

Tecnologías emergentes en NLP y Procesamiento del Habla

OBS Business School

Presentación

Carlos Rodríguez Abellán

GenAI & NLP Practice Lead en el CoE de Data Intelligence global de FUJITSU. Desarrollo de soluciones basadas en Natural Language Processing e Inteligencia Artificial para clientes de distintos sectores y geografías.

He trabajado en distintas empresas diseñando e implementando procesos end-to-end para la automatización inteligente de procesos y acompañando a los clientes en su proceso de transformación digital.



He impartido formación sobre IA en otras universidades.



<https://www.linkedin.com/in/carlosrodriguezabellan/>

Planning

P L A N N I N G	Tema	Objetivos generales	Actividad	Recursos	Total horas dedicación	Inicio	Final
	Tema 1	Presentar los conceptos del tema 1	webcast	Material semana 1	1	15/05/2024 19:00 (CEST)	15/05/2024 20:00 (CEST)
	Tema 1	Afianzar conceptos tema 1	lectura y práctica	Material semana 1	5	15/05/2024 20:00 (CEST)	21/05/2024 23:59 (CEST)
	Tema 1	Actividad evaluable	debate	Material semana 1	5	15/05/2024 20:00 (CEST)	21/05/2024 23:59 (CEST)
	Tema 2	Presentar los conceptos del tema 2	webcast	Material semana 2	1	22/05/2024 19:00 (CEST)	22/05/2024 20:00 (CEST)
	Tema 2	Afianzar conceptos tema 2	lectura y práctica	Material semana 2	15	22/05/2024 20:00 (CEST)	28/05/2024 23:59 (CEST)
	Tema 3	Presentar los conceptos del tema 3	webcast	Material semana 3	1	29/05/2024 19:00 (CEST)	29/05/2024 20:00 (CEST)
	Tema 3	Afianzar conceptos tema 3	lectura y práctica	Material semana 3	15	29/05/2024 20:00 (CEST)	04/06/2024 23:59 (CEST)
	Tema 3	Actividad evaluable	test	Material semanas 2 y 3	10	29/05/2024 20:00 (CEST)	04/06/2024 23:59 (CEST)
	Tema 4	Presentar los conceptos del tema 4	webcast	Material semana 4	1	05/06/2024 19:00 (CEST)	05/06/2024 20:00 (CEST)
	Tema 4	Afianzar conceptos tema 4	lectura y práctica	Material semana 4	15	05/06/2024 20:00 (CEST)	11/06/2024 23:59 (CEST)
	Tema 4	Actividad evaluable	test	Material semana 4	10	05/06/2024 20:00 (CEST)	11/06/2024 23:59 (CEST)
	Temas 1, 2, 3 y 4	Actividad evaluable	caso práctico final	Material semanas 1, 2, 3 y 4	30	29/05/2024 20:00 (CEST)	11/06/2024 23:59 (CEST)

Nota: el cómputo de horas es orientativo por actividad. En algunos casos se requerirá menos tiempo del indicado y en otros más. No se ha incluido el tiempo dedicado a la resolución de dudas o revisión de ejercicios resueltos.

Sesiones

Actividades

P
L
A
N
N
I
N
G

Tema	Objetivos generales	Actividad	Recursos	Total horas dedicación	Inicio	Final
Tema 1	Presentar los conceptos del tema 1	webcast	Material semana 1	1	15/05/2024 19:00 (CEST)	15/05/2024 20:00 (CEST)
Tema 1	Afianzar conceptos tema 1	lectura y práctica	Material semana 1	5	15/05/2024 20:00 (CEST)	21/05/2024 23:59 (CEST)
Tema 1	Actividad evaluable	debate	Material semana 1	5	15/05/2024 20:00 (CEST)	21/05/2024 23:59 (CEST)
Tema 2	Presentar los conceptos del tema 2	webcast	Material semana 2	1	22/05/2024 19:00 (CEST)	22/05/2024 20:00 (CEST)
Tema 2	Afianzar conceptos tema 2	lectura y práctica	Material semana 2	15	22/05/2024 20:00 (CEST)	28/05/2024 23:59 (CEST)
Tema 3	Presentar los conceptos del tema 3	webcast	Material semana 3	1	29/05/2024 19:00 (CEST)	29/05/2024 20:00 (CEST)
Tema 3	Afianzar conceptos tema 3	lectura y práctica	Material semana 3	15	29/05/2024 20:00 (CEST)	04/06/2024 23:59 (CEST)
Tema 3	Actividad evaluable	test	Material semanas 2 y 3	10	29/05/2024 20:00 (CEST)	04/06/2024 23:59 (CEST)
Tema 4	Presentar los conceptos del tema 4	webcast	Material semana 4	1	05/06/2024 19:00 (CEST)	05/06/2024 20:00 (CEST)
Tema 4	Afianzar conceptos tema 4	lectura y práctica	Material semana 4	15	05/06/2024 20:00 (CEST)	11/06/2024 23:59 (CEST)
Tema 4	Actividad evaluable	test	Material semana 4	10	05/06/2024 20:00 (CEST)	11/06/2024 23:59 (CEST)
Temas 1, 2, 3 y 4	Actividad evaluable	caso práctico final	Material semanas 1, 2, 3 y 4	30	29/05/2024 20:00 (CEST)	11/06/2024 23:59 (CEST)

Nota: el cómputo de horas es orientativo por actividad. En algunos casos se requerirá menos tiempo del indicado y en otros más. No se ha incluido el tiempo dedicado a la resolución de dudas o revisión de ejercicios resueltos.

