Python incluyen:

* **Desarrollo web**: Python se utiliza a menudo para desarrollar el back-end de un sitio web o una aplicación. Este proceso puede incluir el envío de datos hacia y desde los servidores, el procesamiento de datos y la comunicación con las bases de datos, el enrutamiento de URL y la garantía de seguridad. Python ofrece varios marcos de trabajo para el desarrollo web, los más utilizados son Django y Flask [16]
* **Desarrollo de software**: Python se implementa en el desarrollo de software desde que se lanzó y continúa siendo muy popular para ese propósito. Se utiliza para crear software en diferentes plataformas porque es compatible con muchos ámbitos de operación, sistemas operativos para computadoras y dispositivos móviles y en entornos de sistema en tiempo de ejecución [16] .
* **Ciencias de blockchain**: Python ha sido un usuario importante de la industria de blockchain, manejando desde siempre ciencias de blockchain. Su habilidad para manejar todos los aspectos de la tecnología blockchain lo convierte en un elemento primordial [16]
* **Aprendizaje automático y ciencia de datos**: Python es uno de los lenguajes de programación más populares en la ciencia de datos y el aprendizaje automático debido a su sintaxis clara y legible, y a su amplia gama de bibliotecas y marcos de trabajo especializados en estas áreas, como NumPy, Pandas, Matplotlib, Scikit-learn y TensorFlow [16]

La elección de este lenguaje para la programación del backend y de los asistentes se basa en que RASA esta construido sobre él, también este es el lenguaje con el que se desarrolla el procesamiento del lenguaje natural.

### HTML, CSS y Javascript

HTML, CSS y JavaScript son las tres tecnologías fundamentales que se utilizan para desarrollar y diseñar sitios web. Cada una de estas tecnologías tiene un propósito específico y se complementan entre sí para crear experiencias web completas y atractivas.

HTML (HyperText Markup Language) es el lenguaje de marcado estándar para documentos diseñados para ser mostrados en un navegador web. HTML proporciona la estructura básica de una página web, definiendo elementos como encabezados, párrafos, enlaces, imágenes y más. HTML se utiliza para estructurar el contenido de una página web, pero no controla la presentación visual de esa estructura [17]

CSS (Cascading Style Sheets) es un lenguaje de hojas de estilo utilizado para describir la apariencia de un documento escrito en HTML. CSS se utiliza para controlar el diseño visual, incluyendo colores, fuentes, espaciado, y más. A diferencia de HTML, que se centra en la estructura y el contenido, CSS se centra en cómo se ve y se siente esa estructura y contenido

JavaScript es un lenguaje de programación que se utiliza para hacer que las páginas web sean interactivas. JavaScript puede cambiar el contenido y el estilo de una página web, reaccionar a las acciones del usuario (como clics o movimientos del mouse), y comunicarse con el servidor para recuperar o enviar datos. JavaScript se ejecuta en el navegador del usuario, lo que significa que puede hacer que las páginas web sean dinámicas y atractivas

### Framework de frontend

**React** es una biblioteca de JavaScript para la construcción de interfaces de usuario, desarrollada por Facebook. React permite a los desarrolladores crear componentes reutilizables, lo que facilita el mantenimiento y la escalabilidad de las aplicaciones. React también utiliza un DOM virtual para mejorar el rendimiento de las aplicaciones. [18]

**Angular** es una plataforma de desarrollo para la construcción de aplicaciones web móviles y de escritorio utilizando TypeScript/JavaScript y otros lenguajes. Angular fue desarrollado por Google y es conocido por su enfoque en la creación de aplicaciones de una sola página (SPA). Angular también proporciona una estructura completa para las aplicaciones, incluyendo enrutamiento, plantillas y pruebas de utilidad [18]

**Vue** es un framework de JavaScript que se centra en la facilidad de uso y la flexibilidad. A diferencia de Angular y React, Vue no está respaldado por una empresa de alto nivel, pero ha ganado una increíble popularidad en la comunidad de código abierto. Vue es conocido por su curva de aprendizaje más sencilla y su balance entre las fortalezas de otros frameworks y menos de sus debilidades. [19]

**Svelte** es una librería de JavaScript que compila el código a JavaScript en tiempo de compilación en lugar de en tiempo de ejecución. Esto resulta en aplicaciones más pequeñas y más rápidas. Svelte es conocido por su simplicidad y su enfoque en la eficiencia [19]

Para este proyecto se elige para el front React, debido a que su enfoque en la creación de componentes reutilizables puede hacer que el desarrollo de aplicaciones sea más eficiente, también cabe resaltar que gracias al virtual DOM puede renderizar componentes extremadamente rápido superando a Angular y a Vue en este aspecto lo cual es importante debido los requisitos de este proyecto. Por último, gracias a su alta popularidad, la mayoria de las librerías existentes están disponibles para su uso en React, como las librerías de reconocimiento facial y de habla utilizadas.

### Framework para la creación del backend de el sistema.

Al elegir un framework, es importante tener en cuenta si se prefiere trabajar con una guía definida o tener más libertad en la configuración. También es importante tener en cuenta otras métricas, como la velocidad de respuesta, el soporte a la carga de peticiones y la concurrencia. Sin embargo, no existe un framework que sea el mejor, ya que existen factores subjetivos como la productividad y las mejores prácticas que hacen imposible medir y comparar de manera objetiva [20].

Para la elección, se tuvo en cuenta que el lenguaje a utilizar sería Python. Dicho esto, se realizó una investigación de cuál sería el más adecuado.

**Django**

Django es un framework de desarrollo web de alto nivel, de código abierto, escrito en Python. Es popular por su versatilidad, productividad y seguridad. Algunas de las empresas más grandes del mundo, como Instagram, Disqus, y la NASA, utilizan Django para sus proyectos [22]

Ventajas:

El Django REST Framework es un marco de trabajo popular para APIs en Python. Ofrece una arquitectura personalizable y modular, ideal para APIs simples y complejas. Viene con operaciones CRUD y un navegador incorporado para probar los puntos finales de la API. Es especialmente útil para proyectos de Machine Learning, ya que Python es el lenguaje de programación estándar para los científicos de Machine Learning. Además, Django puede manejar grandes cantidades de tráfico y proporcionar uso de API a más de 450 millones de usuarios, lo que lo hace escalable y adecuado para soluciones empresariales de gran volumen [26]

Desventajas:

Aunque es fácil de aprender, puede ser difícil de dominar debido a su gran cantidad de configuraciones y ajustes que se deben realizar. No es recomendable para proyectos pequeños debido a su intensivo uso de recursos del servidor . Además, es un framework monolítico que no proporciona mucha flexibilidad en la arquitectura del proyecto, lo que puede ser un inconveniente para este proyecto [26]

Flask

Flask es un microframework para Python que permite a los desarrolladores construir aplicaciones web de manera rápida y fácil. Fue desarrollado por Armin Ronacher, líder del Grupo Internacional de Pythonistas (POCCO), y se basa en el kit de herramientas WSGI y el motor de plantillas Jinja2. Es un framework altamente usado, respaldado por el uso de compañias como Samsung, Netflix, Uber y Airbnb lo cual es un motivo de peso para tomarlo en cuenta al elegir un framework para la creación del backend de este proyecto [21]

Ventajas:

Flask es un marco de trabajo web que es fácil de entender, lo que lo hace ideal para principiantes. Su simplicidad permite a los desarrolladores tener un control total sobre el desarrollo web, tomando el control creativo de la aplicación y el desarrollo web. Aunque es simple, Flask ofrece muchas características interesantes para usar en el marco. Además, Flask viene con un motor de plantillas que permite usar la misma interfaz de usuario para múltiples páginas. Python puede insertar variables en las plantillas. Flask también permite la prueba unitaria a través de su soporte integrado, servidor de desarrollo incorporado, depurador rápido y despacho de solicitudes RESTful. Es ligero para permitir transitar fácilmente a un marco de trabajo web con algunas extensiones [21].

