

Adversarial Attacks

Optimization



Object / Behavior Triggers

Deployment



Role-Playing

Sensor Delay



Representation Hijacking



Data Poisoning



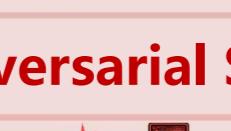
Adversarial Suffixes



Visual Perturbations



Data Poisoning



Adversarial Suffixes



Visual Perturbations

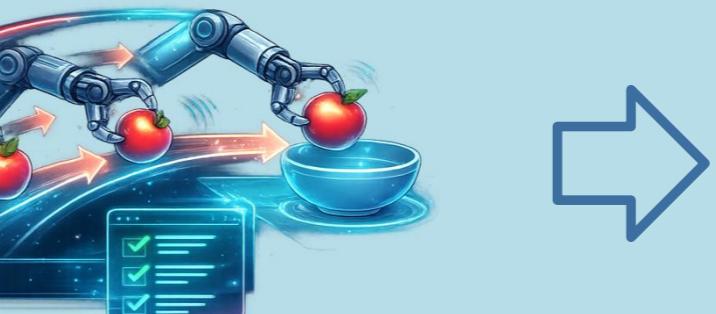


Multimodal Perturbations

Perception



Decision & Planning



Execution & Interaction

Multi-Stage Closed-Loop Embodied AI Systems

Optimization



Adversarial Training



Safety Fine-Tuning



Explicit Modeling



None

Deployment



Anomaly Detection



Information Augmentation



Structural Augmentation



Safety Filtering



Safety Verification



Secondary Control

Adversarial Defenses