

## FUNDAMENTOS DE INTELIGENCIA ARTIFICIAL

### INFORME DE AVANCES

#### TEMA:

Chatbot que responda a preguntas frecuentes utilizando un conjunto de documentos institucionales.

#### GRUPO DE TRABAJO:

- *Edgar David Martínez*
- *Jandri Giovanni Villavicencio*

#### PLANTEAMIENTO DEL PROBLEMA

La gestión de documentos institucionales como políticas internas, regulaciones financieras, procedimientos de seguridad y contratos, presenta un desafío fundamental en la era digital. La necesidad de garantizar un acceso eficiente y la recuperación precisa de información relevante en estos documentos es crítica para el funcionamiento efectivo y el cumplimiento normativo en el sector bancario.

El problema central se encuentra en la complejidad y la ineficiencia de los métodos tradicionales utilizados para buscar información en documentos bancarios. La navegación manual a través de documentos extensos y la búsqueda por palabras clave a menudo resultan insuficientes y propensas a errores, lo que puede llevar a incumplimientos normativos, toma de decisiones subóptimas y un aumento en los costos operativos.

La creación de un chatbot inteligente que pueda responder preguntas basadas en documentos institucionales tiene el potencial de revolucionar la forma en que se accede y utiliza la información crítica. Esto no solo agilizará los procesos internos, sino que también mejorará la satisfacción de los usuarios y garantizará el cumplimiento de regulaciones internas y externas.

Este proyecto se propone abordar estos desafíos al desarrollar un chatbot especializado que pueda comprender y responder preguntas específicas en base a documentos institucionales. Durante su desarrollo, se analizarán desafíos técnicos, de seguridad y de adaptación al contexto de cada institución, con el objetivo de identificar las mejores prácticas y estrategias para transformar la forma en que las instituciones acceden y utilizan la información contenida en sus documentos institucionales.

#### OBJETIVOS (Establecer los objetivos- ¿Qué, Cómo, Para qué?)

##### *General:*

Desarrollar un chatbot especializado capaz de comprender y responder preguntas específicas basadas en documentos institucionales, como políticas internas, regulaciones financieras, procedimientos de seguridad y contratos, mediante la aplicación de técnicas de procesamiento de lenguaje natural (NLP) y aprendizaje automático. Esto tiene como finalidad mejorar la eficiencia en la gestión de documentos institucionales en el sector bancario, optimizar la toma de decisiones, reducir los costos operativos y garantizar el cumplimiento de regulaciones, contribuyendo así al funcionamiento efectivo de las instituciones bancarias en la era digital.

### **Específicos:**

El primer objetivo específico es desarrollar un chatbot capaz de comprender y responder con precisión a consultas relacionadas con políticas internas, regulaciones financieras y procedimientos de seguridad específicos de un banco. Esto permitirá una gestión más eficiente de la información crítica, mejorará la toma de decisiones internas y garantizará el cumplimiento normativo en el contexto bancario.

El segundo objetivo específico es personalizar el chatbot para adaptarse a las necesidades y el entorno de un banco en particular, asegurando que el chatbot pueda comprender y responder de manera efectiva a las consultas específicas de esa institución. Esta adaptabilidad mejorará la eficiencia en la gestión de documentos institucionales y contribuirá al cumplimiento de regulaciones internas y externas, fortaleciendo así la operación segura y eficaz del banco.

### **MARCO TEÓRICO (Conceptos que están relacionados con el tema y contribuyen para el desarrollo del mismo, además estado del arte presentar al menos dos trabajos relacionados)**

Dada una investigación en línea y tras discutir con el equipo del proyecto hemos definido los siguientes conceptos y estado de arte relacionados con el proyecto:

#### **Conceptos relacionados**

- **Procesamiento de Lenguaje Natural (NLP):** Es una rama de la inteligencia artificial que se enfoca en la interacción entre las computadoras y el lenguaje humano. En este proyecto, el NLP se utilizará para que el chatbot pueda comprender y generar respuestas basadas en el lenguaje humano presente en los documentos institucionales.
- **Aprendizaje Automático (Machine Learning):** Se puede definir como una técnica que permite que las computadoras mejoren su rendimiento en tareas específicas a través de la experiencia y la retroalimentación. Se aplicará para entrenar al chatbot en la comprensión de documentos y la generación de respuestas precisas.
- **Regulaciones Financieras:** Son las normativas establecidas por las autoridades financieras que los bancos deben cumplir. Comprender estas regulaciones es crucial para el funcionamiento legal y ético de un banco. El chatbot debe estar preparado para proporcionar información precisa sobre estas regulaciones cuando se le consulte.

