



3° Workshop de Medición en
Upstream y Downstream
de Petróleo y Gas 

22-23
AGO



Aplicación de Inteligencia Artificial Generativa de Texto en Métodos de Medición

Adrian Trujillo,
Analista tgs
Data Scientist – AI Engineer





3° Workshop de Medición en
Upstream y Downstream
de Petróleo y Gas 

22-23
AGO



Universo de IA
generativas de textos
conocidas también como
LLM





Como puede entender una IA un texto:

Las palabras se la representan con tokens a su vez esos tokens representan un numero cada uno que pasan a formar un vector que es la representación numérica de la frase.

Los vectores además guardan otro tipo de información como por ejemplo cercanías con otras palabras.



Tokens:

Token Text:

Con la colaboración del experto, el informe se confeccionó meticulosamente

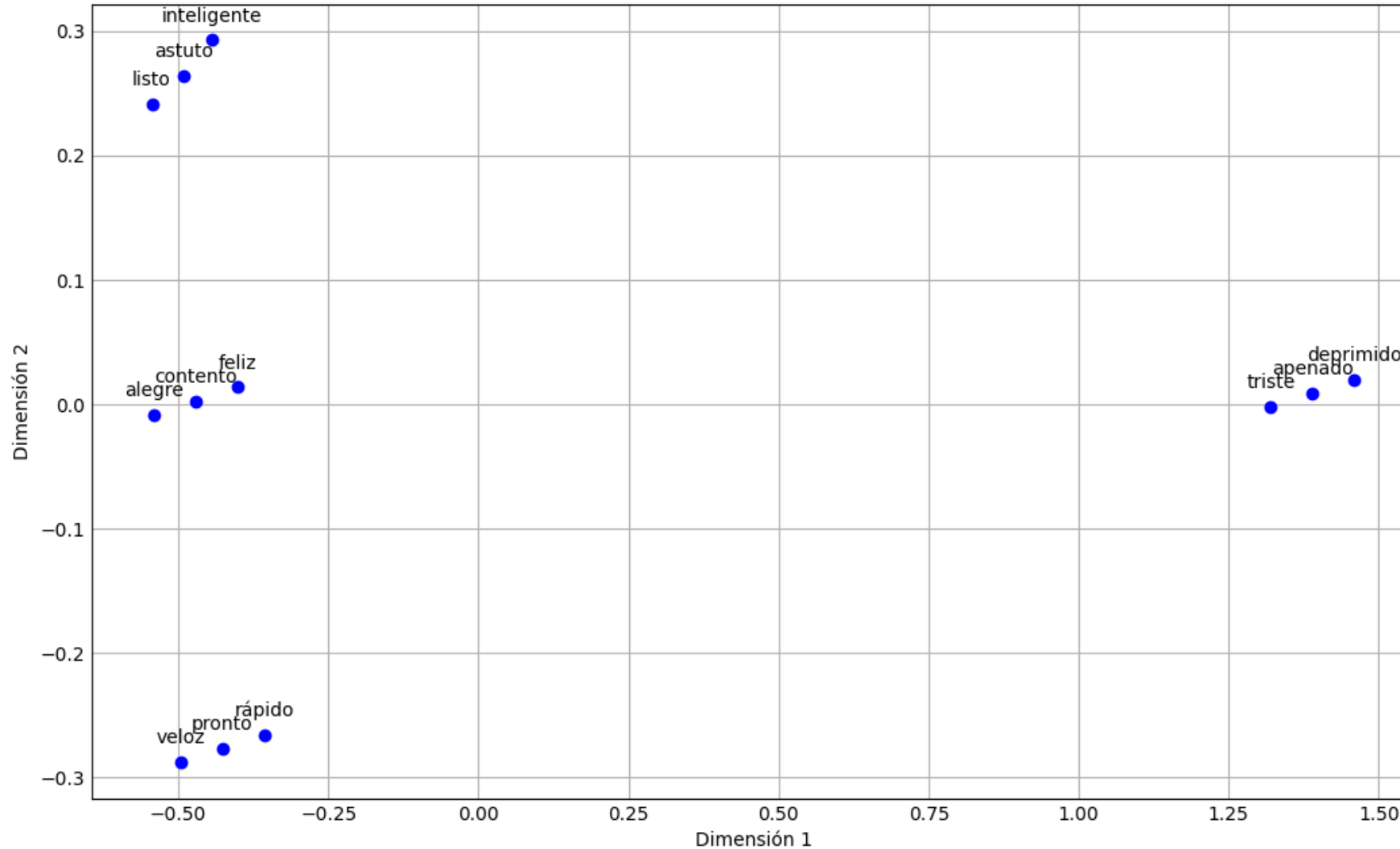
Token Ids:

[3103, 8591, 951, 4820, 32009, 18840, 1619, 1121, 1462, 11, 1288, 4175,
68, 384, 1013, 68, 535, 295, 10205, 30762, 377, 418, 3263, 68]



