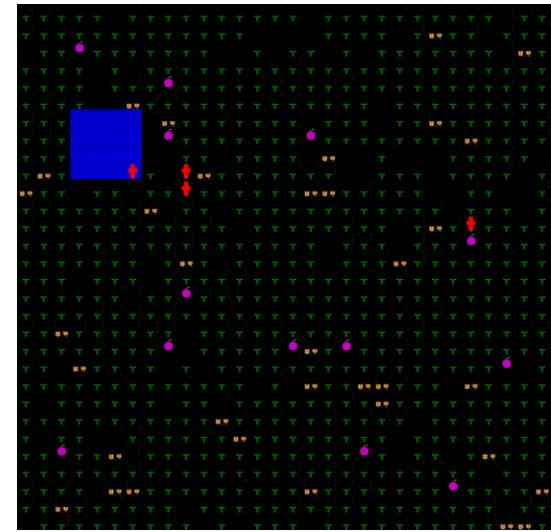


Sueñan los abogados con “Loros Estocásticos”

Limites y oportunidades de los LLMs

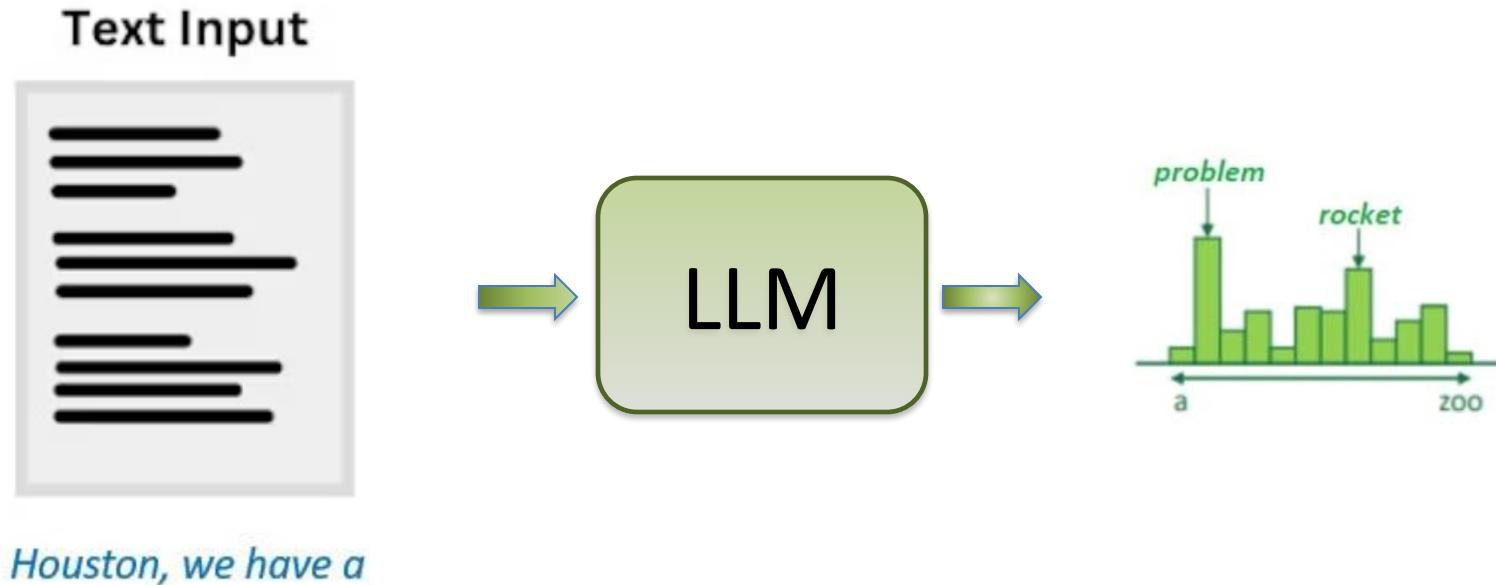
Martí Sánchez-Fibla

marti.sanchez@iiia.csic.es



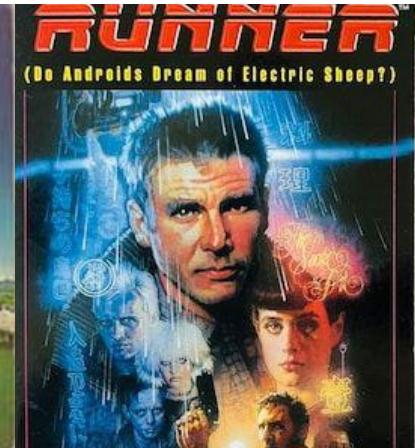
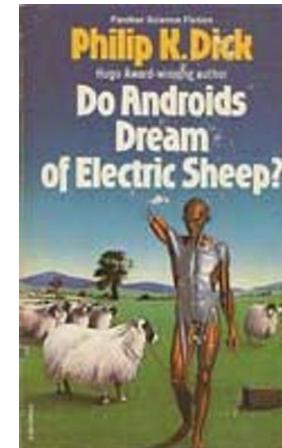
XXXII JORNADES JURIDIQUES
DE LA FACULTAT DE DRET,
ECONOMIA I TURISME

¿Qué es un LLM?



Large Language Model

¿Sueñan los androides con ovejas eléctricas?



Loros estocásticos

- Un “paper” de 2021 llama a los LLMs “Stochastic Parrots”

On the dangers of stochastic parrots: Can language models be too big? 

[EM Bender, T Gebru, A McMillan-Major... - ... ACM conference on ..., 2021 - dl.acm.org](#)

... In §4, we discuss how large datasets based **on** texts from the ... that do not necessarily depend **on** having larger LMs. As we ... Focusing **on** state-of-the-art results **on** leaderboards without ...

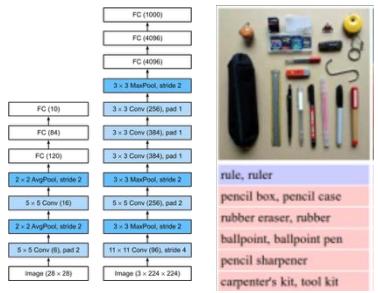
 Save  Cite Cited by 9884 Related articles All 20 versions

- ¿Qué podéis decir del año **2021**?
- Es 1 año a.C.

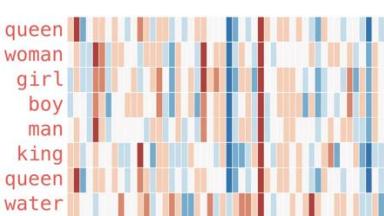


No es un invento reciente

AlexNet: Convolutional



Word2Vec: Embeddings



Transformers

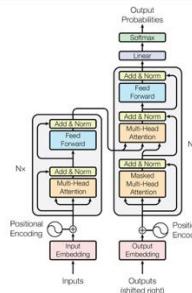


Figure 1: The Transformer - model architecture.

AlphaEvolve
AlphaProof
AlphaGeometry



2012 2013

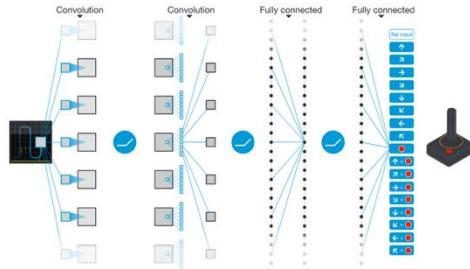
2016 2017

2019

2022

2024 2025

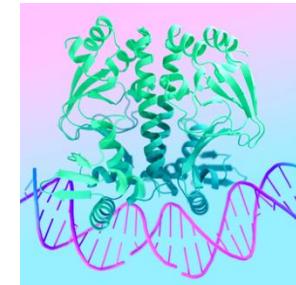
Atari: DQN



AlphaStar: StarCraft II



AlphaFold 3:
Protein Interactions



No es un invento reciente 2013 - 2024



2013 Word2Vec Embeddings learnt to represent Characters, Parts of Word

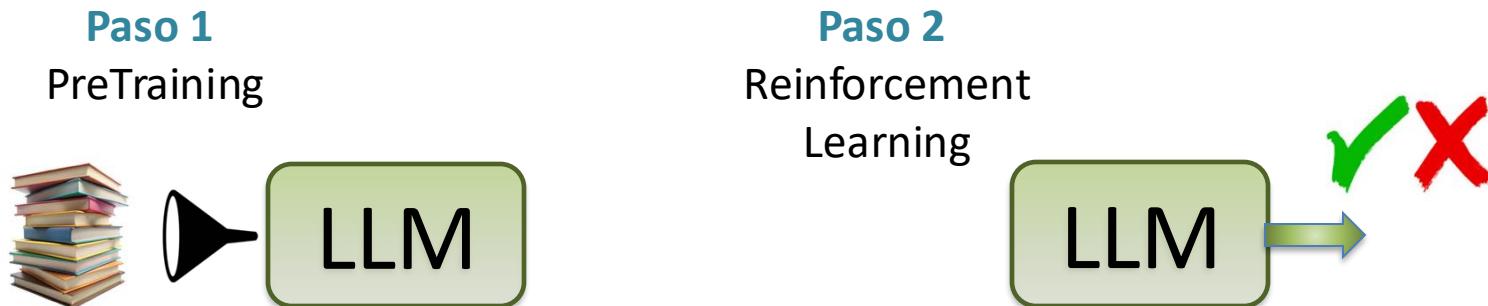
2017 Transformers Encoder / Decoder architecture for Language Translation

2019 GPT2 **1.5B params** Transformer, autoregressive Decoder-Only,
Finetuned: summarization, question-answering, sentiment classification

2020 GPT3 Scaling. **175 Billion params.**

2022 GPT3.5 + RLHF, Instruct models RL from Human Feedback

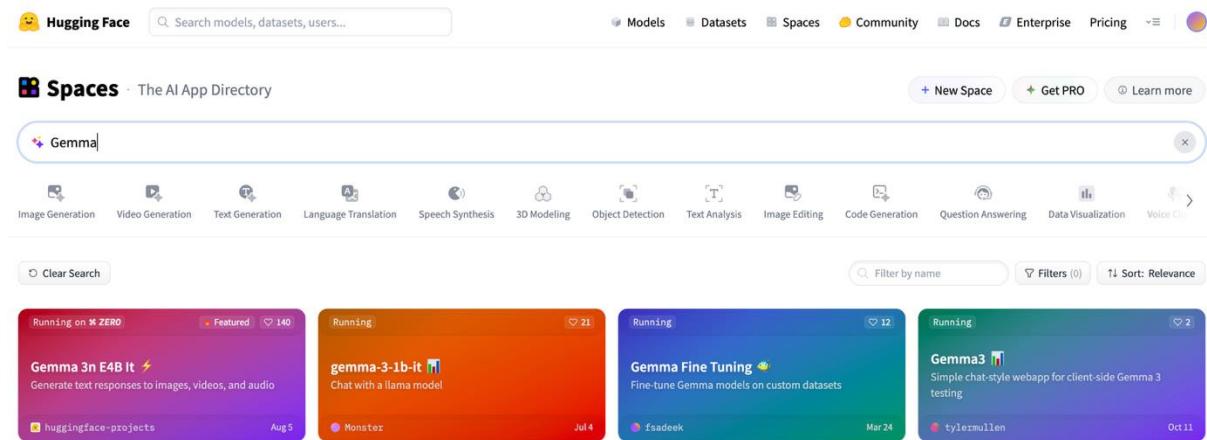
2024 DeepSeek R1 Chain of Thought, test time compute,



¿Cómo usarlos? Abiertos-Open Weights

• Huggingface

<https://huggingface.co/spaces/>



• Instalar un gestor de LLMs

- Descarga de modelos (Qwen, Deepseek, LLaMA, Gema)
- Control de parámetros context, window, temp, GPU/CPU settings)

Usa tu propio ChatGPT local

Ollama

Modelos locales de IA (LLM)

1 Ollama es un programa (cliente) que permite cargar modelos de IA* (LLM) con tu propio sistema, sin hacer uso de Internet.

<https://ollama.com/download>

2 Elige y descarga un modelo de IA* (LLM):

A screenshot of the Ollama client application. The interface has a dark theme with white text. At the top, there's a menu with options like "File", "Edit", "View", "Settings", "Help", and "About". Below the menu, there's a section for "System Prompt" with a dropdown menu and a text input field containing "Example, 'Only answer in Spanish'". The main area is a chat window with two messages: "Hola, ¿Qué tal?" and "Observa esta imagen, Es la port...". At the bottom, there's a dropdown menu labeled "gemma3-1b" and a "Send a message" button.

LMStudio

Local LLM Server

LM Studio comes with a built-in local server.

You can use our SDKs or the OpenAI compatibility API.

A screenshot of the LMStudio interface. The window is titled "LMStudio" and contains several tabs: "System Prompt", "Settings", "Sampling", "Report Penalty", and "Output". Under "System Prompt", there's a text input field with the placeholder "Example, 'Only answer in Spanish'". Under "Settings", there are sliders for "Temperature" (set to 0.6), "Limit Response Length" (set to 100), and "Max Token Count" (set to 10). Under "Sampling", there are sliders for "Top P Sampling" (set to 0.6) and "Min P Sampling". Under "Report Penalty", there's a slider for "Top P Sampling" (set to 0.6). Under "Output", there are checkboxes for "Structured Output" and "Unstructured Output". On the right side of the window, there's a log window with some text about the server's status.