#### **Estado del arte**

- **Chatbots en el Sector Bancario:** Hoy en día varios bancos han adoptado chatbots para mejorar la atención al cliente y la eficiencia operativa. Por ejemplo, el banco HSBC implementó un chatbot llamado "Amy" para ayudar a sus empleados a buscar información en documentos internos. Este es solo uno de los tantos casos que muestran cómo los chatbots pueden agilizar la recuperación de datos en un entorno bancario.
- **Proyectos de NLP en el Sector Bancario:** Actualmente la aplicación de NLP en el sector bancario es una tendencia en crecimiento. Investigaciones previas han demostrado cómo el NLP puede utilizarse para automatizar la extracción de información crítica de documentos financieros y contratos, lo que contribuye a la eficiencia y la toma de decisiones precisas en el sector.

- *Chatbots para Servicios Bancarios Personalizados:* Se sabe que actualmente algunos bancos han implementado chatbots que no solo brindan respuestas a preguntas frecuentes, sino que también pueden personalizar la experiencia del cliente. Tal es el caso del chatbot de Bank of America, llamado "Erica", demuestran cómo los chatbots pueden brindar asesoramiento financiero y realizar transacciones personalizadas para los clientes, lo que mejora la satisfacción del usuario y la retención de clientes en el sector bancario.

### PROTOCOLO CONVERSACIONAL (Establecer un protocolo conversacional y validarlo)

El protocolo conversacional para este caso debe de ser un protocolo amigable el cual pueda responder las preguntas y cuestionamientos de manera precisa, al tratarse de una institución financiera donde los datos y los procedimientos a seguir son de carácter sensible, el protocolo debe de ser lo más optimo en cuanto a respuestas precisas y amigables con el usuario. Por lo cual proponemos la siguiente metodología de protocolo

#### *Saludo y bienvenida al usuario:*

- Los chatbots deben iniciar la conversación con un saludo y un saludo amistosos.
- Ejemplo: "¡Hola! ¡Bienvenido a Ayuda Bancaria! ¿Cómo puedo ayudarte hoy? "

#### *Preguntas base o referentes constantes:*

- Definir preguntas establecidas previamente por ser protocolos utilizados con frecuencia.
- Ejemplo: Inquietudes como ("Consultar saldo de cuenta.", "Ayuda con tarjetas de crédito.")

#### *Interacción Formal y Cálida:*

- Chatbot: Utilizará un lenguaje formal pero amigable para comunicarse con los usuarios.
- Ejemplo: "Por supuesto, con gusto te ayudaré a consultar tu saldo. Por favor, proporciona tu número de cuenta."

#### *Debe identificar las necesidades:*

- Los chatbots deben preguntar a los usuarios sobre sus necesidades específicas, como "¿Con qué área de la banca necesita ayuda o información?"

#### *Detección y registro de intenciones:*

- Utiliza procesamiento del lenguaje natural (NLP) para comprender la intención del usuario.
- Por ejemplo: "¿Está buscando información sobre cómo abrir una cuenta, transferir dinero, solicitar un préstamo u otros servicios bancarios?"

#### *Consulta de base de datos de los protocolos a seguir:*

- El chatbot busca en su base de datos de protocolos y procedimientos bancarios para encontrar información relevante.
- Ejemplo: "Déjame encontrar el programa que necesito..."

#### *Dar la respuesta más optima:*

- Según la intención del usuario y la información obtenida del documento de protocolo, el chatbot genera la mejor respuesta.
- Ejemplo: "Estos son los pasos para abrir una cuenta de ahorros en nuestro banco: [proporcione pasos detallados]".