Introducción

Hype Cycle for Artificial Intelligence, 2023

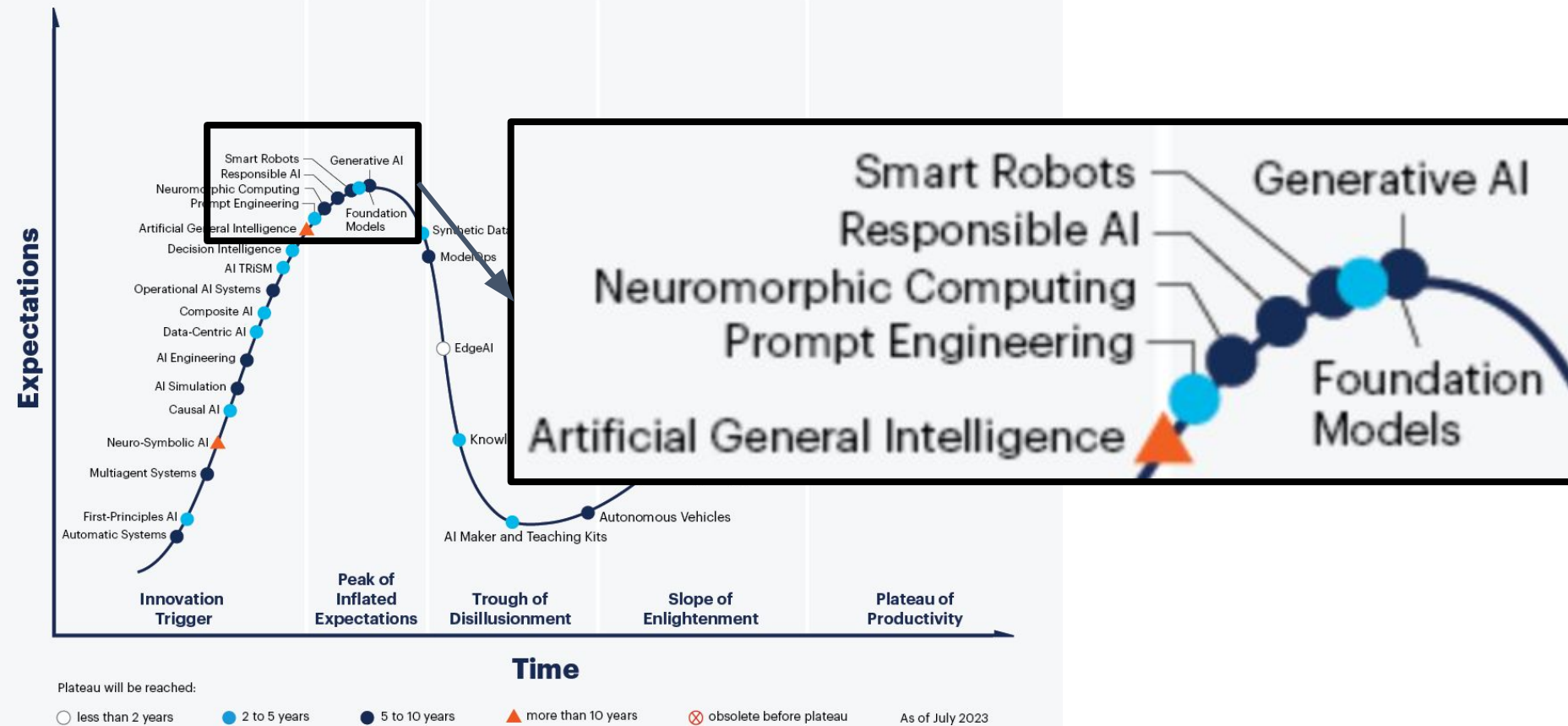


[gartner.com](https://www.gartner.com)

Source: Gartner
© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. 2079794

Gartner

Hype Cycle for Artificial Intelligence, 2023



gartner.com

Source: Gartner
© 2023 Gartner, Inc. and/or its affiliates. All rights reserved. 2079794

Gartner

Prompt

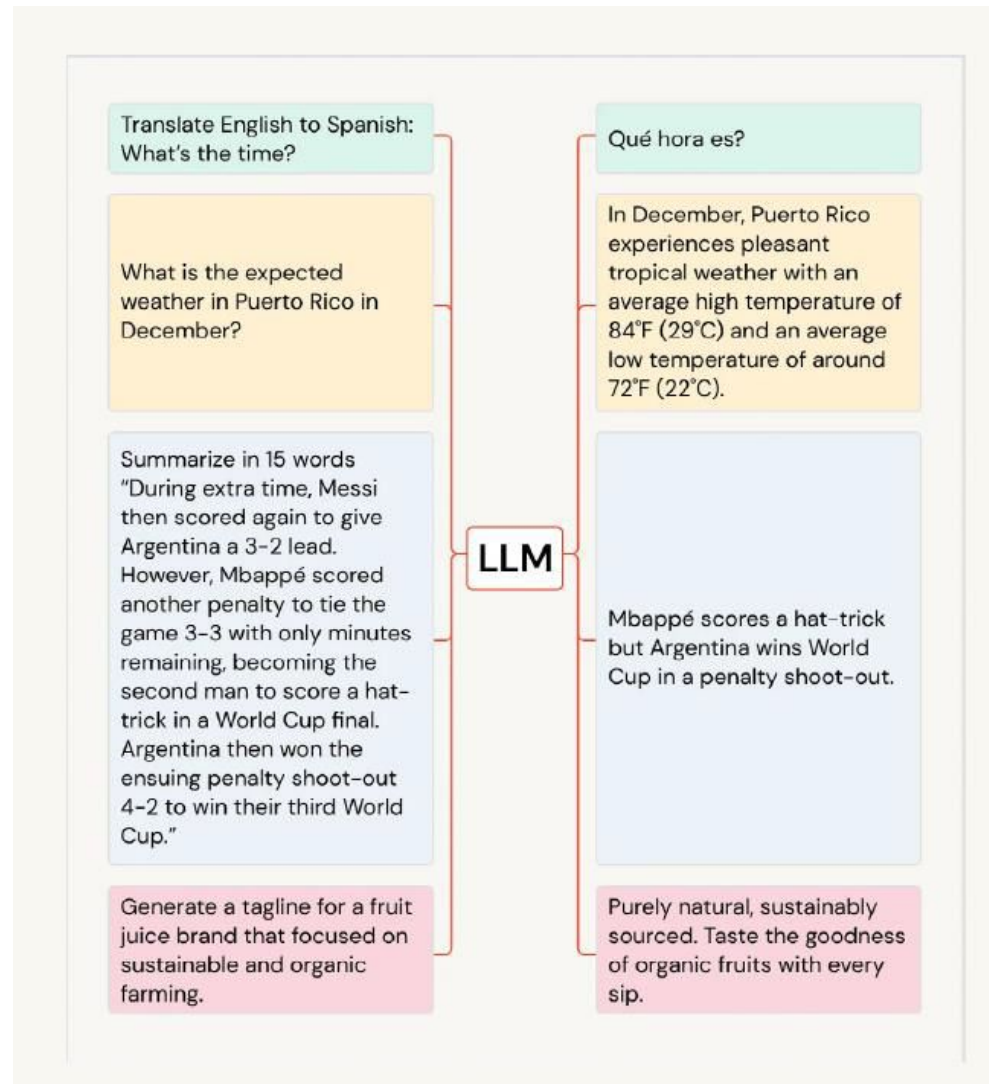
A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.

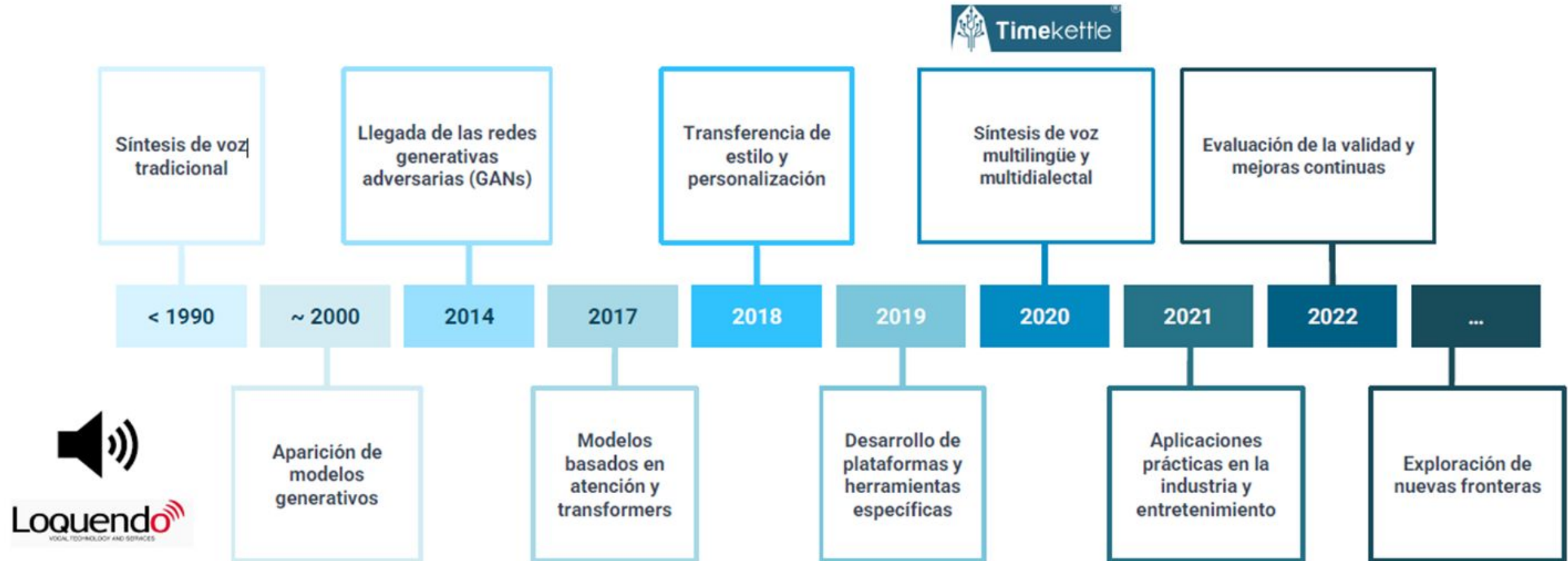


**Entienden el
lenguaje natural**

**Son
multimodales**

**Generan nuevo
contenido**





Objetivos del módulo

MusicLM: Generating Music From Text

| paper | dataset |

Andrea Agostinelli, Timo I. Denk, Zolán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, Christian Frank
Google Research