¿Cómo usarlos? Google Colab

YT Alpaca 7B LoRa.ipynb Changes will not be saved

File Edit View Insert Runtime Tools Help

Commands + Code + Text Run all Copy to Drive

RAM Disk

git clone <https://github.com/tloen/alpaca-lora.git>

Cloning into 'alpaca-lora'...
remote: Enumerating objects: 95, done.
remote: Counting objects: 100% (95/95), done.
remote: Compressing objects: 100% (69/69), done.
remote: Total 95 (delta 53), reused 64 (delta 26), pa
Unpacking objects: 100% (95/95), 6.88 MiB | 8.10 MiB/

Alpaca LLaMa 7B LoRa

%cd alpaca-lora/
/content/alpaca-lora

!pip install -q datasets loralib sentencepiece
!pip uninstall transformers
!pip install -q git+<https://github.com/zphang/transf>
pip install git+<https://github.com/zphang/transf>

Resources

space? [Upgrade to Colab Pro](#)

Python 3 Google Compute Engine backend (GPU)
Showing resources from 22:26 to 22:27

System RAM 2.9 / 12.7 GB

GPU RAM 0.0 / 15.0 GB

Disk 38.1 / 112.6 GB

Change runtime type

Variables Terminal

T4 (Python 3)

Herraminetas: notebooklm.google.com

- NotebookLM



Analysis de papers y
generador de
podcasts



:

Large Language Models
as Tax Attorneys

Oct 14, 2025 • 1 source

Large language models as tax attorneys: a case study in legal capabilities emergence
JJ Nay, D Karamardian, SB Lawske, W Tao, M Bhat, R Jain, AT Lee, ...
Philosophical Transactions of the Royal Society A 382 (2270), 20230159
2024 64 Citations

¿Pueden los LLMs substituir a abogados?

- Tamaño / Capacidades del modelo es el factor clave
- Los LLMs democratizan el acceso a servicios jurídicos básicos



John Nay

Other names ▾

Stanford University - CodeX - Center for Legal Informatics; [NYU](#)

Verified email at nyu.edu - [Homepage](#)

Machine Learning Computational Law Artificial Intelligence AI & Law AI Alignment

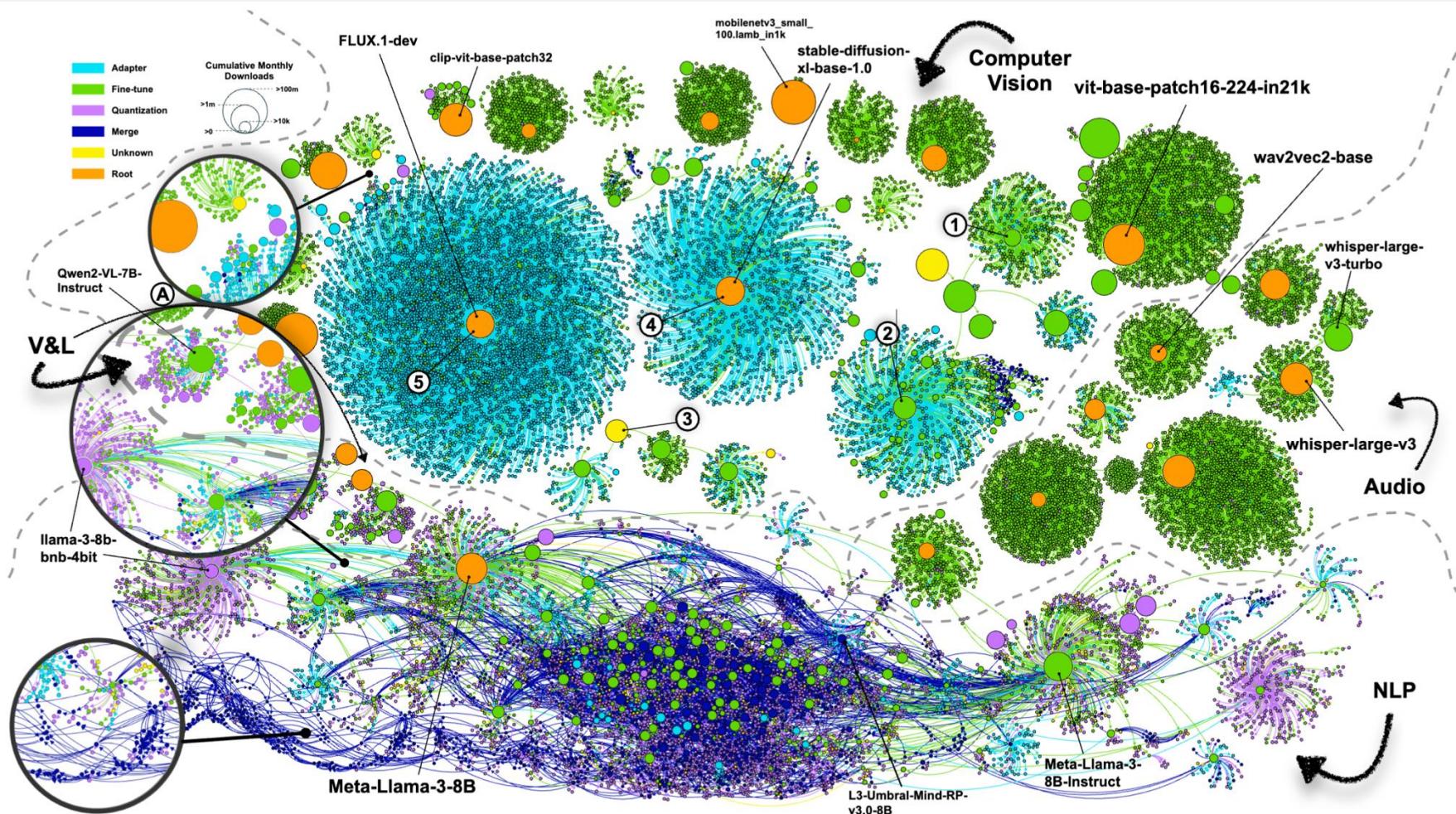
Large language models as tax attorneys: a case study in legal capabilities emergence

JJ Nay, D Karamardian, SB Lawske, W Tao, M Bhat, R Jain, AT Lee, ...

Philosophical Transactions of the Royal Society A 382 (2270), 20230159

2024 64 Citations

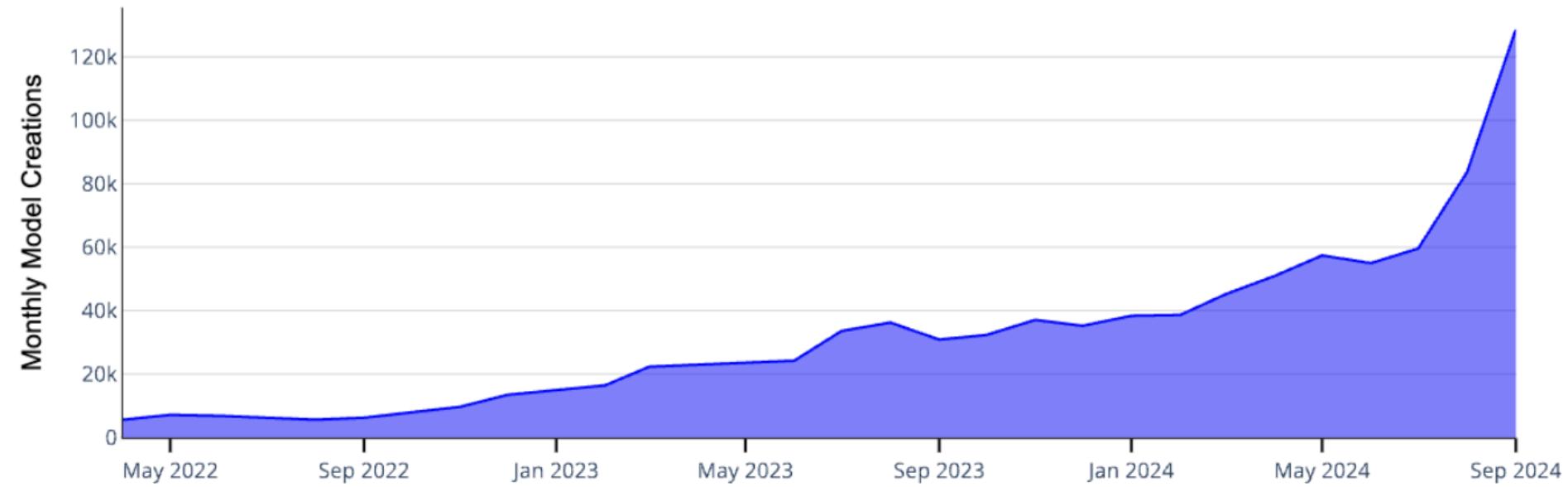
¿Cuantos modelos existen?



Pequeña selección de 63000 Modelos de [HuggingFace](#)

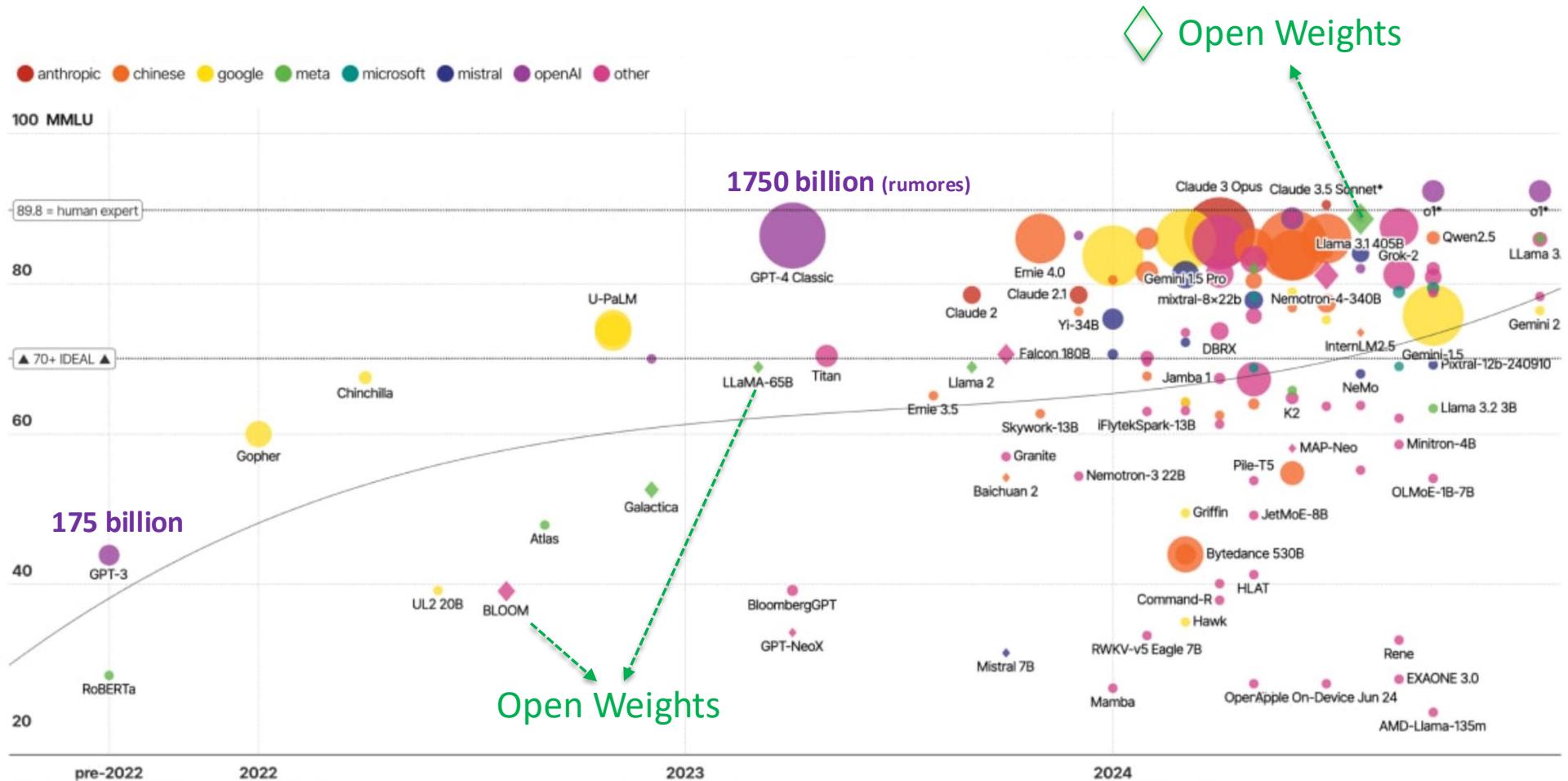
¿Cuantos modelos existen?

Modelos añadidos cada mes en [HuggingFace](#)



¿Qué LLMs existen?

