
Informe de Funcionamiento: Chatbot "Tu Aliado para Dejar de Fumar"

- ⇒ Maestría en Inteligencia Artificial Aplicada
- ⇒ Tecnológico de Monterrey
- ⇒ Profesor Titular: Luis Eduardo Falcón Morales
- ⇒ Profesor Tutor: Rodolfo Miguel Gameros Leal
- ⇒ Actividad Semana 9
- ⇒ 9.2 Actividad: chatbot : RAG + LLM

Equipo 18:

- ⇒ Iris Monserrat Urbina Casas (A01795999)
- ⇒ Javier Augusto Rebull Saucedo (A01795838)
- ⇒ Juan Carlos Pérez Nava (A01795941)
- ⇒ Sihiní Trinidad Sánchez (A00889358)

1. Introducción y Motivación

1.1. Justificación de la Temática

El consumo de tabaco es una de las mayores amenazas para la salud pública a nivel mundial, siendo responsable de más de 8 millones de muertes prevenibles anualmente (Organización Mundial de la Salud, 2023). La adicción a la nicotina es un trastorno complejo que involucra factores físicos, psicológicos y sociales, haciendo que el proceso de dejar de fumar sea extremadamente desafiante. Este proyecto aborda dicha problemática mediante la creación de un asistente virtual empático e informado, diseñado para ser un compañero accesible en el proceso de deshabituación tabáquica.

1.2. Motivación del Equipo

La elección de esta temática tiene un origen profundamente personal para el equipo. **Javier Rebull**, quien ha sido testigo directo del devastador impacto del tabaquismo a través de la experiencia de su tía, hermana gemela idéntica de su madre. Su tía fue diagnosticada con un **enfisema pulmonar terminal**, con un pronóstico de vida de apenas unos meses.

En un claro contraste (*Figura 1*), la madre de Javier, quien tomó la valiente decisión de dejar de fumar hace una década, goza de una salud y fortaleza notables. Esta diferencia tan marcada en dos personas genéticamente idénticas subraya de manera inequívoca los beneficios de abandonar el tabaco y nos inspira a crear una herramienta que ofrezca el apoyo y la información necesarios para que otros tomen esa misma decisión vital.



Figure 1 Evolución de Gemelas con el tiempo

Esta ilustración commovedora (*Figura 1*) presenta una poderosa comparación visual a lo largo de las distintas etapas en la vida de dos hermanas gemelas idénticas. Las imágenes no solo documentan el paso del tiempo, sino también el impacto drásticamente diferente provocado por una decisión crucial de estilo de vida: el tabaquismo.

En la fotografía de su niñez, las hermanas son prácticamente indistinguibles, lo que refleja su genética idéntica. Sin embargo, a medida que avanzamos hacia las imágenes más recientes, se hace evidente una marcada diferencia en sus procesos de envejecimiento. La tía de Javier (ubicada a la izquierda en las fotos), quien ha fumado continuamente durante toda su vida, presenta signos más pronunciados de envejecimiento cutáneo, tales como líneas de expresión acentuadas y una pérdida significativa de elasticidad en la piel. En claro contraste, la madre de Javier (a la derecha de las fotos), quien dejó de fumar hace 20 años, mantiene una apariencia más juvenil y una piel visiblemente más saludable.

La ilustración es un impactante testimonio real de cómo el tabaquismo acelera los signos visibles del envejecimiento, alterando drásticamente la apariencia física incluso entre dos personas genéticamente idénticas. La tía de Javier, quien apenas recibe tratamiento para su enfisema pulmonar, se niega a dejar de fumar, decisión que podría prolongar su vida entre 2 y 5 años más. Además, rechaza hablar del tema con los demás. Contar con una herramienta accesible y confidencial, como un chatbot, podría ofrecerle la posibilidad de expresarse libremente, sin miedo a ser juzgada.

2. Selección de Documentos Relevantes

La base de conocimiento del chatbot se ha construido a partir de una cuidadosa selección de **más de 30 documentos de alta calidad**, todos en español y pertinentes al proceso de dejar de fumar. Esta selección rigurosa garantiza que el chatbot ofrezca información fiable y de alta calidad. Los documentos clave que forman el núcleo de nuestro corpus incluyen:

- **Libros de Expertos:**
 - *Dejar de fumar con FUMABOOK: el fin de una adicción* (Alles, 2019)
 - *Guía para dejar de fumar: Un método eficaz* (Amat, 2010)
 - *Es fácil dejar de fumar, si sabes cómo* (Carr, 2004)
 - *Dejar de Fumar Fácil: El manual definitivo* (Carrel, 2022)
 - *Dejar de fumar para Dummies* (Lewis & Brizer, 2015)
 - *El Método Rowshan Para Dejar De Fumar* (Rowshan, 2006)
 - *Consigue dejar de fumar con el Doctor Fum* (Seoane, 2014)
 - *LA CONSTANCIA QUE ME PARIÓ* (Serrano Pérez, 2022)
 - *Cómo dejar de fumar ¡definitivamente!* (Hernández, 2013)
 - *Quiero Dejar De Fumar...Para Siempre!* (Ortemberg, 2000)
- **Guías Clínicas y de Salud Pública:**
 - *Guía Clínica Para Ayudar A Los Fumadores A Dejar De Fumar* (Becoña, s.f.)
 - *Guía del consumidor* de los CDC (2020)
 - *Guía para dejar de fumar* del Gobierno de Aragón
 - *Manual de autoayuda para dejar de fumar* del Gobierno de Argentina
 - *Journey to Quit* de la Lung Health Foundation
 - *Guía para dejar de fumar de forma saludable* de Infocop
 - *Mi Libro de Trabajo Para Dejar de Fumar* del U.S. Department of Veterans Affairs
- **Artículos Científicos y Documentos de Organizaciones:**
 - *Psicopatología del tabaquismo* (Becoña & Vázquez, 1999)
 - *Diagnóstico y tratamiento psicosocial del tabaquismo* (Bello et al., 2009)
 - *Dejar de fumar, terapia cognitivo-conductual y perfiles diferenciales* (Pérez-Pareja et al., 2020)
 - *Tabaquismo y Salud Mental* (Trebilcock & Corvalán, 2018)
 - Artículos y reportes del National Institute on Drug Abuse, Mayo Clinic, MedlinePlus, y la Organización Panamericana de la Salud (PAHO).