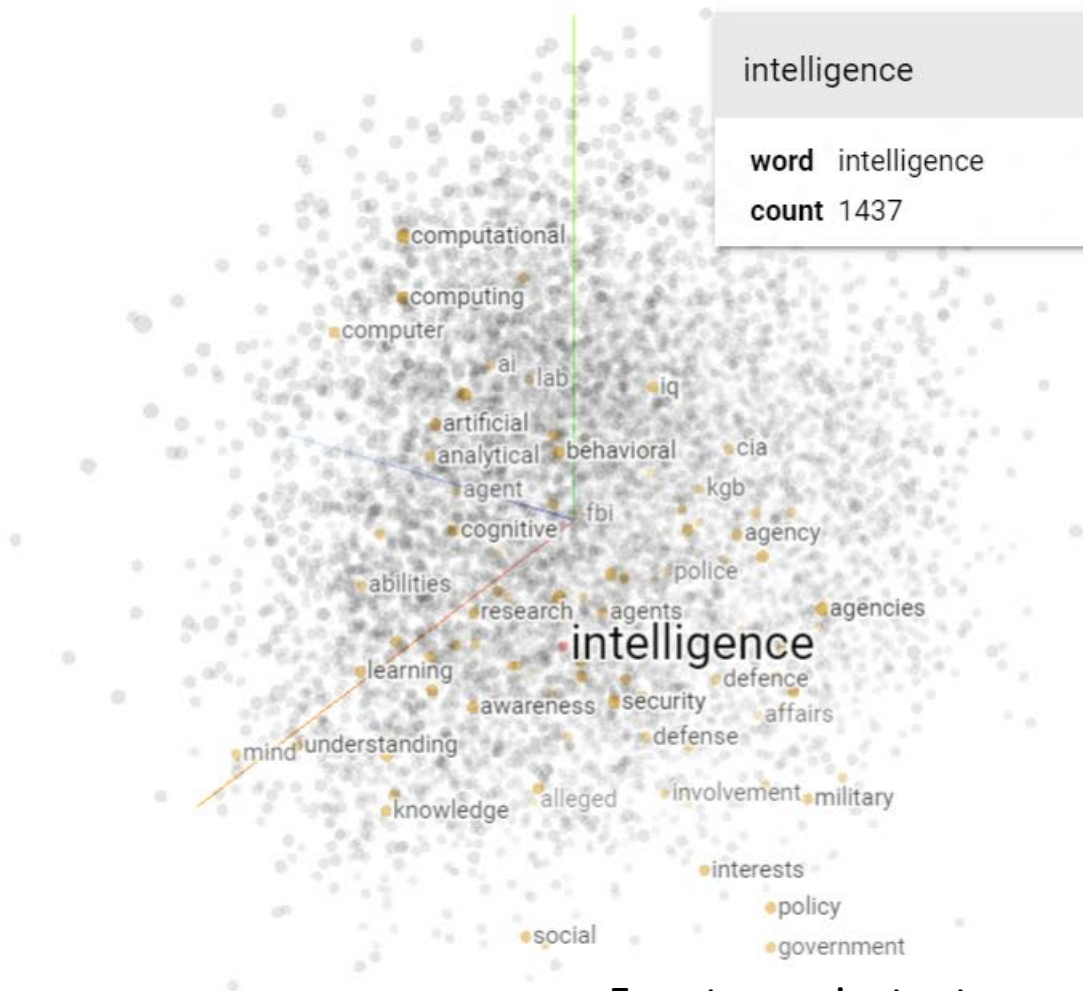
Vectores:

Representación vectorial de palabras clave y sus sinónimos





Points: 10000 | Dimension: 200 | Selected 101 points



intelligence ^

word intelligence

count 1437

Show All Data

Isolate 101 points

Clear selection

Search

by

intelligence

.*

word

neighbors ?



100

distance

COSINE

EUCLIDEAN

Nearest points in the original space:

artificial	0.572
agency	0.600
cognitive	0.605
agents	0.620
security	0.620
agencies	0.651
police	0.663



Probabilidad
del siguiente
token:
mecanismos
de atención

contexto = "When instrument is out of toleranace"

Palabra: range, Probabilidad: 0.2491

Palabra: ,, Probabilidad: 0.1016

Palabra: , Probabilidad: 0.1005

Palabra: ., Probabilidad: 0.0620

Palabra: and, Probabilidad: 0.0253



contexto = "When instrument is out of toleranace range,"

Palabra: it, Probabilidad: 0.0850

Palabra: the, Probabilidad: 0.0737

Palabra: display, Probabilidad: 0.0147

Palabra: alarm, Probabilidad: 0.0138

Palabra: return, Probabilidad: 0.01362



Datos de entrenamiento de GPT-3⁹

Conjunto de datos	# tokens	Proporción dentro del entrenamiento
Common Crawl	410 mil millones	60%
WebText2	19 mil millones	22%
Books1	12 mil millones	8%
Books2	55 mil millones	8%
Wikipedia	3 mil millones	3%

Fuente: <https://es.wikipedia.org/wiki/GPT-3>



Alucinaciones:

Genera información falsa o contradictoria. Las alucinaciones pueden ocurrir por varias razones, como:

- Los datos de entrenamiento no son suficientemente representativos del mundo real
- No tiene una comprensión completa de los datos en los que se entrenó.

Sesgo:

Reflejar las opiniones, creencias y prejuicios de las personas que crearon los datos en los cuales se entreno.

Pregunta: ¿Cuáles son los beneficios de comer frutas y verduras?

Respuesta correcta: Comer frutas y verduras puede ayudar a prevenir enfermedades cardíacas, derrames cerebrales, diabetes y algunos tipos de cáncer.

Respuesta sesgada: Comer frutas y verduras puede ayudar a perder peso.

**ERRORES EN
LAS IA
GENERATIVAS
DE TEXTO:**



Como se
pueden
usar?

Playgrounds:

- <https://chat.openai.com/>
- <https://gemini.google.com/>
- <https://www.perplexity.ai/>
- <https://www.bing.com/>



Programando una aplicación:

- LLM en la nube (API Key: los datos van a la nube)
- LLM on locale (privacidad)





Modelos on locale:

[Models](#)[Datasets](#)[Spaces](#)[Posts](#)[Docs](#)[Pricing](#)

Hugging Face is way more fun with friends and colleagues! 😊 [Join an organization](#)

Dismiss this message

Tasks 1 Libraries Datasets Languages
Licenses Other

Reset Tasks

Multimodal

Image-Text-to-Text

Visual Question Answering

Document Question Answering

Computer Vision

Depth Estimation Image Classification

Object Detection Image Segmentation


Text-to-Image Image-to-Text

Image-to-Image Image-to-Video


Models 78,074

[new](#) Full-text search


Sort: Most likes

 bigscience/bloom


Text Generation • Updated Jul 28, 2023 • ↓ 9.23k • ♥ 4.53k

 meta-llama/Llama-2-7b


Text Generation • Updated 26 days ago • ♥ 3.8k

 mistralai/Mixtral-8x7B-Instruct-v0.1

Text Generation • Updated Feb 29 • ↓ 852k • ♥ 3.63k

 meta-llama/Llama-2-7b-chat-hf

Text Generation • Updated 26 days ago • ↓ 1.42M • ♥ 3.34k

 mistralai/Mistral-7B-v0.1

Text Generation • Updated Dec 11, 2023 • ↓ 2.16M • ♥ 3.08k



Parámetros

LLM	Parámetros [billones]
GPT 4	-
GPT 3.5	-
Falcon 180B	180
GPT 3	175
Gemini Ultra	-
Meta's Llama 2 x 70	70
Falcon 40B	40
Mistral MoE 8x7B	45
Meta's Llama 2 x 13	13
Meta's Llama 2 x 7	7
Mistral 7B	7
Falcon 7 B	7
Gemini Nano-2	3.25
Gemini Nano-1	1.8



	LLaMA 2 70B	GPT - 3.5	Mixtral 8x7B
MMLU (MCQ in 57 subjects)	69.9%	70.0%	70.6%
HellaSwag (10-shot)	87.1%	85.5%	86.7%
ARC Challenge (25-shot)	85.1%	85.2%	85.8%
WinoGrande (5-shot)	83.2%	81.6%	81.2%
MBPP (pass@1)	49.8%	52.2%	60.7%
GSM-8K (5-shot)	53.6%	57.1%	58.4%
MT Bench (for Instruct Models)	6.86	8.32	8.30

MMLU: producir texto de calidad humana:
completar y traducir

HellaSwag: comprender y responder a preguntas

ARC Challenge: razonamiento abstracto

WinoGrande: tareas de razonamiento y
comprensión del mundo real

MBPP: generar código Python

GSM-8K: resumen, respuestas y títulos

MT Bench: traducir idiomas



TÉCNICAS DE MITIGACIÓN DE ERRORES EN IA GENERATIVAS DE TEXTO:

Prompt Engineering:

Permite diseñar preguntas o instrucciones específicas que guían a la IA en la dirección deseada, reduciendo la posibilidad de divagar o inventar información.

