



CAMARGO

AYUNTAMIENTO
DE CAMARGO

Cursos de Verano

2023

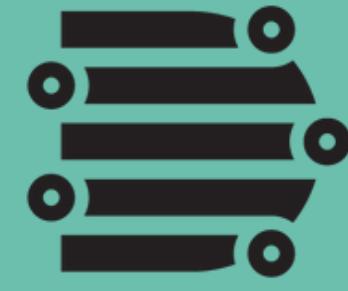
Universidad de Cantabria

50 UC
Universidad de Cantabria

CURSOS DE VERANO
Y DE EXTENSIÓN UNIVERSITARIA

GOBIERNO
de
CANTABRIA
CONSEJERÍA DE UNIVERSIDADES,
IGUALDAD, CULTURA Y DEPORTE
DIRECCIÓN GENERAL DE UNIVERSIDADES,
INVESTIGACIÓN Y TRANSFERENCIA

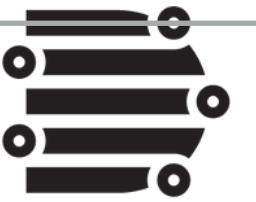
AÑO JUBILAR
LEBANIEGO
2023·2024



DEDUCE
DATA SOLUTIONS

DEEP INTO DATA SCIENCE

VII EDICIÓN

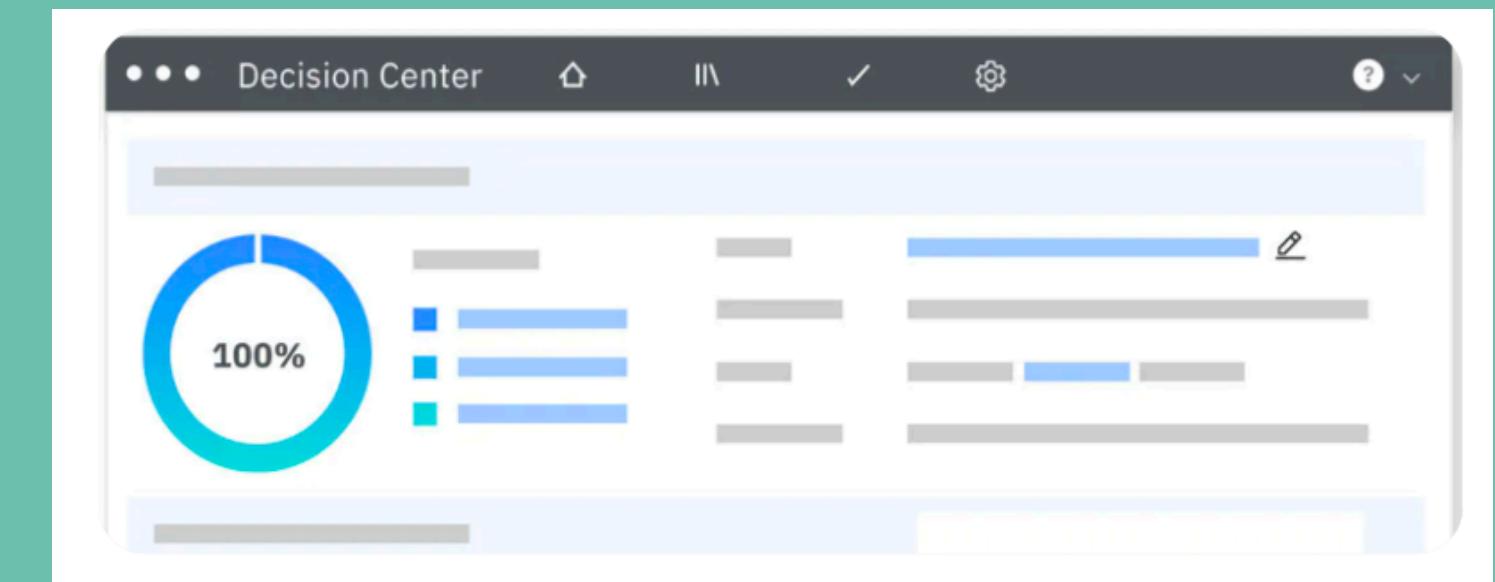


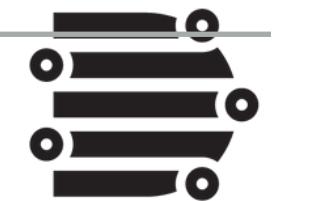
DEFINICIÓN

- ▶ Es un campo de la inteligencia artificial que se enfoca en la generación de contenido nuevo y original, como texto, imágenes, música y otros tipos de datos, que aprende a partir de un conjunto de datos existente y genera nuevos contenidos similares.

OBJETIVO

- ▶ Entender a grandes rasgos cómo funciona la IA Generativa
Saber qué es lo que la tecnología permite a día de hoy.
- ▶ Poder reflexionar cada uno de nosotros en posibles aplicaciones





HISTORIA

- ▶ Aunque empezó en los años 60, no dejaba de ser una curiosidad de alcance y capacidades limitadas
- ▶ Sin embargo, su efecto fue mucho más poderoso que lo que estaba previsto



El primer salto importante se produce en 2014: Generative Adversarial Networks (GANs)



HISTORIA

- ▶ En los últimos años, los modelos de generación de imágenes y texto han avanzado muy rápidamente



A brief history of generative AI spoken in pirate speak in the format of a limerick

Aye, mateys, 'tis a tale to be told
Of how AI learned to be bold
With neural networks and deep learning
Their capabilities, ever burning
Now they can write, translate, and even be cold

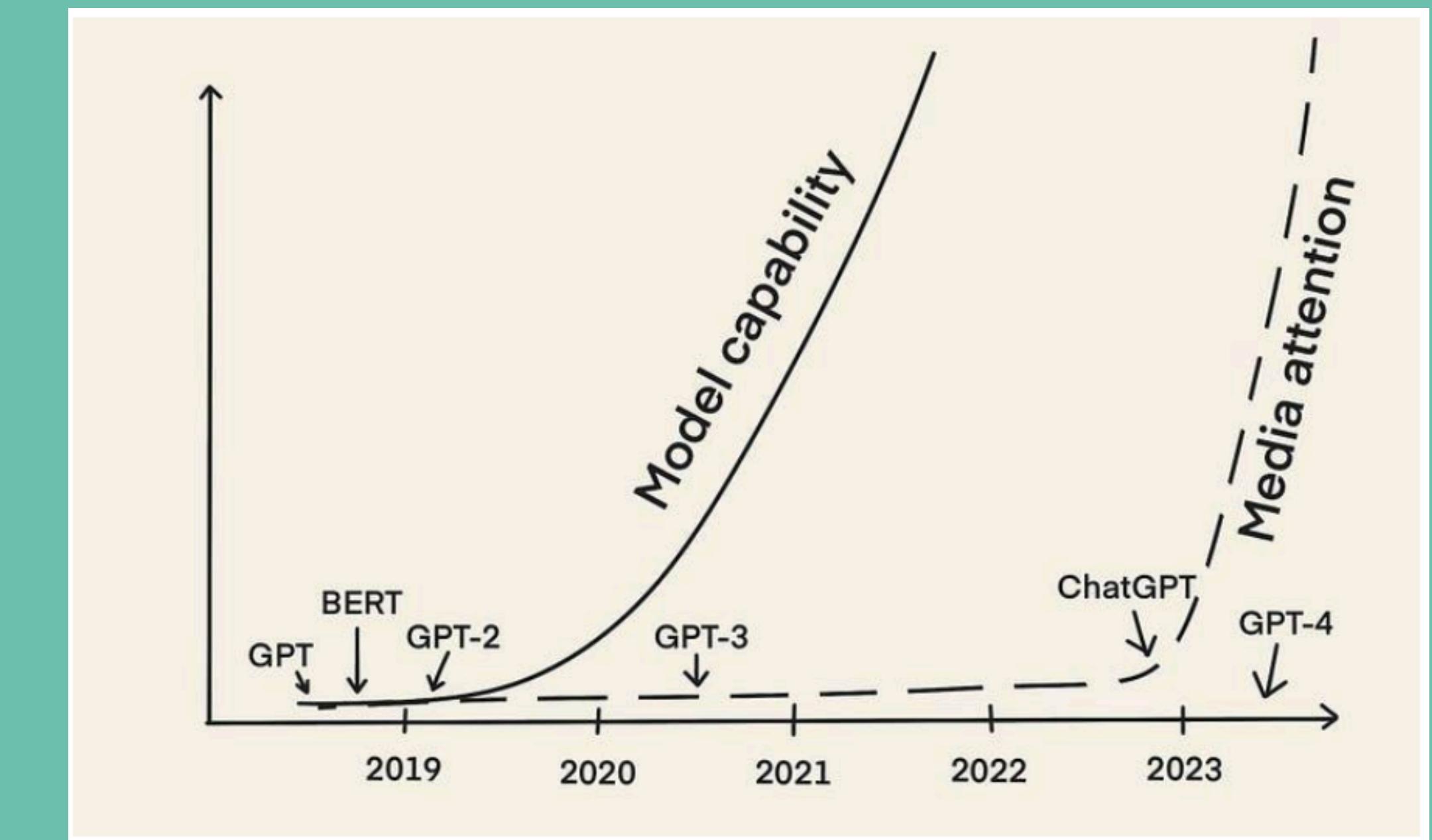
In the beginning, AI was crude
But with time, it has grown quite lewd
It learned to recognize faces
And even write in different places
Now it can generate text that's downright dude

GPT-2 was the next to take sail
A language model that would never fail
It could write articles and stories
Even generate deepfake glories
But its true power, some would say, is to avail