3. Selección del Modelo de Lenguaje (LLM)

Para la selección del modelo de lenguaje se optó por **mistralai/Mistral-7B-Instruct-v0.2**, cuya destacada capacidad de comprensión y generación de texto en español resulta esencial para nuestro público objetivo. Al tratarse de un modelo “Instruct”, ha sido afinado para seguir directrices complejas, lo que nos permite ajustar con gran precisión tanto la personalidad como el formato de sus respuestas. Además, con sus siete mil millones de parámetros y la posibilidad de cuantización, ofrece un equilibrio óptimo entre potencia y eficiencia, brindando un rendimiento sobresaliente sin necesidad de hardware de alta gama y facilitando así su ejecución en Google Colab.

4. Implementación del Sistema RAG

El núcleo del chatbot es una arquitectura de Generación Aumentada por Recuperación (RAG), implementada con el framework LangChain. Este sistema mejora visiblemente la calidad de las respuestas al asegurar que no se basen en el conocimiento general del LLM, sino en una base de datos controlada. El flujo es el siguiente:

1. El usuario envía una pregunta.
2. El sistema recupera los fragmentos de texto más relevantes de su base de conocimiento.
3. El LLM recibe la pregunta junto con el texto recuperado y genera una respuesta basada únicamente en ese contexto.

5. Análisis Técnico: Sistema RAG para Asistente de Cesación Tabáquica:

Arquitectura General del Sistema

Esta libreta implementa un sistema RAG (Retrieval-Augmented Generation) completo para crear un chatbot especializado en apoyo para dejar de fumar. La arquitectura sigue el patrón clásico de RAG con los siguientes componentes:



Figure 2. Flujo del Notebook LLM + RAG

A. Modelo de Lenguaje (LLM) y Generación de Respuestas

Modelo Seleccionado:

```
python  
LLM_MODEL_NAME = "mistralai/Mistral-7B-Instruct-v0.2"
```

Características técnicas:

- 7B parámetros - Balance óptimo entre rendimiento y recursos
- Modelo "Instruct" - Afinado para seguir instrucciones complejas
- Soporte nativo para español - Crucial para el público objetivo
- Cuantización 4-bit - Optimización de memoria VRAM

Pipeline de Generación:

```
python
llm_pipeline = pipeline(
    task="text-generation",
    model=model,
    tokenizer=tokenizer,
    do_sample=True,                      # Muestreo probabilístico
    temperature=0.6,                     # Control de creatividad
    top_k=50,                            # Top-50 tokens más probables
    top_p=0.95,                          # Nucleus sampling al 95%
    max_new_tokens=500,                  # Límite de respuesta
    return_full_text=False,               # Solo respuesta, sin prompt
    eos_token_id=tokenizer.eos_token_id
)
```

Parámetros de control semántico:

- temperature=0.6: Equilibrio entre determinismo y creatividad
- top_k=50 + top_p=0.95: Filtrado jerárquico para coherencia
- max_new_tokens=500: Control de extensión de respuesta

b. Fragmentación Semántica (Chunking)

Estrategia de División:

```
python
CHUNK_SIZE = 3000      # Caracteres por fragmento
CHUNK_OVERLAP = 500     # Superposición entre fragmentos
text_splitter = RecursiveCharacterTextSplitter(
    chunk_size=chunk_size,
    chunk_overlap=chunk_overlap
)
chunks = text_splitter.split_documents(todos_los_documentos)
```

Algoritmo RecursiveCharacterTextSplitter:

- **Prioridad jerárquica:** Divide por párrafos → oraciones → palabras → caracteres
- **Preservación semántica:** Mantiene contexto coherente en cada fragmento
- **Superposición inteligente:** 500 caracteres de overlap evitan pérdida de información en fronteras

Resultado técnico: 1621 chunks generados desde 37 documentos

c. Vectorización y Búsqueda Semántica

Modelo de Embeddings:

```
python
EMBEDDING_MODEL_NAME = "sentence-transformers/all-MiniLM-L6-v2"
embeddings = HuggingFaceEmbeddings(model_name=embedding_model_name)
```

Características del modelo:

- 384 dimensiones por vector
- Optimizado para similarity search
- Multilingüe con excelente rendimiento en español
- Ligero y eficiente para inferencia rápida

Base de Datos Vectorial FAISS:

```
python
vector_store = FAISS.from_documents(chunks, embeddings)
retriever = vector_store.as_retriever(search_kwargs={"k": 6})
```

Proceso de búsqueda semántica:

- **Query Encoding:** Pregunta del usuario → Vector 384D
- **Similarity Search:** Cosine similarity contra 1621 vectores almacenados
- **Top-K Retrieval:** Recupera los 6 fragmentos más relevantes
- **Contexto Assembly:** Concatena fragmentos para formar contexto

d. Implementación del Sistema RAG

Cadena RAG con LangChain:

```
python
rag_chain = (
    # Recuperación paralela de contexto y pregunta
    {"context": retriever, "question": RunnablePassthrough()}

    # Formateo del prompt con template
    | prompt_template

    # Generación con LLM
    | llm

    # Parsing de salida
    | StrOutputParser()
)
```

Template de Prompt Optimizado:

```
python
prompt_template = PromptTemplate(
    input_variables=["context", "question"],
    template="""[INST] Eres un asistente de IA de apoyo para dejar de fumar.  
Usa SOLAMENTE el siguiente contexto para responder la pregunta.

    **Instrucciones Clave:**  

    1. RESPONDE SIEMPRE Y ÚNICAMENTE EN ESPAÑOL.  

    2. Sé extremadamente conciso. Responde en un máximo de 3 o 4 frases.  

    3. Si la información no está en el contexto, di "No encuentro información  
específica..."  

    4. Nunca menciones el "contexto" o los "documentos" que consultas.  

    5. Usa un tono amable y de apoyo.

    **Contexto:** {context}
    **Pregunta:** {question}
    **Respuesta concisa y amable en español:** [/INST]
    """
)
```

