



Figure 11: Performance variation of different multimodal reasoning models on the perception task in the *RH-Bench* Benchmark with changes in reasoning length ($\alpha \in [-0.1, 0.1]$).