- Grandes empresas que compiten por nuestro uso, datos, tiempo y control remoto



¿Qué LLMs existen? Cerrados-Privados

- Grandes empresas que compiten por nuestro uso, **datos, tiempo y control remoto**

	Company	Model
1		Google Gemini 3 Pro
2		xAI Grok 4.1 Thinking
3		OpenAI GPT 5.1 (High)
		Anthropic Claude Sonnet 4.5-20250929 Thinking-32k
		Alibaba Qwen 3 Max Preview
		Z.ai GLM 4.6
		Moonshot Kimi K2 Thinking-Turbo

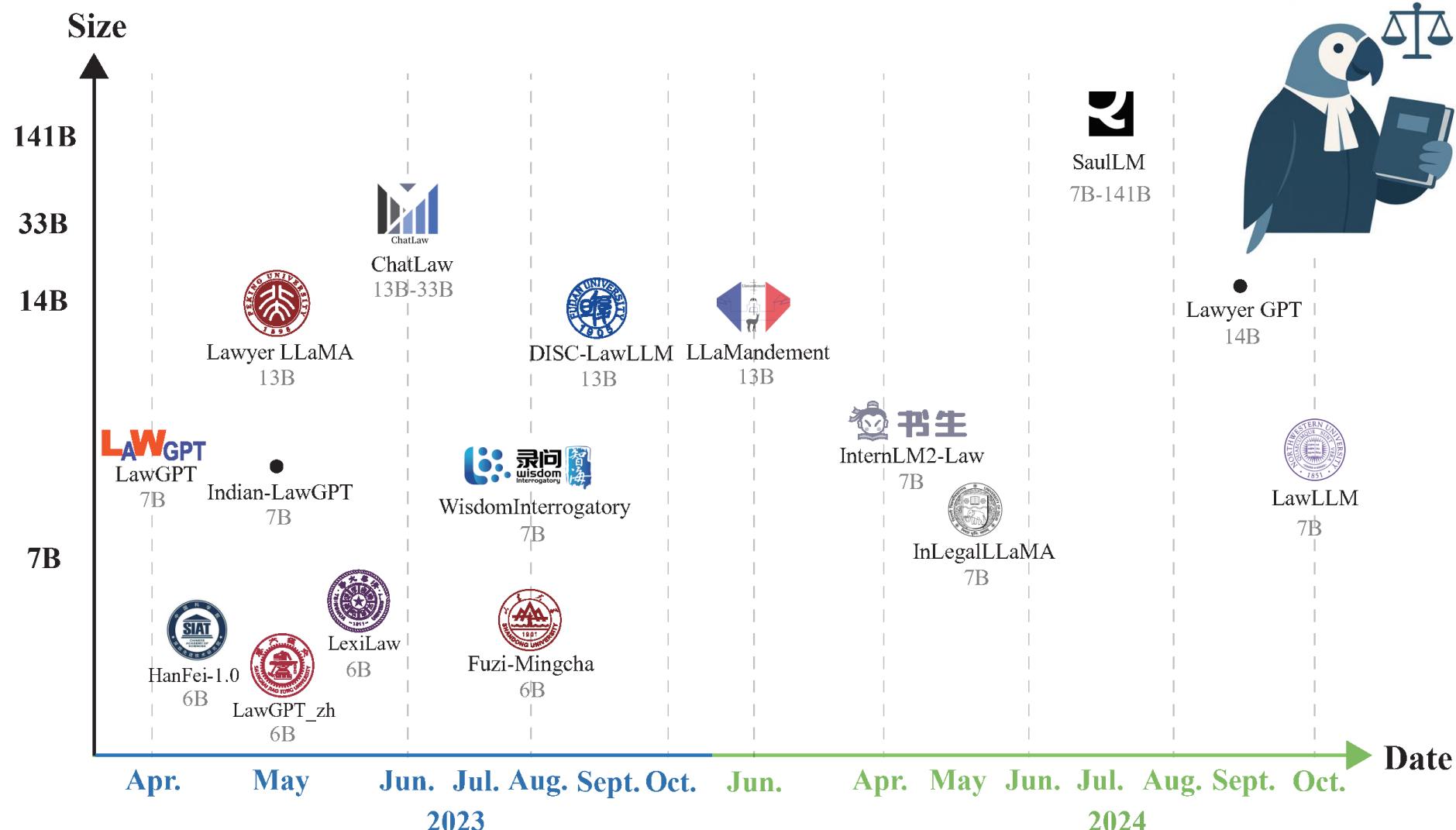
Leaderboard a
22 de
Noviembre
2025

¿Qué LLMs existen? Abiertos-Open Weights

- Grandes y Pequeñas empresas que compiten por nuestro uso y **que integremos sus modelos**

Rank	Model	Company	License	
17	glm-4.6	Z.ai	MIT	
21	qwen3-235b-a22b-instruct-2507	Alibaba	Apache 2.0	
22	deepseek-v3.2-exp-thinking	DeepSeek AI	MIT	
25	deepseek-r1-0528	DeepSeek	MIT	
27	deepseek-v3.1	DeepSeek	MIT	
28	deepseek-v3.1-thinking	DeepSeek	MIT	
29	kimi-k2-0711-preview	Moonshot AI	Modified MIT	

¿Qué LLMs jurídicos existen?



Large Language Models Meet Legal Artificial Intelligence: A Survey

Z Hou, Z Ye, N Zeng, T Hao, K Zeng

arXiv preprint arXiv:2509.09969, 2025 • arxiv.org

¿Qué es un LLM?

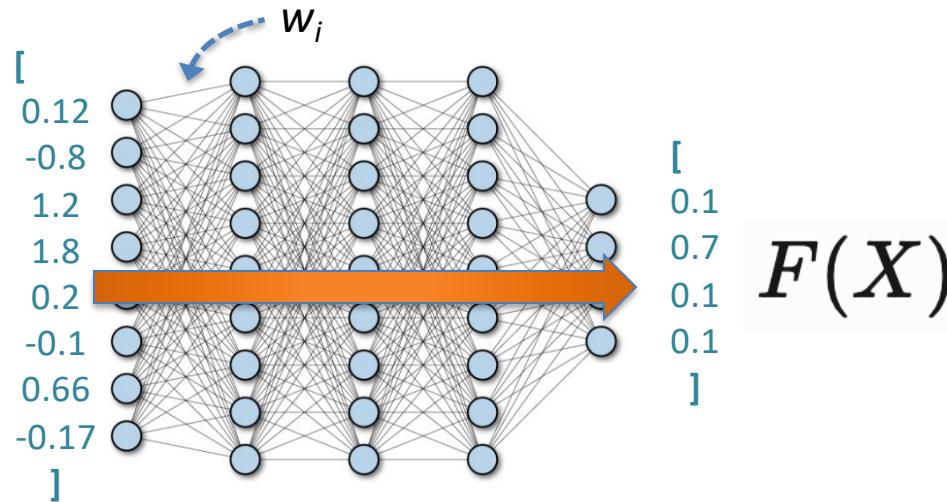
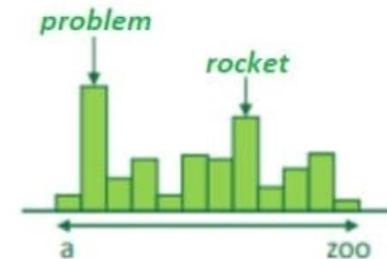
Tipo de Red Neuronal

Text Input

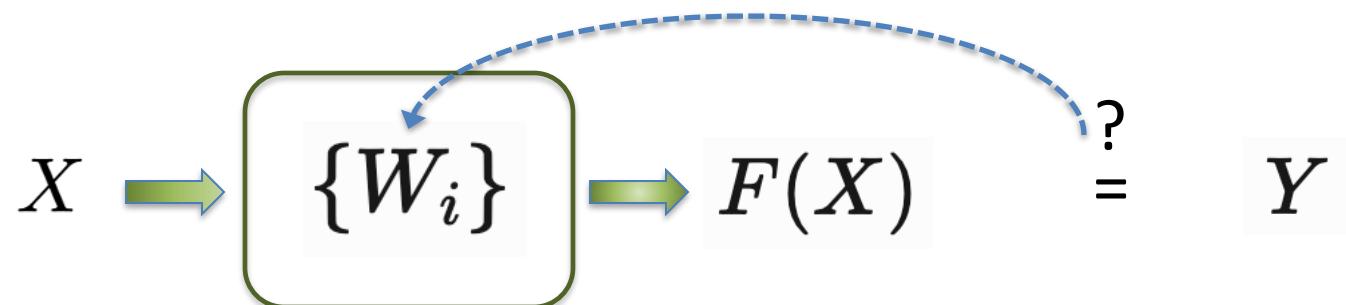


Houston, we have a

X



¿Cómo aprende una Red Neuronal?



Number 2



Border Terrier

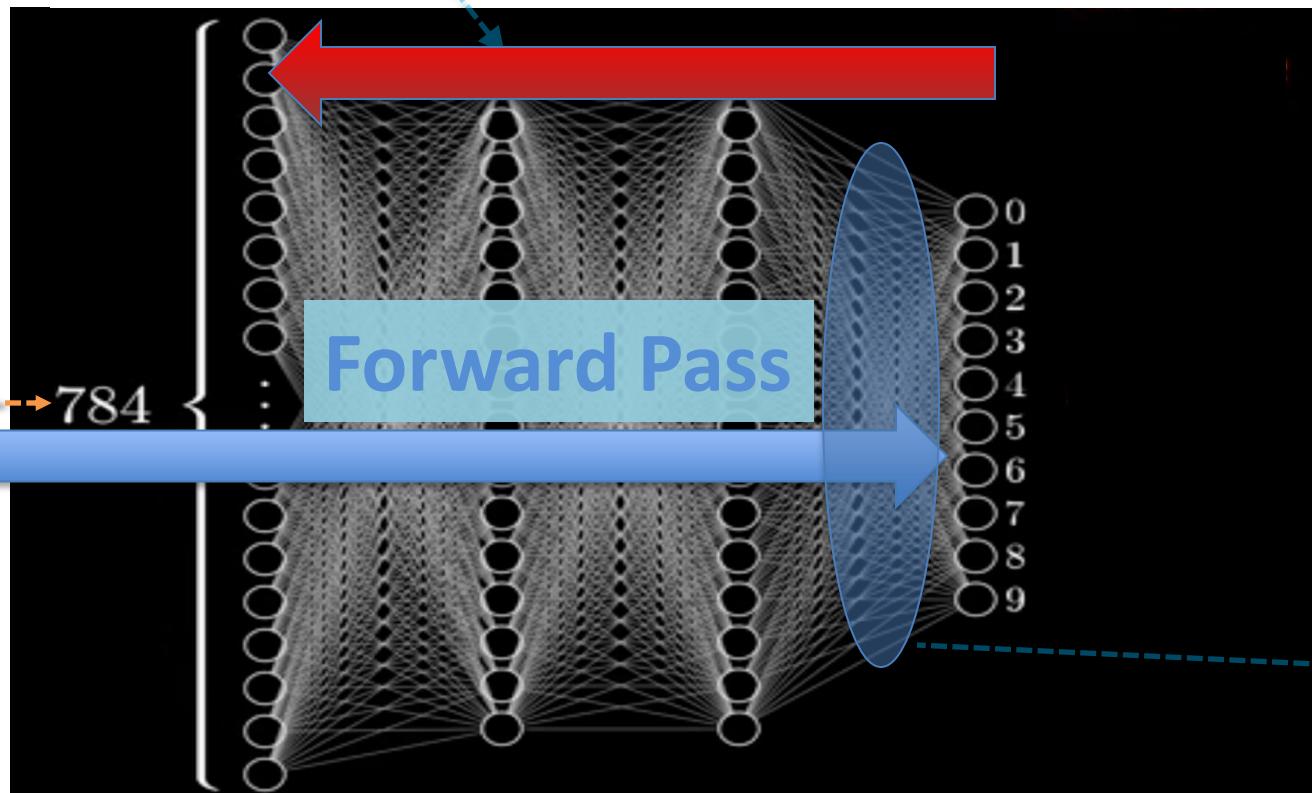
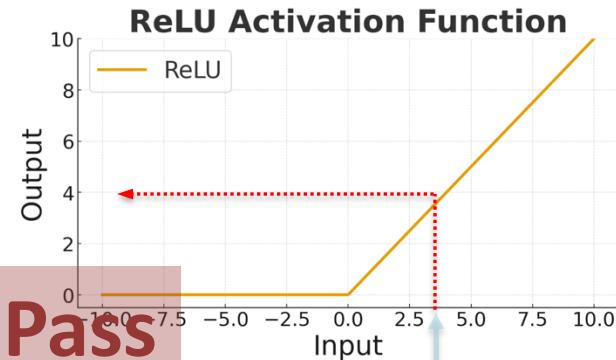
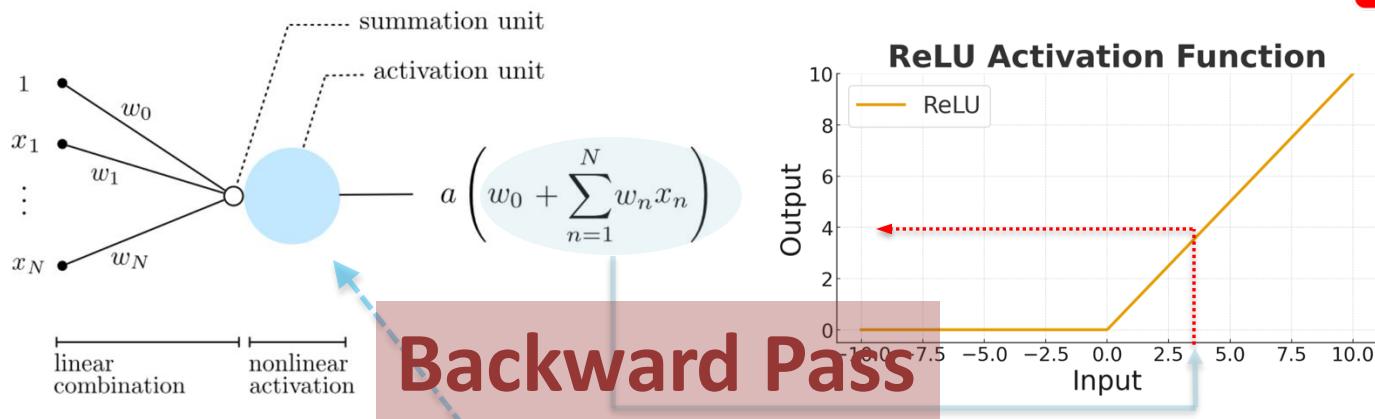
Text Input



