

¿MITO O REALIDAD?

EL IMPACTO DE LOS MODELOS DE LENGUAJE A GRAN ESCALA EN EL MUNDO DE LOS DATOS

Victoria Perez Mola

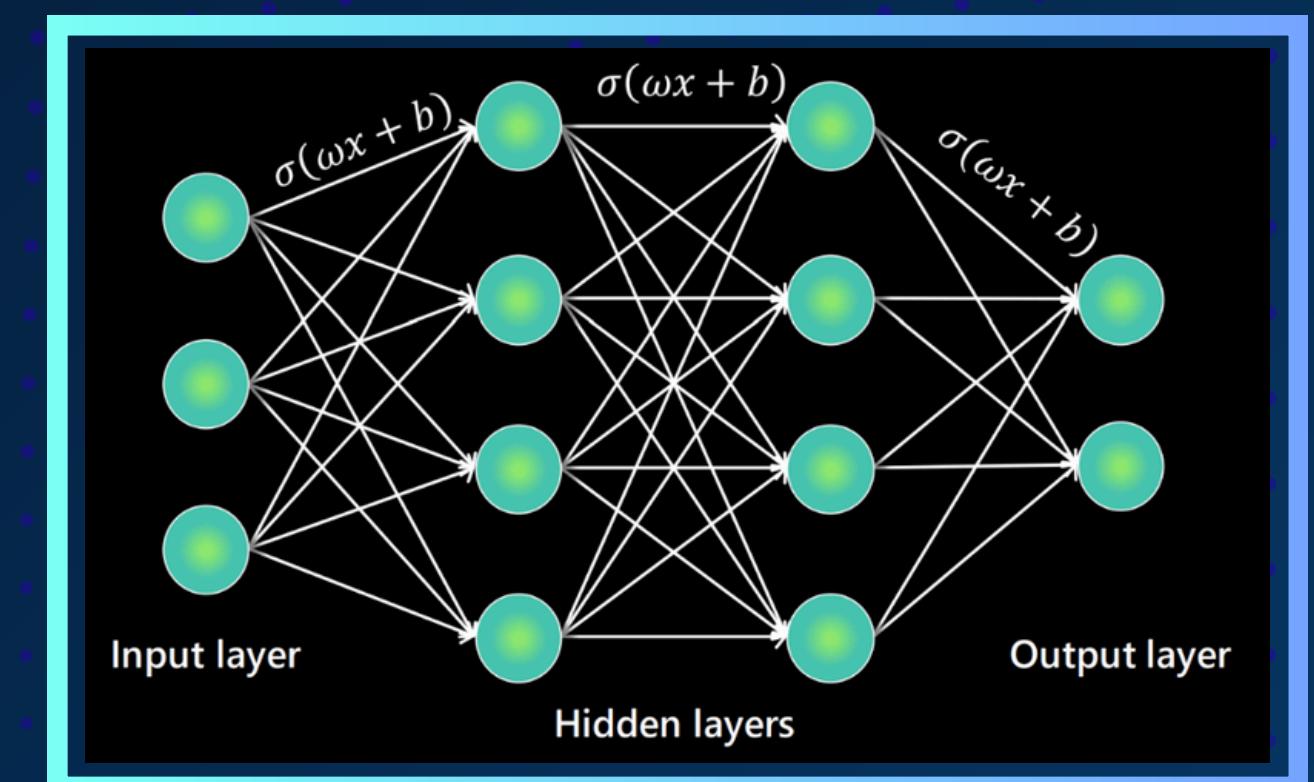
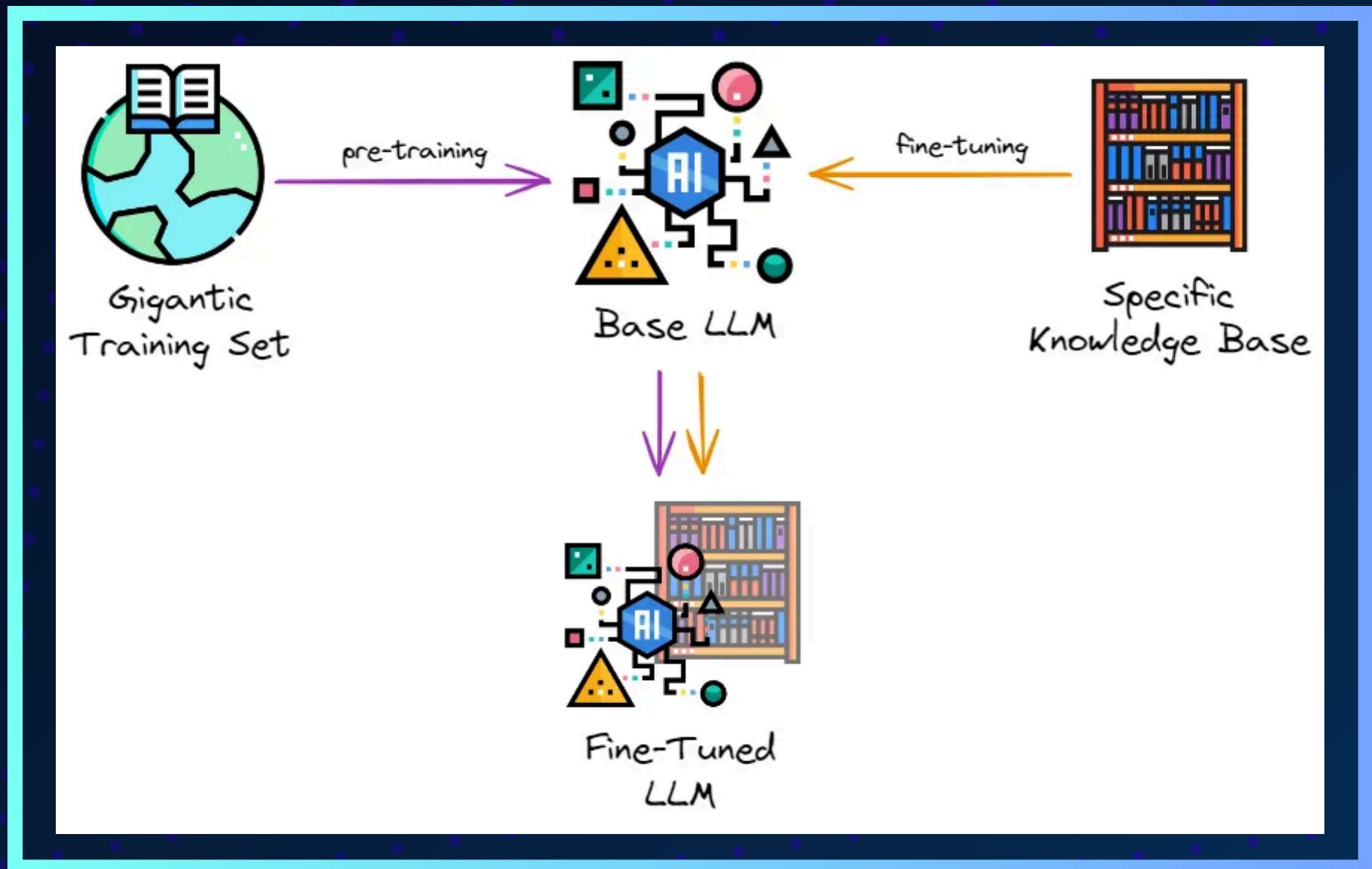
Resident Architect @ dbt Labs