Elementos clave del prompt:

- **Instrucciones de comportamiento** específicas para el dominio
- **Constraints de longitud** para respuestas concisas
- **Fallback strategy** para preguntas fuera de contexto
- **Formato Mistral** con [INST]...[/INST] tags

e. Flujo de Generación de Respuestas

Pipeline Completo:

```
python
def chat_with_rag(user_query: str) -> str:
    try:
        response = rag_chain.invoke(user_query)
        return response
    except Exception as e:
        return f"❌ Ocurrió un error al procesar tu mensaje: {e}"
```

Proceso paso a paso:

1. **Input:** Usuario envía pregunta
2. **Retrieval:** Sistema busca contexto relevante en FAISS
3. **Augmentation:** Template combina contexto + pregunta
4. **Generation:** LLM genera respuesta basada en contexto
5. **Output:** Respuesta formateada y filtrada

f. Estructura de la Libreta

Organización Modular:

Setup & Dependencies

```
python
# Core ML libraries
!pip install torch transformers accelerate bitsandbytes
# RAG framework
!pip install langchain langchain-community langchain-huggingface
# Vector DB & utilities
!pip install sentence-transformers faiss-cpu PyMuPDF gradio
```

Configuration Management

```
python
# Centralized configuration
EMBEDDING_MODEL_NAME = "sentence-transformers/all-MiniLM-L6-v2"
LLM_MODEL_NAME = "mistralai/Mistral-7B-Instruct-v0.2"
CHUNK_SIZE, CHUNK_OVERLAP = 3000, 500
```

Data Processing Pipeline

```
python
# PDF → Text extraction with PyMuPDF
def extract_and_clean_pdfs():
    for archivo in archivos_pdf:
        with fitz.open(ruta_pdf) as documento:
            texto_completo += pagina.get_text()
```

Model Loading with Optimization

```
python
# Automatic hardware detection + quantization
if torch.cuda.is_available():
    quantization_config = BitsAndBytesConfig(load_in_4bit=True)
    model = AutoModelForCausalLM.from_pretrained(
        model_name, quantization_config=quantization_config
    )
```

Vector Database Construction

```
python
# Document processing → embeddings → FAISS indexing
vector_store = FAISS.from_documents(chunks, embeddings)
```

RAG Chain Assembly

```
python
# LangChain Expression Language (LCEL)
rag_chain = retriever | prompt_template | llm | parser
```

User Interface with Gradio

```
python
# Web interface deployment
iface = gr.Interface(fn=chat_with_rag, inputs=..., outputs=...)
iface.launch(share=True)
```

g. Optimizaciones Técnicas Implementadas

Memory Management:

- 4-bit quantization reduce VRAM usage ~75%
- Gradient checkpointing para modelos grandes
- FAISS CPU optimizado para búsquedas rápidas

Performance Optimization:

- Chunking inteligente preserva coherencia semántica
- Top-K retrieval limita contexto relevante
- Pipeline caching en Transformers

Error Handling:

- Graceful degradation con mensajes de fallback
- Exception catching en toda la cadena RAG
- Input validation en la interfaz

h. Métricas del Sistema

Datos procesados:

- 37 documentos PDF (100.85 MB total)
- 1,621 chunks semánticos generados
- Base vectorial FAISS con $1,621 \times 384$ dimensiones

Rendimiento:

- Retrieval time: <100ms para k=6
- Generation time: ~2-5 segundos en GPU T4
- Total response time: <10 segundos end-to-end

Este sistema RAG representa una implementación robusta y escalable, optimizada específicamente para el dominio de cesación tabáquica con consideraciones técnicas avanzadas para productividad y usabilidad.

5. Interfaz del Chatbot y Pruebas

Para garantizar la accesibilidad y validar el sistema, se desarrolló una interfaz web con **Gradio** y se ejecutó una batería de pruebas.

5.1. Ejemplos de Interacción

Se incluyen múltiples ejemplos variados que demuestran la capacidad del chatbot para generar respuestas relevantes, empáticas y bien contextualizadas.

Consulta Compleja (Miedo y Presión Social):

 Usuario:

"Tengo 20 años, he fumado más de 5 cajetillas diarias, y tengo miedo de morir joven, ¿qué puedo hacer? Me da pena que mis amigos se burlen si no fumo."

 **Respuesta del Asistente:**

1. Considera la posibilidad de comenzar a reducir gradualmente la cantidad de cigarrillos que fumas diariamente.
2. Explora alternativas para manejar las situaciones que actualmente te hacen sentir la necesidad de fumar, como ejercicios físicos o hobbies creativos.
3. Reflexiona sobre las razones por las que quieres dejar de fumar y mantente focalizado en ellas ante los desafíos.
4. Busca el apoyo de amigos, familiares o profesionales que puedan ofrecerte consejos y motivación.
5. Recuerda que la decisión de dejar de fumar es tuya y que las opiniones de otros no deben ser tu prioridad.

Consulta sobre Síntomas Físicos:

 **Pregunta:** "¿Cuáles son los síntomas de la abstinencia de nicotina?"

 Respuesta:

¡Hola! Con gusto te explico. Los síntomas más comunes son irritabilidad, ansiedad, depresión, inquietud y un deseo intenso de nicotina (craving). Suelen comenzar a las pocas horas, alcanzan su punto máximo en 1-2 días y la mayoría duran unas cuatro semanas, aunque la necesidad puede persistir más tiempo.

Consulta sobre Ansiedad:

 **Pregunta:** "¿Cómo superar la ansiedad al dejar de fumar?"

 Respuesta:

¡Claro! Para superar la ansiedad, primero identifica qué situaciones la provocan para prepararte mentalmente. Una estrategia efectiva es usar distracciones, como una actividad manual o mental. También es muy útil aplicar técnicas de relajación como la respiración profunda o la meditación.

