



HACKERSPACE
VALENCIA



Taller Práctico:



deepseek

Instalación Local

Configuración paso a paso para un rendimiento óptimo

Optimización

Técnicas avanzadas de ajuste y personalización

Casos Prácticos

Ejemplos y aplicaciones prácticas

Sábado 08 Feb • 11:30h • Hackerspace Valencia

DeepSeek R1: Democratizando la IA



¿Cómo va a funcionar el taller?

1. Introducción y Contexto
2. Aspectos Técnicos de DeepSeek R1
3. Herramientas y Aplicaciones Prácticas
4. Benchmarks y ¿Cómo saber qué modelo usar?
5. Manos a la obra
6. Recursos y Próximos Pasos
7. Discusión y Q&A



<https://hackvlc.es/workshops/deepseek/>



Welcome to
**HACKERSPACE
VALENCIA**





```
$ whoami
```

Ignacio Delgado

```
$ cat role.txt
```

Software Engineer

```
$ cat contact.txt
```

```
web      ignaciodelgado.com
```

```
email    me@igdel.com
```



BY

SA

¿Qué es un LLM?

1. Introducción y Contexto
2. Aspectos Técnicos de DeepSeek R1
3. Herramientas y Aplicaciones Prácticas

MODELOS “ABIERTOS” vs CERRADOS

1. Introducción y Contexto
2. Aspectos Técnicos de DeepSeek R1
3. Herramientas y Aplicaciones Prácticas

STARGATE

1. Introducción y Contexto
2. Aspectos Técnicos de DeepSeek R1
3. Herramientas y Aplicaciones Prácticas

MODELOS “ABIERTOS” vs CERRADOS

- Los gigantes tecnológicos (OpenAI, Anthropic, Google) dominaban el mercado.
- Inversiones multimillonarias en hardware requeridas
- Secretismo extremo sobre arquitecturas
- Modelos abiertos significativamente menos capaces

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

El Impacto Inmediato de DeepSeek

1. Introducción y Contexto
2. Aspectos Técnicos de DeepSeek R1
3. Herramientas y Aplicaciones Prácticas

AI optimism supercharged Wall Street in recent years

— Magnificent 7 market capitalization



Source: LSEG

Created by Thomson Reuters

**Nvidia's \$589 Billion DeepSeek Rout Is Largest in
Market History - Yahoo Finance
(Un tercio del PIB de España)**

Contenidos ocultar

[Inicio](#)

Listas según FMI, Banco Mundial y The World Factbook (2000-2023)

[PIB según reportes de cada país \(2021\)](#)

[PIB en Europa según Eurostat](#)

[Véase también](#)




[Referencias](#)

[Enlaces externos](#)

20	 Suiza	905 684
21	 Polonia	842 172
22	 Taiwán	751 930
23	 Bélgica	627 511
24	 Argentina	621 833
25	 Suecia	597 110
26	 Irlanda	589 569
27	 Noruega	546 768
28	 Austria	526 182
29	 Israel	521 688
30	 Tailandia	512 193
31	 Emiratos Árabes Unidos	509 179
32	 Singapur	497 347
33	 Bangladés	446 349
34	 Filipinas	435 675
35	 Vietnam	433 356
36	 Malasia	430 895
37	 Dinamarca	420 800
38	 Egipto	398 397
39	 Nigeria	390 002
40	 Hong Kong	385 546
41	 Sudáfrica	380 906
42	 Irán	366 438
43	 Colombia	363 835
44	 Rumanía	350 414
45	Chile	344 400

19	 Suiza	991 114
—	 República de China	850 528
20	 Suiza	807 706
21	 Polonia	688 176
22	 Argentina	632 770
23	 Suecia	585 939
24	 Noruega	579 267
25	 Bélgica	578 604
26	 Irlanda	529 244
27	 Israel	522 033
28	 Emiratos Árabes Unidos	507 534
29	 Tailandia	495 340
30	 Nigeria	477 836

