

# Week2.2 More on LLMs

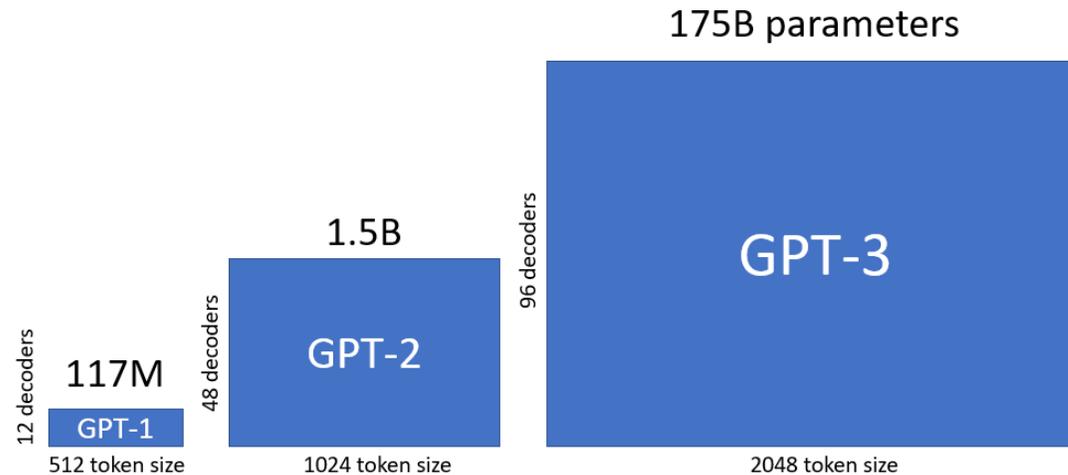
2024 Spring GenAI Risk & Benefits

Dr. Yanjun Qi

20240125

# Last Class:

- GPT1 / 2 / 3
- Emergent Abilities of Large Language Models
- Scaling Instruction-Finetuned Language Models
- On the Opportunities and Risks of Foundation Models

