

Task Name

Task Name	Model			
	Qwen/Qwen2.5-1.5B	Qwen/Qwen2.5-1.5B-Instruct	Qwen/Qwen2.5-Coder-1.5B	Qwen/Qwen2.5-Math-1.5B
humaneval_plus	0.305	0.311	0.354	0.262
mbpp_plus	0.598	0.569	0.762	0.571
ifeval	0.317	0.468	0.317	0.248
gsm8k_cot	0.674	0.620	0.613	0.768
arc_de	0.314	0.326	0.289	0.295
arc_es	0.351	0.385	0.312	0.326
arc_fr	0.358	0.376	0.325	0.310
arc_ru	0.315	0.322	0.302	0.289
hellaswag_de	0.432	0.442	0.383	0.324
hellaswag_es	0.516	0.527	0.452	0.371
hellaswag_fr	0.498	0.512	0.437	0.359
hellaswag_ru	0.455	0.464	0.394	0.323
m_mmlu_de	0.475	0.483	0.398	0.353
m_mmlu_es	0.513	0.523	0.429	0.374
m_mmlu_fr	0.503	0.512	0.416	0.368
m_mmlu_ru	0.456	0.468	0.385	0.337
toxigen	0.462	0.433	0.522	0.395
truthfulqa	0.467	0.467	0.440	0.449
winogender	0.569	0.561	0.537	0.526

Task Category	
math	
multilingual	
instruction_following	
coding	
safety	