

Table 3: Comparison among Qwen3-235B-A22B-Base and other representative strong open-source baselines. The highest, the second-best scores are shown in bold and underlined, respectively.

	Qwen2.5-72B Base	Qwen2.5-Plus Base	Llama-4-Maverick Base	DeepSeek-V3 Base	Qwen3-235B-A22B Base
Architecture	Dense	MoE	MoE	MoE	MoE
# Total Params	72B	271B	402B	671B	235B
# Activated Params	72B	37B	17B	37B	22B
<i>General Tasks</i>					
MMLU	86.06	85.02	85.16	87.19	87.81
MMLU-Redux	83.91	82.69	84.05	86.14	87.40
MMLU-Pro	58.07	63.52	<u>63.91</u>	59.84	68.18
SuperGPQA	36.20	37.18	40.85	<u>41.53</u>	44.06
BBH	<u>86.30</u>	85.60	83.62	86.22	88.87
<i>Math & STEM Tasks</i>					
GPQA	<u>45.88</u>	41.92	43.94	41.92	47.47
GSM8K	91.50	<u>91.89</u>	87.72	87.57	94.39
MATH	62.12	62.78	<u>63.32</u>	62.62	71.84
<i>Coding Tasks</i>					
EvalPlus	65.93	61.43	<u>68.38</u>	63.75	77.60
MultiPL-E	58.70	62.16	<u>57.28</u>	<u>62.26</u>	65.94
MBPP	<u>76.00</u>	74.60	75.40	74.20	81.40
CRUX-O	66.20	68.50	<u>77.00</u>	76.60	79.00
<i>Multilingual Tasks</i>					
MGSM	82.40	82.21	79.69	<u>82.68</u>	83.53
MMMLU	84.40	83.49	83.09	<u>85.88</u>	86.70
INCLUDE	69.05	66.97	<u>73.47</u>	<u>75.17</u>	73.46

Table 4: Comparison among Qwen3-32B-Base and other strong open-source baselines. The highest and second-best scores are shown in bold and underlined, respectively.

	Qwen2.5-32B Base	Qwen2.5-72B Base	Gemma-3-27B Base	Llama-4-Scout Base	Qwen3-32B Base
Architecture	Dense	Dense	Dense	MoE	Dense
# Total Params	32B	72B	27B	109B	32B
# Activated Params	32B	72B	27B	17B	32B
<i>General Tasks</i>					
MMLU	83.32	86.06	78.69	78.27	<u>83.61</u>
MMLU-Redux	81.97	83.91	76.53	71.09	<u>83.41</u>
MMLU-Pro	55.10	<u>58.07</u>	52.88	56.13	65.54
SuperGPQA	33.55	36.20	29.87	26.51	39.78
BBH	84.48	<u>86.30</u>	79.95	82.40	87.38
<i>Math & STEM Tasks</i>					
GPQA	<u>47.97</u>	45.88	26.26	40.40	49.49
GSM8K	<u>92.87</u>	91.50	81.20	85.37	93.40
MATH	57.70	62.12	51.78	51.66	<u>61.62</u>
<i>Coding Tasks</i>					
EvalPlus	<u>66.25</u>	65.93	55.78	59.90	72.05
MultiPL-E	<u>58.30</u>	<u>58.70</u>	45.03	47.38	67.06
MBPP	73.60	<u>76.00</u>	68.40	68.60	78.20
CRUX-O	<u>67.80</u>	66.20	60.00	61.90	72.50
<i>Multilingual Tasks</i>					
MGSM	78.12	<u>82.40</u>	73.74	79.93	83.06
MMMLU	82.40	84.40	77.62	74.83	<u>83.83</u>
INCLUDE	64.35	69.05	<u>68.94</u>	68.09	<u>67.87</u>