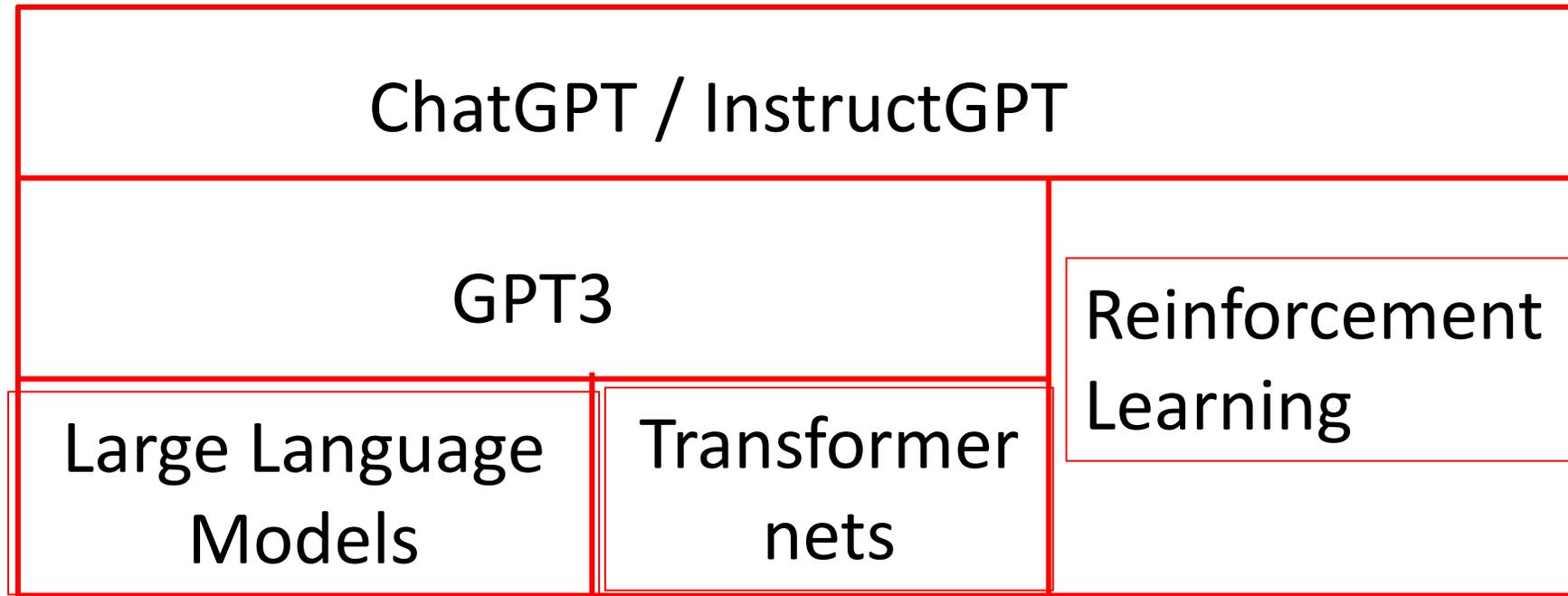


- No paper / Just a blog / Released Nov 30 2022

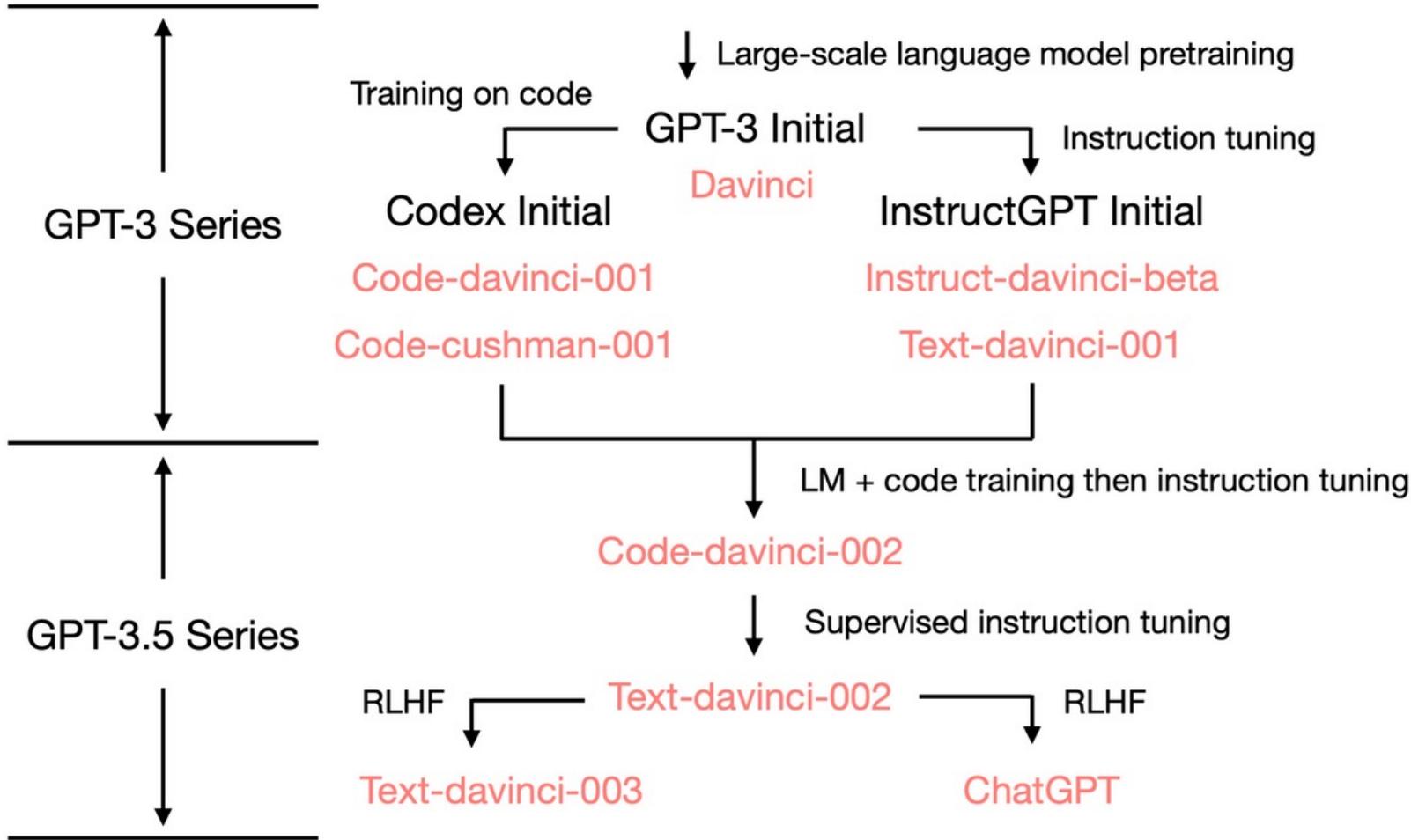
- It took 5 days to reach 1M users

# **CHATGPT: OPTIMIZING LANGUAGE MODELS FOR DIALOGUE” BY OPENAI**

# Concepts that ChatGPT builds on



# Family of GPT-3.5



# Results from InstructGPT paper

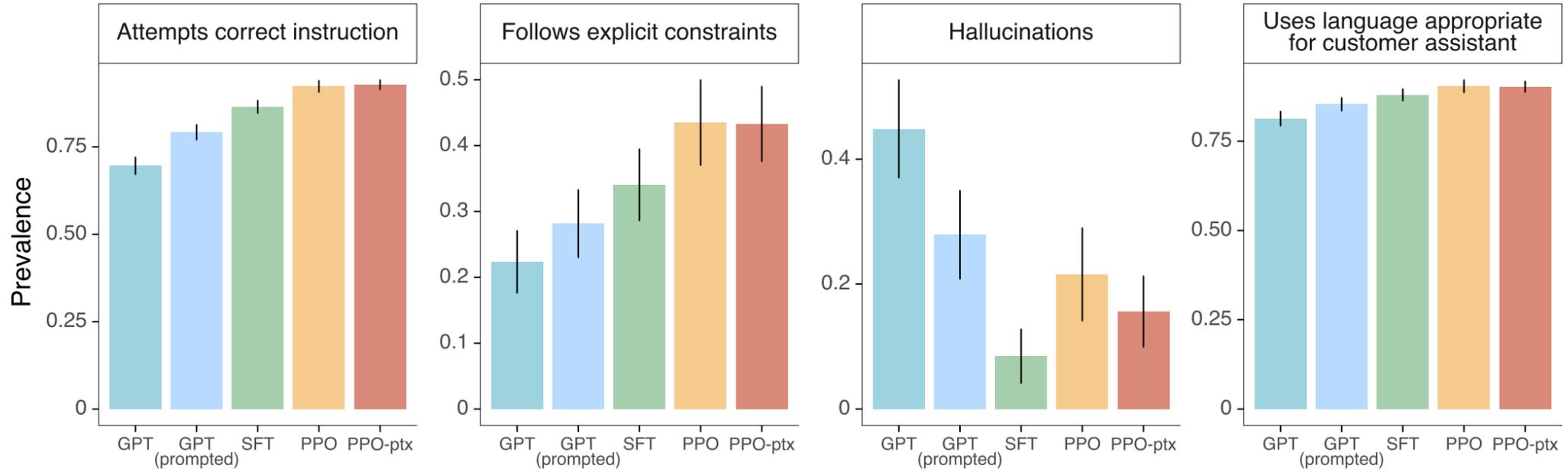


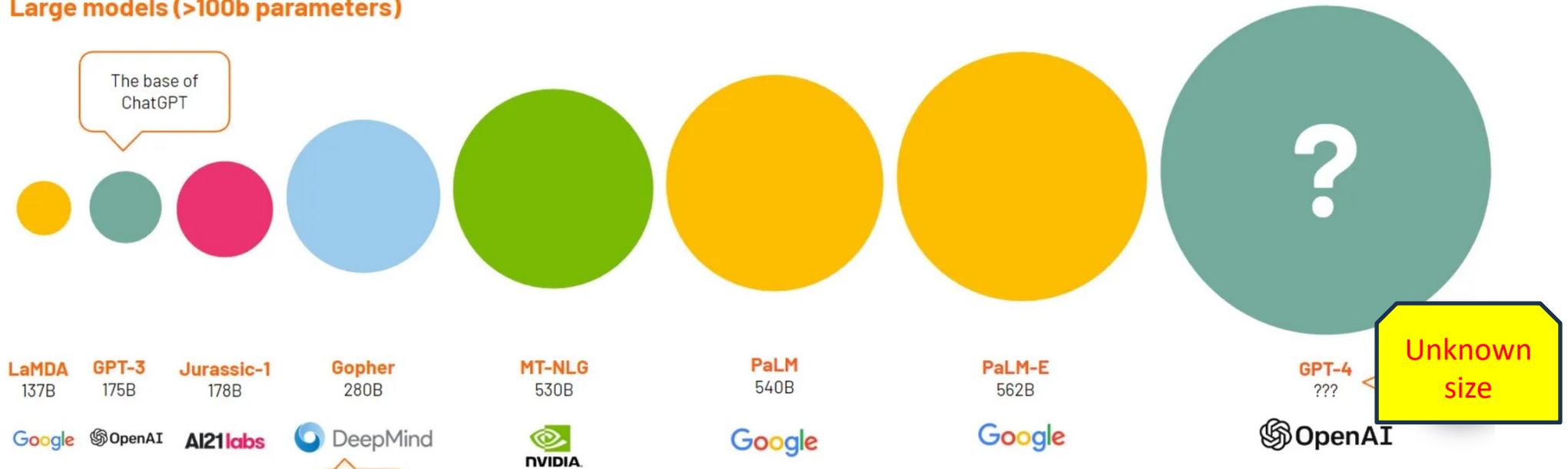
Figure 4: Metadata results on the API distribution. Note that, due to dataset sizes, these results are collapsed across model sizes. See Appendix E.2 for analysis that includes model size. Compared to GPT-3, the PPO models are more appropriate in the context of a customer assistant, are better at following explicit constraints in the instruction and attempting the correct instruction, and less likely

