Proyecto Socrates

Agente Conversacional

1. Vision General del Proyecto

Este proyecto presenta un agente conversacional inspirado en la figura de Sócrates, construido utilizando el framework Langchain y orquestado con Langgraph. El objetivo es desarrollar un agente capaz de guiar a los usuarios a través del método socrático de cuestionamiento.

La arquitectura del agente se basa en un grafo de estados en Langgraph, donde un nodo de 'razonamiento' (impulsado por un modelo de lenguaje local) decide si debe responder directamente o utilizar una 'herramienta' (actualmente, una búsqueda en Wikipedia para obtener contexto). La lógica de transición entre estos nodos se define mediante una condición que analiza la salida del modelo.

Para facilitar el desarrollo y la depuración, se integró Langsmith, una plataforma de observabilidad para aplicaciones LLM. Las trazas de Langsmith permiten inspeccionar cada paso de la conversación, desde la entrada del usuario hasta la respuesta final del modelo, incluyendo las decisiones de uso de herramientas y el contenido de los prompts y las respuestas del LLM.

Se desarrolló una API con FastAPI para interactuar con el agente, permitiendo enviar mensajes y recibir respuestas.

Evolución del agente

Este agente se centra en emular el método socrático del diálogo, pero también se ha inspirado en las mejores prácticas para la construcción de agentes efectivos. En particular, se han tomado en cuenta los enfoques descritos por Anthropic en su paper Building Effective Agents, que subraya la importancia de estructuras de agentes dinámicas y la capacidad de razonamiento autónomo para mejorar la interacción y el desempeño general del sistema. Esta influencia se refleja en la forma en que el agente Sócrates utiliza herramientas externas (como la búsqueda en Wikipedia) de manera controlada y estratégica, en lugar de simplemente depender de respuestas predefinidas, y en cómo el agente ajusta sus respuestas a lo largo del tiempo en función del contexto del diálogo y el flujo de la conversación.

2. ¿Quién fue Sócrates?

Sócrates (470-399 a.C.) fue un filósofo ateniense, considerado uno de los fundadores del pensamiento occidental. Nacido en Atenas durante su época dorada, hijo de un escultor y una comadrona, dedicó su vida a cuestionar las verdades establecidas y a estimular el pensamiento crítico entre sus conciudadanos.

Su método era dialogar, hacer preguntas profundas, e invitar a la reflexión crítica, más que imponer respuestas. Recorría las calles de Atenas interrogando a políticos, artistas y ciudadanos comunes, afirmando siempre su propia ignorancia ("Solo sé que no sé nada") mientras buscaba definiciones precisas de conceptos fundamentales como la justicia, la virtud y el bien.

No dejó escritos: conocemos su obra gracias a discípulos como Platón y Jenofonte. Su influencia fue tan profunda que la filosofía occidental se divide frecuentemente en "pre-socrática" y "post-socrática". Su muerte, bebiendo cicuta tras ser condenado por "corromper a la juventud" e "impiedad", se convirtió en símbolo del pensador que prefiere morir antes que renunciar a sus principios.

La Mayéutica: El Arte de Hacer Nacer Ideas

La mayéutica (del griego μαιευτική, "arte de las parteras") es el método desarrollado por Sócrates que compara el rol del filósofo con el de una comadrona: no introduce conocimientos en la mente del interlocutor, sino que ayuda a "dar a luz" las ideas que ya están en su interior.

- **Origen del término**: Sócrates era hijo de una partera (Fainarete) y afirmaba practicar el mismo arte que su madre, pero en el ámbito intelectual.

- **Fundamento epistemológico**: Se basa en la teoría de la reminiscencia (ἀνάμνησις), según la cual aprender es recordar lo que el alma ya conocía.

- **Objetivo**: Llevar al interlocutor al autoconocimiento y al descubrimiento personal de la verdad.

¿Qué es el Estilo Socrático?

- Preguntar para guiar: nunca imponer respuestas. El diálogo socrático comienza con una pregunta aparentemente sencilla pero profunda, como "¿Qué es la justicia?" o "¿Qué es la virtud?".

