


Policy Validation

Generative Policy $\pi(A|O)$

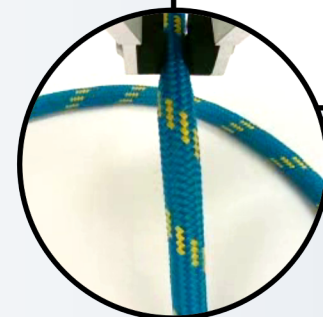


Successful Rollouts 

Score Learning & Calibration

Learning + Calibration

Observation-Based Score



Calibration

Action-Based Score



$$F_A(\tau:t) \wedge F_O(\tau:t)$$



FIPER

Failure Prediction at Runtime

Policy Deployment

