Milestone Papers

Date	keywords	Institute	Paper	Publication
2017-06	Transformers	Google	Attention Is All You Need	NeurIPS
2018-06	GPT 1.0	OpenAI	Improving Language Understanding by Generative Pre-Training	
2018-10	BERT	Google	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	NAACL
2019-02	GPT 2.0	OpenAI	Language Models are Unsupervised Multitask Learners	
2019-09	Megatron-LM	NVIDIA	Megatron-LM: Training Multi-Billion Parameter Language Models Using Model Parallelism	
2019-10	T5	Google	Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer	JMLR

Date	keywords	Institute	Paper	Publication
2019-10	ZeRO	Microsoft	ZeRO: Memory Optimizations Toward Training Trillion Parameter Models	SC
2020-01	Scaling Law	OpenAI	Scaling Laws for Neural Language Models	
2020-05	GPT 3.0	OpenAI	Language models are few-shot learners	NeurIPS
2021-01	Switch Transformers	Google	Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity	JMLR
2021-08	Codex	OpenAI	Evaluating Large Language Models Trained on Code	
2021-08	Foundation Models	Stanford	On the Opportunities and Risks of Foundation Models	
2021-09	FLAN	Google	Finetuned Language Models are Zero-Shot Learners	ICLR

Date	keywords	Institute	Paper	Publication
2021-10	T0	HuggingFace et al.	Multitask Prompted Training Enables Zero-Shot Task Generalization	ICLR
2021-12	GLaM	Google	GLaM: Efficient Scaling of Language Models with Mixture-of-Experts	ICML
2021-12	WebGPT	OpenAI	WebGPT: Browser-assisted question-answering with human feedback	
2021-12	Retro	DeepMind	Improving language models by retrieving from trillions of tokens	ICML
2021-12	Gopher	DeepMind	Scaling Language Models: Methods, Analysis & Insights from Training Gopher	
2022-01	COT	Google	Chain-of-Thought Prompting Elicits Reasoning in Large Language Models	NeurIPS
2022-01	LaMDA	Google	LaMDA: Language	

Date	keywords	Institute	Paper	Publication
			Models for Dialog Applications	
2022-01	Minerva	Google	Solving Quantitative Reasoning Problems with Language Models	NeurIPS
2022-01	Megatron-Turing NLG	Microsoft&NVIDIA	Using Deep and Megatron to Train Megatron-Turing NLG 530B, A Large-Scale Generative Language Model	
2022-03	InstructGPT	OpenAI	Training language models to follow instructions with human feedback	
2022-04	PaLM	Google	PaLM: Scaling Language Modeling with Pathways	
2022-04	Chinchilla	DeepMind	An empirical analysis of compute-optimal large language model training	NeurIPS
2022-05	OPT	Meta	OPT: Open Pre-trained Transformer Language Models	

Date	keywords	Institute	Paper	Publication
2022-05	UL2	Google	Unifying Language Learning Paradigms	ICLR
2022-06	Emergent Abilities	Google	Emergent Abilities of Large Language Models	TMLR
2022-06	BIG-bench	Google	Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models	
2022-06	METALM	Microsoft	Language Models are General-Purpose Interfaces	
2022-09	Sparrow	DeepMind	Improving alignment of dialogue agents via targeted human judgements	
2022-10	Flan-T5/PaLM	Google	Scaling Instruction-Finetuned Language Models	
2022-10	GLM-130B	Tsinghua	GLM-130B: An Open Bilingual Pre-trained Model	ICLR
2022-11	HELM	Stanford	Holistic Evaluation of	

Date	keywords	Institute	Paper	Publication
			Language Models	
2022-11	BLOOM	BigScience	BLOOM: A 176B-Parameter Open-Access Multilingual Language Model	
2022-11	Galactica	Meta	Galactica: A Large Language Model for Science	
2022-12	OPT-IML	Meta	OPT-IML: Scaling Language Model Instruction Meta Learning through the Lens of Generalization	
2023-01	Flan 2022 Collection	Google	The Flan Collection: Designing Data and Methods for Effective Instruction Tuning	ICML
2023-02	LLaMA	Meta	LLaMA: Open and Efficient Foundation Language Models	
2023-02	Kosmos-1	Microsoft	Language Is Not All You Need: Aligning Perception with Language Models	

Date	keywords	Institute	Paper	Publication
2023-03	PaLM-E	Google	PaLM-E: An Embodied Multimodal Language Model	ICML
2023-03	GPT 4	OpenAI	GPT-4 Technical Report	
2023-04	Pythia	EleutherAI et al.	Pythia: A Suite for Analyzing Large Language Models Across Training and Scaling	ICML
2023-05	Dromedary	CMU et al.	Principle-Driven Self-Alignment of Language Models from Scratch with Minimal Human Supervision	NeurIPS
2023-05	PaLM 2	Google	PaLM 2 Technical Report	
2023-05	RWKV	Bo Peng	RWKV: Reinventing RNNs for the Transformer Era	EMNLP
2023-05	DPO	Stanford	Direct Preference Optimization: Your Language Model is Secretly a Reward Model	Neurips
2023-05	ToT	Google&Princeton	Tree of Thoughts: Deliberate Problem Solving	NeurIPS

Date	keywords	Institute	Paper	Publication
			with Large Language Models	
2023-07	LLaMA 2	Meta	Llama 2: Open Foundation and Fine-Tuned Chat Models	
2023-10	Mistral 7B	Mistral	Mistral 7B	
2023-12	Mamba	CMU&Princeton	Mamba: Linear-Time Sequence Modeling with Selective State Spaces	ICML
2024-03	Jamba	AI21 Labs	Jamba: A Hybrid Transformer-Mamba Language Model	