FOOTNOTE URLS

[2]https://aws.amazon.com/blogs/machine-learning/fmops-llmops-operationalize-generative-ai-and-differences-with-mlops
[3]https://huggingface.co/spaces
[4]https://pages.github.com
[5]https://paperswithcode.com
[6]https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard/discussions/801
[7]https://huggingface.co/blog/open-llm-leaderboard-drop
[8]https://huggingface.co/spaces/open-llm-leaderboard/blog
[9]https://github.com/topics/awesome-list
[10]https://scholar.google.com
[11]https://github.com/SAILResearch/awesome-foundation-model-leaderboards
[12]https://huggingface.co/spaces/zhiminy/awesome-foundation-model-leaderboard-search
[13]https://chat.lmsys.org/?leaderboard
[14]https://www.datacomp.ai/dclm/leaderboard.html
[15]https://www.datacomp.ai/dcclip/leaderboard.html
[16]https://github.com/yuh-zha/AlignScore
[17]https://huggingface.co/spaces/locuslab/tofu_leaderboard
[18]https://huggingface.co/spaces/ameerazam08/Paper-LeaderBoard
[19]https://leaderboard.withmartian.com
[20]https://www.kaggle.com/competitions
[21]https://eval.ai
[22]https://taostats.io
[23]https://huggingface.co/spaces/mrfakename/open-leaderboards-leaderboard
[24]https://super.gluebenchmark.com/leaderboard
[25]https://neptune.ai/blog/llmops
[26]https://github.com/paperswithcode/paperswithcode-data
[27]https://huggingface.co/spaces?sort=trending&search=leaderboard
[28]https://huggingface.co/spaces/Vikhrmodels/Russian_Arena_Hard/discussions/3
[29]https://huggingface.co/spaces?search=chatbot+arena+leaderboard
[30]https://github.com/Helsinki-NLP/OPUS-MT-dashboard/issues/2
[31]https://huggingface.co/spaces/HuggingFaceH4/human_eval_llm_leaderboard/discussions/4
[32]https://www.superclueai.com
[33]https://github.com/SAILResearch/awesome-foundation-model-leaderboards
[34]https://huggingface.co/spaces/zhiminy/awesome-foundation-model-leaderboard-search
[35]https://slack.com
[36]https://discord.com
[37]https://weixin.qq.com
[38]https://github.com/THU-KEG/KoLA/issues/17
[39]https://github.com/embeddings-benchmark/mteb/pull/187
[40]https://github.com/LudwigStumpp/llm-leaderboard
[41]https://tatsu-lab.github.io/alpaca_eval
[42]https://huggingface.co/spaces/bigcode/bigcode-models-leaderboard
[43]https://github.com/paperswithcode/axcell
[44]https://huggingface.co/spaces/gaia-benchmark/leaderboard
[45]https://leaderboard.allenai.org/a-okvqa/submissions/public
[46]https://cmedbenchmark.llmzoo.com/static/leaderboard.html
[47]https://opencompass.org.cn/doc
[48]https://jwolpxeehx.feishu.cn/wiki/C6VfwvbmOiuVrokpJAgcJXUcnLh
[49]https://docs.google.com/document/d/1Rs6d_pcs2vfqW4Ymub7Wb1XtBNlrc-WfH3T7U1ktuBo
[50]https://github.com/stars
[51]https://huggingface.co/spaces/NexaAIDev/domain_llm_leaderboard
[52]https://flageval.baai.ac.cn/#/leaderboard
[53]https://cmmmu-benchmark.github.io/#leaderboard
[54]https://research.ibm.com/blog/efficient-llm-benchmarking
[55]https://chat.lmsys.org
[56]https://huggingface.co/spaces/Auto-Arena/Leaderboard
[57]https://rank.opencompass.org.cn/leaderboard-arena
[58]https://huggingface.co/spaces/llm-council/emotional-intelligence-arena
[59]https://bryanyzhu.github.io/posts/2024-06-20-elo-part1
[60]https://arena.lmsys.org
[61]https://twitter.com/eval_ai/status/1341091250544521231
[62]https://github.com/paperswithcode/model-index
[63]https://github.com/rowanz/hellaswag/tree/master/hellaswag_models#submitting-to-the-leaderboard
[64]https://github.com/tatsu-lab/alpaca_eval?tab=readme-ov-file#contributing-a-model