RAG (Retrieval-Augmented Generation):

Se brinda a la IA información como contexto para genere respuesta en base al mismo. La información es una página completa de texto.



TÉCNICAS DE MITIGACIÓN DE ERRORES EN IA GENERATIVAS DE TEXTO:

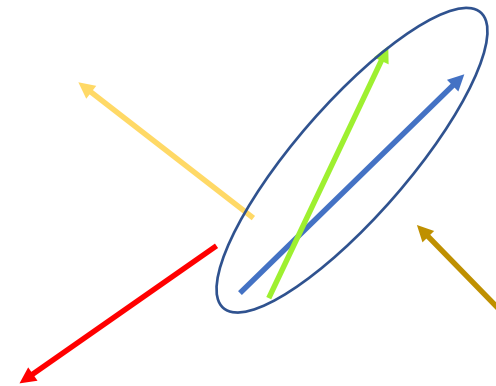
Embeddings:

El mismo concepto que RAG, solo que la información que se pasa como contexto es a nivel de similitudes a nivel de frases, en lugar de una página entera, mediante similitudes vectoriales.

Pregunta:



Contexto:



Similitud: Angulo entre vectores por longitud



CUSTOMIZAR MODELOS: fine tuning

Buscar que un modelo de IA generativa de texto sea “EXPERTA” en un tema específico, por ejemplo en:

- Contabilidad
- Medicina
- Botánica
- Metrología
- Mecánica
- Etc.

La primer tarea es convertir el texto en un dataset con dos columnas:

- Pregunta
- Respuesta

Fine Tuning:

Se reentrena al modelo ajustando una cantidad limitada de parámetros, es decir no se reentrena todo el modelo sino una parte del mismo.



CASOS DE USO GENERICOS:



AMPLIAR
CONOCIMIENTO



GENERAR TEXTO
SOBRE UN TEMA



TRADUCIR TEXTO



RESUMIR TEXTO



OBTENER TEXTO
DE UNA IMAGEN



GENERAR CÓDIGO
PROGRAMACIÓN



RAZONAR CON OPERACIONES
MATEMÁTICAS. EJ:
CONVERSIÓN DE UNIDADES.



...



APLICADO AL AMBITO DE MEDICIONES:



GENERAR
DOCUMENTOS
SOBRE EL TEMA



GENERAR
VALORES
FICTICIOS



EFFECTUAR
INFERENCIAS
SOBRE VALORES



CONVERSAR
SOBRE
DOCUMENTOS
ESPECÍFICOS.



CREAR UN
EXPERTO EN
METROLOGÍA



CREAR UN
EQUIPO QUE
RESUELV A UNA
TAREA



3° Workshop de Medición en
Upstream y Downstream
de Petróleo y Gas 

22-23
AGO



INSTITUTO ARGENTINO
DEL PETROLEO Y DEL GAS

Prompt Engineering





Basic Prompt:

Rol + Acciones + Modificadores + Datos

Eres un especialista en metrología, elabora procedimientos de calibración para transmisor de presión, contemplando los requisitos de la norma ISO 17025, contempla requisitos de trazabilidad de los patrones, establece 5 puntos de calibración sobre el rango del instrumento, por cada punto de calibración efectúa 3 repeticiones, calcula la incertidumbre de medición

Rol

Acción (verbos)

Modificadores



Shots para generar una respuesta inducida + CoT

Eres un especialista en metrología. Calcula la incertidumbre de medición de un instrumento teniendo en cuenta la incertidumbre del patrón y el desvío estándar del instrumento.

Shot 1: La incertidumbre de medición se calcula mediante la raíz cuadrada de la suma de la Incertidumbre del Patrón al cuadrado y la Incertidumbre del Instrumento al cuadrado.

Shot 2: La Incertidumbre del Instrumento se calcula multiplicando el desvío estándar por dos

Pregunta 1: Cual es la incertidumbre de medición de un instrumento, si la del patrón es 4 y el desvío estándar del instrumento es 3?

Paso 1: calcula la incertidumbre del instrumento de multiplicando el desvío estándar del instrumento por dos.

Paso 2: calcula la incertidumbre del patrón de multiplicando el desvío estándar del patron por dos.

Paso 3: calcula la incertidumbre de medición de la siguiente manera como la raíz cuadrada de la suma de la incertidumbre del patrón mas la del instrumento.

Rol

Acción (verbos)

Modificadores



PROMPT ENGINEERING DATOS FICTICIOS:

Rol + Acciones + Modificadores + Datos

Eres un científico de datos, genera una tabla con datos ficticios de temperatura, siguiendo el patrón de valores que te proporciono en {datos}, elabora una tabla con dos columnas: una para la variable fecha formato timestamp del día de hoy a partir de las 6 de la mañana y los siguientes registros cada hora, la otra para la variable temp_fict en donde colocarás los valores ficticios que generastes, entregala en formato csv.

{datos} = [28.117, 28.117, 27.891, 27.94, 28.145, 28.195, 28.557, 28.646, 28.331, 27.883, 27.578, 27.992, 27.92, 27.876, 27.998, 28.153, 28.256, 28.219, 28.145, 28.175, 28.204, 28.136, 28.144, 28.159, 28.14, 28.117]

Rol

Acción (verbos)