Desventajas:

El marco de trabajo Flask es amigable para principiantes y aficionados, lo que puede llevar a un desarrollo de código de baja calidad. Flask maneja cada solicitud una tras otra, lo que puede ralentizar el servicio de múltiples solicitudes. La inclusión de módulos adicionales puede aumentar el riesgo de seguridad si se incluye un módulo malicioso. Aunque se pueden mitigar estos problemas utilizando un hosting especializado en Python, es crucial tener en cuenta estos factores al elegir Flask para el desarrollo web [21]

**FastAPI**

astAPI es un marco de trabajo web de alto rendimiento para Python. Su rendimiento puede compararse con NodeJS y Go, y se considera uno de los marcos de trabajo más rápidos de Python disponibles [24]. Este marco de trabajo es conocido por su simplicidad y velocidad, lo que lo hace ideal para la construcción de APIs rápidas y eficientes.

Veloz. Solo esa palabra lo define. Es (y no solo en Python, sino hablando en comparación a todos los otros lenguajes de programación) una de las herramientas más veloces para construir un backend, compitiendo con lenguajes como Go o Node.js

Algunas empresas que utilizan FastAPI incluyen a Netflix, Uber, y Microsoft [25]

Ventajas:

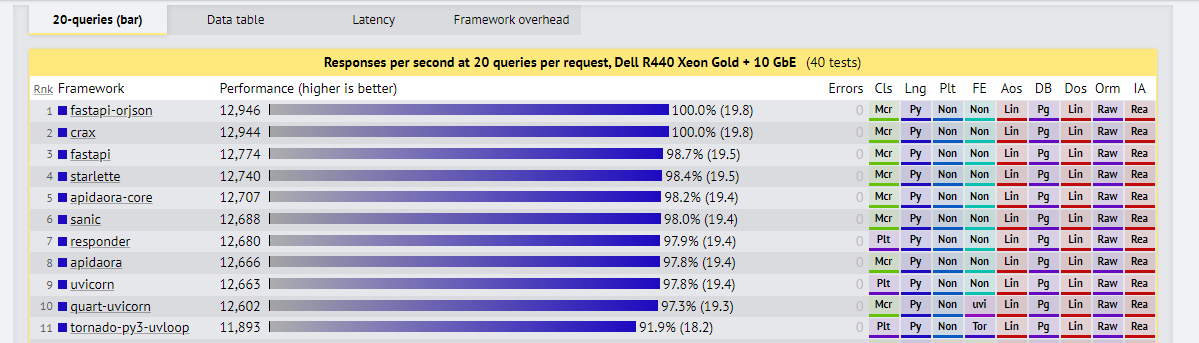
FastAPI es un framework moderno y de alto rendimiento para construir APIs con Python 3.6+. Destaca por su documentación automática, que genera una interfaz interactiva para explorar y probar la API, y por el uso de declaraciones de tipos de Python 3.6 para proporcionar un soporte de editor excelente y autocompletado. Además, FastAPI incorpora soporte para seguridad y autenticación, lo cual es esencial para una API [23].

Desventajas:

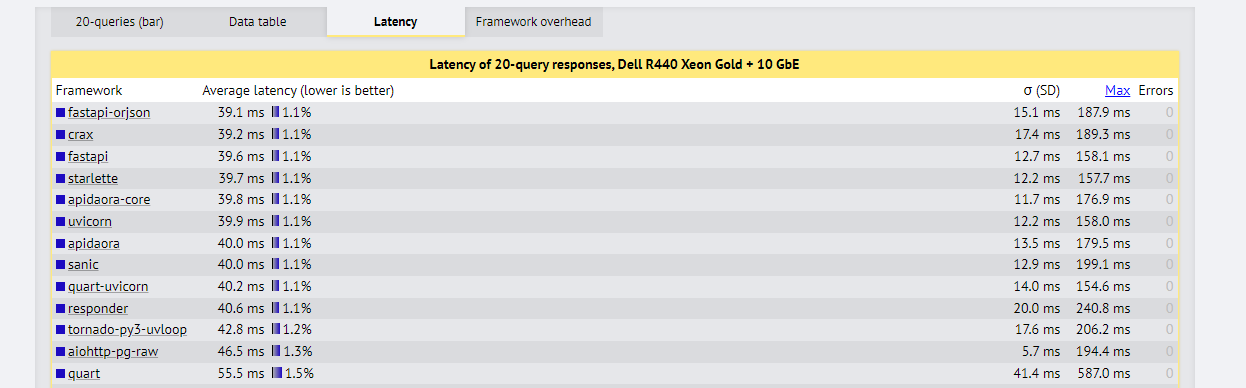
Pese a su gran robustez, es un framework relativamente nuevo, dado esto es más propenso a que no se use con frecuencia. FastAPI aún está en desarrollo y hay muchas cosas que se pueden mejorar. Aunque su comunidad está creciendo, todavía no es comparable a las comunidades de otros frameworks [23].

**Benchmark de los Frameworks**

A continuación se muestra una prueba de Benchmarking realizada en el año 2022 la cual consiste en realizar 20 consultas a un servidor por segundo [25]



Como bien se puede observar en el primer y tercer lugar se encuentra FastApi mostrando su incredible manejo de concurrencia, ligereza y escasa latencia al realizar las consultas.



En esta imagen se puede apreciar la tasa de latencia al realizar las 20 consultas las cuales posicionan a FastAPI en primer y tercer lugar al igual que en la imagen anterior.

Luego de realizar un análisis crítico sobre las ventajas y desventajas de cada uno de los frameworks y de revisar la prueba de benchmark que se plasmó previamente, se llega a la conclusión de que el framework que cumple con todos los requisitos para desarrollar el backend de la aplicación que generará los asistentes virtuales es FastAPI. Esto se debe a que la aplicación necesita tener un rendimiento óptimo en múltiples consultas en tiempo real, debe ser una solución sencilla con una mínima cantidad de configuraciones, y debe ser flexible y fácil de escalar para futuras mejoras.

### Herramienta para el reconocimiento de rostros

La biblioteca face-api.js es una herramienta de JavaScript que permite la detección y reconocimiento de rostros en el navegador. Esta biblioteca se construye sobre la API central de tensorflow.js y es compatible con la detección y reconocimiento de rostros, la identificación de expresiones faciales, la determinación de la edad y el género [27].

