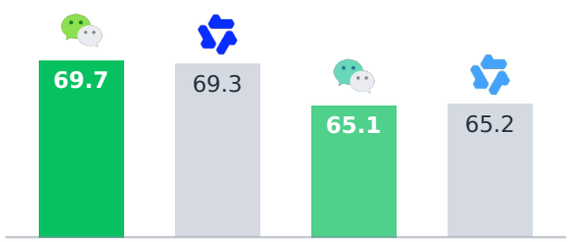
General Benchmarks · WeMM-Ecom-9B & 4B vs. Dense Peers

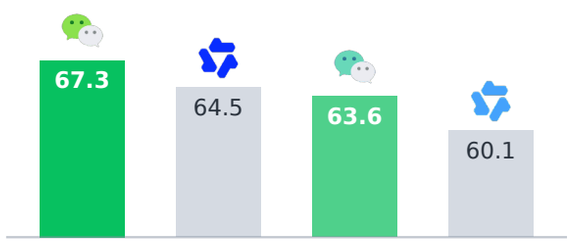
Thinking Mode

Overall (64) · WeMM-Ecom-9B: 74.18 · WeMM-Ecom-4B: 70.49

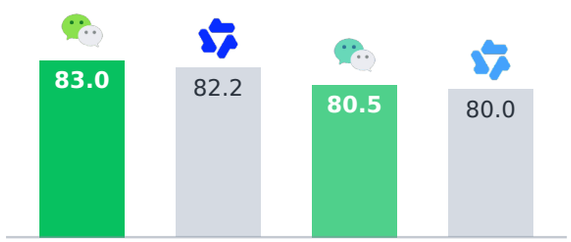
STEM Reasoning
12 Benchmarks



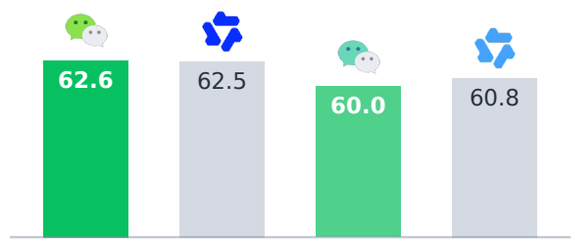
General VQA
11 Benchmarks



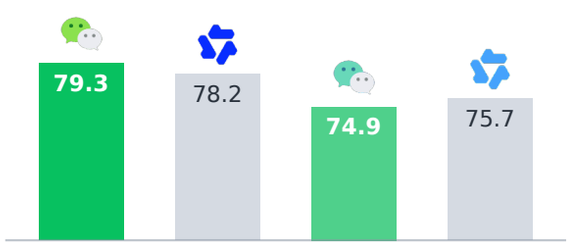
Document & Chart
11 Benchmarks



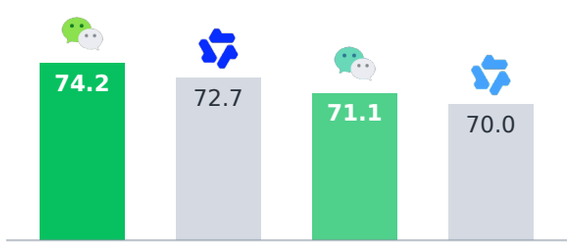
2D / 3D Spatial
6 Benchmarks



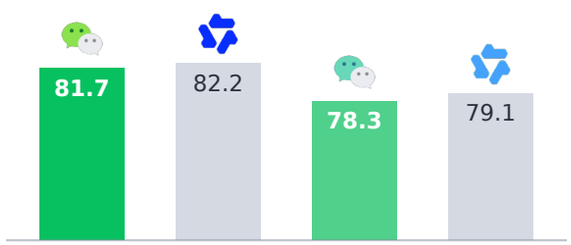
Grounding & Counting
8 Benchmarks



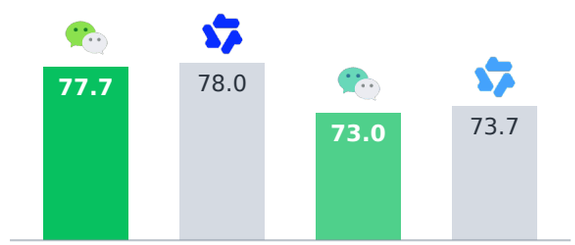
Multi-Image
4 Benchmarks



Instruction Following
5 Benchmarks



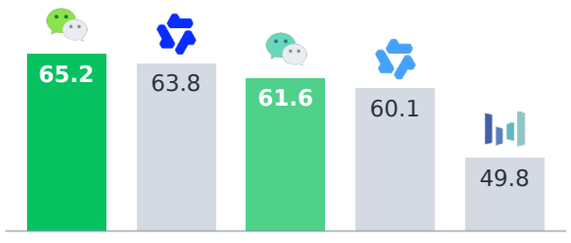
Text-Centric
7 Benchmarks



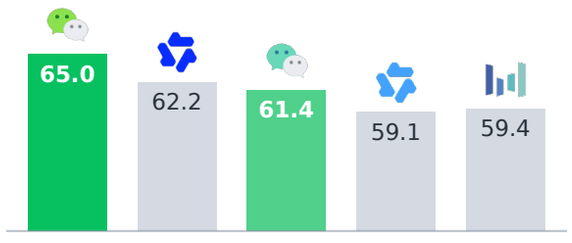
