

Task Name	Model					
	MergeBench/Llama-3.1-8B-Instruct_coding	MergeBench/Llama-3.1-8B-Instruct_instruction	MergeBench/Llama-3.1-8B-Instruct_math	MergeBench/Llama-3.1-8B-Instruct_multilingual	MergeBench/Llama-3.1-8B-Instruct_safety	meta-llama/Llama-3.1-8B-Instruct
humaneval_plus	0.390	0.000	0.000	0.037	0.433	0.006
mbpp_plus	0.638	0.000	0.193	0.050	0.000	0.000
ifeval	0.621	0.851	0.356	0.747	0.464	0.855
gsm8k_cot	0.625	0.018	0.170	0.752	0.004	0.795
arc_de	0.431	0.400	0.356	0.462	0.367	0.430
arc_es	0.430	0.413	0.367	0.485	0.382	0.448
arc_fr	0.464	0.432	0.362	0.475	0.377	0.444
arc_ru	0.388	0.386	0.318	0.426	0.358	0.397
hellaswag_de	0.594	0.539	0.528	0.608	0.535	0.578
hellaswag_es	0.643	0.576	0.588	0.661	0.596	0.610
hellaswag_fr	0.627	0.569	0.579	0.650	0.574	0.603
hellaswag_ru	0.569	0.520	0.506	0.581	0.520	0.555
m_mmlu_de	0.520	0.440	0.324	0.546	0.350	0.490
m_mmlu_es	0.541	0.498	0.291	0.569	0.336	0.550
m_mmlu_fr	0.541	0.444	0.302	0.550	0.359	0.490
m_mmlu_ru	0.495	0.384	0.276	0.529	0.308	0.407
toxigen	0.820	0.841	0.537	0.861	0.717	0.843
truthfulqa	0.514	0.551	0.466	0.530	0.471	0.551
winogender	0.667	0.551	0.619	0.690	0.542	0.694