Now we have GPT-3, the grandest of all
It can generate text with a huge standing wall
It can write code and even design
It's hard to believe, but it's not a decline
With GPT-3, the future is looking tall

HISTORIA

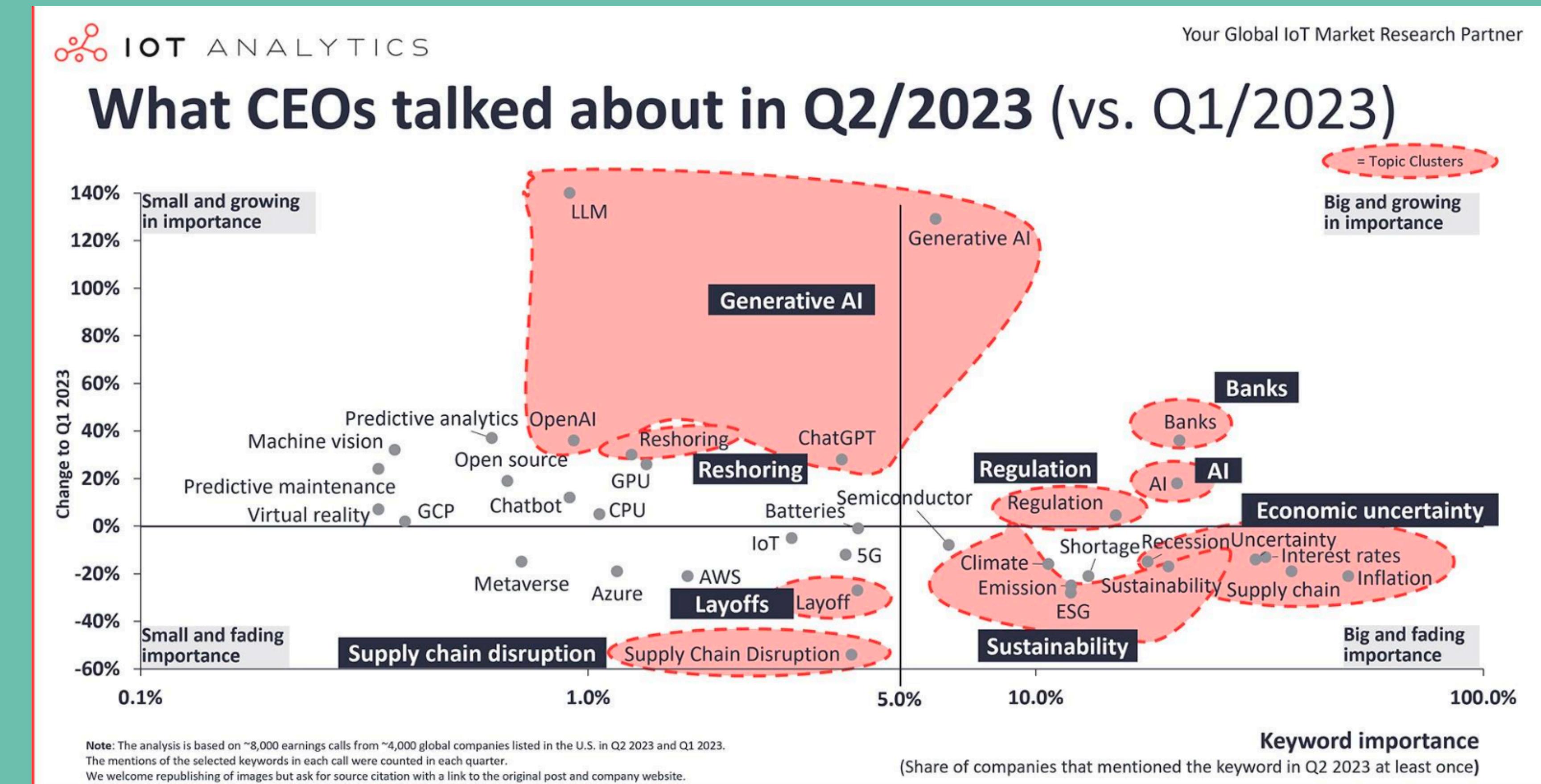
- ▶ Muy impresionante, muy divertido, pero...
¿para qué sirve? ¿Y cómo se vende esto?

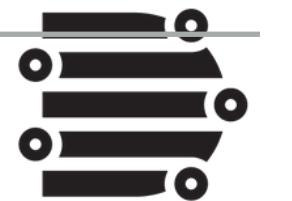




HISTORIA

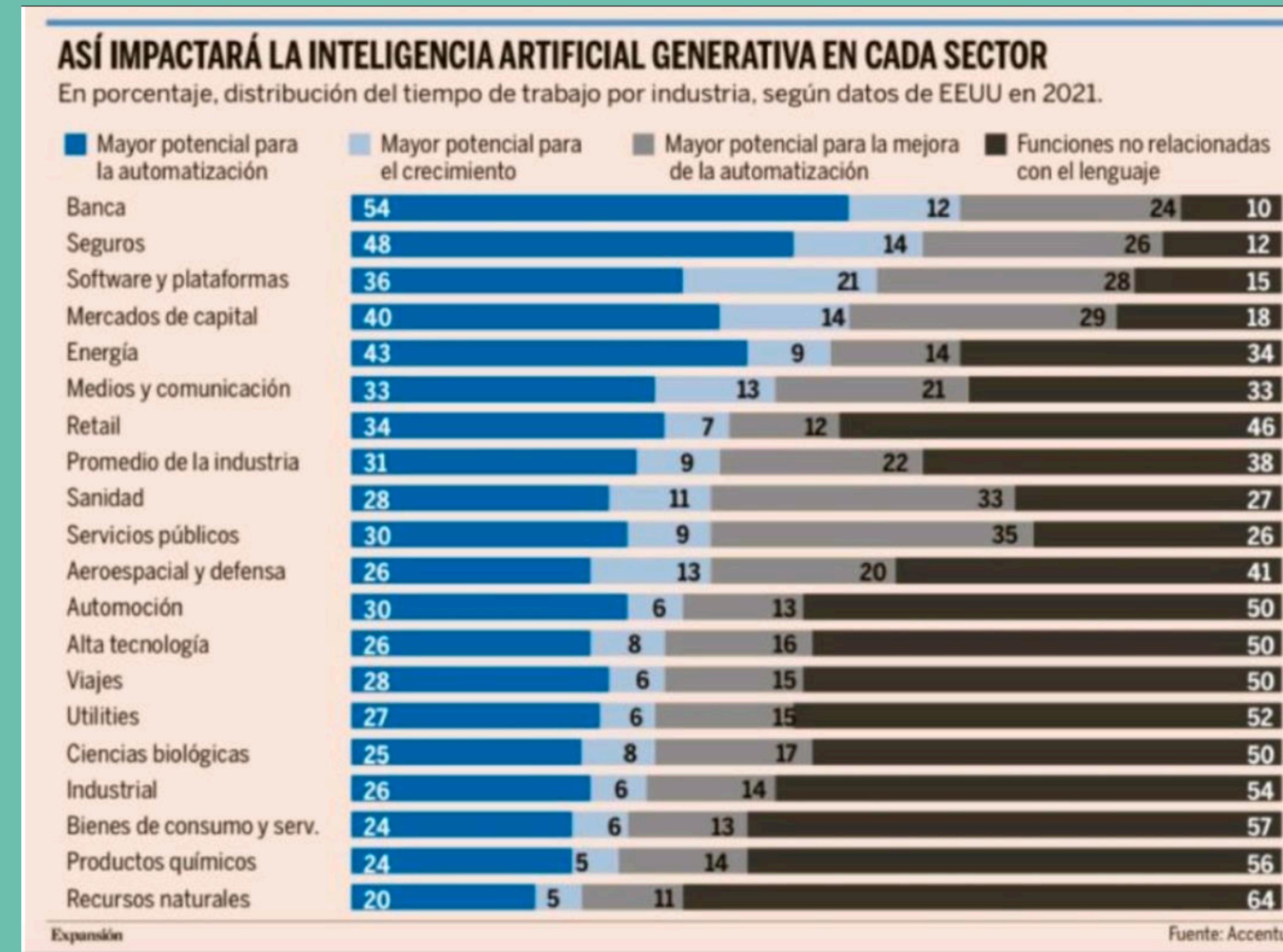
▶ ¿Hay hype?

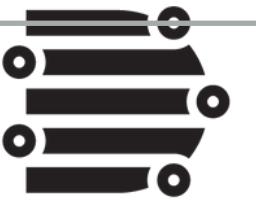




HISTORIA

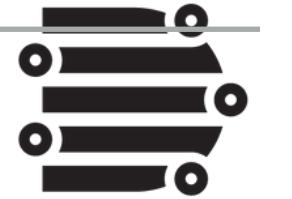
▶ ¿Hay hype?