19	 Suiza	680 600	2017
20	 Arabia Saudita	678 500	2017
21	 Argentina	616 900	2017
22	 Taiwán	571 500	2017
23	 Suecia	541 900	2017
24	 Polonia	510 000	2017
25	 Bélgica	491 700	2017
26	 Tailandia	437 800	2017
27	 Irán	427 700	2017
28	 Austria	409 300	2017
29	 Nigeria	394 800	2017
30	 Noruega	392 100	2017
31	 Emiratos Árabes Unidos	378 700	2017
32	 Israel	348 000	2017

31	 Emiratos Árabes Unidos	509 179
32	 Singapur	497 347
33	 Bangladés	446 349



[https://es.wikipedia.org/wiki/Anexo:Pa%C3%ADses_por_PIB_\(nominal\)](https://es.wikipedia.org/wiki/Anexo:Pa%C3%ADses_por_PIB_(nominal))

Democratización de la IA

- Primer modelo abierto comparable a sistemas comerciales
- Rendimiento similar a OpenAI o1
- Código completamente abierto y documentado
- Ejecutable en hardware de consumo

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

Podríamos hablar de muchos más temas

1. Introducción y Contexto
2. Aspectos Técnicos de DeepSeek R1
3. Herramientas y Aplicaciones Prácticas

Si da tiempo, o después de la charla.

Fin del Monopolio

- El Momento Sputnik de la IA
- Similar al impacto del Sputnik en la carrera espacial
- Estados Unidos descubre que no tiene el monopolio
- Posible pivote de la industria hacia mayor apertura

Implicaciones para el Futuro

Un Nuevo Paradigma

- Desarrollo más abierto y eficiente de IA
- Menor dependencia de grandes centros de datos
- Democratización del acceso a IA avanzada
- Posible transformación del modelo de negocio actual

Preguntas Pendientes

- Costos reales de desarrollo
- Acceso a hardware no declarado
- Cumplimiento de restricciones comerciales
- Sostenibilidad del modelo abierto
- PRIVACIDAD

Arquitectura Base

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

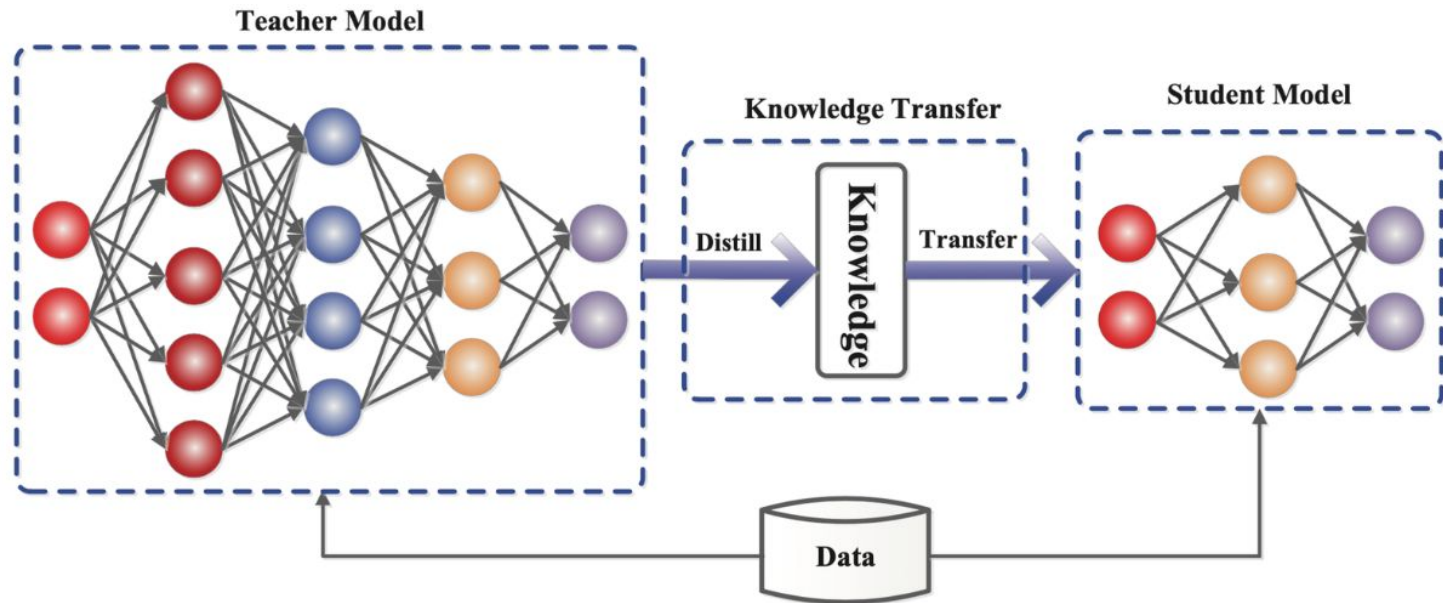
3. Herramientas y Aplicaciones Prácticas