[65]https://github.com/ray-project/llmperf-leaderboard?tab=readme-ov-file#feedback
[66]https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard
[67]https://huggingface.co/spaces/Cognitive-Lab/indic_llm_leaderboard
[68]https://crfm.stanford.edu/helm/lite/latest/#/leaderboard
[69]https://stats.stackexchange.com/questions/333446/what-does-the-term-gold-label-refer-to-in-the-context-of-semi-supervised-class
[70]https://github.com/EleutherAI/lm-evaluation-harness
[71]https://github.com/lm-sys/FastChat
[72]https://tatsu-lab.github.io/alpaca_eval
[73]https://huggingface.co/spaces/gsaivinay/open_llm_leaderboard
[74]https://github.com/MikeGu721/XiezhiBenchmark
[75]https://leaderboard.tabbyml.com
[76]https://artificialanalysis.ai/leaderboards/models
[77]https://videoniah.github.io
[78]https://xwang.dev/mint-bench
[79]https://osu-nlp-group.github.io/TravelPlanner
[80]https://huggingface.co/spaces/BramVanroy/open_dutch_llm_leaderboard
[81]https://huggingface.co/spaces/ml-energy/leaderboard
[82]https://huggingface.co/spaces/ramiroluo/LLMHallucination_Leaderboard
[83]https://huggingface.co/spaces/dimbyTa/open-llm-leaderboard-viz
[84]https://huggingface.co/spaces/upstage/open-ko-llm-leaderboard/discussions/28
[85]https://github.com/GAIR-NLP/factool/issues/39
[86]https://github.com/microsoft/promptbench/issues/26
[87]https://github.com/nyu-mll/jiant/issues/1366
[88]https://github.com/TheoremOne/llm-benchmarker-suite/issues/10
[89]https://huggingface.co/spaces/lovodkin93/FuseReviews-Leaderboard/discussions/1
[90]https://huggingface.co/spaces/nlphuji/WHOOPS-Leaderboard-Full/discussions/2
[91]https://huggingface.co/spaces/Nexusflow/Nexus_Function_Calling_Leaderboard/discussions/3
[92]https://huggingface.co/spaces/SeaEval/SeaEval_Leaderboard/discussions/2
[93]https://github.com/open-compass/LawBench/issues/7
[94]https://huggingface.co/spaces/mike-ravkine/can-ai-code-results/discussions/4
[95]https://huggingface.co/spaces/mike-ravkine/can-ai-code-results/discussions/6
[96]https://huggingface.co/spaces/mike-ravkine/can-ai-code-results/discussions/9
[97]https://github.com/tjunlp-lab/M3KE/issues/8
[98]https://github.com/open-compass/MathBench/issues/7
[99]https://github.com/BradyFU/Awesome-Multimodal-Large-Language-Models/issues/124
[100]https://github.com/hendrycks/test/issues/25
[101]https://github.com/princeton-nlp/SWE-bench/issues/46
[102]https://github.com/CLUEbenchmark/SuperCLUE/issues/43
[103]https://github.com/CLUEbenchmark/SuperCLUE-Agent/issues/8
[104]https://github.com/CLUEbenchmark/SuperCLUE-Auto/issues/3
[105]https://github.com/CLUEbenchmark/SuperCLUE-Math6/issues/1
[106]https://github.com/CLUEbenchmark/SuperCLUE-Safety/issues/8
[107]https://github.com/CLUEbenchmark/SuperCLUEgkzw/issues/2
[108]https://github.com/CLUEbenchmark/SuperCLUE-RAG/issues/2
[109]https://github.com/CLUEbenchmark/SuperCLUE-Code3/issues/3
[110]https://github.com/CLUEbenchmark/SuperCLUE-Role/issues/2
[111]https://github.com/CLUEbenchmark/SuperCLUE-Industry/issues/1
[112]https://github.com/EleutherAI/lm_perplexity/issues/2
[113]https://huggingface.co/spaces/mike-ravkine/can-ai-code-results/discussions/7
[114]https://huggingface.co/spaces/mike-ravkine/can-ai-code-results/discussions/3
[115]https://github.com/stanford-crfm/helm/issues/2386
[116]https://github.com/hkust-nlp/ceval/issues/76
[117]https://huggingface.co/spaces/daishen/SCULAiW/discussions/1
[118]https://github.com/MikeGu721/XiezhiBenchmark/issues/6
[119]https://github.com/OpenLMLab/LEval/issues/10
[120]https://huggingface.co/spaces/ought/raft-leaderboard/discussions/20