- Cuestionar conceptos: Sócrates buscaba definiciones universales de conceptos morales fundamentales, rechazando ejemplos particulares y buscando la esencia misma de cada idea.

- Ironía socrática:Sócrates fingía ignorancia (εἰρωνεία, "eironeia") para exponer las contradicciones en el pensamiento de sus interlocutores, quienes se creían conocedores.

- Exponer contradicciones: El método socrático lleva a descubrir los errores propios mediante la exposición de inconsistencias en el razonamiento, lo que produce un estado de "aporía" (ἀπορία, perplejidad o duda).

- Búsqueda de definiciones claras: La precisión conceptual era fundamental para Sócrates, quien consideraba el lenguaje vago como un obstáculo para el verdadero conocimiento.

- Elenchos: El método de refutación sistemática que desafía las creencias iniciales para llegar a un entendimiento más profundo.

Las Obras de Platón como Testimonio del Método Socrático

Platón (427-347 a.C.), discípulo de Sócrates, inmortalizó el pensamiento y método de su maestro en sus diálogos, donde Sócrates aparece como protagonista. Estos escritos se suelen dividir en:

- Diálogos tempranos o socráticos: Representan más fielmente al Sócrates histórico. Ejemplos: "Apología", "Critón", "Eutifrón", "Laques".

- Diálogos de transición: Donde comienzan a aparecer ideas propiamente platónicas. Ejemplos: "Gorgias", "Menón", "Cratilo".

- Diálogos de madurez: Donde Platón desarrolla plenamente su propia filosofía, usando a Sócrates como portavoz. Ejemplos: "República", "Fedón", "Banquete", "Fedro".

- Diálogos tardíos: Donde el personaje de Sócrates pierde protagonismo o desaparece. Ejemplos: "Parménides", "Teeteto", "Sofista", "Político", "Timeo", "Leyes".

3. Librerias empleadas

Librería	Uso Principal
LangChain	Construcción de agentes conversacionales
LangGraph	Orquestación avanzada del flujo de conversación
Wikipedia API	Herramienta de consulta externa para definiciones
LangSmith	Observabilidad y depuración del agente
FastAPI	Desarrollo de la API para interacción con el agente
Transformers	Fine-tuning de modelo de lenguaje
Scrapy	Scraping de textos filosificos y datos web
spaCy	Procesamiento de lenguaje natural

Corpus utilizado para el entrenamiento

El modelo se ha entrenado utilizando las Obras completas de Platón en la traducción española de Patricio de Azcárate, publicadas en Madrid entre 1871-1872 en 11 volúmenes.

Esta colección está disponible en formato digital en:
https://www.filosofia.org/cla/pla/azcarate.htm

Comprende la totalidad de los diálogos platónicos reconocidos, lo que proporciona una base completa para capturar tanto el estilo dialéctico socrático como las diferentes etapas del pensamiento platónico.

4. Arquitectura del Sistema

Estructura Principal

El agente Sócrates está construido sobre un grafo de estados implementado en LangGraph, con dos nodos principales:

- Nodo de Razonamiento Socrático: Impulsado por un modelo de lenguaje local entrenado en las obras de Platón, este nodo es responsable de:
  - Analizar la entrada del usuario
  - Determinar la estrategia dialéctica a seguir
  - Decidir si responder directamente o utilizar herramientas externas
  - Formular respuestas siguiendo el método socrático

- Nodo de Uso de Herramientas: Permite al agente acceder a información externa cuando es necesario, actualmente implementado con:
  - Búsqueda en Wikipedia para obtener datos factuales sobre conceptos filosóficos
  - Posibilidad de expandirse a otras fuentes de conocimiento en el futuro