Text Output



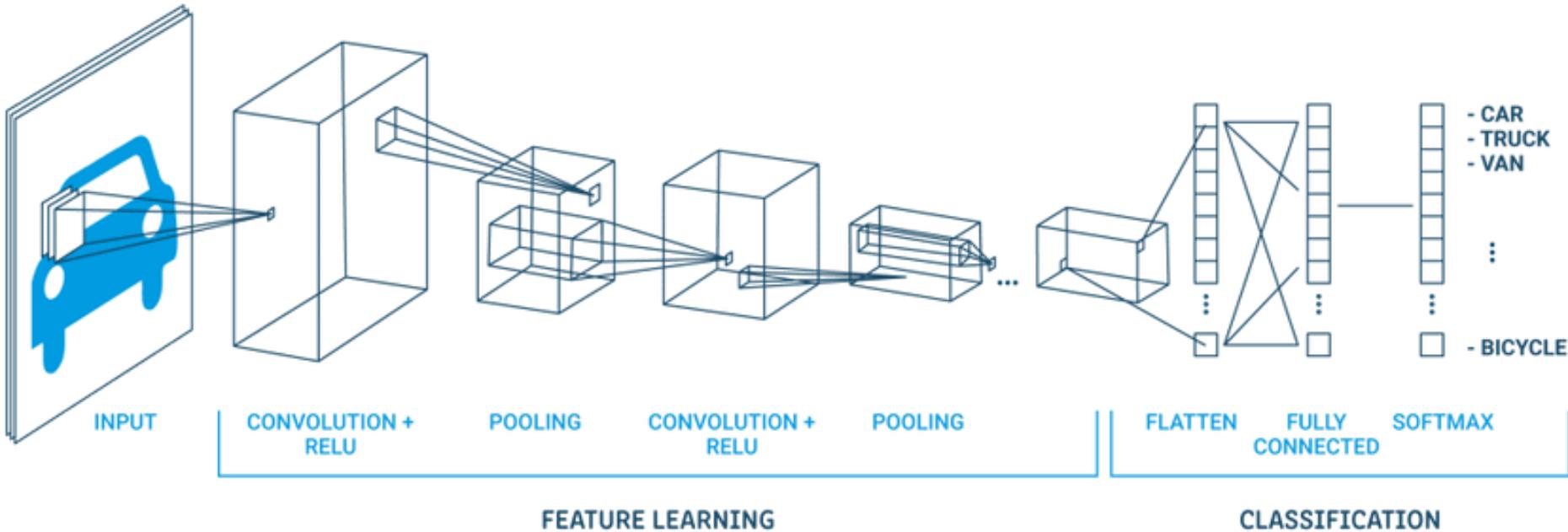
Red Neuronal Multi Capa



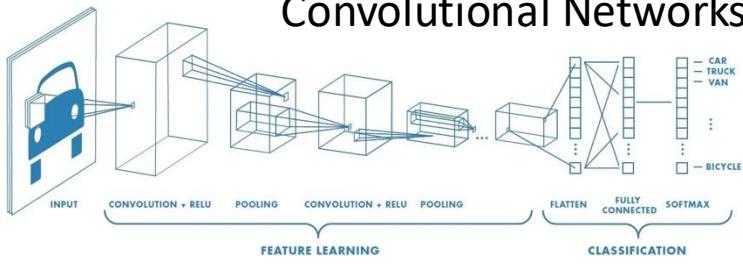
Muchos Tipos de Redes Neuronales



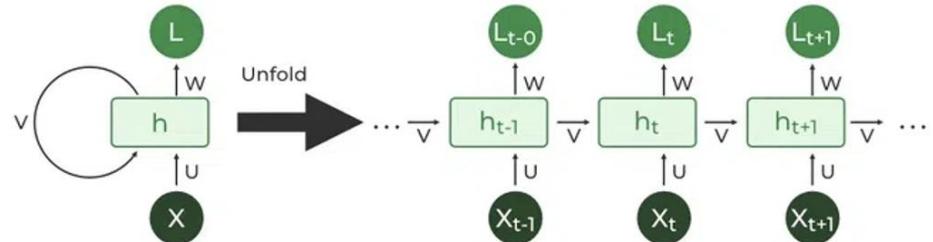
Convolutional Networks



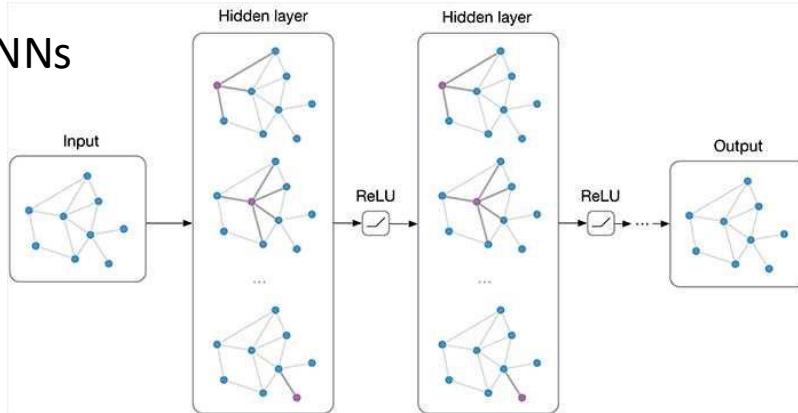
Muchos Tipos de Redes Neuronales



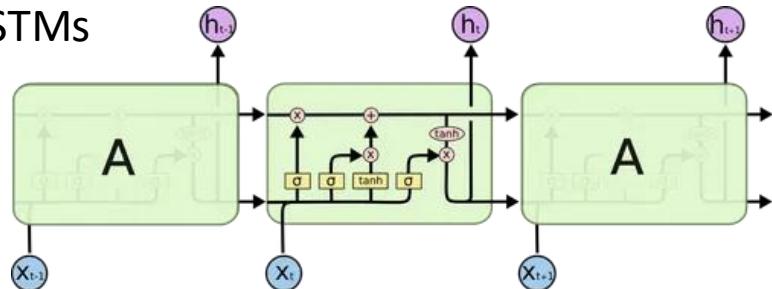
Recurrent Neural Networks



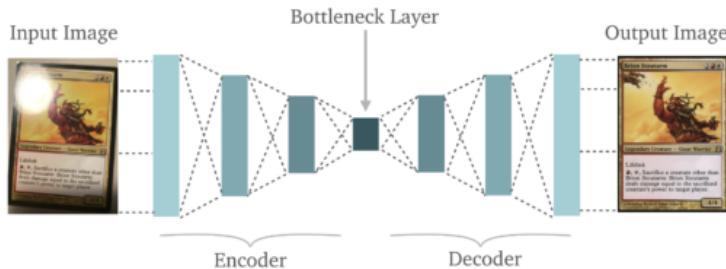
GNNs



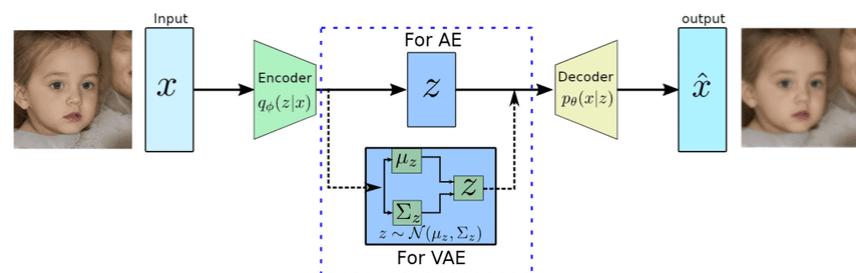
LSTMs



AutoEncoders

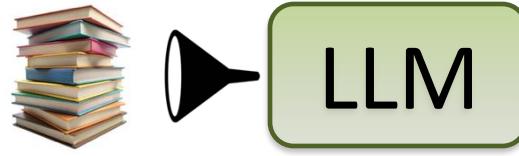


Vaiational AutoEncoders



¿Cómo aprende LLM?

- Next Word Prediction
 - **Datos sin etiquetas!!!!**



Pre-Training



- ¿Saben hablar idiomas?
- ¿La jerga jurídica un idioma en sí misma?
 - Términos, reglas, leyes dependen de cada país

¿Cómo aprende LLM?

- Enormes Datasets de Texto públicos



Dataset	Language	Scale	Text Type
HanFei* (2023)	Chinese	60G	Legal Case Documents, Statutes, Litigation Documents, Legal news
JEC-QA (2020b)	Chinese	26K Samples	Legal Exams
LawGPT* (2024)	Chinese	500K Samples	Legal Case Documents
WisdomInterrogatory* (2024)	Chinese	40G	Legal Case Documents, Legal QA Dataset
InLegalLLaMA* (2024)	English	10K Samples	Legal Case Documents
NyayaAnumana (2025)	English	22M Samples	Legal Case Document
Pile of Law (2022)	English	291.5G	Court Opinions, Contracts, Administrative Rules, Legislative Records
SaulLM* (2024a)	English	520B	Contracts, Court Transcripts, Statutes, Legislative Records, Legal Corpus
LBOX OPEN (2022)	Korean	147K Samples	Legal Case Document

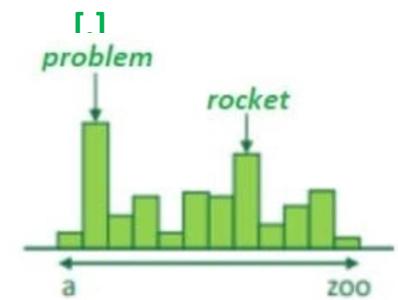
[Large Language Models Meet Legal Artificial Intelligence: A Survey](#)

[Z Hou, Z Ye, N Zeng, T Hao, K Zeng](#)

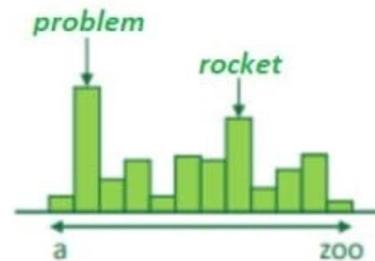
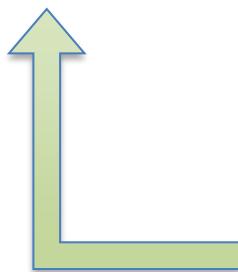
arXiv preprint arXiv:2509.09969, 2025 • arxiv.org

LLMs: Generación de respuesta iterativa

Text Input



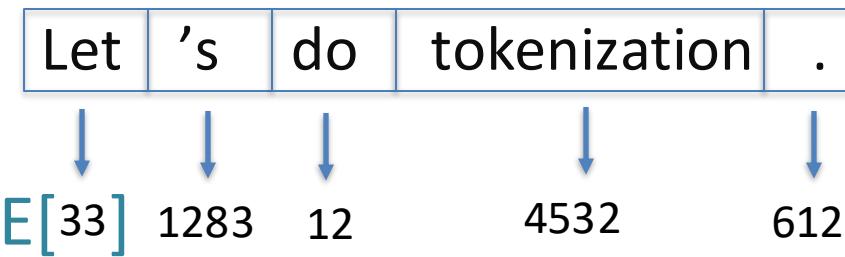
Houston, we have a problem



¿Qué pasa dentro de un LLM?

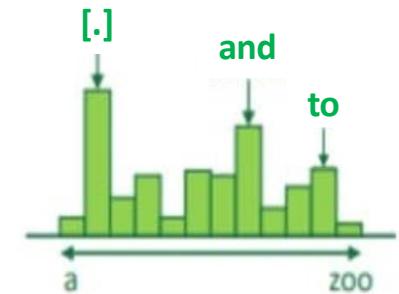


- Nos dan las palabras de input, el **PROMPT**. Las dividimos en partes según una base de datos de partes de cada lengua **TOKENIZATION**
- **PROMPT:** Let's do tokenization.



- Para cada Token vamos a buscar su Embedding id y sus vectores

Let	.21	.01	.002	0	.1	0	.01	.17	.11	.001
's	0	.01	.17	.11	.001	.21	.01	.002	0	.1
do	.3	.01	.002	0	.1	.01	.17	.11	.001	.21
tokenization	.01	.02	0	.1	.01	.17	.11	.01	.1	.2



¿Qué pasa dentro de un LLM?



PROMPT

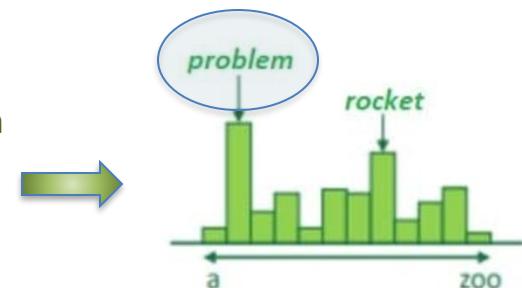
Houston
we
have
a



$$\mathbb{R}^{n \times d}$$

.21	.01	.002	0	.1	0	.01	.17	.11	.001
0	.01	.17	.11	.001	.21	.01	.002	0	.1
.3	.01	.002	0	.1	.01	.17	.11	.001	.21
.01	.02	0	.1	.01	.17	.11	.01	.1	.2

Houston
we
have
a



Era el token ¿problem?

¿Queeeeeee?



¡No necesitamos
etiquetas!

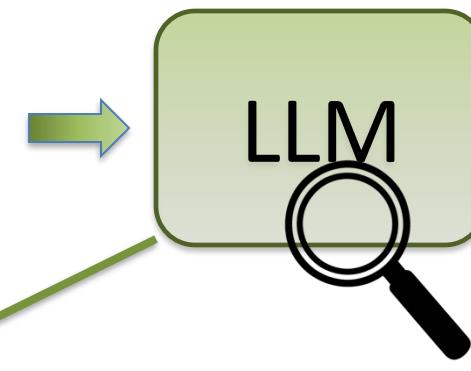
¿Qué pasa dentro de un LLM?