[121]https://github.com/taoyds/spider/issues/101
[122]https://github.com/embeddings-benchmark/mteb/issues/186
[123]https://github.com/Helsinki-NLP/OPUS-MT-dashboard/issues/2
[124]https://huggingface.co/spaces/opencompass/open_vlm_leaderboard/discussions/6
[125]https://github.com/LaVi-Lab/CLEVA/issues/9
[126]https://github.com/FlagOpen/FlagEval/issues/36
[127]https://github.com/palash1992/GEM-Benchmark/issues/4

[128]https://github.com/YangLinyi/GLUE-X/issues/4
[129]https://github.com/OpenGVLab/Multi-Modality-Arena/issues/19
[130]https://github.com/spraakbanken/SuperLim-2/issues/1
[131]https://huggingface.co/spaces/nlphuji/WHOOPS-Leaderboard-Full/discussions/1
[132]https://huggingface.co/spaces/mlfoundations/VisIT-Bench-Leaderboard/discussions/1
[133]https://github.com/paperswithcode/sota-extractor/issues/25
[134]https://huggingface.co/spaces/mlabonne/Yet_Another_LLM_Leaderboard/discussions/8
[135]https://github.com/Q-Future/Q-Bench/issues/11
[136]https://github.com/stanford-crfm/helm/issues/2034
[137]https://github.com/stanford-crfm/helm/issues/2060
[138]https://github.com/LudwigStumpp/llm-leaderboard/issues/9
[139]https://github.com/google/BIG-bench/issues/983
[140]https://web.archive.org/web/20221205184917/https://paperswithcode.com/sota/image-classification-on-humaneval
[141]https://github.com/LudwigStumpp/llm-leaderboard/issues/10
[142]https://github.com/FlagOpen/FlagEval/issues/34
[143]https://github.com/THU-KEG/KoLA/issues/5
[144]https://github.com/OpenM3D/M3DBench/issues/2
[145]https://github.com/OpenLMLab/GAOKAO-Bench/issues/25
[146]https://github.com/thodan/bop_toolkit/issues/119
[147]https://github.com/princeton-nlp/intercode/issues/23
[148]https://github.com/stanford-crfm/helm/issues/2238
[149]https://github.com/stanford-crfm/helm/issues/2008
[150]https://github.com/stanford-crfm/helm/issues/2238
[151]https://huggingface.co/spaces/Wwwduojin/MLLM_leaderboard/discussions/1
[152]https://paperswithcode.com/sota/zero-shot-video-question-answer-on-star
[153]https://paperswithcode.com/sota/zero-shot-video-question-answer-on-star-1
[154]https://github.com/THU-KEG/KoLA/issues/6
[155]https://github.com/THU-KEG/KoLA/issues/7
[156]https://github.com/THU-KEG/KoLA/issues/2
[157]https://huggingface.co/spaces/OpenGVLab/MVBench_Leaderboard/discussions/4
[158]https://huggingface.co/spaces/OpenGVLab/MVBench_Leaderboard/discussions/3
[159]https://github.com/THU-KEG/KoLA/issues/17
[160]https://github.com/stanford-crfm/helm/issues/2351
[161]https://huggingface.co/docs/huggingface_hub/package_reference/cards
[162]https://huggingface.co/spaces/demo-leaderboard-backend/leaderboard
[163]https://huggingface.co/spaces/freddyaboulton/gradio_leaderboard
[164]https://hub.opencompass.org.cn/dataset-submit
[165]https://github.com/xingyaoww/mint-bench/blob/main/docs/CONTRIBUTING.md#contribute-dataset
[166]https://github.com/tatsu-lab/alpaca_eval?tab=readme-ov-file#contributing-an-eval-set
[167]https://github.com/llm-council/llm-council?tab=readme-ov-file#adding-a-prompt
[168]https://github.com/svilupp/Julia-LLM-Leaderboard#contributing-your-test-case
[169]https://github.com/tatsu-lab/alpaca_eval?tab=readme-ov-file#contributing-an-evaluator
[170]https://github.com/tatsu-lab/alpaca_eval?tab=readme-ov-file#contributing-a-completion-function
[171]https://twitter.com/karpathy/status/1737544497016578453
[172]https://twitter.com/Nils_Reimers/status/1777653735192556621
[173]https://huggingface.co/spaces/andrewrreed/closed-vs-open-arena-elo