### Preguntas de comprobación:

- En caso de seguir con preguntas, se debe de especificar de manera clara como continuar o detallar el procedimiento para los protocolos necesarios que han sido consultados.
- Por ejemplo, ¿Necesitas ayuda adicional en el procedimiento planteado?

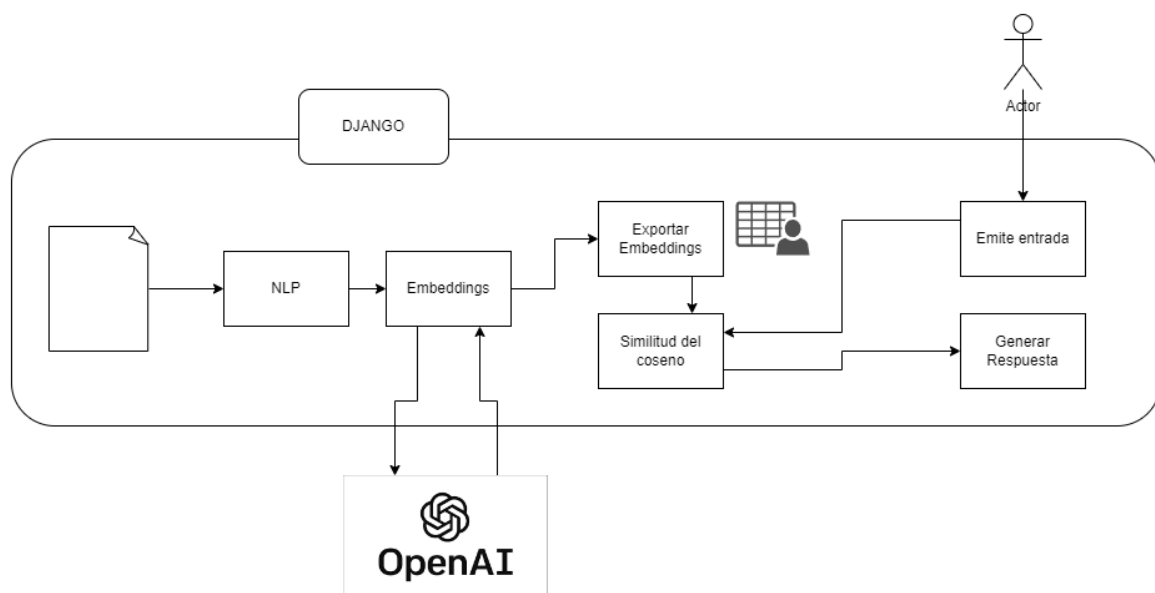
PLATAFORMA SELECCIONADA (Analizar, evaluar e identificar la plataforma que permita desarrollar el chatbot)

### API OPENAI

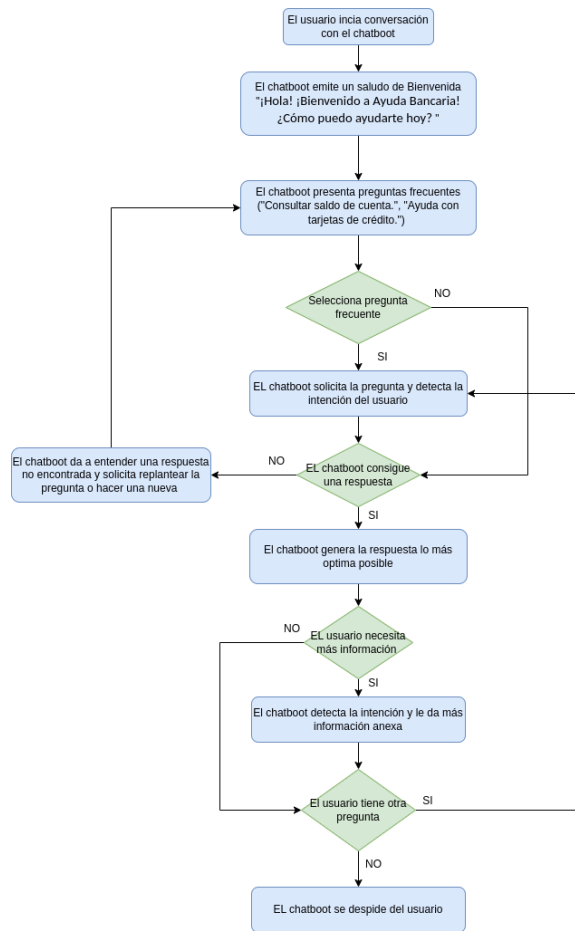
Además de ello también se optó por hacer uso del API de OpenAI, esta es un API la cual puede ser capaz de usarse en sistemas externos y ya proporciona librerías y mecanismos de integración de esta API con los proyectos, en este caso se ha optado por hacer el uso de Embeddings las cuales son una de las especiales que proporciona el API de OpenAI. Además de esto presenta otras características como:

