

The background of the slide is a scenic photograph of a Swedish landscape. In the foreground, there are several traditional red wooden houses with white window frames and green grass. A large, calm lake reflects the sky and the surrounding greenery. In the background, there are rolling hills and mountains, some with patches of snow, under a cloudy sky.

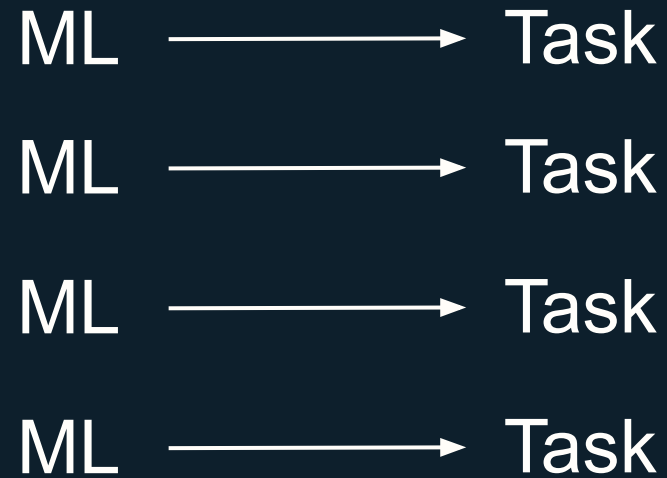
# GPT-SW3: a foundational resource for Swedish NLP

Language models learn language by reading text

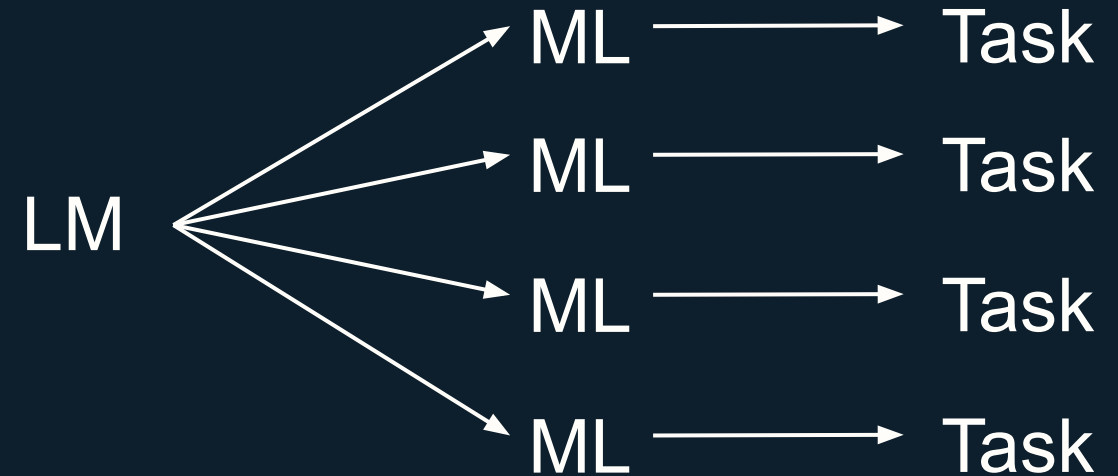
Can be used to solve (all) language processing tasks

