

Transformers

History

Jun 2017 - Transformers
 Jan 2018 - ULMFiT
 Jun 2018 - GPT
 Oct 2018 - BERT
 Feb 2019 - GPT-2
 Oct 2019 - BART, T5, DistilBERT
 May 2020 - GPT-3
 May 2021 - GPT-Neo
 Jun 2021 - GPT-J
 Feb 2022 - GPT-NeoX

Encoder-Only Model

- ① Understanding of input
- Sentence Classification
- Named Entity Recognition
- Ex: BERT, RoBERTa, DistilBERT

Decoder Only Model

- ① Generative task
- Ex: GPT, GPT-2.

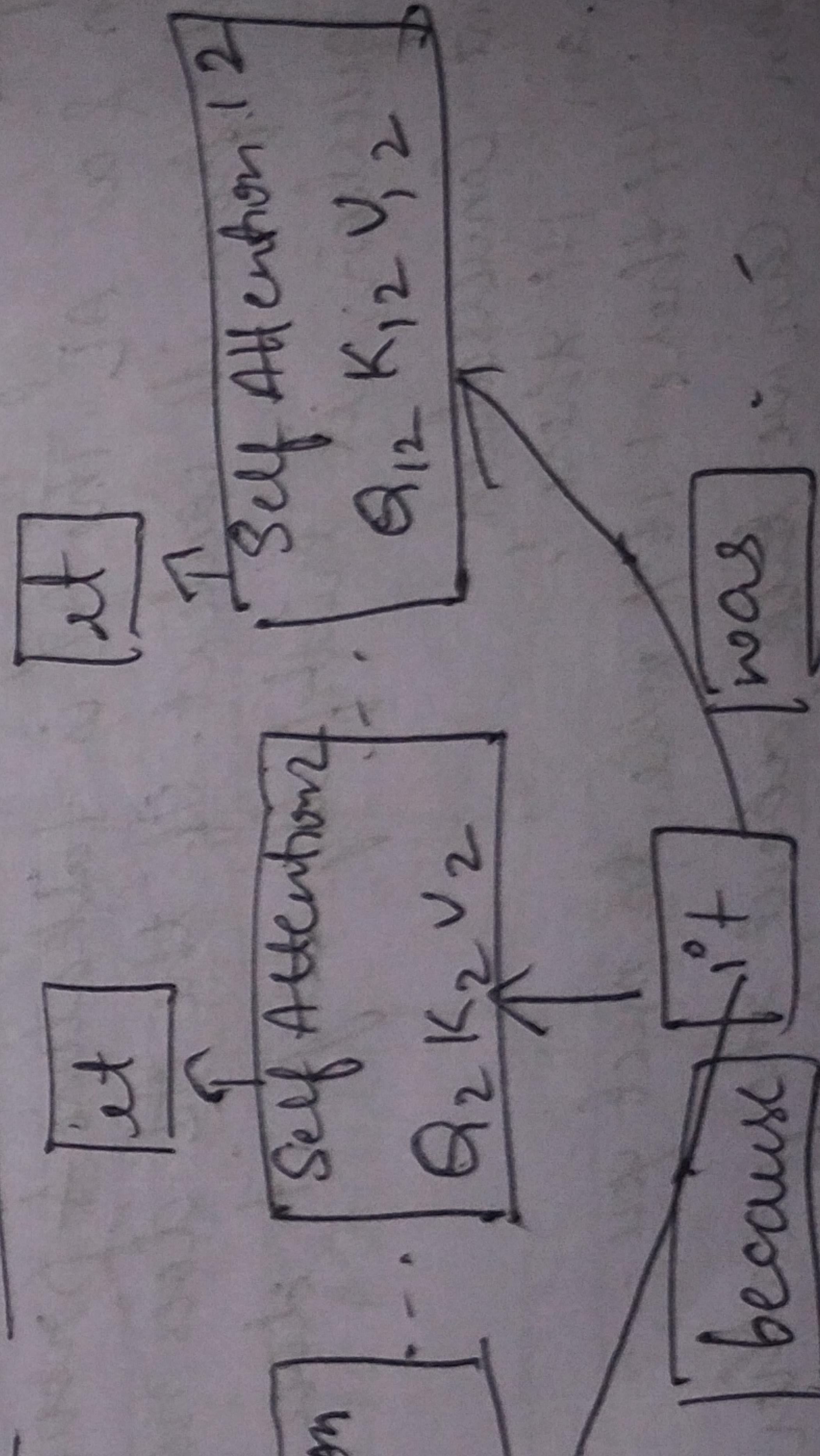
Self Attention

Q (Query), K (Key) V (Value) → dot product of Q & K

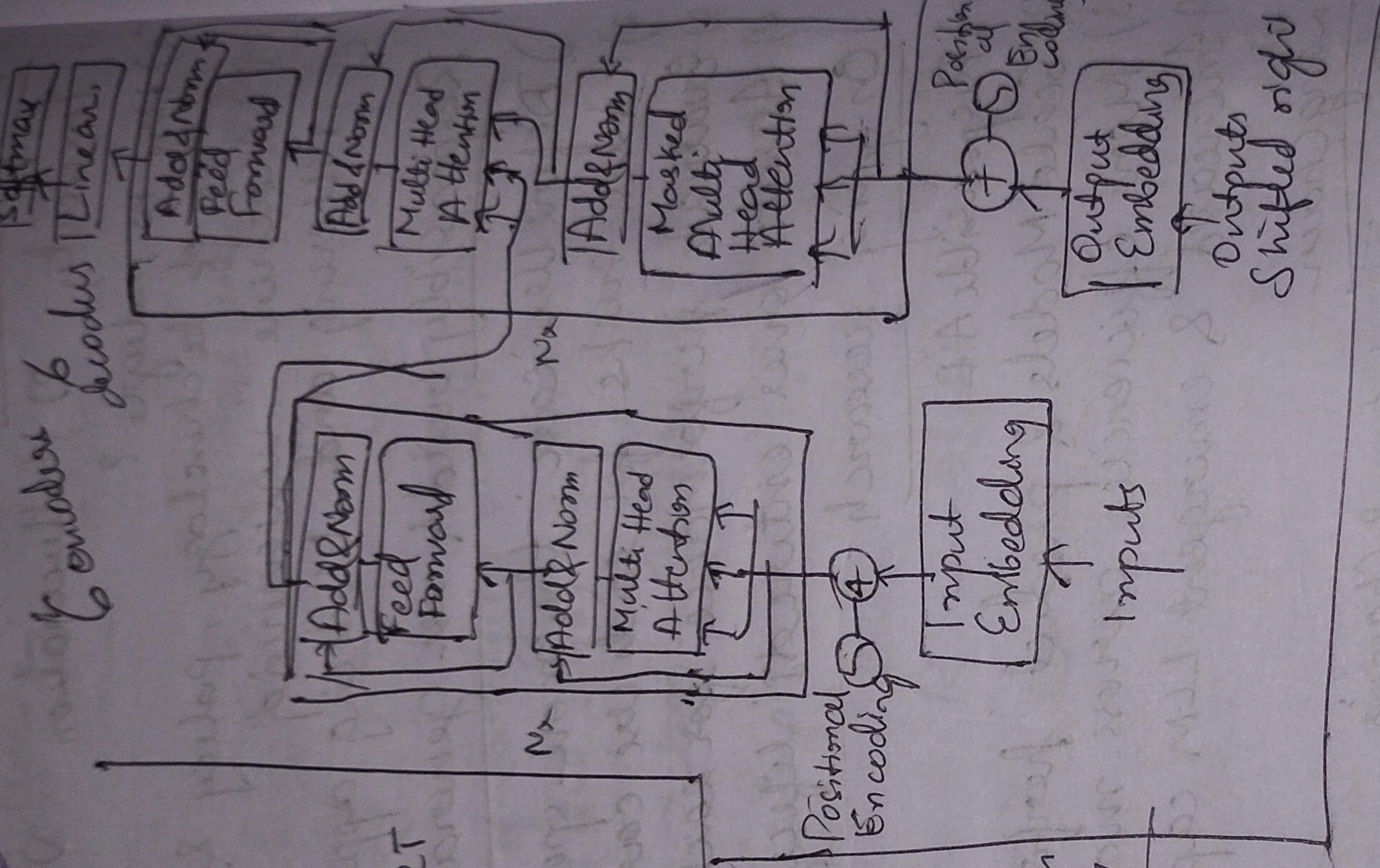
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{n}}\right)V$$