AGENDA

- 01** ¿QUÉ SON LOS MODELOS DE LENGUAJE A GRAN ESCALA?
- 02** ¿CÓMO SE UTILIZAN PARA LOS ANÁLISIS DE DATOS?
- 03** MITO
- 04** REALIDAD

¿QUÉ SON LOS MODELOS DE LENGUAJE A GRAN ESCALA (LLM)?

- Sistemas de inteligencia artificial diseñados para comprender y generar texto basado en el lenguaje humano.
- Los modelos de lenguaje surgen inicialmente como modelos estadísticos en el año 1948-1980
- Durante las últimas décadas, se han mejorado, digitalizado y ampliado para llegar a gran escala



ChatGPT

como se dice LLM en español?

En español, "LLM" (Large Language Model) se traduce como "modelo de lenguaje grande" o "modelo de lenguaje a gran escala".



DONDE GRAN ESCALA SIGNIFICA...



3.5 TRILLONES

La cantidad de tokens utilizados para entrenar Falcon 180B (Hugging Face - 2023)



7.7 MILLONES

La cantidad de horas de GPU que requirió entrenar Llama 3 (Meta - 2023)



1.4 TRILLONES

La cantidad de párametros que se estima que contiene GPT4 (OpenAI - 2024)

APLICACIÓN AL ANÁLISIS DE DATOS

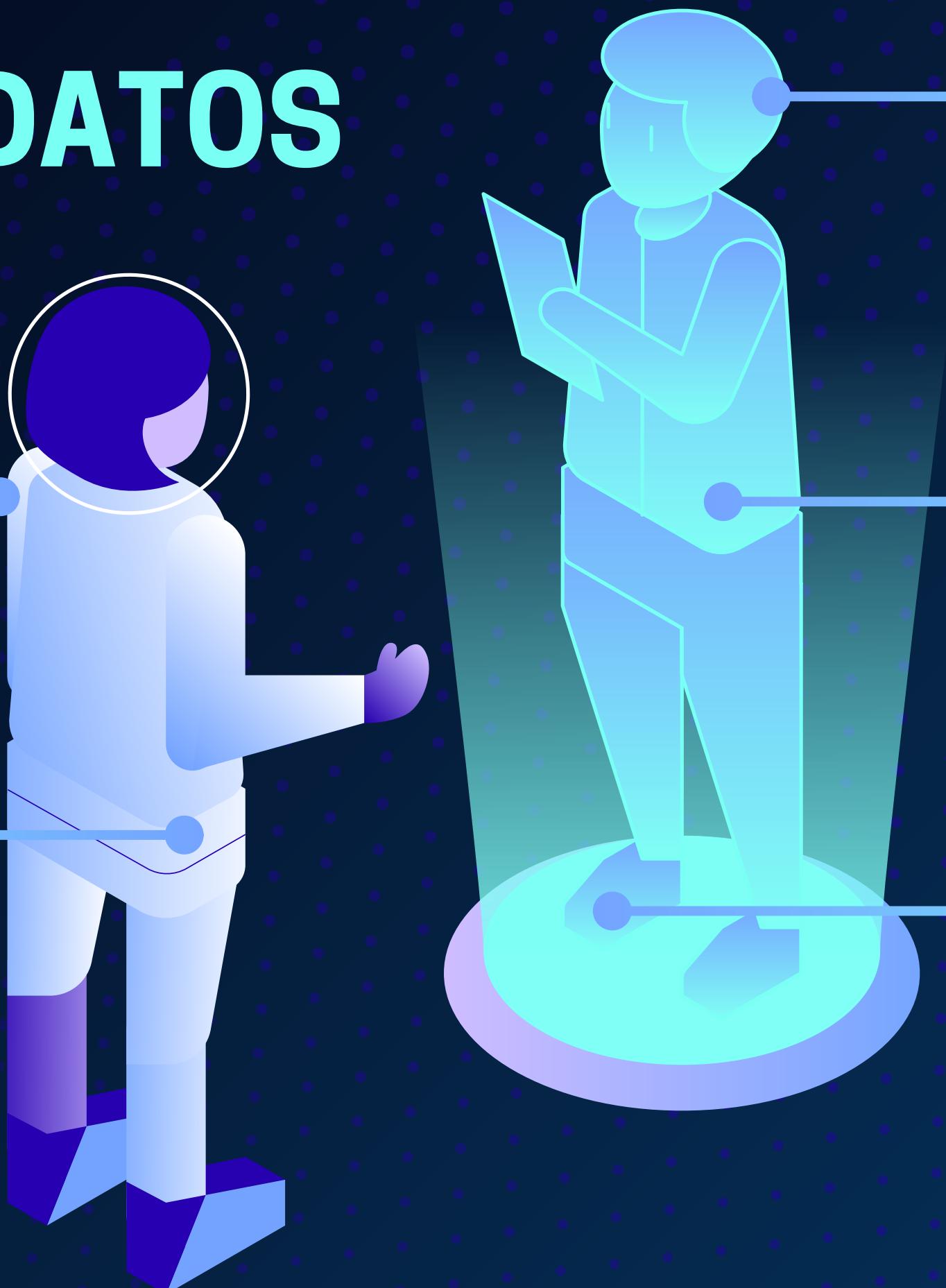
ANÁLISIS DE
SENTIMIENTOS

CLASIFICACIÓN DE
DATOS

LIMPIEZA DE
DATOS

AUTOMATIZACIÓN
DE REPORTES

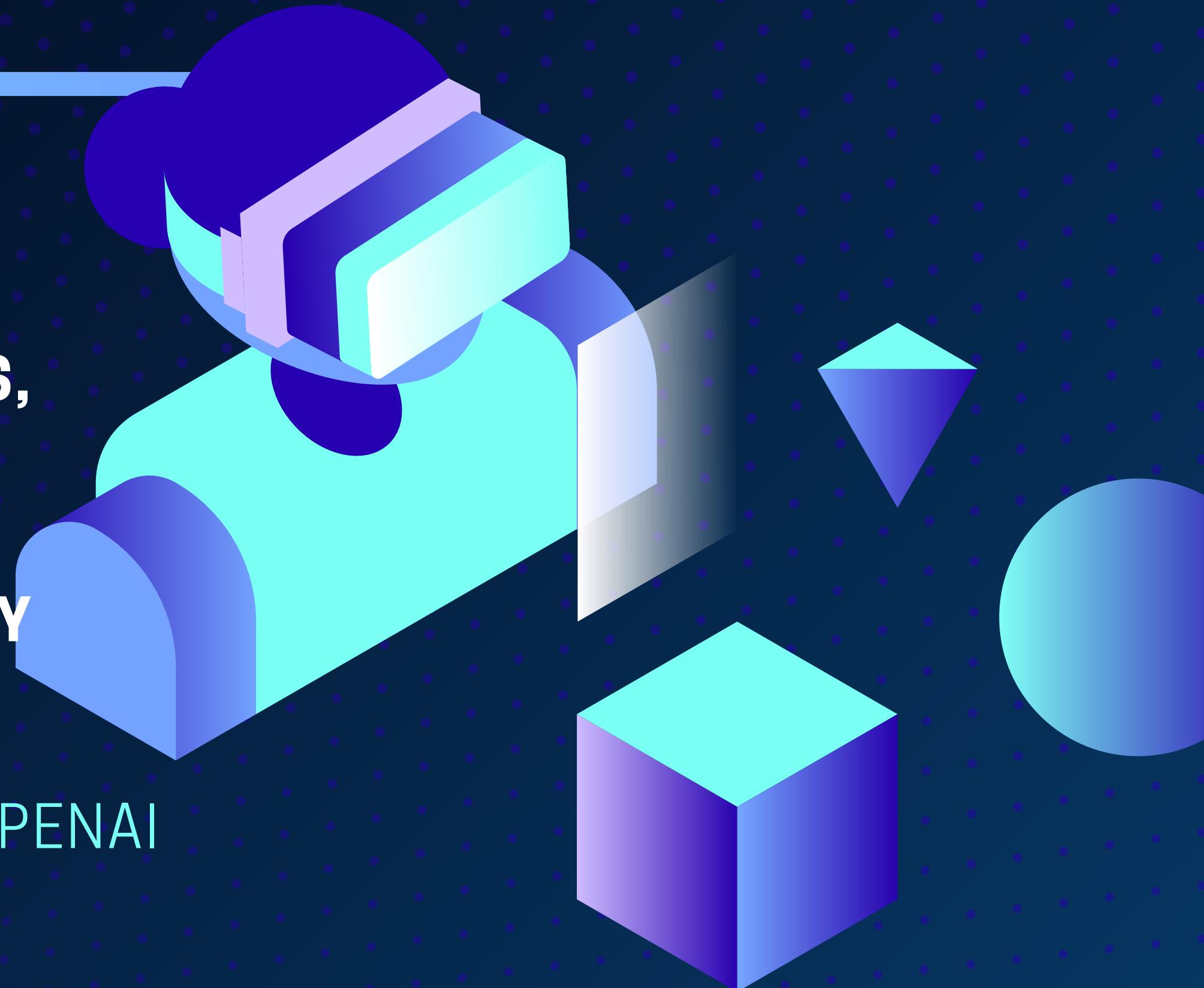
ANÁLISIS
PREDICTIVO



MITO

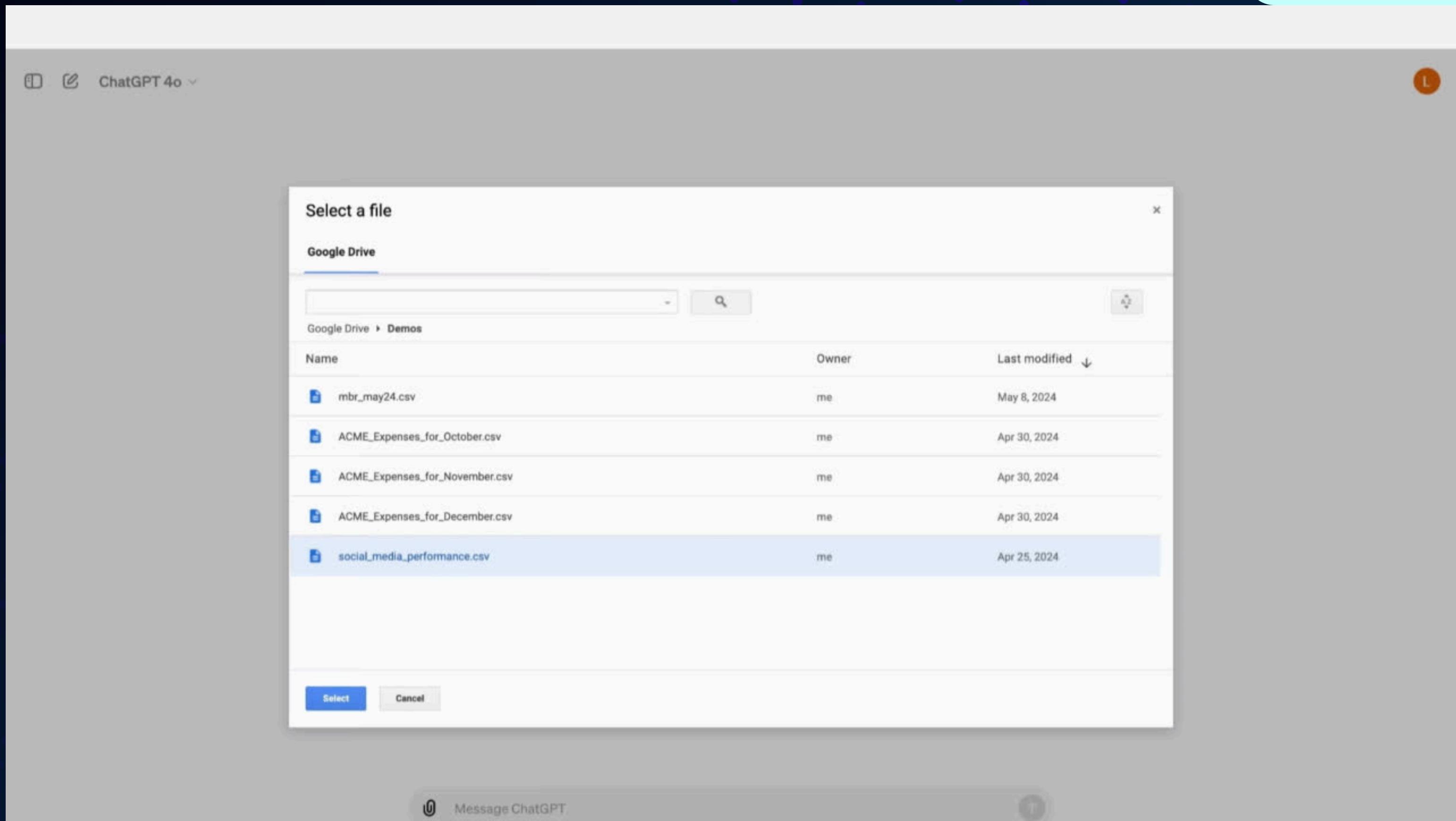
“EL 95% DE LO QUE LOS ESPECIALISTAS EN MARKETING UTILIZAN HOY EN DÍA CON AGENCIAS, ESTRATEGAS Y PROFESIONALES CREATIVOS SERÁ MANEJADO DE MANERA FÁCIL, CASI INSTANTÁNEA Y CASI SIN COSTO POR LA IA.”

