AI Tools Based on Pre-Trained Language Models

Miguel Domingo

midobal@prhlt.upv.es

Departamento de Sistemas Informáticos y Computación
Universitat Politècnica de València

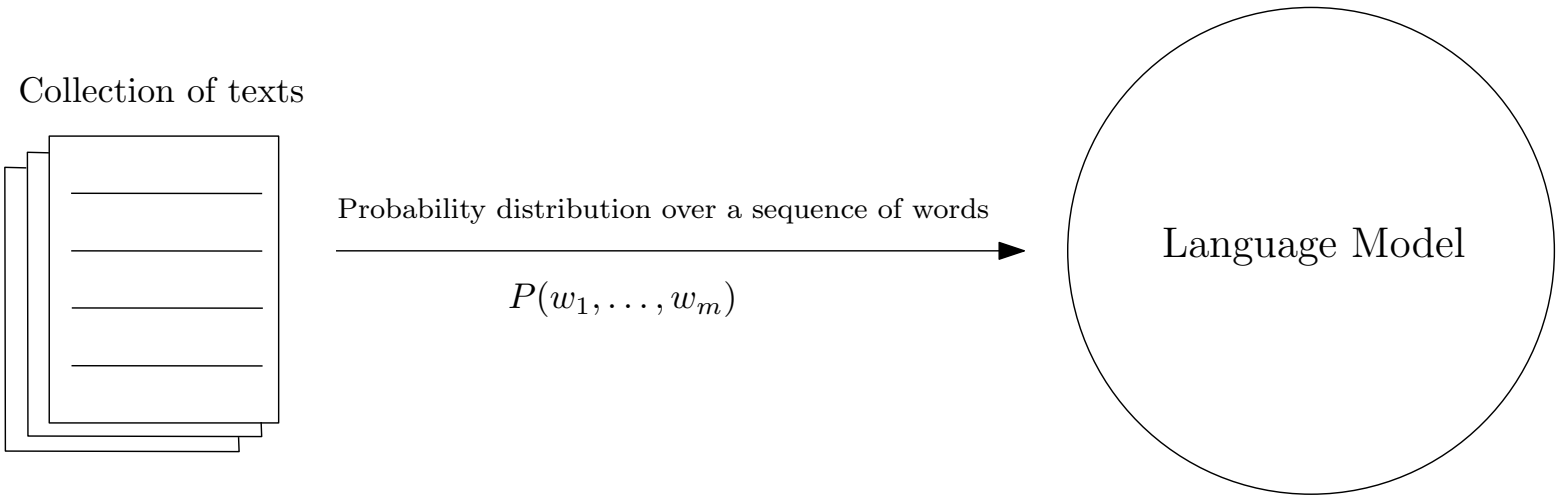
Tools and Applications of Artificial Intelligence

MIARFID, February 27, 2024

Outline

Outline

What is a Language Model?



What is a Language Model?

N-Gram

A language model that models sequences of words as a Markov process.

$$P(w_1, \ldots, w_m) \approx \prod_{i=2}^{m} P(w_i | w_{i-(n-1)}, \ldots, w_{i-1})$$

$$P(w_i | w_{i-(n-1)}, \ldots, w_{i-1}) = \frac{\text{count}(w_{i-(n-1)}, \ldots, w_{i-1}, w_i)}{\text{count}(w_{i-(n-1)}, \ldots, w_{i-1})} + \text{smoothing}$$

What is a Language Model?