Attention(Q, K, V) = softmax($\frac{QK^T}{\sqrt{n}}$)V

Multi Headed Attention & feed forward Network



Current input word it because



OLP

Encoder Decoder Linear

Add & Norm
Feed Forward

Add & Norm
Multi Head Attention

Generative

① Needs to trace back

② Causal & traced

③ Needs to trace back

④ Prompt to Book

⑤ Way to zero &

⑥ One shot expected

⑦ Few shots

⑧ example challenge

⑨ Bias

⑩ Enigma

⑪ Carbon

⑫ 1/3rd

⑬ large

⑭ type

⑮ dense

⑯ decod

⑰ claim

⑱ no

⑲ 64B/64G

Company	Open AI	Google
Training	200B	280B
Scope	Learning with Large model	Crafting fine model
Model	Machine support	Machine support

Scrophularia 1

- 1) Released by DeepMind from 44M to 280B params
 - 2) Six Models, from 44M to 280B params
 - 3) Massive Text Dataset.
 - = 4) Tested on 152 tasks
 - Tan22 → Gopher → Model performance across a range of model sizes & tasks. In general, larger models & mathematical reasoning tasks better except on logical & param tokens
 - 280B → 300B → DeepMind / google.

occurring laws - model size
presently.

most of our performances all over
modemly

→ has competitive advantage & influence

Chinchilla paper from
Alcommin clothier

• Financial budget in comprehensive financial management

Model size
Scalable in several ways
Tokyo training
No. 1

وَمَنْ

Chinchilla
Gives a fine
giving
Cirqueent
Cloud Moon
Big
-bam

Heads
OPT 8 Dec 1251 π
B 100 100 π
3 1 1 π π π

Dindilla → DeepMind → 70B params - 1.4T tokens
Given a fixed flop budget, how should one trade off model size & the number of training tokens?

Current language models are under-trained]
[use more training tokens to reduce model params]
Bio-bench (Beyond the Imitation Game Benchmark)

Results
→ Model outperformed the best performing human on any task
→ On some task, the next-best performing model beat the average human

PALM :- (April 2022) Pathways Language Models (500B)
Size: 760B tokens
→ 100 languages
→ 100% in English
→ range of perform tasks

Chain of Thought prompting
→ Model trained on Pathway hardware infrastructure - best overall performance on benchmarks to date.

→ 400B tokens

OPT & BLOOM : Open pre-trained transformers → Facebook

→ Decoder only transformer model
→ 125M to 66B shared → open source
→ 145B model available.
BLOOM
→ Tragging face & Montreal AI Ethics Institute
→ 176B decoder-only transformer model
→ Everything openly available
→ Intermediate checkpoints
→ 46 natural languages
→ 13 programming languages

- First language model with 100B+ parameters for many languages (Spanish, French, Arabic)
- Still requires expensive h/w accelerators

Natural Language Processing

- 1) Natural Language Understanding: Mapping the given input from Natural language into formal representation and analysing it.
- 2) Natural Language Generation: Producing meaningful phrases & sentences in the form of Natural language from some internal representation

Applications

Responsible AI

- 1) Be socially beneficial
 - 2) Avoid creating or reinforcing unfair bias & discrimination
 - 3) Be built & tested for safety
 - 4) Be accountable to people
 - 5) Incorporate privacy design principles
 - 6) Uphold high standards of scientific excellence.
 - 7) Be made available for uses that accord with these principles.
- Decisions made at all stages in a project make an impact on responsible AI.
- Transparency, fairness, accountability & privacy -
- Responsible AI practice can help build trust with customers & stakeholders

is

High
1) Fou
lot y
inter