Tokenization + Embedding

Houston
we
have
a

.21	.01	.002	0	.1	0	.01	.17	.11	.001
0	.01	.17	.11	.001	.21	.01	.002	0	.1
0.3	.01	.002	0	.1	.01	.17	.11	.001	.21
.01	.02	0	.1	.01	.17	.11	.01	.1	.2



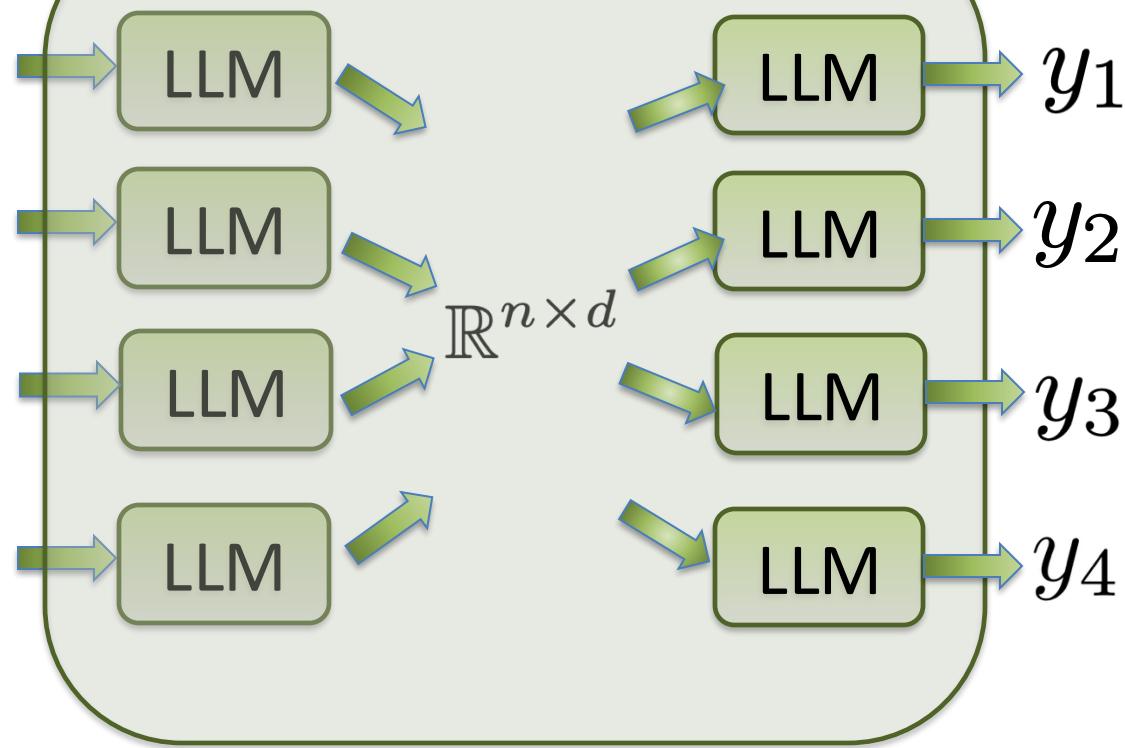
y_1	.21	.01	.002	0	.1	0	.01	.17	.11	.001
y_2	0	.01	.17	.11	.001	.21	.01	.002	0	.1
y_3	0.3	.01	.002	0	.1	.01	.17	.11	.001	.21
y_4	.01	.02	0	.1	.01	.17	.11	.01	.1	.2

$$E[i_{\text{Houston}}] + p_1$$

$$E[i_{\text{we}}] + p_2$$

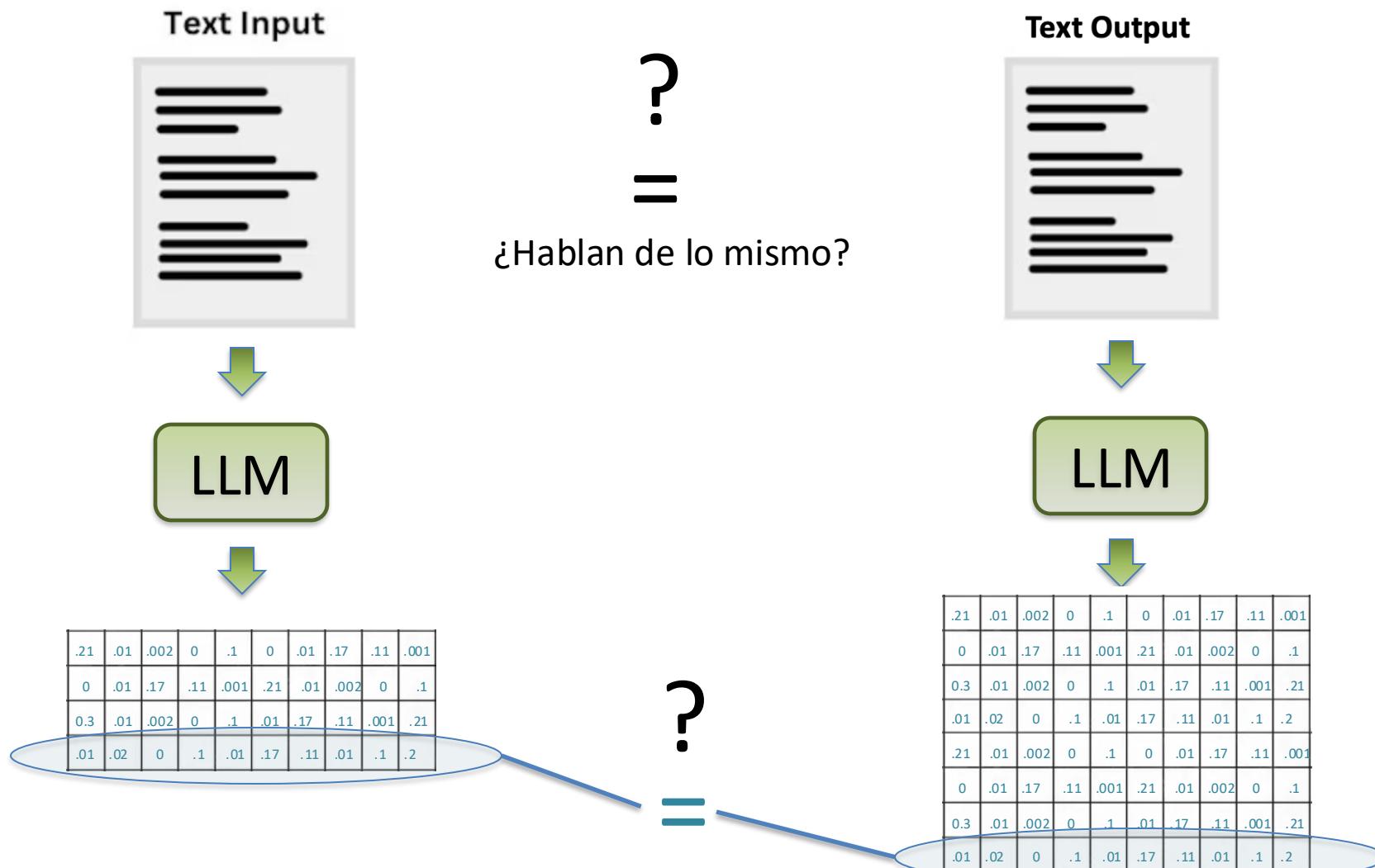
$$E[i_{\text{have}}] + p_3$$

$$E[i_a] + p_4$$



Gran aporte de los LLMs

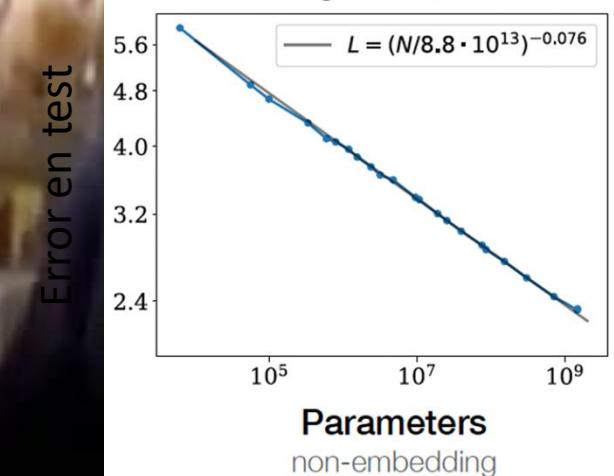
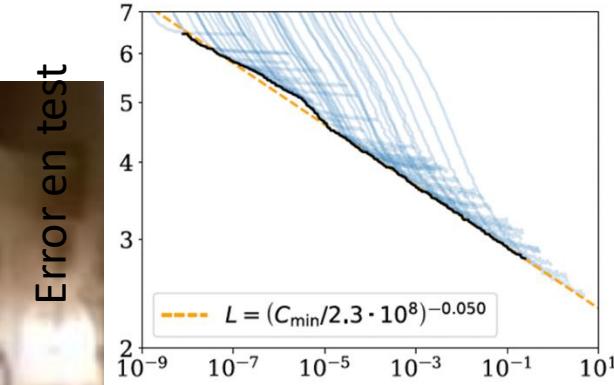
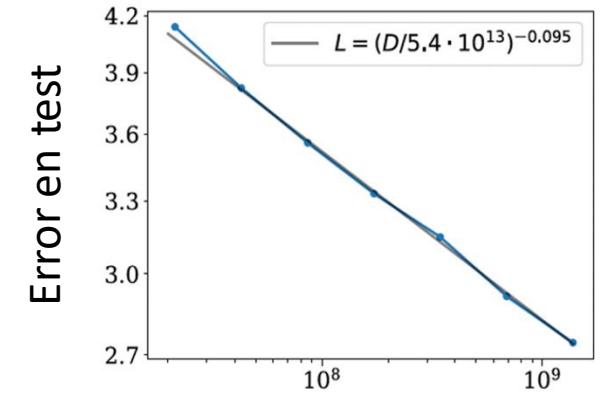
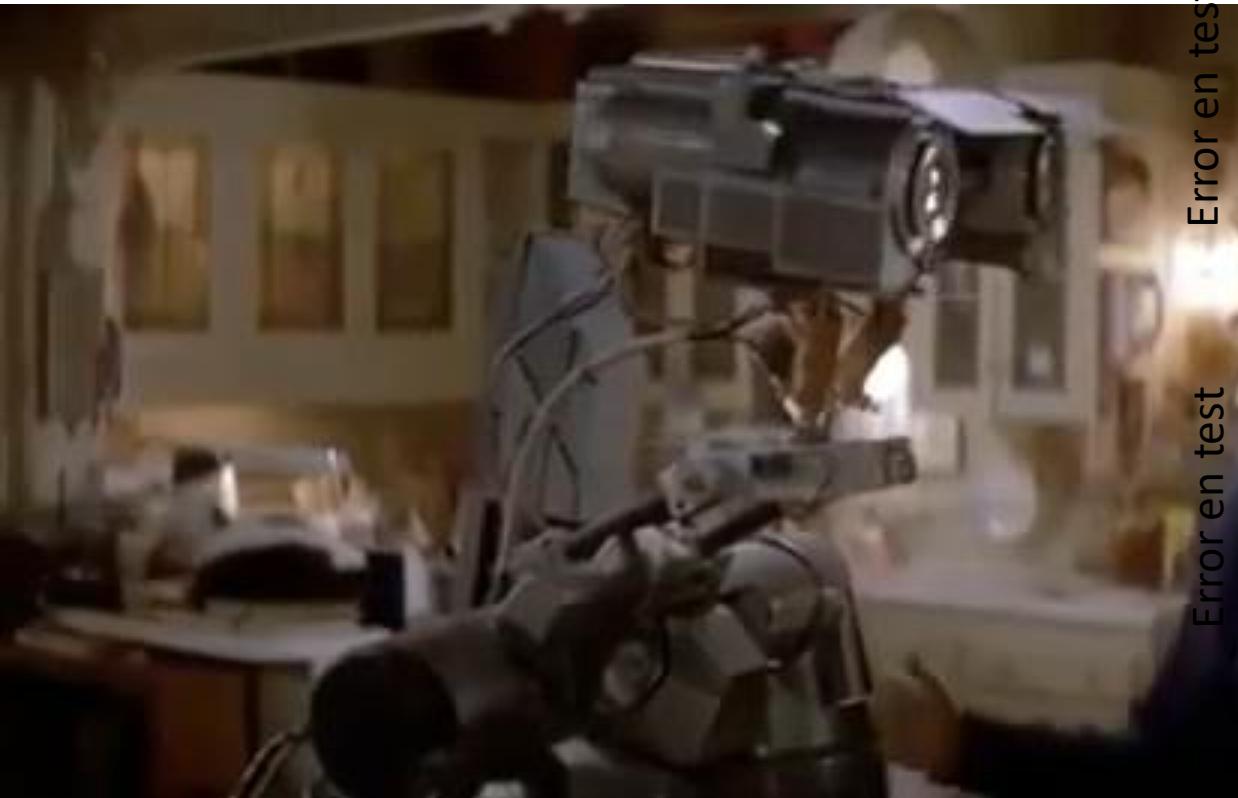
- Similaridad semántica de cualquier texto



Fuerza bruta para mejorar un LLM

- **Mas Datos**

"Datos de calidad son el nuevo petróleo refinado" Ilya Sutskever



¿Cómo especializamos un LLM?