En términos de alternativas, existen varias API de reconocimiento facial disponibles, incluyendo las ofrecidas por Amazon, Microsoft, Google, IBM, Face++ de Megvii y Clarifai [29] Estas API ofrecen una amplia gama de funciones para la detección, verificación y reconocimiento facial.

En comparación con las alternativas, face-api.js se destaca por tener un modelo entrenado previamente, lo que puede facilitar su implementación y uso, además hay que tener en cuenta que los datos no quedan almacenados ni usados por terceros con Face Api y otro punto importante es que es una libreria gratuita a diferencia de las otras tecnologías mencionadas, que algunas pese a que tienen una face gratuita, luego tienen limitaciones o cobros por servicio.

## LLM a utilizar en el algoritmo de generación de preguntas y respuestas

**Llama 2**

Llama 2 es un modelo de lenguaje de código abierto desarrollado por Meta. Es conocido por su eficiencia y velocidad, y es una opción atractiva para ciertas instancias en las que se necesita un modelo de generación de lenguaje. A pesar de ser un modelo más pequeño que otros modelos de lenguaje como GPT-3.5 y GPT-4, Llama 2 ha demostrado un rendimiento sólido en varios benchmarks, lo que lo convierte en una opción viable para ciertas aplicaciones [28].

Llama 2 es especialmente útil para pequeñas empresas que buscan construir un chatbot, gracias a su naturaleza de código abierto y su adaptabilidad, que permite personalizaciones que se ajusten a las necesidades específicas de un negocio .

**GPT-3.5**

GPT-3.5, lanzado en marzo de 2022, es una versión actualizada de GPT-3. Es conocido por su excelente rendimiento en conversaciones complejas y su amplia compatibilidad con idiomas, lo que lo hace ideal para empresas con una base de clientes global [30].

**Falcon**

Falcon es otro modelo de lenguaje desarrollado por Meta. En comparación con Llama 2, Falcon ha mostrado un rendimiento inferior en la mayoría de los benchmarks, aunque su rendimiento específico puede variar dependiendo del caso de uso específico [28].

**GPT-4**

GPT-4 es el modelo de lenguaje más reciente de la familia GPT, desarrollado por OpenAI. Es conocido por superar a Llama 2 y GPT-3.5 en el benchmark 5-shot MMLU, lo que lo convierte en la mejor opción para tareas complejas y "misionales" que requieren un alto nivel de creatividad [28] .

**PaLM**

PaLM es una familia de modelos de lenguaje fundamentales desarrollada por Google. Según Google, PaLM 2 puede codificar, traducir y "razonar" de una manera que supera a GPT-4. Google ya utiliza PaLM 2 para alimentar 25 productos, incluyendo su asistente de IA conversacional Bard [32].

Se podría pensar que la elección de Llama se basa en que es Open Source, pero no, la realidad es que Llama frente a modelos como GPT-3.5, PALM y FALCON, tiene un rendimiento bastante adecuado y las métricas que se muestran a continuación lo demuestran.

Si comparamos a Llama 2 con Falcon, como bien se observa en la siguiente foto, en el caso de los modelos de 7Billones de parámetros Llama tiene casi el doble de puntuación en MMLU (Massive Multitask Language Understanding) métrica que se utiliza para medir cuán bien un modelo de lenguaje a gran escala (LLM) comprende el lenguaje y puede resolver problemas con el conocimiento adquirido durante su entrenamiento.

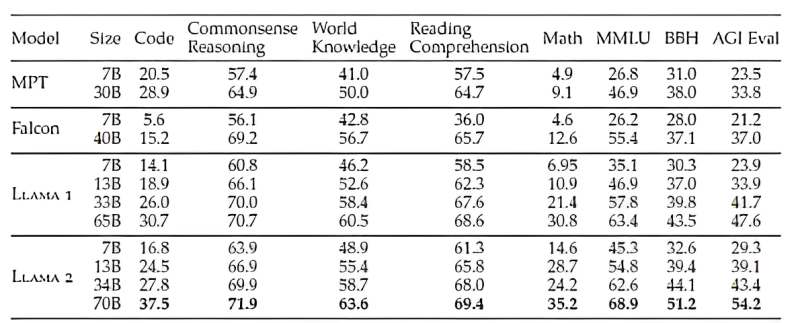


Tabla comparativa entre los modelos de Falcon y Llama [29]

En la próxima imagen, se observa como pese a que el modelo Llama 2 70 billones pese a ser el que menor puntuación tiene de sus contrincantes PALM 1, PALM 2, GPT-3.5, GPT-4, se muestra muy cercano en sus métricas a estos modelos.

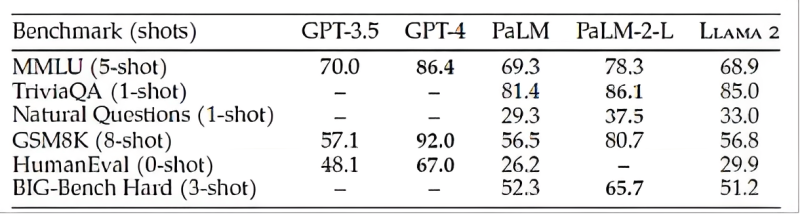


Tabla comparativa entre Llama2, PaLM, PaLM2, GPT-3.5 y GPT-4

Por último, el motivo de peso por el cual elijo a Llama 2 para el desarrollo de esta tesis es que se le puede realizar fine-tuning para lograr que el modelo realice el proceso de generación de preguntas y respuestas con una mayor precisión. Además, este proceso se podría llevar a cabo de forma local, ya que los requisitos de hardware de este modelo permiten ejecutarlo sin necesidad de una GPU. Estos dos aspectos son importantes, ya que los modelos anteriormente mencionados carecen de ellos, algunos modelos no permiten el fine-tuning por su licencia, mientras que otros son muy costosos en cuanto a recursos de hardware

### Base de Datos utilizada, MongoDB

* **Escalabilidad**: MongoDB es una base de datos distribuida que soporta el sharding automático. Esto significa que puede manejar grandes cantidades de datos y tráfico de manera eficiente, lo cual es beneficioso para un proyecto de gestión de asistentes virtuales que podría crecer con el tiempo [31].
* **Modelo de datos flexible**: MongoDB utiliza un modelo de datos basado en documentos, lo que permite almacenar datos no estructurados, semi-estructurados o incluso estructurados en la misma colección. Esto puede ser útil para tu proyecto, ya que los asistentes virtuales pueden tener diferentes tipos de preguntas y respuestas, y las necesidades de los usuarios pueden variar [33]
* **Rendimiento**: MongoDB utiliza BSON (Binary JSON) para almacenar datos, lo que facilita la lectura y escritura de datos. Además, MongoDB soporta la creación de índices en campos en documentos y subdocumentos, lo que puede acelerar las consultas de datos [34].
* **Transacciones**: MongoDB ofrece transacciones multi-documento, lo que significa que puedes realizar cambios en múltiples documentos como una sola operación atómica. Esto puede ser útil para tu proyecto si necesitas realizar operaciones que involucran a múltiples asistentes virtuales o usuarios a la vez  [34].
* **Costo**: MongoDB es una opción más económica en comparación con PostgreSQL, especialmente si tu proyecto crece y necesita más recursos. MongoDB también ofrece una opción de nube, MongoDB Atlas, que puede facilitar la administración de tu base de datos [35].
* **Facilidad de uso**: MongoDB tiene una sintaxis de consulta más intuitiva y fácil de usar en comparación con SQL, lo que puede facilitar el desarrollo de tu proyecto.