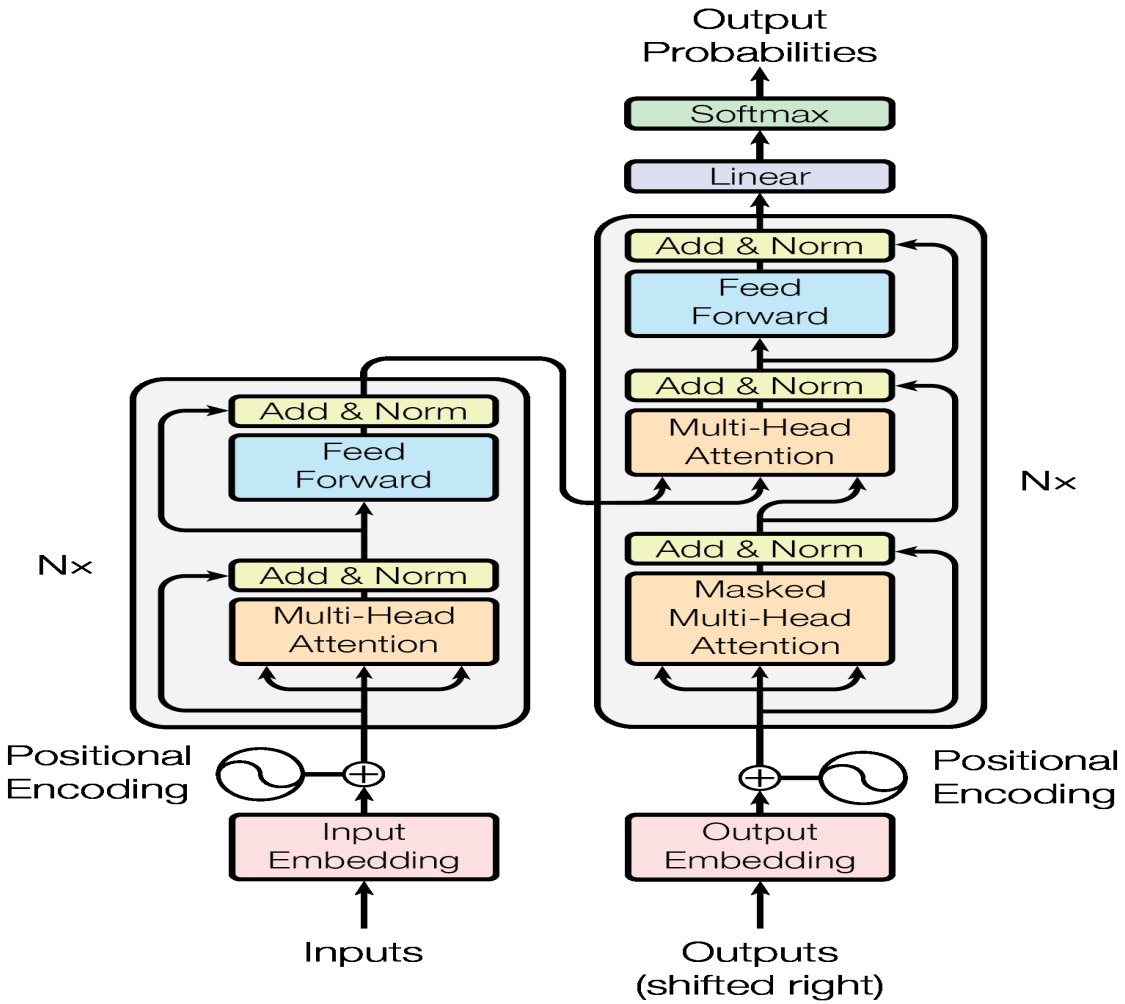
Neural Network

A language model that makes predictions using a continuous representation or embeddings of words.