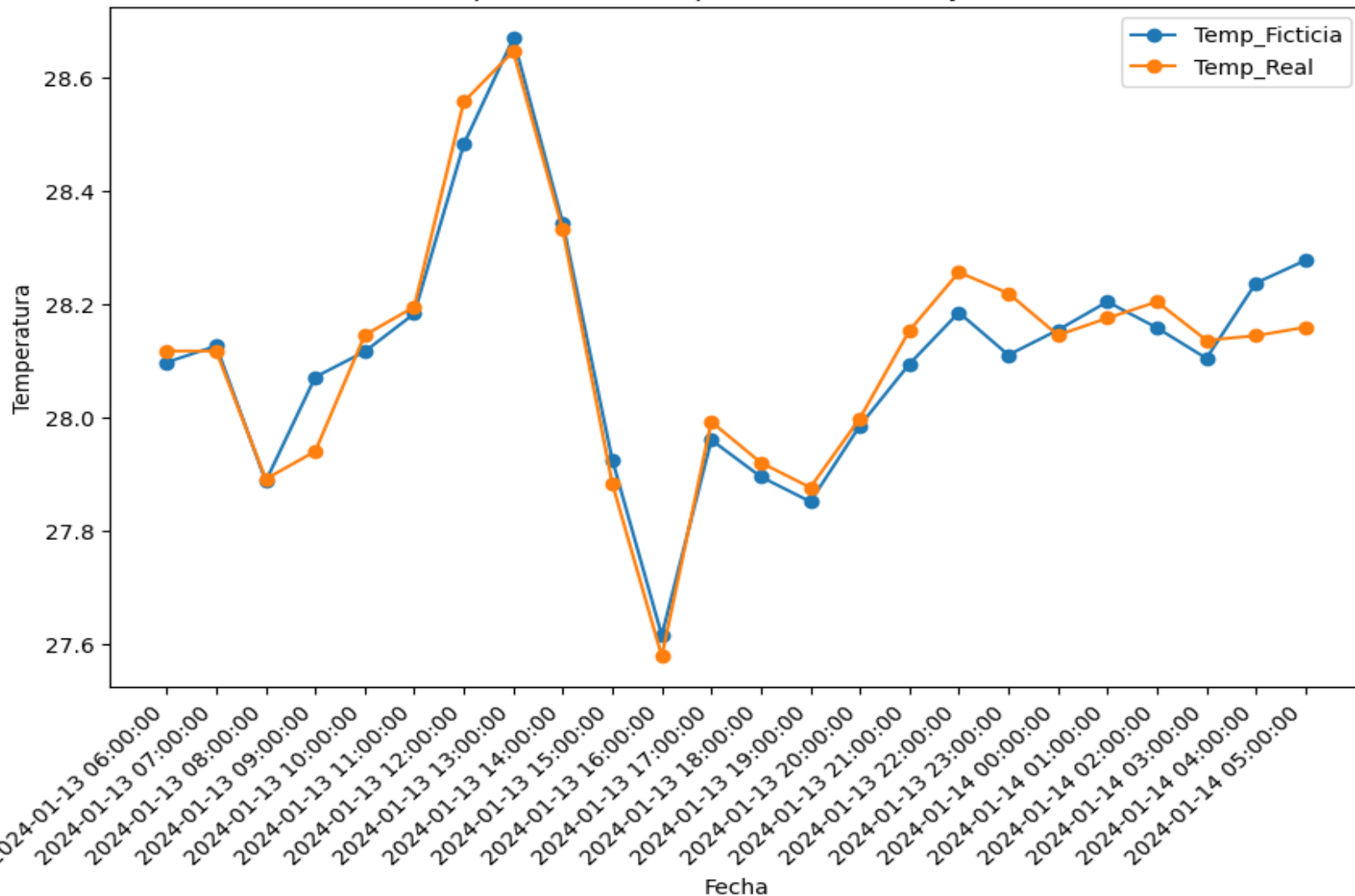
Modificadores

Datos

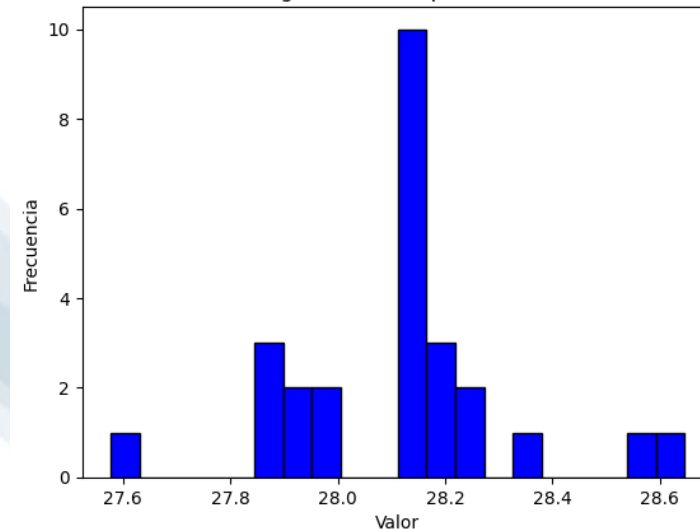


PROMPT ENGINEERING DATOS FICTICIOS: :

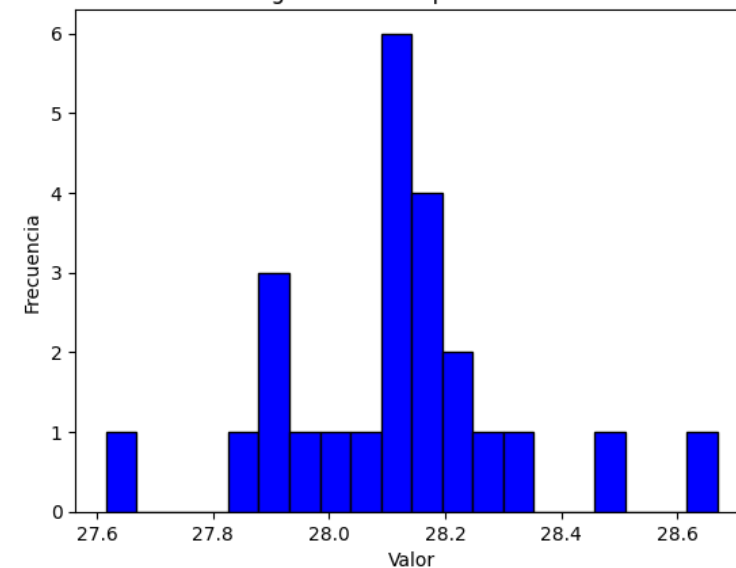
Comparación de Temperaturas Ficticias y Reales



Histograma de temperatura real



Histograma de Temperatura ficticia





EFFECTUAR INFERENCIAS SOBRE VALORES:



Gas Turbine CO and NOx Emission Data Set

Donated on 11/28/2019

The dataset contains 36733 instances of 11 sensor measures aggregated over one hour, from a gas turbine located in Turkey for the purpose of studying flue gas emissions, namely CO and NOx.



UC Irvine
Machine Learning
Repository

Variable Information

The explanations of sensor measurements and their brief statistics are given below.