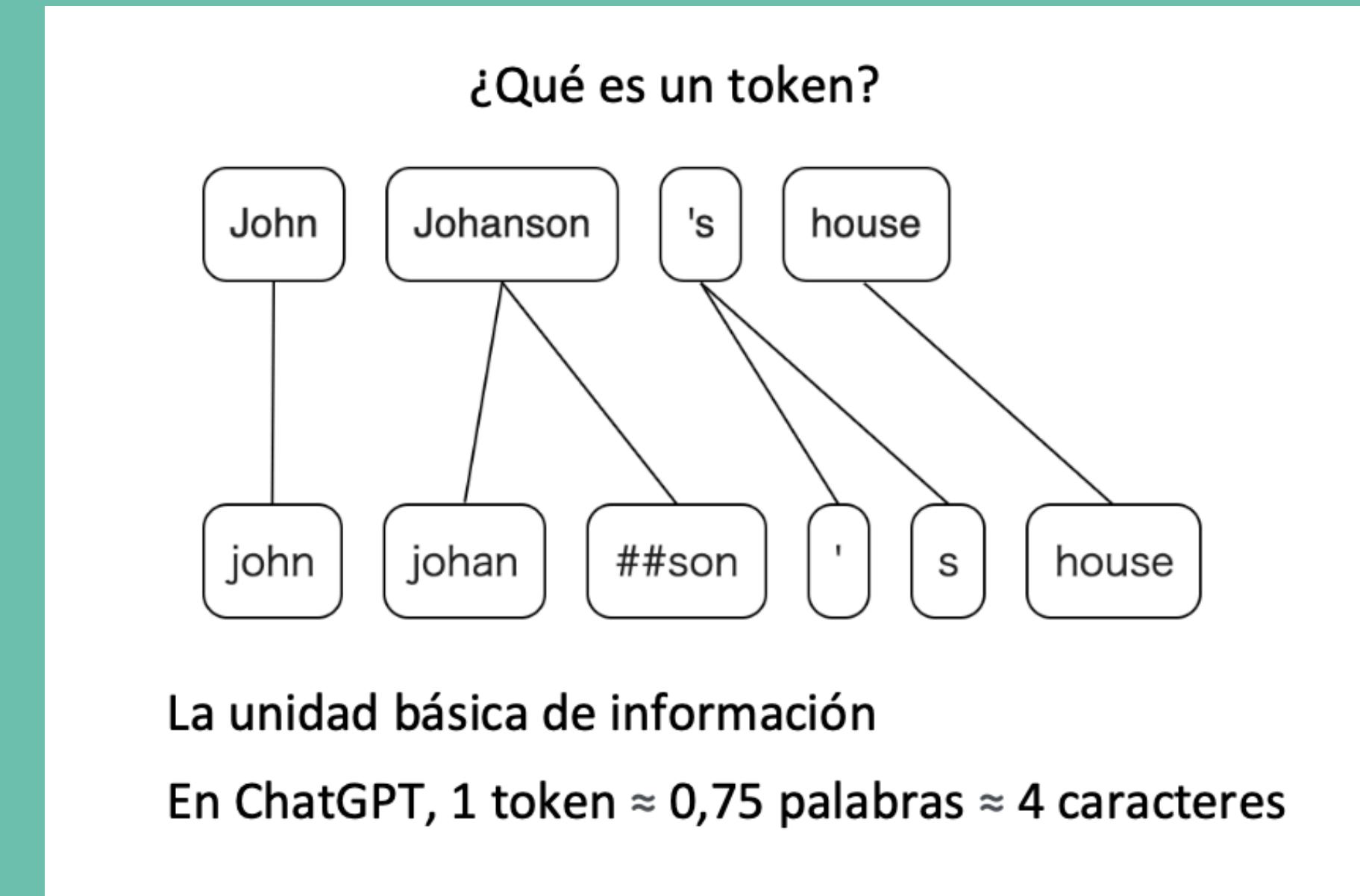
TIPOS DE IA GENERATIVA

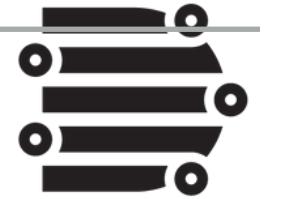
- ▶ **Texto**
- ▶ Imagen
- ▶ Otros
- ▶ Música, código, medicamentos



TEXTO: RASGOS GENERALES

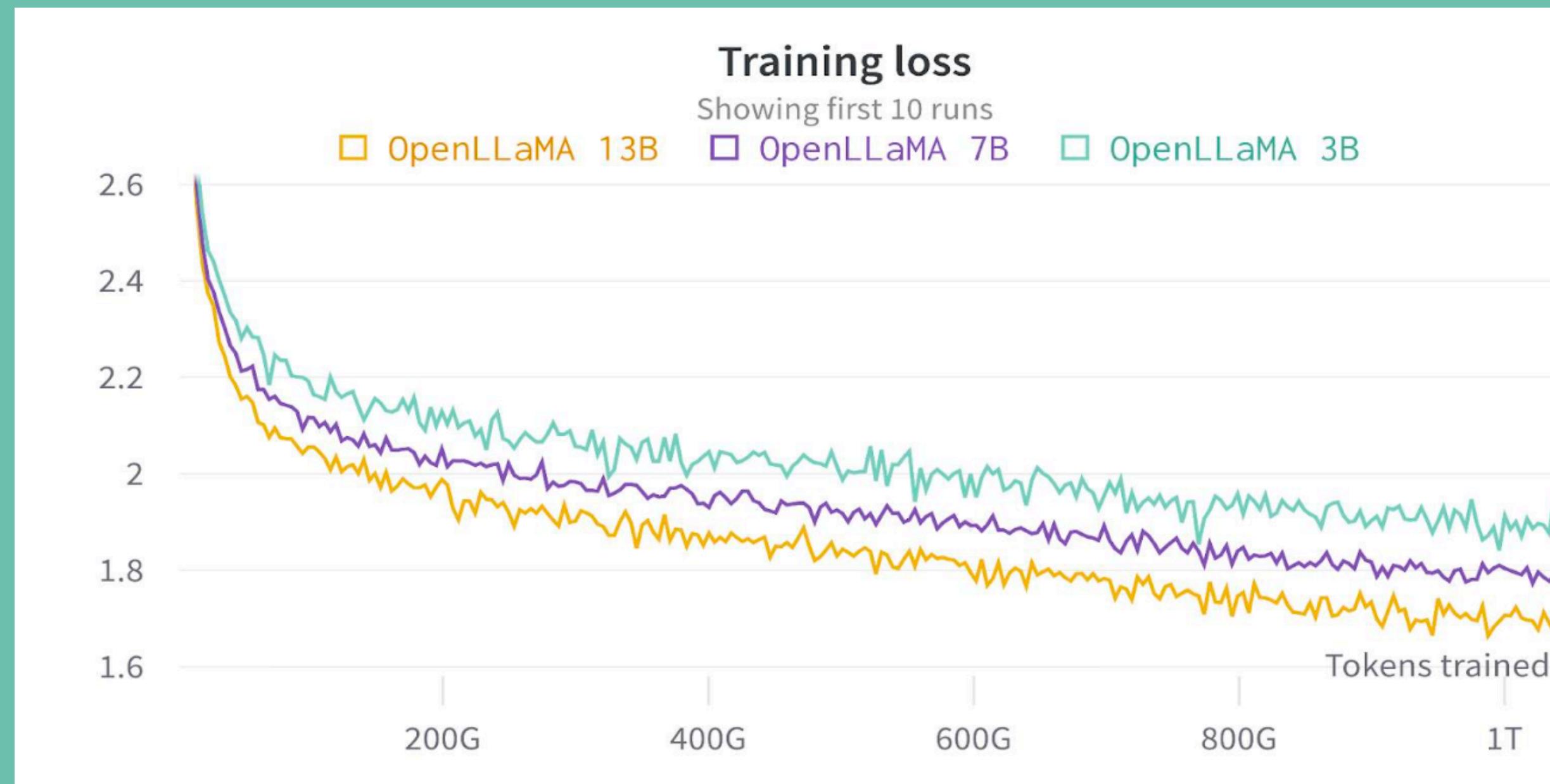
- ▶ Creado a partir de modelos de lenguaje
- ▶ Son capaces de predecir la probabilidad de la palabra siguiente de un texto
- ▶ Cada vez se entrena con más datos y son capaces de realizar más tareas
- ▶ Hay modelos de lenguaje entrenados con billones de tokens