DeepSeek-V3-Base (671B parámetros)

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas



Source: arXiv

data science dojo
— data science for everyone —

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

deepseek-r1

DeepSeek's first-generation of reasoning models with comparable performance to OpenAI-o1, including six dense models distilled from DeepSeek-R1 based on Llama and Qwen.

1.5b 7b 8b 14b 32b 70b 671b

↓ 10.7M Pulls ⌚ Updated 13 hours ago

7b



🔖 29 Tags

ollama run deepseek-r1



Updated 2 weeks ago

0a8c26691023 · 4.7GB

model	arch qwen2 · parameters 7.62B · quantization Q4_K_M	4.7GB
-------	--	-------

params	{ "stop": ["< begin_of_sentence >", "< end_of_sentence..."	148B
--------	---	------

template	{{- if .System }}{{ .System }}{{ end }} {{- range \$i, \$_ :...	387B
----------	---	------

license	MIT License Copyright (c) 2023 DeepSeek Permission is here...	1.1kB
---------	---	-------

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

Hardware mínimo y recomendado

Alternativas a CUDA

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

DeepSeek R1 Across the M4 Armada



1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

No hace falta usar el modelo de más de 600B

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

Modelo	Base	VRAM (Base × 1.2)	Explicación
1.5b	1.1GB	1.32GB	$1.1 \times 1.2 = 1.32\text{GB}$
7b	4.7GB	5.64GB	$4.7 \times 1.2 = 5.64\text{GB}$
8b	4.9GB	5.88GB	$4.9 \times 1.2 = 5.88\text{GB}$
14b	9.0GB	10.8GB	$9.0 \times 1.2 = 10.8\text{GB}$
32b	20GB	24GB	$20 \times 1.2 = 24\text{GB}$
70b	43GB	51.6GB	$43 \times 1.2 = 51.6\text{GB}$
671b	404GB	484GB	$404 \times 1.2 = 484\text{GB}$

- 1. Introducción y Contexto
- 2. Aspectos Técnicos de DeepSeek R1
- 3. Herramientas y Aplicaciones Prácticas

[Discord](#)[GitHub](#)[Models](#)

deepseek-r1

DeepSeek's first-generation of reasoning models with comparable performance to OpenAI-o1, including six dense models distilled from DeepSeek-R1 based on Llama and Qwen.

[1.5b](#) [7b](#) [8b](#) [14b](#) [32b](#) [70b](#) [671b](#)

↓ 10.7M Pulls ⌚ Updated 13 hours ago

7b

29 Tags

ollama run deepseek-r1

1.5b	1.1GB	0a8c26691023 · 4.7GB
7b	4.7GB	Parameters 7.62B · quantization Q4_K_M 4.7GB
8b	4.9GB	begin_of_sentence >", "< end_of_sentence...
14b	9.0GB	}}{ .System }}{{ end }} {{- range \$i, \$_ :...
32b	20GB	copyright (c) 2023 DeepSeek Permission is here...
70b	43GB	
671b	404GB	

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

**Solo hemos tocado la superficie
Si queréis muchos más detalles
Podéis mirar el vídeo o hablar conmigo
después de la charla**



<https://www.youtube.com/watch?v=CiS9gDfYZ-w>

1. Introducción y Contexto

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

Ecosistema de herramientas

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

Ecosistema de herramientas

- Ollama como base
- Web UIs disponibles
- Extensiones y plugins

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

Casos de uso entre muchos

- Análisis de código
- Procesamiento de documentos
- RAG local

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

Herramientas destacadas

- Page-Assist (Chrome)
- Open Web-UI
- Roo-Code (VS Code)
- RAG personalizado

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

Integración con proyectos existentes

- LangChain
- Frameworks populares
- APIs propias

2. Aspectos Técnicos de DeepSeek R1

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

Selección de Modelos

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

Comparativas con modelos comerciales

Métricas clave

Interpretación de resultados

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

Creación de benchmarks propios (Para un futuro taller)

- Métricas personalizadas
- Evaluación de precisión
- Medición de rendimiento

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

¿Qué modelo usamos?