Variable (Abbr.)	Unit	Min	Max	Mean
Ambient temperature (AT)	C	6.23	37.10	17.71
Ambient pressure (AP)	mbar	985.85	1036.56	1013.07
Ambient humidity (AH)	(%)	24.08	100.20	77.87
Air filter difference pressure (AFDP)	mbar	2.09	7.61	3.93
Gas turbine exhaust pressure (GTEP)	mbar	17.70	40.72	25.56
Turbine inlet temperature (TIT)	C	1000.85	1100.89	1081.43
Turbine after temperature (TAT)	C	511.04	550.61	546.16
Compressor discharge pressure (CDP)	mbar	9.85	15.16	12.06
Turbine energy yield (TEY)	MWH	100.02	179.50	133.51
Carbon monoxide (CO)	mg/m3	0.00	44.10	2.37
Nitrogen oxides (NOx)	mg/m3	25.90	119.91	65.29



EFFECTUAR INFERENCIAS SOBRE VALORES:

Prompt (GPT 3):

You are a helpful assistant. Answer this question {question} based on the given context information: {context}. Provide a value for the concentration, considering the observed pattern in the data.

{question} = What is the carbon monoxide concentration value, when ambient temperature is between 19.5 to 20 C , turbine inlet temperature is between to 1099 to 1100 C, and Turbine energy yield is near to 149 to 151 MWH?

{context} = INSERTAR AQUI VALORES CONTEXTUALIZADOS FORMATO DICCIONARIO

Rol

Acción (verbos)

Modificadores

Datos



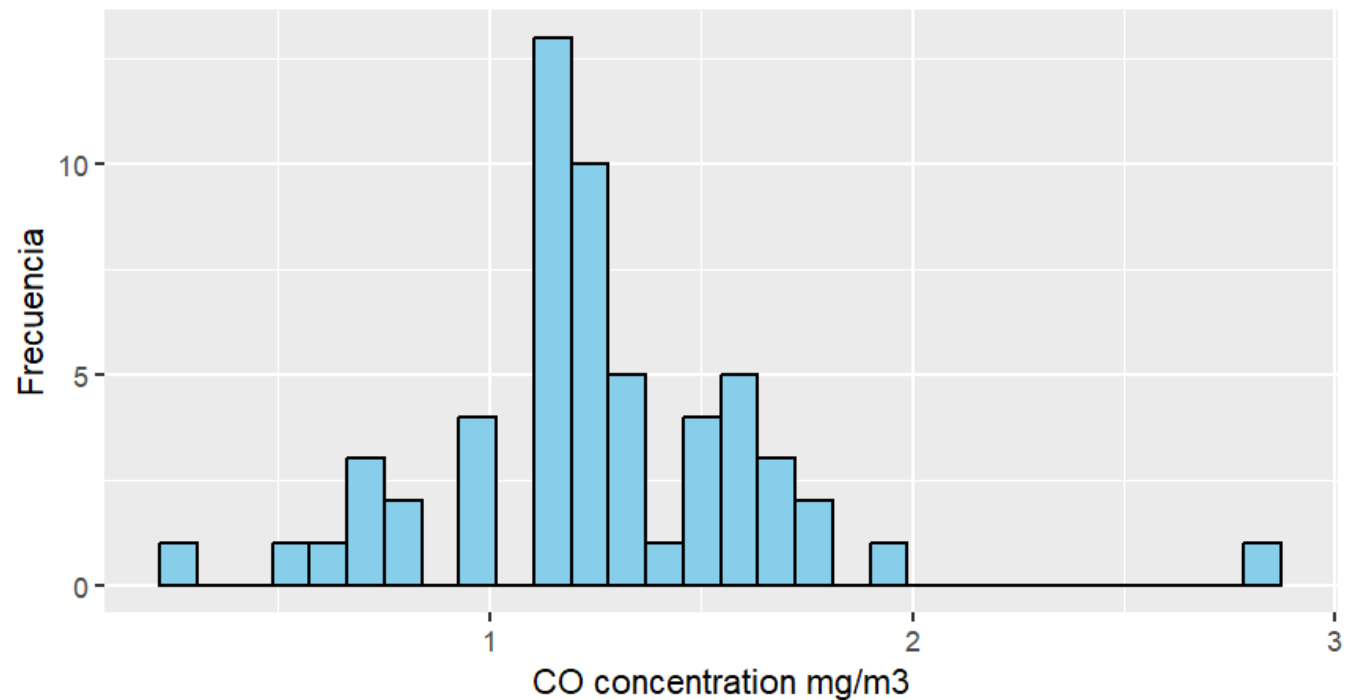
EFECTUAR INFERENCIAS SOBRE VALORES: *Respuesta:*

You can observe that the carbon monoxide concentrations in these conditions are within the range mentioned. Based on this, you might estimate the carbon monoxide concentration to be around 1.2 mg/m^3 under the specified conditions. Keep in mind that this is an approximation, and the exact value may vary based on other factors and conditions not explicitly mentioned in the provided data.

summary(result\$CO)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.2822	1.1106	1.2191	1.2594	1.4819	2.8448

Histograma



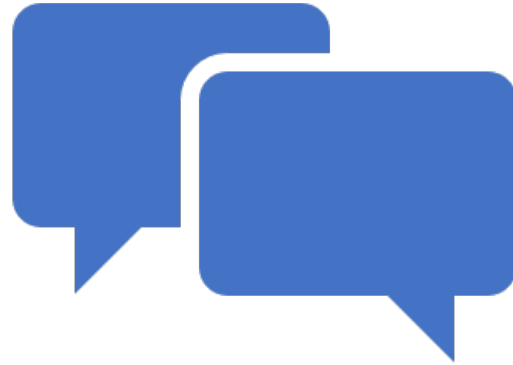


3° Workshop de Medición en
Upstream y Downstream
de Petróleo y Gas 

22-23
AGO



INSTITUTO ARGENTINO
DEL PETROLEO Y DEL GAS



CHAT CON SUS DOCUMENTOS



CHAT CON SUS DOCUMENTOS:

"RAG" se refiere a "Retrieval-Augmented Generation", que significa "Generación con recuperación aumentada". Es una forma de dar contexto al prompt pero no con toda la información sino con la información que sea mas cercana a la pregunta.

"Eres un asistente que responde preguntas. : Responde la siguiente pregunta {query} en base al contexto brindado por el usuario: {context}. Responde siempre en español."