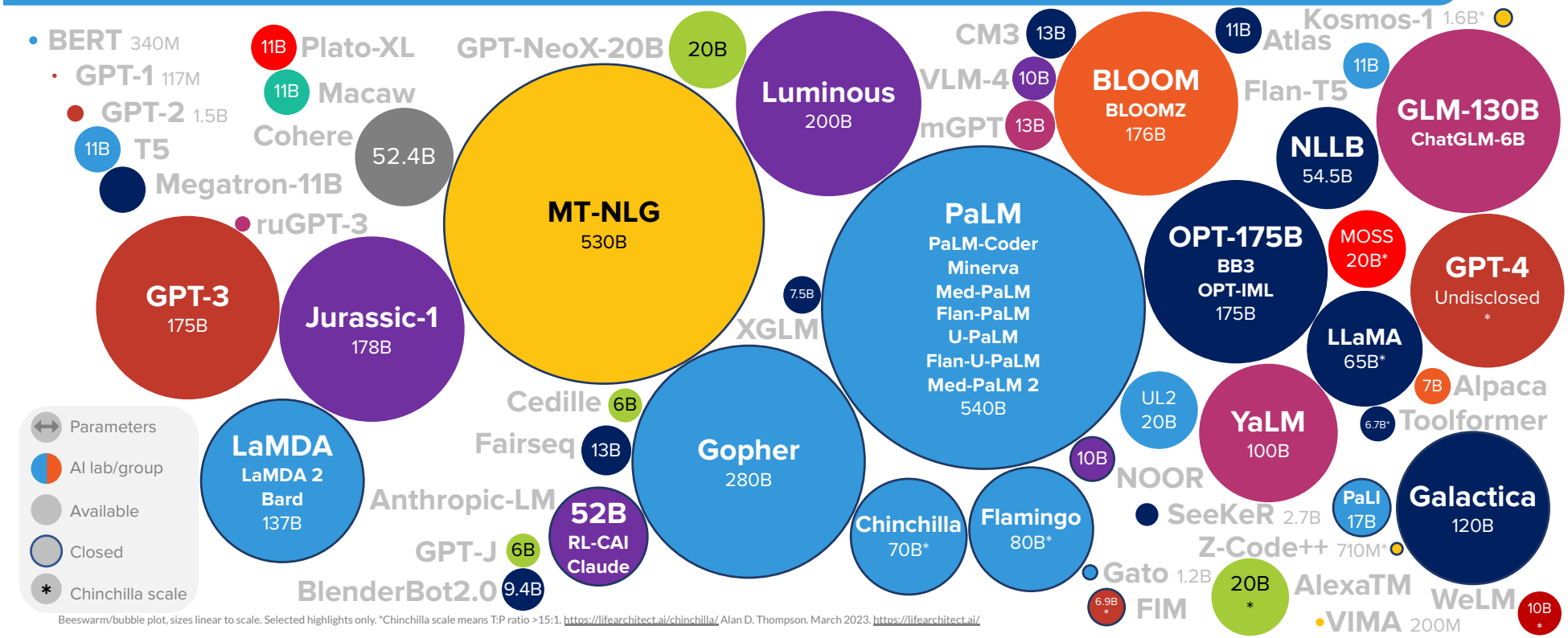
$$P(w_t|\text{context})\forall t \in V$$

Architecture

Transformer (Vaswani 2017)