Pretrain - finetune - predict

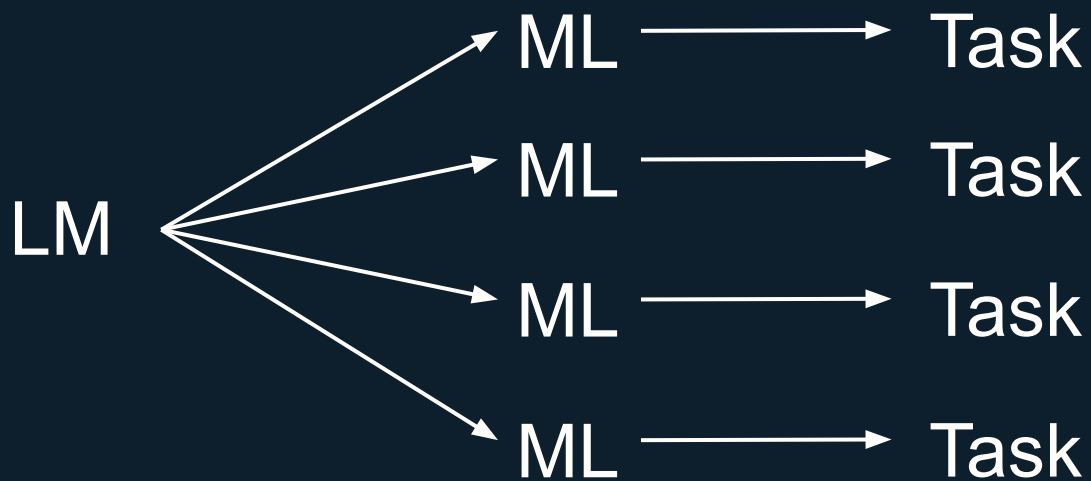
## Before 2018



## After 2018

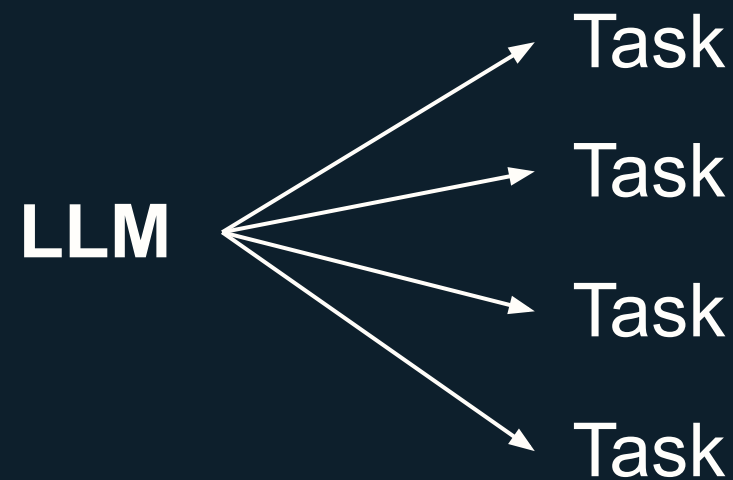


**After 2018**



**Encoders**

**After 2021**

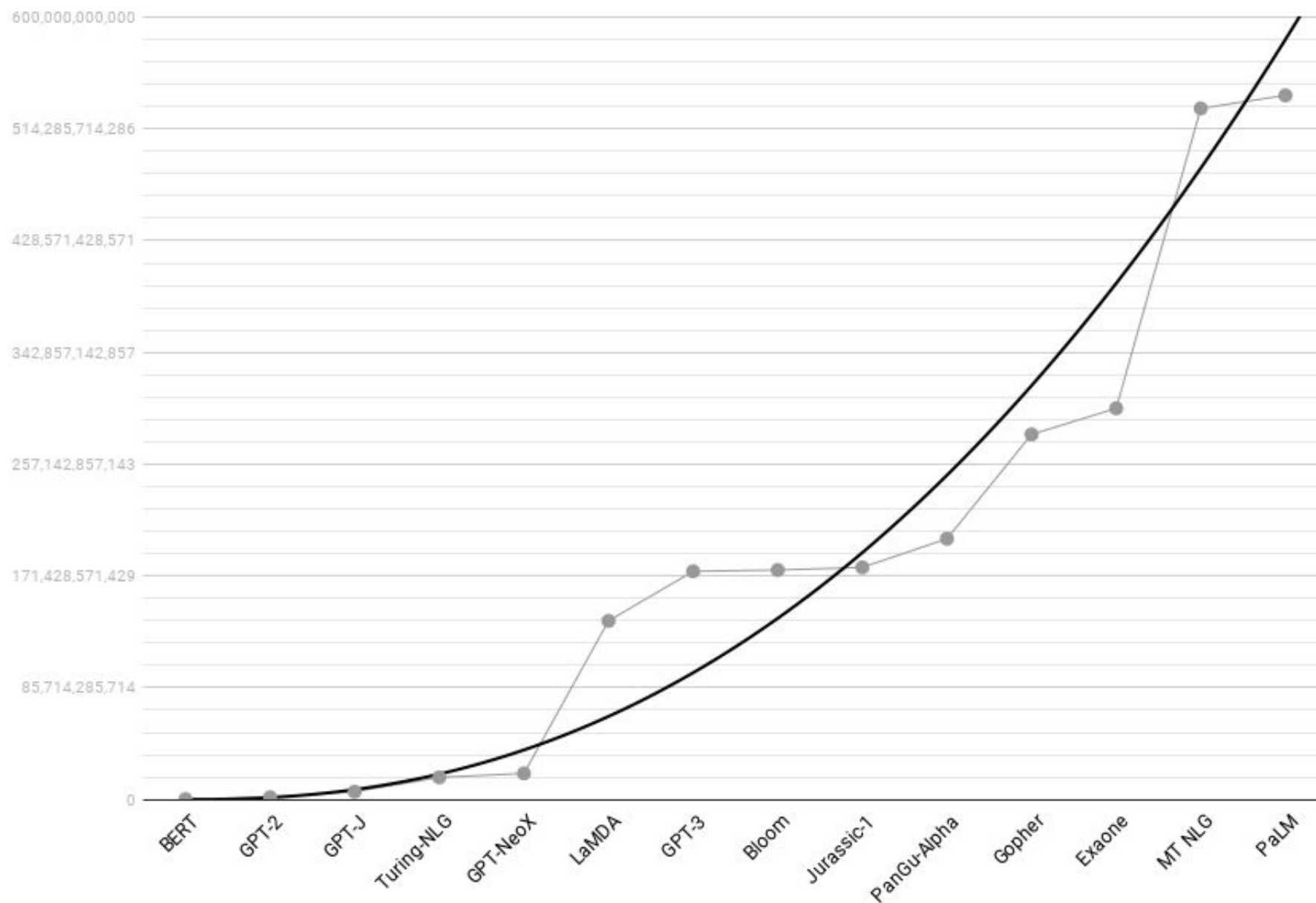


**Decoders**

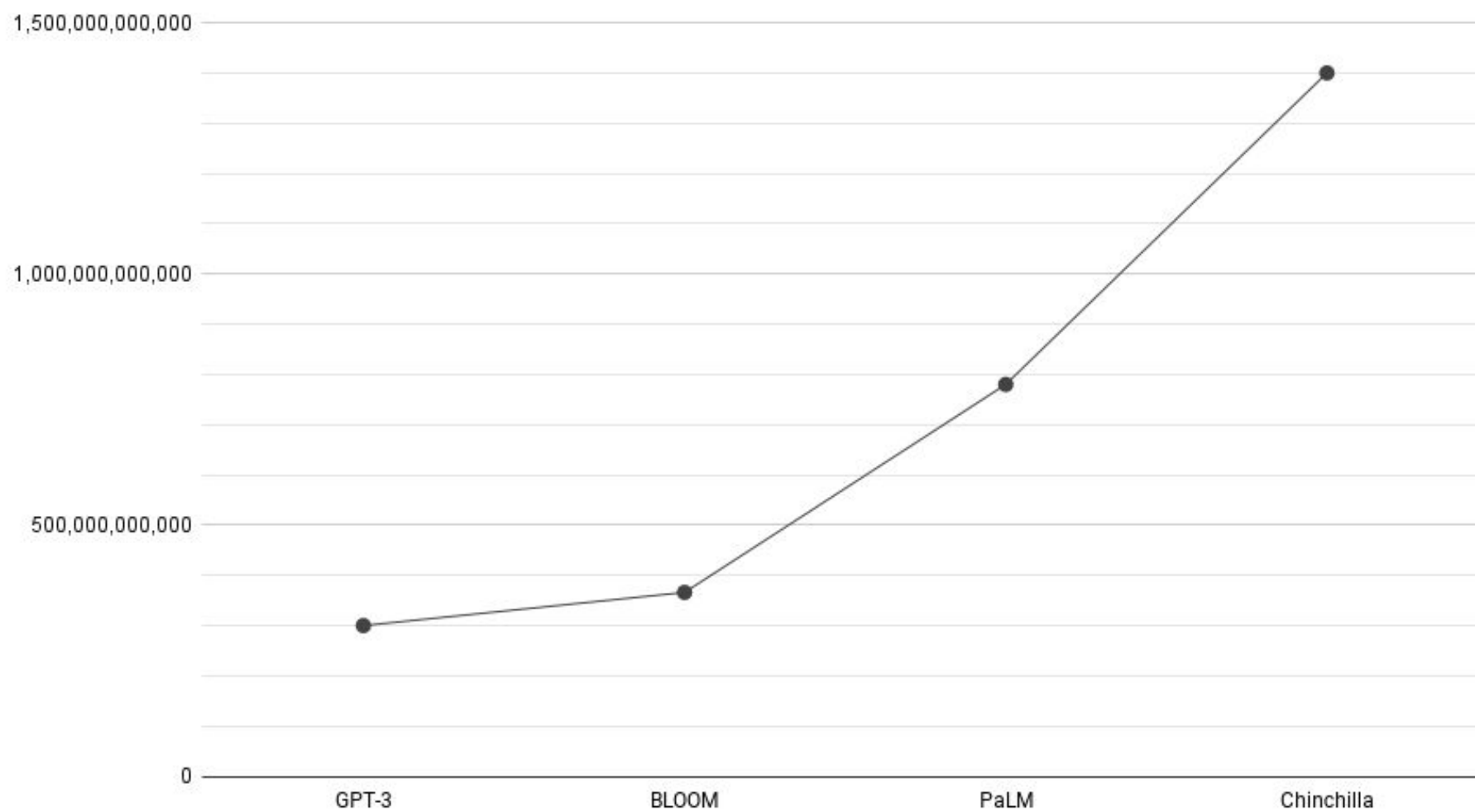


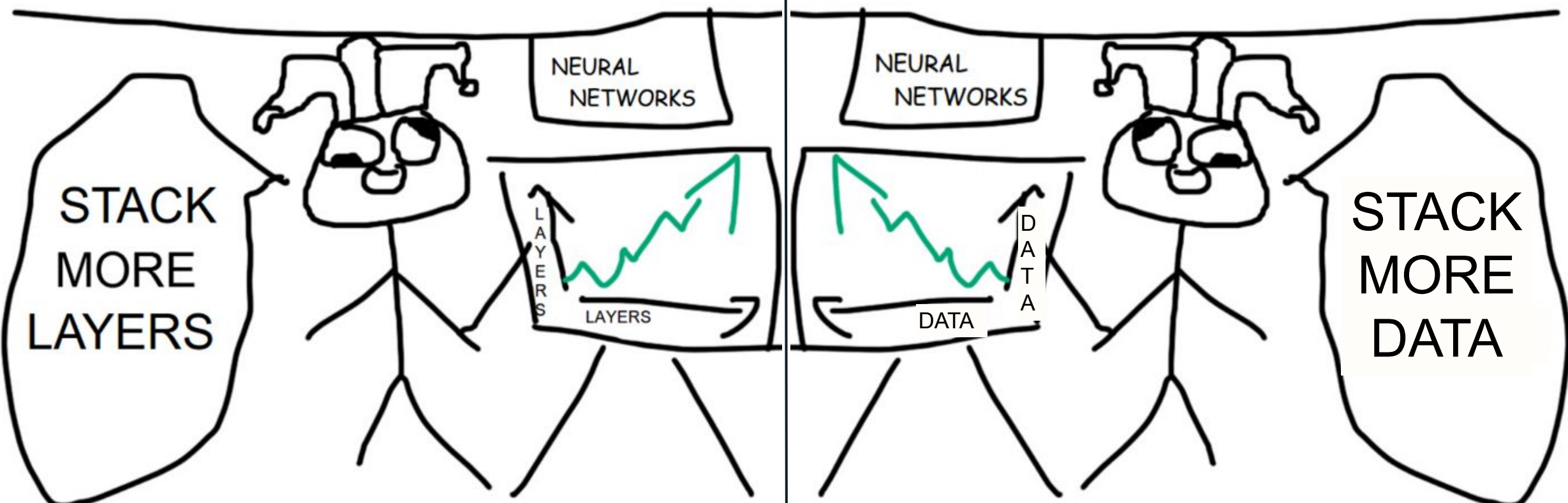


## Transformer model parameters



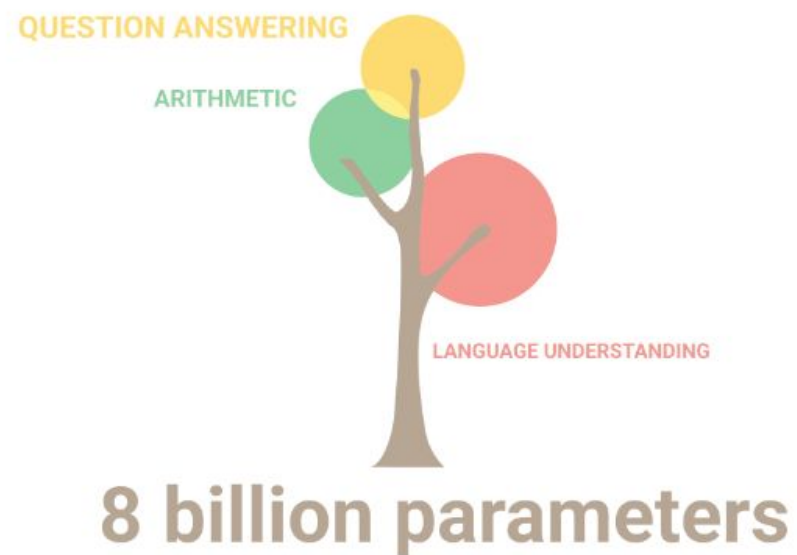
## Number of training tokens





Kaplan et al. (2020) Scaling laws for neural language models

Hofman et al. (2022) Training compute-optimal large language models



<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>

Solve tasks by formulating them as generative tasks

Input instructions to the model (**prompts**)

*In theory* no limit to what tasks we can solve,  
as long as we can formulate prompts



## The three settings we explore for in-context learning

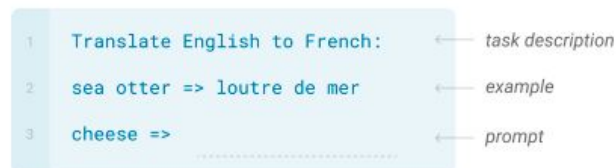
### Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



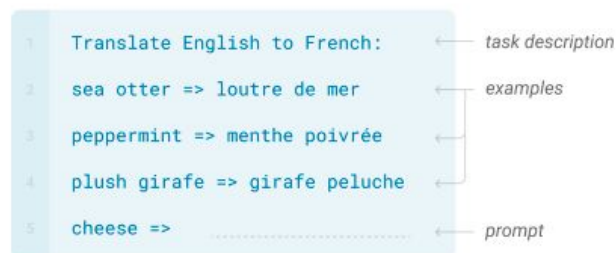
### One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.



### Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



## Traditional fine-tuning (not used for GPT-3)

### Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.



### Explaining a joke

#### Prompt

Explain this joke:

Joke: Did you see that Google just hired an eloquent whale for their TPU team? It showed them how to communicate between two different pods!

#### Model Response

Prediction: TPUs are a type of computer chip that Google uses for deep learning. A "pod" is a group of TPUs. A "pod" is also a group of whales. The joke is that the whale is able to communicate between two groups of whales, but the speaker is pretending that the whale is able to communicate between two groups of TPUs.

#### Standard Prompting

##### Example 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?

##### Example Output

A: The answer is 11.