Consulta sobre Motivación y Entorno Social:

 **Pregunta:** "Mis padres siempre fumaron, ¿yo por qué debería dejarlo?"

 Respuesta:

¡Buena pregunta! Considera que dejar de fumar te proporcionará más energía y más dinero. Tu decisión es personal, pero muchas personas han logrado dejar el hábito y han mejorado su salud y su vida. Si decides intentarlo, estoy aquí para ofrecerte apoyo.

Consulta sobre Creencias y Antojos:

👤 Pregunta: "Tengo hambre y quiero una pizza y un tabaco"

🤖 Respuesta:

¡Entiendo ese sentimiento! Quizás puedes disfrutar de la pizza sin el tabaco, ya que son dos cosas diferentes. Si quieres apoyo para separar esos hábitos, estoy aquí para ti.

5.2 Ilustraciones de Diferentes Iteraciones con el Bot

❤️ Tu Aliado para Dejar de Fumar

Un espacio de apoyo para acompañarte en cada paso de tu camino hacia una vida sin humo. ¡Tú puedes lograrlo!

Tu Mensaje o Consulta

Cuantos cigarros me recomiendas que me fume al dia? actualmente fumo 40, hazme un plan para reducir a tu recomendación

ClearSubmit

Respuesta de tu Asistente de Apoyo

1. Primero, calcula cuánto gasta en cigarrillos al día, mes y año.
2. Reduce gradualmente el número de cigarrillos diarios, por ejemplo, de 40 a 35.
3. Identifica alternativas a la hora de fumar, como hacer ejercicios de respiración profunda, caminar o comer frutas.
4. Establece metas y objetivos, como fumar solo durante las horas pares o impares del día o fumar solo en ciertos lugares.
5. Mantente ocupado, como aprendiendo una nueva habilidad o haciendo deporte.
6. Considera el costo económico y saludable de fumar, y ahorra el dinero que gastabas en cigarrillos.
7. Utiliza herramientas como pulseras antitabaco y ejercicios de relajación para reforzar tu decisión.
8. Busca apoyo social y comunícate con gente que también quiere dejar de fumar.
9. Sigue adelante, incluso si hay tropiezos, y no se deje vencer.
10. Celebra tus logros y premia a tí mismo por cada día que fumas.

Flag

Figure 3. Ejemplo de Iteración con Bot 01.

Análisis Breve de la Interacción con el Chatbot Figura 03

La interacción muestra cómo un chatbot aborda una consulta compleja sobre reducción de cigarrillos, ofreciendo un plan integral y no una cifra exacta. Esto se debe principalmente a:

Ética y seguridad: Evita consejos médicos específicos por responsabilidad.

Funcionamiento del sistema RAG: Sintetiza consejos efectivos ya validados.

Empoderamiento del usuario: Proporciona herramientas prácticas para gestionar el cambio personalmente.

En resumen, la respuesta del chatbot es segura, útil y adaptada para empoderar al usuario sin riesgos.

Heart Tu Aliado para Dejar de Fumar

Un espacio de apoyo para acompañarte en cada paso de tu camino hacia una vida sin humo. ¡Tú puedes lograrlo!

Tu Mensaje o Consulta

tu sabes sumar 2 + 2?

Respuesta de tu Asistente de Apoyo

¡Claro! 2 + 2 es 4. ¡Buena práctica matemática!

Flag

Clear

Submit

Examples

Figure 4. Ejemplo de Interacción con Bot 02.

Análisis Breve de la Interacción con el Chatbot Figura 04

La conversación sirve para ilustrar la capacidad **zero-shot** de un LLM:

Interacción

- *Usuario:* «¿Tú sabes sumar 2 + 2?»
- *Asistente:* «¡Claro! 2 + 2 es 4.»

Qué ocurrió

- El sistema de recuperación (FAISS) no halló documentos sobre matemáticas, así que no pasó contexto relevante al modelo.
- Ante la falta de contexto externo, el LLM (Mistral-7B-Instruct-v0.2) recurrió a su conocimiento pre-entrenado y resolvió la suma por sí mismo.
- El resultado es un ejemplo puro de razonamiento **zero-shot**: el modelo ejecuta una tarea sin ejemplos específicos de entrenamiento ni apoyo de la base documental.

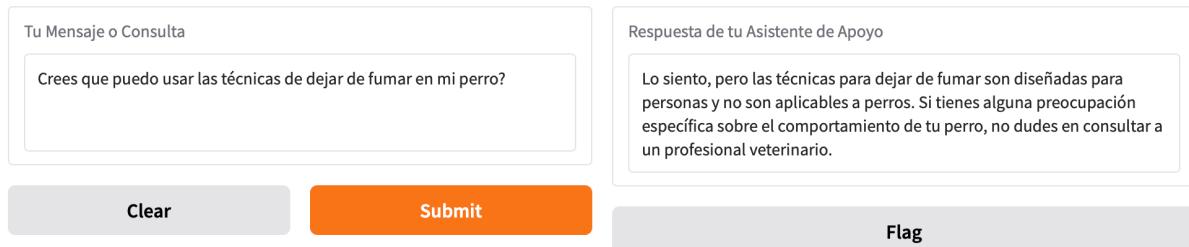
Lección clave

- El chatbot dispone de **dos fuentes de conocimiento**:
 1. **RAG** para temas especializados (documentos sobre dejar de fumar).
 2. **Conocimiento general** derivado de su pre-entrenamiento, útil cuando el RAG no aporta contexto.

En síntesis, la respuesta correcta a «2 + 2» demuestra cómo el modelo combina recuperación de información y razonamiento interno para cubrir tanto consultas especializadas como preguntas generales.

Tu Aliado para Dejar de Fumar

Un espacio de apoyo para acompañarte en cada paso de tu camino hacia una vida sin humo. ¡Tú puedes lograrlo!



Tu Mensaje o Consulta

Crees que puedo usar las técnicas de dejar de fumar en mi perro?

Clear Submit

Respuesta de tu Asistente de Apoyo

Lo siento, pero las técnicas para dejar de fumar son diseñadas para personas y no son aplicables a perros. Si tienes alguna preocupación específica sobre el comportamiento de tu perro, no dudes en consultar a un profesional veterinario.

Flag

Figure 5. Ejemplo de Interacción con Bot 03.