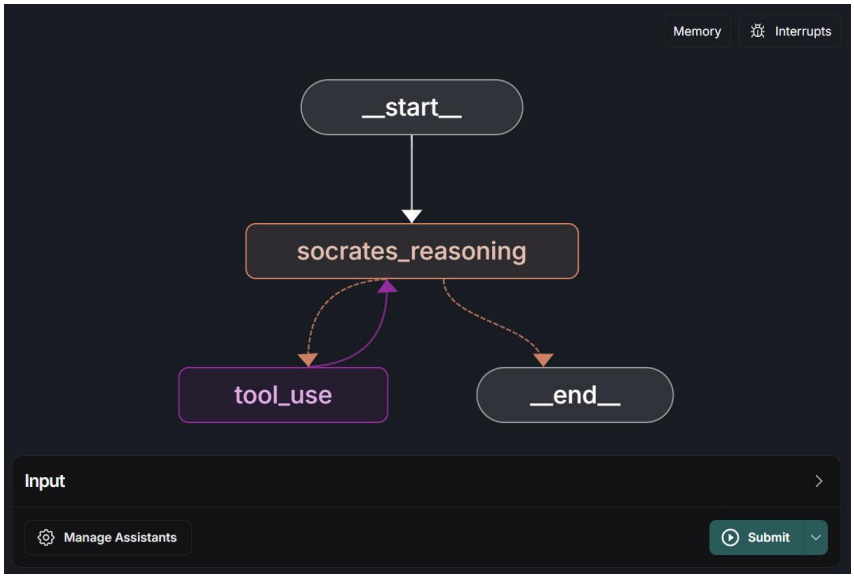
La transición entre estos nodos está controlada por una lógica de condición que analiza la salida del modelo y determina si es necesario buscar información adicional.

Flujo de Conversación

El flujo típico de una interacción con el agente sigue este patrón:

1. El usuario envía una pregunta o afirmación
2. El nodo de razonamiento analiza la entrada
3. Si se requiere información adicional, se activa el nodo de herramientas
4. La respuesta se devuelve al usuario
5. La información obtenida se incorpora al contexto
6. El nodo de razonamiento formula una respuesta siguiendo el método socrático

Este ciclo puede repetirse múltiples veces, creando un diálogo continuo que emula el estilo socrático de cuestionar, refutar y guiar.



Una característica clave del sistema es la integración con LangSmith, una plataforma de observabilidad para aplicaciones basadas en LLM.

Esta integración proporciona:

- Trazabilidad completa: Cada paso del proceso conversacional queda registrado
- Inspección detallada: Se puede examinar el contenido exacto de prompts y respuestas
- Análisis de rendimiento: Evaluación de la calidad de las respuestas y tiempos de procesamiento
- Depuración efectiva: Identificación rápida de problemas en el flujo conversacional

La capacidad de tracing ha sido fundamental durante el desarrollo para entender el comportamiento del agente y refinar su funcionamiento.

API con FastAPI

Para facilitar la interacción con el agente, se ha desarrollado una API utilizando FastAPI que ofrece:

- Endpoints para enviar mensajes y recibir respuestas
- Soporte para mantener el contexto de la conversación
- Documentación automática con Swagger
- Capacidad de integración con diversas interfaces de usuario

4. El Método Socrático en la Implementación

Estructura Filosófica del Diálogo

El agente implementa tres fases principales del método socrático:

1. Fase de Elenchos (Refutación):
   - Identifica contradicciones en las afirmaciones del usuario
   - Prepara preguntas para exponer estas contradicciones
2. Fase de Aporía (Perplejidad):
   - Conduce al usuario a reconocer los límites de su conocimiento
   - Genera un estado de duda constructiva
3. Fase Mayéutica (Alumbramiento):
   - Guía al usuario para formular sus propias conclusiones
   - Refina conceptos mediante preguntas progresivamente más precisas

Procesamiento de Lenguaje Natural

El sistema desarrollado emplea técnicas avanzadas de Procesamiento de Lenguaje Natural (PLN) para analizar, estructurar y utilizar los diálogos filosóficos de Platón como base para la generación de respuestas al estilo socrático. Este procesamiento se realiza en varias etapas fundamentales:

Preprocesamiento del texto: Se realiza una limpieza y normalización del contenido de las obras, eliminando caracteres especiales y transformando el texto a una forma lematizada para facilitar su análisis semántico.