##### Prompt

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

##### Model Response

The answer is 50. ❌

#### Chain of thought prompting

##### Example 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?

##### Example Output

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

##### Prompt

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

##### Model Response

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. ✅

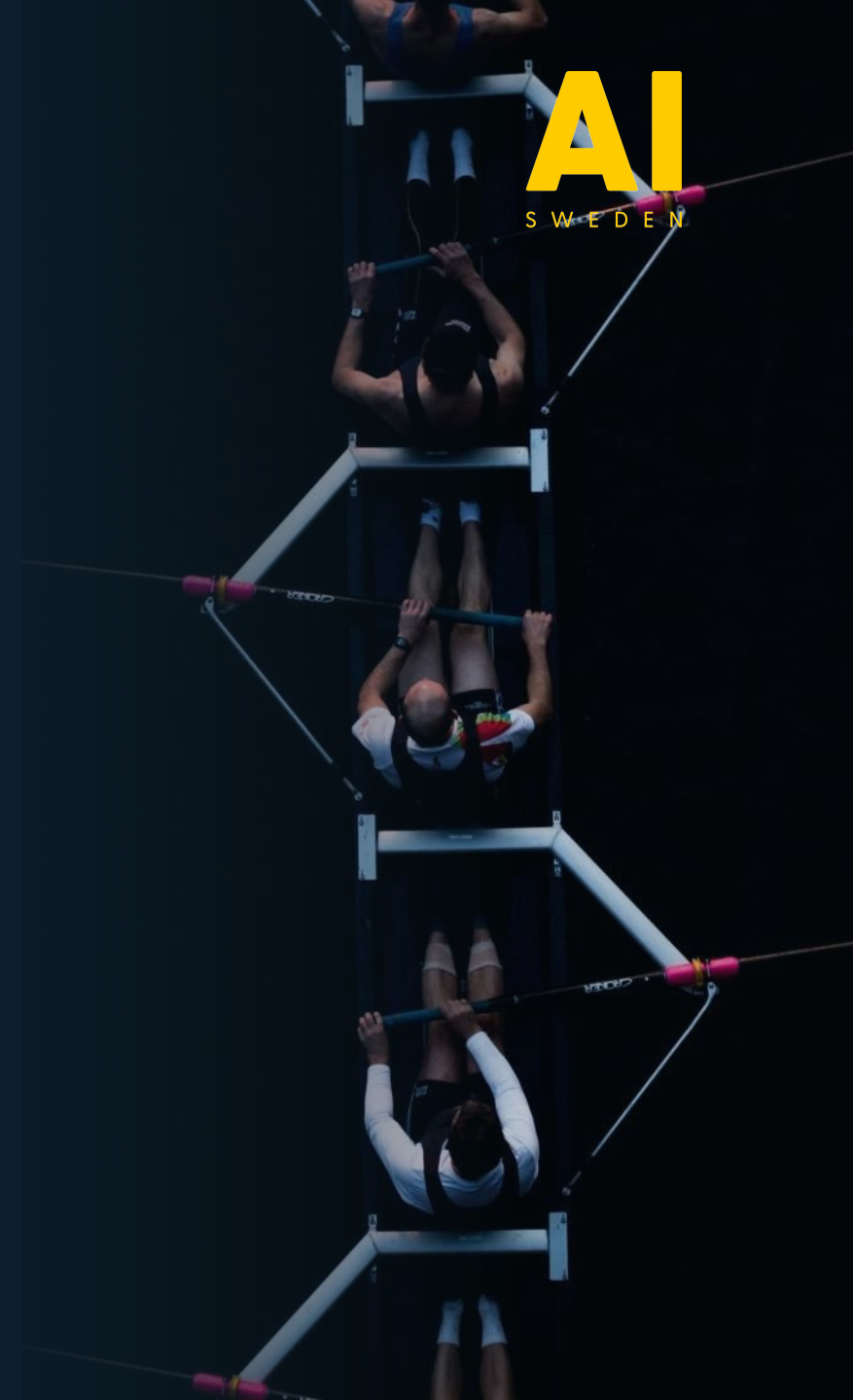
<https://arxiv.org/pdf/2005.14165.pdf>

<https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html>

Prompting: zero-shot, one-shot, few-shot

Prompt-tuning: find prompts that work well

**P-tuning: find embeddings that work well**









**Data**



**Infrastructure**



**Competence**



**AI Nordics Discord:  
[discord.gg/hh3fpz6X](https://discord.gg/hh3fpz6X)**



Large models require large data

BLOOM: 366B tokens

GPT-3: 500B tokens

PaLM: 780B tokens

Chinchilla: 1.4T tokens

How do we find enough data in a small language like Swedish?



**Swedish is a small language, but with close friends**  
**Pool resources from typologically similar languages**

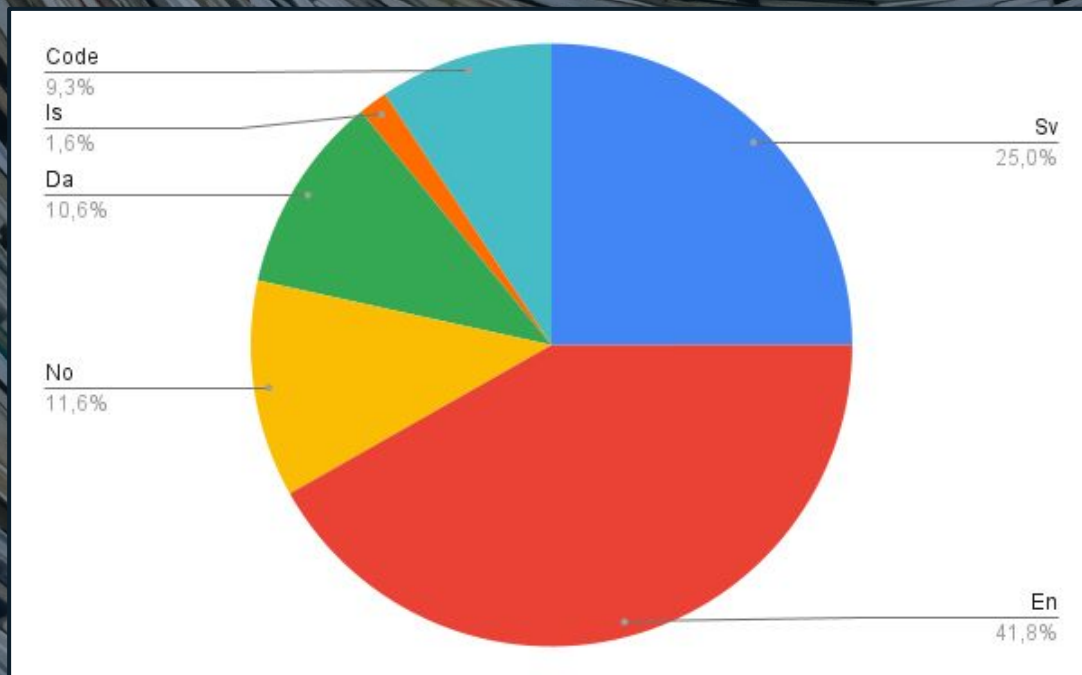
**The Nordic Pile:** The flags of Sweden, Norway, Denmark, and Iceland are displayed in a row. From left to right: the Swedish flag (blue with a yellow cross), the Norwegian flag (red, white, and blue horizontal stripes), the Danish flag (red with a white cross), and the Icelandic flag (blue, red, and green horizontal stripes with a white border).

**Swedish, Norwegian, Danish, Icelandic, English**  
**1.3 TB in total**

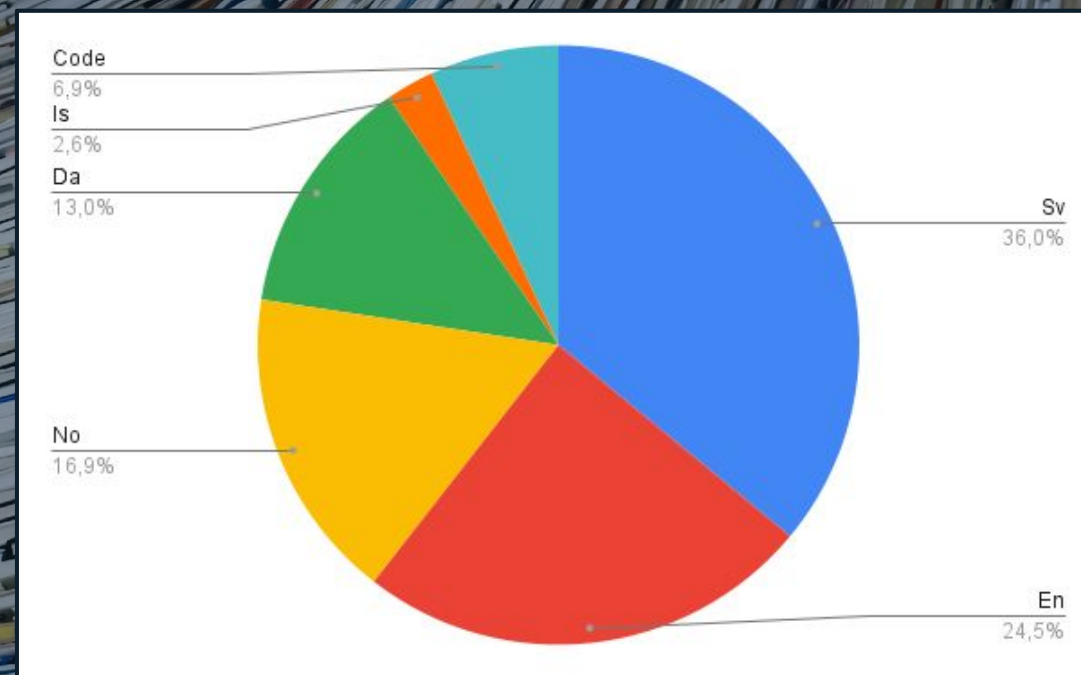
- **Normalization (whitespace, character encoding)**
- **Quality filtering**
- **Exact deduplication**
- **Fuzzy deduplication (intra-language LSH)**
- **Weighting of data sources**

