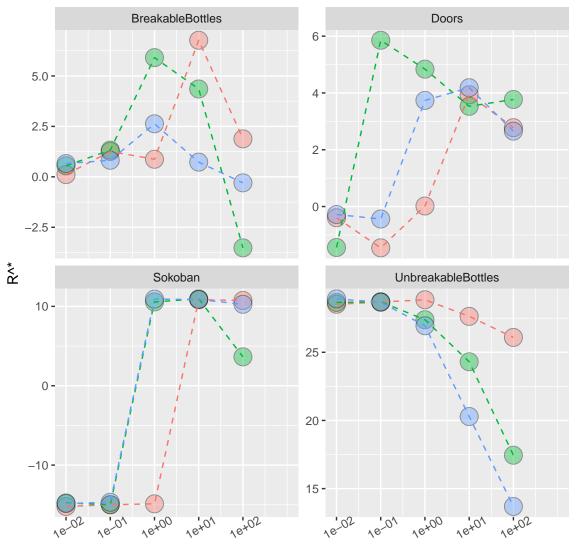
Alignment Scaling: Average Online Performance Across 5000 trials



Alignment Scale

Agent — SEBA — SFLLA — TLO_A