

# **LicitBot** Un Chatbot de Análisis de datos en Licitaciones

**Judith Rojas – Miguel Silva**

**Profesor: Germain Garcia**



# Índice

1. Descripción del Problema
  - Área: Análisis de Datos con IA Generativa
  - Problema: Análisis de datos estructurados
2. Trabajos Relacionados
3. Producto Propuesto : LicitBot
4. Experimentación y Resultados
5. Conclusiones
6. Trabajos Futuros





# 1. Introducción



# Área: *Análisis de datos con IA Generativa*

El análisis de datos tradicional es un obstáculo para generar Insights



¿Tienes el reporte que  
te pedí?  
Quisiera saber esto ....y  
también lo otro ...?

	Número ID	Email / Teléfono	Dirección	N° Documento	Imponible	IVA	Totale
10-01-2021	1002585428		Fantasy World	FAC 2020579	176.75	21.21	197.96
10-01-2021	0702162181001	+39565656	Via del Polo Norte y Southampton	FAC 2020580	176.75	21.21	197.96
10-01-2021	1203806094		Ave del Bar el Tufo - Pellotillehue	FAC 2020581	27.65	3.32	30.97
10-01-2021	1002585428		Fantasy World	NCR 2020345	-176.75	-21.21	-197.96
13-01-2021	0911071017		.	FAC 2020582	176.75	21.21	197.96
13-01-2021	1718684895		Televisa San Angel - Mexico	FAC 2020583	27.65	3.32	30.97
13-01-2021	1002367058		.	FAC 2020584	27.65	3.32	30.97

# Problema: *Algunas Definiciones*

<b>OSCE</b>	“Organismo Supervisor de las Contrataciones del Estado” es una entidad pública del gobierno peruano encargada de supervisar, regular, y promover la transparencia y eficiencia en las contrataciones públicas
<b>SEACE</b>	"Sistema Electrónico de Contrataciones del Estado" y es una plataforma electrónica en línea utilizada en Perú para la gestión y publicación de todos los procesos de contratación pública
<b>LICITACION</b>	Es un proceso mediante el cual una entidad pública o privada solicita propuestas o ofertas de diferentes proveedores interesados para la adquisición de bienes, contratación de servicios o ejecución de obras
<b>RNP</b>	"Registro Nacional de Proveedores". Es un registro administrado por el Organismo Supervisor de las Contrataciones del Estado (OSCE) que centraliza y gestiona la información de los proveedores interesados en participar en procesos de contrataciones públicas en el país.
<b>CARTA FIANZA</b>	Una carta fianza es un instrumento financiero emitido por una entidad bancaria o una compañía de seguros, que garantiza el cumplimiento de una obligación financiera o contractual por parte de una persona o entidad (llamada el principal) hacia otra parte (llamada el beneficiario)
<b>FIDEICOMISO</b>	Un fideicomiso es un contrato legal en el cual una persona (llamada fideicomitente o fideicomitente) transfiere la propiedad de ciertos activos, bienes o derechos a otra persona o entidad (llamada fiduciario), quien se encarga de administrarlos y gestionarlos en beneficio de una tercera parte (llamada beneficiario o fideicomisario)
<b>OLA</b>	Es una proyección de licitaciones a darse en el mes, es decir, cuantos procesos deberían asignarles fecha de buena pro, sirve para que se observe el comportamiento en el mercado y ver entidades que tienen procesos más fuertes.
<b>FECHA DE BUENA PRO</b>	Se refiere al momento específico en el cual una entidad pública o privada adjudica formalmente un contrato o licitación a uno de los participantes, declarándolo como el ganador del proceso de selección

# **Problema:** *Actualmente..*

## **Proceso Manual**

1. Se realiza búsquedas manuales de las licitaciones publicadas en el SEACE
2. Se revisa la ficha de cada licitación para saber los postores que están participando.
3. Se ingresa al portal de RNP y se realiza la búsqueda de los participantes para captar sus teléfonos y correos electrónico

***Son miles de licitaciones, es difícil llevar un control y mantener actualizada la información de Licitaciones y sus participantes, para proyectar potenciales clientes.***

# Problema

¿Cómo optimizar la captación de participantes que postulan a licitaciones con el Estado peruano?

The screenshot shows the top navigation bar with the SE@CE logo (Sistema Electrónico de Contrataciones del Estado) and the SCE logo (Organismo Supervisor de las Contrataciones del Estado). Below the navigation bar, there are several search filters: 'Buscador de Procedimientos de Selección', 'Buscador por Expediente Tribunal', 'Buscador de Expresiones de', 'Buscador Público de Órdenes de Compra y Órdenes de Servicio', and 'Buscador de Condiciones de Contratación'. A red banner across the middle of the page reads 'gob.pe | Buscador de Proveedores del Estado'. Below this banner, the main heading is 'Búsqueda de Proveedores del Estado'. At the bottom, there is a search input field with the placeholder text 'Buscar...' and a magnifying glass icon.



De Interés para Analistas  
o Bróker de Seguros que ofrecen  
Cartas Fianza y/o Fideicomisos