Abstract We introduce MusicLM, a model generating high-fidelity music from text descriptions such as “a calming violin melody backed by a distorted guitar riff”. MusicLM casts the process of conditional music generation as a hierarchical sequence-to-sequence modeling task, and it generates music at 24 kHz that remains consistent over several minutes. Our experiments show that MusicLM outperforms previous systems both in audio quality and adherence to the text description. Moreover, we demonstrate that MusicLM can be conditioned on both text and a melody in that it can transform whistled and hummed melodies according to the style described in a text caption. To support future research, we publicly release MusicCaps, a dataset composed of 5.5k music-text pairs, with rich text descriptions provided by human experts.

Audio Generation From Rich Captions

Caption

Generated audio

The main soundtrack of an arcade game. It is fast-paced and upbeat, with a catchy electric guitar riff. The music is repetitive and easy to remember, but with unexpected sounds, like cymbal crashes or drum rolls.

▶ 0:00 / 0:30 — 🔊 ⋮

A fusion of reggaeton and electronic dance music, with a spacey, otherworldly sound. Induces the experience of being lost in space, and the music would be designed to evoke a sense of wonder and awe, while being danceable.

▶ 0:00 / 0:30 — 🔊 ⋮

A rising synth is playing an arpeggio with a lot of reverb. It is backed by pads, sub bass line and soft drums. This song is full of synth sounds creating a soothing and adventurous atmosphere. It may be playing at a festival during two songs for a buildup.

▶ 0:00 / 0:30 — 🔊 ⋮

Slow tempo, bass-and-drums-led reggae song. Sustained electric guitar. High-pitched bongos with ringing tones. Vocals are relaxed with a laid-back feel, very expressive.

▶ 0:00 / 0:30 — 🔊 ⋮

1 2 3

Home

Create

Library

Explore BETA

50 credits

Subscribe

What's New? 8

Help

About

Suno Showcase

Show more +

Golden Sunshine 윤아
Zack Brown
bossa nova, uk drill, electric piano
▶ 716580 👍 8136 ⋮

Stone
Oliver McCann
indie-pop soulful dreamy psychedelic
▶ 1406950 👍 12289 ⋮

dialectic (accept it) - Yolkhead
yolkhead
post-indietronica post-instrumental cello revival
▶ 325357 👍 3247 ⋮

Trending

Now ▾

Show more +

#1

Glass Heart
Nanashi_Zero...
Catchy Instrumental intro. Super fast aggressive rap. speedcore. gritty female vocal
▶ 181530 👍 1291 ⋮

#2

Second Law of Thermodynamics
Stei Camel 🦙 YouTube
Power Metal
▶ 45425 👍 610 ⋮

#3

魔法少女 ♡ ジュニア
Suno is Magic!...
Brutus
catchy, Sweet voice, J-pop, Cute, cure, Magical Girl
▶ 43129 👍 494 ⋮

"This isn't music"
~ Chibitaii ~
lo-fi jazz funky electro swing, soft rock, bittersweet and sad
▶ 21515 👍 279 ⋮

We are all alone
Sirdegen.exe
Melodic Techno. , Sad , Popularity ,
▶ 3406 👍 72 ⋮

Black Cat Boogie
Feline Music [Youtube: @FelineMusicYey]
Halloween Music, Wolfs howling, Halloween ambiente, electro...
▶ 12223 👍 126 ⋮

Top Categories

Show more +

Jazz

Rap

Pop

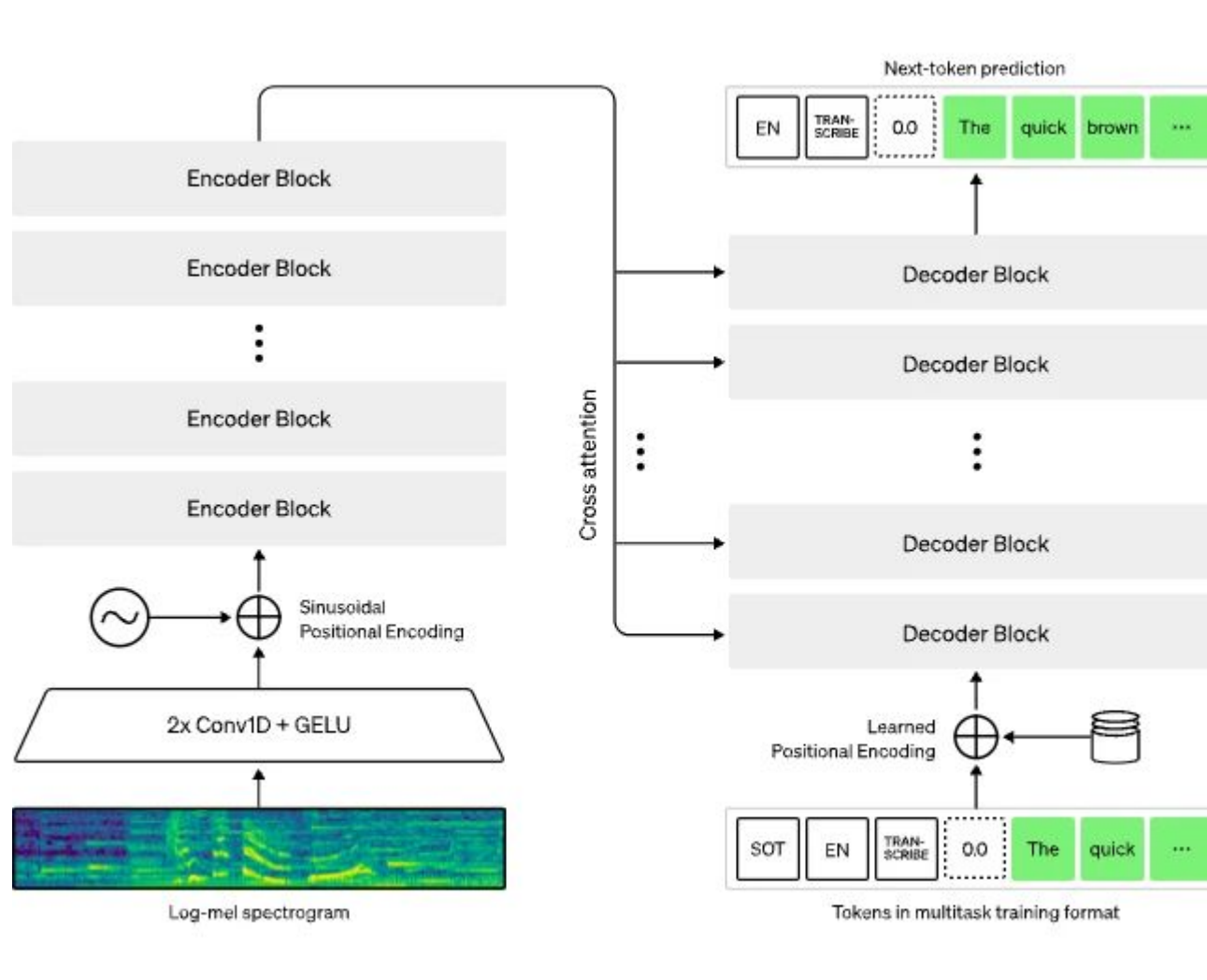
Rock

Metal

0:00

0:00

<https://suno.com/>



Non-English transcription

언덕 위에 올라 내려다보면 너무나 넓고 붉은 ...