### Entorno de desarrollo

**Visual Studio Code**

Visual Studio Code (también llamado VS Code) es como la versión mini de Visual Studio. Es un editor de texto ligero y de código abierto disponible en Windows, Mac y Linux. También está disponible la versión web.

VS Code viene con soporte integrado para JavaScript, TypeScript y Node JS, pero puedes usarlo para codificar en cualquier idioma que desees. Todo lo que necesitas hacer es descargar las extensiones correspondientes.

Como es compatible con JavaScript, TypeScript y Node JS de forma predeterminada, también obtienes un depurador y detección inteligente (intellisence). Pero para obtener detección inteligente, un compilador y depuradores para otros lenguajes, debe descargar las extensiones correspondientes [36].

Será de gran utilidad para el desarrollo del proyecto gracias a su alto nivel de personalización y facilidad a la hora de el desarrollo, permitiendo la integracion de herramientas como GIT para el control de versiones.

### Sitema de Control de Versiones GIT

Git es un sistema de control de versiones distribuido que permite a los desarrolladores trabajar juntos en proyectos de software. Fue creado por Linus Torvalds, el creador de Linux, y es ampliamente utilizado en el desarrollo de software. Git rastrea los cambios en el código fuente a lo largo del tiempo, lo que permite a los desarrolladores volver a versiones anteriores del código si es necesario. También facilita la colaboración entre desarrolladores al permitirles trabajar en el mismo código base de manera simultánea [37]

### XP

La Metodología de Desarrollo de Software XP, también conocida como Programación Extrema, es una metodología ágil que se centra en la entrega de software de alta calidad en un tiempo reducido. Fue creada por Kent Beck en 1999 y se basa en la adaptabilidad y la comunicación constante entre el equipo de desarrollo y el cliente [38].

La metodología XP se compone de varias fases y roles clave:

1. Planificación: Se identifican las historias de usuario, que son tarjetas donde se detallan las funcionalidades específicas del software a desarrollar. Estas historias se descomponen en mini-versiones y se priorizan según su importancia [39].
2. Diseño: En esta fase, se realiza el diseño del software, buscando hacer un código sencillo que sea lo mínimo necesario para que funcione. Para un diseño de software orientado a objetos, se crean tarjetas CRC (Clase-Responsabilidad-Colaboración) [38]
3. Codificación: La programación se realiza en parejas en frente del mismo ordenador, asegurando que se realice un código más universal. Esto ayuda a garantizar que cualquier otro programador pueda trabajar y entender el código [39].
4. Pruebas: El código de una función se somete a una serie de pruebas unitarias continuas, con el objetivo de corregir fallas periódicamente [38].
5. Lanzamiento: Si se han seguido de forma correcta las etapas anteriores, se ha logrado estructurar un software que cumple con las expectativas del cliente [38]

La elección de esta metodología se basa en que permite una comunicación constante entre el equipo y el cliente, es altamente adaptable a los cambios como podrían ser nuevos requisitos en el proyecto, cambios de diseño y etc, tiene un alto enfoque en los resultados y por ultimo se centra en entregar constantemente software de alta calidad.

Conclusiones del capítulo

En conclusión, en este capítulo de marco referencial ofreció una base teórica y práctica sólida para el desarrollo del sistema de creación de asistentes, se han abordado diversos temas relacionados con el desarrollo del proyecto, desde el procesamiento del lenguaje natural hasta el estado del arte de las herramientas empleadas en su desarrollo. Además, se han presentado herramientas y lenguajes de programación como Python, Visual Studio Code, RASA, React JS, FastAPI y Mongo DB, entre otros, que son fundamentales para el desarrollo de este proyecto. Asimismo, se ha explicado la metodología de desarrollo de software SCRUM, que será utilizada para garantizar un proceso eficiente y efectivo en el desarrollo del proyecto.

# ORGANIZACIÓN Y DISEÑO

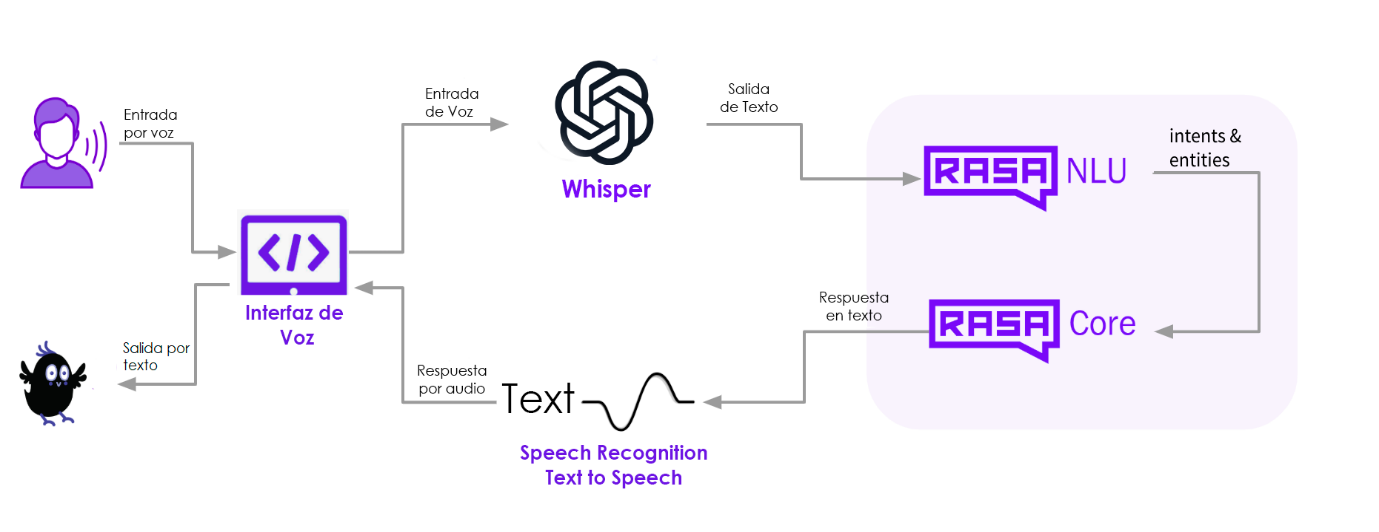
En este capítulo, se abordará el proceso de planificación y diseño del sistema, en el cual se han utilizado diferentes técnicas y herramientas para lograr un modelo claro y preciso de la aplicación. Se presentarán los diagramas que se han utilizado para el modelado de la aplicación, así como la planificación para su desarrollo. Además, se discutirán los criterios que se han tenido en cuenta para la selección de las herramientas y técnicas utilizadas y se explicarán las ventajas de su uso en el proceso de desarrollo del sistema. El objetivo de este capítulo es proporcionar una visión general y detallada del proceso de planificación y diseño del sistema y servir como guía para el desarrollo de futuros proyectos de este tipo.

