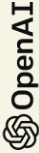






# Comparison of LLM Specifications


20  
24  
MARCH


	Model	Parameters Size	Structure	Training Data (Date)	Output Token	Context Window	Type	Release Date
	GPT-3.5	20B	Transformer	45T ~2021/09	4k	4k~16k	close	2022/3/15
	GPT-4	8*220B Plus, Team, Enterprise	MoE 8 experts	?	4k	32k Plus, Team 128K Enterprise	close	2023/4/4
	PaLM 2	340B Unicorn, Bison, Otter, Gecko	Transformer	3.6T ~2023/8	1k~8k	8K~32K	close	2023/5/10
	Gemma	2B, 7B	Transformer	2T, 6T	?	8K	open source	2024/2/21
	Gemini 1.0	Nano1 1.8B, Nano2 3.3B, Pro, Ultra	Transformer	?	2k	8K	close	2024/2/8
	Gemini 1.5	Pro	MoE	?	8k	128k~1m	close	2024/2/15
	Claude 2.1	200B	Evolved Transformer	? ~2023/Early	4k	200K	close	2023/11/21
	Claude 3	20B, 70B, 2T Haiku, Sonnet, Opus	Sparse Transformer	40T ~2023/8	4k	200K	close	2024/3/4
	Llama2	7B, 13B, 70B	Transformer	2T ~2023/7	4k	4k	open source	2023/7/18
	Mistral Tiny	7B	Transformer	?	?	32k	open source	2023/9/27
	Mixtral 8x7B Small	45B use 12B	SMoE 8 experts	?	?	32k	open source	2023/12/11
	Mistral	Medium, Large	?	?	?	32k	close	2024/2/26

 Certain

 Uncertain

If you find any errors or wish to make an addition, please issue or PR on the Github repo: [wsxqaza12/Comparison-of-LLM-Specifications](#).  
The Github repo includes source links for information that may be uncertain.

 [wsxqaza12](#)

 [ChiChieh Huang](#)