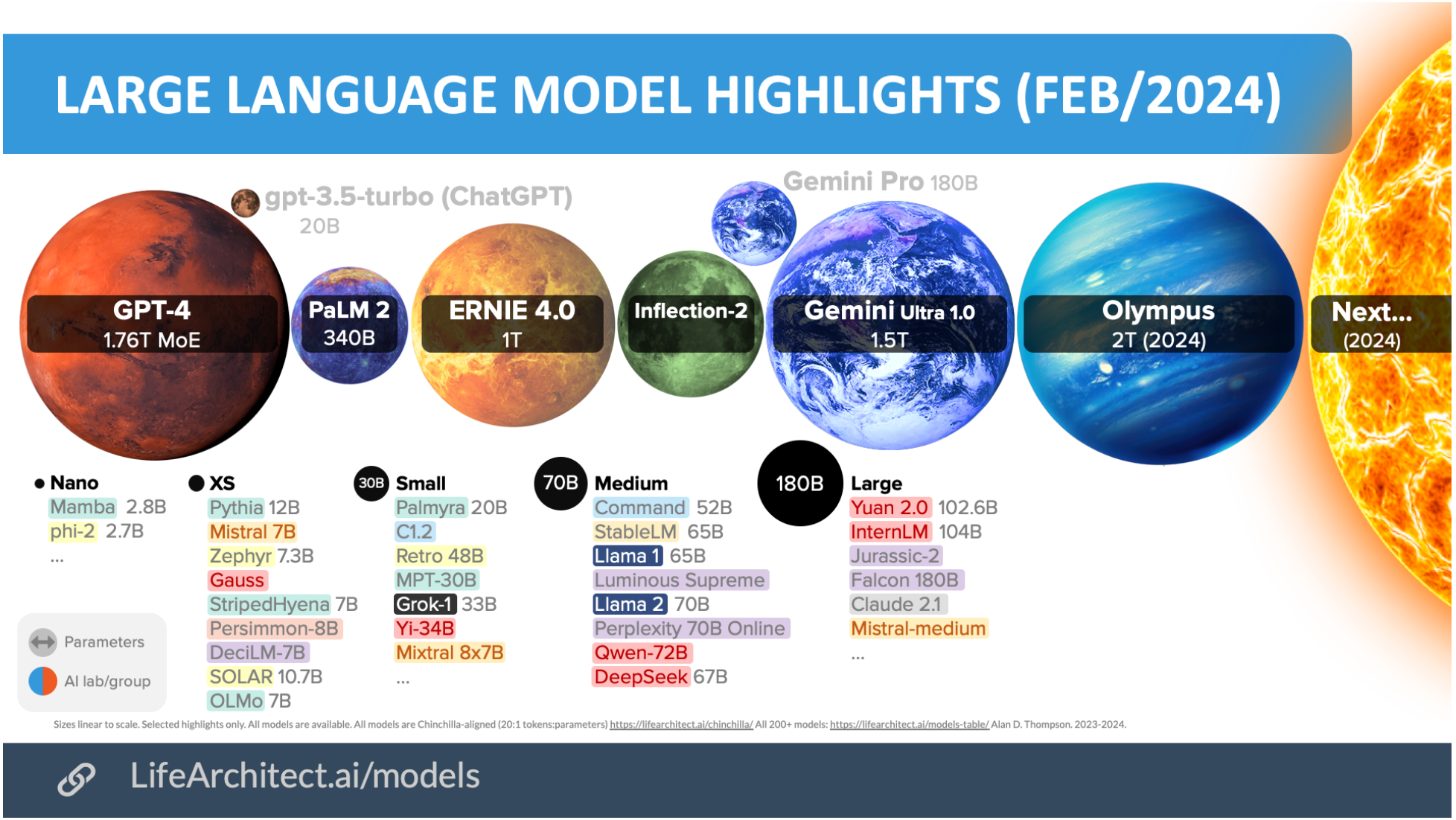


Models



LifeArchitect.ai/models

Models


LARGE LANGUAGE MODEL HIGHLIGHTS (FEB/2024)
gpt-3.5-turbo (ChatGPT) 20B
Gemini Pro 180B
GPT-4 1.76T MoE
PaLM 2 340B
ERNIE 4.0 1T
Inflection-2
Gemini Ultra 1.0 1.5T
Olympus 2T (2024)
Next... (2024)
Nano
Mamba 2.8B
phi-2 2.7B
...
XS
Pythia 12B
Mistral 7B
Zephyr 7.3B
Gauss
StripedHyena 7B
Persimmon-8B
DeciLM-7B
SOLAR 10.7B
OLMo 7B
30B Small
Palmyra 20B
C1.2
Retro 48B
MPT-30B
Grok-1 33B
Yi-34B
Mixtral 8x7B
...
70B Medium
Command 52B
StableLM 65B
Llama 1 65B
Luminous Supreme
Llama 2 70B
Perplexity 70B Online
Qwen-72B
DeepSeek 67B
180B Large
Yuan 2.0 102.6B
InternLM 104B
Jurassic-2
Falcon 180B
Claude 2.1
Mistral-medium
...
Parameters
AI lab/group
Sizes linear to scale. Selected highlights only. All models are available. All models are Chinchilla-aligned (20:1 tokens:parameters) https://lifearchitect.ai/chinchilla/ All 200+ models: https://lifearchitect.ai/models-table/ Alan D. Thompson. 2023-2024.
LifeArchitect.ai/models

Outline

Programming

Goal

Assist in the generation of source code while programming.

Input: Write a function that receives a person's name and says hi to them.
Output:

```
def hi(person):
    print(f'Hi, {person}!')
```

Programming

Applications

- GitHub Copilot.
- Chatbots.

We will see more about them at the next units.

Image Generation

Goal

React to a text input (known as *prompt*) by generating new related images.

Image Generation

Applications

- Stable Diffusion.
- Dall-e.
- Midjourney.
- Adobe Firefly.
- DreamFusion.
- Leonardo AI.
- Sora.

We will see more about them at the next units.

Text Generation

Goal

React to a text input (known as *prompt*) by generating new related text. The most trendy applications are *chat bots*.