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

Por capacidad de hardware

Por caso de uso

Por requerimientos de velocidad/precisión

3. Herramientas y Aplicaciones Prácticas

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

¡MANOS A LA OBRA!

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Hemos instalado Ollama previamente (Si no, lo hacemos ahora)

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

ollama.com

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Análisis de código

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/RooVetGit/Roo-Code>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Web UI

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/open-webui/open-webui>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Para los “Power users”

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Click slider numbers to input manually.

Text Completion presets

Default

Response (tokens)

350

Context (tokens)

8192

Streaming

Unlocked

Neutralize Samplers

Sampler Select

Temperature

0,5

Top K

0

Top P

0,9

Typical P

1

Min P

0

Top A

0

TFS

1

Repetition Penalty

1,1

Rep Pen Range

0

Frequency Penalty

0

Presence Penalty

0

Mirostat

Mode

0

Tau

5

Eta

0,1

Skip Special Tokens

Seed

-1

Banned Tokens/Strings

some text as tokens
[420, 69, 1337]
"Some verbatim string"

User

September 21, 2024 8:42 PM

Tell me a joke about cats.

AI Assistant

September 21, 2024 8:42 PM

Why don't cats play poker in the jungle?

User

September 21, 2024 8:42 PM

Why?

AI Assistant

September 21, 2024 8:42 PM

Too many cheetahs! 🐾

User

September 21, 2024 8:43 PM

That was very funny!
Can you help me solve this equation?
$$x^2 + 2x + 4 = 0$$

AI Assistant

September 21, 2024 8:43 PM

I can help you with that!

To solve this quadratic equation, we can use the quadratic formula:
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Where:

- a = 1
- b = 2
- c = 4

Let's plug in the values:

Favorite characters to add them to HotSwaps

AI Assistant

2 Tokens
(2 Permanent)

More...

Search / Create tags

Creator's Notes

Description

Describe your character's physical and mental traits here.

Tokens: 0

First message

This will be the first message from the character that starts every chat.

Tokens: 0

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/SillyTavern/SillyTavern>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/ggerganov/llama.cpp>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