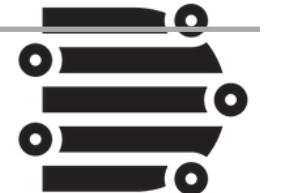
TEXTO: RASGOS GENERALES

▶ Datasets gigantes

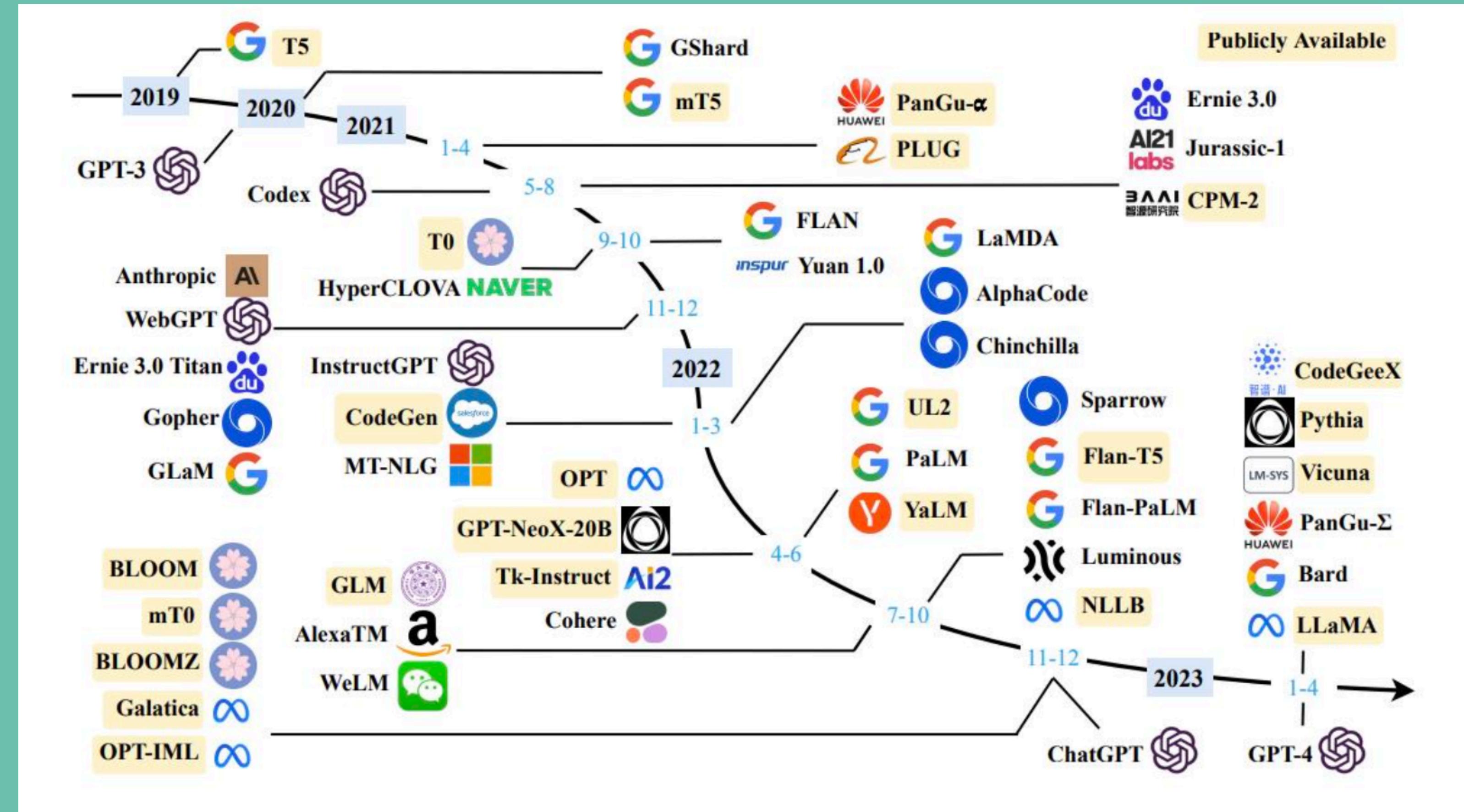


The websites in Google's C4 dataset

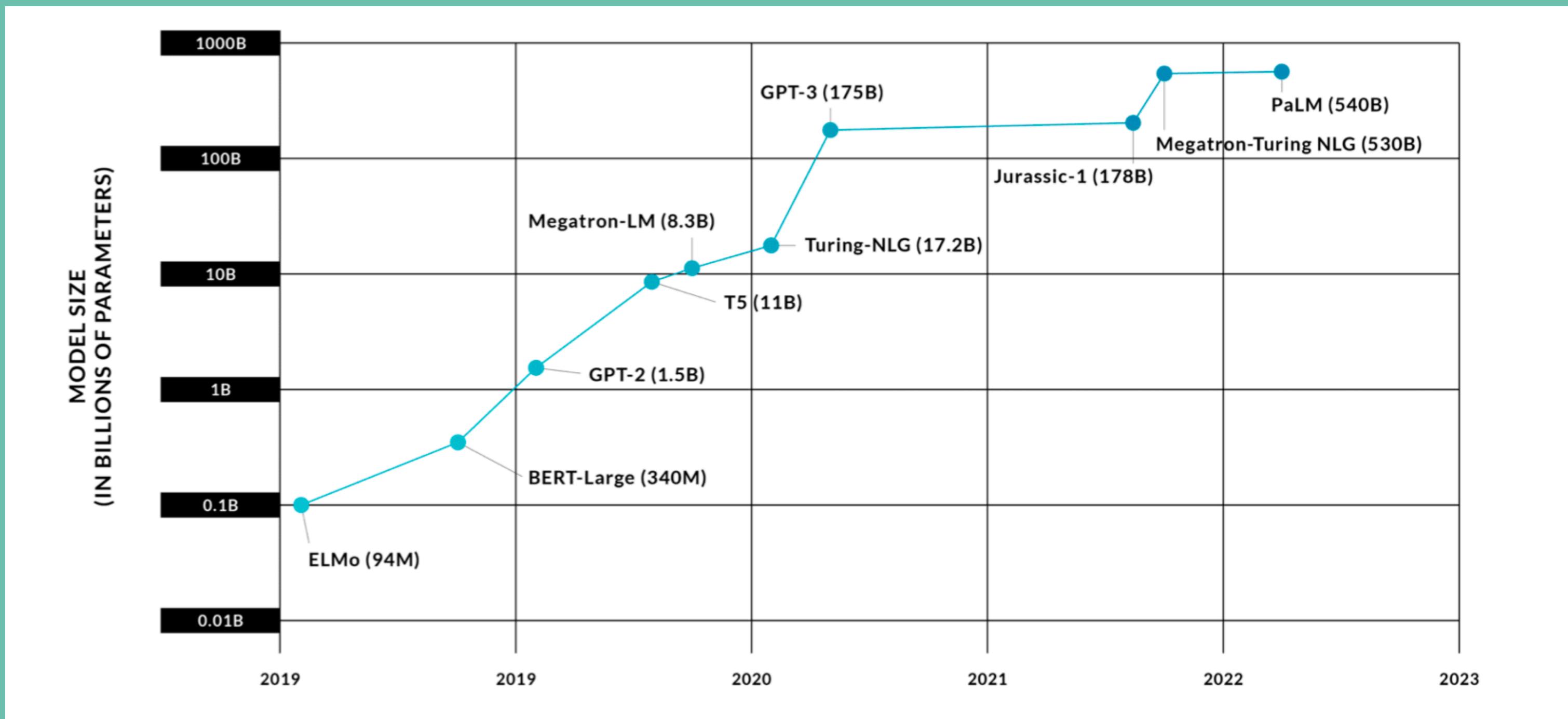
| RANK | DOMAIN | CATEGORY | PERCENT OF ALL TOKENS |
|------|--------------------|-----------------------|-----------------------|
| 1 | patents.google.com | Law & Government | 0.46% |
| 2 | wikipedia.org | News & Media | 0.19% |
| 3 | scribd.com | News & Media | 0.07% |
| 4 | nytimes.com | News & Media | 0.06% |
| 5 | journals.plos.org | Science & Health | 0.06% |
| 6 | latimes.com | News & Media | 0.05% |
| 7 | theguardian.com | News & Media | 0.05% |
| 8 | forbes.com | News & Media | 0.05% |
| 9 | huffpost.com | News & Media | 0.04% |
| 10 | patents.com | Law & Government | 0.04% |
| 11 | washingtonpost.com | News & Media | 0.03% |
| 12 | coursera.org | Jobs & Education | 0.03% |
| 13 | fool.com | Business & Industrial | 0.03% |
| 14 | frontiersin.org | Science & Health | 0.03% |
| 15 | instructables.com | Technology | 0.03% |

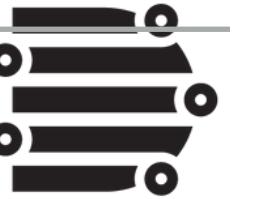


TEXTO: EVOLUCIÓN



TEXTO: PARÁMETROS





TEXTO: LLMS SEGÚN LICENCIA

MODELOS CERRADOS

- ▶ GPT3 y superior
- ▶ Bard
- ▶ Lamda

MODELOS ABIERTOS

- ▶ LLaMa
- ▶ GLM
- ▶ NeMO

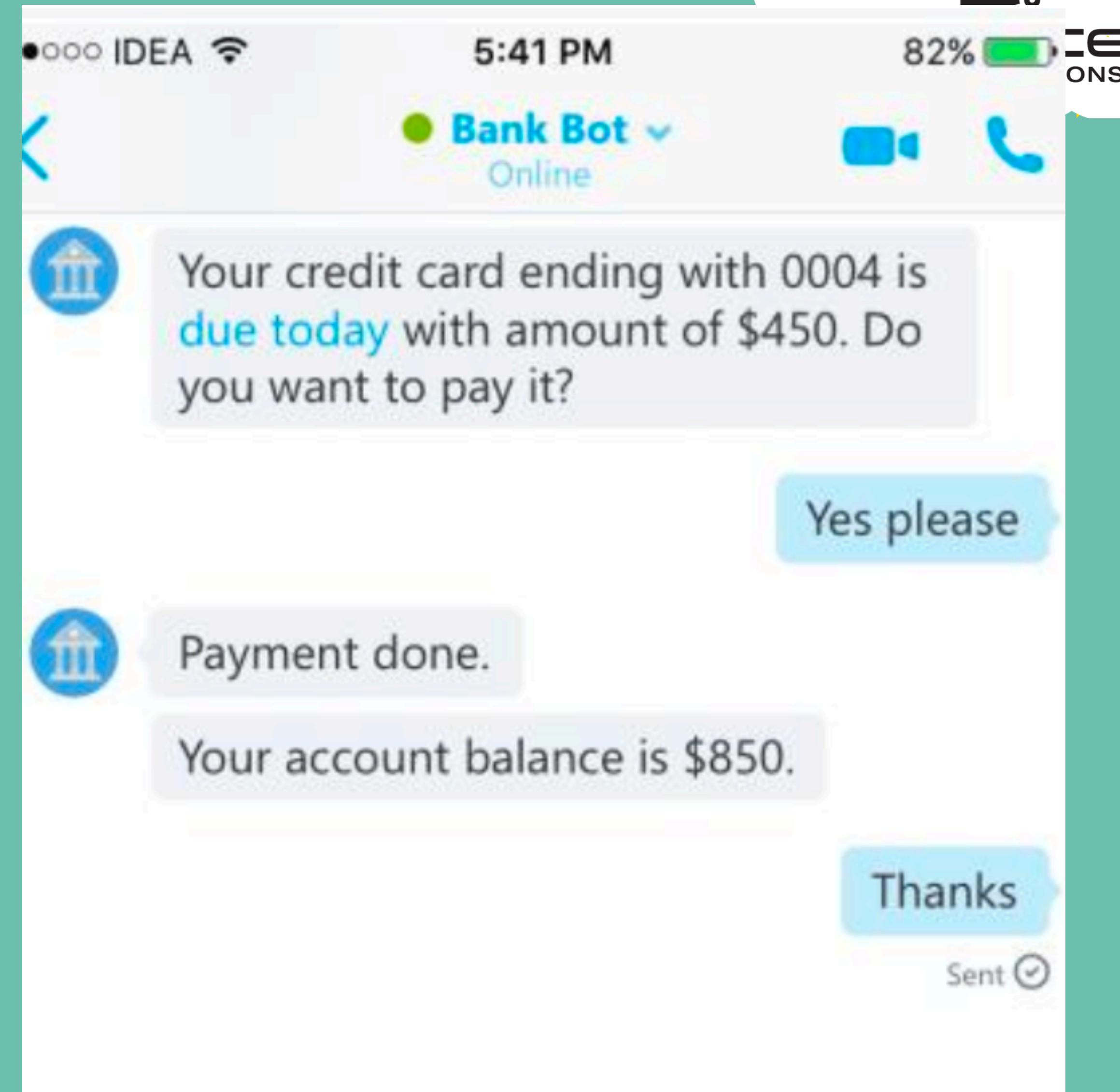
MODELOS ABIERTOS CON LICENCIAS PERMISIVAS