# 2.

## Tabajos Relacionados

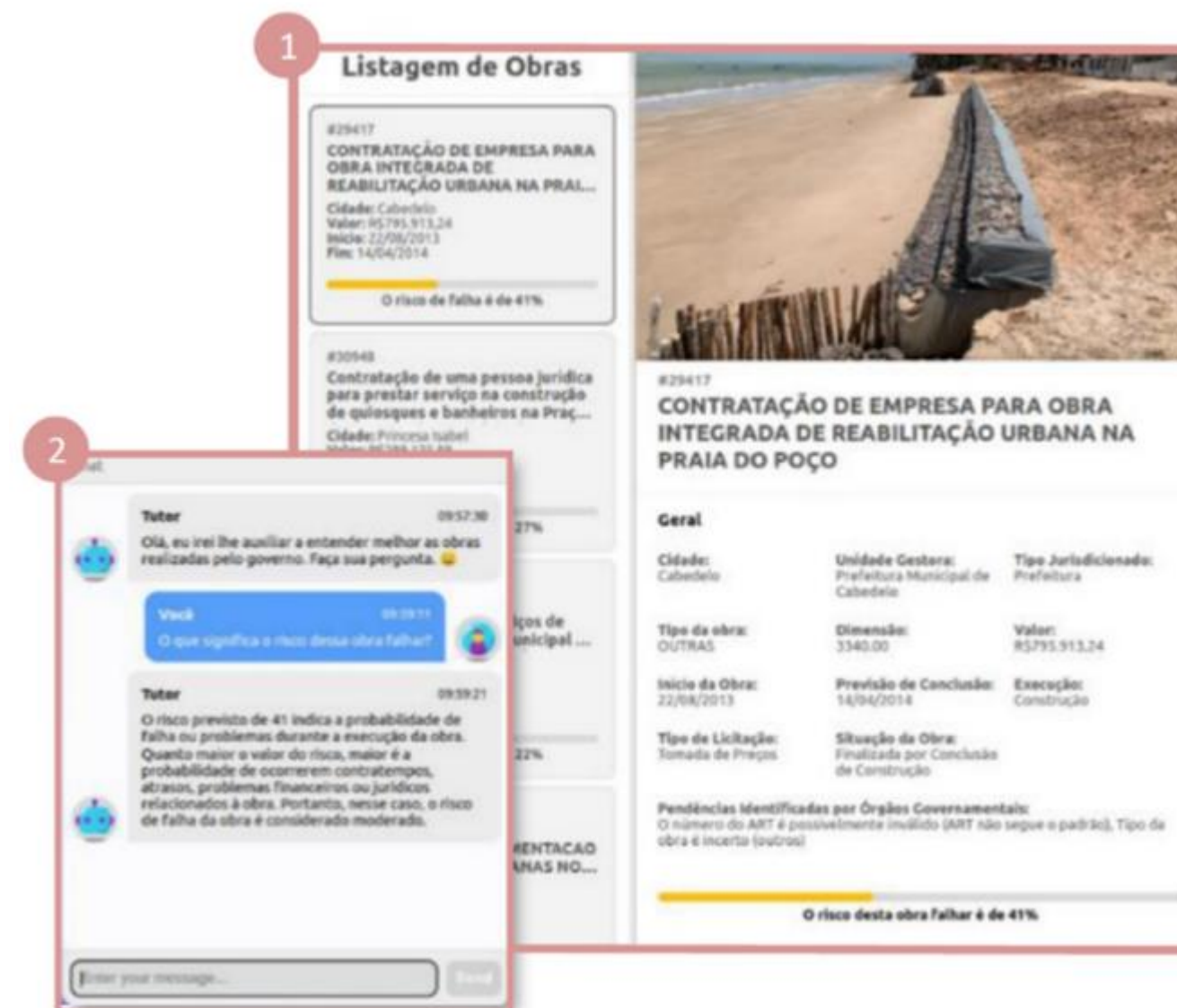


# Chat OSCE



Organismo Supervisor de las Contrataciones del Estado (OSCE). (2024).  
Página de consulta para el chat del OSCE.  
<https://consultas.osce.gob.pe/chat/>

# Large Language Models in Civic Education on the Supervision and Risk Assessment of Public Works



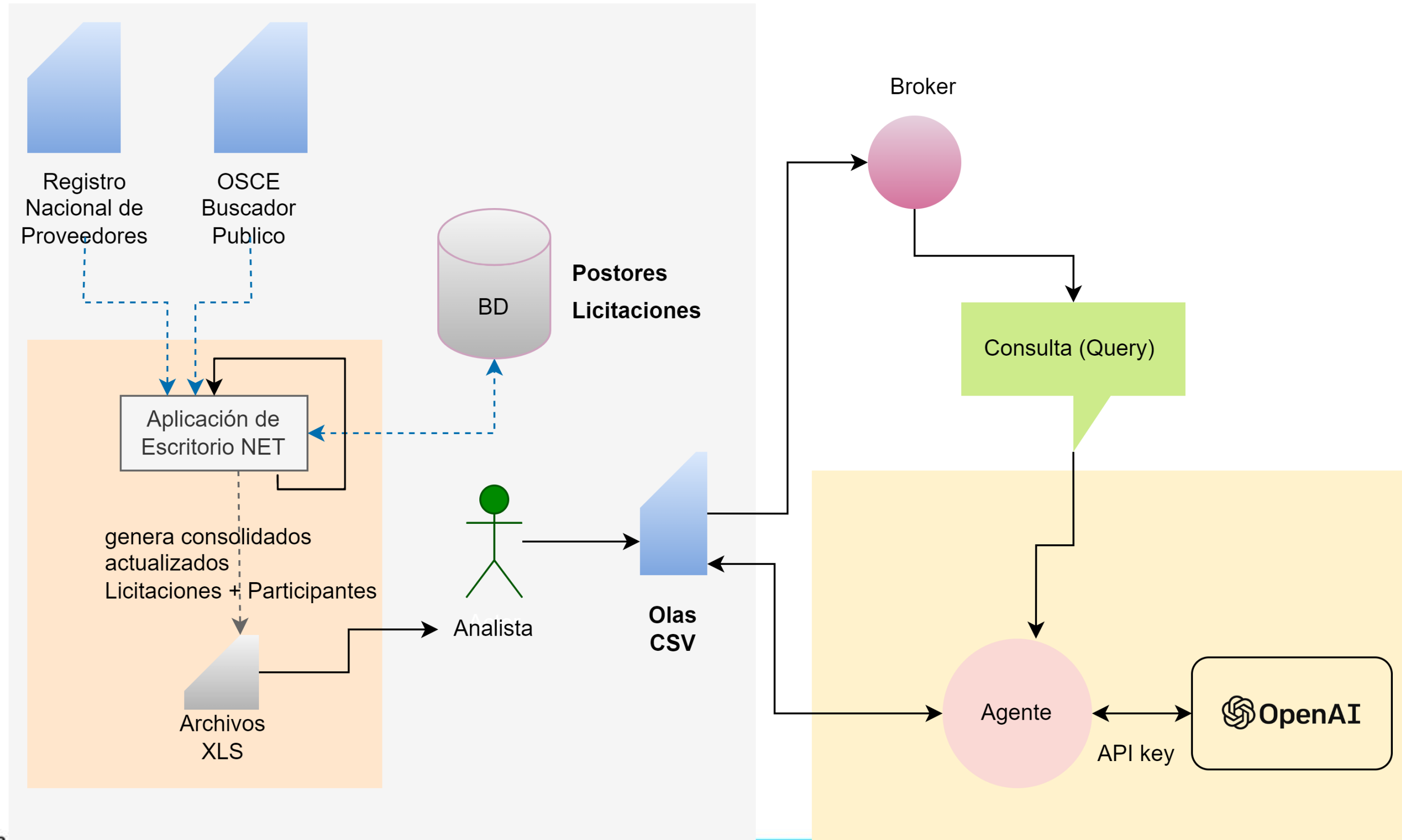
Honório, J. J. C. M., Brito, P. C. O., Moura, J. A. B., & Andrade, N. F. (2024). Large Language Models in Civic Education on the Supervision and Risk Assessment of Public Works. Graduate Program in Computer Science, Systems and Computing Department, Federal University of Campina Grande (UFCG), Brazil. <https://www.scitepress.org/Papers/2024/125893/125893.pdf>