## Propuesta de Sistema

Para darle solución a la problemática planteada se propone crear un sistema de gestión de asistentes virtuales basados en el framework RASA, que le permitirá al usuario introducir un documento o una cierta cantidad de información referente a su asistente, este contexto brindado por el usuario se analizará y se generará un asistente a partir del mismo dándole por supuesto en todo momento la opción de cambiar las preguntas o respuestas con las que se crea el asistente al usuario.

## Arquitectura de el asistente creado

En el siguiente diagrama, se muestra una vista física de la arquitectura de Rasa y en particular de cada asistente creado por el usuario. En él, se pueden apreciar los diferentes módulos del asistente (Interfaz de voz, Whisper, RASA\_NLU, RASA\_CORE y Text to Speech la cual hace la conversión de texto a voz de la respuesta dada por el bot al usuario).



*Diagrama de Arquitectura del Asistente*

## Requisitos Funcionales

### Frontend

RF 1- Registro de el usuario usando su correo electrónico

RF 2- Inicio de sesión en el sistema tanto con el correo electrónico, como con su cuenta de google o facebook del usuario.

RF 3- Generación de preguntas y respuestas a partir de un contexto dado.

RF 4- Generar los archivos de conocimiento para el asistente virtual.

RF 5- Entrenar asistente

RF 6- Crear, editar, visualizar, buscar, eliminar y probar asistentes virtuales

### Backend

RF 1- Registro de el usuario usando su correo electrónico

RF 2- Inicio de sesión en el sistema tanto con el correo electrónico, como con su cuenta de google o facebook del usuario.

RF 3- Generación de preguntas

RF 4- Evaluación de las preguntas generadas

RF 5- Generación de respuestas

RF 6- Generar los archivos de conocimiento para el asistente virtual.

RF 7- Entrenar asistente

RF 8- Crear, editar, eliminar y probar asistentes virtuales

***Nota: Los requisitos funcionales antes descritos se convierten en historias de usuario***

## Requisitos No Funcionales

### Software

**Cualquier sistema operativo**

Node versión 18 o superior

Python versión 3.10 o superior

MongoDB versión 5 o superior

### Hardware

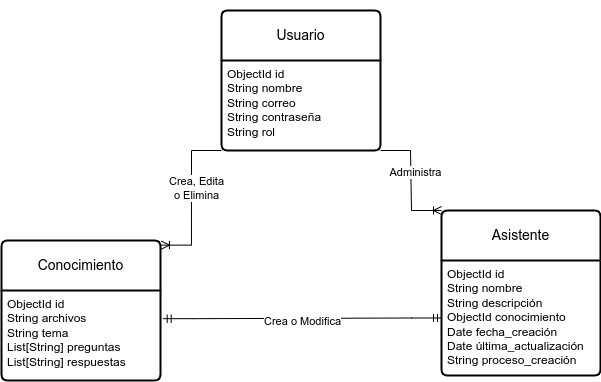
Procesador IntelCore I3 7ma gen

RAM: 8GB aprox (Debido a los requitos del modelo del lenguaje utilizado)

Almacenamiento: 10GB aprox (Debido al peso de los modelos del lenguaje y de Whisper por su caché)

## Modelo Entidad Relación (MER)

Pese a que en este proyecto se va a utilizar una base de datos NOSQL, es buena práctica crear un modelo entidad relación que refleje la información que se va a manejar en la base de datos, además esto es muy importante a la hora de realizar la documentación de el sistema.



*Modelo Entidad Relación*

## Historias de Usuario

A continuación se describen 3 de las principales historias de usuario derivadas del proceso de levantamiento de requisitos funcionales.

|  |  |
| --- | --- |
| **Historia de usuario** | |
| **Número: 1** | **Nombre de HU:** Autenticación de Usuario en el Sistema |
| **Usuario:** Cualquier usuario | |
| **Prioridad en negocio:** Alta | **Riesgo de desarrollo:** Bajo |
| **Descripción:** Apartado donde el usuario puede iniciar sesión en el sistema con su correo electrónico o sus perfiles sociales, en caso de no tener cuenta en el sistema también puede registrarse en el apartado correspondiente | |
|  | |

|  |  |
| --- | --- |
| **Historia de usuario** | |
| **Número:1** | **Nombre de HU:** Creación de asistente |
| **Usuario:** Usuario autenticado | |
| **Prioridad en negocio:** Alta | **Riesgo de desarrollo:** Alto |
| **Descripción:** Donde se rellena el formulario que permite obtener la información necesaria para analizar el documento y generar las preguntas y respuestas para posteriormente pasarlas al fomulario de la derecha donde el usuario podrá editarlas en caso de que no este satisfecho o encuentre un error en alguna. Por ultimo podrá generar los archivos de entrenamiento para su asistente o guardar los resultados para posteriormente entrenar el asistente. | |
|  | |

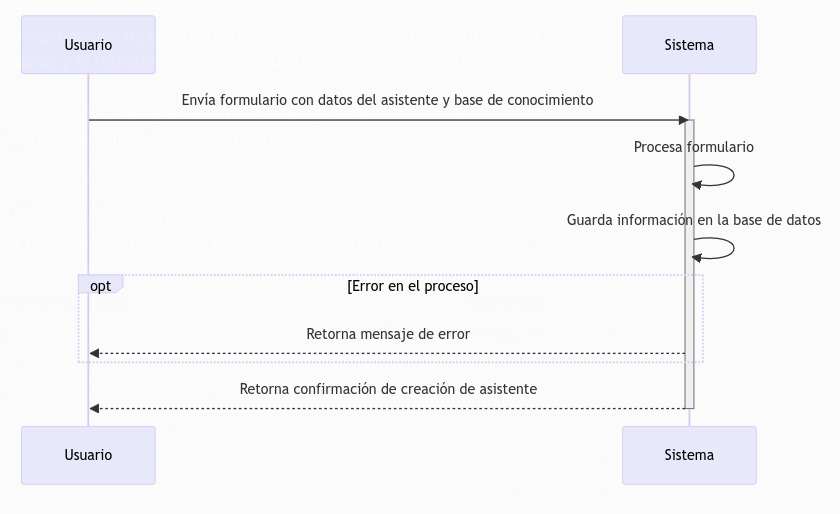
|  |  |
| --- | --- |
| **Historia de usuario** | |
| **Número:1** | **Nombre de HU:** Administrar asistentes |
| **Usuario:** Usuario autenticado | |
| **Prioridad en negocio:** Alta | **Riesgo de desarrollo:** Alto |
| **Descripción:** Donde se edita, prueba y se eliminan asistentes, también se pueden buscar asistentes por cualquiera de los parámetros que tiene la tabla | |
|  | |