Instruct Mode (Non-Thinking)

Overall (64) · WeMM-Ecom-9B: 69.69 · WeMM-Ecom-4B: 66.59

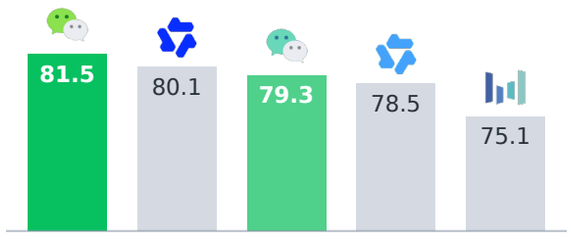
STEM Reasoning
12 Benchmarks



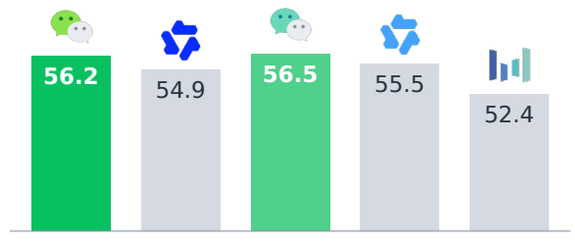
General VQA
11 Benchmarks



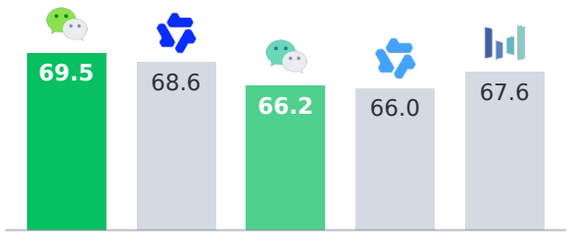
Document & Chart
11 Benchmarks



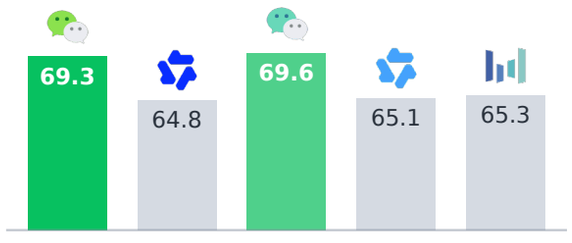
2D / 3D Spatial
6 Benchmarks



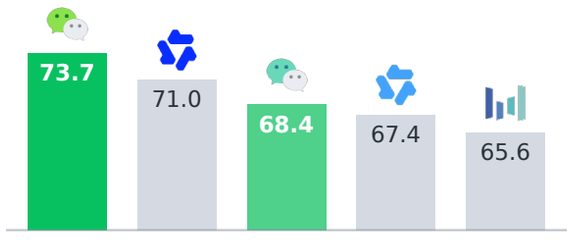
Grounding & Counting
8 Benchmarks



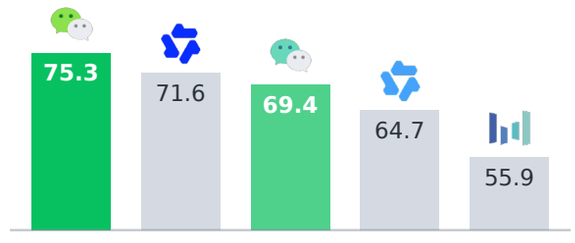
Multi-Image
4 Benchmarks



Instruction Following
5 Benchmarks



Text-Centric
7 Benchmarks



WeMM-Ecom-9B · Qwen3.5-9B · **WeMM-Ecom-4B** · Qwen3.5-4B · Valley3-8B*

* Valley3-8B is evaluated in Instruct mode only. All scores are group averages across the 64-benchmark general suite.