

****1 Billion & UNDER****

1. Qwen3.5-0.8B (262k)
2. Granite 4.0 1B (128k)
3. Llama 3.2-1B Instruct (128k)
4. MiniCPM4-0.5B (128k)
5. Granite 3.1 1B-A400M (128k)
6. Gemma 3-1B-Instruct (128k)
7. DeepSeek-V3.2-Lite-1B (128k)
8. Qwen3-1.5B-Coder (128k)
9. OLMo 2-1B (128k)
10. Qwen3-0.6B (128k)
11. Gemma 3 1B IT (262k)
12. Hunyuan-0.5B-Pretrain (256k)
13. Gemma 3 270M (128k)
14. SmolLM3-0.5B (128k)
15. Phi-3.5-mini (128k)
16. Qwen3.5-0.5B (128k)
17. StableLM-2-1.6B (128k)
18. OLMo 2-0.5B (128k)
19. DeepSeek-V3-Lite-0.5B (128k)
20. MiniCPM4-0.4B (128k)

****3 Billion to 5 Billion****

1. Nanbeige4.1-3B (1M)
2. Phil 4 Mini 3.4B (128k)
3. SmolLM3-3B (128k)
4. Qwen3.5-3B (128k)
5. Ministral-3-3B-Instruct (128k)
6. Gemma 3-4B (128k)
7. Phi-4-mini 3.8B (128k)
8. Llama 3.2-3B (128k)
9. DeepSeek-V3.2-Lite-3B (128k)

****7 Billion To 10 Billion****

1. Qwen3-8B (128k)
2. DeepSeek-R1-Distill-Qwen-7B (128k)
3. Qwen3.5-8B (128k)
4. Llama 3.3-8B (128k)
5. Ministral-8B (128k)
6. Mistral Small 24B (trimmed 8B active) (128k)
7. Gemma 3-9B (128k)
8. Phi-4-9B (128k)
9. OLMo 2-7B (128k)
10. StableLM-2-7B (128k)

11. InternLM 3-8B (128k)
12. Qwen3.5-7B (128k)
13. DeepSeek-V3-Lite-7B (128k)
14. SmolLM3-7B (128k)
15. Llama 4 Scout-8B (128k)
16. Gemma 3-7B (128k)
17. Phi-4-8B (128k)
18. Qwen2.5-7B (128k)
19. Mistral-7B-Instruct (128k)
20. Granite 3.1-8B (128k)

Top 40 LLMs from Kaggle/GitHub/Hugging Face (finetuned/trained on most respected datasets – first 20; most datasets – last 20)

All open-source, 128k+ context, fine-tunable, top benchmarks.

1. Qwen3.5-0.8B (Sushi-Coder)
2. JetBrains/Mellum-4b-dpo-all
3. SmolLM3-3B (multi-dataset)
4. DeepSeek-R1-Distill-Qwen-7B
5. Gemma 3-1B (FineWeb-Edu)
6. Llama 3.3-8B (Tulu)
7. Phi-4-mini 3.8B
8. Qwen3-8B
9. Ministral-3-3B
10. OLMo 2-7B
11. Granite 3.1-8B
12. InternLM 3-8B
13. Mistral Small 8B active
14. Gemma 3-4B
15. DeepSeek-V3.2-Lite-3B
16. Qwen3.5-3B
17. SmolLM3-7B
18. Llama 4 Scout-8B
19. Phi-4-9B
20. Qwen3.5-7B
21. Tulu-3-8B (multi)
22. OpenHermes-7B (multi)
23. Nemotron-Post-Training-9B (multi)
24. WildChat-4.8M finetune
25. HelpSteer-7B
26. Cosmopedia-3B
27. UltraChat-7B
28. OpenOrca-7B
29. Ling-Coder-SFT-8B
30. smoltalk2-3B

31. FineWeb-Edu-7B
32. RefinedWeb-7B
33. RedPajama-8B
34. Dolma-7B
35. C4-8B
36. The Pile-7B
37. SQuAD-7B
38. NVIDIA HelpSteer-8B
39. BigCode StarCoder2-7B
40. InclusionAI Ling-Coder-8B

****20 MOST RESPECTED DATASETS****

1. bigatuna/Qwen3-0.6B-Sushi-Coder
2. JetBrains/Mellum-4b-dpo-all
3. The Pile
4. C4 (Colossal Clean Crawled Corpus)
5. FineWeb-Edu
6. RefinedWeb
7. RedPajama
8. Dolma
9. Common Crawl (filtered)
10. Wikipedia
11. Books3
12. ArXiv
13. StackExchange
14. UltraChat
15. OpenOrca
16. Tulu
17. HelpSteer
18. Cosmopedia
19. WildChat
20. SQuAD

****20 LLMs FineTuned on multi DATASETS****

1. SmoLLM3-3B
2. Tulu-3-8B
3. OpenHermes-7B
4. Nemotron-Post-Training-9B
5. WildChat finetunes
6. HelpSteer-7B
7. Cosmopedia-3B
8. UltraChat-7B
9. OpenOrca-7B
10. Ling-Coder-SFT-8B

11. smoltalk2-3B
12. FineWeb-Edu-7B
13. RefinedWeb-7B
14. RedPajama-8B
15. Dolma-7B
16. C4-8B
17. The Pile-7B
18. SQuAD-7B
19. NVIDIA HelpSteer-8B
20. BigCode StarCoder2-7B

****Top 5 LLMs best for Mixture-Of-Experts merge****

1. Qwen3.5-122B-A10B (MoE base)
2. DeepSeek-V3.2 (MoE)
3. gpt-oss-120B (MoE)
4. MiMo-V2-Flash (MoE)
5. Llama 4 Maverick (MoE)

****Top 5 LLMs that are the best of the best****

1. DeepSeek-V3.2
2. Qwen3.5-397B-A17B
3. GLM-5
4. Kimi K2.5
5. Gemma 3 27B

https://grok.com/share/bGVnYWN5LWNvcHk_f503fefd-da0d-4241-a3e3-666b5aaf2863