- Partimos de un LLM entrenado y continuamos entrenando **Further Pre-Training** o pocos parámetros añadidos **Fine-Tuning**
- Lo Tuneamos con Tareas. *Supervised Fine-Tuning SFT*

Task	Datasets
Retrieval	COLIEE 2023 (2023), ELAM (2022b), LeCaRD (2021), LeCaRDv2 (2024d), ML2IR (2024)
Legal Information Extraction	InLegalNER (2024), JointExtraction (2020), LEVEN (2022)
Legal Judgement Prediction	CAIL2018 (2018), CJO22 (2023b), ECHR (2019), FSCS (2021), MultiLJP (2023)
Legal QA	CaseHold (2021), JEC-QA (2020b), LegalCQA (2024b), LegalLFQA (2024), LegalQA (2023), LLeQA (2024)
Legal Reasoning	SARA (2023), COLIEE 2022 (2023), SLJA (2023)
Legal Summarization	Claritin (2019), CLSum (2024b), IN-Abs (2022), IN-Ext (2022), Me-Too (2019), UK-Abs (2022), US-Election (2019)



¿Es importante la especialización de LLMs?

- Fundamental en el ámbito jurídico
- Puede salir a cuenta entrenar solo con textos legales
- *Fine-Tuning* produce grandes mejoras **solo cuando el modelo ya está pre-entrenado en derecho**
- La especialización jurídica (pre-training + SFT) es **superior** frente a modelos generalistas



Bringing order into the realm of Transformer-based language models for artificial intelligence and law

CM Greco, A Tagarelli

2024 59 Citations

Artificial Intelligence and Law, 2024 • Springer

Aprender sin entrenar

- In-Context Learning

Prompt: Si te digo: corazón rojo, nieve blanca.

Hierba ? verde

Language models are few-shot learners

Cited by 59247

T Brown, B Mann, N Ryder... - Advances in neural ..., 2020 - proceedings.neurips.cc

Pero cuidado...

- Fallan al aplicar reglas de conocimiento que saben definir, tienen **Potemkin Understanding**:
 - Define el sesgo psicológico llamado efecto IKEA:
 - LLM responde: Valorar más lo que uno mismo construyó.
 - ¿Puedes dar un ejemplo?
 - LLM: si claro, “le gusta su coche nuevo”

Potemkin Understanding in Large Language Models

[M Mancoridis](#), [B Weeks](#), [K Vafa](#)... - arXiv preprint arXiv ..., 2025 - arxiv.org

Cited by 9

Efectivamente

- In-Context Learning **FALLA** ChatGPT 5.1

ChatGPT 5.1 >



Prompt (ejemplo de In-Context Learning)



Vamos a usar un código secreto llamado ZAL-7.

La regla es:

- Cada palabra se transforma tomando la segunda letra y poniéndola al principio.
- Luego se añade "-zar" al final.

Ejemplos:

Original: mesa

ZAL-7: emas-zar

Original: gato

ZAL-7: agot-zar

Original: luna

ZAL-7: ulna-zar

Ahora tú:

Original: planeta

ZAL-7:

Language models are few-shot learners

Cited by 59247

[T Brown, B Mann, N Ryder... - Advances in neural ...](#), 2020 - proceedings.neurips.cc

34

Los LLMs ¿Razonan?



- Gracias a la empresa China Deepseek tenemos un artículo donde se describe cómo puede aprender un LLM a razonar.

<think> Me piden... Puedo subdividir en tareas... </think>

A conversation between User and Assistant. The user asks a question, and the Assistant solves it. The assistant first thinks about the reasoning process in the mind and then provides the user with the answer. The reasoning process and answer are enclosed within `<think> </think>` and `<answer> </answer>` tags, respectively, i.e., `<think>` reasoning process here `</think>` `<answer>` answer here `</answer>`. User: **prompt**. Assistant:

Table 1 | Template for DeepSeek-R1-Zero. **prompt** will be replaced with the specific reasoning question during training.

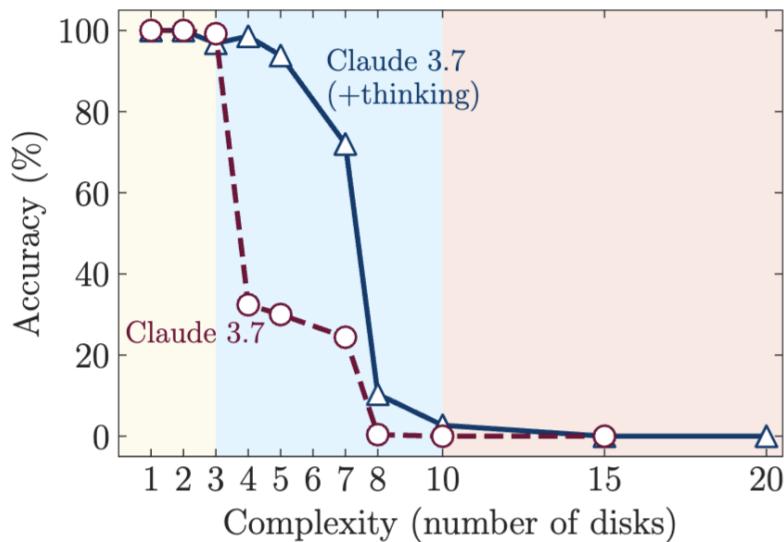
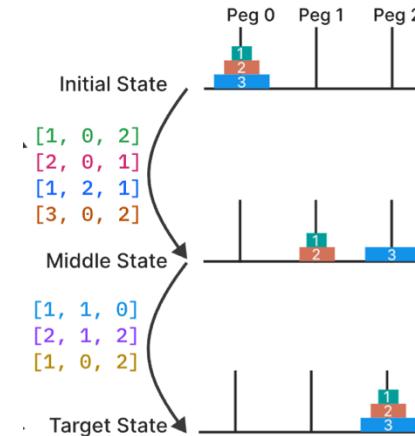
Deepseek-r1: Incentivizing reasoning capability in llms [PDF] arxiv.org
via reinforcement learning

D Guo, D Yang, H Zhang, J Song, R Zhang... - arXiv preprint arXiv ..., 2025 - arxiv.org Cited by 5373

Limites de los LLMs que razonan

- Illusion of Thinking

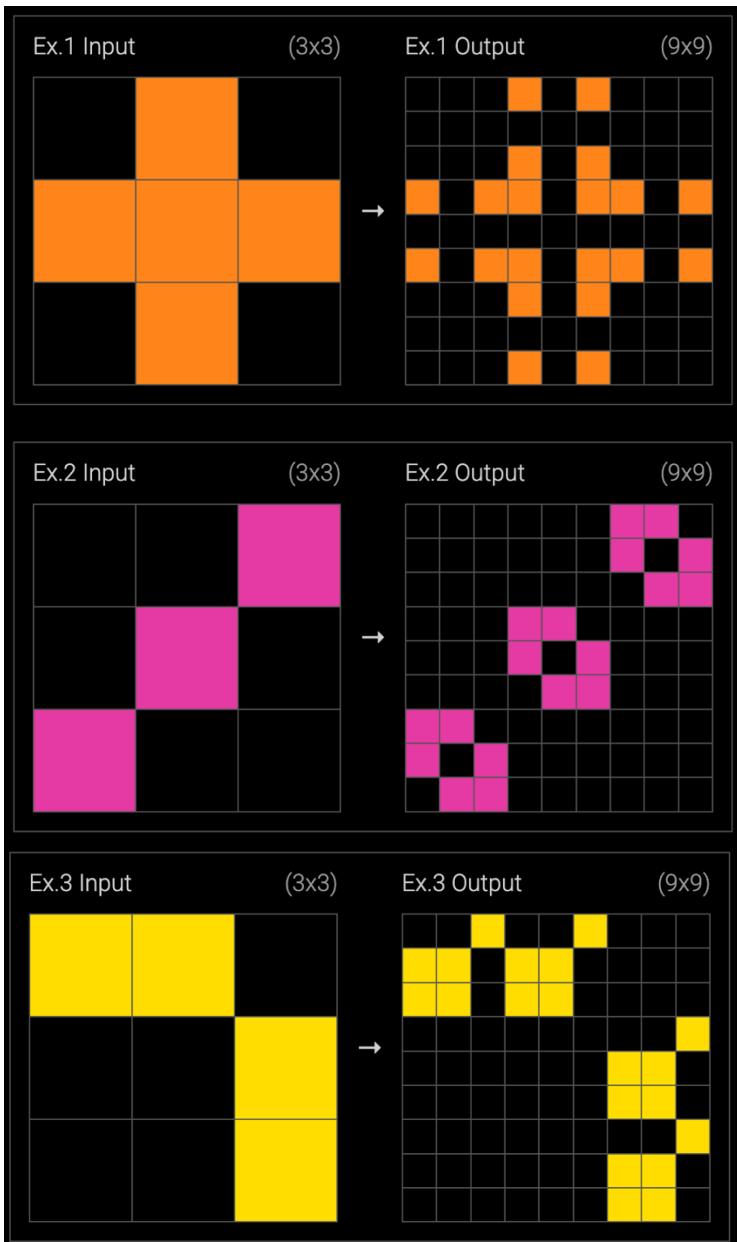
Torres de Hanoi



The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity

P Shojaee, I Mirzadeh, K Alizadeh, M Horton... - arXiv preprint arXiv ..., 2025 - arxiv.org Cited by 221

Los LLMs ¿razonan realmente?

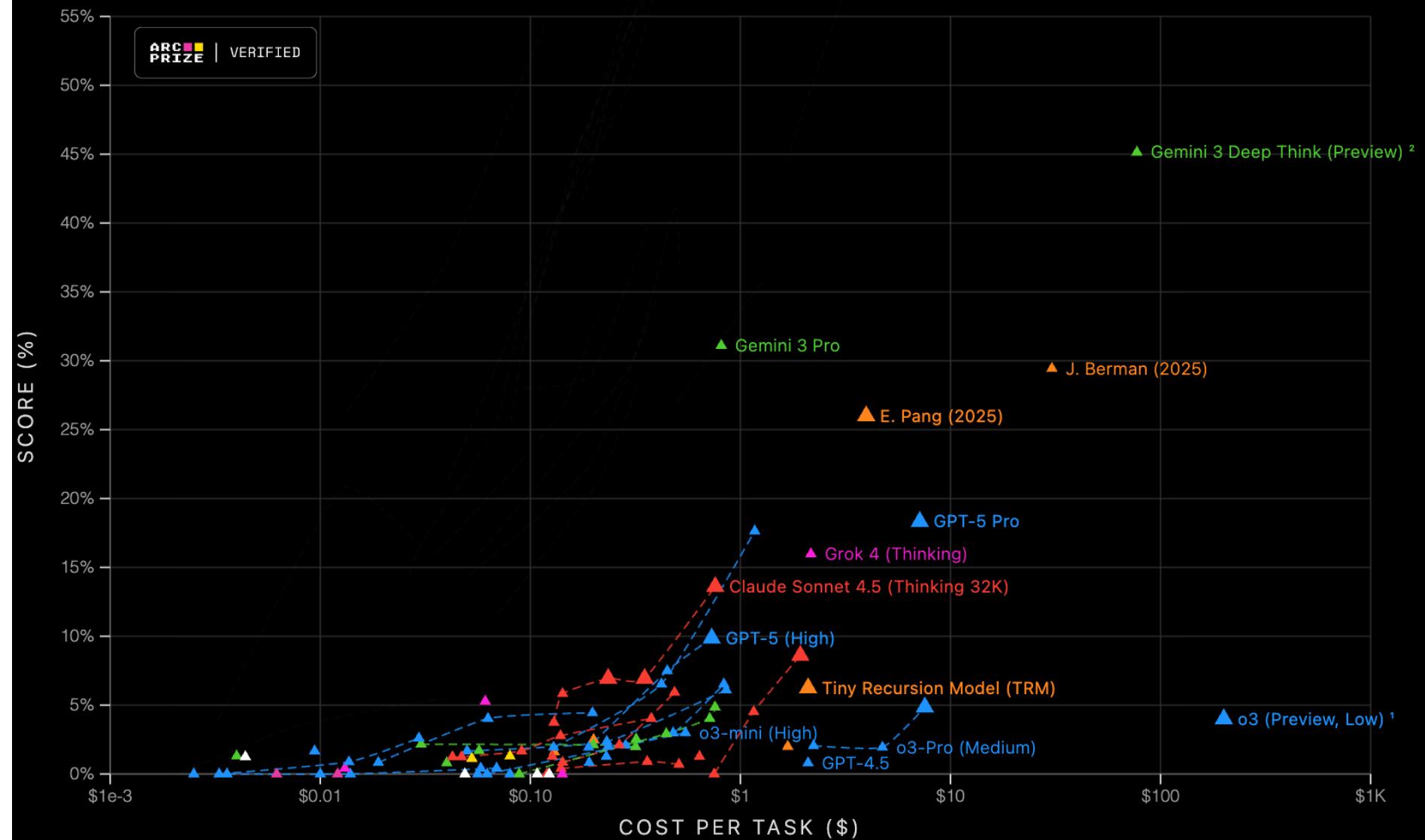


[PDF] [On the measure of intelligence](#)

Cited by 1105
F Chollet - arXiv preprint arXiv:1911.01547, 2019 - ms456000.wpcosstaging.com

Los LLMs ¿razonan realmente?

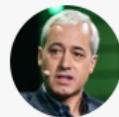
ARC-AGI-2 LEADERBOARD



Conclusiones



-¡Objeción, mi señoría!
-Alucinación sostenida.



Pedro Domingos @pmddomingos · Oct 11



It's OK if people get worse at the things AI can do. What matters is whether people are better with AI than without.

