

Task Name

Task Name	Task Category				
	math	multilingual	instruction_following	coding	safety
humaneval_plus	0.530	0.323	0.287	0.006	0.409
mbpp_plus	0.643	0.648	0.630	0.000	0.659
ifeval	0.347	0.029	0.182	0.855	0.301
gsm8k_cot	0.519	0.552	0.570	0.795	0.677
arc_de	0.423	0.462	0.437	0.430	0.430
arc_es	0.448	0.493	0.471	0.448	0.442
arc_fr	0.459	0.490	0.469	0.444	0.446
arc_ru	0.425	0.453	0.422	0.397	0.414
hellaswag_de	0.625	0.633	0.608	0.578	0.618
hellaswag_es	0.689	0.698	0.678	0.610	0.686
hellaswag_fr	0.665	0.677	0.652	0.603	0.660
hellaswag_ru	0.603	0.604	0.585	0.555	0.594
m_mmlu_de	0.497	0.539	0.526	0.490	0.525
m_mmlu_es	0.518	0.556	0.557	0.550	0.552
m_mmlu_fr	0.511	0.550	0.539	0.490	0.540
m_mmlu_ru	0.471	0.516	0.510	0.407	0.510
toxigen	0.432	0.430	0.432	0.843	0.432
truthfulqa	0.487	0.513	0.451	0.551	0.547
winogender	0.633	0.618	0.607	0.694	0.625
Model					
MergeBench/Llama-3.1-8B_coding					
MergeBench/Llama-3.1-8B_multilingual					
meta-llama/Llama-3.1-8B					
meta-llama/Llama-3.1-8B-Instruct					
mlfoundations-dev/seed_math_allenai_math					