SOTA Models for Instruction Following & Visual Dialogue

Task	Description	SOTA Models
Visual Instruction Following	Perform tasks based on visual input + commands	GPT-4o, MiniGPT-4
Multimodal Chat/Agents	Real-time interaction with images + text + audio	GPT-4o, Gemini 1.5