

Model Performance Summary Table

Model	Neutral-Text Avg	Emotion-Matched Avg	Emotion-Mismatched Avg	Paralinguistic Audio	Overall Avg
Qwen3-instruct	66.6%	33.5%	38.0%	--	34.4%
Qwen2.5-Omni-7B	26.9%	37.6%	38.8%	22.7%	32.8%
Qwen3-Omni-30B	27.3%	42.8%	38.3%	21.0%	33.9%
Baichuan-Omni-1.5	15.9%	36.0%	36.0%	22.7%	28.3%
Gemini2.5-Flash	25.1%	34.8%	36.9%	18.0%	30.2%
Gemini2.5-Pro	38.3%	38.9%	39.8%	15.7%	35.7%