Text Generation

Applications

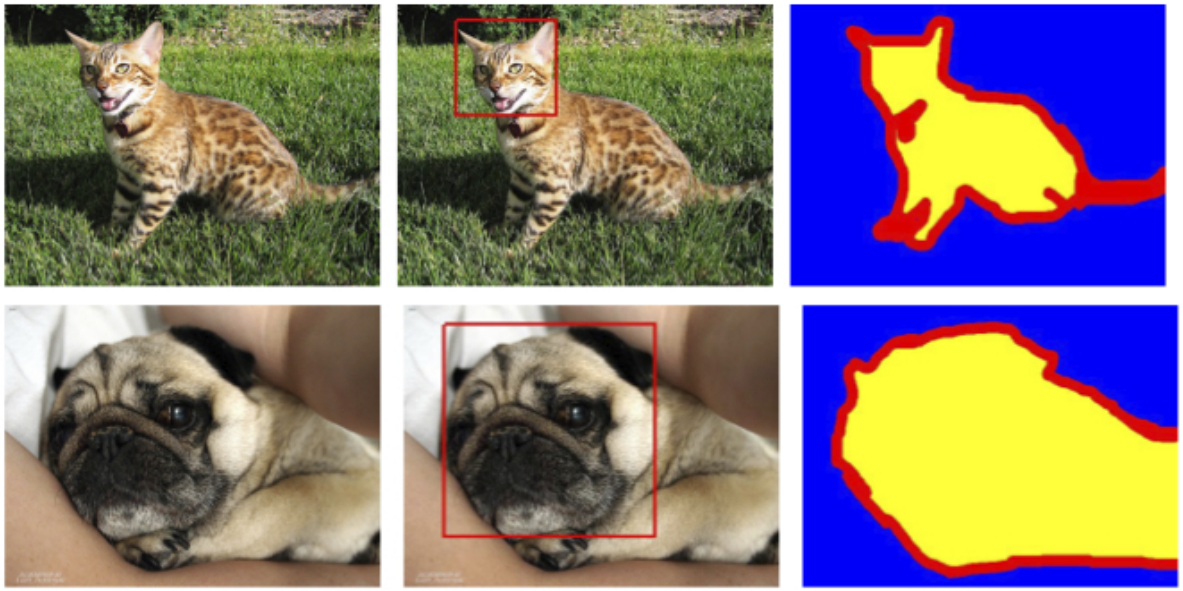
- ChatGPT.
- BingChat.
- ERNIE Bot.
- StableLM.
- AutoGPT.
- Bard/Gemini.
- Copilot.

We will see more about them at the next units.

Image Segmentation

Goal

Detect objects on an image.



Example from *Cats and Dogs* (Parkhi et al., 2012).

Image Segmentation

Applications

- Segment Anything Model.

Segment Anything Model (SAM)

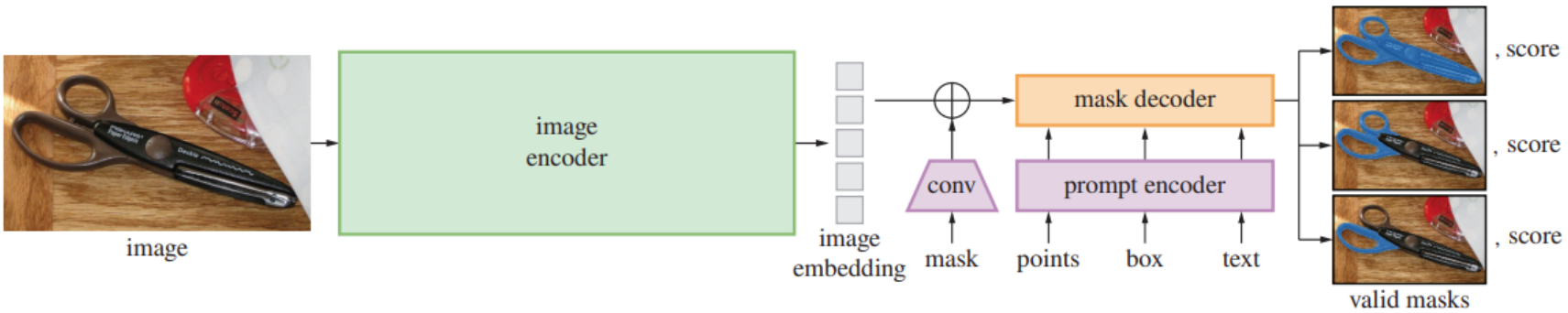
Goal

SAM (Krillov, 2023) is a promptable segmentation system with **zero-shot** generalization to unfamiliar objects and images, without the need for additional training.

