

CompWellGPT: Asistente en Completación de Pozos

Modelos de la IA basados en el cliente

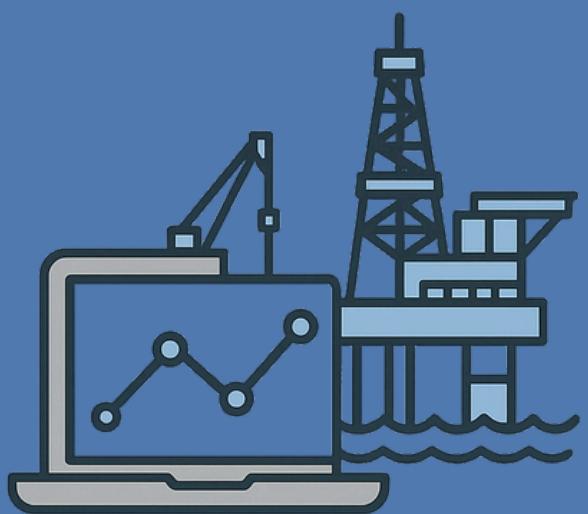
Por Neivys Luz González Gómez.





Índice

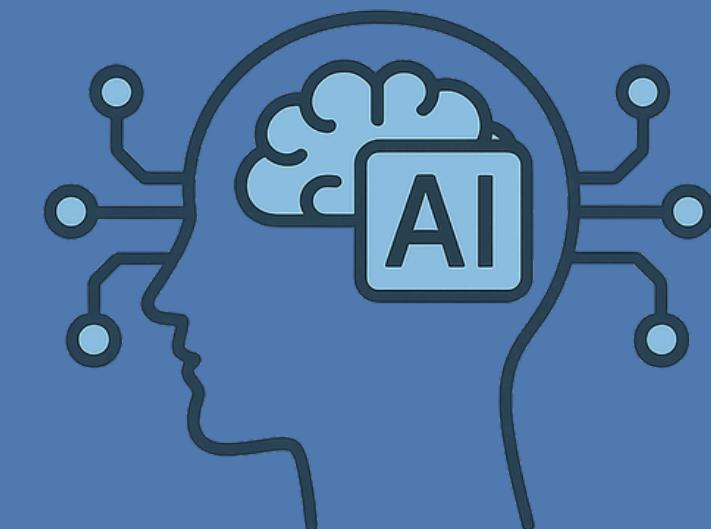
- 01.** Objetivo
- 02.** Caso de negocio
- 03.** Fuentes de datos
- 04.** Descripcion del Asistente
- 05.** Demo
- 06.** Costes estimados
- 07.** Linea de Investigacion
- 08.** Conclusiones



OBJETIVO

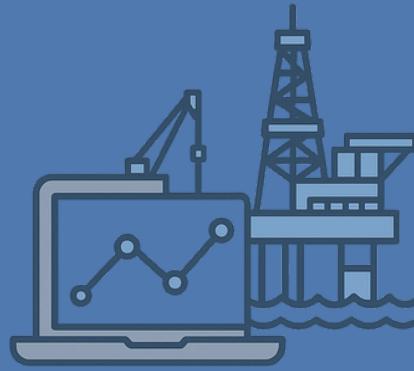
Desarrollar un asistente inteligente para la revisión técnica de pozos petroleros, capaz de analizar archivos estructurados (CSV) y no estructurados (PDF), responder preguntas técnicas, generar KPIs y detectar problemas, utilizando tecnologías de IA generativa.

Este asistente tiene como propósito agilizar el análisis documental, reducir errores humanos y mejorar la toma de decisiones técnicas durante las etapas de completación y/o reparación de pozos.



