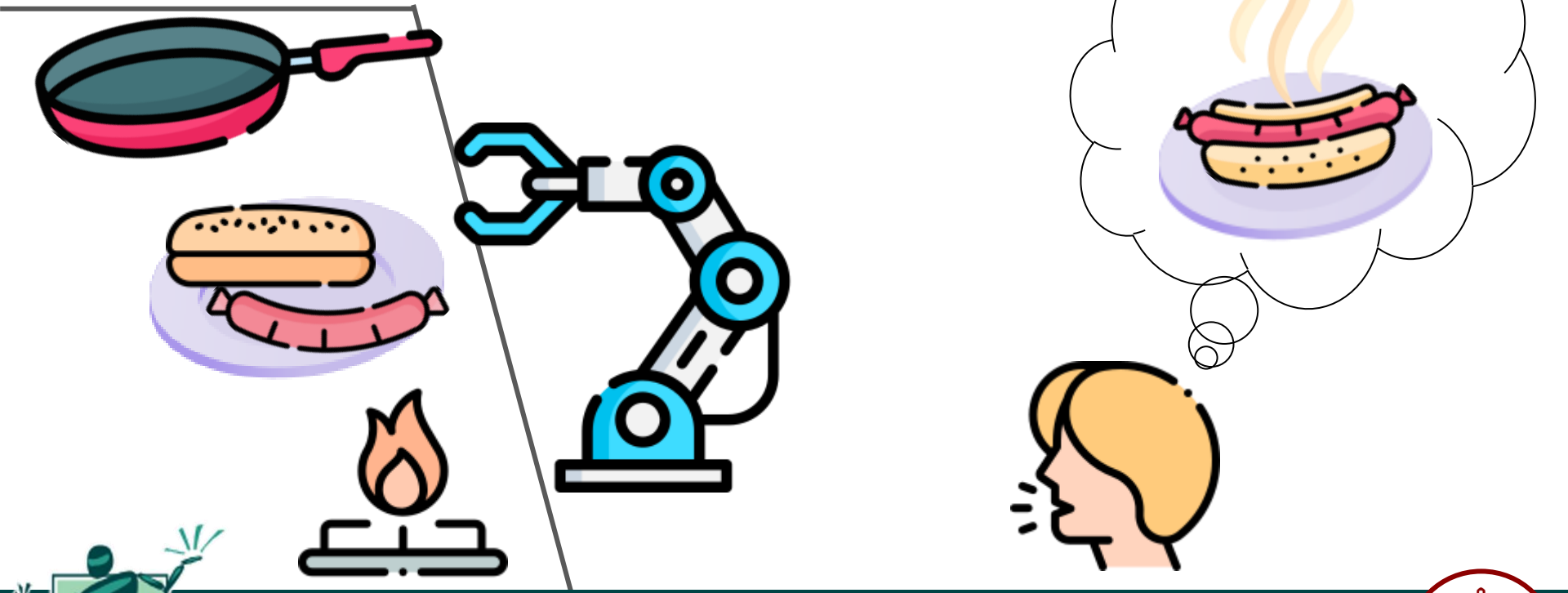
Primitive Skill-based Robot Learning from Human Evaluative Feedback

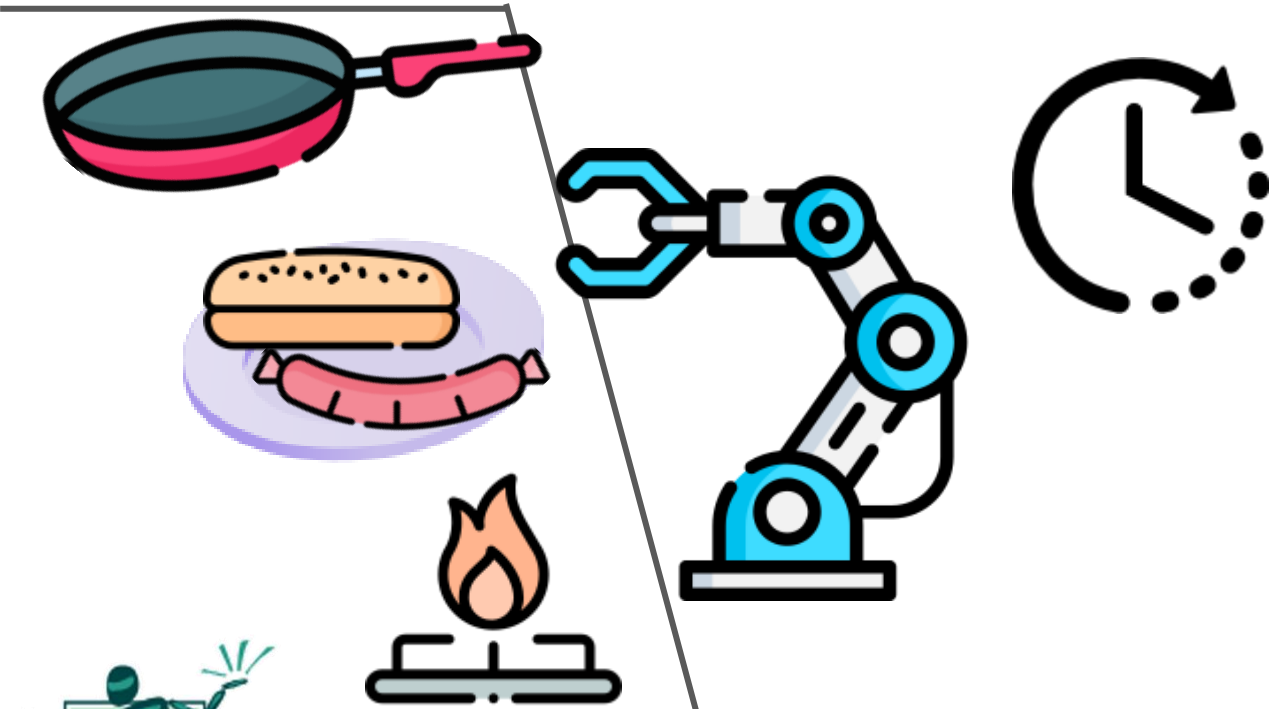
Ayano Hiranaka*, Minjune Hwang*, Sharon Lee, Chen Wang, Li Fei-Fei, Jiajun Wu, Ruohan Zhang
Stanford University