Segment Anything Model (SAM)

Model

- Image encoder: Vision Transformer (ViT) model (Dosovitskiy et al., 2020).
- Prompt encoder.
- Mask decoder.
- Resolving ambiguity: average of multiple valid masks.
- Efficiency: 50ms (CPU).



Segment Anything Model (SAM)

Demo

https://segment-anything.com/demo

Mathematical Expressions

Goal

Understand mathematical expressions, solve problems, prove theorems, etc.

$$4x + 26x + 12 = 0 \longrightarrow x = -6, x = -\frac{1}{2}$$

Mathematical Expressions

Applications

- Minerva.

Minerva

Goal

Solving quantitative reasoning problems (Kirillov et al., 2022).

Question: A line parallel to $y = 4x + 6$ passes through $(5, 10)$. What is the y-coordinate of the point where this line crosses the y-axis?

Model output: The line is parallel to $y = 4x+6$, which means that it has the same slope as $y = 4x+6$. Since the line passes through $(5, 10)$, we can use the point-slope form of a line to find the equation of the line:

$$\begin{aligned} y - 10 &= 4(x - 5) \\ y - 10 &= 4x - 20 \\ y &= 4x - 10 \end{aligned}$$

Now that we have the equation of the line, we can find the y-coordinate of the point where the line crosses the y-axis by substituting $x = 0$ into the equation:

$$y = 4 \cdot 0 - 10 = \boxed{-10}.$$

Minerva

Features

- Based on PaLM (Chowdhery et al., 2022).
- Focused on collecting relevant training data.
- Use of *best-in-class* inference techniques.
- Mixture of natural language and mathematical notation.
- Combines:
  - Few-shot prompting.
  - Scratchpad prompting.
  - Majority voting.

Minerva

Weaknesses and limitations

Question: If $\sqrt{400} = \sqrt{81} + \sqrt{n}$, then what is the value of n?

Model output: We have that $\sqrt{400} = \sqrt{81} + \sqrt{n}$. Therefore, $400 = 81 + n$, which gives $319 = n$.
Final Answer: The final answer is 319.

Calculation mistake: The model incorrectly cancels the square root on both sides of the equation.

Minerva

Weaknesses and limitations

Question: At each basketball practice last week, Jenny made twice as many free throws as she made at the previous practice. At her fifth practice she made 48 free throws. How many free throws did she make at the first practice?

Model output: At the first practice, Jenny made $\frac{48}{2} = 24$ free throws. Final Answer: The final answer is 24.

Reasoning mistake: The model computes the number of free throws at the fourth practice, but then uses this number as the final answer for the first practice.

Minerva

Weaknesses and limitations

Main limitation: automatic verification of the answer and the intermediate steps.

Minerva

Demo

https://minerva-demo.github.io/#category=Algebra&index=1

Speech Recognition

Goal

Transcription of fragments of speech.

Speech Recognition

Applications

- Whisper.

Whisper

Goal

Speech processing system (Radford et al., 2022) trained for:

- English transcription.
- Any-to-English speech translation.
- Non-English transcription.
- No speech detection.
- Language detection.