# LLMs Size changes

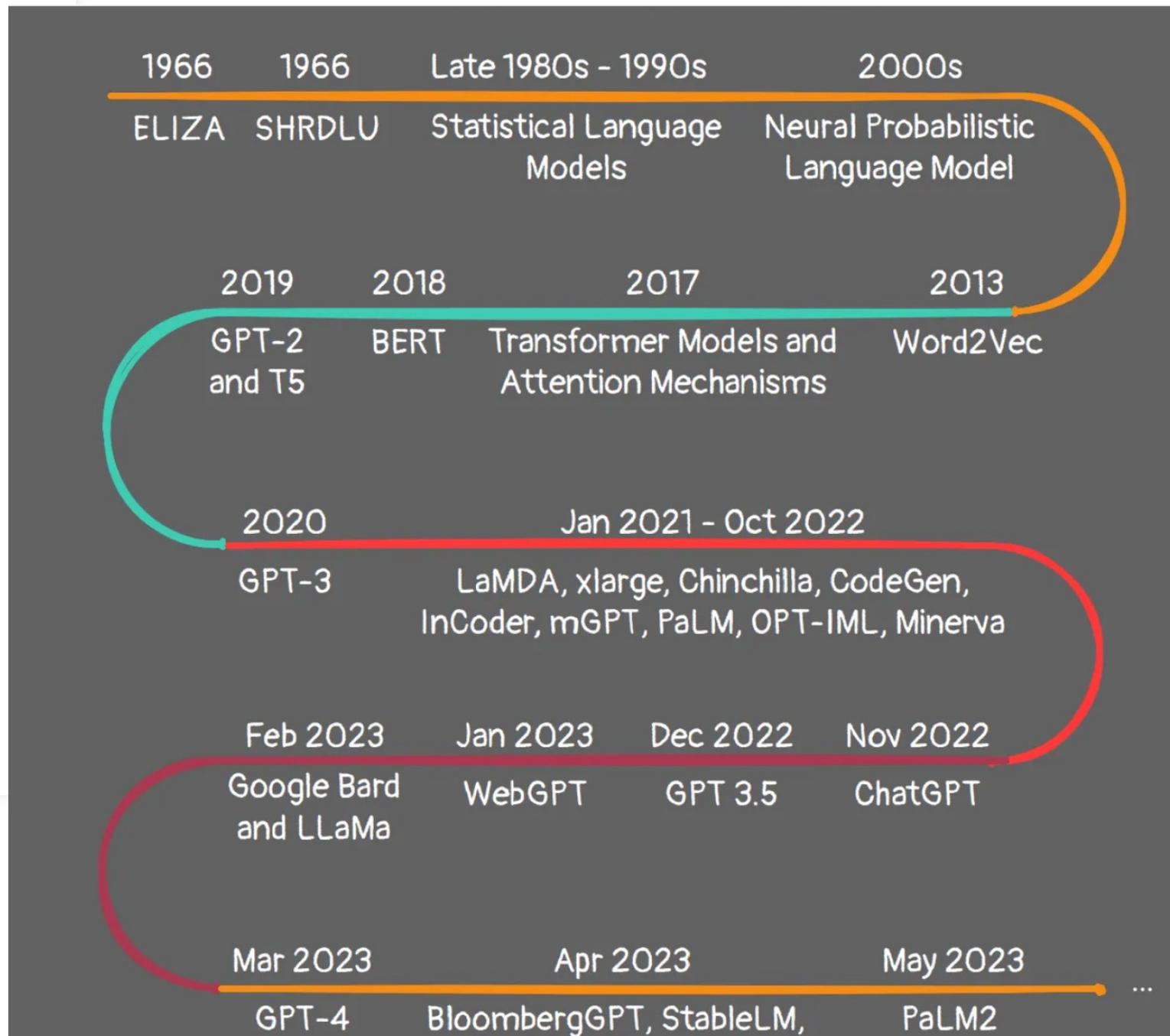
## Small models (<= 100b parameters)



## Large models (>100b parameters)



# Many new LLMs in 2022-2023



# LMSYS Chatbot Arena Leaderboard

[Vote](#) | [Blog](#) | [GitHub](#) | [Paper](#) | [Dataset](#) | [Twitter](#) | [Discord](#) |

LMSYS [Chatbot Arena](#) is a crowdsourced open platform for LLM evals. We've collected over 200,000 human preference votes to rank LLMs with the Elo ranking system.

Arena Elo Full Leaderboard

Total #models: 55. Total #votes: 230875. Last updated: Jan 18, 2024.

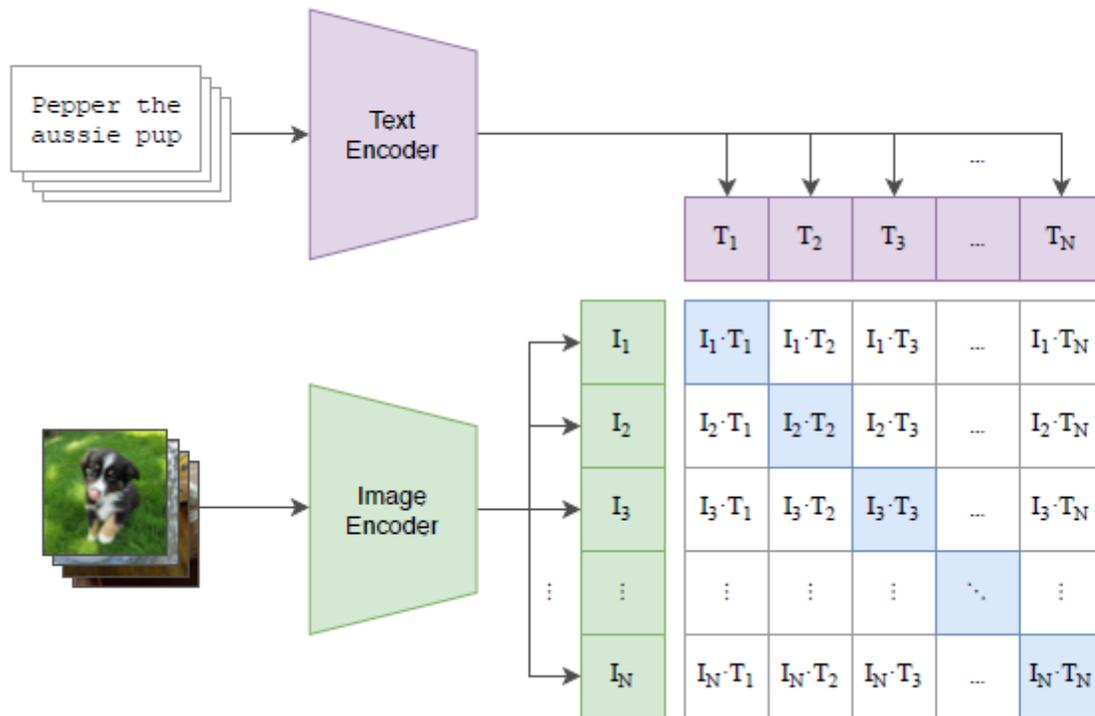
Contribute your vote at [chat.lmsys.org](https://chat.lmsys.org)! Find more analysis in the [notebook](#).

Rank	Model	★ Arena Elo	📊 95% CI	🗳️ Votes	Organization	License
1	<a href="#">GPT-4-Turbo</a>	1249	+14/-13	27399	OpenAI	Proprietary
2	<a href="#">GPT-4-0314</a>	1191	+15/-14	17372	OpenAI	Proprietary
3	<a href="#">GPT-4-0613</a>	1160	+13/-13	24888	OpenAI	Proprietary
4	<a href="#">Claude-1</a>	1150	+14/-13	17333	Anthropic	Proprietary
5	<a href="#">Mistral_Medium</a>	1148	+14/-13	9370	Mistral	Proprietary
6	<a href="#">Claude-2.0</a>	1131	+14/-13	11475	Anthropic	Proprietary
7	<a href="#">Mixtral-8x7b-Instruct-v0.1</a>	1124	+15/-13	13485	Mistral	Apache 2.0
8	<a href="#">Gemini_Pro_(Dev)</a>	1121	+15/-15	5304	Google	Proprietary