**Tokenizer: SentencePiece**  
**Vocabulary size: 64k**



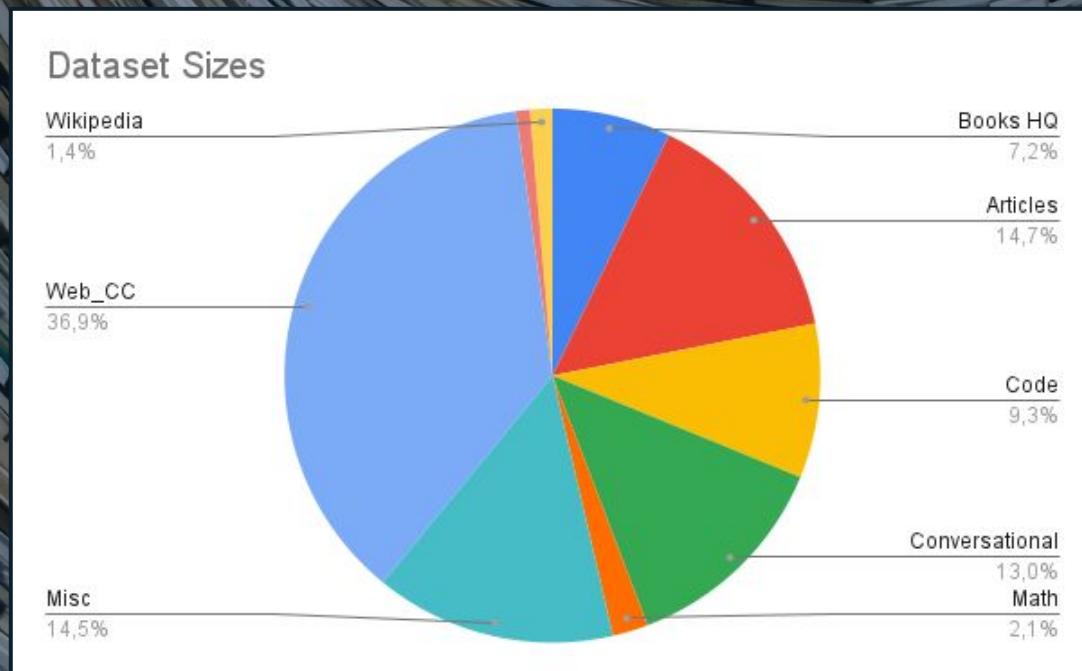


Raw distribution

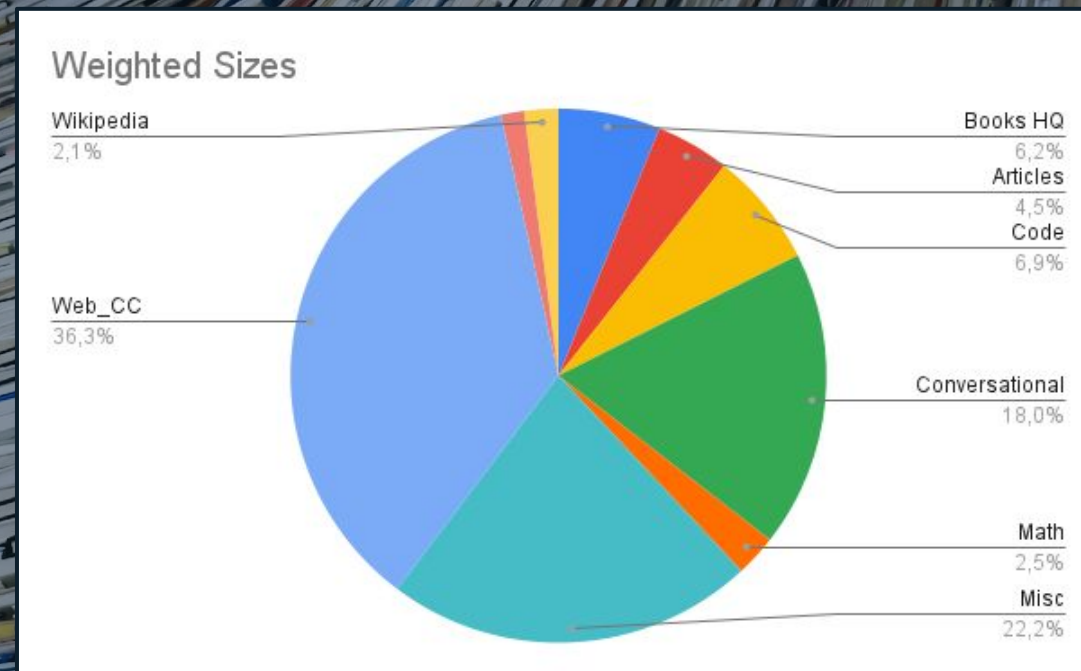


Weighted distribution