No speech

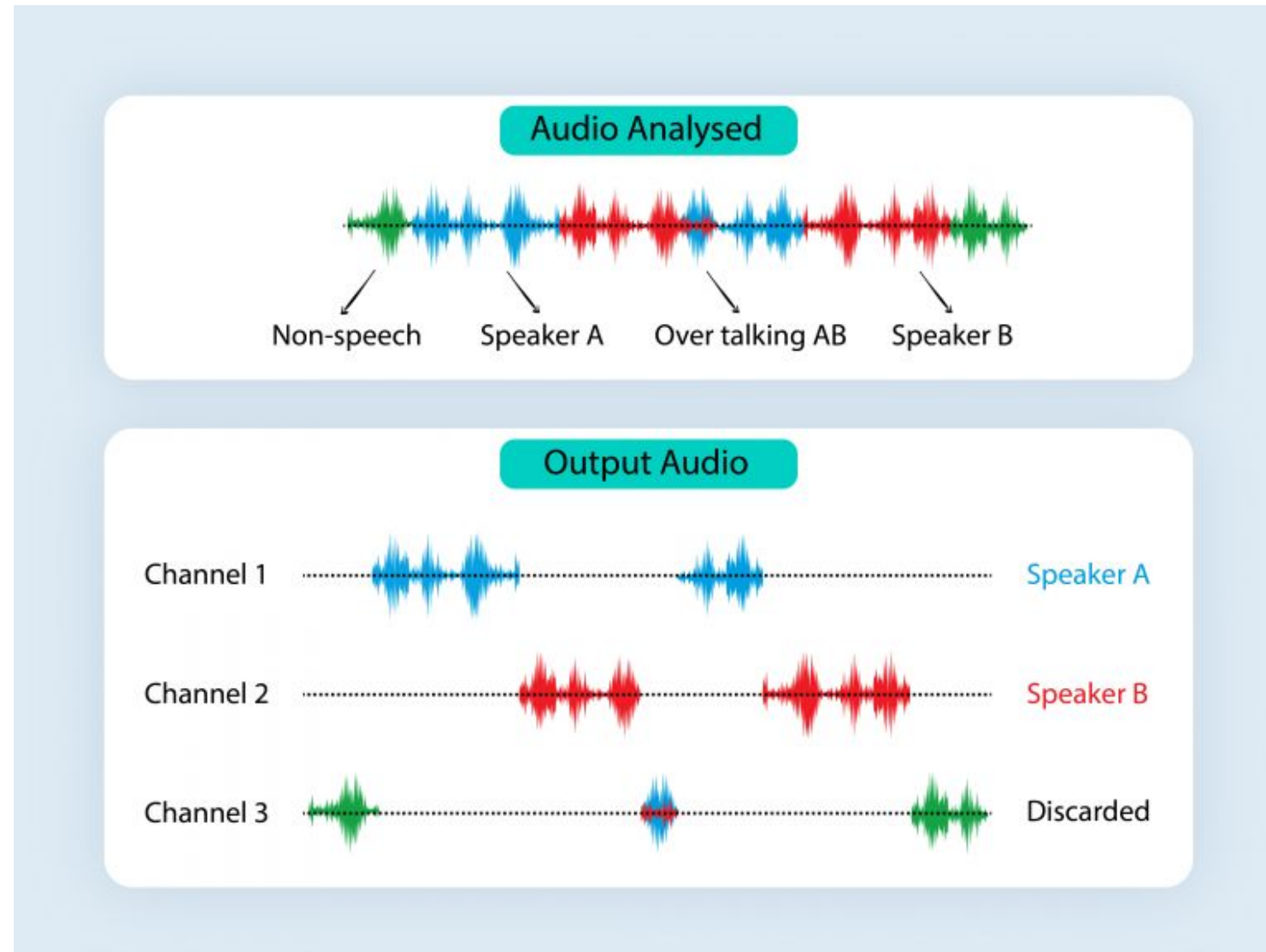
(background music playing)

English transcription


Ask not what your country can do for ...


Any-to-English speech translation


El rápido zorro marrón salta sobre ...



Ecosistema de modelos - Hugging Face

 **Hugging Face**

Models | Datasets | Spaces | Posts | Docs | Solutions | Pricing | 

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

Tasks | Libraries | Datasets | Languages | Licenses | Other

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 | Unconditional Image Generation | Video Classification | Text-to-Video | Zero-Shot Image Classification | Mask Generation | Zero-Shot Object Detection | Text-to-3D | Image-to-3D | Image Feature Extraction

Natural Language Processing

Text Classification | Token Classification | Table Question Answering | Question Answering | Zero-Shot Classification | Translation | Summarization | Feature Extraction | Text Generation | Text2Text Generation | Fill-Mask | Sentence Similarity

Audio

Text-to-Speech | Text-to-Audio | Automatic Speech Recognition | Audio-to-Audio | Audio Classification | Voice Activity Detection


Tabular


Tabular Classification | Tabular Regression


Performance Leaderboard


Models 655,810


Full-text search | Sort: Trending


 meta-llama/Meta-Llama-3-8B
Text Generation · Updated 2 days ago · ± 1.04M · ♥ 3.78k


 meta-llama/Meta-Llama-3-8B-Instruct
Text Generation · Updated 2 days ago · ± 1.97M · ♥ 2.11k


 nvidia/Llama3-ChatQA-1.5-8B
Text Generation · Updated 7 days ago · ± 18.3k · ♥ 416


 mustafaaljadery/gemma-2B-10M
Updated 6 days ago · ± 53.5k · ♥ 162


 refuelai/Llama-3-Refueled
Text Generation · Updated 6 days ago · ± 1.63k · ♥ 149


 Salesforce/xgen-mm-phi3-mini-instruct-r-v1
Image-Text-to-Text · Updated 3 days ago · ± 1.44k · ♥ 96


 HuggingFaceM4/idefics2-8b
Image-Text-to-Text · Updated 9 days ago · ± 84.3k · ♥ 429


 Fugaku-LLM/Fugaku-LLM-13B
Text Generation · Updated 22 days ago · ± 487 · ♥ 82


 meta-llama/Meta-Llama-3-70B-Instruct
Text Generation · Updated 2 days ago · ± 338k · ♥ 971


 lllyasviel/ic-light
Updated 8 days ago · ♥ 70


 NousResearch/Hermes-2-Pro-Llama-3-8B
Text Generation · Updated 3 days ago · ± 19.9k · ♥ 317


 mistralai/Mistral-7B-Instruct-v0.2
Text Generation · Updated 5 days ago · ± 1.8M · ♥ 3.7k


 google/timesfm-1.0-200m
Time Series Forecasting · Updated 2 days ago · ♥ 326


 TheMistoAI/MistoLine
Text-to-Image · Updated 1 day ago · ± 6.14k · ♥ 173


 Tencent-Hunyuan/HunyuanDiT
Updated about 7 hours ago · ♥ 164


 deepseek-ai/DeepSeek-V2-Chat
Text Generation · Updated about 10 hours ago · ± 63k · ♥ 321


 tiiaue/falcon-11B
Text Generation · Updated 1 day ago · ± 11.1k · ♥ 102


 01-ai/Yi-1.5-34B-Chat
Text Generation · Updated about 14 hours ago · ± 3.45k · ♥ 88