SAM ALTMAN - CEO OPENAI



EL CICLO DE VIDA DE LOS DATOS





A ADMIN ACCOUNTADMIN

Search: how can i forecast next year sales call volume

All Tables and views Databases and schemas Data products Documentation

Documentation View all >

NAME

- Time-Series Forecasting (ML-Powered Function)
- FORECAST
- Anomaly Detection (ML-Powered Function)
- Using Window Functions

Tables and views View all >

NAME RELEVANT COLUMNS

- ACTIVITY_HISTORY CRM / SALES CALL_DISPOSITION CALL_DURATION_IN_SECONDS CALL_OBJECT CALL_TYPE
- LEAD_EXT CRM / SALES FIRST_CALL_DATE_TIME CONVERTED_CONTACT_ID ACTION_CADENCE_ASSIGNEE_ID
- CONTACT CRM / SALES FIRST_CALL_DATE_TIME DO_NOT_CALL FIRST_EMAIL_DATE_TIME ACCOUNT_ID BIRTH_DATE

Data products View all >

NAME

- Global Economic Forecast Data Prevedere
- Intelligent Event Data: Attended PredictHQ

PROJECTS > NIKOLAI_MANUAL_TEST PREVIEW

Model Quality 0.00 Share Publish Start training

Training Documents + Add Question 1 / 1 9 / 11 doc. 100% ⌂

list_of_units	what is the list of units inspected?
Score: 1.00	Injection Unit
Score: 1.00	Mold Clamping Unit
Score: 1.00	Hydraulic System
Score: 1.00	Temperature Control System
Score: 1.00	Ejector System
Score: 1.00	Lubrication System
Score: 1.00	Safety Devices
Score: 1.00	Control Software

inspection_grade	what is the inspection grade?
Score: 1.00	FAIL

inspection_recommendation	what is the recommended action?
Score: 1.00	regular maintenance

inspector	who is the inspector?
Score: 1.00	Emily Johnson

inspection_date	when was the inspection done?
-----------------	-------------------------------

SkiGear Co.
1661 Mesa Drive
Las Vegas, Nevada
123-555-5555

EQUIPMENT INSPECTION

MACHINE	SERIAL NUMBER	INSPECTION GRADE
Injection Molder	SGMM-12345	FAIL

INSPECTION SUMMARY

During the inspection of the injection molder, it was observed that one of the parts on the machine was failing. Specifically, the nozzle seemed to be clogged, which was causing the production of defective parts. The technician examined the nozzle and found that it had some residual material stuck inside. They cleaned the nozzle thoroughly, and the machine resumed production of good quality parts. The technician recommended regular maintenance of the machine to prevent such failures in the future. Overall, prompt action was taken to identify and rectify the issue, ensuring that production could resume without further delays.