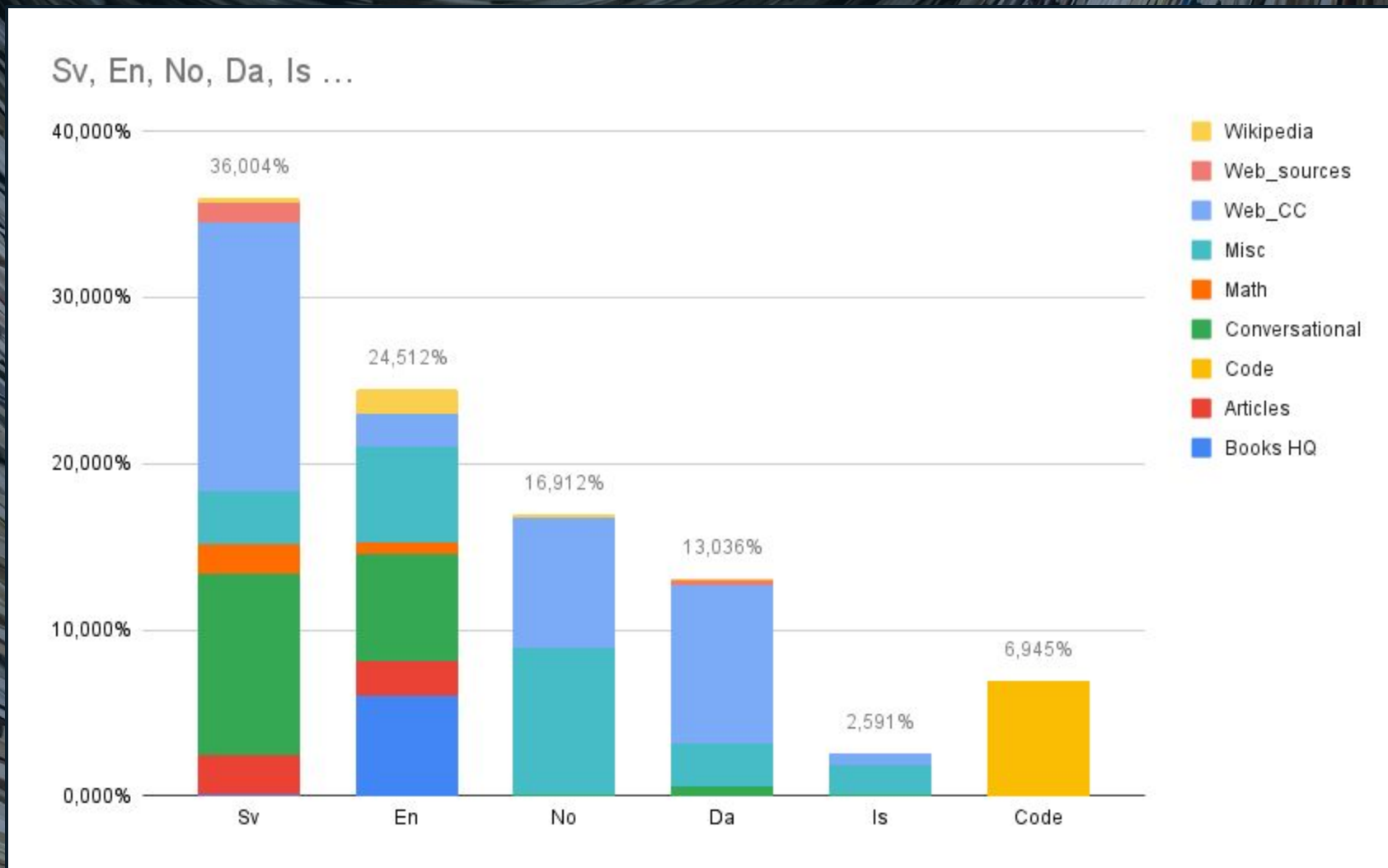


Raw distribution



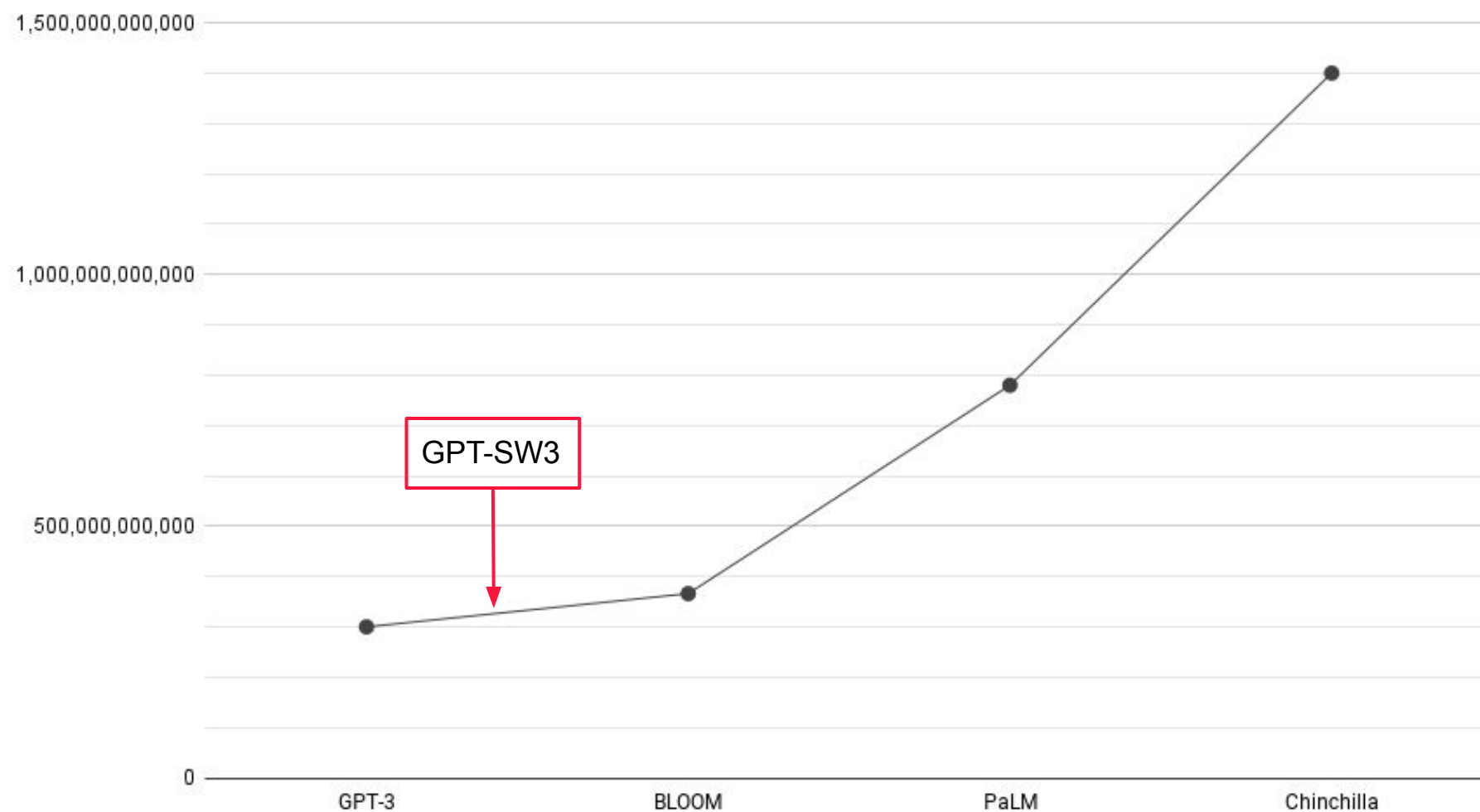
Weighted distribution







## Number of training tokens





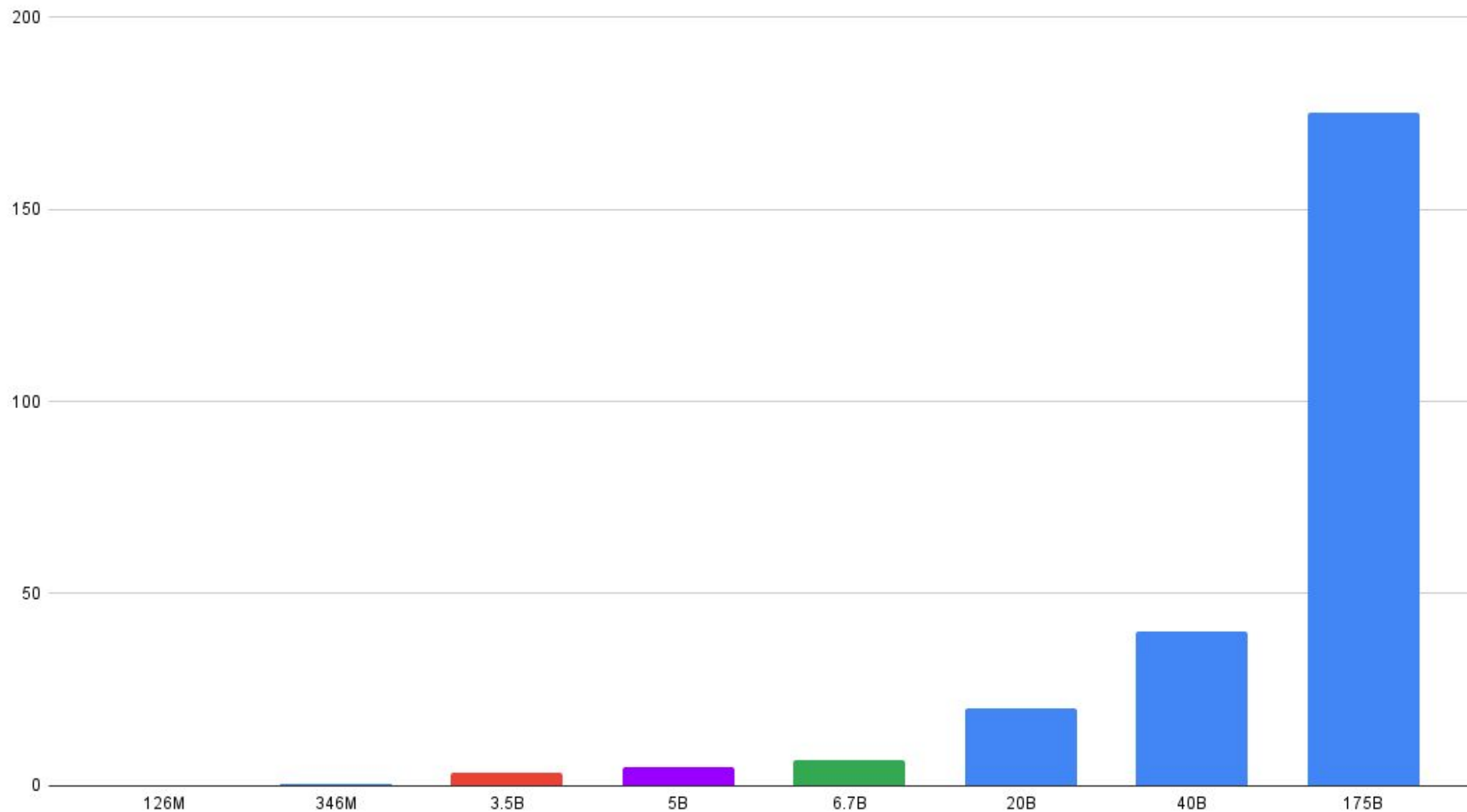