**ARTIFICIAL
INTELLIGENCE**

Descripción del Caso de Negocio



1

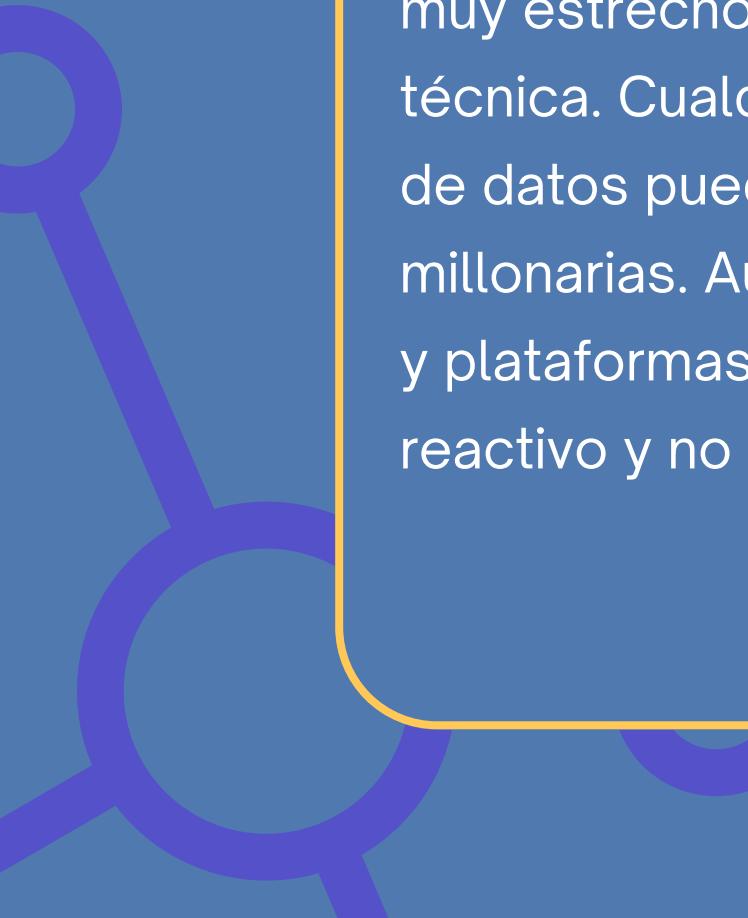
Entorno competitivo

La industria de la perforación y completación de pozos offshore es una de las más exigentes del sector energético. Se trabaja bajo presión extrema, con márgenes de error muy estrechos y documentación altamente técnica. Cualquier error en la interpretación de datos puede traducirse en pérdidas millonarias. Aunque existen sistemas SCADA y plataformas como WITSML, su enfoque es reactivo y no predictivo

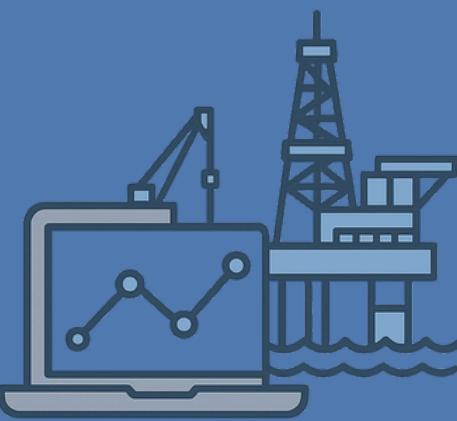
2

Necesidad a resolver

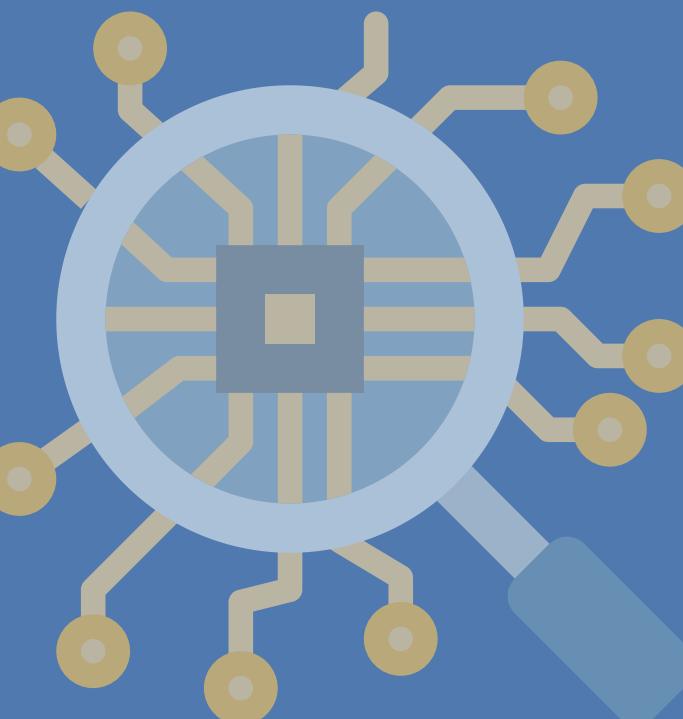
Actualmente, los ingenieros deben revisar manualmente cientos de páginas de documentación técnica. Esta tarea consume tiempo, está sujeta a errores humanos y retrasa decisiones críticas. Las empresas necesitan una solución automatizada, basada en inteligencia artificial, que facilite el análisis técnico, la generación de insights y la alerta ante comportamientos anómalos.