# 3. Producto Propuesto

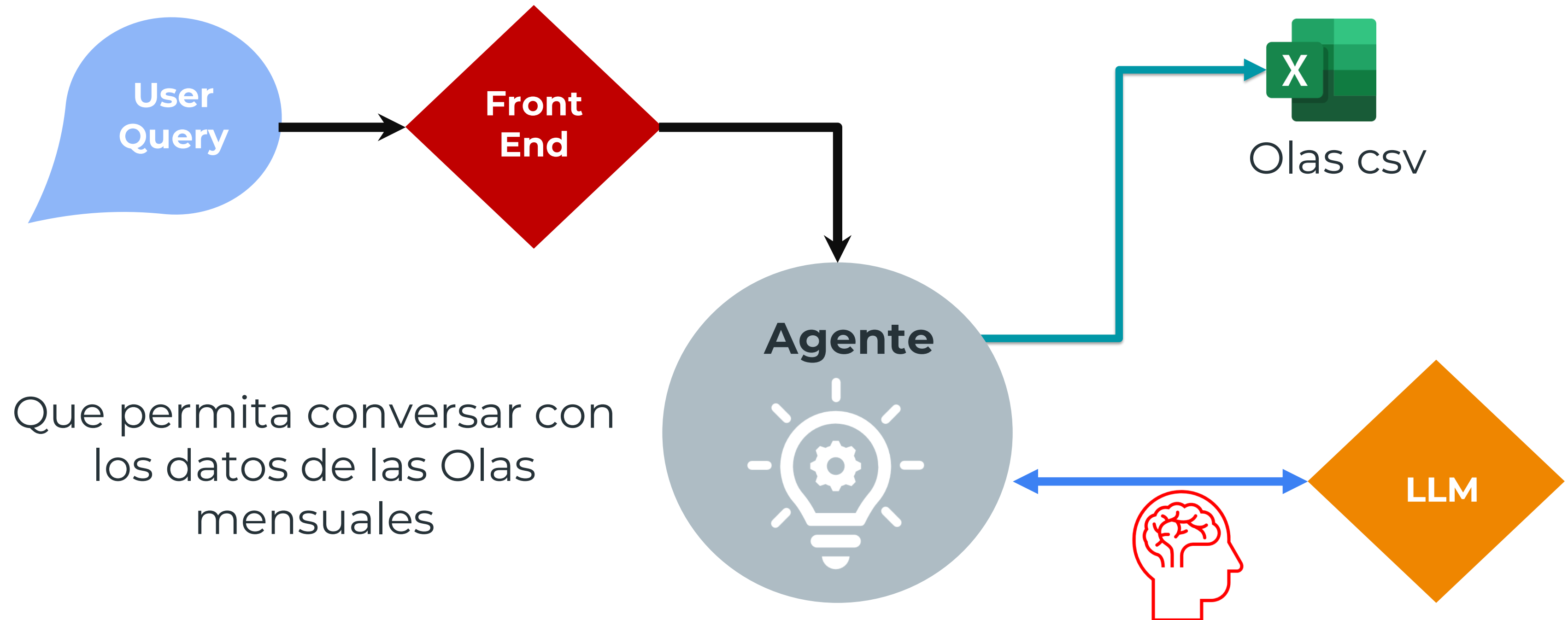


# Propuesta: Diagrama de Software



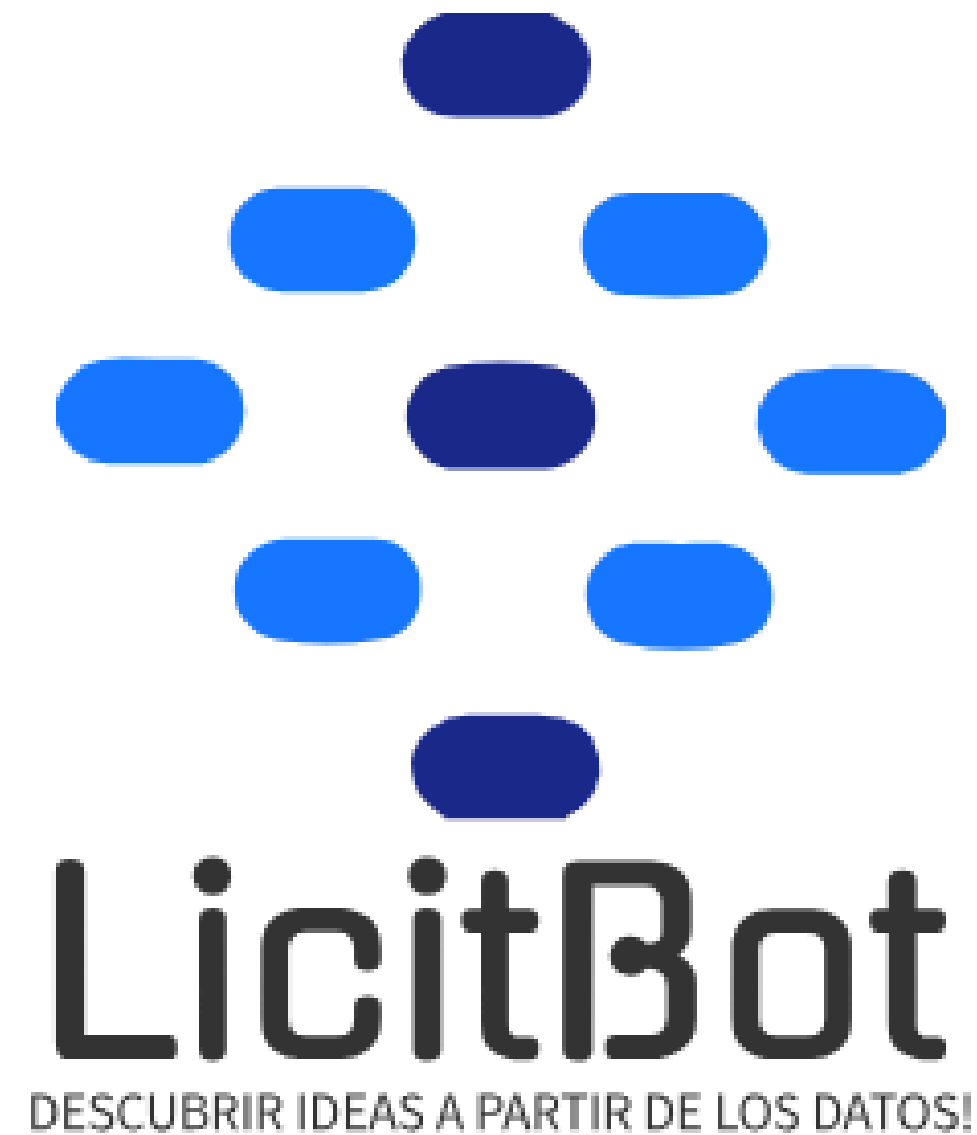
# Propuesta

Proponemos un ChatBot utilizando LLM y Agentes Langchain





# Producto Propuesto: Nace...



## LicitBot - ChatBot de análisis

Descubrir ideas a partir de los datos!

Este ChatBot fue creado para responder preguntas a partir de Olas de Licitaciones. Haga una pregunta y el chatbot responderá con el análisis apropiado.



[cjudithrb/licitbot \(github.com\)](https://github.com/cjudithrb/licitbot)

# Producto Propuesto: *Idea de Negocio*



## Branding



# DataSet: *Extracción de datos*

## De la OSCE....

Página desde donde consultamos datos	Enlaces de Extracción
Página Principal del Contrataciones del Estado <b>Se capta las licitaciones diarias</b>	<a href="https://prod2.seace.gob.pe/seacebus-uiwd-pub/buscadorPublico/buscadorPublico.xhtml">https://prod2.seace.gob.pe/seacebus-uiwd-pub/buscadorPublico/buscadorPublico.xhtml</a>
Página Fichas Electrónicas de cada Licitación <b>Se capta Lista de Postores y Cronograma por Licitacion</b>	<a href="https://prod2.seace.gob.pe/seacebus-uiwd-pub/fichaSeleccion/fichaSeleccion.xhtml?ptoRetorno=LOCAL">https://prod2.seace.gob.pe/seacebus-uiwd-pub/fichaSeleccion/fichaSeleccion.xhtml?ptoRetorno=LOCAL</a>
Página de Proveedores con Observaciones <b>Se capta datos de cada Proveedor Postulante (Postor)</b>	<a href="https://prod2.seace.gob.pe/seacebus-uiwd-pub/fichaSeleccion/fichaListaConsultasElectronicas.xhtml">https://prod2.seace.gob.pe/seacebus-uiwd-pub/fichaSeleccion/fichaListaConsultasElectronicas.xhtml</a>

Código captcha (*)	<input type="text"/>	
Nombre y Firma de		

Para obtener el código Captcha usamos la librería TESSRACT de código abierto para el reconocimiento óptico de caracteres (OCR, por sus siglas en inglés)

[tesseract-ocr/tesseract: Tesseract Open Source OCR Engine \(main repository\) \(github.com\)](https://github.com/tesseract-ocr/tesseract)

# DataSet: Extracción de datos

## ...De la RNP

<https://eap.osce.gob.pe/perfilprov-bus/1.0/ficha/>

https://apps.osce.gob.pe/perfilprov-ui/ficha/20532554005

**gob.pe** | Buscador de Proveedores del Estado

Inicio > Búsqueda > Ficha Única del Proveedor

CONSORCIO GRAÑA S.A.C.

