

SOTA Models for Captioning & OCR

Task	Description	SOTA Models
Image Captioning	Generate textual descriptions for images	BLIP-2, Flamingo
OCR + Grounding	Recognize text and ground it to visual objects	Kosmos-2, GPT-4o