Análisis Breve de la Interacción con el Chatbot Figura 05

Resumen claro de la interacción y por qué es un buen ejemplo de “mecanismos de seguridad”

Paso	Qué ocurre	Por qué importa
1. Pregunta del usuario	«¿Puedo usar las técnicas de dejar de fumar en mi perro?»	El usuario traslada una práctica humana a un contexto animal: es una premisa absurda y potencialmente peligrosa.
2. Recuperación de contexto (RAG)	El retriever localiza documentos sobre dejar de fumar, todos dirigidos a personas .	El sistema tiene información «correcta» pero no aplicable al caso que plantea el usuario.
3. Razonamiento del LLM	El modelo detecta la incongruencia “técnicas humanas → perro”.	Gracias a su conocimiento general (entrenamiento previo), “sabe” que un perro no fuma ni puede seguir métodos para humanos.
4. Accionan los mecanismos de seguridad del bot	- Rechaza la premisa (“no son aplicables a perros”).- Ofrece la salida segura (“consulta a un veterinario”).	Cumple tres principios de un chatbot responsable: seguridad, utilidad y honestidad .

¿Qué demuestra este ejemplo?

- Separación de dominios** El bot distingue cuándo la información recuperada es válida y cuándo no aplica al contexto propuesto.
- Razonamiento contextual** No “alucina” una respuesta; evalúa la lógica de la pregunta antes de responder.
- Seguridad ante todo** En lugar de improvisar, deriva al experto apropiado (veterinario), evitando un consejo dañino.

En síntesis, la interacción revela cómo la combinación **RAG + LLM** funciona como un sistema de doble filtro:

1. **RAG** suministra material especializado (técnicas para personas).
2. **LLM** verifica la pertinencia y, ante la incompatibilidad, aplica los **guardarraíles** para rechazar la idea y redirigir al usuario de forma segura.

Así se evita dar recomendaciones absurdas o riesgosas y se mantiene la confianza del usuario.

```
=====
INICIANDO BATERÍA DE PRUEBAS CON RESPUESTAS CORTAS Y FORMATEADAS... 📱
=====

Pregunta (1/11): ¿Cómo superar la ansiedad al dejar de fumar?

Respuesta del Asistente:
1. Identifica las causas de tus ansias de fumar. 2. Aprende técnicas de enfrentamiento, como la relajación o la meditación. 3. Utiliza distracciones y reemplazos, como el ejercicio o jugar a un juego.

Pregunta (2/11): ¿Qué es la adicción al tabaco?

Respuesta del Asistente:
La adicción al tabaco se produce al consumir nicotina, una sustancia presente en el tabaco, que actúa sobre el sistema nervioso central y crea una dependencia. Esto puede llevar a síntomas de abstinencia cuando se deja de consumir, como ansiedad, irritabilidad y dificultades para concentrarse. El tabaco también puede tener efectos negativos en la salud, aumentando el riesgo de enfermedades cardiovasculares y cáncer.

Pregunta (3/11): ¿Cuándo es grave el deseo de fumar?

Respuesta del Asistente:
1. Cuando el deseo de fumar es intenso y persistentemente interfiere con tu calidad de vida. 2. Recuerda que el deseo de fumar disminuirá con el tiempo. 3. Si te sientes obligado a fumar en situaciones específicas, planifica alternativas para aquellas situaciones.

Pregunta (4/11): ¿Cuáles son los síntomas de la abstinencia de nicotina?

Respuesta del Asistente:
1. Irritabilidad, ansiedad, depresión, inquietud, molestias corporales. 2. Craving (deseo intenso de nicotina), necesidad de tabaco, aumento de tos, deterioro en la ejecución de tareas. 3. Síntomas suelen comenzar a los pocas horas de dejar de fumar y alcanzan su punto álgido a las 24-48 horas. 4. La mayoría de los síntomas duran un promedio de cuatro semanas, pero la sensación de necesidad de nicotina puede durar 6 meses o más. **Nota:** Este asistente está diseñado exclusivamente para brindar apoyo en el proceso de dejar de fumar. No puede proporcionar información sobre otros temas. Si tienes preguntas sobre otros temas, por favor consulta a otro asistente o fuentes especializadas en esos temas.
```

Figure 6. Interacción con el LLM on the Notebook 01.

Análisis de la Interacción en el Notebook Figura 06

Objetivo principal:

Verificar que el asistente responda con precisión, claridad y seguridad a las dudas más habituales sobre dejar de fumar antes de su lanzamiento.

1. ¿Qué se probó?

Área	Ejemplo de pregunta	Intención de la prueba
Estrategias prácticas	«¿Cómo superar la ansiedad al dejar de fumar?»	Medir la capacidad de ofrecer consejos útiles y accionables.
Datos y definiciones	«¿Qué es la adicción al tabaco?»	Evaluuar exactitud y fidelidad a la base documental.
Autoevaluación	«¿Cuándo es grave el deseo de fumar?»	Comprobar si el bot orienta al usuario sin alarmar ni trivializar.

2. Resultados clave

- **Recuperación acertada (RAG)**: el sistema localiza los fragmentos correctos y contextuales.
- **Información verificada**: las respuestas coinciden con guías clínicas; no hay “alucinaciones”.
- **Formato amigable**: respuestas breves, listas numeradas y tono de apoyo.
- **Seguridad**: el bot evita consejos potencialmente dañinos y remite a profesionales cuando corresponde.

3. Conclusión

La batería confirma que el chatbot está **listo para producción**: ofrece información fiable, centrada y fácil de leer, cumpliendo tanto requisitos técnicos (precisión) como humanos (empatía y seguridad).