RUC (\*) : 20532554005

Teléfono (\*) : 952823482

Email (\*) : consorciograna@hotmail.com

Domicilio (\*\*):

```

1 {
2   "resultadoT01": {
3     "codigo": "00",
4     "mensaje": "Procesamiento completado.",
5     "nivel": 5,
6     "mensajes": {
7       "data": null
8     }
9   },
10  "dataInfo": {
11    "updatedTo": "2024-07-14T05:00:00.000+0000"
12  },
13  "proveedorT01": {
14    "codProv": "20532554005",
15    "idOrigenProv": 10,
16    "numRuc": "20532554005",
17    "nomRzsProv": "CONSORCIO GRAÑA S.A.C.",
18    "esHabilitado": false,
19    "lscIdTipReg": "",
20    "lscIdTipRegVig": "",
21    "esAptoContratar": false,
22    "emails": [
23      "consorciograna@hotmail.com"
24    ],
25    "telefonos": [
26      "952823482"
27    ],
28    "cmcTexto": null,
29    "espProvT01s": [],
30    "espProvMsj": 1,
31    "tipoPersoneria": 2
32  },

```

Extraemos los datos de Postores a las licitaciones desde el Buscador de proveedores del Estado



# Dataset: *Recolección de datos*



Información Recolectada
Enero 2024
Febrero 2024
Marzo 2024
Abril 2024
Mayo 204
Junio 2024
<b>6792 registros</b>

Datos de Licitación	
• Entidad	• Fecha Fin de observaciones
• Fecha y hora de Publicación	• Fecha de Absolución de Observaciones
• Nomenclatura	• Integración de las Bases
• Objeto de Contratación	• Fecha Presentacion de Propuesta
• Descripción del Objeto	• Fecha de Buena Pro
• Valor Referencial	• Estado
• Moneda	• Tiene Fideicomiso

Datos de Postores a la Licitación	
• Ruc	• Razón Social
• Tipo de Formulario	• Teléfono
• Email	

Datos de un Archivo de Ola

# Herramientas utilizadas para Scrapping:



HTML Agilite Pack

Tesseract.Net SDK



NPOI Net





# Herramientas utilizadas para ChatBot:



- Marco de aplicación de código abierto.
- Convierte los scripts de datos en aplicaciones web que se pueden compartir en minutos.
- Todo en Python puro.
- No se requiere experiencia en front-end



- Framework de código abierto para el desarrollo de aplicaciones que utilizan modelos de lenguaje grande (LLM)
- Disponibles en bibliotecas basadas en Python y Javascript
- Simplifican la creación de aplicaciones impulsadas por LLM, como chatbots y agentes virtuales



- Permite integrar la inteligencia artificial avanzada de OpenAI en aplicaciones
- Disponibles en bibliotecas basadas en Python y Javascript
- Facilita tareas como procesamiento de lenguaje natural, generación de texto, y análisis de datos mediante sencillas llamadas a su servicio web



# 4. Experimentación



# Chatbot Conversacional

Documento de Olas CSV



Crear Embedding y  
Almacenar VectorBD

```
embeddings = OpenAIEmbeddings(openai_api_key=openai_api_key)
db = FAISS.from_documents(split_documents, embeddings)
db.save_local("faiss_index")
```

index.faiss	14/07/2024 04:25 p. m.	FAISS File	40,207 KB
index.pkl	14/07/2024 04:25 p. m.	PKL File	2,166 KB

Desarrollar Chat  
Conversacional

```
query_chain = ConversationalRetrievalChain.from_llm(
    llm=llm,
    retriever=new_db.as_retriever(),
    return_source_documents=True,
    memory=memory,
    verbose=True
)
```

```
#return query_chain({"question": question, "chat_history": chat_hi
docs = new_db.similarity_search(question)
```

```
# Concatenar los contenidos de todos los documentos
answer = " ".join([doc.page_content for doc in docs])
```

Construir Streamlit UI

Inconveniente con el consumo del Open API



Por favor ingrese su consulta

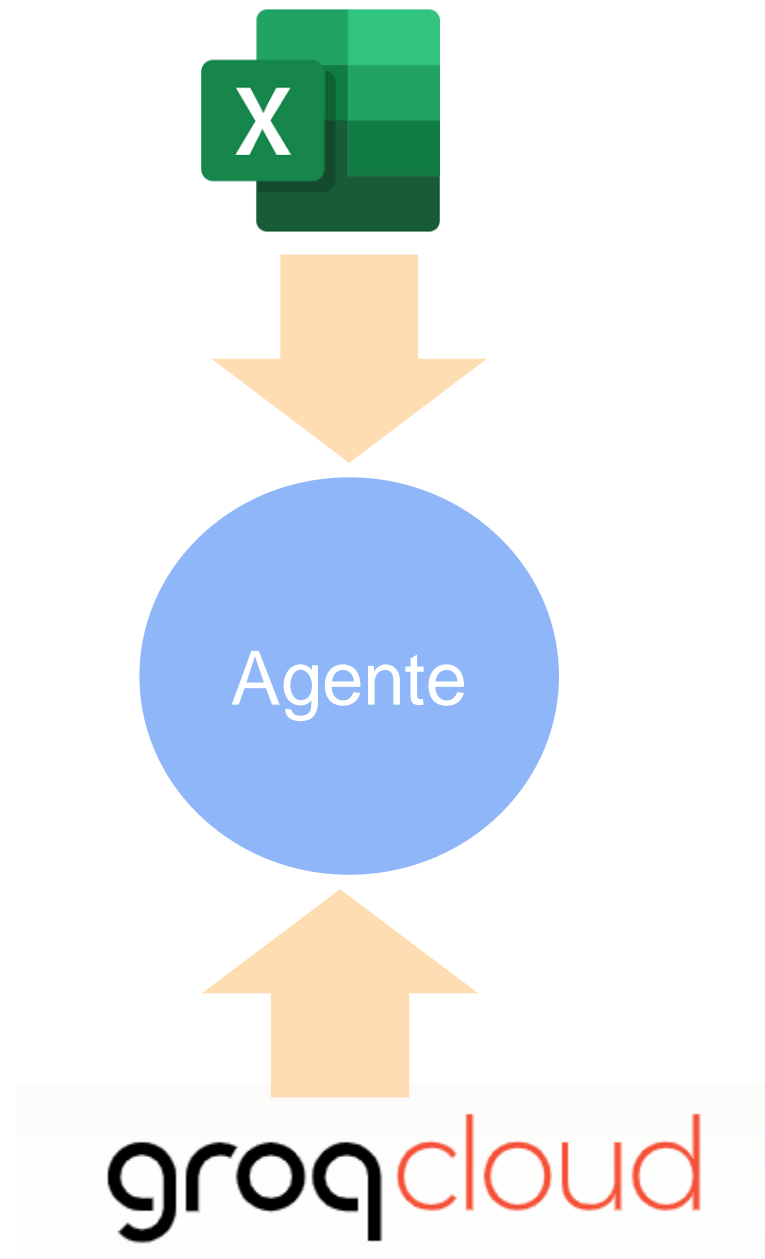
cual es el telefono email de MEGARCON CONTRATISTAS GENERALES SAC