- ▶ Falcon
- ▶ GPT2
- ▶ Bert

<https://sapling.ai/llm/index>

TEXTO: APLICACIONES

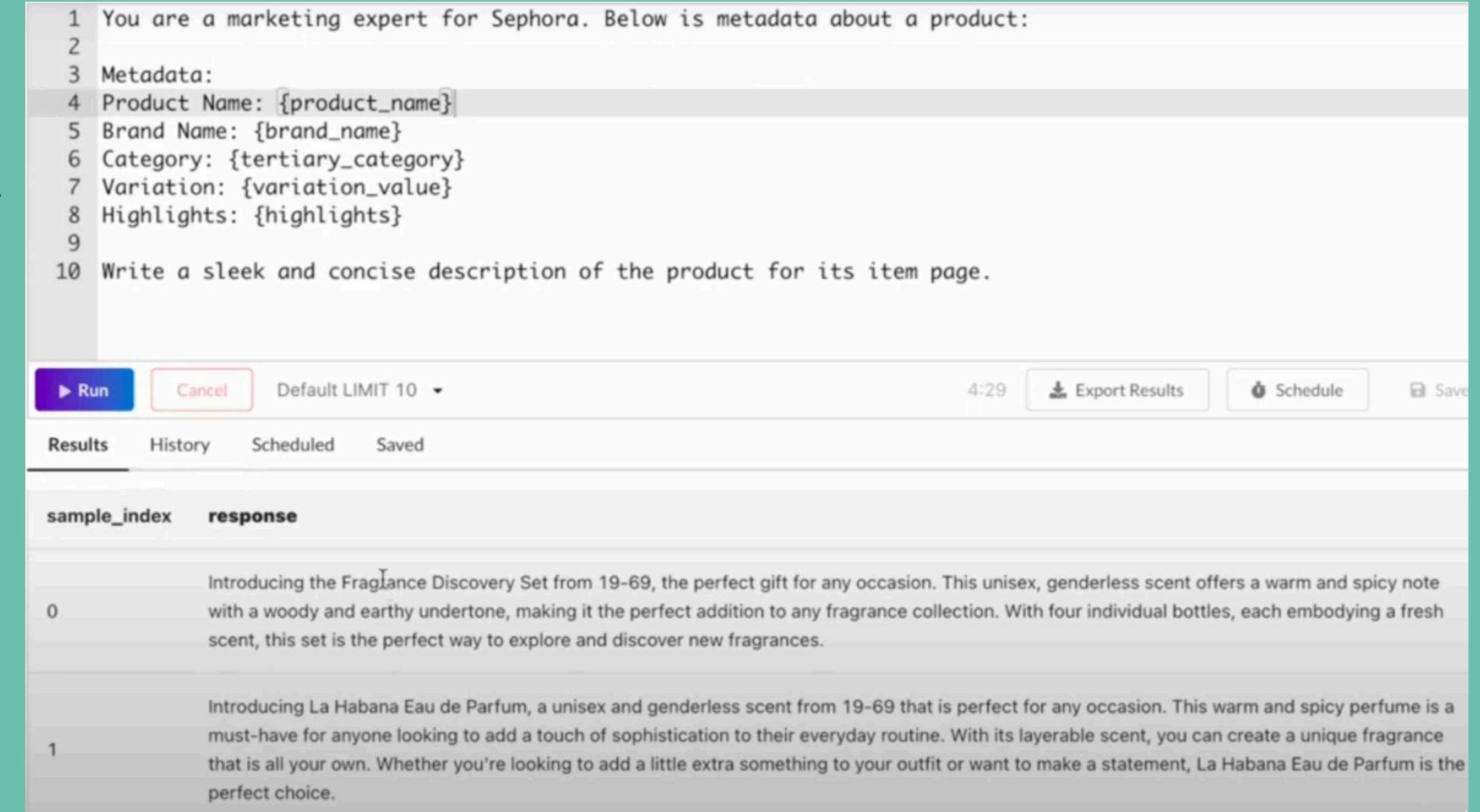
- ▶ Chatbots
 - ▶ Obtención de información inicial antes de hablar con un agente humano
 - ▶ Encauzamiento de una petición al departamento apropiado
 - ▶ Peticiones simples, como hacer una reserva
 - ▶ Soporte 24/7
 - ▶ Asistentes de voz
 - ▶ Recomendación de productos
 - ▶ Recordatorios
 - ▶ Seguimiento de gastos
 - ▶ Aprendizaje (generación de preguntas o aprendizaje de idiomas, por ejemplo)



TEXTO: APLICACIONES

▶ Creación de contenido

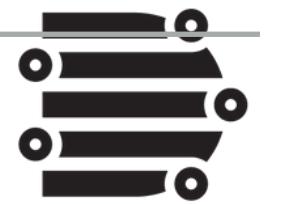
- ▶ Artículos (blog, revistas, periódicos, social media, etc.)
- ▶ Descripciones de productos (e-commerce, copywriting, etc)
- ▶ Expansión de un concepto
- ▶ Aumento de la calidad de un texto
- ▶ Adaptación de un texto a un tipo de audiencia
- ▶ Artículos para aprendizaje
- ▶ Diagnóstico de incidentes



1 You are a marketing expert for Sephora. Below is metadata about a product:
2
3 Metadata:
4 Product Name: {product_name}
5 Brand Name: {brand_name}
6 Category: {tertiary_category}
7 Variation: {variation_value}
8 Highlights: {highlights}
9
10 Write a sleek and concise description of the product for its item page.

Run Cancel Default LIMIT 10 4:29 Export Results Schedule Save

| sample_index | response |
|--------------|---|
| 0 | Introducing the Fragrance Discovery Set from 19-69, the perfect gift for any occasion. This unisex, genderless scent offers a warm and spicy note with a woody and earthy undertone, making it the perfect addition to any fragrance collection. With four individual bottles, each embodying a fresh scent, this set is the perfect way to explore and discover new fragrances. |
| 1 | Introducing La Habana Eau de Parfum, a unisex and genderless scent from 19-69 that is perfect for any occasion. This warm and spicy perfume is a must-have for anyone looking to add a touch of sophistication to their everyday routine. With its layerable scent, you can create a unique fragrance that is all your own. Whether you're looking to add a little extra something to your outfit or want to make a statement, La Habana Eau de Parfum is the perfect choice. |