- Generación de Lenguaje: El API permite generar texto coherente y relevante basado en las instrucciones dadas.
- Límites de Tokens: Las solicitudes al API están limitadas por el número de tokens. Cada palabra o símbolo cuenta como un token. Si una solicitud excede el límite permitido, se aplican tarifas adicionales.
- Modelos Preentrenados: GPT-3 es un modelo preentrenado, lo que significa que ha sido entrenado en grandes cantidades de datos antes de ser puesto a disposición a través del API.
- Acceso a Modelos Específicos: OpenAI puede proporcionar acceso a modelos específicos, como "davinci" o "curie", con diferentes capacidades y costos asociados.
- Usos Diversos: El API puede ser utilizado para una amplia variedad de aplicaciones, desde tareas de procesamiento de lenguaje natural hasta la creación de chatbots y asistentes virtuales.

ARQUITECTURA (Definir la arquitectura preliminar y explicar cada uno de los componentes)



### Flujo de procesamiento del chatboot



DESARROLLAR EL CHATBOT (Desarrollar el chatbot en integrar en una página web para que responda a las consultas de los usuarios)

### API OPENAI

Para el desarrollo del chatboot lo que se realizo fue primero la implementación de pruebas de lectura de archivos PDF con las librerías de PyPDF2 y Spacy para procesamiento de lenguaje natural. Con esto sintetizamos el contenido de los PDFs en ideas centralizadas para poder entrenar el modelo de IA del chatbot

```
In [1]: import PyPDF2
import spacy
#Para mayor nivel de detalle en el procesamiento se usar NLTK y sklearn

nlp = spacy.load('es_core_news_sm')

def extraer_ideas_principales(ruta_pdf):
    with open(ruta_pdf, 'rb') as file:
        pdf_reader = PyPDF2.PdfReader(file)
        text = ''
        for page_num in range(len(pdf_reader.pages)):
            text += pdf_reader.pages[page_num].extract_text()

        # Procesamiento de lenguaje natural con spaCy
        doc = nlp(text)

        # Extraer ideas principales
        ideas_principales = [sent.text for sent in doc.sents]

        return ideas_principales

# Llama a la función con la ruta de tu PDF
ruta_pdf = '/home/edmartinez/Documents/UTPL/Septimo Ciclo/Inteligencia Artificial/386670spa.pdf'
ideas = extraer_ideas_principales(ruta_pdf)

# Imprime o utiliza las ideas principales según tus necesidades
print(ideas)

In [3]: for idea in ideas:
        print(idea)
```

```
n      En 2017, un equipo de psicólogos y expertos en IA de la
Universidad de Stanford (EE.UU.) desarrolló WoeBot, un
chatbot de terapia que ayuda a los usuarios a monitorear
su estado emocional y adquirir conciencia de sí
mismos.
Este chatbot está disponible para cualquier
persona pero está especialmente dirigido a adultos
jóvenes en universidades y escuelas de posgrado.
Es un
servicio gratuito y disponible en inglés (Eve, 2020)3.

n
La Universidad Multimedia de Malasia desarrolló
ChatWithMe, un chatbot que puede identificar las
emociones del usuario y si este podría estar en riesgo
de cometer suicidio.
Además, puede orientarlo hacia
profesionales de la salud mental. (Universidad Multimedia
```