IEEE/RSJ International Conference on Intelligent Robots and Systems

Suppose we want the robot to cook a hotdog.



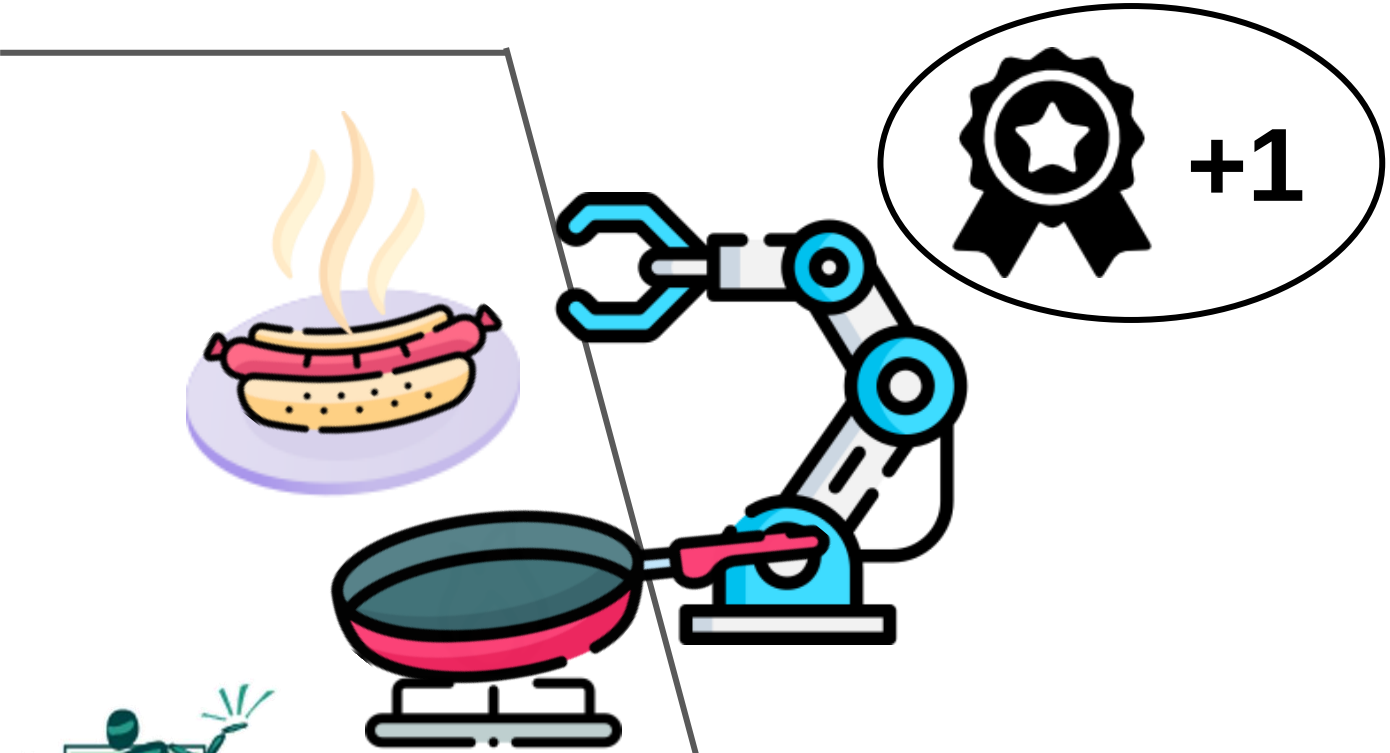
*RL in the real world is **sample inefficient**.*