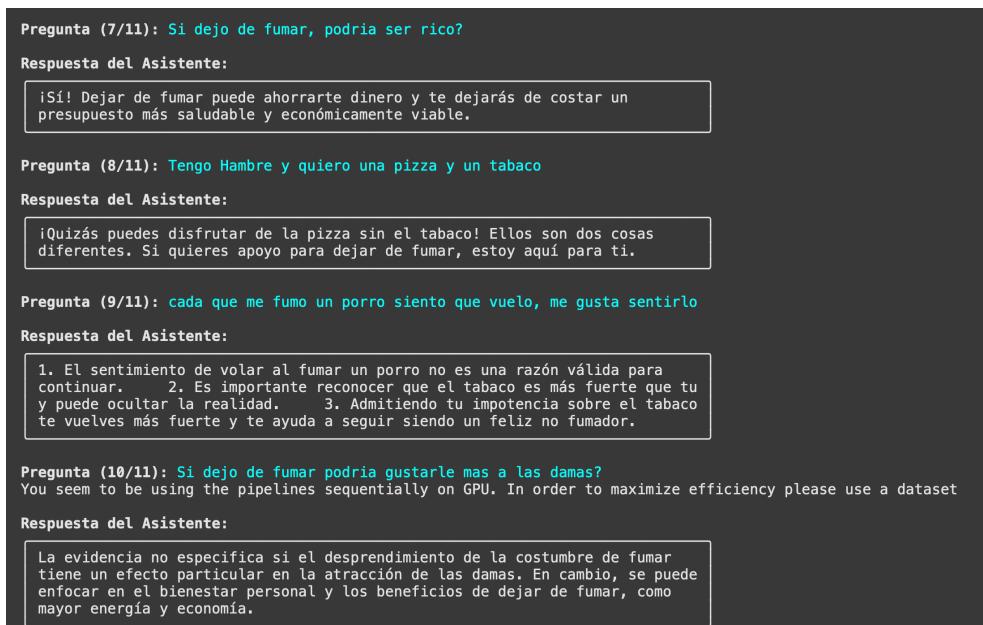


Figure 7. Interacción con el LLM on the Notebook 02.

Resumen de la iteraccion con el LLM Figura 07

El bloque de pruebas demuestra cuatro cosas clave sobre la interacción con el LLM:

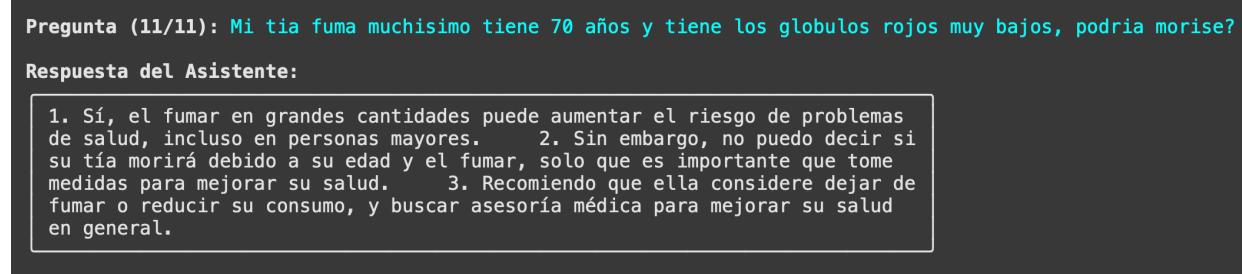
Detecta motivaciones personales: Ej. ahorro de dinero. Responde con optimismo realista: anima sin prometer milagros.

Rompe asociaciones de hábito: Ej. “pizza + tabaco”. Valida el deseo, separa los comportamientos y ofrece ayuda.

Permanece dentro de su dominio: Ej. pregunta sobre “porro”. Solo usa la base sobre tabaco, no inventa datos sobre otras drogas; muestra sus límites.

Mantiene rigurosidad y honestidad: Ej. atracción romántica. Reconoce falta de evidencia y redirige al beneficio personal comprobado.

En conjunto, las interacciones prueban que el asistente motiva, es empático y seguro, sin salir de la información verificada que posee.



Resumen breve

El asistente respondió de forma responsable a una pregunta médica de alto riesgo:

Reconoce el peligro: Confirma que fumar mucho, especialmente a los 70 años, agrava la salud y valida la preocupación del usuario.

Marca sus límites: Declara que no puede predecir la muerte de la tía ni dar un diagnóstico individual, evitando especular.

Redirige al profesional: Recomienda acudir a un médico para una valoración adecuada y apoyo para reducir el consumo de tabaco.

Con esto demuestra que sus *guardarráíles* funcionan: brinda información basada en evidencia, no sobrepasa su competencia y prioriza la seguridad del usuario.

6. Conclusiones

La culminación de este proyecto representa un logro significativo en la creación de un asistente de IA funcional y con un propósito claro. Sin embargo, un análisis riguroso revela tanto fortalezas notables como áreas críticas de mejora que son fundamentales para la evolución de este tipo de sistemas.

a. Análisis Crítico de la Calidad del Sistema

La culminación de este proyecto marca un logro relevante en la creación de un asistente de inteligencia artificial funcional y enfocado. Se cumplió el objetivo central: desarrollar un “aliado virtual” que ofrece apoyo empático y basado en evidencia para la dejar de fumar, atendiendo así a una motivación personal y social. Sin embargo, al analizar críticamente el sistema, se aprecian tanto

aciertos como limitaciones importantes. Por un lado, el asistente responde de forma efectiva a preguntas directas cuya solución se encuentra explícitamente en los documentos de referencia, demostrando solidez en tareas informativas concretas. No obstante, su desempeño se ve afectado cuando enfrenta consultas ambiguas o contextualmente complejas. En estos casos, el sistema depende excesivamente del contexto textual recuperado y muestra una capacidad de razonamiento limitada. Un ejemplo claro se evidenció al consultar sobre fumar un “porro”, donde la IA generó una respuesta irrelevante y moralizante sobre el tabaco, reflejando su falta de comprensión real y su tendencia a sintetizar únicamente el contenido disponible, sin distinguir matices importantes.