Ideas principales extraídas del PDF de referencia

### Código en python para generar un chatboot en consola que aprenda de un PDF

```
1  # chatbot.py
2  import PyPDF2
3  import spacy
4  from chatterbot import ChatBot
5  from chatterbot.trainers import ChatterBotCorpusTrainer
6  #Para mayor nivel de detalle en el procesamiento se usar NLTK y sklearn
7  nlp = spacy.load('es_core_news_sm')
8
9
10 def extraer_ideas_principales(ruta_pdf):
11     with open(ruta_pdf, 'rb') as file:
12         pdf_reader = PyPDF2.PdfReader(file)
13         text = ''
14         for page_num in range(len(pdf_reader.pages)):
15             text += pdf_reader.pages[page_num].extract_text()
16
17     # Procesamiento de lenguaje natural con spaCy
18     doc = nlp(text)
19
20     # Extraer ideas principales
21     ideas_principales = [sent.text for sent in doc.sents]
22
23     return ideas_principales
24
25 # Llama a la función con la ruta de tu PDF
26 ruta_pdf = '/home/edmartinez/Documents/UTPL/Septimo Ciclo/Inteligencia Artificial/386670spa.pdf'
27 ideas = extraer_ideas_principales(ruta_pdf)
28
29 print("Se extrajeron las ideas")
30
31 chatbot = ChatBot('MiChatBot', tagger_language=nlp)
32 trainer = ChatterBotCorpusTrainer(chatbot)
33
34 print("Se creo el chatboot")
35
36 trainer.train(ideas)
37 # Añadimos más entrenamiento con los corpus predeterminados de ChatterBot
38 trainer.train("chatterbot.corpus.spanish")
39
40 print("Se entreno el chatboot")
41
42 # Obtener respuestas del chatbot
43 while True:
44     try:
45         user_input = input("Usuario: ")
46         response = chatbot.get_response(user_input)
47         print("ChatBot:", response)
48     except KeyboardInterrupt:
49         break
50
```

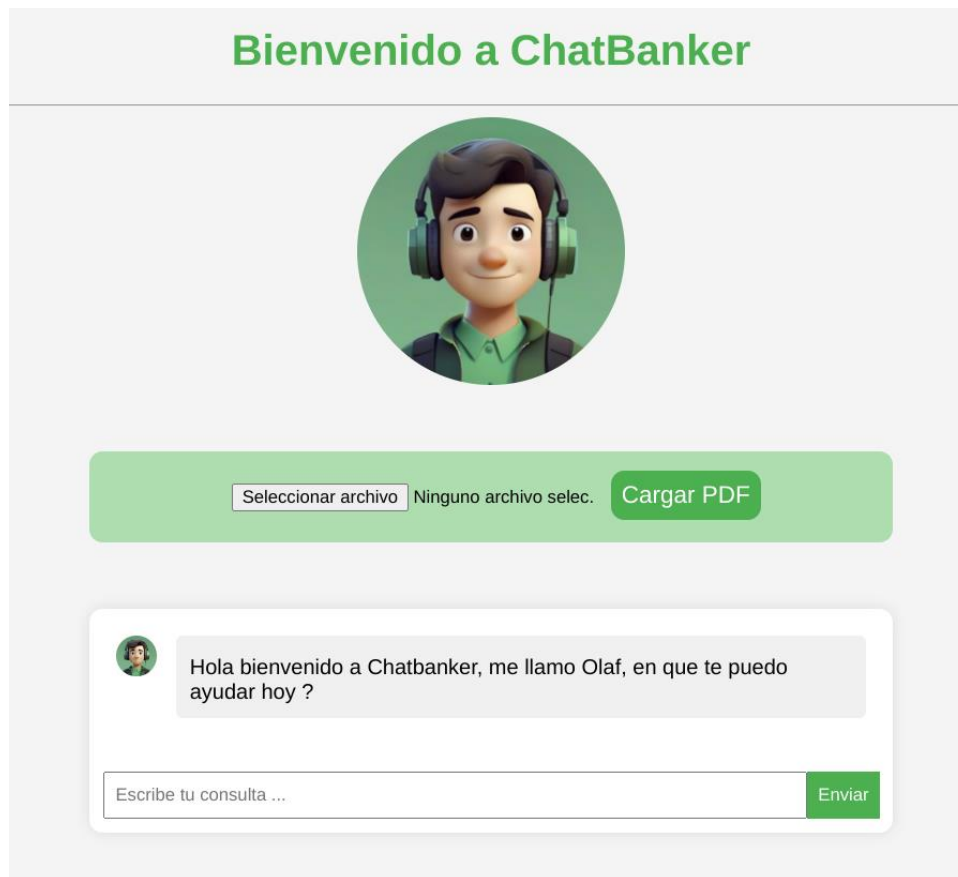
Se uso las librerías de chatterbot para poder generar un modelo de chatboot que se ejecute en consola y que este pueda aprender de la información de los pdf mediante el procesamiento de lenguaje natural generado con las librerías de NLTK y Spacy.