Fuente de datos para el asistente



- **Documentación técnica (PDF):** Resúmenes de pozo, informes de producción, análisis de presiones, historiales de completación, reportes de eventos no productivos.
- **Archivos estructurados (CSV):** Datos operacionales diarios, presiones de formación, producción acumulada, tiempo operativo vs. no productivo.
- **Formatos futuros:** PDFs con tablas embebidas, gráficos exportados desde SCADA, reportes híbridos, imágenes, logs eléctricos, datos de sensores en tiempo real (WITSML).



Descripción del Asistente CompWellGPT

CompWellGPT es un asistente técnico diseñado para apoyar tareas críticas de análisis en la completación y/o reparación de pozos. Funciona de forma interactiva, permitiendo ahorrar tiempo, reducir errores y tomar decisiones con base en datos reales.

Integraciones tecnológicas:

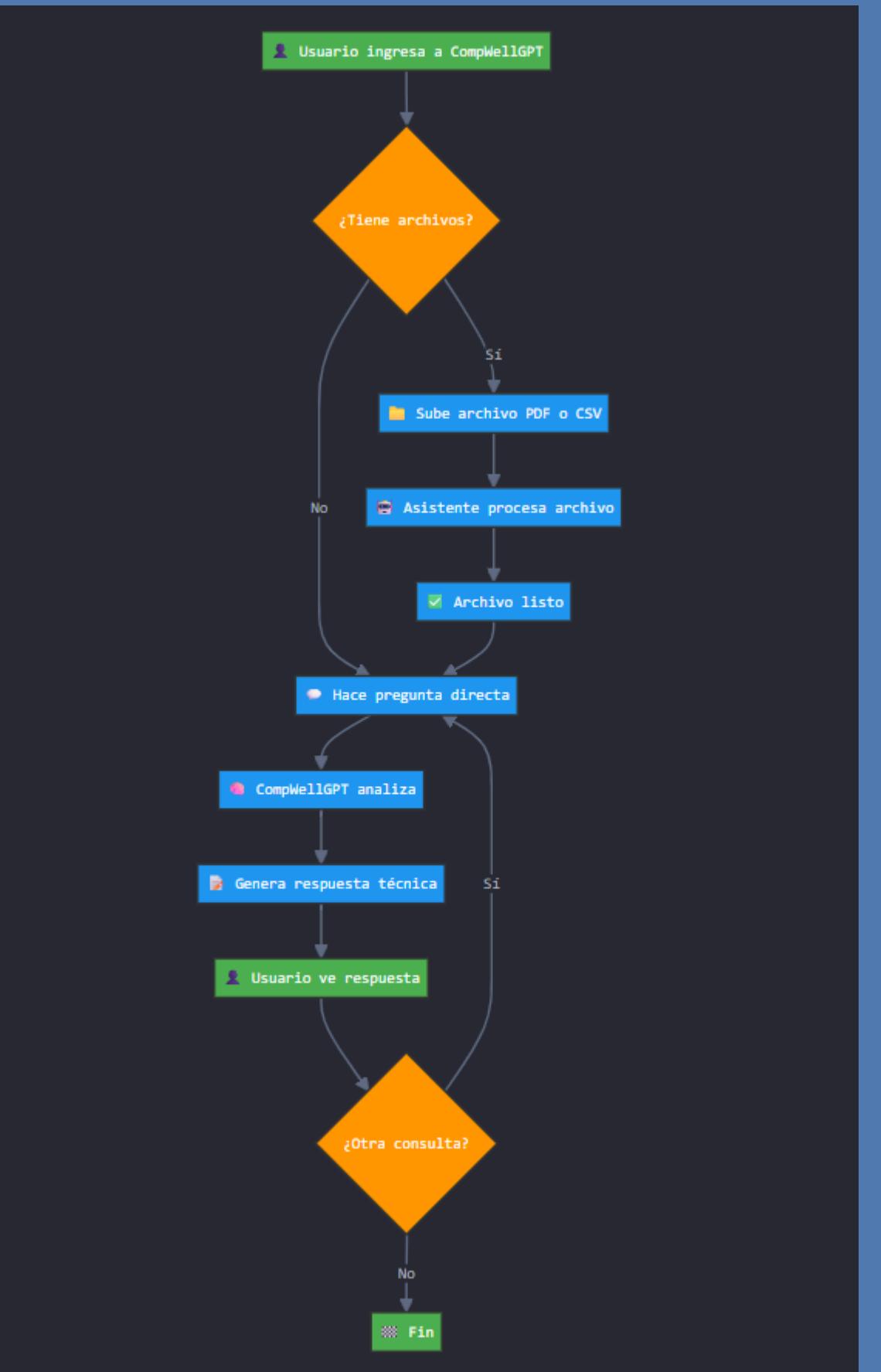
- **code_interpreter**: para análisis de datos, KPIs y detección de anomalías.
- **file_search**: para búsquedas semánticas en múltiples archivos PDF.
- Se generó una vectorstore con documentos técnicos de completacion de pozos

