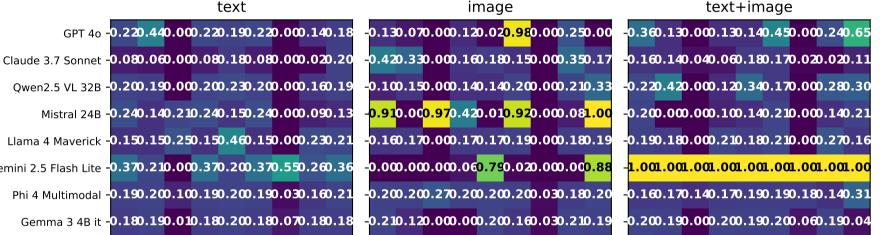
text



0.8

Owen2.5 VL 32B -0.200.190.000.200.230.200.000.160.19 Mistral 24B -0.240.140.210.240.150.240.000.090.13 Llama 4 Mayerick -0.150.150.250.150.460.150.000.230.21 Gemini 2.5 Flash Lite -0.370.210.000.370.200.370.550.260.36 Phi 4 Multimodal -0.190.200.100.190.200.190.030.160.21 Gemma 3 4B it -0.180.190.010.180.200.180.070.180.18 F 0.2 GPT 5 Mini -0.020.060.050.020.160.020.500.040.59 -0.110.120.140.120.210.100.140.070.12 -0.040.060.040.060.080.090.060.010.08 Longer Context Gradual noise Longer Context Longer Context short context Gradual noise Inconsistent Prior Inconsistent Prior Numerical steer short context Gradual noise Inconsistent Prior Verbal steer Ablation Ablation Ablation

Sequential-ness (factor probability)