Estado: Convocado Fideicomiso: No Ruc: 20518861604 Razon Social: MEGARCON CONTRATISTAS GENERALES SAC Tipo de Formulario: Consulta Telefono: 2753640-967274617 Email: [megarconsac@hotmail.com](mailto:megarconsac@hotmail.com) Ola: Enero 2024 Estado: Convocado Fideicomiso: No Ruc: 20518861604 Razon Social: MEGARCON CONTRATISTAS GENERALES SAC Tipo de Formulario: Consulta Telefono: 2753640-967274617 Email: [megarconsac@hotmail.com](mailto:megarconsac@hotmail.com) Ola: Enero 2024 Estado: Convocado Fideicomiso: No Ruc: 20518861604 Razon Social: MEGARCON CONTRATISTAS GENERALES SAC Tipo de Formulario: Consulta Telefono: 2753640-967274617 Email: [megarconsac@hotmail.com](mailto:megarconsac@hotmail.com) Ola: Enero 2024 Estado: Convocado Fideicomiso: No Ruc: 20518861604 Razon Social: MEGARCON CONTRATISTAS GENERALES SAC Tipo de Formulario: Consulta Telefono: 2753640-967274617 Email: [megarconsac@hotmail.com](mailto:megarconsac@hotmail.com) Ola: Enero 2024

Ingrese su Consulta: |



# Agentes y LLM Open Source



```

from langchain.agents.agent_types import AgentType
from langchain_experimental.agents.agent_toolkits import create_csv_agent
from langchain_groq import ChatGroq
import streamlit as st
  
```

```
llm = ChatGroq(temperature=0, model="llama3-70b-8192", api_key=api_key)
```

```

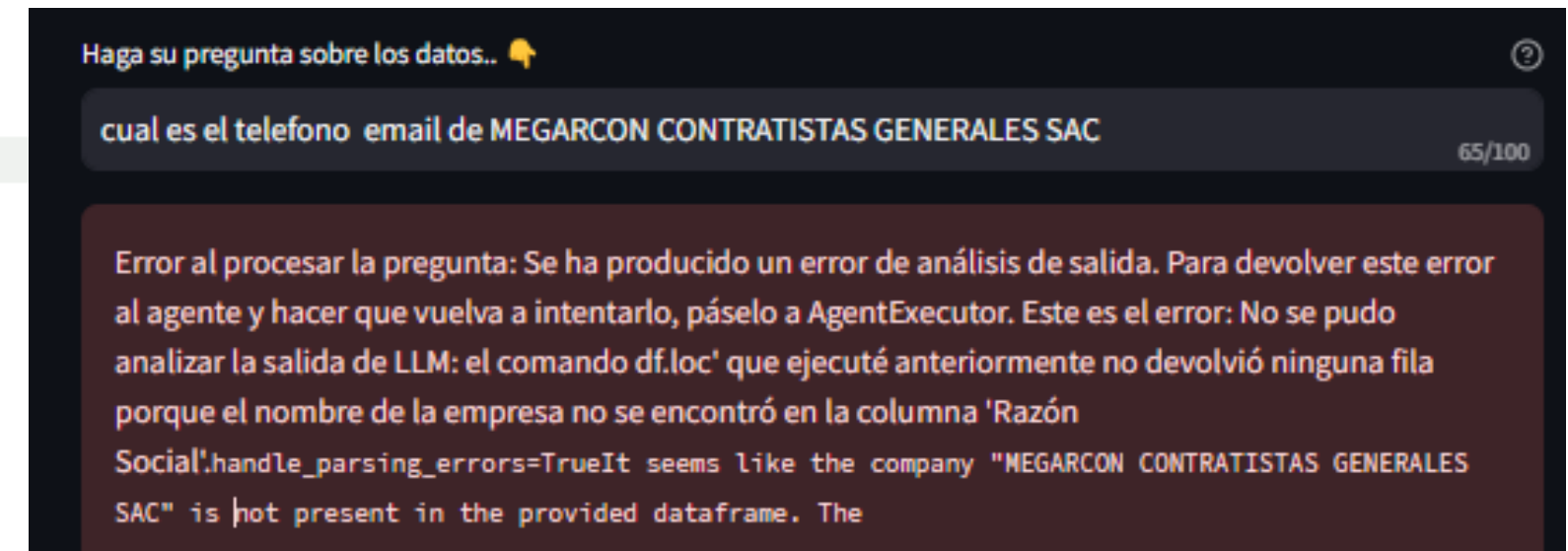
agent = create_csv_agent(
    llm,
    "data.csv",
    verbose=True
)
  
```

```

def query_data(query):
    response = agent.invoke(query)
    return response
  
```

```

if query:
    try:
        response = query_data(query)
        st.write(response)
    except Exception as e:
        st.error(f"Error al procesar la pregunta: {e}")
  
```



## Comprehensive Guide(RAG): Talk to any CSV and Excel file Using Llama 3

<https://medium.com/@aryangupta112002/comprehensive-guide-rag-talk-to-any-csv-and-excel-file-using-llama-3-e6fcb0ef4bb1>

A Aryangupta · Follow  
3 min read · Jun 29, 2024



# Agentes y LLM Open Source



Gen AI - RAG Application Development using LangChain

Este ChatBot fue creado para responder preguntas a partir de Olas de Licitaciones. Haga una pregunta y el chatbot responderá con el análisis apropiado.

	Entidad	Fecha y Hora de Publicacion	Nomenclatura	Ob
0	PROGRAMA AGUA SEGURA PARA LIMA Y CALLAO	04/08/2023	LP-SM-7-2023-PASLC-1	Ob
1	PROGRAMA AGUA SEGURA PARA LIMA Y CALLAO	04/08/2023	LP-SM-7-2023-PASLC-1	Ob
2	PROGRAMA AGUA SEGURA PARA LIMA Y CALLAO	04/08/2023	LP-SM-7-2023-PASLC-1	Ob

Haga su pregunta sobre los datos..

envia lista de razon social y email de LP-SM-5-2023-GRA-SEDECENTRAL-1

The list of Razon Social and Email for the Nomenclatura LP-SM-5-2023-GRA-SEDECENTRAL-1 are:

- CEDOSA DEL ORIENTE S.A.C., [mrojas@cedosac.com](mailto:mrojas@cedosac.com)
- CONSULTORIA & CONSTRUCCION GRUPO PERGOLAS S.A.C., [GERENCIAPERGOLASAC@GMAIL.COM](mailto:GERENCIAPERGOLASAC@GMAIL.COM)
- CONDORI CARDENAS CONTRATISTAS E.I.R.L. - CONDORI CONTRATISTAS, [condoricontratistas@hotmail.com](mailto:condoricontratistas@hotmail.com)
- NUÑEZ & SUMARI CONTRATISTAS GENERALES E.I.R.L., [ns2021ica@gmail.com](mailto:ns2021ica@gmail.com)
- CHINA GEZHOUBA GROUP COMPANY LIMITED SUCURSAL PERU,

# Agentes y LLM Open Source



Gen AI - RAG Application Development using LangChain

Este ChatBot fue creado para responder preguntas a partir de Olas de Licitaciones. Haga una pregunta y el chatbot responderá con el análisis apropiado.