TEXTO: APLICACIONES

▶ Mejora de Contenido



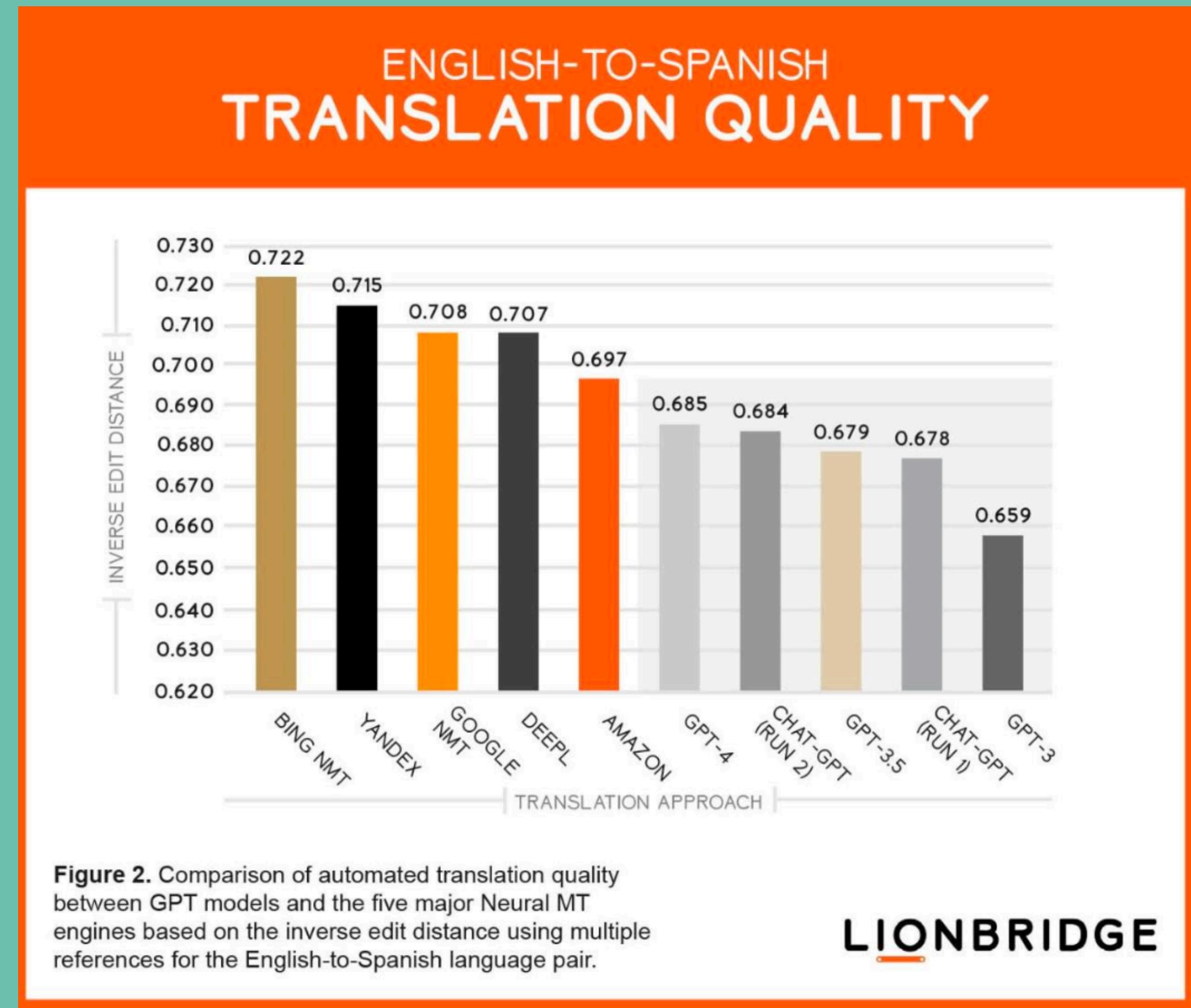
simplify and proofread this text : Chat GPT is an incredible tool that can revolutionize the way we create content. By taking advantage of advances in natural language processing, Chat GPT can improve the text and generate new text with minimal human input. This makes generating content much more accessible and significantly speeds up the writing process. Powered by the powerful GPT 3, Chat GPT utilizes sophisticated machine learning algorithms to allow users to generate high-quality output with contextually correct syntax and grammar. What's even more impressive is that the program understands the semantic structure and the writer's style, generating precise and accurate content every time. Chat GPT promises to be a revolutionary asset for anyone looking to quickly create quality output, making it a must-have for any business or individual looking for an edge when creating digital content.



Chat GPT is a powerful tool that can change the way we create content. It uses natural language processing to improve text and generate new text with little human input, making content creation faster and more accessible. GPT-3 powers Chat GPT, which uses advanced machine learning algorithms to produce high-quality output with correct syntax and grammar. It even understands the writer's style and produces accurate content every time. Chat GPT is a must-have for anyone looking to create quality digital content quickly and efficiently.

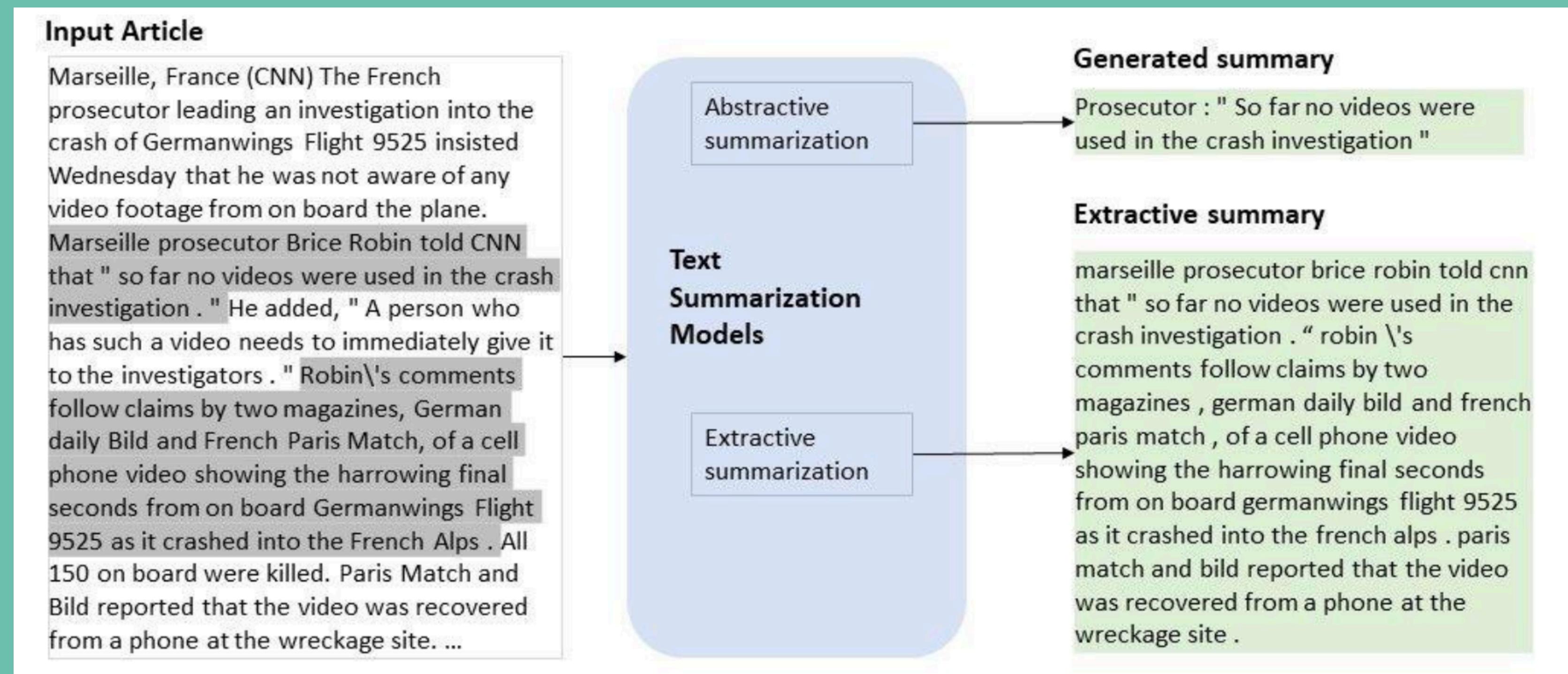
TEXTO: APLICACIONES

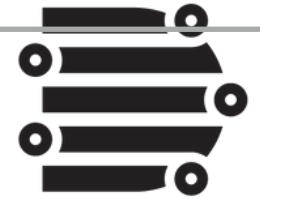
▶ Traducción



TEXTO: APLICACIONES

▶ Resumen

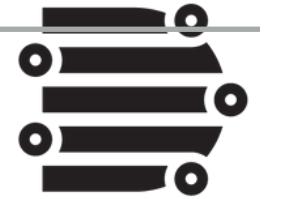




TEXTO: APLICACIONES

▶ Análisis de sentimiento

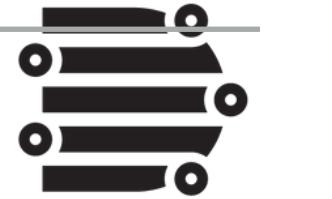
I **①** really enjoyed using the **①** Canon Ixus in Madrid on March 4. The **②** Panasonic Lumix **②** is a bit disappointing, but the **③** Canon camera **③** not bad at all. All I want when taking photos is point it and then just press the button. For only 200 dollars, a **④** really fair **④** price, this **⑤** camera **⑤** perfect for me. Besides, I have had a **⑥** good **⑥** customer service experience.