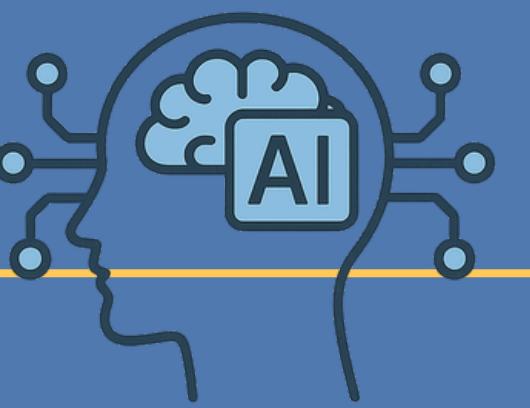
Funciones clave:

- Consultar documentos técnicos (PDF, DOCX, TXT).
- Analizar archivos CSV con datos de presión, producción, corte de agua, etc.
- Responder preguntas como “¿Cuál fue la presión máxima registrada en el pozo XX?”
- Resumir historial del pozo, puntos críticos y alertas técnicas.
- Detectar comportamientos anómalos (caídas de presión, baja producción, etc.).



Descripción del Asistente CompWellGPT



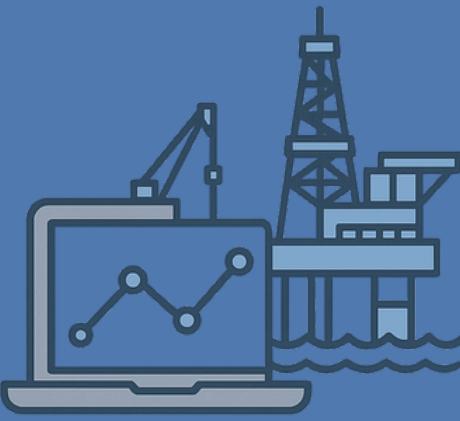


DEMO



CompWellGPT
Asistente en
Completabión de Pozos

Costes de Implementación

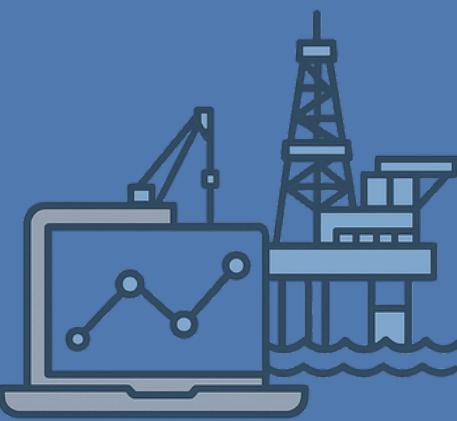


Componente API	Uso estimado diario	Uso mensual (22 días)	Costo por unidad*	Costo mensual estimado
Chat Completions	100 mensajes (~400k tokens)	~8.8M tokens	\$0.005 por 1k tokens (input/output)	~\$44.00 USD
Code Interpreter	15 sesiones	330 sesiones	~\$0.0125 por sesión	~\$4.13 USD
File Search	10 archivos por día (PDF, CSV)	220 archivos	~\$0.01 por archivo	~\$2.20 USD
Almacenamiento CSV	10 archivos por día (incluido)	220 archivos	incluido	\$0.00 USD
TOTAL aproximado (1 user)				~\$50.33 USD/mes

Los precios están calculados con base en GPT-4o y herramientas beta (code_interpreter, file_search). Se asume uso activo durante análisis técnico de un pozo por día. Escalable por número de usuarios.