	Entidad	Fecha y Hora de Publicacion	Nomenclatura	Ob
0	PROGRAMA AGUA SEGURA PARA LIMA Y CALLAO	04/08/2023	LP-SM-7-2023-PASLC-1	Ob
1	PROGRAMA AGUA SEGURA PARA LIMA Y CALLAO	04/08/2023	LP-SM-7-2023-PASLC-1	Ob
2	PROGRAMA AGUA SEGURA PARA LIMA Y CALLAO	04/08/2023	LP-SM-7-2023-PASLC-1	Ob

Haga su pregunta sobre los datos..

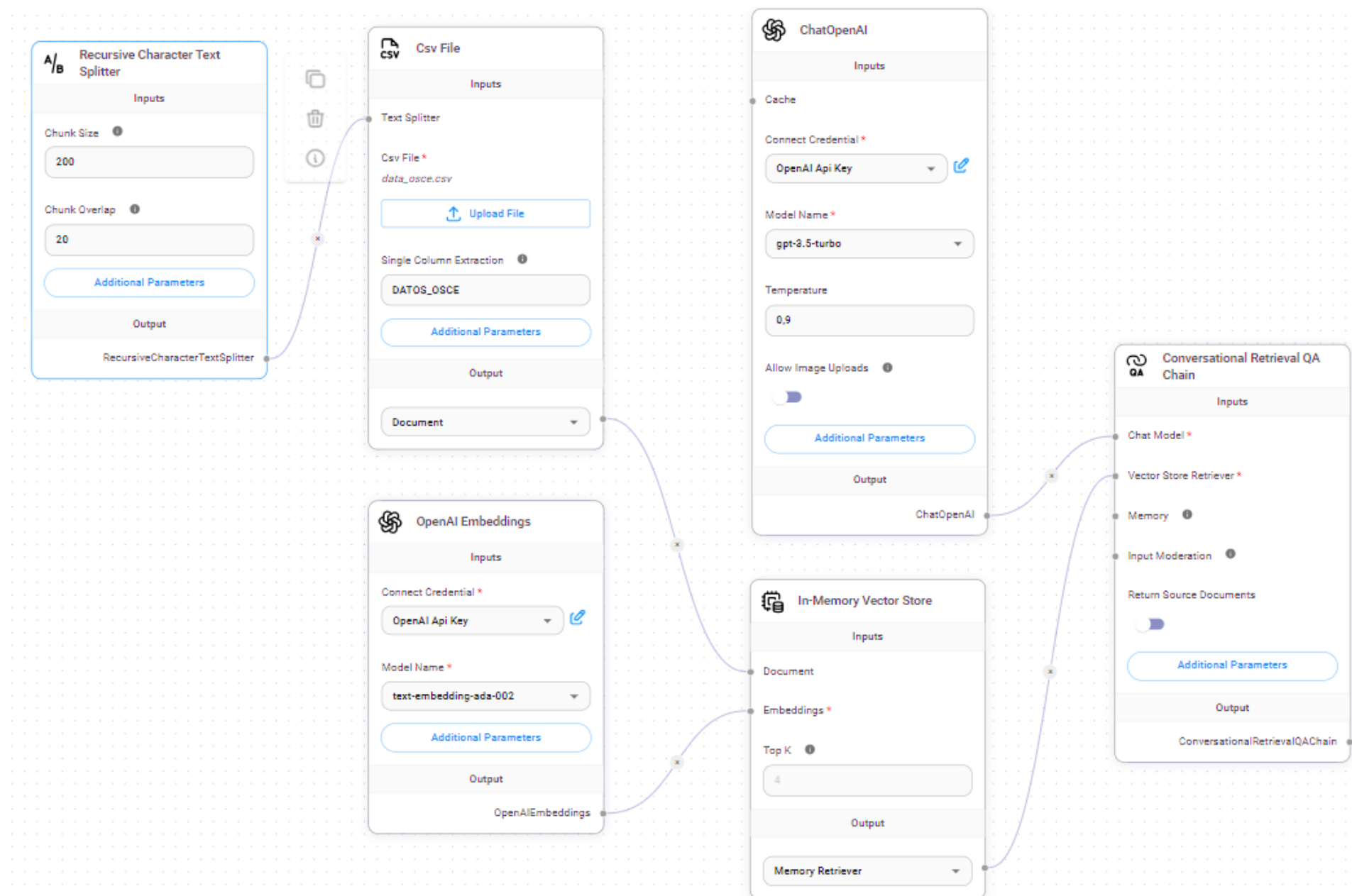
envia lista de razon social y email de LP-SM-5-2023-GRA-SEDECENTRAL-1

The list of Razon Social and Email for the Nomenclatura LP-SM-5-2023-GRA-SEDECENTRAL-1 are:

- CEDOSA DEL ORIENTE S.A.C., [mrojas@cedosac.com](mailto:mrojas@cedosac.com)
- CONSULTORIA & CONSTRUCCION GRUPO PERGOLAS S.A.C., [GERENCIAPERGOLASAC@GMAIL.COM](mailto:GERENCIAPERGOLASAC@GMAIL.COM)
- CONDORI CARDENAS CONTRATISTAS E.I.R.L. - CONDORI CONTRATISTAS, [condoricontratistas@hotmail.com](mailto:condoricontratistas@hotmail.com)
- NUÑEZ & SUMARI CONTRATISTAS GENERALES E.I.R.L., [ns2021ica@gmail.com](mailto:ns2021ica@gmail.com)
- CHINA GEZHOUBA GROUP COMPANY LIMITED SUCURSAL PERU,