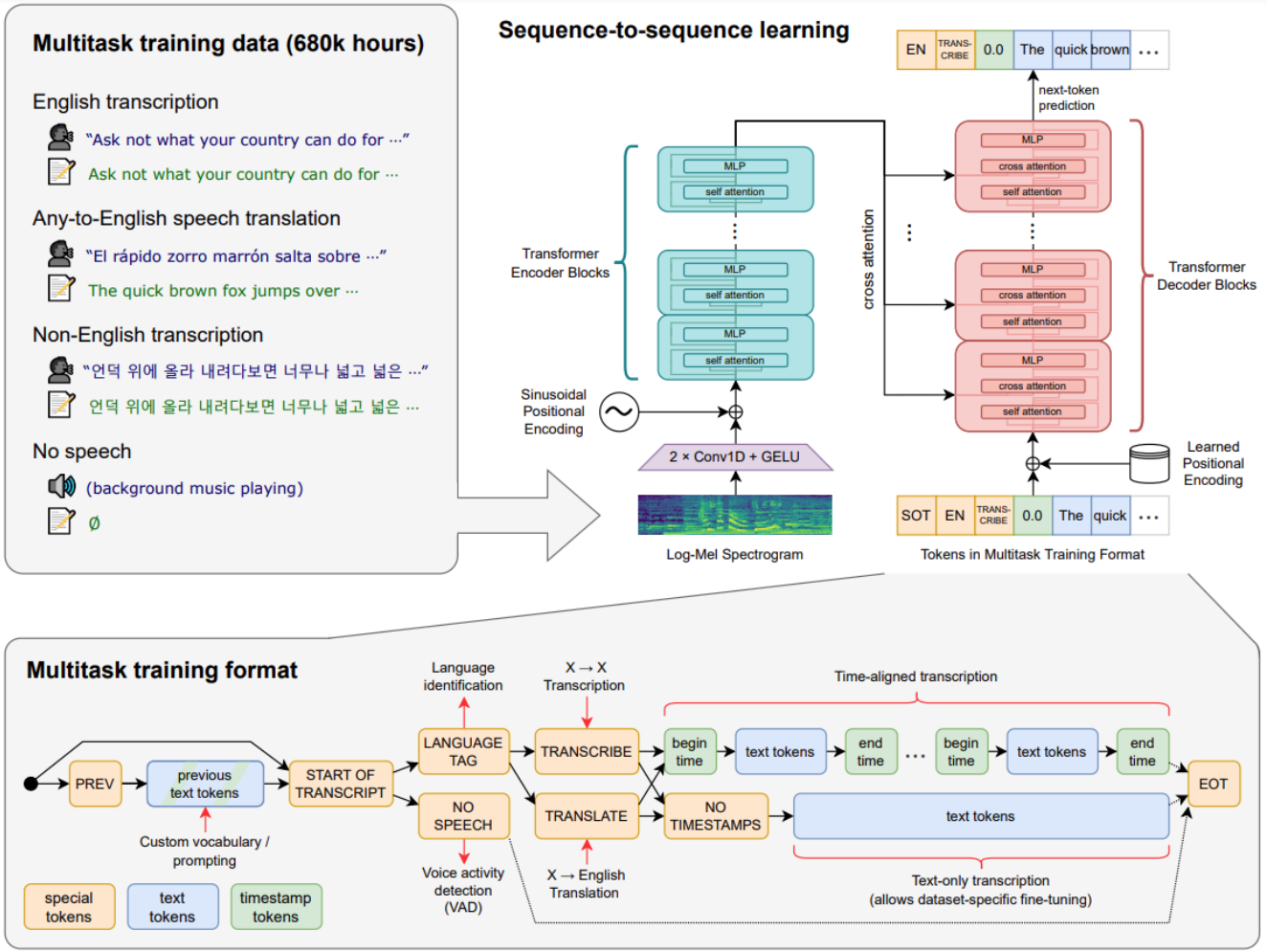
Whisper

Features

- Automatic speech recognition (ASR) system.
- Trained on 680,000 hours of multilingual and multitask supervised data.
- Data quality.
- Basic Transformer architecture.