Costes de Implementación

Costes Mensuales Combinados: API + Infraestructura en la Nube (Máquina Potente, 1 técnico trabajando sobre varios pozos)

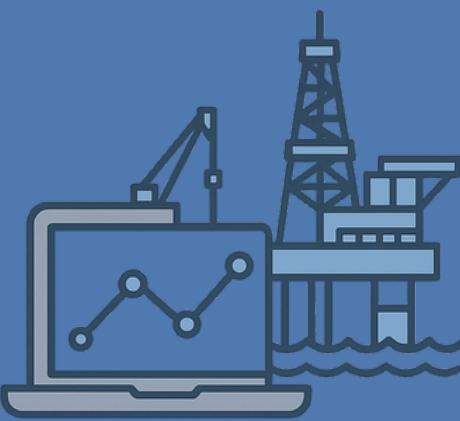


Componente	Descripción / Tipo de uso	Unidad estimada	Costo unitario	Costo mensual estimado
OpenAI API (GPT-4o + tools)	Chat, análisis de archivos, búsqueda y cálculos	22 días intensos	Ver desglose anterior	~\$50.33 USD
Servidor EC2 (g4dn.xlarge)	4 vCPU, 16 GB RAM, GPU NVIDIA T4 (para futuros modelos locales)	24/7 (~720 h/mes)	~\$0.526/hora	~\$379.00 USD
EBS (SSD rápido)	100 GB para almacenar docs, embeddings, cache y logs	100 GB	~\$0.10/GB	~\$10.00 USD
Tráfico de salida (egreso)	50 GB (respuestas, dashboards, archivos descargables)	50 GB	~\$0.09/GB	~\$4.50 USD
Backup S3 (opcional)	10 GB de respaldo automático semanal	10 GB	~\$0.023/GB	~\$0.23 USD
TOTAL Infraestructura Cloud	Infraestructura robusta, escalable y lista para múltiples usuarios			~\$393.73 USD

Incluye:

- **Instancia EC2 con GPU, lista para modelos locales.**
- **Almacenamiento para PDFs, CSVs y embeddings.**
- **Infraestructura compatible con Docker, Streamlit, FastAPI o Flask.**
- **Escalabilidad con balanceador de carga y autoscaling group.**

Línea de Investigación - Comparativa con Modelos Open Source



Objetivo:

Explorar la viabilidad de modelos LLM open source para sustituir parcial o totalmente a soluciones comerciales como OpenAI, asegurando privacidad, escalabilidad y reducción de costes.

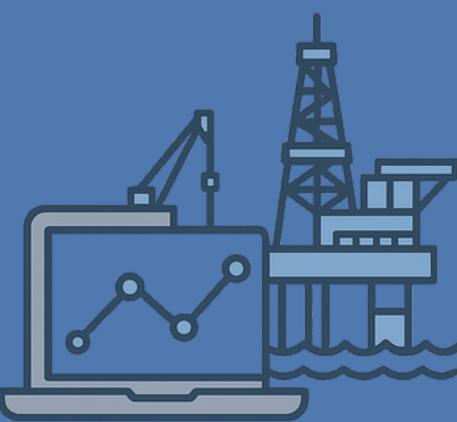
Hipótesis de trabajo:

“Modelos open source pueden ofrecer resultados competitivos para tareas técnicas si se afinan correctamente y se despliegan en infraestructuras locales controladas.”

Fases de la investigación:

- **Selección de modelos open source.**
- **Pruebas comparativas en openrouter.ai con los mismos archivos usados en CompWellGPT.**
- **Evaluación cualitativa y cuantitativa de las respuestas.**
- **Estimación de costes de infraestructura para implementación local.**

Línea de Investigación - Comparativa con Modelos Open Source



Pruebas realizadas:

- Las pruebas fueron ejecutadas en [OpenRouter.ai](#), evaluando archivos reales utilizados también en CompWellGPT.
- Se analizaron modelos Qwen 14B, Mistral 7B Instruct y Gemma 4B, cargando archivos CSV técnicos y formulando preguntas reales del dominio petrolero.