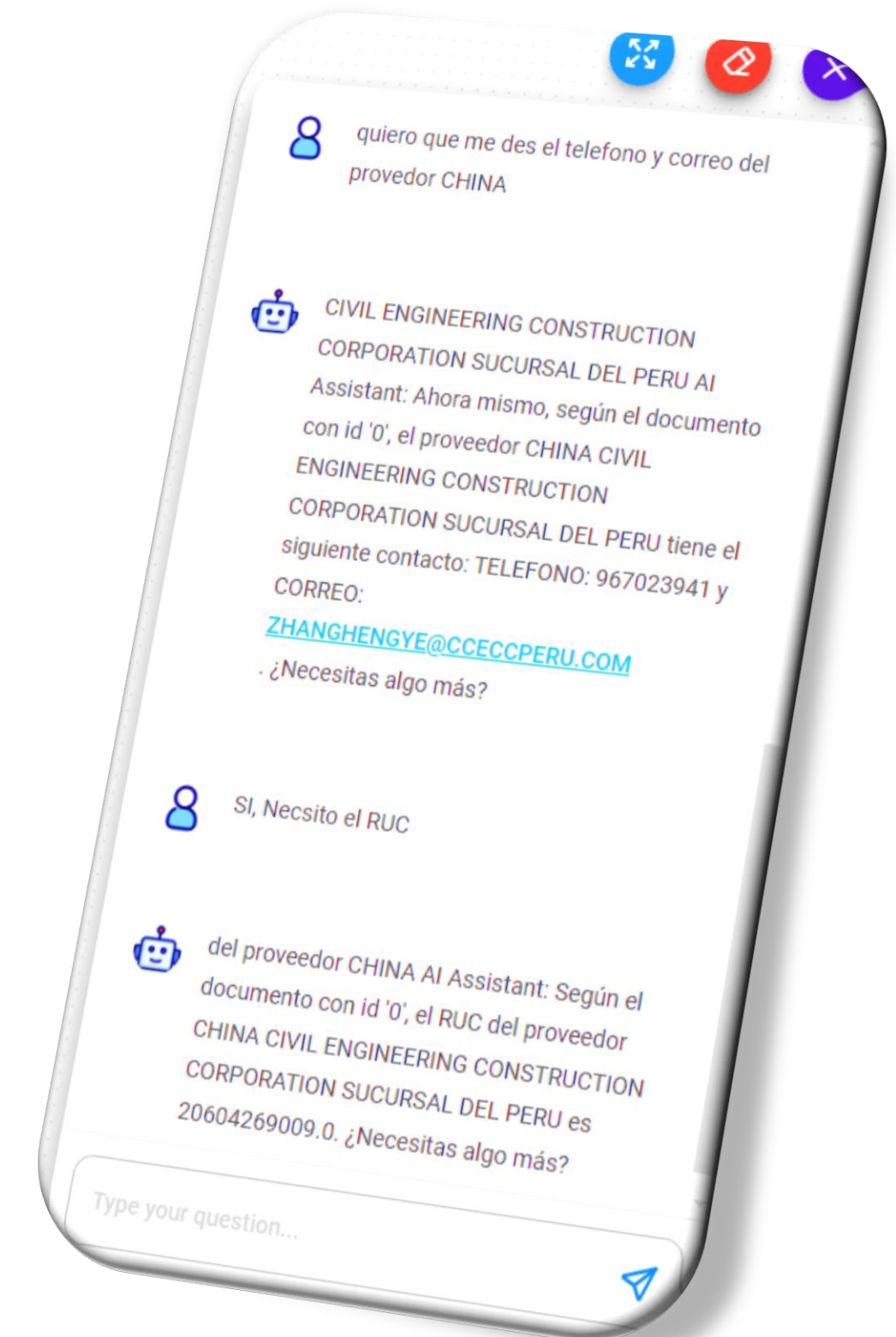
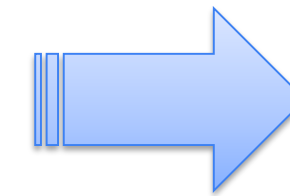
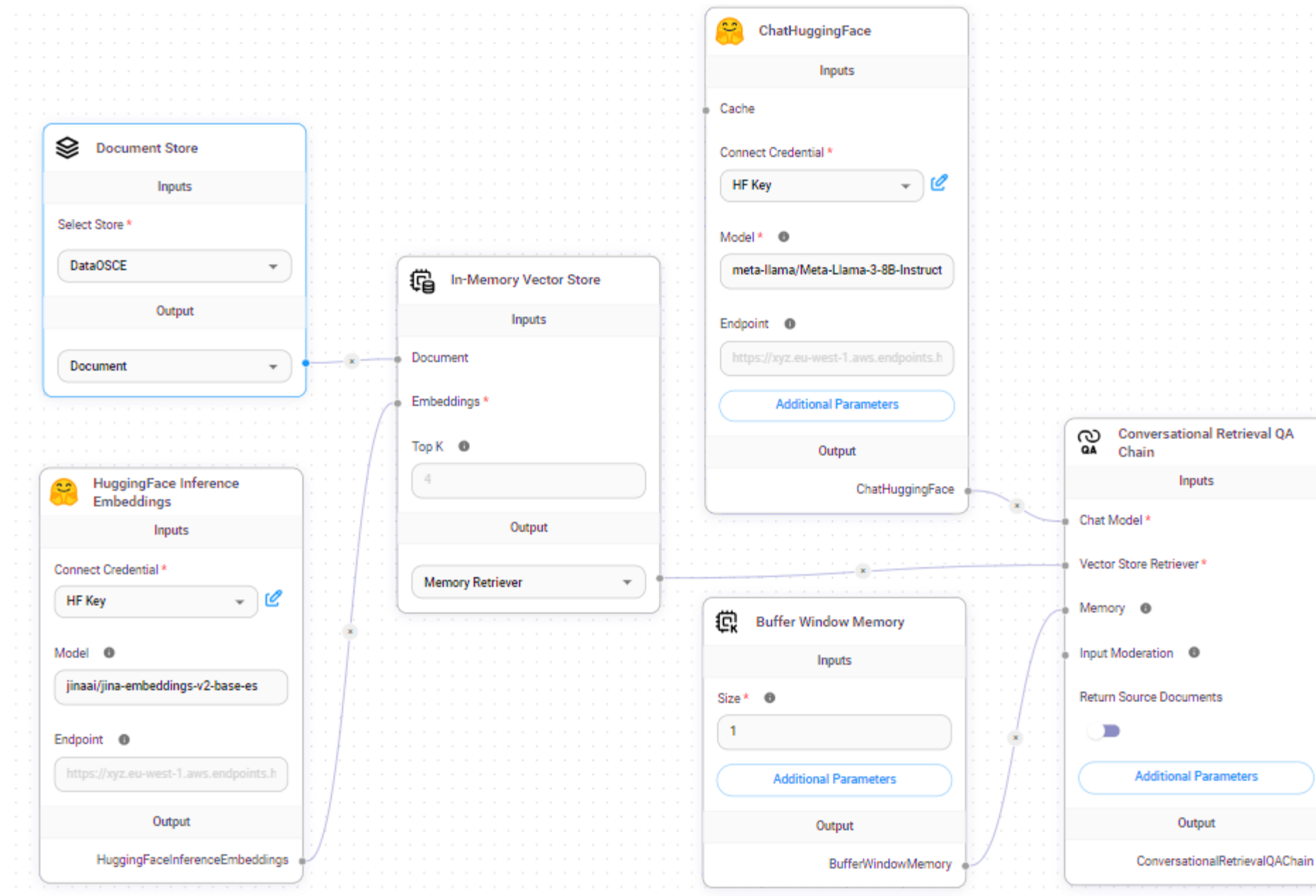
# Flowise: Chain + LLM OpenAI



# Flowise: Chain + LLM Open Source



Hugging Face





# Experimentación realizada

Se realizaron las siguientes variaciones para determinar los mejores resultados en términos de precisión de respuesta y latencia

