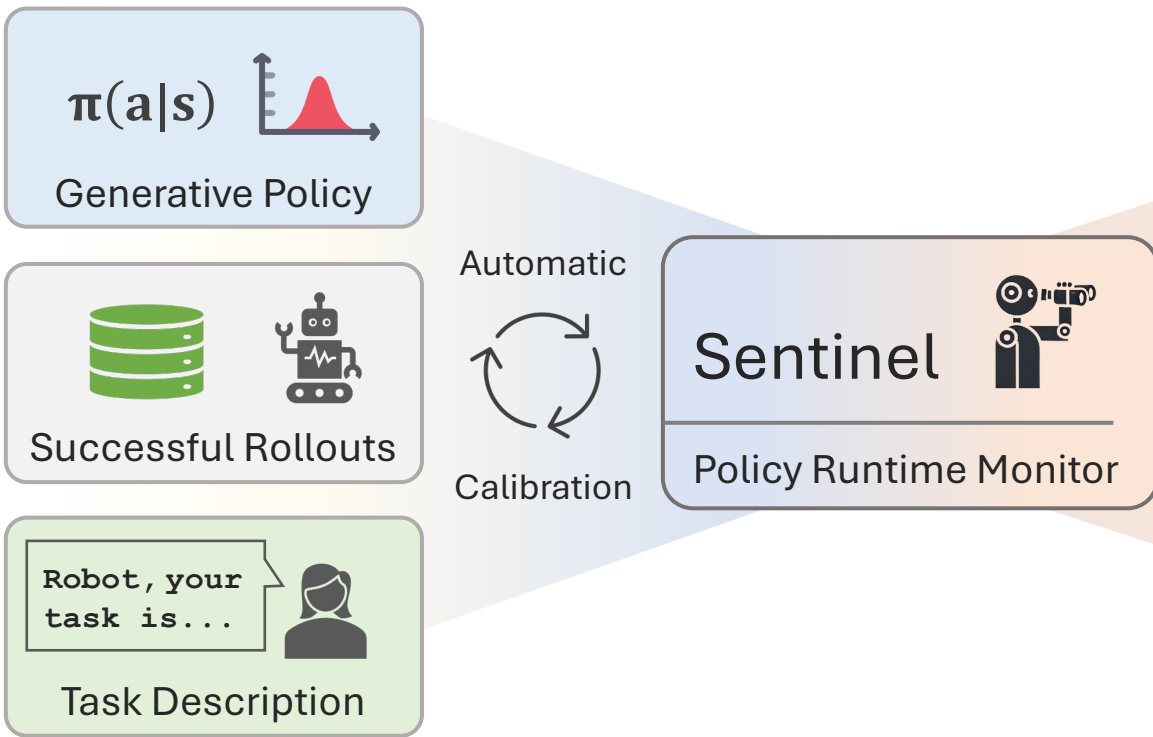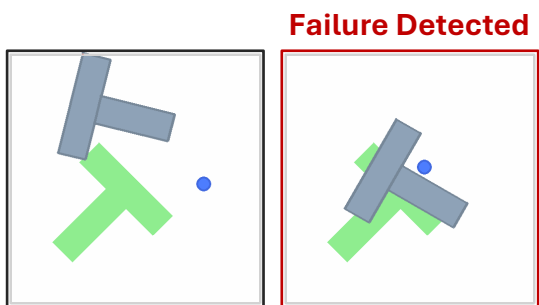# Policy Training & Validation
In-Distribution Scenarios
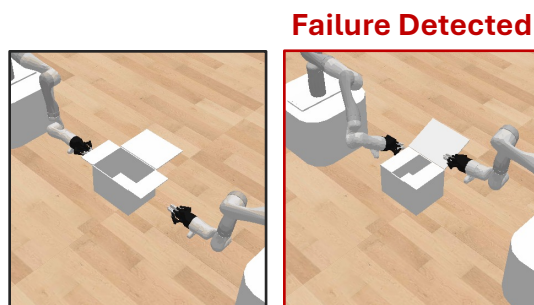
Policy Deployment
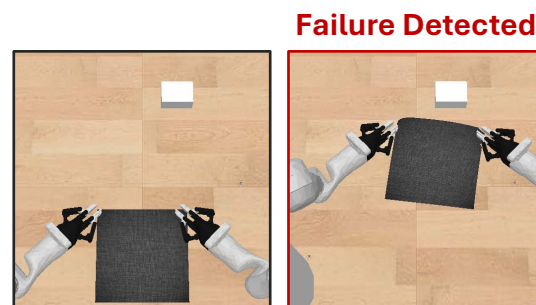## Out-of-Distribution Scenarios

$\pi(a|s)$
Generative Policy

Successful Rollouts

Task Description

Robot, your task is...

Automatic ↻ Calibration

Sentinel 🤖
Policy Runtime Monitor

(a) Consistency Checking

$\pi(a|s)$
Policy Rollout

ϵ

Statistical **T**emporal **A**ction **C**onsistency

High Temporal Error
↓
Policy Failing

(b) Online Video QA

Prompt

Video

Vision **L**anguage **M**odel

Explanation: The robot is failing to make progress on the task...

Failure Detected

PushT: Action Multimodality

Failure Detected

Box Closing: Scale Shift

Failure Detected

Object Covering: Pose Shift

Failure Detected

Chair Tucking: Dynamics Shift