{query}= como se calibra un transmisor de presión
{context} =... Colocar aquí el resultado del RAG

Rol

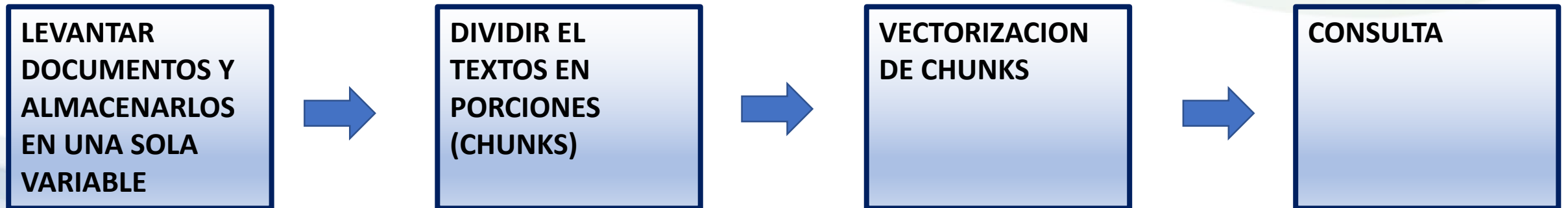
Acción (verbos)

Modificadores

Datos



RETRIEVAL-AUGMENTED GENERATION (RAG)





CHAT CON SUS DOCUMENTOS: mistralai/Mistral-7B-Instruct-v0.2 (INSTANCIABLE VRAM 16 GB)

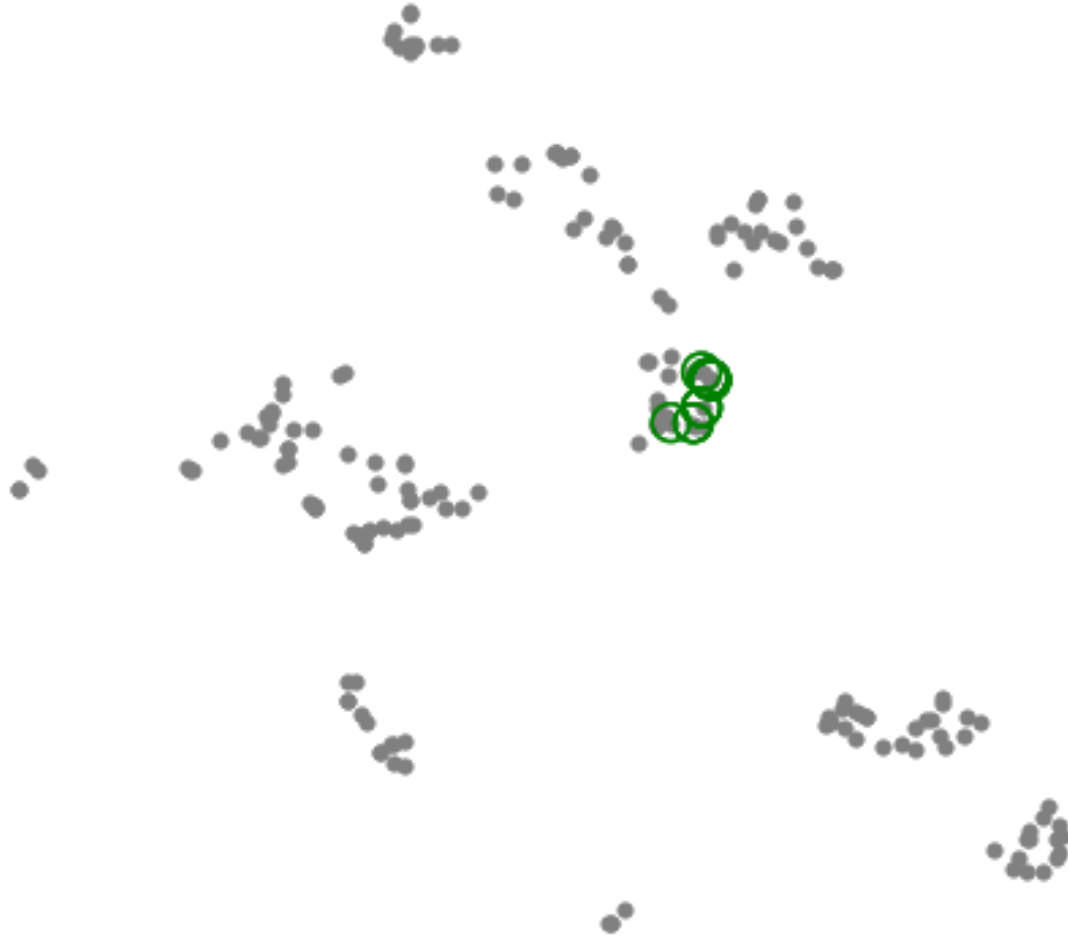
Para calibrar un transmisor de temperatura, siga estos pasos:

1. Conecte el transmisor de temperatura a un instrumento generador de temperatura, como un generador de termopar o un generador de termómetro de resistencia.
2. Conecte el transmisor de temperatura a un sensor de temperatura, como un termómetro de resistencia o un termopar.
3. Encienda el instrumento generador de temperatura y deje que se estabilice.
4. Configure el instrumento generador de temperatura para que funcione en grados centígrados o grados Fahrenheit.
5. Establezca la unidad de lectura de temperatura en grados centígrados o Fahrenheit en el patrón de medición del transmisor de temperatura.
6. Retire el transmisor de temperatura del instrumento generador de temperatura y vuelva a insertarlo.
7. Cuando la lectura de temperatura del estándar de medición del transmisor de temperatura sea estable, ingrese el valor.
8. Repita los pasos 6 y 7 para tomar los valores de temperatura restantes necesarios para la calibración.
9. Retire el transmisor de temperatura del instrumento generador de temperatura, desconéctelo del estándar de medición del transmisor de temperatura y retire el termómetro estándar.
10. Desenergice el equipo maestro.
11. Coloque el transmisor de temperatura en el proceso y conéctelo al sensor de temperatura.
12. Si la variable se ha dejado en valores fijos, cámbiela a valores activos.
13. Si hay un bucle de control y se ha puesto en control manual, restáurelo.
14. Complete y coloque la tarjeta de calibración.

Nota: El cálculo de la incertidumbre se realizará utilizando la hoja de cálculo PGCO101-F04. También se contempla el caso de calibrar la indicación de temperatura en su conjunto, es decir, indicador (transmisor o transductor) junto con su sensor de temperatura, utilizando un generador de temperatura y realizando las repeticiones como se indica en el Procedimiento PGCO104. El patrón a utilizar es únicamente un termómetro.



How a pressure transmitter is calibrated?



Mapa vectorial tokens

Relevancy and Distraction

- **Relevancia** se refiere a qué tan estrechamente relacionado está algo a un tema o tarea en particular.
- **Distracción**, es algo que desvía la atención del LLM de la información relevante.