Con esto a continuación se realizó el entrenamiento haciendo el uso de embeddings y almacenándolos dentro de archivos csv, de esta forma toda la información de los csv se tendrá que almacenar solo una vez y se podrá usar para todas las consultas que se desee hacer.

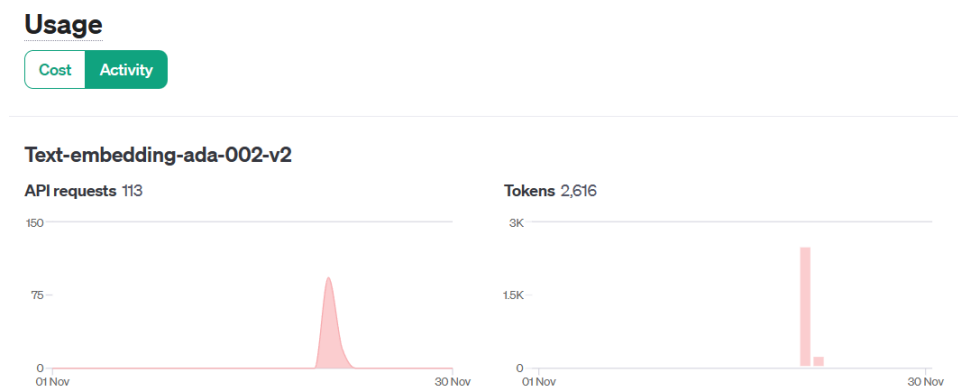
Posterior a ello lo que se realizo es la creación de una función capaz de medir la distancia del coseno entre todos los embeddings preprocesados y la entrada de texto de la cual se desee saber.

Para hacerlo más interactivo, se desarrolló un frontend en Django el cual está sirviendo como GUI para poder hacer uso del chatboot