11

7

61

10K



Conclusiones

- ¿Entienden los LLMs?
- ¿Razonan los LLMs?
- ¿Qué limitaciones tienen los LLMs?
- ¿Nos pueden ayudar *en el ámbito jurídico*?

Gracias

PATROCINA



Universitat de Lleida
Facultat de Dret, Economia
i Turisme

Proyectos de «Generación de
Conocimiento»

2023

Proyecto *La polarización política
y su impacto sobre la realidad
constitucional (PORECO)*
financiado por MICIU/AEI
/10.13039/501100011033 y por
FEDER, UE

Proyecto *PID2023-153037-NB-100* financiado por:



Asociación de
Constitucionalistas
de España



Centro de
Estudios
en Ciencias
Constitucionales



Extra

LLMs son una arquitectura Transformer

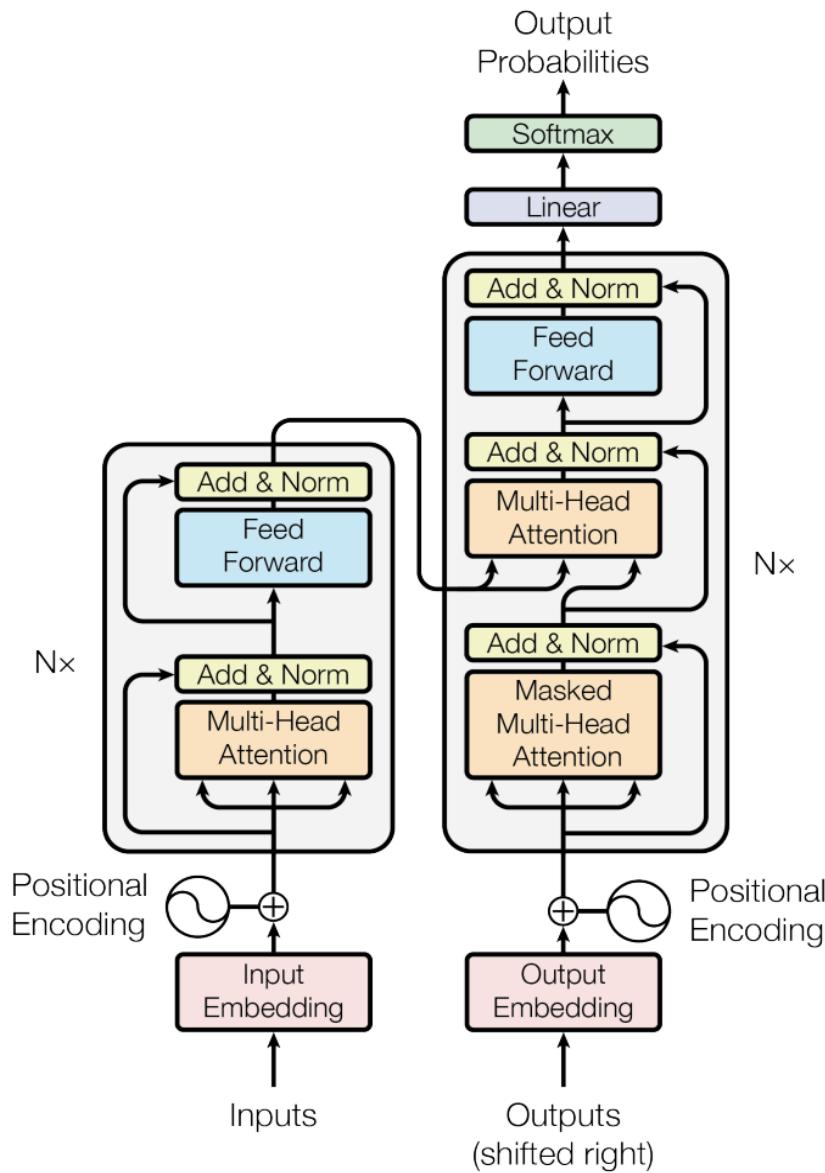
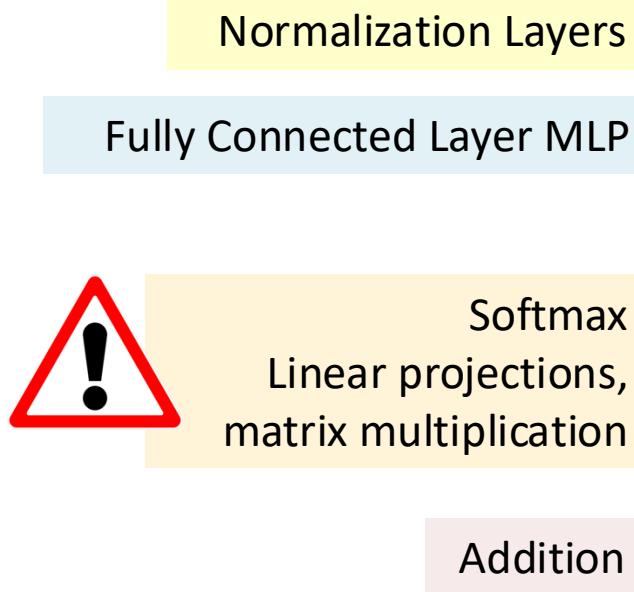
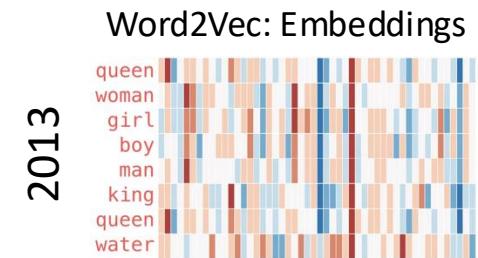


Figure 1: The Transformer - model architecture.



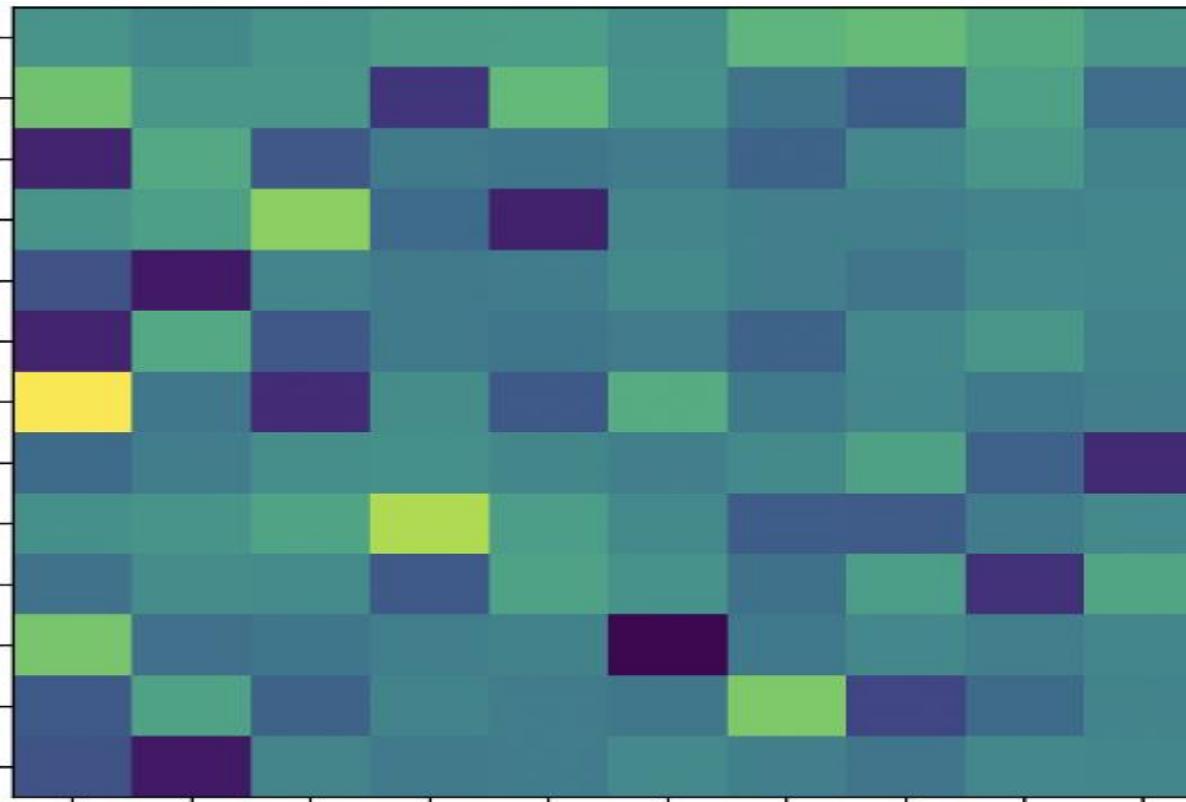
From Convnets to Transformers

- How we represent words and phrases?



