Mean Bayesian Reference-Relative Efficiency 102.41 100 93.05 87.04 80 77.65 77.22 Mean Efficiency (%) 69.77 68.08 58.74 60 40 31.59 20 0 Llama 4 Maverick AB

Gemma 3 4B it

Phi 4 Multimodal GPT 40

Claude 3.7 Sonnet

Owen 2.5 VL 32B

Gemini 2.5 Flash Lite