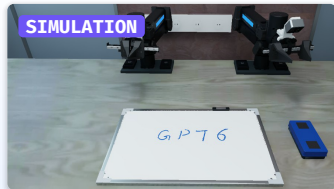


a Matched Sim-Real Evaluation

16 tasks · 3 levels · 3 trials per task and environment



GPT-6 Astra

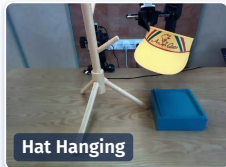
zero-shot

Shared tasks and success criteria



b The PhysLadder

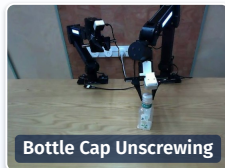
Each step adds one interaction requirement



L1 5 tasks
Selection & alignment



L2 7 tasks
+ Force regulation
Selection & alignment



L3 4 tasks
+ Reconfiguration
+ Force regulation
Selection & alignment

c Success by Level