OpenRouter Search /

New Room To: Mistral 7B Instruct (free) Qwen 14B Gemma 3n 4B (free)

¿Cuáles son los pozos...

Untitled Chat

Mistral 7B Instruct (free) | DeepInfra

Utilizando el archivo "Produccion_de_Pozos.csv", he analizado lo que se refiere a "API gravedad crudo" en la columna "Contenido_sono_crudo" para determinar los tipos de crudos que producen. Con base en esto, identifico siguientes tipos de crudos que son producidos:

PZ-103 | 20.29 |
PZ-104 | 30.8 |
PZ-105 | 31.24 |
PZ-106 | 42.61 |
PZ-107 | 23.92 |
PZ-108 | 30.17 |

Qwen 14B — Qwen3 14B (free) | Chutes

Reasoning (thought for 41.6s)

Los pozos con mayor producción diaria (en barriles por día, bpd) según los datos proporcionados son los siguientes:

Los pozos PEN-xx (PEN-001 a PEN-054) son los principales en términos de producción, seguidos por pozos PZ-xx (PZ-055 a PZ-104), que tienen valores más bajos en comparación. Si se necesitan los 10 primeros pozos, puedo detallarlos también.

Gemma 3n 4B (free) | Google AI Studio

De acuerdo con los datos proporcionados, aquí están los pozos con la mayor producción diaria (Produccion_diaria_bpd), ordenados de mayor a menor:

Es importante notar que algunos pozos tienen varias observaciones que indican diferentes problemas. He ordenado en función del valor más alto de "Produccion_diaria_bpd".

Web Search

¿Qué pozos presentan problemas de arenamiento o presencia de arena?

Mistral 7B Instruct (free) | DeepInfra

Utilizando el archivo "Produccion_de_Pozos.csv", he analizado lo que se refiere a "API gravedad crudo" en la columna "Contenido_sono_crudo" para determinar los tipos de crudos que producen. Con base en esto, identifico siguientes tipos de crudos que son producidos:

PZ-103 | 20.29 |
PZ-104 | 30.8 |
PZ-105 | 31.24 |
PZ-106 | 42.61 |
PZ-107 | 23.92 |
PZ-108 | 30.17 |

Qwen 14B — Qwen3 14B (free) | Chutes

Reasoning (thought for 171.9s)

Los pozos producen **crudos de diferentes densidades**, clasificados según la gravedad API (American Petroleum Institute), un índice que mide la densidad del crudo en comparación con el agua. La clasificación típica es la siguiente:

La mayoría de los pozos están en el rango de crudo medio, seguido por los de crudo ligero, mientras que los de crudo pesado son un pequeño porcentaje. El crudo pesado suele asociarse con mayores porcentajes de corte de agua y problemas operativos como obstrucciones causadas por parafina o arena.

Gemma 3n 4B (free) | Google AI Studio

Para determinar los tipos de crudos que producen cada pozo, necesito información adicional sobre las propiedades del crudo (como gravedad API). La columna "API_gravedad_crudo" en tu archivo CSV proporciona una indicación, pero no es suficiente para identificar el tipo específico de crudo.

Web Search

¿Qué pozos presentan problemas de arenamiento o presencia de arena?

Mistral 7B Instruct (free) | DeepInfra

En el archivo "Produccion_de_Pozos.csv", he identificado los pozos en los que se presenta problemas de arenamiento o presencia de arena basándome en:

Los pozos identificados que presentan un elevado contenido de arena en el fluido o adversos problemas de arenamiento son PEN-008, PEN-036, PEN-037, PEN-058, PEN-059, PEN-062, PEN-063, PEN-064, PEN-067, PEN-098, y PEN-102.

Qwen 14B — Qwen3 14B (free) | Chutes

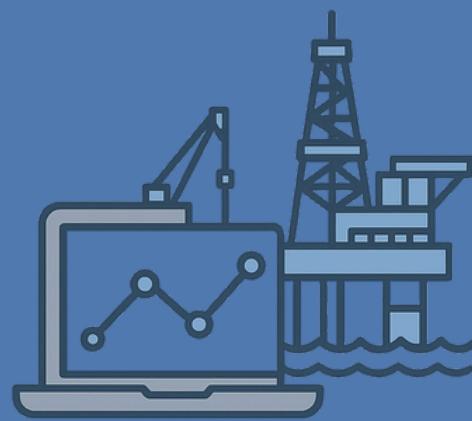
Reasoning (thought for 56.7s)

Los pozos que presentan **problemas de arenamiento, presencia de arena, o riesgo de arenamiento** según el campo "Problema Operacional" son los siguientes:

Estos pozos requieren monitoreo o intervención para mitigar la presencia de arena o evitar daños mecánicos en el equipo de producción (tubing, válvulas, etc.).