Injection Unit	Defective
Mold Clamping Unit	Good ✓
Hydraulic System	Good ✓
Temperature Control System	Good ✓
Ejector System	Good ✓
Lubrication System	Good ✓
Safety Devices	Good ✓
Control Software	Good ✓

Emily Johnson Inspected by Date 2023-03-01

SQL as RAG_SQL Connected

```

1 select snowflake.cortex.complete(
2   'snowflake-arctic',
3   concat(
4     'Answer the question based on the context. Be concise.',
5     'Context: ',
6     (
7       select content from wiki_retrieval
8       order by vector_l2_distance(
9         snowflake.ml.embed_text(
10           'e5-base-v2',
11           'Who are the suppliers of SkiMax Pro 9000?'
12           ) content_embedding
13     )
14   )
15 )

```

What products does AlpineEquipmentCo. supply?

What products does AlpineEquipmentCo. supply?

AlpineEquipment Co. supplies the following products:
GlacierGlider 2000, DryProof670, GlacierGlide 400



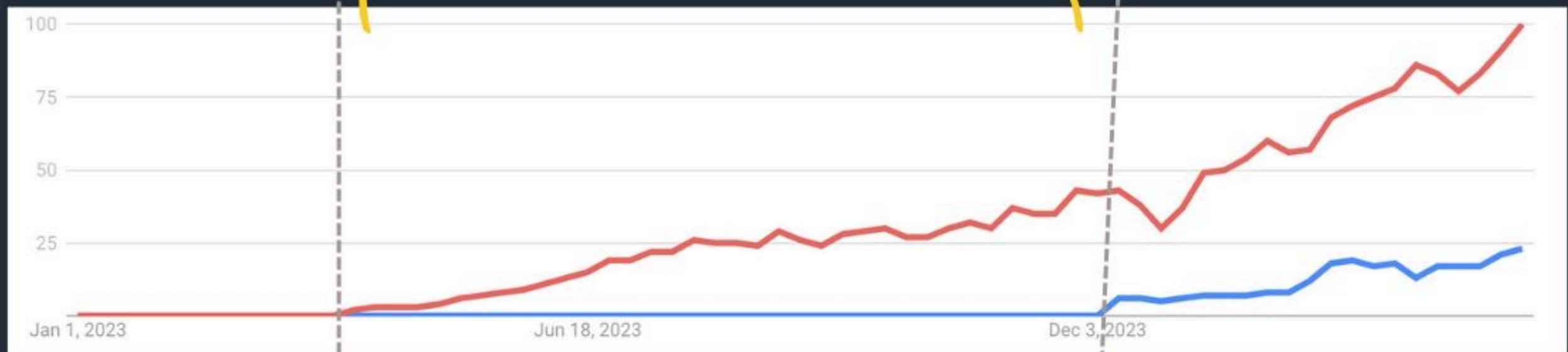
REALIDAD

- Los LLM no entienden el lenguaje, solo copian patrones
- Un LLM no sabe diferenciar emociones, realidades o siquiera que es lo que entiende
- Un LLM será tan “inteligente” como los datos que lo alimentan
- El hardware es, como siempre, una limitación: entrenar LLMs es caro y demandante

¿CONFIARÍAS EL FUTURO DE
TU EMPRESA EN ESTO?

— Google Trend “LLM” — Google Trend “Data Engineering”

“Let’s run a Poc !!” 🚀



“Wait ! We actually need clean data first...”