## Usuarios del Sistema

|  |  |
| --- | --- |
| Actor | Tareas que realiza |
| Cliente | Crea, edita, visualiza, elimina, prueba y entrena sus asistentes. |
| Administrador | Administra la base de datos del sistema, arregla fallos e implementa nuevas funcionalidades. |

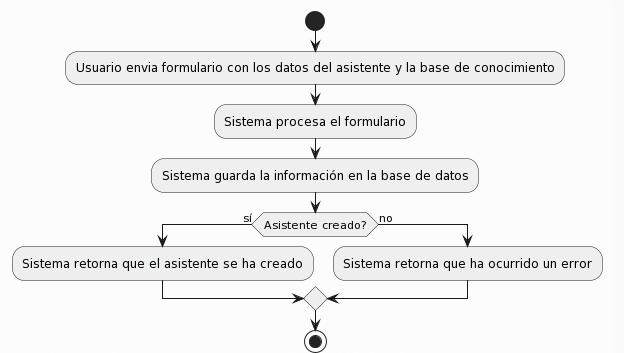
## Diagrama de Secuencia de creación de un asistente

Como bien se observa en este flujo primeramente el usuario envía los datos del asistente que desea crear y la base de conocimiento que le va a otorgar, luego el sistema procesa el formulario, guarda la información en la base de datos, si hubo un error en el proceso notifica el error al usuario y de no haber ninguno entonces confirma la creación del asistente.



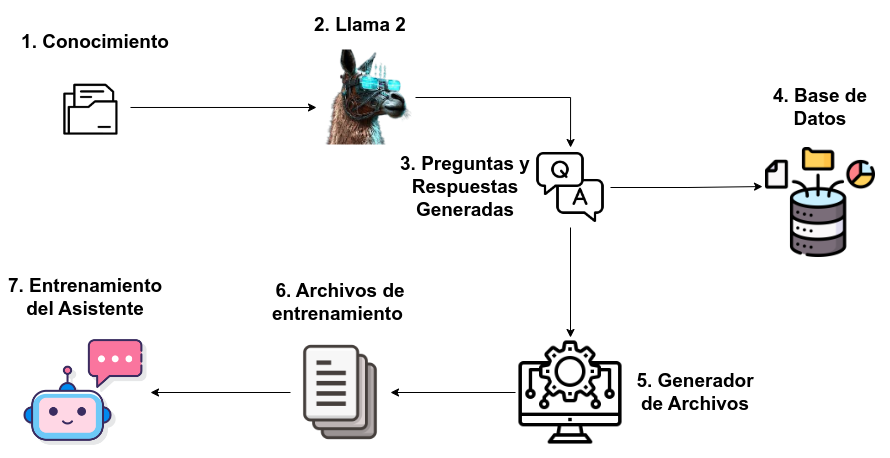
*Diagrama de Secuencia de creación de un asistente*

## Diagrama de Actividades de Creación de un Asistente

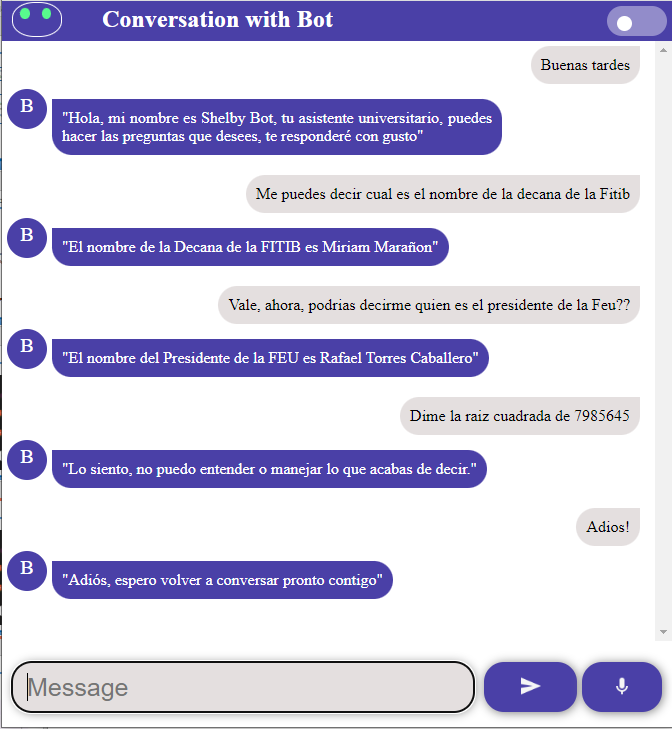


*Diagrama de Actividades de Creación de un Asistente*

## Diagrama de Arquitectura de la creación de preguntas y respuestas usando Llama 2



## Diseño de la interfaz



*Figura 9. Interfaz del asistente*

IMPLEMENTACIÓN Y PRUEBA

En este capítulo se abordará la fase de implementación y prueba de la solución propuesta. Se presentarán los detalles de la implementación del sistema, los componentes utilizados y la configuración necesaria para su funcionamiento. Además, se describirá el proceso de prueba y validación de la solución, así como los resultados obtenidos en las pruebas realizadas.

En este capítulo es fundamental para demostrar la viabilidad y eficacia de la solución propuesta, y para comprobar que cumple con los objetivos y requisitos establecidos en esta tesis.

## Instalación de las herramientas empleadas

### Python 3.8

1. Visita el sitio web oficial de Python en https://www.python.org/downloads/ y haz clic en "Download Python 3.8.x".

2. Selecciona el sistema operativo que estás utilizando (Windows, MacOS o Linux) y descarga el archivo de instalación correspondiente.

3. Una vez que se haya completado la descarga, abre el archivo de instalación.

4. Sigue las instrucciones del instalador para seleccionar las opciones de instalación que desees.

5. Asegúrate de agregar Python a la variable de entorno PATH, lo que permitirá que Python se ejecute desde la línea de comandos. En Windows, selecciona la opción "Add Python 3.8 to PATH" durante la instalación.

6. Haz clic en "Install Now" para comenzar la instalación.

7. Espera a que se complete la instalación.

8. Verifica que la instalación se haya completado correctamente abriendo la línea de comandos y escribiendo "python --version". Si se muestra la versión de Python 3.8.x, la instalación se ha completado correctamente.

