Large Language Models

Year	Model	# parameters	dataset size
2018	ELMo	$9.36 * 10^7$	$\approx$ 6GB *
2019	BERT	$3.4 * 10^8$	16 GB
2019	Megatron-LM	$8.3 * 10^9$	174 GB
2020	GPT-3	$1.75 * 10^{11}$	570 GB
2020	GShard	$6.00 * 10^{11}$	†
2021	Switch-C	$1.57 * 10^{12}$	745 GB
2021	Gopher	$2.8 * 10^{11}$	$\approx$ 1800 GB ‡
2022	PaLM	$5.4 * 10^{11}$	$\approx$ 4680 GB $

* 1 billion words

† 25 billion training examples (100 languages)

‡ 300 billion tokens

$ 780 billion tokens

How Much Data is that?

Someone who reads a 400 page book each day will take how long to read the same amount of information as GPT-3 was trained on?

How Much Data is that?

Someone who reads a 400 page book each day will take how long to read the same amount of information as GPT-3 was trained on?

(a) 2 years

(b) 20 years

(c) 200 years

(d) 2000 years

How Much Data is that?

Someone who reads a 400 page book each day will take how long to read the same amount of information as GPT-3 was trained on?

(a) 2 years

(b) 20 years

(c) 200 years

(d) 2000 years

Answer: 2000 years
Someone who reads a 400 page book each week takes about 30 years to read 180 million words, which is approximately one GB of text.