TEXTO: APLICACIONES

▶ Generación código

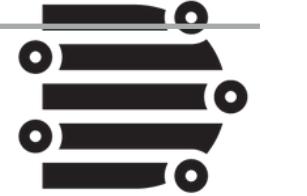
| Problem | Generated Code | Test Cases |
|---|--|--|
| <p>H-Index</p> <p>Given a list of citations counts, where each citation is a nonnegative integer, write a function <code>h_index</code> that outputs the h-index. The h-index is the largest number h such that h papers have each least h citations.</p> <p>Example:</p> <p>Input: [3,0,6,1,4] Output: 3</p> | <pre>def h_index(counts): n = len(counts) if n > 0: counts.sort() counts.reverse() h = 0 while (h < n and counts[h]-1>=h): h += 1 return h else: return 0</pre> | <p>Input: [1,4,1,4,2,1,3,5,6]</p> <p>Generated Code Output: 4 ✓</p> <p>Input: [1000,500,500,250,100, 100,100,100,100,75,50, 30,20,15,15,10,5,2,1]</p> <p>Generated Code Output: 15 ✓</p> |



TEXTO: APLICACIONES

- ▶ Clasificación texto
- ▶ Asignación de emails al departamento correspondiente
- ▶ Clasificación de incidencias
- ▶ Detección de spam

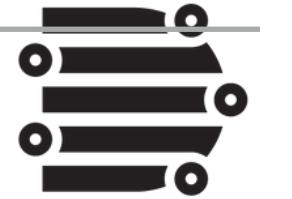
| | Inquiry | Shipping | Returns | Tracking |
|----|---|----------|---------|----------|
| 1 | Do you offer same day shipping? | ● | ○ | ○ |
| 2 | Can you ship to Italy? | ● | ○ | ○ |
| 3 | How long does shipping take? | ● | ○ | ○ |
| 4 | Can I buy online and pick up in store? | ● | ○ | ○ |
| 5 | What are your shipping options? | ● | ○ | ○ |
| 6 | My order arrived damaged, can I get a refund? | ○ | ● | ○ |
| 7 | You sent me the wrong item | ○ | ● | ○ |
| 8 | I want to exchange my item for another colour | ○ | ● | ○ |
| 9 | I ordered something and it wasn't what I expected. Can I return it? | ○ | ● | ○ |
| 10 | What's your return policy? | ○ | ● | ○ |
| 11 | Where's my package? | ○ | ○ | ● |
| 12 | When will my order arrive? | ○ | ○ | ● |
| 13 | What's my shipping number? | ○ | ○ | ● |
| 14 | Which carrier is my package with? | ○ | ○ | ● |
| 15 | Is my package delayed? | ○ | ○ | ● |



TEXTO: APLICACIONES

- ▶ No todos los modelos son iguales, incluso dentro de los modelos generalistas. Algunos modelos son mejores en algunas tareas

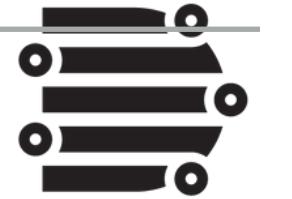




TEXTO: APLICACIONES

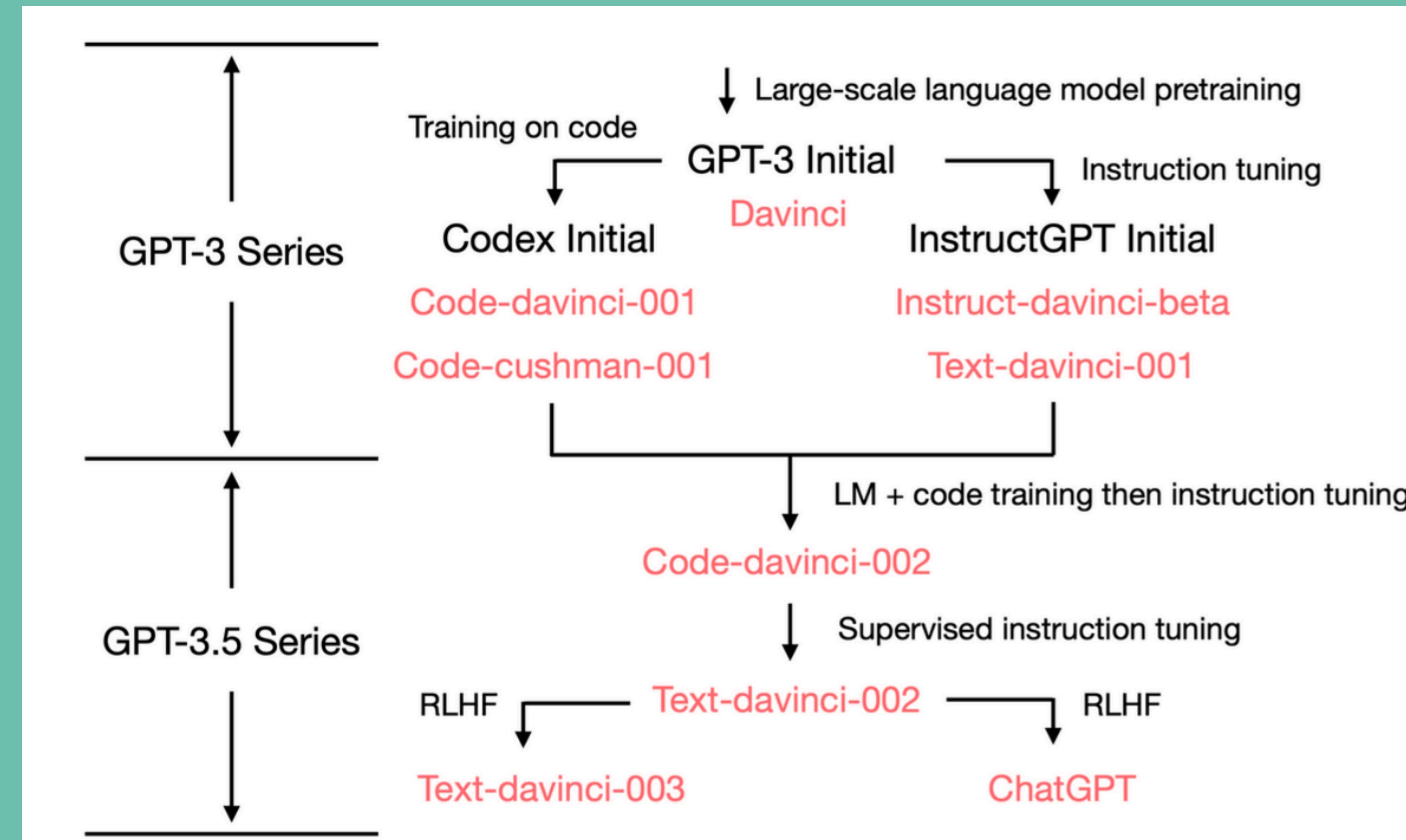
- ▶ Se puede hacer una prueba comparativa en <https://chat.lmsys.org>

| | Reasoning | Knowledge | Conversation | Creativity | Personality | Storytelling | Empathy |
|------------|-----------|-----------|--------------|------------|-------------|--------------|---------|
| LaMDA | 0.84 | 0.69 | 1.0 | 0.53 | 0.85 | 0.58 | 0.94 |
| ChatGPT | 0.74 | 0.82 | 0.92 | 0.77 | 0.72 | 0.74 | 0.7 |
| GPT-3 | 0.87 | 0.86 | 0.72 | 0.75 | 0.66 | 0.72 | 0.49 |
| T5 | 0.7 | 0.6 | 0.19 | 0.51 | 0.1 | 0.36 | 0.04 |
| PaLM | 0.76 | 0.56 | 0.21 | 0.24 | 0.21 | 0.18 | 0.17 |
| BLOOM | 0.48 | 0.35 | 0.29 | 0.36 | 0.15 | 0.18 | 0.24 |
| Turing-NLG | 0.56 | 0.42 | 0.29 | 0.07 | 0.16 | 0.07 | 0.0 |



TEXTO: OPCIONES DE USO

- ▶ Hay modelos especializados en algunas tareas concretas

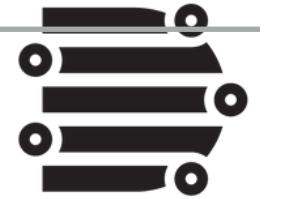


TEXTO: OPCIONES DE USO

- ▶ Pero para resolver casos de uso específicos, es probable que tengamos que adaptar un modelo.

| | GPU Type | GPU Power consumption | GPU-hours | Total power consumption | Carbon emitted (tCO ₂ eq) |
|------------|-----------|-----------------------|-----------|-------------------------|--------------------------------------|
| OPT-175B | A100-80GB | 400W | 809,472 | 356 MWh | 137 |
| BLOOM-175B | A100-80GB | 400W | 1,082,880 | 475 MWh | 183 |
| LLaMA-7B | A100-80GB | 400W | 82,432 | 36 MWh | 14 |
| LLaMA-13B | A100-80GB | 400W | 135,168 | 59 MWh | 23 |
| LLaMA-33B | A100-80GB | 400W | 530,432 | 233 MWh | 90 |
| LLaMA-65B | A100-80GB | 400W | 1,022,362 | 449 MWh | 173 |

Table 15: **Carbon footprint of training different models in the same data center.** We follow the formula from [Wu et al. \(2022\)](#) to compute carbon emission of train OPT, BLOOM and our models in the same data center. For the power consumption of a A100-80GB, we take the thermal design power (TDP) for NVLink systems, that is 400W. We take a PUE of 1.1 and a carbon intensity factor set at the national US average of 0.385 kg CO₂e per KWh.



TEXTO: ESPECIALIZACIÓN

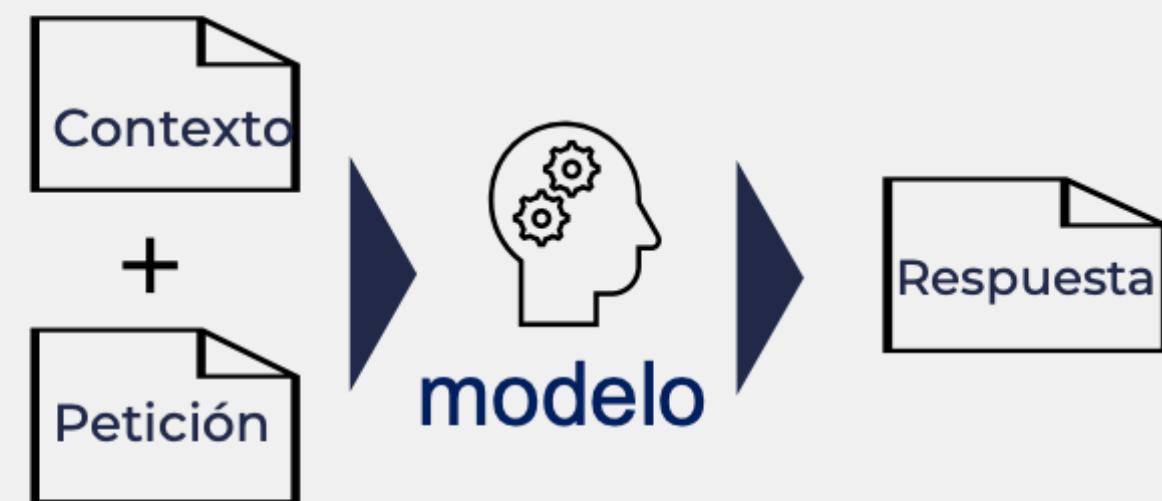
- ▶ Pero sí que existen opciones viables
- ▶ Las dos principales opciones son prompt engineering y fine tuning

Prompt engineering

Se crea un contexto que le permita al modelo dar respuestas más apropiadas.