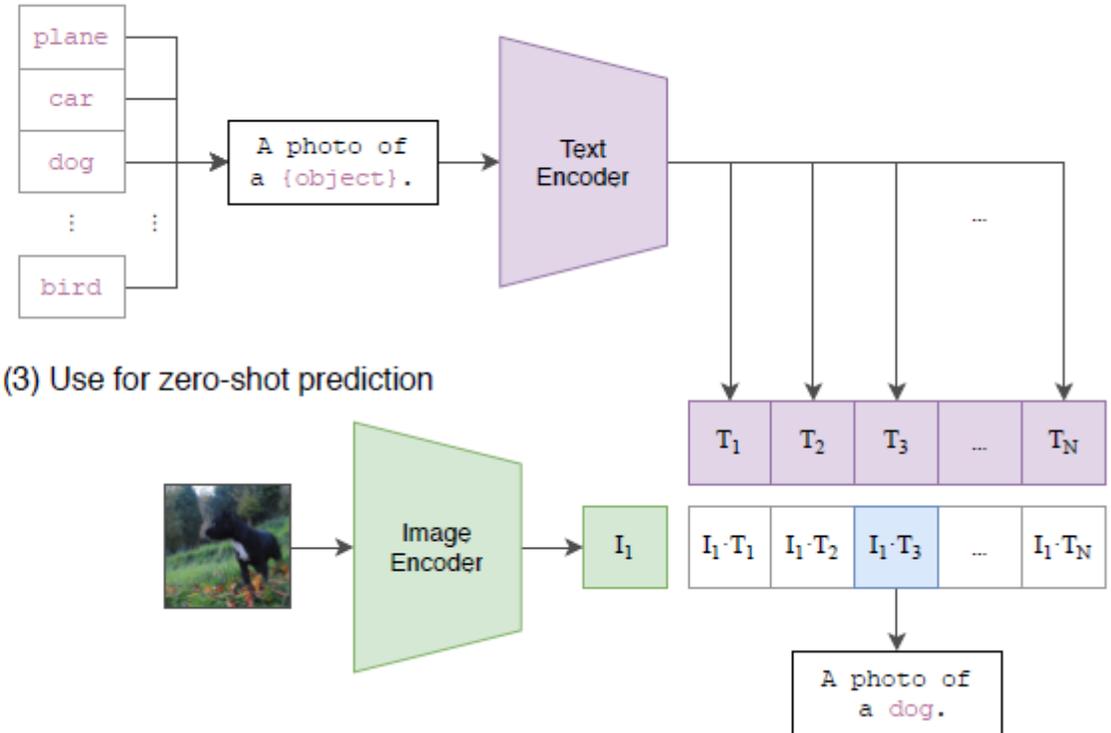
# **CLIP: CONTRASTIVE LANGUAGE-IMAGE PRETRAINING FOR VISION**

# CLIP: Learning Transferrable Models from Natural Language Supervision (Radford et al. 2021)

(1) Contrastive pre-training



(2) Create dataset classifier from label text



(3) Use for zero-shot prediction

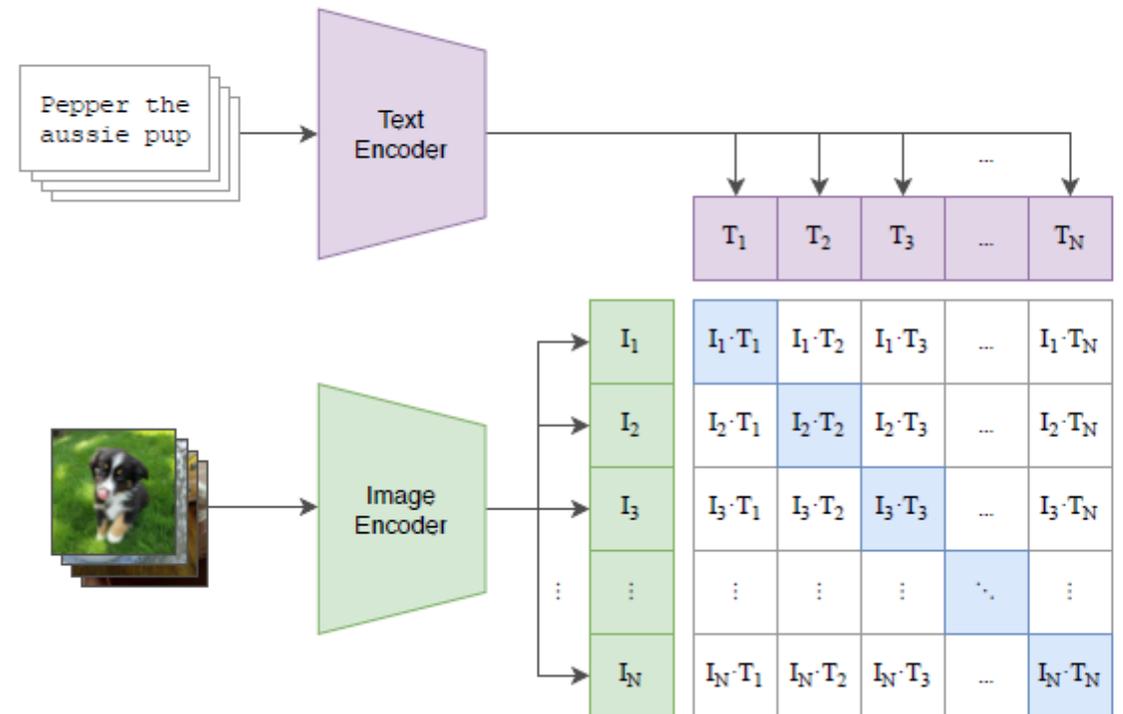
use a text encoder as a classifier

# Second key idea(s): **contrastively match text to image**

- Use small transformer language model (76M parameters for base)

“The largest ResNet model RN50x64, took 18 days to train on 592 V100 GPUs, while the largest Vision Transformer took 12 days on 256 V100 GPUs”

(1) Contrastive pre-training



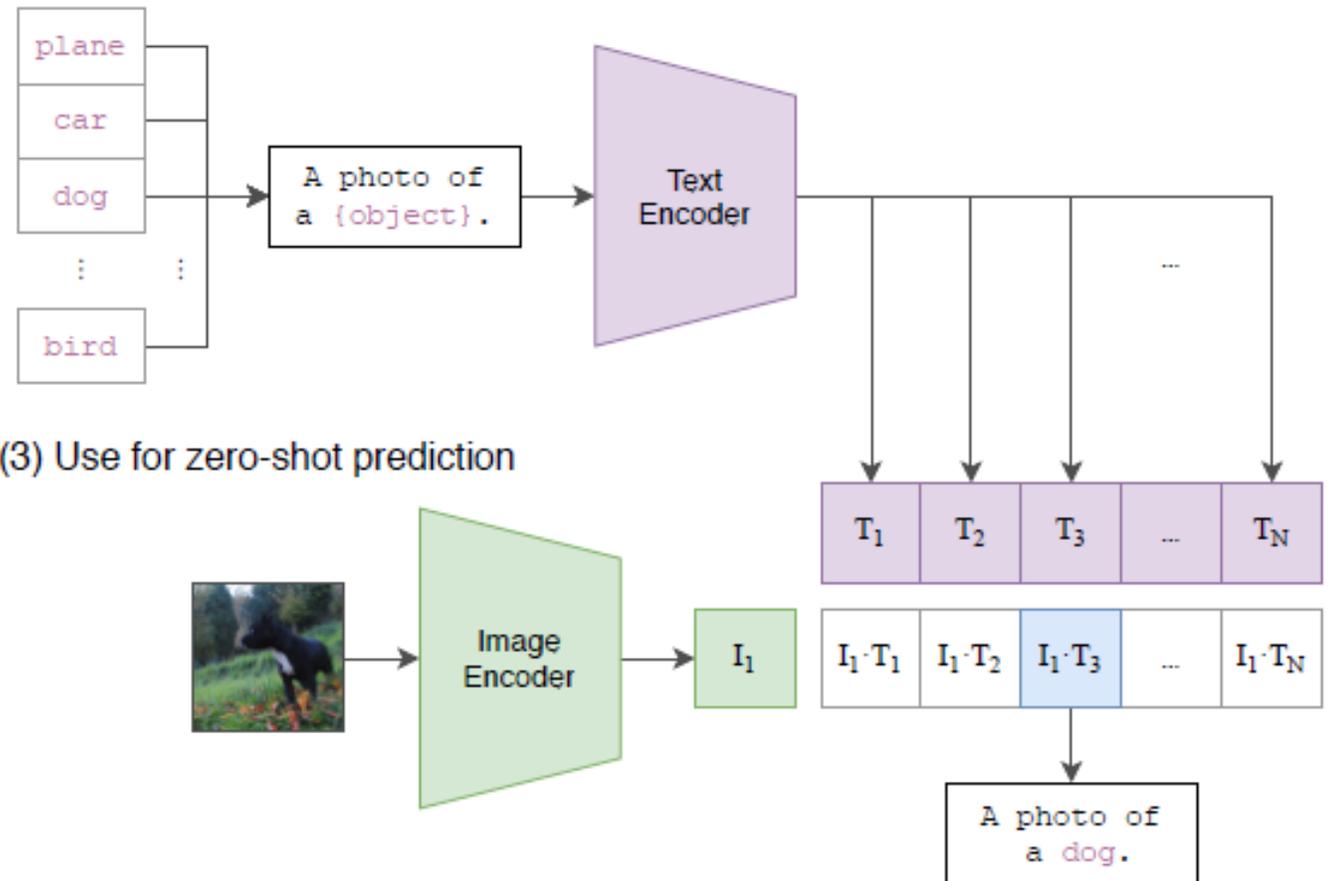
**Contrastive formulations is a good general pretrain way to learn when exact target is unpredictable**