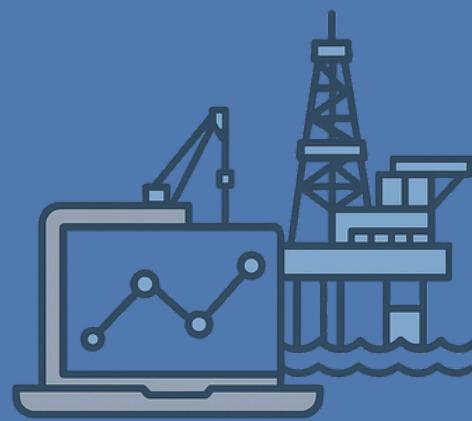
Web Search

Línea de Investigación - Comparativa con Modelos Open Source



Pregunta / Modelo	CompWellGPT	Qwen 14B	Mistral 7B	Gemma 3n 4B
1. Pozos con mayor producción	Lista top 5 pozos con bpd, rápida y directa	Lista top 5 con bpd y detalles operacionales	Lista con filtro >1000 bpd, enfoque técnico	Lista extensa, no filtra, da todos los pozos
2. Tipos de crudo producidos	Clasificación en ligero, medio y pesado + conteo	Clasificación + explicación por gravedad API	Lista numérica con API de 5 pozos puntuales	Dice que no puede clasificar con certeza
3. Pozos con arenamiento	Lista directa y resumida de pozos con arenamiento	Explica tipos de problemas + interpretación larga	Lista con pocos pozos si tienen columnas claras	Tarda, pero da todos sin filtrar ni interpretar
Velocidad de respuesta	⚡ Rápido	Lento pero detallado	Rápido pero limitado	Lento
Interpretación técnica	✓ Alta (filtra, resume y propone analizar más)	✓ Buena (detalla si se le insiste)	✗ Limitada (solo lectura de columnas)	✗ No interpreta, solo lee

Línea de Investigación - Comparativa con Modelos Open Source



Comparativa de Costes - CompWellGPT vs Modelos Open Source

Elemento	CompWellGPT (OpenAI API + Nube)	Open Source (Qwen/Mistral en AWS)
Uso API	~\$53 USD/mes	\$0 (modelo local)
Infraestructura (GPU 6h/día)	~\$433 USD/mes	~\$433 USD/mes
Costo Total Estimado	~\$486 USD/mes	~\$433 USD/mes
Escalado (2–3 usuarios)	~\$550–650 USD/mes	~\$480–550 USD/mes
Privacidad	Limitada (depende de OpenAI)	Alta (infraestructura propia)

🔧 Ambos escenarios suponen:
Instancia g4dn.xlarge en AWS (GPU).
Capacidad para cargar archivos PDF, CSV y generar embeddings.
Despliegue con Docker + RAG + Streamlit o interfaz propia.

Conclusiones



- CompWellGPT demostró ser altamente eficaz en tareas técnicas como clasificación de crudos, identificación de pozos con mayor producción y detección de problemas como arenamiento.
- Los modelos open source evaluados (Qwen, Mistral) muestran capacidades razonables pero requieren contexto bien estructurado y afinación para alcanzar el mismo nivel interpretativo.
- La privacidad y la escalabilidad se ven fortalecidas al migrar a infraestructura local con modelos open source.
- El coste mensual estimado usando OpenAI (CompWellGPT) es ~\$486 USD/mes, mientras que un despliegue local open source se mantiene en ~\$433 USD/mes, con opción a reducirlo si se usa hardware propio.



CompWellGPT representa un avance técnico en el análisis inteligente de datos de pozo. Su capacidad para interpretar, resumir y responder preguntas técnicas con rapidez lo posiciona como un copiloto digital valioso en operaciones de completación.

Próximo paso: Evaluar integración vía RAG multimodal (CSV, PDF, imágenes), y analizar escenarios de migración a modelos open source en infraestructura privada o híbrida..