Whisper

Architecture



Whisper

Evaluation

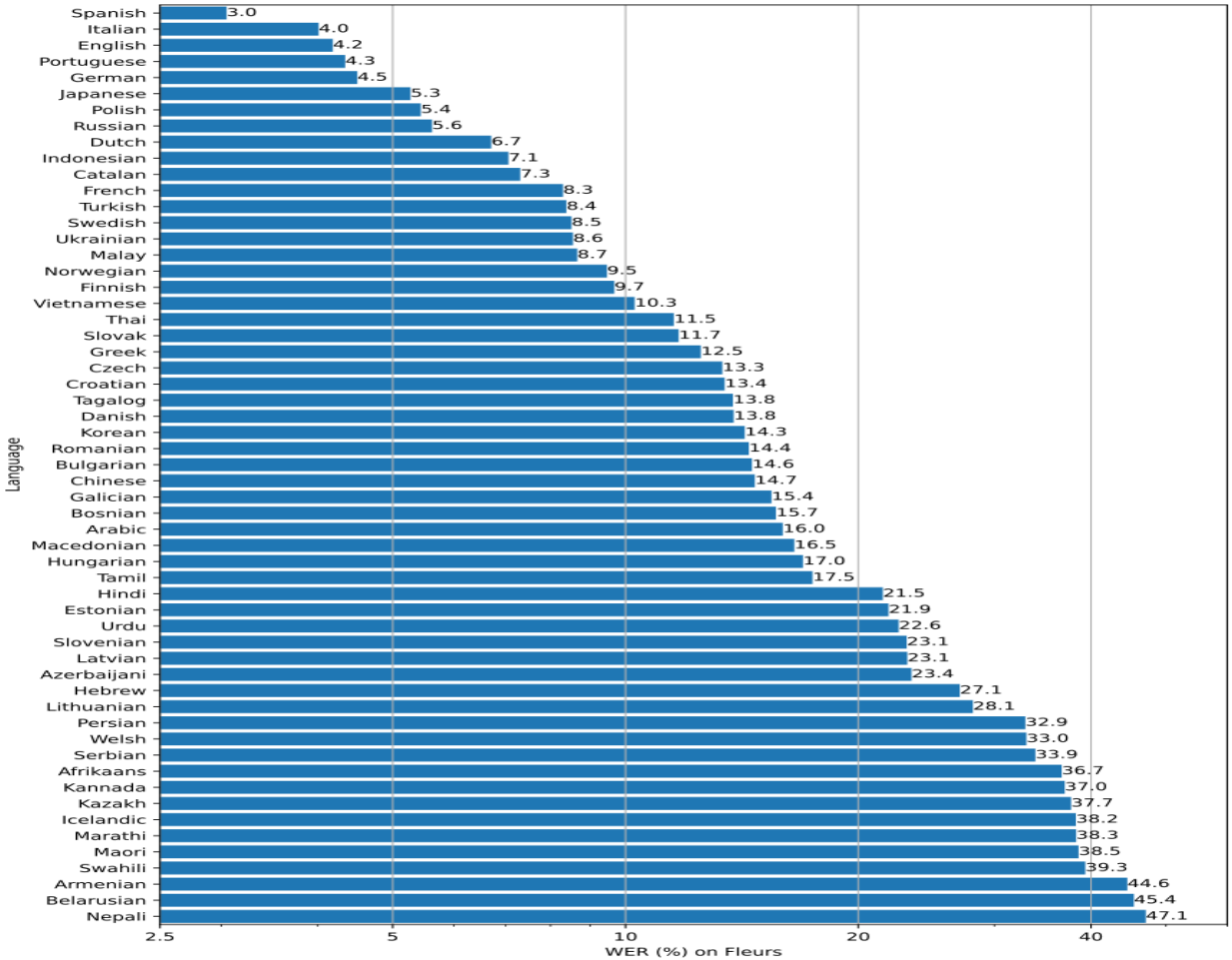


Figure from Whisper's GitHub.

Whisper

Demo

- Original demo: `https://openai.com/research/whisper`.
- Google Colab: `https://colab.research.google.com/drive/1CvvYPAFemIZdSOt9fhN541esSlZR7Ic6`.

Music Generation

Applications

- Suno.

Suno

Features

- Song generation system (tune and lyrics) from a prompt[1].
- Developed by a team of musicians and AI experts from Cambridge.
- Integrated into Microsoft's Copilot.
- Artist protection.
- Monetization.
  - 5 free prompts per day.
  - Billing system for various prompts per month.

[1] `https://www.suno.ai/`.

Suno

Demo

Make a song: `https://app.suno.ai/`.

Bibliography

- Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, ..., & Fiedel, N. (2022). **Palm: Scaling language modeling with pathways**. *arXiv preprint arXiv:2204.02311*.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. & Houlsby, N. (2020). **An image is worth 16x16 words: Transformers for image recognition at scale**. *arXiv preprint arXiv:2010.11929*.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, Y., Dollár, P. & Girshick, R. (2023). **Segment anything**. *arXiv preprint arXiv:2304.02643*.
- Lewkowycz, A., Andreassen, A., Dohan, D., Dyer, E., Michalewski, H., Ramasesh, V., Slone, A., Anil, C., Schlag, I., Gutman-Solo, T., Wu, Y., Neyshabur, B., Gur-Ari, G., & Misra, V. (2022). **Solving quantitative reasoning problems with language models**. In *Advances in Neural Information Processing Systems*, 3843–3857.

Bibliography

- Parkhi, O. M., Vedaldi, A., Zisserman, A. & Jawahar, C. V. (2012). **Cats and Dogs**. In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). **Robust speech recognition via large-scale weak supervision**. *arXiv preprint arXiv:2212.04356*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). **Attention is all you need**. In *Advances in Neural Information Processing Systems*, 5998–6008.