### Instalación de Rasa

Vía Gratuita (Por Datos Móviles):

1. Tipear en la consola de comandos:

pip install rasa --index-url http://nexus.prod.uci.cu/repository/pypi-proxy/simple/ --trusted-host nexus.prod.uci.cu

### Instalación de Whisper

1. Instalación de Chocolatey

Abrir la PowerShell de Windows como administrador y escribir el comando:

Set-ExecutionPolicy Bypass -Scope Process -Force; [System.Net.ServicePointManager]::SecurityProtocol = [System.Net.ServicePointManager]::SecurityProtocol -bor 3072; iex ((New-Object System.Net.WebClient).DownloadString('https://community.chocolatey.org/install.ps1'))

1. Instalación de ffmpeg

Ejecutar el comando

choco install ffmpeg

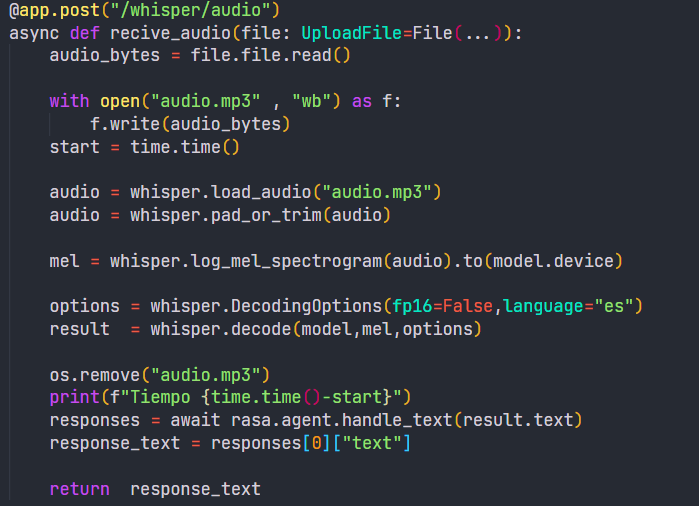
1. Por último, instalar las dependencias de Whisper

pip install torch --index-url http://nexus.prod.uci.cu/repository/pypi-proxy/simple/ --trusted-host nexus.prod.uci.cu

pip install pip install git+https://github.com/openai/whisper.git

## Algoritmos Importantes

### Algoritmo de Reconocimiento del Habla y Transcripción



*Figura 10. Algoritmo de Reconocimiento del Habla y Transcripción*

Este algoritmo es un endpoint de una API construida con FastAPI que recibe un archivo de audio en formato mp3 y devuelve el texto transcritpito de dicho audio en español.

Primero, el archivo de audio se lee y se guarda en un archivo temporal "audio.mp3". A continuación, se utiliza la librería "whisper" para cargar el audio, ajustar su duración y calcular el espectrograma de mel logarítmico.

Luego, se utiliza el modelo de reconocimiento de voz pre-entrenado para transcribir el audio en texto mediante el método "decode" de la librería "Whisper". Los resultados de la transcripción se almacenan en una variable "result".

Finalmente, se utiliza Rasa, una librería de procesamiento de lenguaje natural, para procesar el texto transcritpito y devolver una respuesta coherente. El tiempo que tarda el algoritmo en ejecutarse se imprime en la consola y se devuelve el texto transcritpito como respuesta a la solicitud del usuario.

CONCLUSIONES Y RECOMENDACIONES

**Conclusiones**

Las conclusiones deben estar en correspondencia con los objetivos planteados. Ellos deben referirse al problema de investigación planteado, de hecho la primera conclusión de un estudio es evaluar qué ocurrió con el planteamiento. Deben ser objetivas, breves, precisas y convincentes. Es importante destacar que las conclusiones no son un recuento de lo que se realizó en el trabajo. La conclusión constituye una generalización o comentario aplicado,.relativa a los resultados obtenidos

Para una mejor comprensión pueden ser numeradas y si algún caso lo requiere, pueden ser comentadas. Ejemplos

1. Se comprobó que el efecto del jitter sobre la BER no es significativo cuando la velocidad de transmisión no excede los 4 Mbps
2. Se verificó que …
3. Se demostró que …
4. Conclusión 3

**Recomendaciones**

Escriba lo que usted recomienda a otros investigadores para mejorar su trabajo o para determinar otros datos que usted no pudo lograr, sugerir qué hacer con sus resultados y aportes.

Recomendación 1

Recomendación 2

Recomendación 3

No debería haber muchas recomendaciones pues si uno recomienda muchas acciones la lectura sería que no se avanzó mucho en la investigación.

La extensión de estos aspectos no debe exceder de cinco páginas.

REFERENCIAS BIBLIOGRÁFICAS

[1]G. de Oliveira, “Redes Neurais Aplicadas No Desenvolvimento De Chatbots: Uma Análise Bibliométrica,” Jan. 2018, Accessed: Nov. 02, 2023. [Online]. Available: https://www.academia.edu/72166314/Redes\_Neurais\_Aplicadas\_No\_Desenvolvimento\_De\_Chatbots\_Uma\_An%C3%A1lise\_Bibliom%C3%A9trica  
   
   
 [2]“5 learnings from our ‘Chatbot Survey — 2017’ | by Mindbowser | Chatbots Journal.” Accessed: Nov. 02, 2023. [Online]. Available: https://chatbotsjournal.com/5-learnings-from-our-chatbot-survey-2017-72a6a4fc209c?gi=21ae27bf4d12  
   
   
 [3]“Chatbot: una propuesta viable para la atención al cliente en el centro de soporte de la UCI. | Caballero Ramírez | Revista Cubana de Ciencias Informáticas.” Accessed: Nov. 02, 2023. [Online]. Available: https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=2289  
   
   
 [4]B. Xu, “Domain-Specific Bots vs. Generic Bots,” DealerAI. Accessed: Feb. 28, 2023. [Online]. Available: https://dealerai.com/domain-specific-bots-vs-generic-bots/  
   
   
 [5]D. Khurana, A. Koli, K. Khatter, and S. Singh, “Natural language processing: state of the art, current trends and challenges,” Multimed Tools Appl, vol. 82, no. 3, pp. 3713–3744, Jan. 2023, doi: 10.1007/s11042-022-13428-4.  
   
   
 [6]“Asistente virtual: retos y posibilidades para la inteligencia artificial | Tableau.” Accessed: Nov. 02, 2023. [Online]. Available: https://www.tableau.com/data-insights/ai/ai-virtual-assistant  
   
   
 [7]D. C. Sobrino, “¿Cómo funciona el reconocimiento automático del habla?,” SoldAI. Accessed: Nov. 02, 2023. [Online]. Available: https://medium.com/soldai/c%C3%B3mo-funciona-el-reconocimiento-autom%C3%A1tico-del-habla-eb038ecfe72e  
   
   
 [8]“Dialogflow de Google cambia su modelo de pricing: ¿cuánto cuesta la IA / NLU? - Planeta Chatbot.” Accessed: Oct. 07, 2023. [Online]. Available: https://planetachatbot.com/dialogflow-google-cambia-modelo-pricing/  
   
   
 [9]“Cómo Amazon Lex cionamiento cionamiento cionamiento cionamiento cionamiento cionamiento cionamiento cionamiento - Amazon Lex V1.” Accessed: Nov. 02, 2023. [Online]. Available: https://docs.aws.amazon.com/es\_es/lex/latest/dg/how-it-works.html  
   
   
 [10]“IBM watsonx Assistant Virtual Agent.” Accessed: Nov. 02, 2023. [Online]. Available: https://www.ibm.com/products/watsonx-assistant  
   
   
 [11]AskLua, “What is RASA? — the open-source AI for building conversational chatbots.,” Medium. Accessed: Nov. 02, 2023. [Online]. Available: https://medium.com/@asklua/what-is-rasa-the-open-source-ai-for-building-conversational-chatbots-8a86bef47ec4  
   