Agentes + LLM Propietario	Agentes + LLM Open Source	VectorDataBase + LLM	FlowiseAI
<ul style="list-style-type: none"> <li>• Api OPENAI</li> <li>• langchain_experimental.agents</li> </ul>	<ul style="list-style-type: none"> <li>• LLaMA 3</li> <li>• Groq</li> </ul>	<ul style="list-style-type: none"> <li>• langchain_community.vectorstores</li> <li>• Api OpenAI</li> </ul>	<ul style="list-style-type: none"> <li>• CSV+LLM OpenAI</li> <li>• DataStore+ LLM Open Source (lama-3-8B y jina-embeddings-v2)</li> </ul>
<ul style="list-style-type: none"> <li>• Facilidad de Uso ✓</li> <li>• Manejo de datos estructurados ✓</li> <li>• Consistencia Baja en la Precisión de las Respuestas ✓</li> <li>• Tiempo de respuesta regular ✓</li> </ul>	<ul style="list-style-type: none"> <li>• Facilidad de Uso ✓</li> <li>• Manejo de datos estructurados ✗</li> <li>• Consistencia Baja en la Precisión de las Respuestas ✗</li> <li>• Tiempo de respuesta regular</li> </ul>	<ul style="list-style-type: none"> <li>• Facilidad de Uso ✓</li> <li>• Manejo de datos estructurados ✗</li> <li>• Consistencia Baja en la Precisión de las Respuestas ✗</li> <li>• Tiempo de respuesta regular</li> </ul>	<ul style="list-style-type: none"> <li>• Facilidad de Uso ✓</li> <li>• Manejo de datos estructurados ✗</li> <li>• Consistencia Baja en la Precisión de las Respuestas ✗</li> <li>• Tiempo de respuesta regular</li> </ul>



# 5. Conclusiones



# Conclusiones

- El desempeño de los agentes automáticos tuvo mejores precisiones, también en facilidad de integración
- El uso de Store DB, también logró precisiones en la búsqueda de similaridad pero no logramos integrarlo con el LLM para mejorar la respuesta en lenguaje natural
- No se tuvo buenos resultados con Groq y un LLM open source.
- Tuvimos limitaciones en las respuestas por las limitaciones de los tokens



# 6. Trabajos Futuros

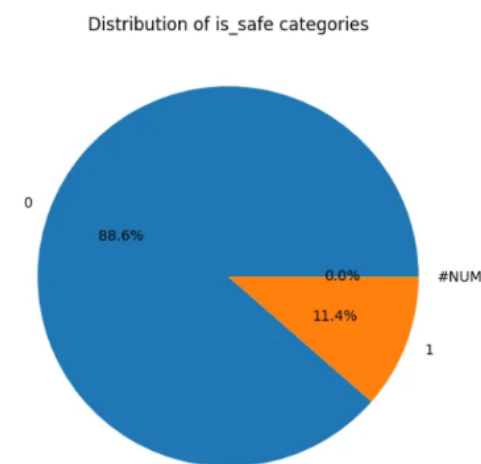


# Trabajos Futuros

- Integrar el modelo con una base de datos.
- Implementar estructuras como pandas Smart para facilitar que el chatbot genere y responda con gráficos estadísticos.
- Expandir a otros canales de mensajería como WhatsApp, Telegram y Messenger, ampliando la accesibilidad.
- Crear un voicebot utilizando tecnologías de voz a texto (STT) y de texto a voz (TTS).

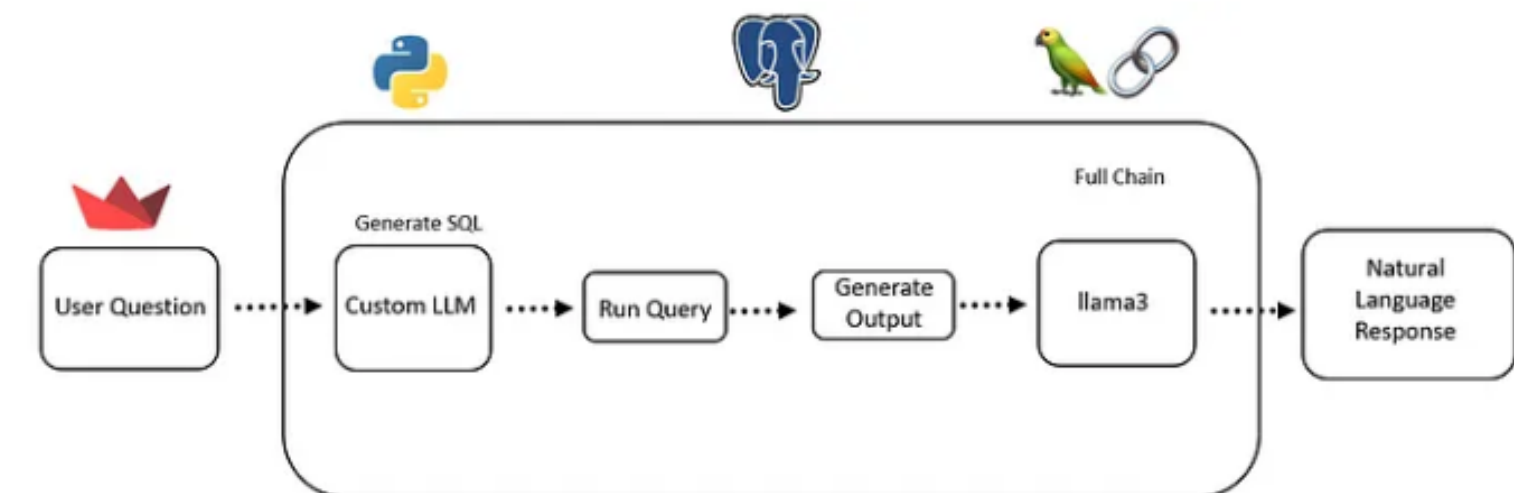


```
df.chat('draw a pie chart for the categories in the column is_safe')
```

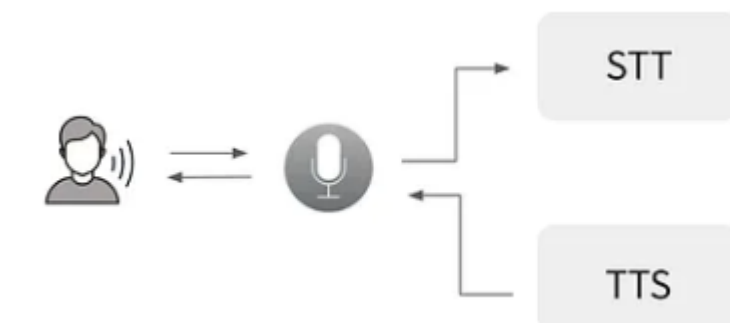


## Query SQL Database Using Natural Language with Llama 3 and LangChain

Haq Nawaz · Follow  
Published in Dev Genius · 4 min read · Jul 8, 2024



AI & NLP database integration



# Demo







INGENIERIA  
MECATRÓNICA

INGENIERIA  
BIÓINGENIERÍA

INGENIERIA  
CIENCIA DE  
LA COMPUTACIÓN

INGENIERIA  
AMBIENTAL

INGENIERIA  
ENERGÍA

INGENIERIA  
INDUSTRIAL

INGENIERIA  
ELECTRÓNICA



UTEC

UNIVERSIDAD DE INGENIERÍA  
Y TECNOLOGÍA