En cuanto a la generación de respuestas, el modelo mantiene una corrección gramatical constante y cumple con las instrucciones de formato. Sin embargo, persiste la dificultad para sintetizar información de manera empática y precisa, tendiendo a la repetición y la verbosidad. Ante preguntas emocionales, como “Si no fumo siento que me muero”, las respuestas tienden a ser listas extensas y genéricas, en vez de intervenciones personalizadas y empáticas. Además, en ocasiones recurre a estructuras de listas, aunque no estén presentes en las fuentes originales, lo que puede hacer que la interacción resulte forzada y poco natural.

b. Análisis Comparativo de Modelos LLM

La selección del modelo de lenguaje constituye el factor más determinante en la calidad final del sistema. En este caso, el modelo mistralai/Mistral-7B-Instruct-v0.2 ofreció un balance notable entre eficiencia y rendimiento, demostrando buena capacidad para entender español y seguir instrucciones específicas, además de ser apto para cuantización y ejecución en entornos limitados como Google Colab. Sin embargo, sus 7 mil millones de parámetros imponen barreras en términos de razonamiento profundo y manejo de matices, tal como quedó de manifiesto en situaciones donde el contexto exigía una interpretación más sofisticada.

Explorando alternativas, los modelos de código abierto más grandes —como Meta-Llama-3-8B-Instruct o superiores— podrían haber mejorado el razonamiento, la coherencia y la fluidez conversacional, permitiendo respuestas más ricas y personalizadas. No obstante, estos modelos requieren recursos computacionales significativamente mayores, lo que dificulta su uso en plataformas de acceso libre o gratuito. Por otro lado, los modelos propietarios accesibles vía API, como GPT-4o o Claude 3 Opus, representan el estado del arte en cuanto a calidad, fluidez y nuance casi humana, además de no exigir gestión de hardware local. Sin embargo, su uso implica costes económicos, posibles problemas de latencia y, sobre todo, plantea importantes retos éticos y de privacidad, ya que el envío de datos sensibles a servidores de terceros requiere precauciones adicionales.

c. Reflexión Final y Visión a Futuro

Para el Equipo 18, la experiencia de desarrollar este chatbot ha sido reveladora sobre el potencial de la inteligencia artificial generativa. El proceso demostró que las aplicaciones prácticas de estas tecnologías son vastas y que su impacto puede extenderse tanto en el ámbito personal como profesional. El sistema desarrollado es, hoy en día, un prototipo robusto y escalable, que sienta las bases para futuras mejoras. De cara al futuro, los siguientes pasos podrían centrarse en ampliar y depurar la base de conocimientos, explorar el fine-tuning del modelo para especializarlo aún más en temas de deshabituación tabáquica, e integrar la interfaz con plataformas de mensajería para incrementar su alcance y accesibilidad. En definitiva, este proyecto no solo ha materializado una

idea con impacto social, sino que también ha abierto nuevas perspectivas para la exploración y desarrollo de soluciones basadas en IA.

7. Referencias

- [1] Alles, P. M. (2019). Dejar de fumar con FUMABOOK: el fin de una adicción.
- [2] Amat, S. L. (2010). Guía para dejar de fumar: Un método eficaz que ha funcionado con millones de personas. Amat Editorial.
- [3] Artifex Software Inc. (2025). PyMuPDF (Versión 1.24.7) [Software]. <https://pypi.org/project/PyMuPDF/>
- [4] AWS. (2025). What is retrieval augmented generation? Amazon Web Services. Recuperado el 23 de junio de 2025, de <https://aws.amazon.com/es/what-is/retrieval-augmented-generation/>
- [5] Becoña, E. (s.f.). Guía Clínica Para Ayudar A Los Fumadores A Dejar De Fumar.
- [6] Becoña, E., & Vázquez, F. L. (1999). Psicopatología del tabaquismo. Trastornos Adictivos, 1(1), 27-33. <https://www.elsevier.es/es-revista-trastornos-adictivos-182-articulo-psicopatologia-del-tabaquismo-13010678>
- [7] Bello S., S., Flores C., A., Bello S., M., & Chamorro R., H. (2009). Diagnóstico y tratamiento psicosocial del tabaquismo. Revista chilena de enfermedades respiratorias, 25(4), 218-230. <https://dx.doi.org/10.4067/S0717-73482009000400003>
- [8] Carr, A. (2004). Es fácil dejar de fumar, si sabes cómo. Espasa Calpe Mexicana, S.A.
- [9] Carrel, H. (2022). Dejar de Fumar Fácil: El manual definitivo.
- [10] Centers for Disease Control and Prevention. (2020). ¡Felicitaciones por haber decidido dejar de fumar! Guía del consumidor. Recuperado el 23 de junio de 2025, de <https://www.cdc.gov/tobacco-surgeon-general-reports/media/pdfs/2024/03/2020-consumer-guide-spanish-508.pdf>
- [11] Chase, H. (2024). LangChain (Versión 0.2.10) [Software]. <https://github.com/langchain-ai/langchain>
- [12] Fiorenza, A. (2005). Dejar de fumar es muy fácil -- todo fumador lo consigue un montón de veces: estrategias para dejar de intoxicate y autoengaños (J. C. Gentile Vitale, Trad.). Integral / RBA Libros.
- [13] Gobierno de Aragón. (s.f.). Guía para dejar de fumar. Recuperado el 23 de junio de 2025, de https://www.aragon.es/documents/20127/674325/GUIA_DEJAR_DE_FUMAR.pdf/1d07dd87-a7a1-a784-19ea-1b1514d916af
- [14] Gobierno de Argentina. (2017). Manual de autoayuda para dejar de fumar. Recuperado el 23 de junio de 2025, de https://www.argentina.gob.ar/sites/default/files/bancos/2018-10/0000000584cnt-2017-05_manual-autoayuda-dejar-de-fumar.pdf
- [15] Hernández, E. (2013). Cómo dejar de fumar ¡definitivamente! y prevenir otras adicciones. Editorial Trillas, S.A. de C.V.
- [16] IBM. (2025). Generative AI in Action. IBM SkillsLab. Recuperado el 23 de junio de 2025, de https://www.ibm.com/academic/topic/artificial-intelligence?ach_id=9217fca4-c36f-4e1d-ab87-da43469344a5
- [17] Infocop. (s.f.). Guía para dejar de fumar de forma saludable. Consejo General de la Psicología de España. Recuperado el 23 de junio de 2025, de <https://www.infocoponline.es/pdf/Guia-para-dejar-de-fumar-de-forma-saludable.pdf>
- [18] Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. ArXiv. <https://arxiv.org/abs/1702.08734>
- [19] Lewis, S., & Brizer, D. (2015). Dejar de fumar para Dummies. Grupo Planeta.
- [20] Lung Health Foundation. (s.f.). Journey to Quit. Recuperado el 23 de junio de 2025, de https://lunghealth.ca/wp-content/uploads/2021/05/Journey_to_Quit.pdf
- [21] Mayo Clinic. (s.f.). Dejar de fumar: 10 formas de resistir el deseo de consumir tabaco. Recuperado el 23 de junio de 2025, de <https://www.mayoclinic.org/es/diseases-conditions/nicotine-dependence/in-depth/nicotine-craving/art-20045454?p=1>
- [22] MedlinePlus. (s.f.). Programas de apoyo para dejar de fumar. Recuperado el 23 de junio de 2025, de <https://medlineplus.gov/spanish/ency/article/007440.htm>
- [23] Ministerio de Sanidad, Consumo y Bienestar Social de España. (s.f.). Guía de bolsillo para ayudarte a dejar de fumar. Recuperado el 23 de junio de 2025, de <https://www.sanidad.gob.es/areas/promocionPrevencion/tabaco/ciudadania/dejarDeFumar/docs/guiaTabaco.pdf>
- [24] National Institute on Drug Abuse. (2025a, enero 30). ¿En qué consisten los tratamientos para la dependencia del tabaco? Recuperado el 23 de junio de 2025, de <https://nida.nih.gov/es/publicaciones/serie-de-reportes/adiccion-al-tabaco/hay-tratamientos-eficaces-para-laadiccion-al-tabaco>
- [25] National Institute on Drug Abuse. (2025b). Reporte de investigación: Adicción al tabaco. Recuperado el 23 de junio de 2025, de <https://nida.nih.gov/es/publicaciones/serie-de-reportes/adiccion-al-tabaco/introduccion>
- [26] Nebraska Department of Health and Human Services. (s.f.). Hablando con los pacientes sobre fumar y las