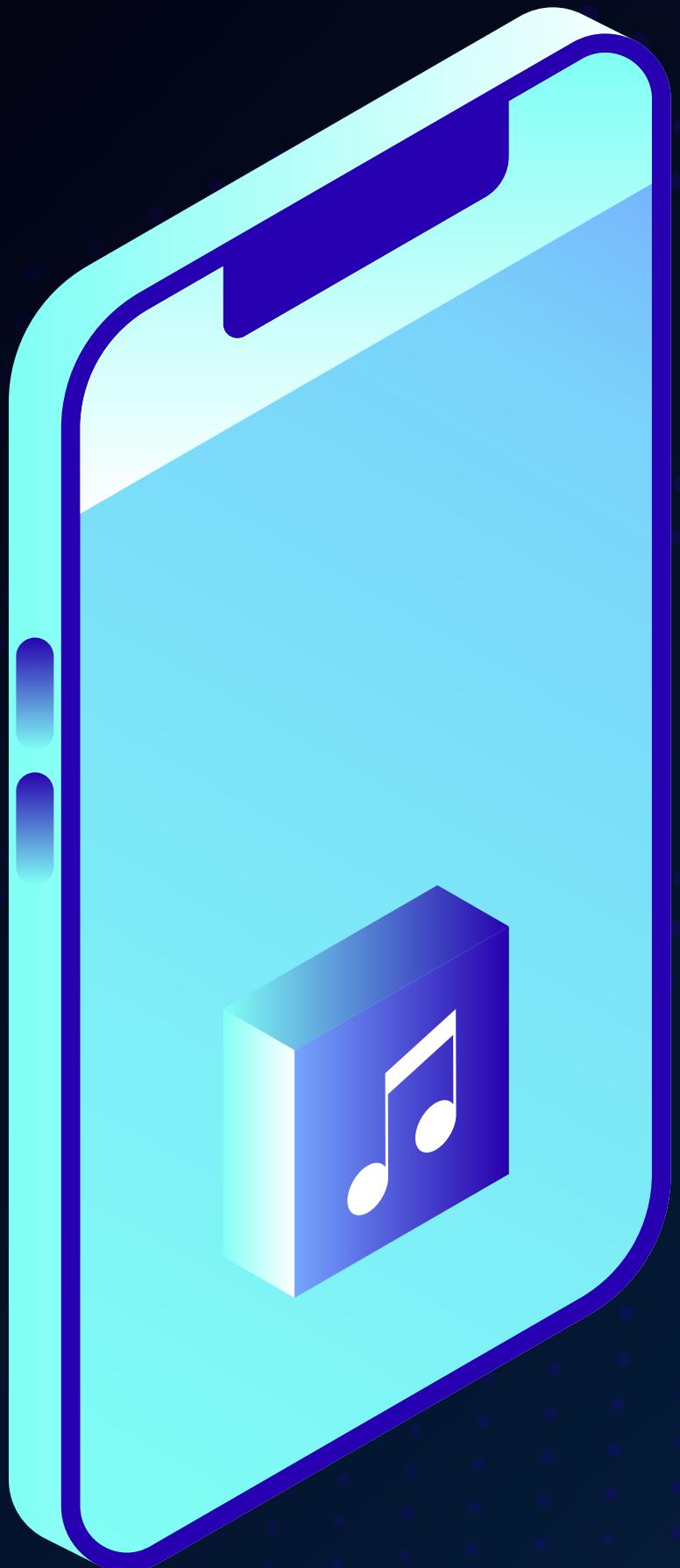
CONCLUSIONES

LOS LLM SON UNA HERRAMIENTA MUY PODEROSA

Hoy en día hay muchas aplicaciones en uso de LLMs aplicados al mundo de los datos, nos facilitan nuestro trabajo y ayudan a automatizar.

EL PENSAMIENTO CRÍTICO SIGUE SIENDO HUMANO

Podemos utilizar LLMs para facilitar las pequeñas tareas repetitivas, para enfocarnos en aquello que requiere de nuestro juicio de valor para obtener resultados de gran impacto.



¿PREGUNTAS?



Victoria Perez Mola