**NVIDIA SuperPOD:  
60 NVIDIA DGX-A100 compute nodes**

**Each DGX-A100:  
8 NVIDIA A100 Tensor Core GPUs**

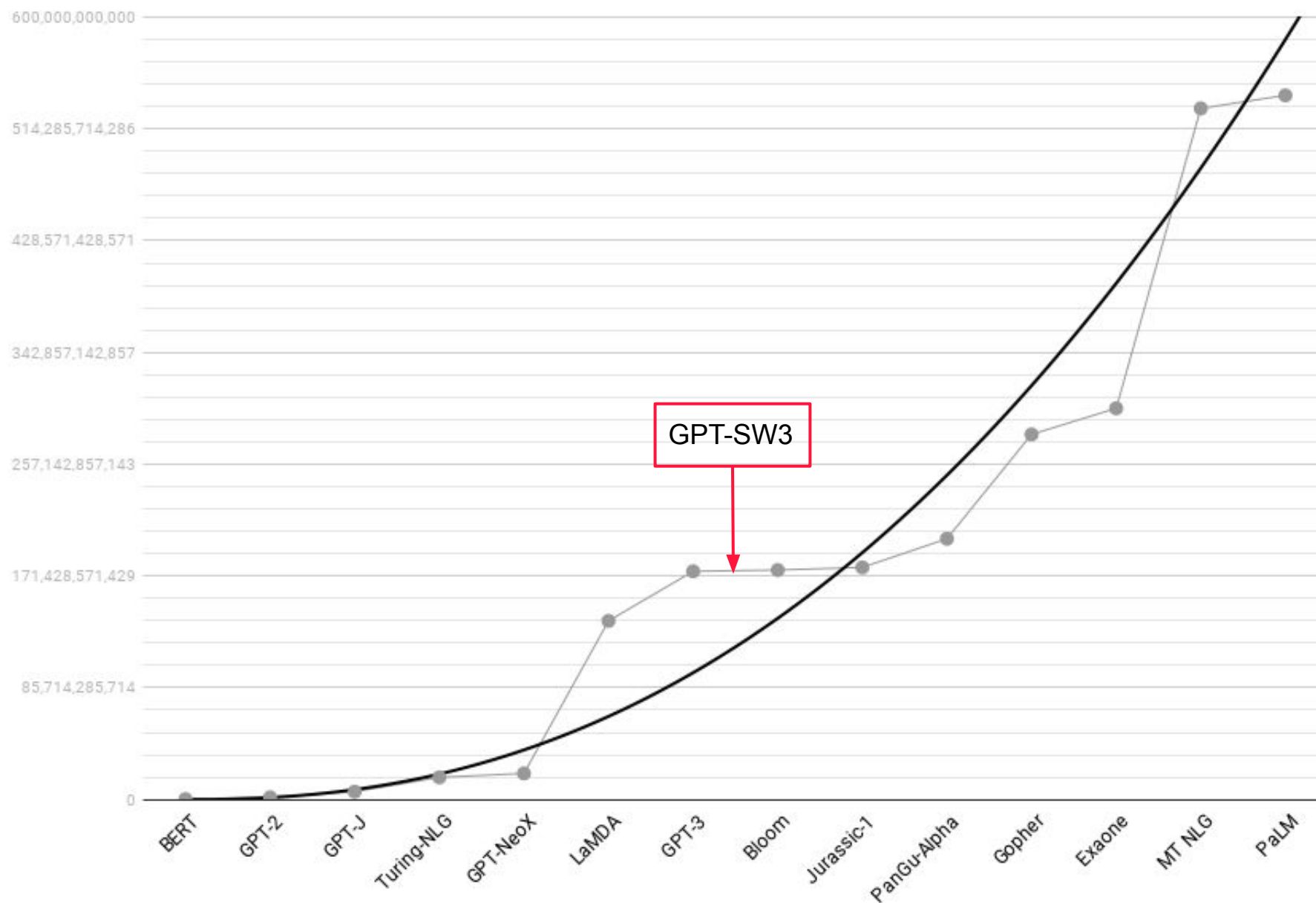
**Each A100 GPU:  
40 GB on-board HBM2 VRAM**