# Key idea 3: zero-shot classification

To create a new classification task:

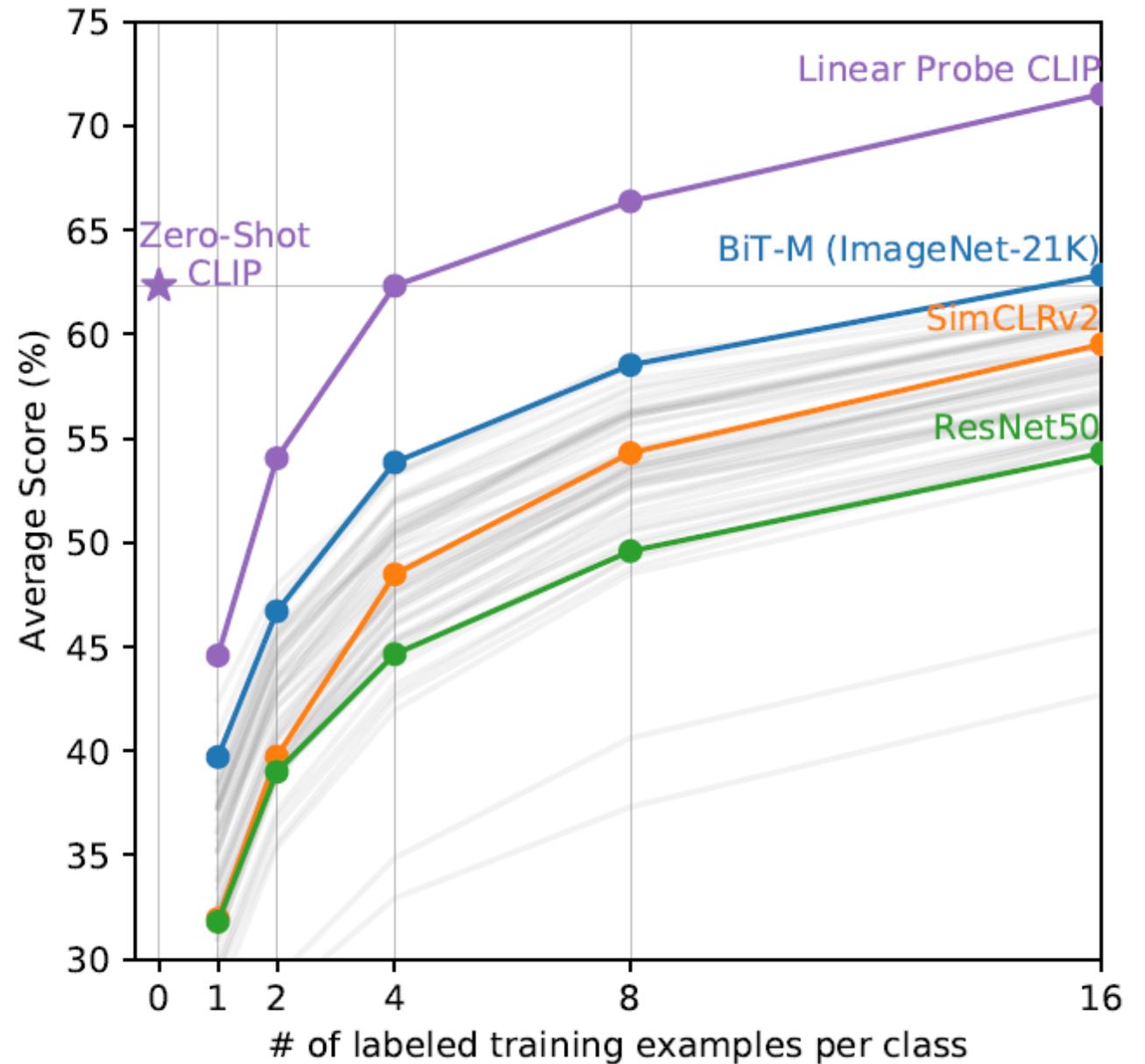
1. Convert class labels into captions and encode the text
2. Encode the image
3. Assign the image to the label whose caption matches best

(2) Create dataset classifier from label text



Every batch of training is like a novel classification task, matching 32K classes to 32K images

Pretrain learning that match images to text produces a good zero-shot classifier and an excellent image encoder



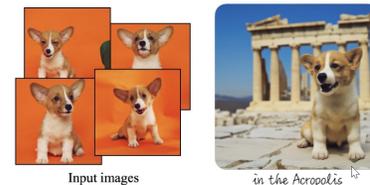
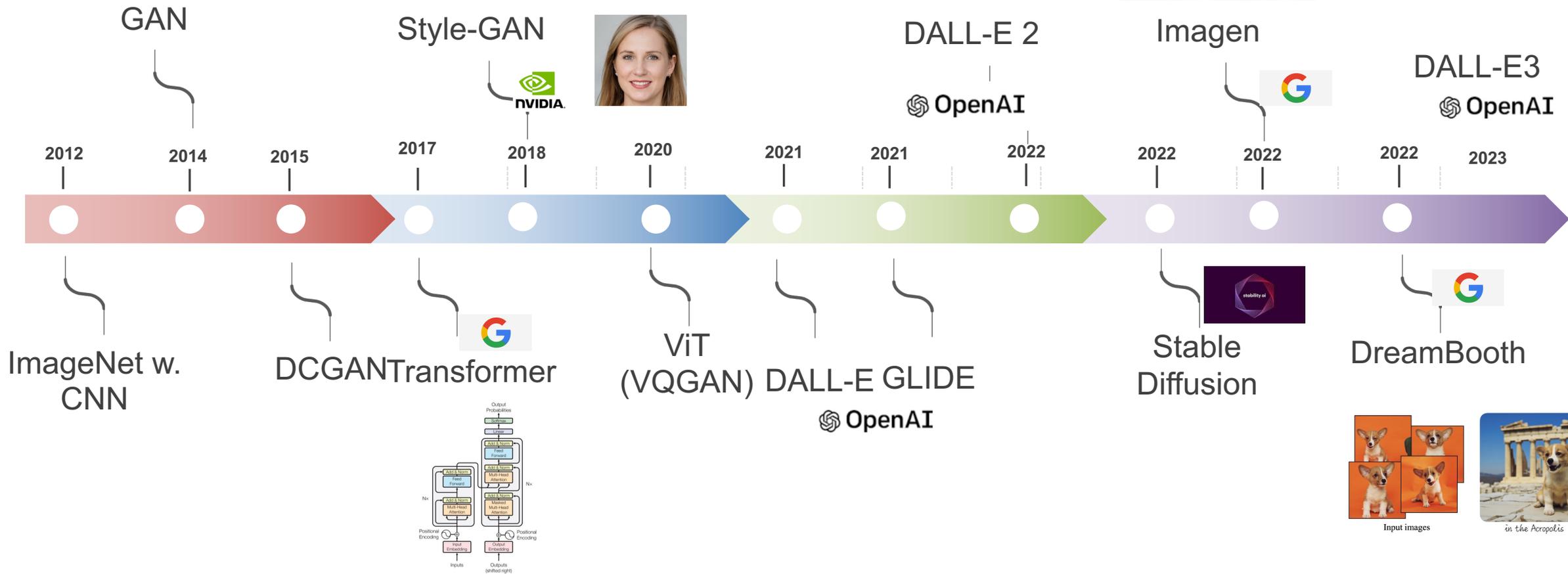
Zero shot to few shot image prediction