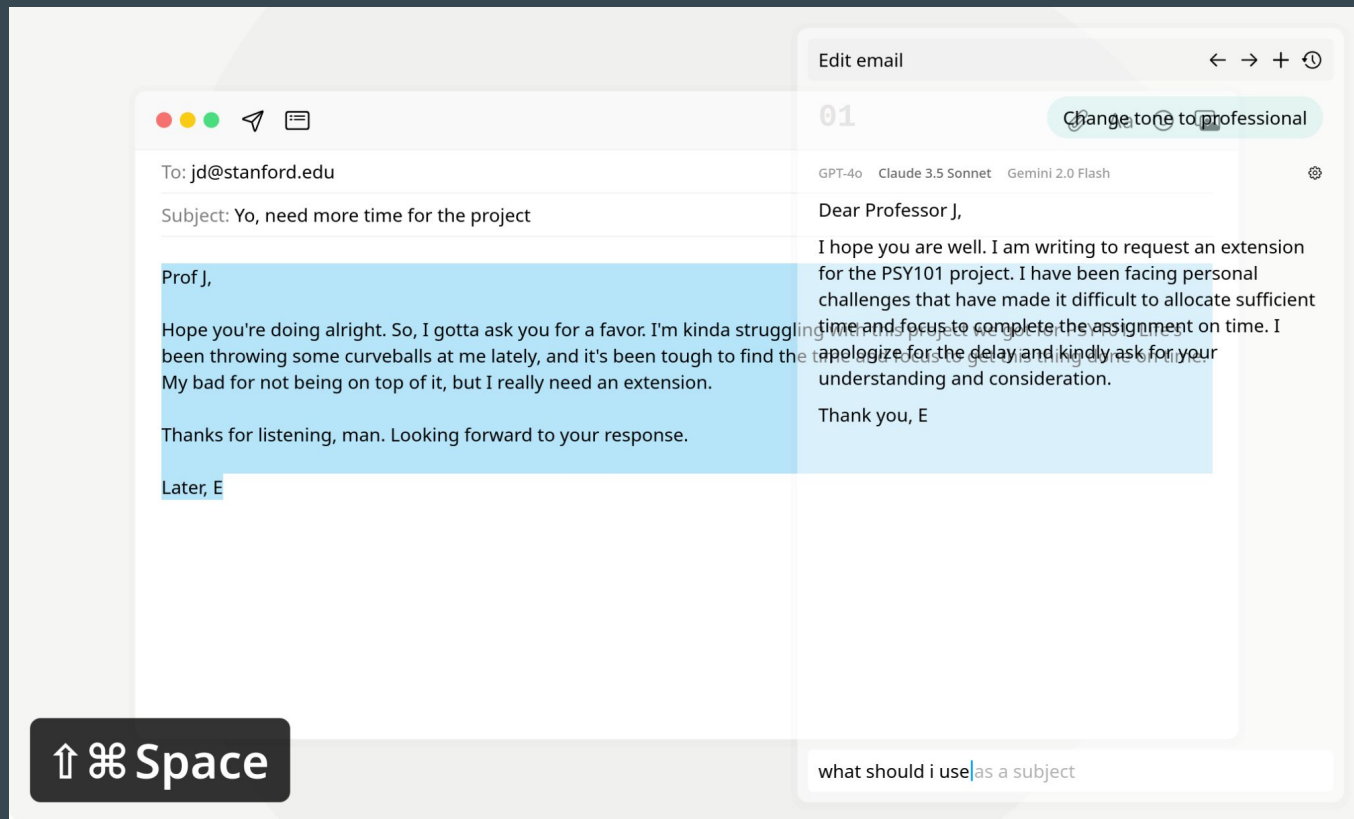
6. Recursos y Próximos Pasos

Para fancy Mac Users

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos



↑ ⌘ Space

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://intellibar.app/>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Para automatizaciones

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/n8n-io/n8n>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Chatear con PDFs, Mínimo 8b parámetros

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/hasan-py/chat-with-pdf-RAG>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

anythingllm.com

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Charlar con una web

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/n4ze3m/page-assist>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Este puede ser muy interesante,
Que la IA use el navegador,
Mínimo 14B parámetros

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

<https://github.com/browser-use/browser-use>

4. Benchmarks y ¿Cómo saber qué modelo usar?

5. Manos a la obra

6. Recursos y Próximos Pasos

Documentación oficial
Recursos de la comunidad, Hackerspace
Próximos desarrollos
Oportunidades de contribución

5. Manos a la obra

6. Recursos y Próximos Pasos

7. Discusión y Q&A

¿Alguna pregunta?

¿Cómo váis a usar estas herramientas?

Discussion Abierta

Si da tiempo, o después de la charla.

Fin del Monopolio

- El Momento Sputnik de la IA
- Similar al impacto del Sputnik en la carrera espacial
- Estados Unidos descubre que no tiene el monopolio
- Posible pivote de la industria hacia mayor apertura

Implicaciones para el Futuro

Un Nuevo Paradigma

- Desarrollo más abierto y eficiente de IA
- Menor dependencia de grandes centros de datos
- Democratización del acceso a IA avanzada
- Posible transformación del modelo de negocio actual

Preguntas Pendientes

- Costos reales de desarrollo
- Acceso a hardware no declarado
- Cumplimiento de restricciones comerciales
- Sostenibilidad del modelo abierto
- Privacidad

Se grabará un video con más información



```
$ whoami
```

Ignacio Delgado

```
$ cat role.txt
```

Software Engineer

```
$ cat contact.txt
```

```
web          ignaciodelgado.com
```

```
email       me@igdel.com
```



Muchas gracias