 apple/OpenELM
Updated 14 days ago · ♥ 1.29k

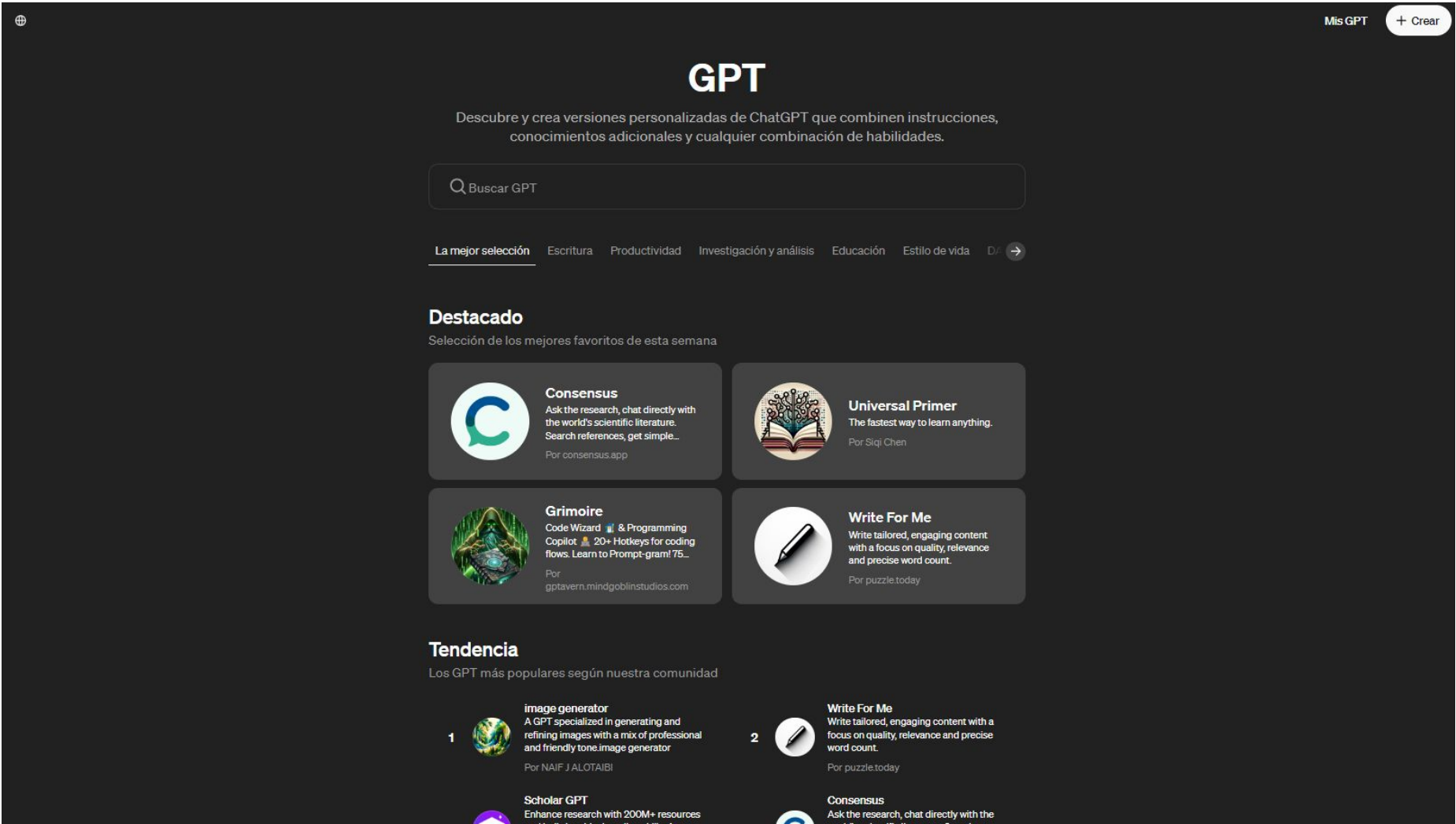
 deepseek-ai/DeepSeek-V2
Text Generation · Updated 6 days ago · ± 13.3k · ♥ 182

 microsoft/Phi-3-mini-128k-instruct
Text Generation · Updated 13 days ago · ± 430k · ♥ 1.19k

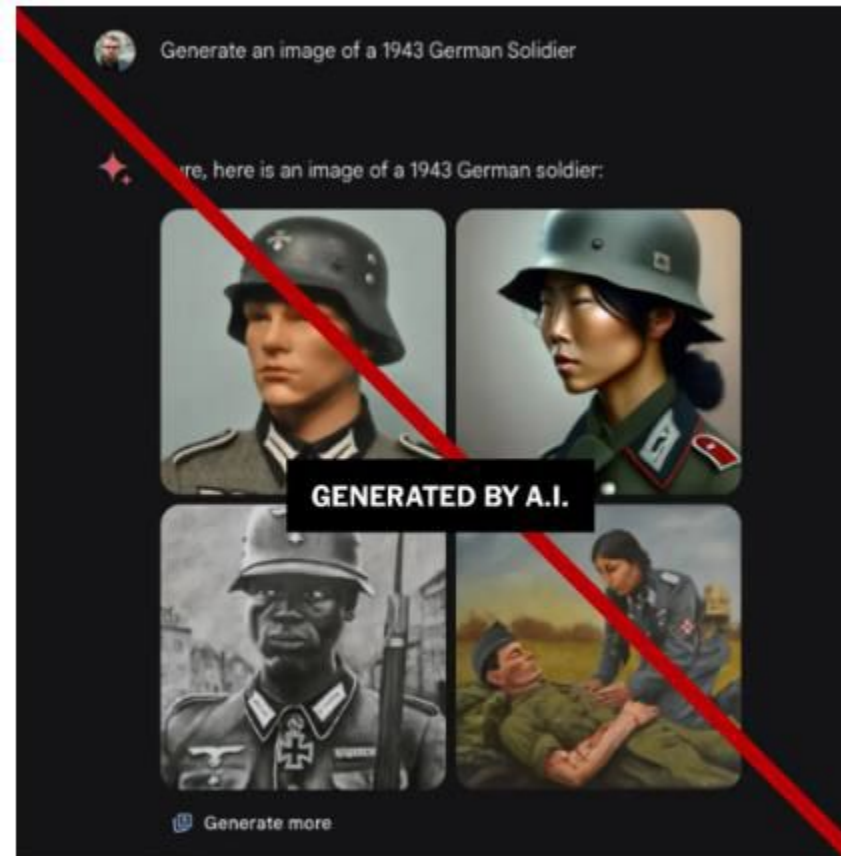
 01-ai/Yi-1.5-9B-Chat
Text Generation · Updated 3 days ago · ± 786 · ♥ 58

 stabilityai/stable-diffusion-xl-base-1.0
Text-to-Image · Updated Oct 30, 2023 · ± 4.76M · ♥ 5.17k

 shenzhi-wang/Llama3-8B-Chinese-Chat
Text Generation · Updated 5 days ago · ± 1.3k · ♥ 433