openAI DALLE-3 with  
prompt = "Bears,  
Beets or Battlestar  
Galactica"

## TEXT 2 IMAGE GENERATION



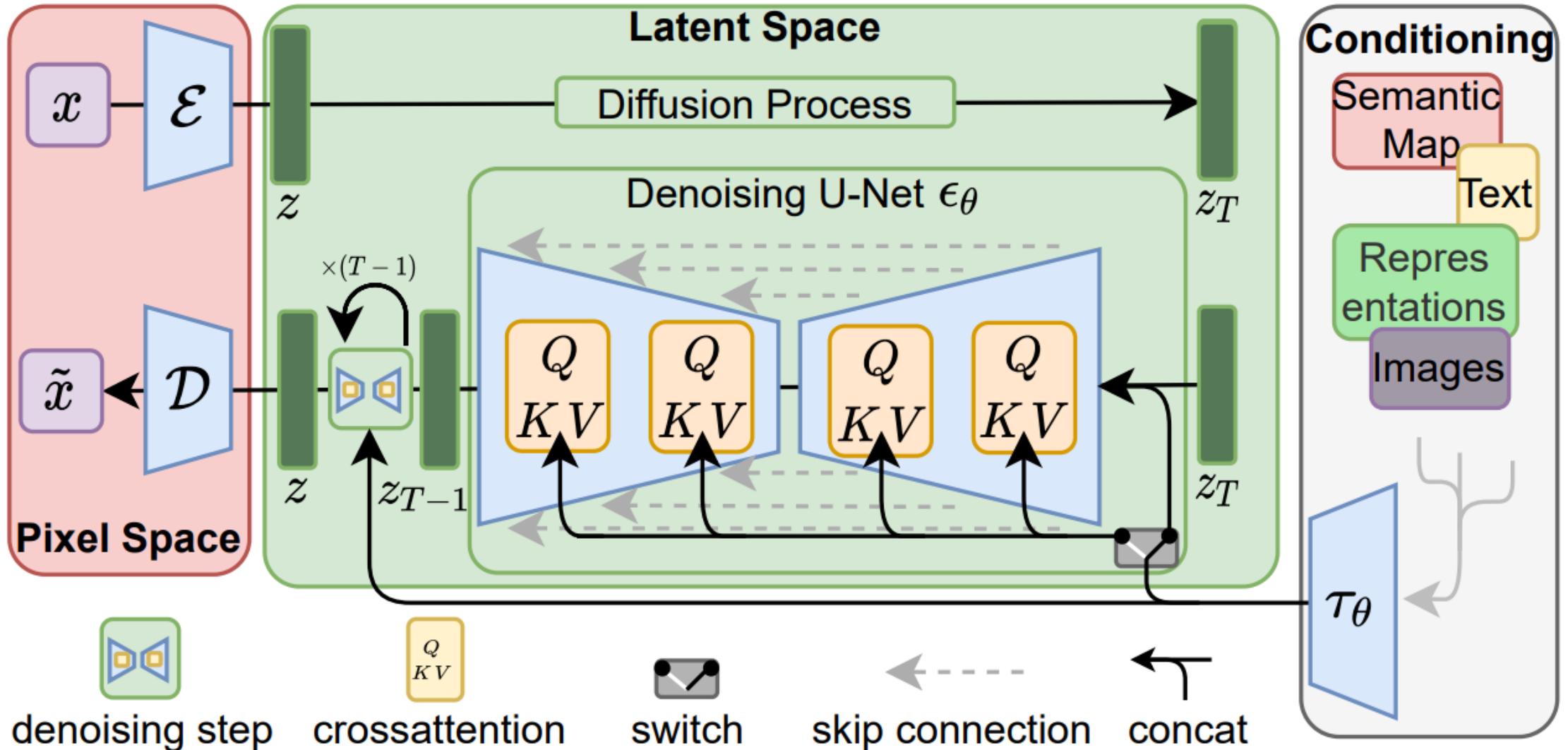
# History of Image Generation

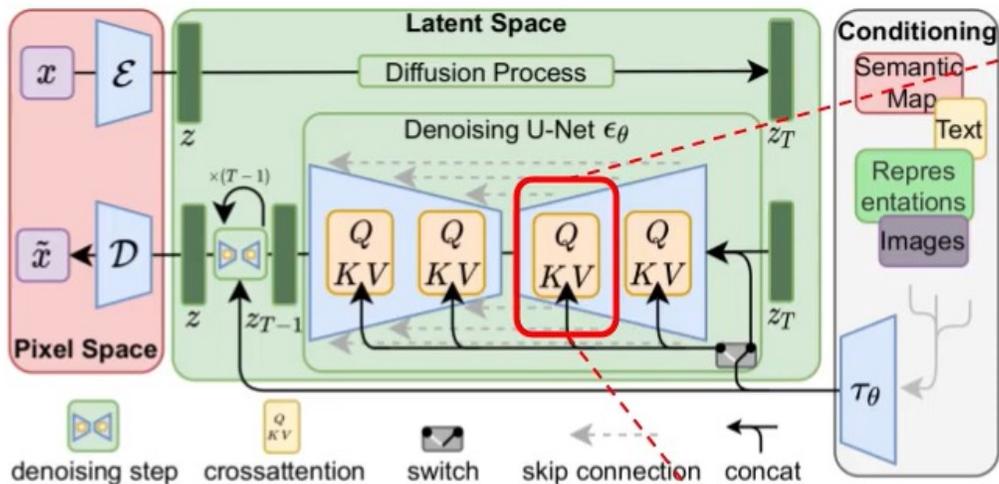


# Stable diffusion (2022-08)

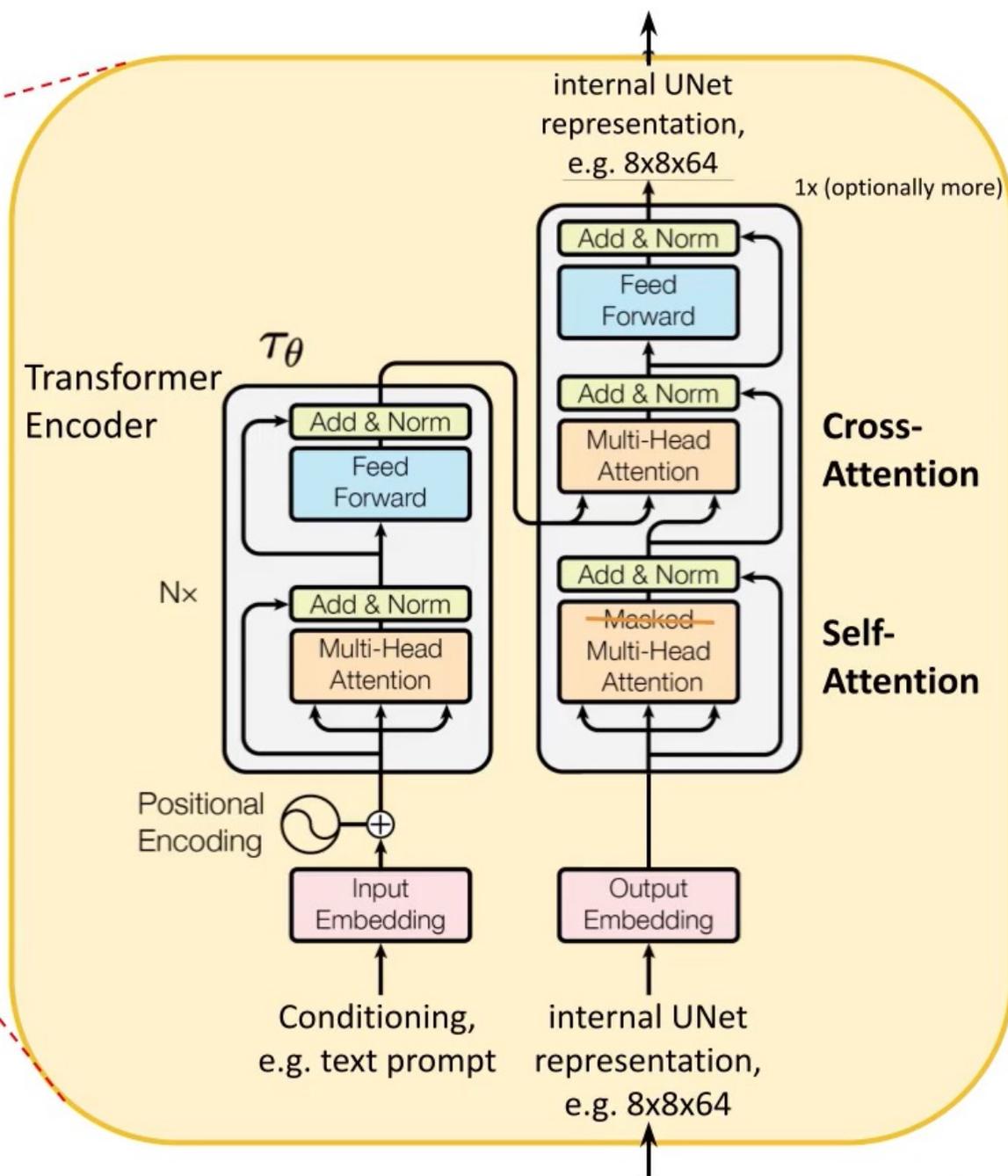
- [High-Resolution Image Synthesis with Latent Diffusion Models](#)  
