SOTA Models for Visual Question Answering

Task	Description	SOTA Models
VQA (Visual QA)	Answer natural language questions about images	Flamingo, BLIP-2
Visual Reasoning (Few-shot)	Logical reasoning on image-based questions	MiniGPT-4, GPT-4o

complementary scenes

Tuple: <girl, walking, bike>

Question: Is the girl walking the bike?