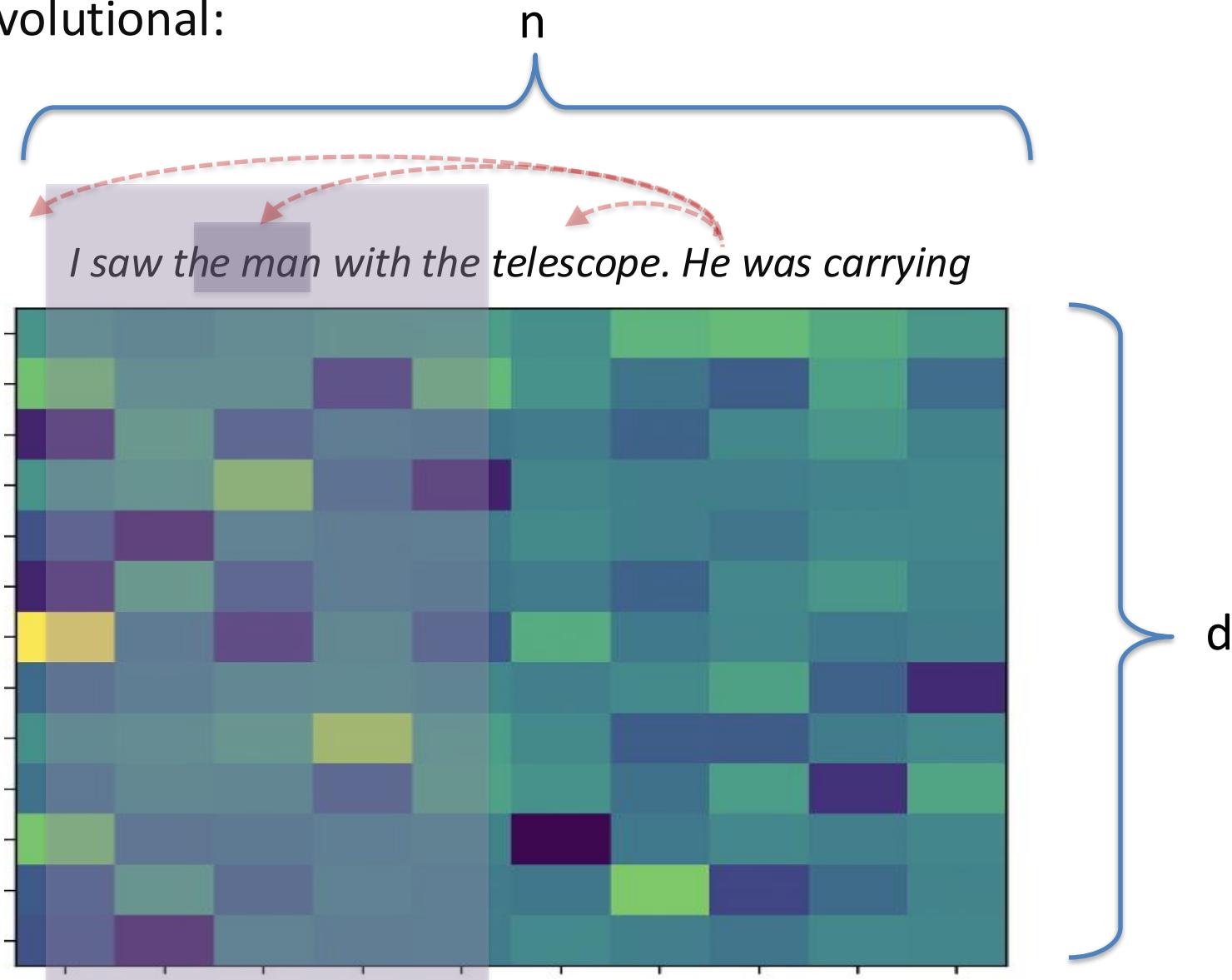
2013

I saw the man with the telescope. He was carrying



Attention Layer

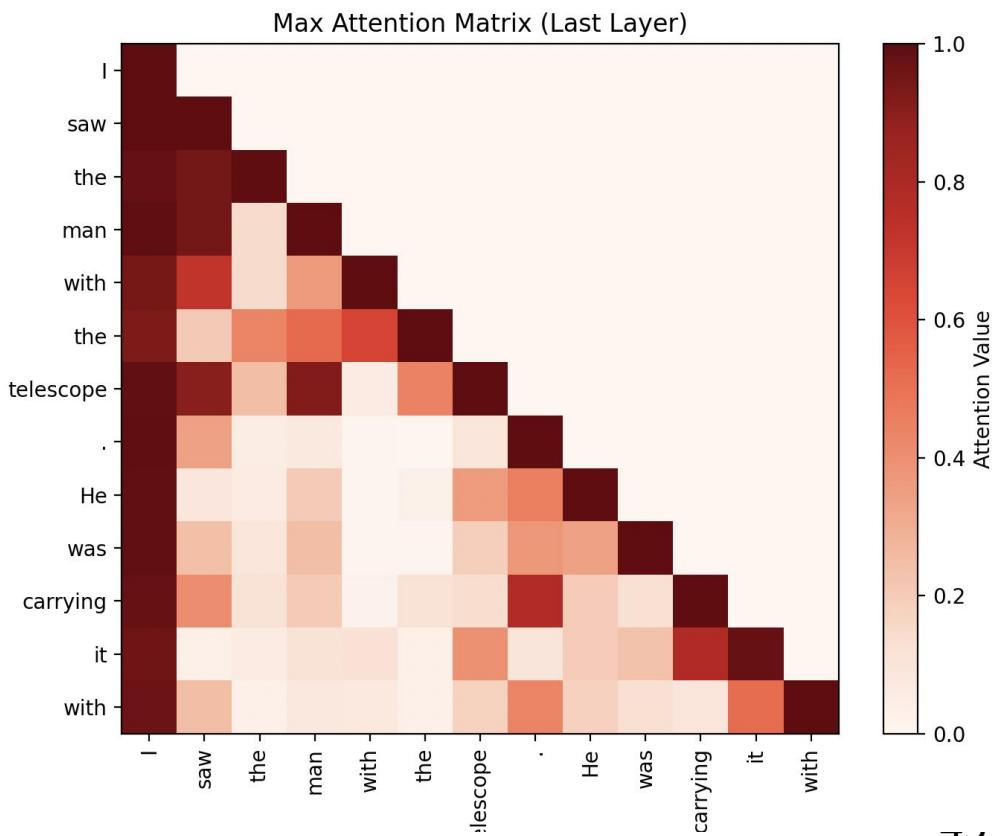
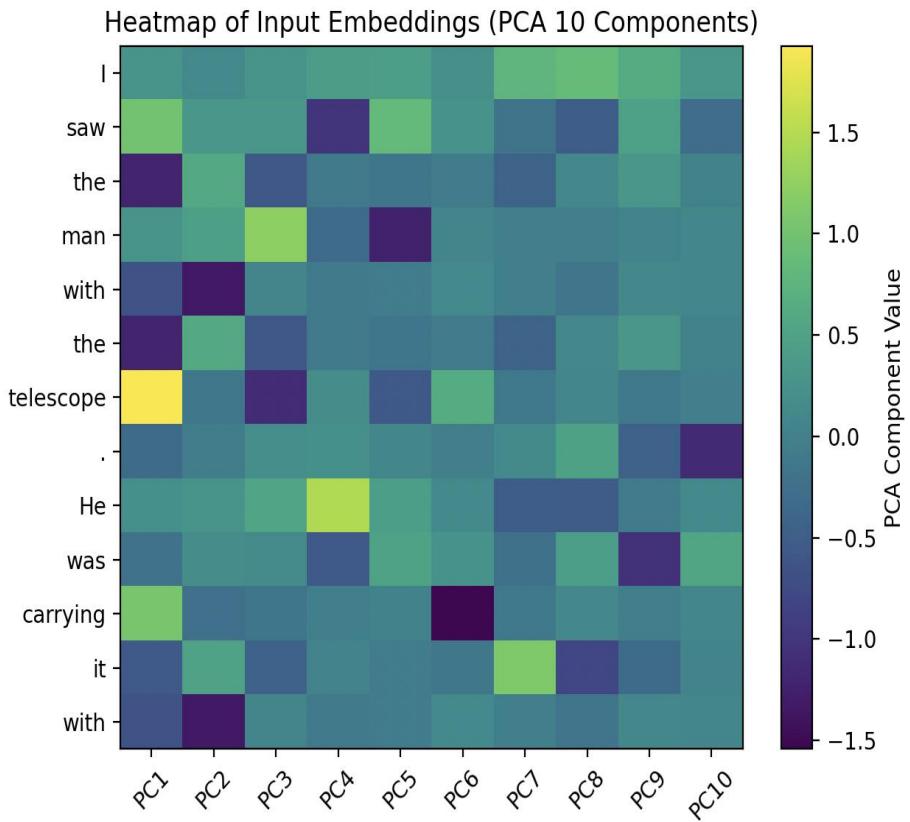
- In Between Fully Connected
and Convolutional:



Attention Layer

- A **scoring function** of all pairs of input tokens

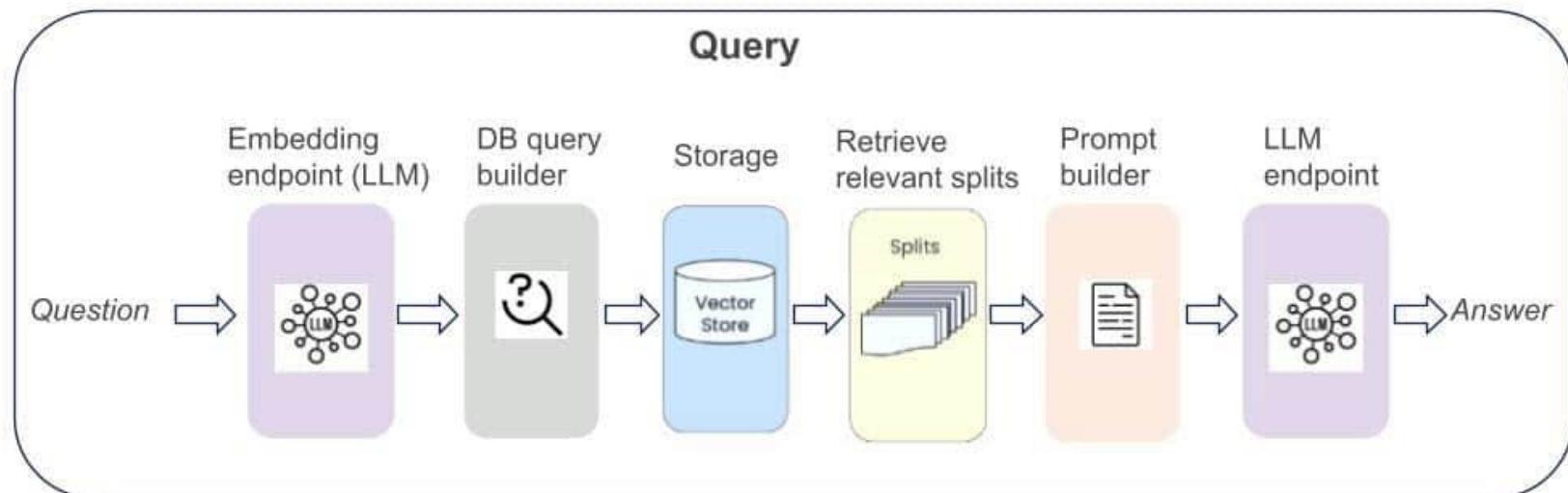
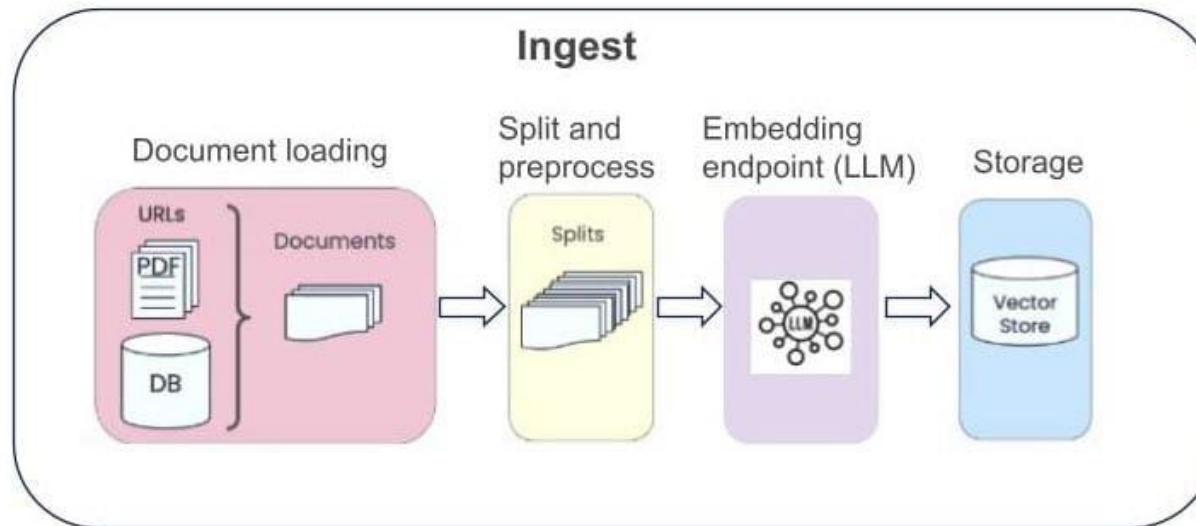
I saw the man with the telescope. He was carrying



Aprender sin entrenar



- Retrieval Augmented Generation (RAG)



¿Cómo entrenar a un AI-bogado?



- Tareas para *Supervised Fine-Tuning*:

Clasificación: La capacidad de predecir etiquetas para textos jurídicos, como determinar el cargo relevante en un caso (predicción de sentencias legales) o clasificar una cláusula específica en un contrato.

Razonamiento: La capacidad de realizar inferencias lógicas, como determinar si una conclusión jurídica se desprende de un conjunto de hechos y normas.

Extracción de información: La habilidad de extraer información específica y estructurada de texto jurídico no estructurado, incluyendo el reconocimiento de entidades y la elaboración de resúmenes documentales.

Recuperación: La tarea de identificar y recuperar documentos jurídicos relevantes, como casos similares o normas aplicables, a partir de una gran base de datos.

Pregunta-respuesta: La capacidad de generar respuestas precisas y pertinentes a preguntas jurídicas, a menudo fundamentadas en documentos fuente proporcionados.

Evaluación del conocimiento: La capacidad de recordar y recitar con precisión conocimientos jurídicos fundamentales, incluidos conceptos, definiciones y artículos legales.

Aritmética: La habilidad de realizar cálculos basados en la información presente en textos jurídicos, como determinar plazos o calcular importes monetarios relacionados con un delito.

[Large Language Models Meet Legal Artificial Intelligence: A Survey](#)

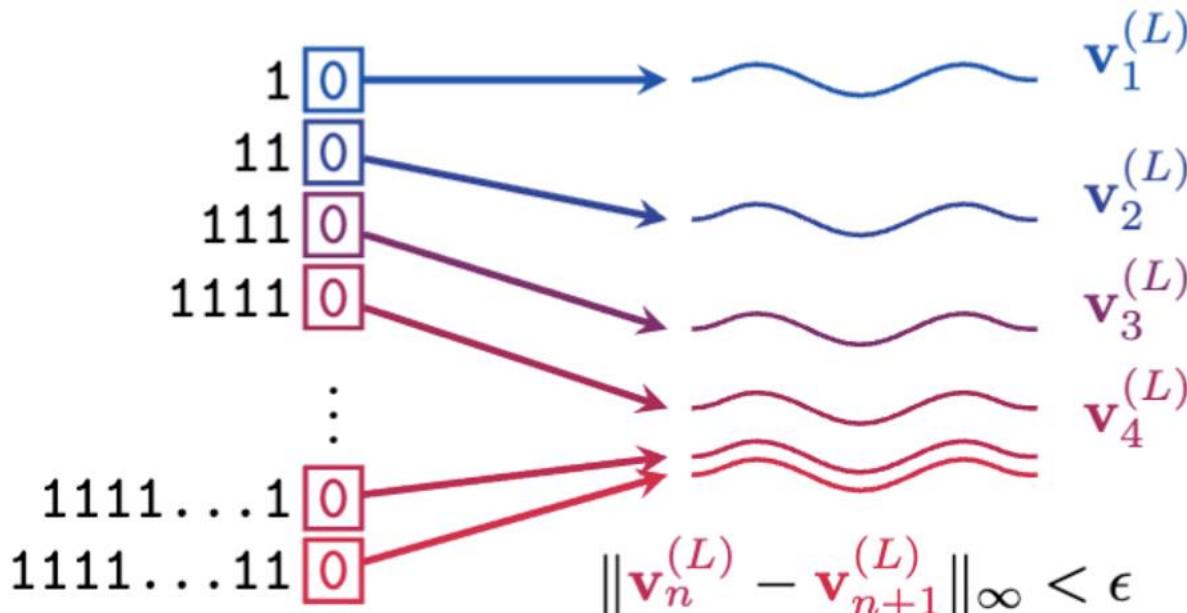
[Z Hou, Z Ye, N Zeng, T Hao, K Zeng](#)

arXiv preprint arXiv:2509.09969, 2025 • arxiv.org

Problemas de los LLMs



- Limites de precisión de su representación



(a) Representational Collapse

Transformers need glasses! information over-squashing in language tasks

F Barbero, A Banino, S Kapturowski, D Kumaran, J Madeira Araújo, O Vitvitskyi, R Pascanu...

Advances in Neural Information Processing Systems, 2024 • proceedings.neurips.cc

Cited by 43

Problemas de los LLMs

- **Alucinaciones**

De conocimiento, de hechos:

- Conocimiento que ha cambiado
- Poco frecuente, Fuera de distribución en el Dataset de entrenamiento,
- Reversos de cadenas deductivas triviales

De Razonamiento:

- Tareas Largas que requieren largas secuencias de razonamientos

Problemas de los LLMs

- Carecen de Explicabilidad: ***No Explainability***
- Poca auto-conocimiento de lo que saben y no saben
- Pésimas habilidades comparativas de fechas y horas
- No razonan, Predicen lo que quieres oír

Razonamiento en los LLMs: Deepseek R1



DeepSeek-R1-Zero average length per response during training