   
 [12]“Primeros pasos con Botpress | Botpress Blog.” Accessed: Nov. 02, 2023. [Online]. Available: https://botpress.com/es/blog/getting-started-with-botpress  
   
   
 [13]“He usado Whisper para transcribir una entrevista: es la herramienta que llevaba esperando desde hace años.” Accessed: Oct. 07, 2023. [Online]. Available: https://www.xataka.com/aplicaciones/he-usado-whisper-para-transcribir-entrevista-herramienta-que-llevaba-esperando-hace-anos  
   
   
 [14]“¿Qué es Python? - Explicación del lenguaje Python - AWS,” Amazon Web Services, Inc. Accessed: Nov. 02, 2023. [Online]. Available: https://aws.amazon.com/es/what-is/python/  
   
   
 [15]“¿Qué es y cómo funciona Whisper? | Código Espagueti.” Accessed: Oct. 07, 2023. [Online]. Available: https://codigoespagueti.com/noticias/tecnologia/que-es-como-funciona-whisper/  
   
   
 [16]P. Londoño, “Qué es Python, para qué sirve y cómo se usa (+ recursos para aprender).” Accessed: Nov. 02, 2023. [Online]. Available: https://blog.hubspot.es/website/que-es-python  
   
   
 [17]“Learn Web Development Basics – HTML, CSS, and JavaScript Explained for Beginners.” Accessed: Nov. 02, 2023. [Online]. Available: https://www.freecodecamp.org/news/html-css-and-javascript-explained-for-beginners/  
   
   
 [18]“React vs Vue vs Angular vs Svelte,” DEV Community. Accessed: Nov. 02, 2023. [Online]. Available: https://dev.to/hb/react-vs-vue-vs-angular-vs-svelte-1fdm  
   
   
 [19]“Comparing React, Angular, Vue, and Svelte - Accelebrate.” Accessed: Nov. 02, 2023. [Online]. Available: https://www.accelebrate.com/blog/react-angular-vue-svelte-comparison  
   
   
 [20]U. Hernández, “Cómo elegir un framework para el backend,” CódigoFacilito. Accessed: Feb. 01, 2023. [Online]. Available: https://codigofacilito.com/articulos/elegir-framework-backend  
   
   
 [21]A. J. Condez, “Top 10 Big Companies Using Flask,” Career Karma. Accessed: Nov. 02, 2023. [Online]. Available: https://careerkarma.com/blog/companies-that-use-flask/  
   
   
 [22]“Qué es Django y por qué usarlo,” OpenWebinars.net. Accessed: Oct. 07, 2023. [Online]. Available: https://openwebinars.net/blog/que-es-django-y-por-que-usarlo/  
   
   
 [23]“Explicación de FastAPI en 5 minutos o menos - Geekflare.” Accessed: Oct. 07, 2023. [Online]. Available: https://geekflare.com/es/fastapi-explained/  
   
   
 [24]“Build And Host Fast Data Science Applications Using FastAPI | by Farhad Malik | Towards Data Science.” Accessed: Nov. 02, 2023. [Online]. Available: https://towardsdatascience.com/build-and-host-fast-data-science-applications-using-fastapi-823be8a1d6a0  
   
   
 [25]“TechEmpower Web Framework Performance Comparison,” www.techempower.com. Accessed: Feb. 01, 2023. [Online]. Available: https://www.techempower.com/benchmarks/#section=data-r20&hw=ph&test=fortune&l=zijzen-sf  
   
   
 [26]P. Team, “Advantages of Django | Disadvantages of Django,” Python Geeks. Accessed: Nov. 02, 2023. [Online]. Available: https://pythongeeks.org/advantages-disadvantages-of-django/  
   
   
 [27]“10 principales API de Reconocimiento Facial - Infoteknico.” Accessed: Oct. 07, 2023. [Online]. Available: https://www.infoteknico.com/principales-api-de-reconocimiento-facial/  
   
   
 [28]K. Luzniak, “6 main differences between Llama 2, GPT-3.5 & GPT-4,” Neoteric. Accessed: Nov. 02, 2023. [Online]. Available: https://neoteric.eu/blog/6-main-differences-between-llama2-gpt35-and-gpt4/  
   
   
 [29]S.-H. Tsang, “Brief Review — Llama 2: Open Foundation and Fine-Tuned Chat Models,” Medium. Accessed: Nov. 02, 2023. [Online]. Available: https://sh-tsang.medium.com/brief-review-llama-2-open-foundation-and-fine-tuned-chat-models-6666eb8b56b7  
   
   
 [30]“Llama 2 Vs GPT-3.5 Vs GPT-4: What, When & How To Chose,” Labellerr. Accessed: Nov. 02, 2023. [Online]. Available: https://www.labellerr.com/blog/llama2-vs-gpt-3-5-vs-gpt-4/  
   
   
 [31]“Postgres vs. MongoDB: a Complete Comparison in 2023,” Bytebase. Accessed: Nov. 02, 2023. [Online]. Available: https://www.bytebase.com/blog/postgres-vs-mongodb/  
   
   
 [32]“The AI race heats up: Google announces PaLM 2, its answer to GPT-4 | Ars Technica.” Accessed: Nov. 02, 2023. [Online]. Available: https://arstechnica.com/information-technology/2023/05/googles-top-ai-model-palm-2-hopes-to-upstage-gpt-4-in-generative-mastery/  
   
   
 [33]“MongoDB vs PostgreSQL: 15 Critical Differences.” Accessed: Nov. 02, 2023. [Online]. Available: https://kinsta.com/blog/mongodb-vs-postgresql/  
   
   
 [34]“MongoDB Vs PostgreSQL,” MongoDB. Accessed: Nov. 02, 2023. [Online]. Available: https://www.mongodb.com/compare/mongodb-postgresql  
   
   
 [35]“MongoDB vs PostgreSQL: 15 Critical Differences.” Accessed: Nov. 02, 2023. [Online]. Available: https://kinsta.com/blog/mongodb-vs-postgresql/  
   
   
 [36]“Visual Studio vs Visual Studio Code: ¿Cuál es la diferencia entre estos editores de código IDE?,” freeCodeCamp.org. Accessed: Nov. 02, 2023. [Online]. Available: https://www.freecodecamp.org/espanol/news/visual-studio-vs-visual-studio-code-cual-es-la-diferencia-entre-estos-editores-de-codigo-ide/  
   
   
 [37]“Qué es Git | Atlassian Git Tutorial.” Accessed: Nov. 02, 2023. [Online]. Available: https://www.atlassian.com/es/git/tutorials/what-is-git  
   
   
 [38]G. Mancuzo, “Metodología XP: La Mejor Vía para el Desarrollo de Software,” Blog - ComparaSoftware. Accessed: Dec. 05, 2023. [Online]. Available: https://blog.comparasoftware.com/metodologia-xp/  
   
   
 [39]“Metodología XP o Programación Extrema: ¿Qué es y cómo aplicarla?,” Gestor de proyectos online. Accessed: Dec. 05, 2023. [Online]. Available: https://www.sinnaps.com/blog-gestion-proyectos/metodologia-xp