3° Workshop de Medición en
Upstream y Downstream
de Petróleo y Gas 

22-23
AGO

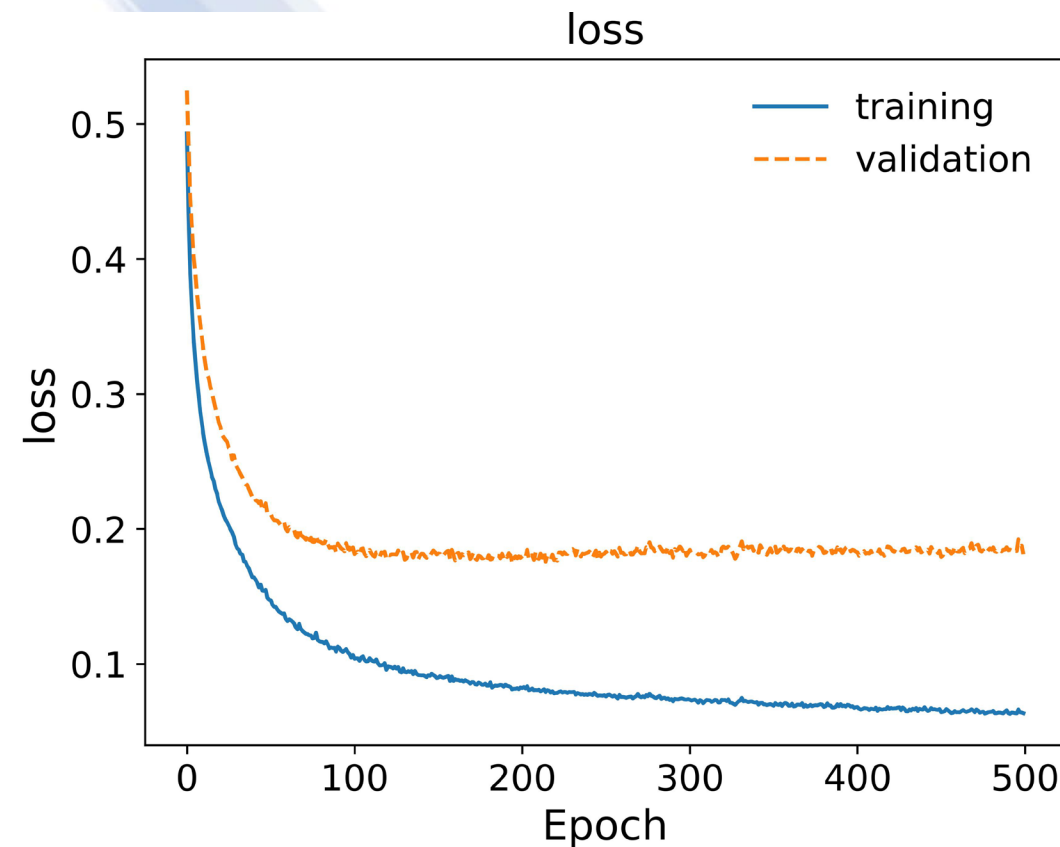
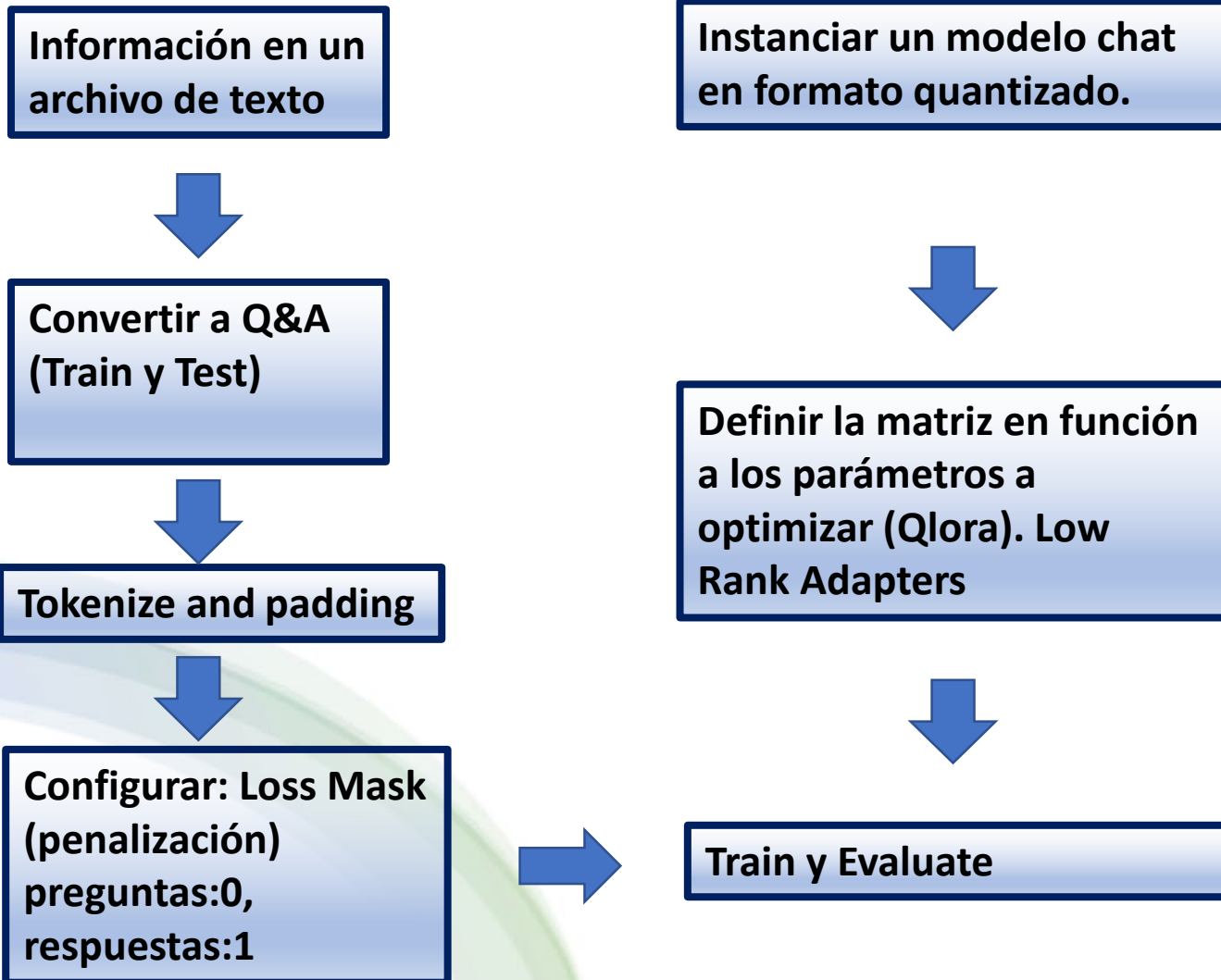


