

Model	ChartEdit Task					Reproduction Task				
	Acc. ↑	F1 ↑	Prec. ↑	Recall ↑	Spec. ↑	Acc. ↑	F1 ↑	Prec. ↑	Recall ↑	Spec. ↑
<i>Proprietary</i>										
Gemini-3-Pro	79.1	81.6	71.4	95.1	64.0	75.6	79.1	68.5	93.6	57.9
GPT-5.2	79.1	81.3	73.7	90.7	69.6	74.3	71.4	79.5	64.7	83.7
Seed1.6-VL	70.8	68.6	71.8	65.6	75.6	70.5	65.7	77.4	57.0	83.7
Claude-Sonnet-4.5	60.9	68.8	56.5	87.7	35.7	62.9	66.7	62.0	72.1	56.1
<i>Open-Source (thinking)</i>										
MiMo-VL-SFT	64.5	72.9	58.6	96.3	34.7	72.7	74.1	69.9	78.8	66.7
Qwen3-VL-30B	66.0	67.4	63.1	72.4	59.9	69.5	71.8	66.3	78.2	61.0
MiMo-VL-RL	59.4	76.2	63.2	96.0	39.6	71.8	71.6	71.3	71.8	71.7
LLaVA-Critic-R1	54.3	62.2	52.1	77.3	32.6	53.3	67.6	51.5	98.1	9.4
ThinkLite-VL-7B	56.1	63.5	53.3	78.5	34.9	61.9	71.7	56.7	97.4	27.0
<i>Open-Source (non-thinking)</i>										
Qwen2-VL-7B	53.1	66.1	51.0	93.9	14.5	52.4	67.4	51.0	99.4	6.3
Qwen2-VL-72B	55.8	68.4	52.5	98.2	15.7	44.9	50.7	34.5	24.6	29.1
DeepSeek-VL	51.6	65.3	50.7	91.8	16.6	49.5	66.2	49.5	100.0	0.0
Qwen2.5-VL-7B	57.6	56.4	56.4	56.4	58.7	63.8	72.3	58.2	95.5	32.7
Qwen2.5-VL-72B	61.5	67.2	57.4	81.0	43.0	68.9	68.4	68.8	68.0	69.8
GLM-4V	60.6	74.0	59.9	96.7	32.2	63.8	78.0	71.7	85.3	67.5
MiMo-VL-SFT (NoThink)	67.2	64.1	68.5	60.1	73.8	52.1	7.4	85.7	3.9	99.4
MiMo-VL-RL (NoThink)	66.6	69.7	62.3	79.1	54.6	55.9	21.5	90.5	12.2	98.7
Kimi-VL	50.1	66.1	49.4	100.0	2.9	51.8	67.2	50.6	100.0	4.4
InternVL2.5-38B	55.2	67.4	52.2	95.1	17.4	57.8	69.6	54.1	97.4	18.9
InternVL2.5-8B	56.7	61.9	54.1	72.4	41.9	52.4	67.5	51.0	100.0	5.7
Molmo	53.7	33.5	55.7	23.9	82.0	50.5	0.0	0.0	0.0	100.0
Molmo2	50.8	62.6	49.6	84.7	18.6	58.7	61.3	57.2	66.0	51.6
Qwen3-VL-8B	66.0	70.3	61.1	82.8	50.0	64.4	72.3	58.9	93.6	35.9
Qwen3-VL-32B	69.0	72.5	63.7	84.0	54.6	71.4	68.1	76.2	61.5	81.1
InternVL3.5-8B	58.2	65.4	54.8	81.0	36.6	65.1	71.4	60.2	87.7	43.4
InternVL3.5-38B	55.2	64.6	52.5	84.0	27.9	68.6	75.2	61.7	96.2	41.5
Average	60.7	66.8	58.3	81.7	41.8	62.3	62.3	61.5	74.7	50.3