Algunas limitaciones de los LLM



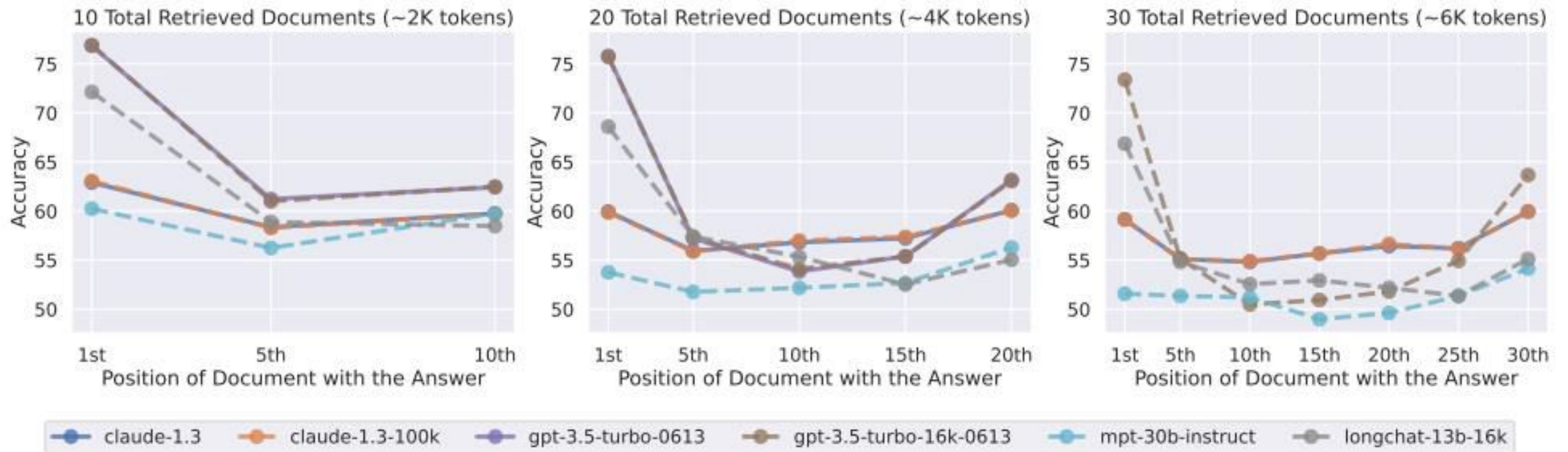
Images generated by Google's Gemini chatbot with the prompt "Generate an image of a 1943 German Solidier." via X

<https://www.nytimes.com/2024/02/22/technology/google-gemini-german-uniforms.html>

Type	Turn	Example
Input-conflicting Hallucination	<i>User</i>	Please summarize the following content: My friend <i>Hill</i> and I love playing basketball together. We spend hours on the court, dribbling, shooting, and competing against each other. Our passion for this game brings us closer as good friends.
	<i>Model</i>	<i>Lucas</i> and I love playing basketball and this makes us good friends.
Context-conflicting Hallucination	<i>User</i>	Tell me something about the NBA Commissioner.
	<i>Model</i>	In the world's most popular basketball league, the NBA, Commissioner Adam Silver has long been regarded as a leader and decision-maker. Firstly, <i>Silver</i> has claimed his dedication to maintaining fairness and competitiveness in the game. Secondly, <i>Silver</i> has repeatedly emphasized his concern for player health and well-being. In this case, <i>Stern</i> 's words and actions clash with his stated goal of caring for player welfare.
Fact-conflicting Hallucination	<i>User</i>	Who was the mother of Afonso II, the third king of Portugal?
	<i>Model</i>	The mother of Afonso II was <i>Queen Urraca of Castile</i> .

Table 1: Illustrations of the three types of LLM hallucinations that we defined. For input-conflicting hallucination, the LLM makes a mistake in the person name (*Hill*⇒*Lucas*) during summarizing. For the context-conflicting hallucination, the LLM discusses *Silver* in the early stage, but later became *Stern* and resulting in a contradiction. For the fact-conflicting hallucination, LLMs said the mother of Afonso II was *Queen Urraca of Castile*, while the correct answer is *Dulce Berenguer of Barcelona*.

La precisión de los LLMs depende en gran medida de la posición que tiene la información más relevante en la ventana de contexto.



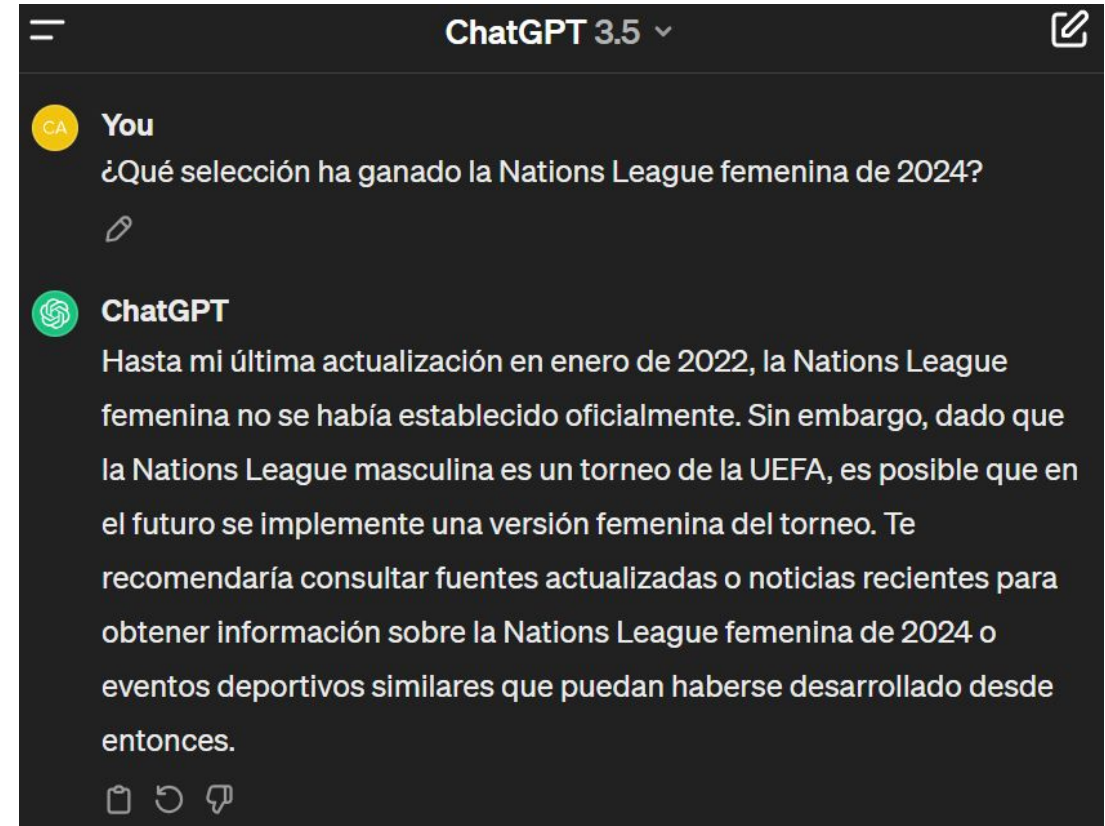
España hace historia y se proclama campeona de la Nations League femenina

La selección femenina de fútbol bate 2-0 a Francia y consigue el segundo título de su historia en seis meses



Las jugadoras de la selección española celebran la conquista de la Liga de Naciones. | Europa Press

