

Table 7: Comparison among Qwen3-4B-Base and other strong open-source baselines. The highest and second-best scores are shown in bold and underlined, respectively.

	Gemma-3-4B Base	Qwen2.5-3B Base	Qwen2.5-7B Base	Qwen3-4B Base
Architecture	Dense	Dense	Dense	Dense
# Total Params	4B	3B	7B	4B
# Activated Params	4B	3B	7B	4B
<i>General Tasks</i>				
MMLU	59.51	65.62	74.16	72.99
MMLU-Redux	56.91	63.68	71.06	72.79
MMLU-Pro	29.23	34.61	45.00	50.58
SuperGPQA	17.68	20.31	26.34	28.43
BBH	51.70	56.30	70.40	72.59
<i>Math & STEM Tasks</i>				
GPQA	24.24	26.26	<u>36.36</u>	36.87
GSM8K	43.97	79.08	<u>85.36</u>	87.79
MATH	26.10	42.64	<u>49.80</u>	54.10
<i>Coding Tasks</i>				
EvalPlus	43.23	46.28	<u>62.18</u>	63.53
MultiPL-E	28.06	39.65	<u>50.73</u>	53.13
MBPP	46.40	54.60	<u>63.40</u>	67.00
CRUX-O	34.00	36.50	<u>48.50</u>	55.00
<i>Multilingual Tasks</i>				
MGSM	33.11	47.53	<u>63.60</u>	67.74
MMMLU	59.62	65.55	<u>71.34</u>	71.42
INCLUDE	49.06	45.90	<u>53.98</u>	56.29

Table 8: Comparison among Qwen3-1.7B-Base, Qwen3-0.6B-Base, and other strong open-source baselines. The highest and second-best scores are shown in bold and underlined, respectively.

	Qwen2.5-0.5B Base	Qwen3-0.6B Base	Gemma-3-1B Base	Qwen2.5-1.5B Base	Qwen3-1.7B Base
Architecture	Dense	Dense	Dense	Dense	Dense
# Total Params	0.5B	0.6B	1B	1.5B	1.7B
# Activated Params	0.5B	0.6B	1B	1.5B	1.7B
<i>General Tasks</i>					
MMLU	47.50	52.81	26.26	<u>60.90</u>	62.63
MMLU-Redux	45.10	51.26	25.99	<u>58.46</u>	61.66
MMLU-Pro	15.69	24.74	9.72	<u>28.53</u>	36.76
SuperGPQA	11.30	15.03	7.19	<u>17.64</u>	20.92
BBH	20.30	41.47	28.13	<u>45.10</u>	54.47
<i>Math & STEM Tasks</i>					
GPQA	24.75	<u>26.77</u>	24.75	24.24	28.28
GSM8K	41.62	59.59	2.20	<u>68.54</u>	75.44
MATH	19.48	32.44	3.66	<u>35.00</u>	43.50
<i>Coding Tasks</i>					
EvalPlus	31.85	36.23	8.98	<u>44.80</u>	52.70
MultiPL-E	18.70	24.58	5.15	<u>33.10</u>	42.71
MBPP	29.80	36.60	9.20	<u>43.60</u>	55.40
CRUX-O	12.10	27.00	3.80	<u>29.60</u>	36.40
<i>Multilingual Tasks</i>					
MGSM	12.07	30.99	1.74	<u>32.82</u>	50.71
MMMLU	31.53	50.16	26.57	<u>60.27</u>	63.27
INCLUDE	24.74	34.26	25.62	<u>39.55</u>	45.57