CVPR'22 Rombach et al.
- [Code and models](#) released Open Source
- [The CreativeML OpenRAIL M license](#)
- by Stability AI and Runway

# Latent Diffusion Model (LDM)





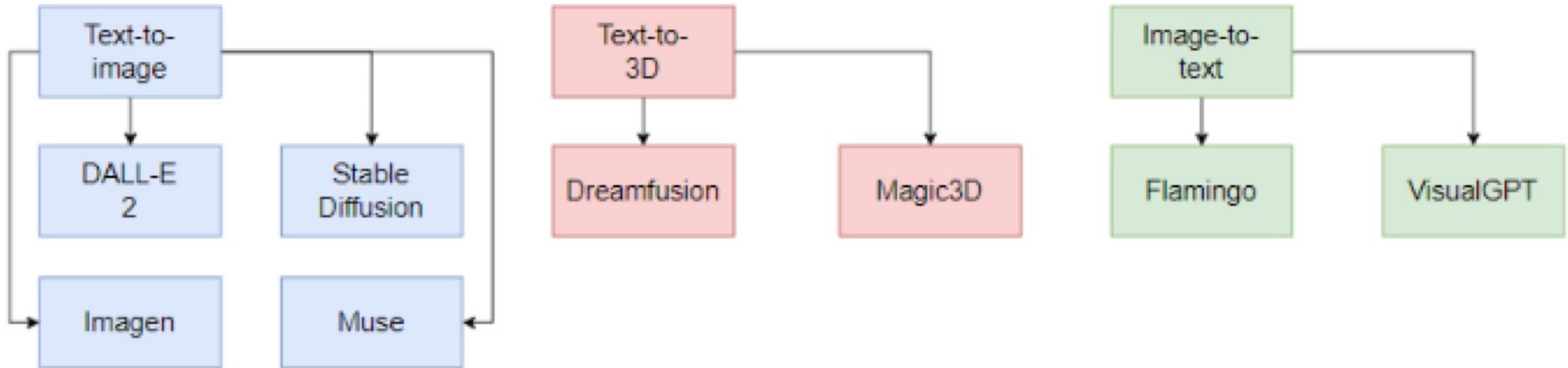
# Conditioning on text



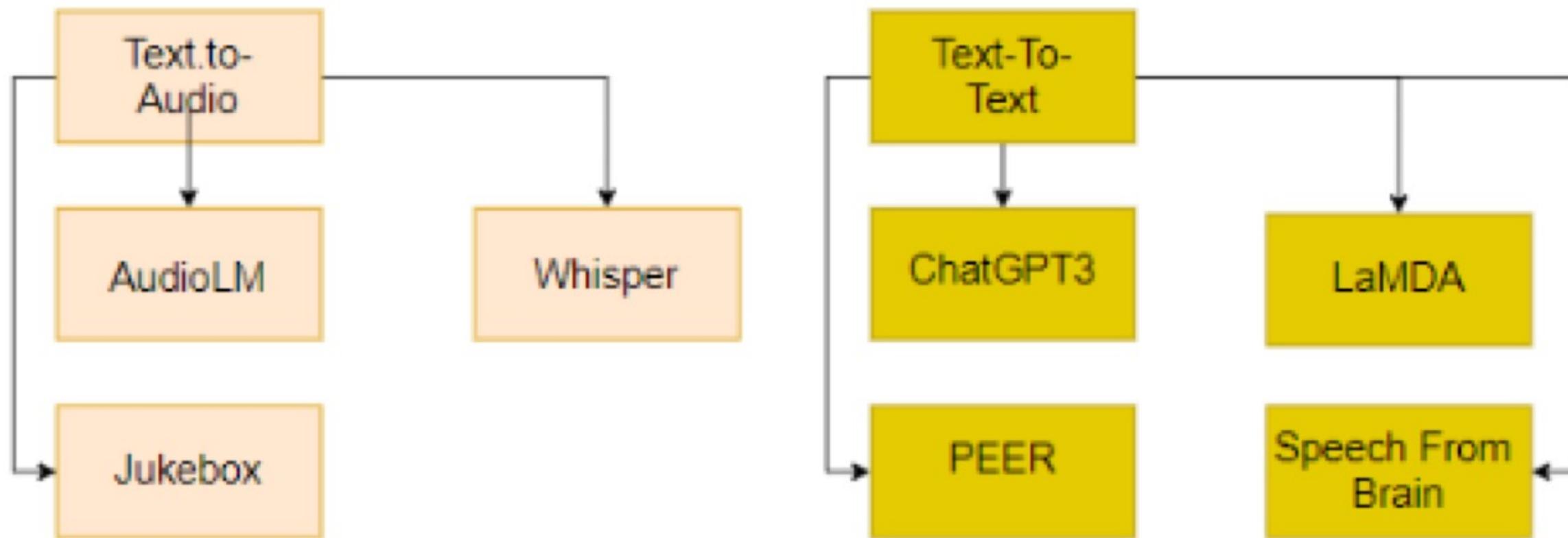
*[Submitted on 11 Jan 2023] --- OLD for GenAI*