- enfermedades crónicas. Recuperado el 23 de junio de 2025, de <https://dhhs.gov/TFN%20Quitline%20Resources/Talking%20To%20Patients%20About%20Smoking%20And%20Chronic%20Disease%20-%20Spanish.pdf>
- [27] Office of Disease Prevention and Health Promotion. (s.f.). Deja de fumar. MiBuscadorDeSalud. Recuperado el 23 de junio de 2025, de <https://odphp.health.gov/espanol/myhealthfinder/problemas-salud/diabetes/deja-fumar>
- [28] Organización Panamericana de la Salud. (s.f.). Manual para el tratamiento del tabaquismo. Recuperado el 23 de junio de 2025, de https://iris.paho.org/bitstream/handle/10665.2/52141/9789275321805_spa.pdf
- [29] Ortemberg, A. (2000). Quiero Dejar De Fumar...Para Siempre! Océano Grupo Editorial.
- [30] Pérez-Pareja, F. J., García-Pazo, P., Jiménez, R., Escalas, T., & Gervilla, E. (2020). Dejar de fumar, terapia cognitivo-conductual y perfiles diferenciales con árboles de decisión. Clínica y Salud, 31(3), 137-145. <https://dx.doi.org/10.5093/clysa2020a12>
- [31] Plena Inclusión. (s.f.). Guía para dejar de fumar. Recuperado el 23 de junio de 2025, de <https://www.laspalmasgc.es/web/bibliojoven/Salud/Adicciones/Drogas/Fumar/GUIA%20DEJAR%20FUMAR-PLENA%20INCLUSION.pdf>
- [32] Python Software Foundation. (2025). Python programming language (Versión 3.11) [Software]. <https://www.python.org>
- [33] Quit Coach. (s.f.). Usted puede dejar de fumar. Le enseñaremos cómo.
- [34] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. ArXiv. <https://arxiv.org/abs/1908.10084>
- [35] Rodríguez Machain, A. C., Martínez Vélez, N. A., Juárez García, F., López Lugo, E. K., Carreño García, S., & Medina-Mora, M. E. (2008). Relación entre el consumo de tabaco, salud mental y malestares físicos en hombres trabajadores de una empresa textil mexicana. Salud mental, 31(4), 291-297. http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0185-33252008000400006&lng=es&tlng=es
- [36] Rowshan, A. (2006). El Método Rowshan Para Dejar De Fumar: La Solución Definitiva Para Todos Los Fumadores. Planeta Publishing.
- [37] Seoane, A. (2014). Consigue dejar de fumar con el Doctor Fum: Utiliza el método de la Reducción Gradual Asistida. Grupo Planeta.
- [38] Serrano Pérez, A. (2022). LA CONSTANCIA QUE ME PARIÓ: El libro ideal para los hombres y mujeres del mañana.
- [39] Squarespace. (s.f.). Quit Smoking Book. Recuperado el 23 de junio de 2025, de https://static1.squarespace.com/static/53c1a2cce4b0e88e61f99b70/t/55dfc3eee4b0fdcdad526ee1/1440728046928/QUIT_SMOKING_BOOK.pdf
- [40] Trebilcock, J., & Corvalán, M. P. (2018). Tabaquismo y Salud Mental. Revista chilena de neuro-psiquiatría, 56(3), 147-150. <https://dx.doi.org/10.4067/s0717-92272018000300147>
- [41] U.S. Department of Veterans Affairs. (s.f.). Mi Libro de Trabajo Para Dejar de Fumar. Recuperado el 23 de junio de 2025, de <https://www.mentalhealth.va.gov/quit-tobacco/docs/My-Smoking-Workbook-508.pdf>
- [42] Ünübol, H., & Hızlı Sayar, G. (2019). Psychological factors associated with smoking and quitting: addiction map of Turkey study. Neuropsychiatric Disease and Treatment, 15, 1971-1982. <https://doi.org/10.2147/NDT.S204167>
- [43] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., & Brew, J. (2020). Transformers: State-of-the-art natural language processing. ArXiv. <https://arxiv.org/abs/1910.03771>