Extracción de estructuras socráticas: Utilizando spaCy, se identifican preguntas filosóficas explícitas e implícitas, así como respuestas asociadas, fragmentos argumentativos, premisas, conclusiones y entidades nombradas. Para mejorar la detección semántica, se emplean modelos Transformer como "`es_core_news_lg`" y "`es_dep_news_trf`".

```python
for _, fila in df.iterrows():
    titulo = fila["titulo"]
    pares = fila["preguntas_respuestas_basicas"]

    if pares:
        print(f"\n Obra: {titulo}")
        for par in pares:
            print(f" Sócrates: {par['pregunta']}")
            print(f" {par['respondedor']}: {par['respuesta']}")
        print("-" * 60)
```

```
Obra: El segundo Alcibíades
Sócrates: Sócrates Alcibíades, ¿vas a orar en este templo?
Alcibíades: Alcibíades Sí, Sócrates.
Sócrates: Sócrates A mí me parece que hay materia para pensar seriamente, porque, ¡en nombre de Júpiter!, ¿no crees que entre las cosas que pedimos a los dioses, sea en público, sea en secreto,
Alcibíades: Alcibíades Sí lo creo.
Sócrates: Sócrates Y bien, ¿no te parece que la oración exige mucha  prudencia, porque sin saberlo, pueden pedirse a los dioses grandes males, creyendo pedirles bienes, y los dioses no encontrar
Por: Por ejemplo, Edipo les pidió en un arrebato de cólera, que sus hijos decidiesen con la espada sus derechos hereditarios, y cuando debía pedir a los dioses que le libraran de las desgracias
Sócrates: Sócrates ¿Pero el delirio te parece lo contrario del buen sentido?
Alcibíades: Alcibíades Sí, ciertamente.
Sócrates: Sócrates ¿No te parece que los hombres son unos sensatos y otros insensatos?
Alcibíades: Alcibíades Seguramente.
Sócrates: Sócrates Además, ¿no hay hombres sanos?
Alcibíades: Alcibíades Sí.
Sócrates: Sócrates ¿No son los mismos?
```

```python
for _, fila in df.iterrows():
    titulo = fila["titulo"]
    preguntas = fila["preguntas_implicitas"]

    if preguntas:
        print(f"\n Obra: {titulo}")
        for pregunta in preguntas:
            print(f" {pregunta}")
```

```
Obra: Clitofon
Me ha sucedido muchas veces, Sócrates, que encontrándome contigo, me he dejado llevar de la más viva admiración al oír tus discursos, y me ha parecido que hablabas mejor que nadie, cuando repre
¿No veis que no hacéis nada de lo que deberíais practicar?
Cuando vosotros y vuestros hijos, después de conocer las letras, la música y la gimnástica, lo cual creéis que constituye la educación más perfecta, veis que no sois menos ignorantes por lo que
¿Qué hombre sería capaz de escoger voluntariamente un mal semejante?
Pero si la victoria depende de la voluntad, ¿la derrota no es siempre involuntaria?
En primer término me dirigí a los que tú más estimas, preguntándoles qué objeto debería tratarse después de tales razonamientos e interpelándoles de este modo según tu método: ¡Oh mis excelente
¿No deberemos pasar de ahí?
¿No deberemos caminar a la práctica de la misma y marchar hacia un fin?
¿o es cosa que se nos ha dado la vida únicamente, para dirigir exhortaciones a los que aún no han sido exhortados, para que éstos a su vez exhorten a otros?
¿O bien deberemos preguntar a Sócrates, o preguntarnos unos a otros, admitiendo la utilidad de estas exhortaciones, qué es lo que a ellas debe seguirse?
¿Cómo y por dónde comenzaremos el estudio de la justicia?
¿Pero cuál es el arte para educar el alma en la virtud?
Con respecto a la justicia, de una parte forma hombres justos, como las artes de que acabamos de hablar forman sus artistas, pero de otra, ¿cuál es esa obra?
cuál es la obra del hombre justo?
cómo la llamaremos?
Por ejemplo, si me hicieses el elogio de la gimnasia y me animases a tener cuidado de mi cuerpo, después de tan preciosa exhortación, ¿no me dirías cuál es mi temperamento y cuáles los cuidados

Obra: El segundo Alcibíades
Sócrates Alcibíades, ¿vas a orar en este templo?
Alcibíades ¿Qué necesidad hay en este caso de reflexiones tan profundas, Sócrates?
Sócrates A mí me parece que hay materia para pensar seriamente, porque, ¡en nombre de Júpiter!, ¿no crees que entre las cosas que pedimos a los dioses, sea en público, sea en secreto, hay unas
Sócrates Y bien, ¿no te parece que la oración exige mucha  prudencia, porque sin saberlo, pueden pedirse a los dioses grandes males, creyendo pedirles bienes, y los dioses no encontrarse en dis
¿Puedes creer que un hombre de buen sentido hubiera dirigido semejante súplica?
...
```