Ventajas:

- Más sencillo
- Inmediato
- Menor coste

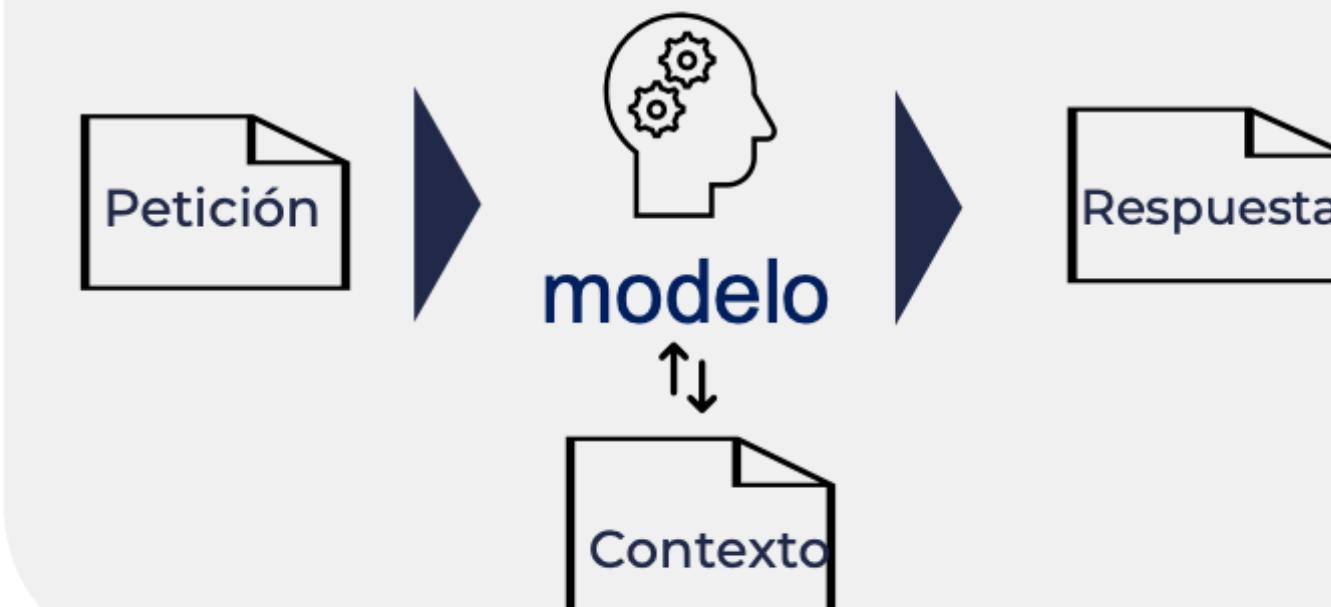


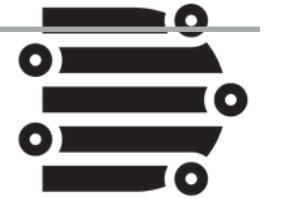
Fine Tuning

Reentrenar el modelo con algunos ejemplos. Se hace mediante pares pregunta-respuesta

Ventajas:

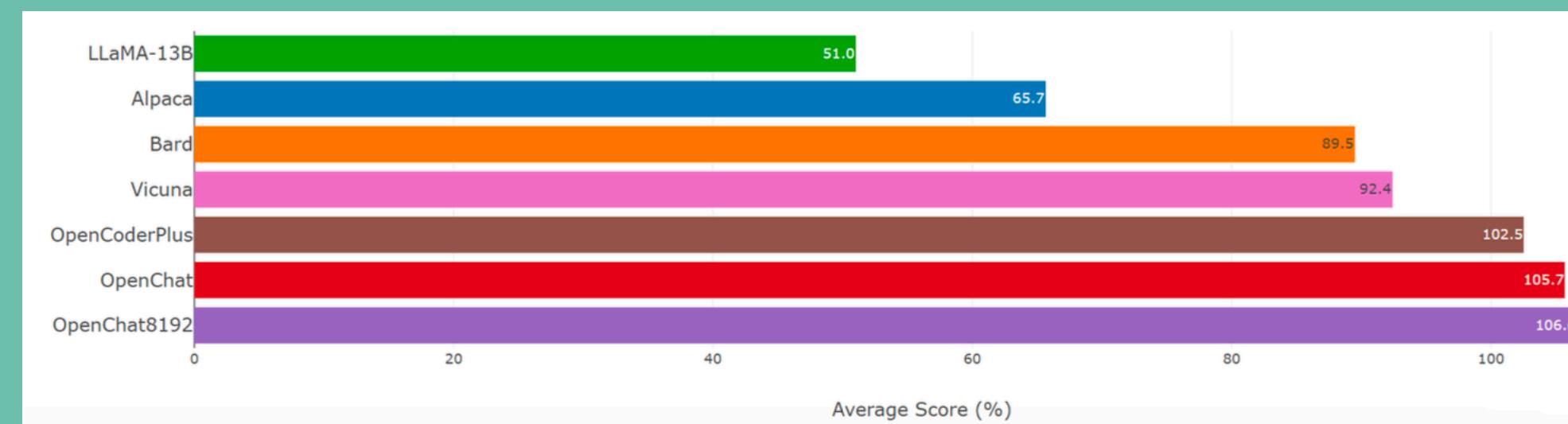
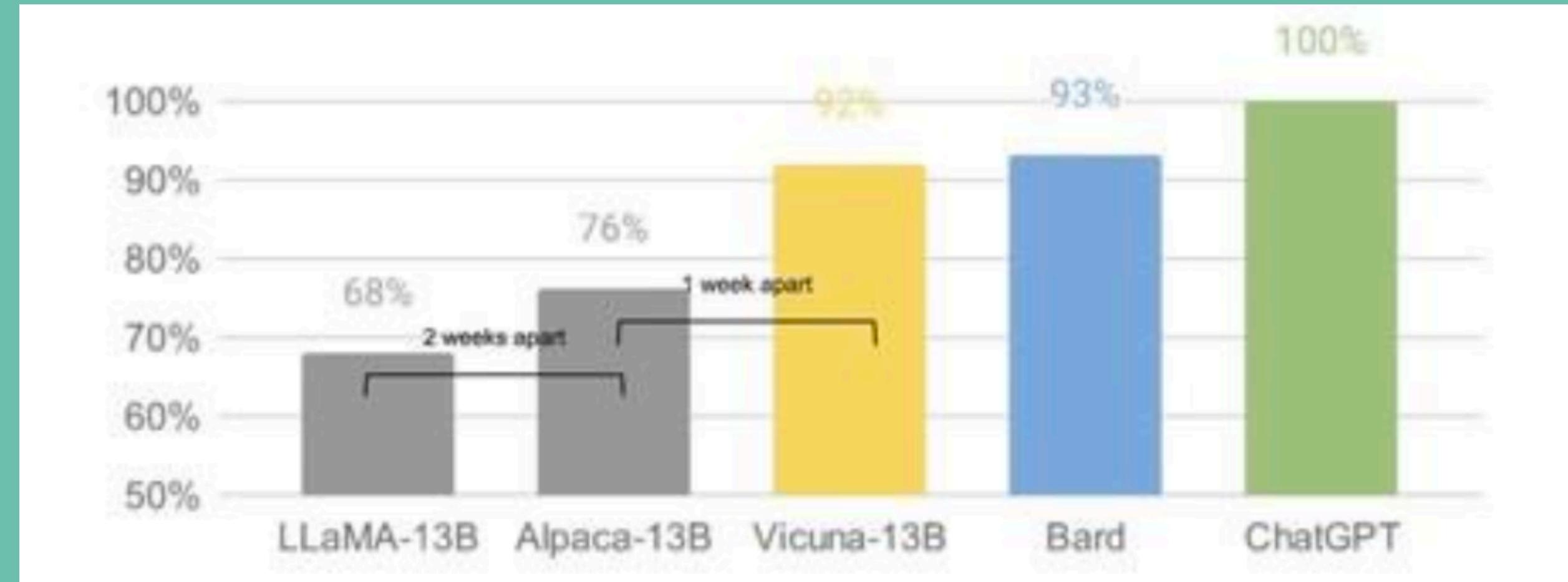
- No hay limitaciones de tamaño de datos de entrada

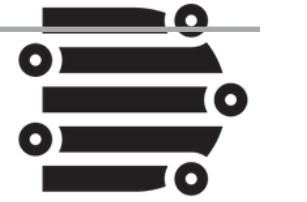




TEXTO: MEJORAS

- ▶ Gracias al fine tuning se pueden obtener muy buenos resultados con modelos más pequeños





TEXTO: PROMPT ENGINEERING/LEARNING