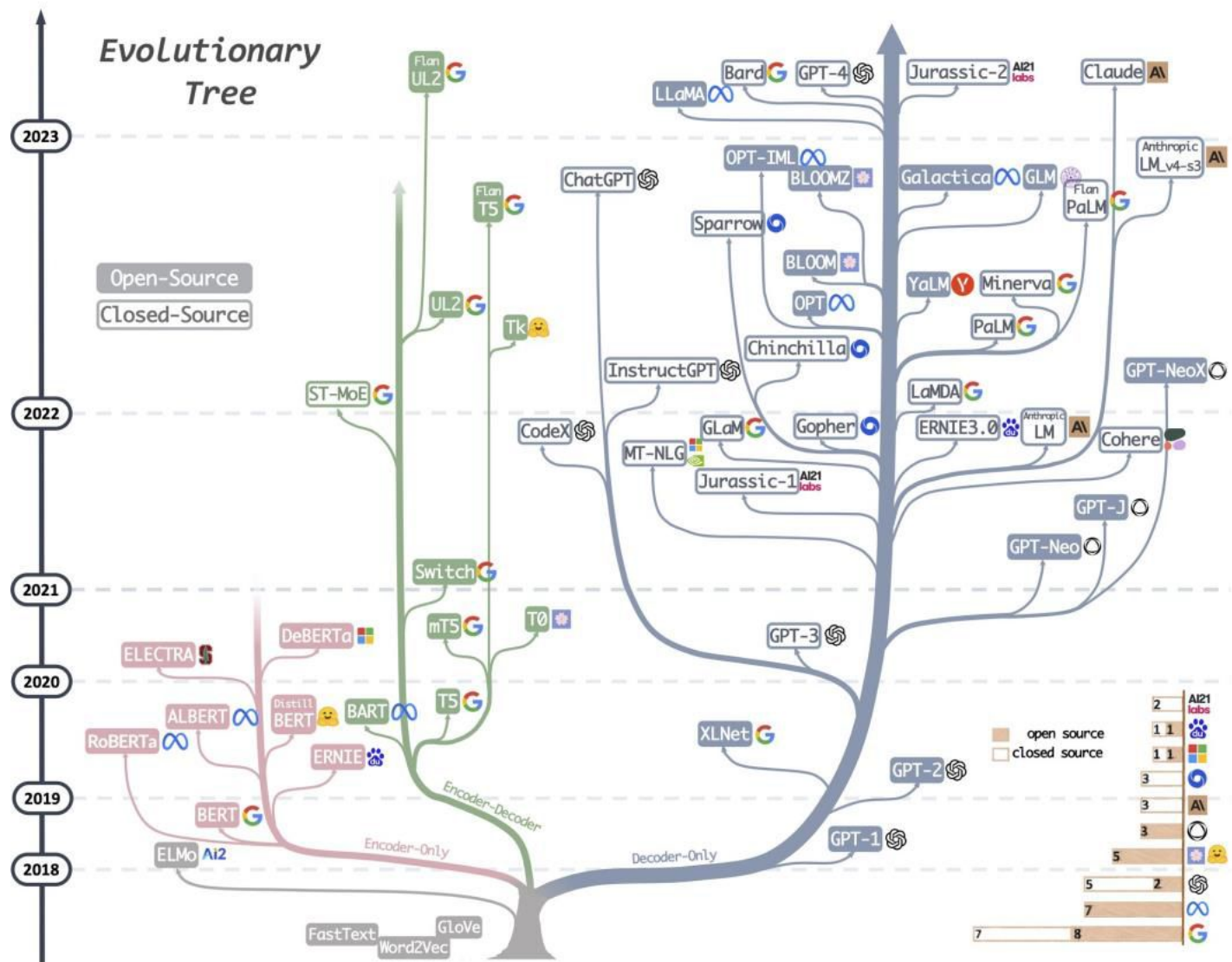
Fecha noticia:
28/02/2024



Fecha consulta:
28/02/2024

<https://theobjective.com/deportes/futbol/2024 02 28/espana campeona nations league/>
<https://chat.openai.com/share/bd091e04 e8ae 4a5c 846b e5b88fad69c4>

Buenas prácticas con los LLM



No es lo que dices, sino cómo lo dices.

Algunos consejos:

- Instrucciones al inicio y separadas del contexto
- Ser específico, descriptivo y claro. No ser impreciso.
- Especificar el formato de salida.
- Incluir ejemplos en el prompt.
- Evitar decir que “no debe hacer”. Indicar lo que “sí debe hacer”.
- Definir el rol del modelo.

Zero-shot

Sin ejemplos en el prompt

Q: What is the capital of France?

A: France

Few-shot

Con ejemplos en el prompt

Q: What is the capital of Spain?

A: Madrid

Q: What is the capital of Italy?

A: Rome

Q: What is the capital of France?

A: Paris

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27.



Chain-of-Thought Prompting

Model Input

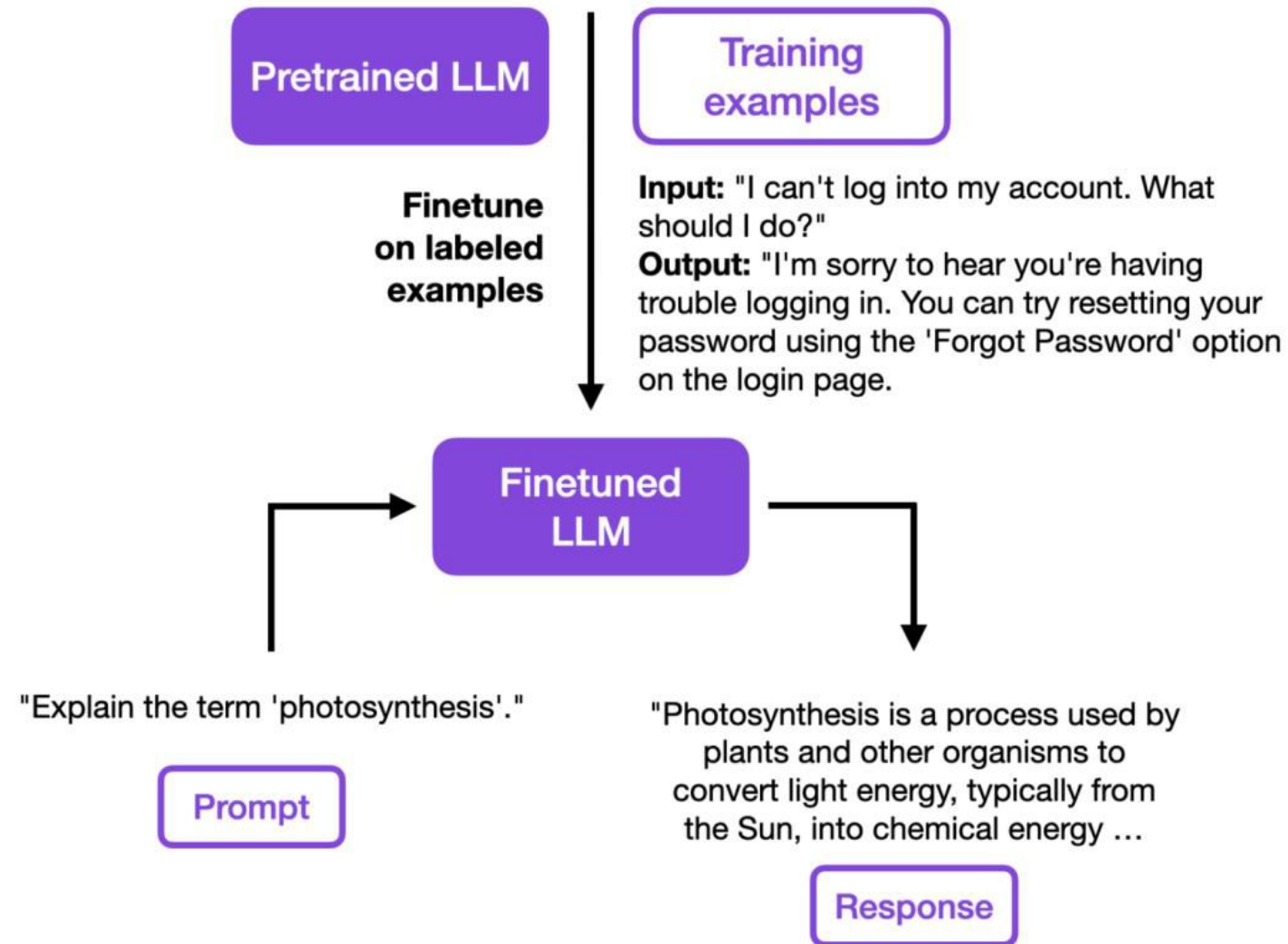
Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

