

keep improving

Benchmark	Mini-InternVL 4B	InternVL2 Pro	InternVL2.5 78B	InternVL3 78B	Qwen2.5-VL 72B	Other Open-Source MLLMs	GPT-4o 20240513	Claude-3.5 Sonnet	Gemini-2.5 Pro
MMMU Multi-discipline	43.3%	62.0%	68.7%	72.2% (3.5 ↑)	70.2%	66.7%	69.1%	66.4%	74.7%
MathVista Math	53.7%	66.3%	76.5%	79.0% (2.5 ↑)	74.8%	70.5%	63.8%	65.1%	80.9%
AI2D Diagrams	76.9%	87.3%	89.3%	89.8% (0.5 ↑)	88.7%	88.1%	84.9%	81.2%	89.5%
ChartQA Charts	81.0%	87.1%	89.1%	89.7% (0.6 ↑)	89.5%	88.3%	85.7%	90.8%	-
DocVQA Documents	87.7%	95.1%	95.0%	94.9% (0.1 ↓)	96.4%	96.5%	92.8%	95.2%	-
InfographicVQA infographics	64.6%	83.3%	84.3%	85.2% (0.9 ↑)	87.3%	84.7%	69.1%	74.3%	-
HallusionBench Hallucination	-	55.2%	58.1%	59.1 % (1.0 ↑)	55.2%	58.1%	55.0%	55.5%	64.1%
MMT-Bench Val Multitask AGI	58.4%	66.9%	70.6%	73.2% (1.9 ↑)	-	71.8%	-	-	-
LongVideoBench Video	-	61.1%	63.6%	65.7% (2.1 ↑)	60.7%	61.3%	66.7%	-	-