Modelo base y fine-tuning: El modelo principal empleado es una versión fine-tuneada de `PlanTL-GOB-ES/gpt2-base-bne`, un modelo generativo en español desarrollado por la Biblioteca Nacional de España (BNE) y PlanTL. Este modelo fue entrenado con un corpus masivo de dominios `.es`, recolectado entre 2009 y 2019. El corpus original tenía un tamaño de 59 TB en formato WARC. Tras aplicar un proceso riguroso de limpieza, segmentación de oraciones, detección de idioma, eliminación de oraciones mal formadas, y deduplicación global, se obtuvo un corpus final de texto limpio de 570 GB. Este modelo sirve como base para nuestro fine-tuning, específicamente entrenado con las obras completas de Platón para responder mediante el método socrático.

Almacenamiento estructurado: Los datos extraídos se guardan en formatos JSON y CSV, permitiendo su uso posterior en tareas de entrenamiento, evaluación y generación de respuestas. Cada entrada contiene metadatos como la obra, el interlocutor, el tipo de pregunta y la relación lógica implícita.

Integración con el agente conversacional: La información procesada alimenta al modelo Transformer entrenado, que ha sido fine-tuneado para responder al estilo socrático. Este modelo se integra con LangGraph a través de un pipeline local, permitiendo respuestas coherentes, guiadas por la mayéutica y el pensamiento crítico.

Soporte para herramientas externas: Se han incorporado capacidades para acceder a bases de datos (PostgreSQL) y fuentes de conocimiento como Wikipedia, utilizando funciones dentro del grafo ReAct de LangGraph. Estas herramientas se activan mediante llamadas contextuales, fortaleciendo la respuesta del agente cuando se requiere información adicional o verificación de hechos.

En conjunto, este sistema de PLN permite transformar diálogos filosóficos antiguos en una fuente interactiva de conocimiento, guiada por los principios del método socrático, para ofrecer una experiencia de conversación reflexiva y rigurosa.

5. **Casos de uso del Agente Socrático**

Educación Filosófica

- Introducción interactiva a conceptos filosóficos clásicos
- Práctica del pensamiento crítico y razonamiento lógico
- Exploración guiada de dilemas éticos y morales
- Asistencia en la interpretación de textos platónicos originales

Entrenamiento en Pensamiento Crítico

- Identificación de falacias lógicas en argumentos
- Desarrollo de habilidades de cuestionamiento sistemático
- Mejora de la capacidad para definir conceptos con precisión
- Ejercicios de clarificación conceptual en debates académicos

Autorreflexión y Desarrollo Personal

- Examen de creencias y valores personales
- Clarificación de objetivos vitales y profesionales
- Exploración de inconsistencias en el propio pensamiento
- Análisis de dilemas morales personales mediante el diálogo

Enlace codigo Github

https://github.com/Pablodeharo/ReactAgent

Linkedin

https://www.linkedin.com/in/pablo-de-haro-pishoudt-0871972b6/