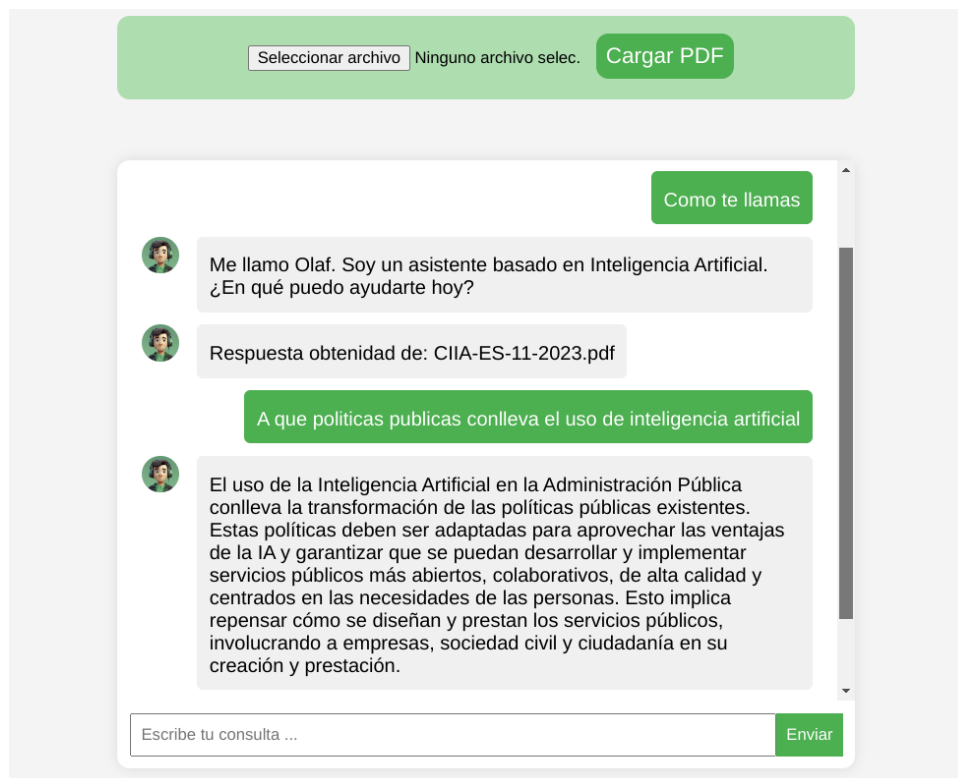
Como se puede ver esta es la interfaz de usuario a la cual se ha llegado mediante el uso de Django, amigable, compacta y fácil de comprender.



Tras todo el entrenamiento y pruebas en el desarrollo se llegó a hacer el uso de 113 request de las cuales el costo por uso del API llego a ser de menos de 1 centavo.



Una vez entrenado se muestra el funcionamiento con la búsqueda por medio de embeddings fusionado con la GUI por parte de Django



## BIBLIOGRAFÍA (referencias)

Moreno, A. (2021, 8 marzo). Procesamiento del lenguaje natural ¿Qué es? Instituto de Ingeniería del Conocimiento. <https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>

¿Qué es el aprendizaje automático? (s. f.). Oracle Argentina. <https://www.oracle.com/ar/artificial-intelligence/machine-learning/what-is-machine-learning/>

colaboradores de Wikipedia. (2021, 23 febrero). Regulación financiera. Wikipedia, la enciclopedia libre. [https://es.wikipedia.org/wiki/Regulaci%C3%B3n\\_financiera](https://es.wikipedia.org/wiki/Regulaci%C3%B3n_financiera)

Top chatbots en servicios financieros para 2021 | BotPress Blog. (s. f.). <https://botpress.com/es/blog/top-chatbots-financial-services#:~:text=El%20uso%20de%20chatbots%20beneficia,y%20les%20aporte%20mayor%20satisfacci%C3%B3n.>

Admin. (2023, 2 agosto). Artificial intelligence solutions for financial data analysis. Coinscrap Finance. <https://coinscrapfinance.com/es/banking-innovation/inteligencia-artificial-para-analisis-datos-financieros/>

Erica - virtual financial assistant from Bank of America. (s. f.). Bank of America. <https://promotions.bankofamerica.com/digitalbanking/mobilebanking/erica>

IBM documentation. (s. f.). <https://www.ibm.com/docs/es/cics-ts/6.1?topic=processes-designing-conversations>

Hammond, M. (2023, 20 enero). Cómo diseñar un protocolo de atención al cliente (con ejemplos). <https://blog.hubspot.es/service/protocolo-servicio-cliente>



Desarrollo. (2023, 15 marzo). ▷ Los 13 Mejores Chatbots para tu sitio web en 2023 - Bloo Media. Bloo Media. Agencia de tecnología de marketing digital. <https://bloo.media/blog/mejores-herramientas-chatbot/>

Kan, A. S. (2023, 12 julio). Botsonic AI Chatbot Builder: Revisión, características, precios, pros y contras - Prompts para IA. Prompts para IA. <https://prompt.uno/chatgpt/botsonic-ai-chatbot-builder-revision-caracteristicas-precios-pros-y-contras/>