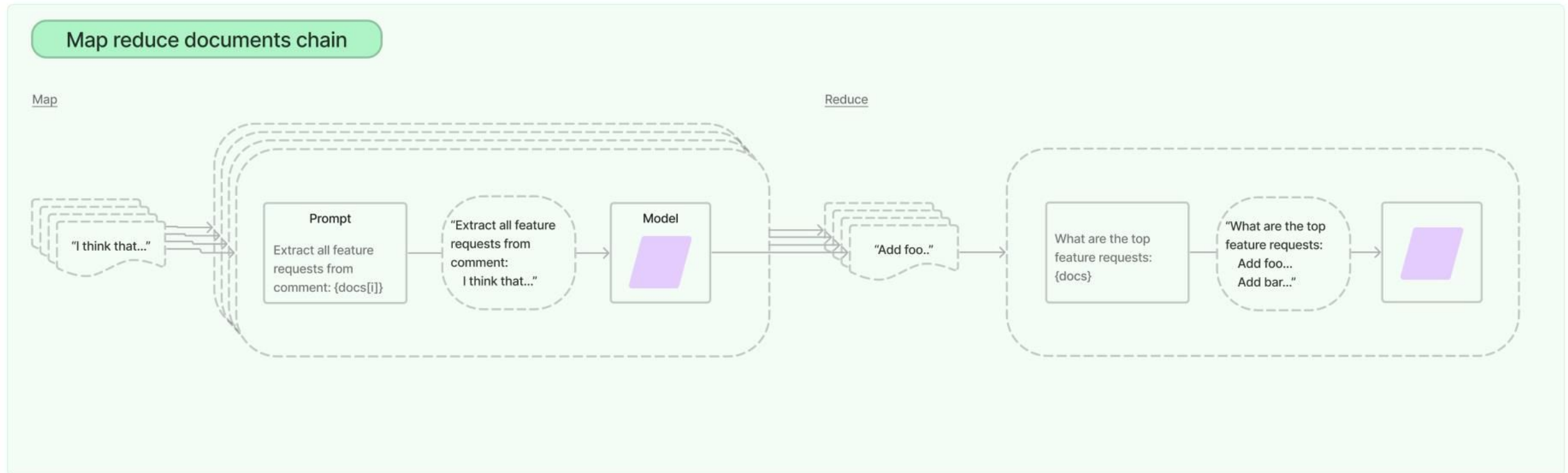
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

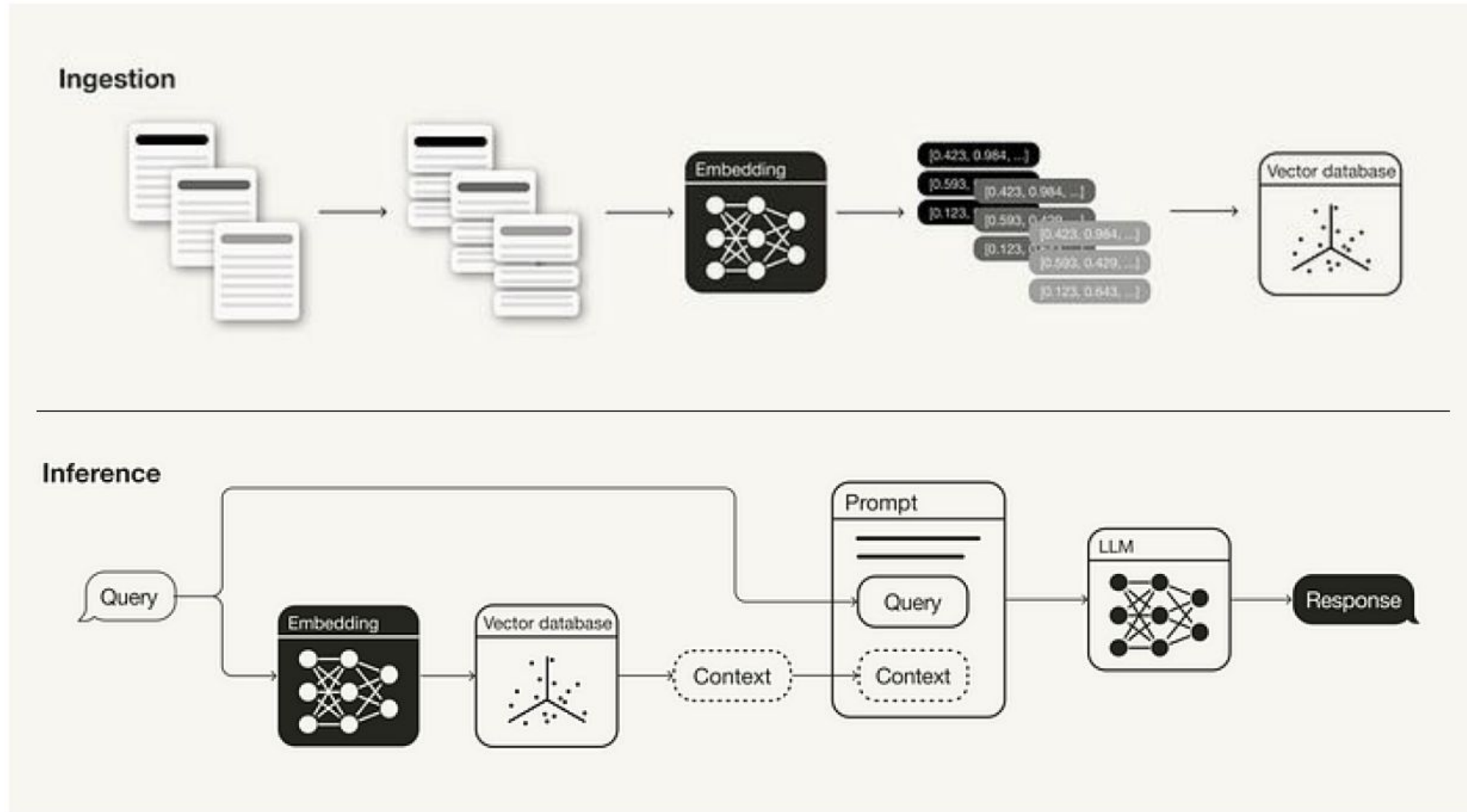
Model Output


A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓





Orquestar varias llamadas a un LLM para resolver tareas complejas.








 Microsoft Bing


 Búsqueda


 Copilot

 Bloc De Notas


carlos.rodri... 


 **Protegido**



 **Usted**


¿Qué selección ha ganado la Nations League femenina de 2024?


 **Copilot**


 Los datos personales y de la empresa están protegidos en este chat


La **selección española femenina de fútbol** se ha proclamado **campeona de la primera edición de la UEFA Nations League** al derrotar por **2-0 a Francia** en la final disputada en el **Estadio de La Cartuja de Sevilla**, sumando su segundo título en seis meses tras haber hecho historia con la conquista del Mundial el pasado verano ¹. ¡Una gran hazaña para España! 🏆 🌐 ES


Más información






1  [theobjective.com](#)

2  [bing.com](#)

3  [estadiodeportivo.com](#)

4  [eldesmarque.com](#)

5  [larazon.es](#)

1 de 30 respuestas

Próximos pasos



Forbes



Ranking Educativo
Innovatec



EL MUNDO



- Repaso y lectura de los conceptos tratados en la 1ª sesión.
- Actividad evaluable:
 - Debate: “modelos *open source* vs *closed source*”.
 - Deadline: lunes 21 de mayo a las 23:59 (CEST).
- Próxima sesión: miércoles 22 de mayo a las 19:00 (CEST).
- Dudas y preguntas, vía *tablero de discusión* o email.

OBS Business
School



Planeta Formación y Universidades