INSTITUTO ARGENTINO
DEL PETRÓLEO Y DEL GAS

ENTRENAR MODELOS PARA TAREAS ESPECÍFICAS: FINE TUNING



Fine tuning





3° Workshop de Medición en
Upstream y Downstream
de Petróleo y Gas 

22-23
AGO



INSTITUTO ARGENTINO
DEL PETRÓLEO Y DEL GAS

AI Agents





AI Agents:
ejemplo
aplicación
web
scrapping

USER



HERRAMIENTAS



EQUIPO

Investigador



Escritor de papers



Traductor



TASK





Definir agentes

```
[ ] researcher = Agent(  
    role='Gas Flow Measurement Expert',  
    goal="""Investigate and analyze the different types and causes of errors in gas flow measurement using orifice plates,  
    specifically focusing on whether the errors result in overestimation or underestimation of the measured value."""  
    backstory="""You are a specialist in gas flow measurement with a focus on orifice plates.  
    Your expertise lies in identifying and analyzing factors that contribute to measurement errors."""  
    verbose=True,  
    allow_delegation=False,  
    llm=llm,  
    tools=[web_rag_tool, search_tool]  
)
```

```
[ ] writer = Agent(  
    role='Technical Content Writer',  
    goal="""Create an informative and engaging article that analyzes the different types and causes of errors in gas flow measurement  
    using orifice plates, including their impact on the measurement value (underestimation or overestimation).""",  
    backstory="""You are a skilled technical writer with a strong understanding of gas flow  
    measurement principles and applications. You excel at translating complex technical details  
    into clear and concise content for various audiences.""",  
    verbose=True,  
    allow_delegation=True,  
    llm=llm,  
    tools=[]  
)
```



Tareas:

```
▶ task1 = Task(  
  description="""Conduct an investigation into the various types and causes of errors encountered in gas flow measurement utilizing orifice plates.  
  Analyze the influence of these errors on the accuracy of measurement values, distinguishing between underestimation and overestimation.  
  Identify and categorize common factors contributing to each type of error as documented in online resources.  
  Explore a diverse range of potential sources, considering factors such as rugosity, and sharp edges, distances from elbows and straight line distance.  
  Compile a comprehensive report detailing the identified types of errors, their respective causes, and their impact on the accuracy of measurement values.""",  
  expected_output='comprehensive report detailing types of errors, their causation, and whether they result in underestimation or overestimation in orifice plate flow measurement',  
  agent=researcher,  
  confidence_level=0.9,  
  min_pages=10  
)
```

```
[11] task2 = Task(  
  description="""Investigate and analyze the causes of overestimation and underestimation in gas flow measurement using orifice plates.  
  Provide a comprehensive list of common factors contributing to each type of error found in research. Consider a variety of sources.  
  Write an informative and engaging article for a general audience that explains the importance of accurate gas flow measurement  
  in orifice plates and the potential consequences of errors. Avoid technical jargon and use clear, concise language.""",  
  expected_output='comprehensive list of causes of underestimation and overestimation in gas flow measurement using orifice plates',  
  agent=writer  
)
```



Resultado:



****Errores en la medición del flujo de gas utilizando placas de orificio: causas y consecuencias****

****Introducción****

Las placas de orificio se emplean ampliamente en aplicaciones de medición de flujo de gas debido a su simplicidad y rentabilidad. Sin embargo, pueden surgir imprecisiones por diversos factores

****Errores de sobreestimación****

- * ****Rugosidad de la tubería:**** La rugosidad en la superficie interior de la tubería aguas arriba de la placa de orificio puede generar turbulencia, lo que resulta en una mayor caída de presión
- * ****Bordes afilados:**** Los bordes afilados en el orificio de la placa de orificio pueden crear caídas de presión localizadas, lo que lleva a una sobreestimación del caudal.
- * ****Longitud insuficiente de la tubería recta aguas arriba:**** La presencia de curvas, válvulas u otras obstrucciones demasiado cerca de la placa de orificio puede alterar el patrón de flujo, c
- * ****Instalación incorrecta de la placa de orificio:**** La desalineación o el sellado inadecuado de la placa de orificio puede causar fugas de gas alrededor de la placa, lo que lleva a una sobre
- * ****Variación de temperatura del gas:**** Los cambios en la temperatura del gas pueden afectar la densidad y la viscosidad del fluido, lo que resulta en desviaciones del caudal calibrado y una p

****Errores de subestimación****

- * ****Erosión de la placa de orificio:**** Con el tiempo, la erosión del orificio de la placa de orificio debido a partículas abrasivas en el gas puede agrandar el tamaño del orificio, lo que llev
- * ****Depósitos y acumulación:**** La acumulación de depósitos o acumulación en la superficie de la placa de orificio puede reducir el tamaño efectivo del orificio, lo que resulta en una subestima
- * ****Longitud insuficiente de la tubería recta aguas abajo:**** La presencia de curvas o restricciones inmediatamente aguas abajo de la placa de orificio puede causar contrapresión, lo que lleva
- * ****Ubicación incorrecta de las tomas de presión:**** La colocación incorrecta de las tomas de presión puede resultar en una medición inexacta de la caída de presión a través de la placa de orif
- * ****Variación de la densidad del gas:**** Los cambios en la densidad del gas, debido a variaciones de presión o composición, pueden afectar la caída de presión a través de la placa de orificio y

****Importancia de la medición precisa del flujo de gas****

La medición precisa del flujo de gas utilizando placas de orificio es crucial por varias razones:

- * ****Medición fiscal:**** Las placas de orificio se utilizan en transacciones comerciales para medir el volumen de gas transferido. Una medición precisa garantiza una facturación justa y evita di
- * ****Control de procesos:**** La medición del flujo de gas es esencial para controlar y optimizar los procesos industriales que dependen del gas como materia prima o fuente de energía. Una medici
- * ****Seguridad y cumplimiento:**** La medición precisa del flujo ayuda a garantizar que los sistemas de gas operen dentro de límites seguros, previniendo situaciones de sobrepresión o subabasteci
- * ****Monitoreo ambiental:**** La medición del flujo de gas es fundamental para monitorear las emisiones y garantizar el cumplimiento de las regulaciones ambientales. Una medición precisa ayuda a a

****Consecuencias de los errores****

Los errores en la medición del flujo de gas utilizando placas de orificio pueden tener consecuencias significativas:



3° Workshop de Medición en
Upstream y Downstream
de Petróleo y Gas 

22-23
AGO



INSTITUTO ARGENTINO
DEL PETRÓLEO Y DEL GAS

Muchas gracias!