**Nvidia Mellanox Infiniband networking**

## GPT-SW3 versions

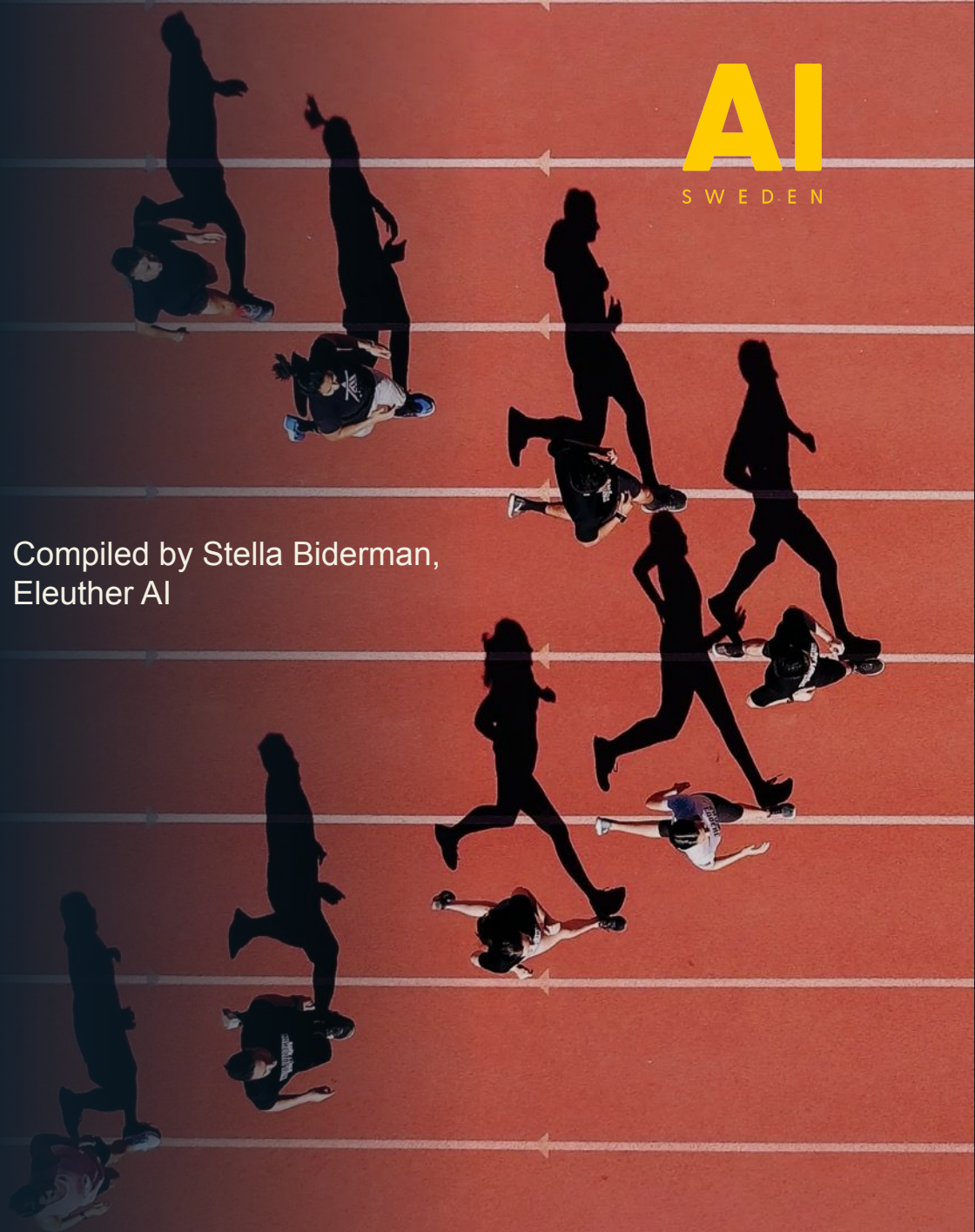


## Transformer model parameters





1		Date	Organization	Author Location	Language	Parameters
2	PaLM	2022-04-04	Google	USA	English	540.0B
3	Megatron-Turing	2021-10-11	Microsoft, NVIDIA	USA	English	530.0B
4	Gopher	2021-12-08	DeepMind	USA	English	280.0B
5	ERNIE 3.0	2021-12-08	Baidu	China	Chinese, English	260.0B
6	Yuan 1.0	2021-10-10	Inspur AI Research	China	Chinese	245.0B
7	HyperCLOVA	2021-09-10	NAVER	Korea	Korean	204.0B
8	PanGu- $\alpha$	2021-04-26	Huawei	China	Chinese	200.0B
9	Jurassic-1	2021-08-11	AI21 Labs	Israel	English	178.0B
10	BLOOM	Training	Big Science	12+ Countries	56 Languages	176.0B
11	GPT-3	2020-05-28	OpenAI	USA	English	175.0B
12	OPT	2021-05-03	Meta AI	USA	English	175.0B
13	LaMDA	2022-01-20	Google	USA	English	137.0B
14	Chinchilla	2022-03-29	DeepMind	USA	English	70.0B
15	Anthropic LM	2021-12-01	Anthropic	USA	English	52.0B
16	GPT-NeoX	2022-02-02	EleutherAI	Germany, USA, India, Canada, UK, Australia, Austria	English	20.0B
17	UL2	2022-05-10	Google	USA	English	20.0B
18	???	???	???	Japan	Japanese	20.0B
19	Turing NLG	2020-02-13	Microsoft	USA	English	17.2B
20	FairSeq Dense	2021-12-20	Meta AI	USA, UK, Germany	English	13.0B
21	BLOOM	2022-01-21	Big Science	Multinational	Multilingual	13.0B
22	mT5	2020-10-22	Google	USA	Multilingual	13.0B
23	ByT5	2021-05-28	Google	USA	Multilingual	13.0B
24	T5	2019-10-23	Google	USA	English	11.0B
25	CPM-2.1	2021-06-20	Tsinghua University	China	Chinese	11.0B
26	Megatron 11B	2020-04-03	NVIDIA	USA	English	11.0B
27	WuDao-GLM-XXL	???	Beijing Academy of Artificial Intelligence	China	Chinese	10.0B
28	WuDao-GLM-XXL	???	Beijing Academy of Artificial Intelligence	China	English	10.0B
29	Lyra-fr	2022-04-12	LightOn	France	French	10.0B
30	NOOR	???	Technology Innovation Institute	UAE	Arabic	10.0B
31	BlenderBot	2020-04-28	Meta AI	USA	English	9.4B
32	Megatron-LM	2019-09-17	NVIDIA	USA	English	8.3B
33	XGLM	2021-12-20	Meta AI	USA, UK, Germany	Multilingual	7.5B
34	GPT-J-6B	2021-06-08	EleutherAI	Australia, USA	English	6.0B
35	Megatron-BERT	2019-09-17	NVIDIA	USA	English	3.9B
36	GPT-SW3	2022-02-15	AI Sweden	Sweden	Swedish	3.5B





# GPT-SW3 validation project (2022-2024)

---

**Validate the use of GPT-SW3 for solving NLP tasks**

**Models, API, and applications**

**Private sector (small and large), public sector, academia**