# **CHATGPT IS NOT ALL YOU NEED. A STATE OF THE ART REVIEW OF LARGE GENERATIVE AI MODELS**

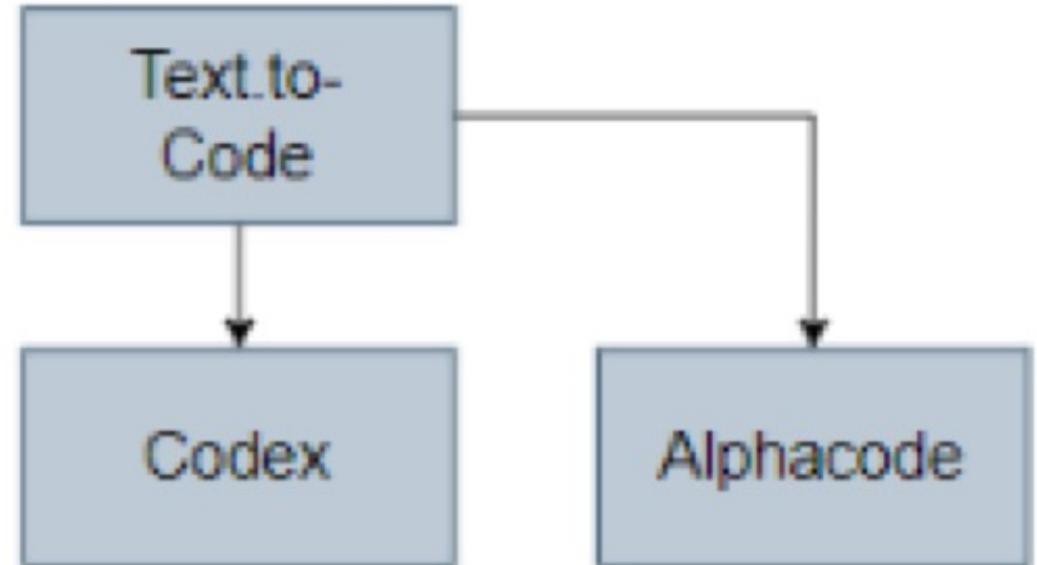
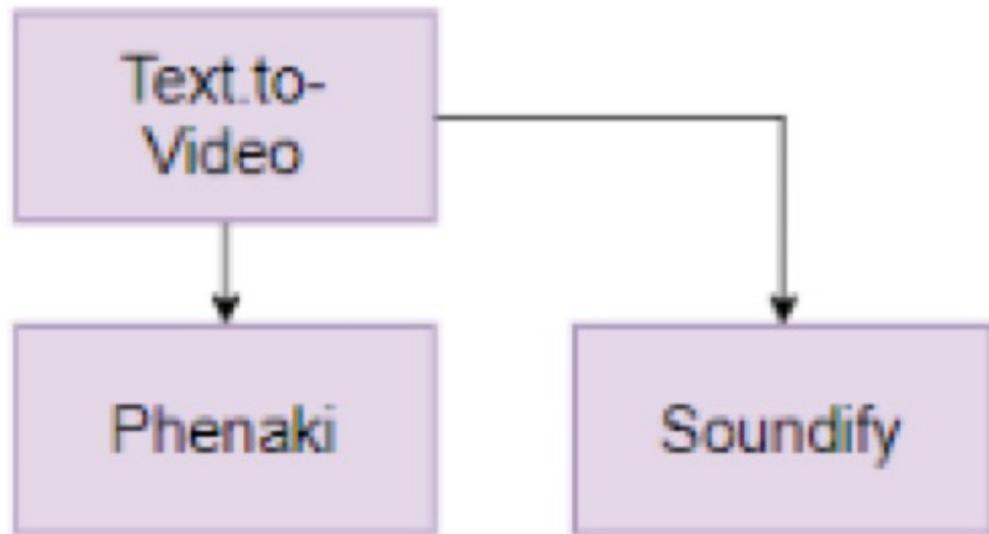
A taxonomy of the most popular generative AI models that have recently appeared classified according to their input and generated formats.



Many new ones out in each type in 2023

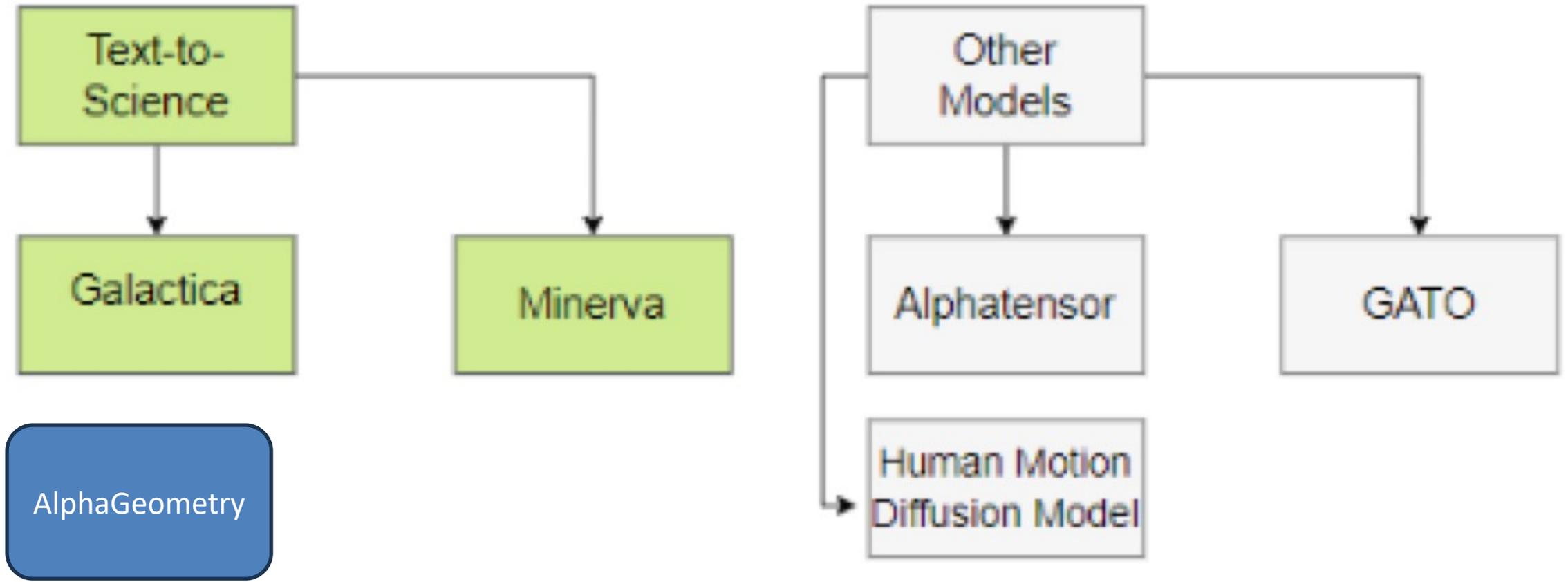


Many new ones out in each type in 2023



**Emu Video**

Many new ones out in each type in 2023



Many new ones out in each type in 2023

# A Comprehensive Overview of Large Language Models

<https://arxiv.org/pdf/2307.06435.pdf>

Chronological display of LLM releases: light blue rectangles represent 'pre-trained' models, while dark blue rectangles correspond to 'instruction-tuned' models. Models on the upper half signify open-source availability, whereas those on the bottom half are closed-source.