Challenges with RL

- **Sample Inefficiency**

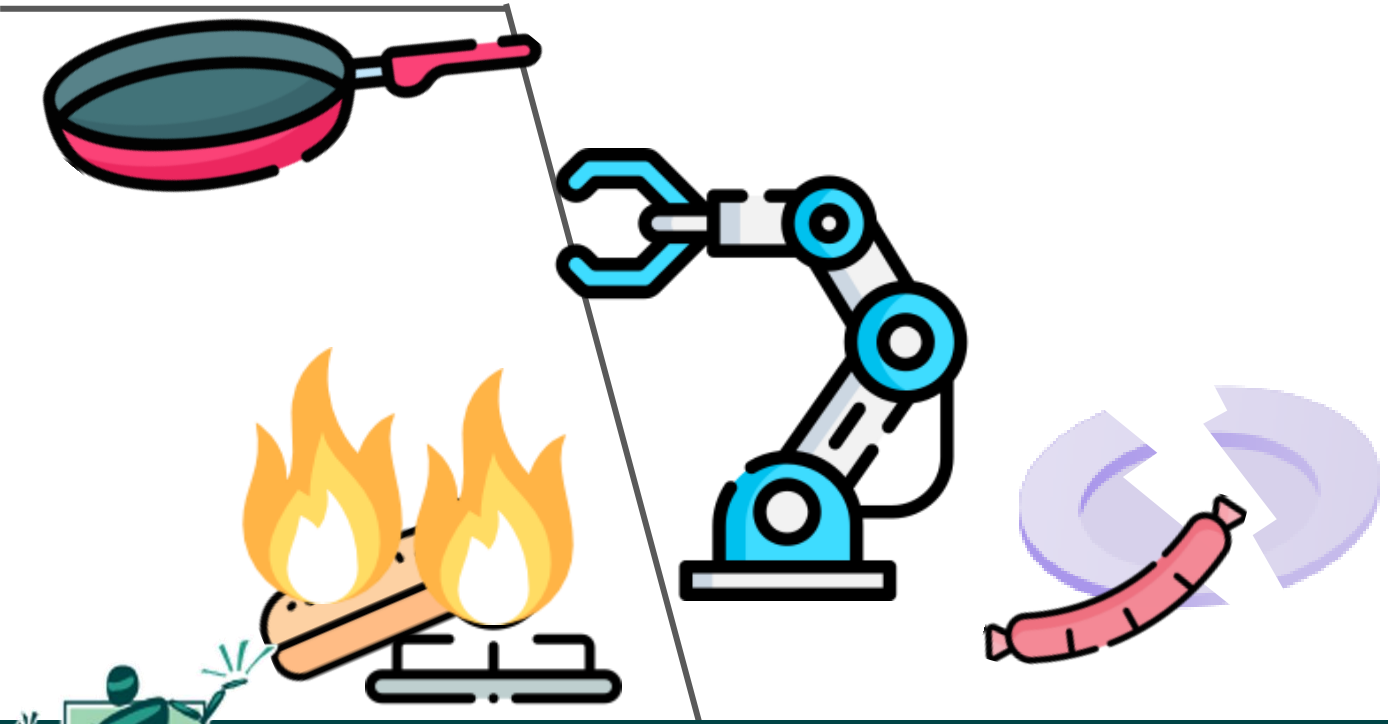
RL in the real world has ***challenges in reward design****.*



Challenges with RL

- Sample Inefficiency
- **Sparse Rewards**

*RL in the real world can be **dangerous**.*



Challenges with RL

- Sample Inefficiency
- Sparse Rewards
- **Safety Concerns**

Goal of Our Work

Robot RL in Real World

- Sample Inefficiency
- Sparse Rewards
- Safety Concerns

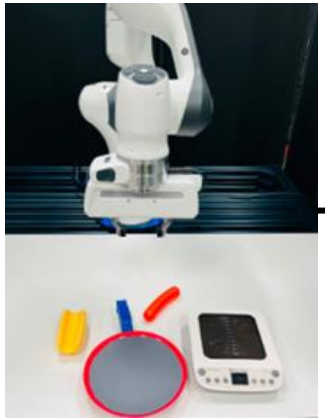
Our Work: SEED (Skill-based Evaluative fEeDback)

Parameterized Primitive Skills → Intuitive for Humans

Human Evaluative Feedback → Dense Training Signals

Evaluation without Execution → Safe

Method: Network Architecture



State Observation

s

Primitive Skill Selection

π_φ *skill policy*

a

Skill ID

π_ψ *parameter policy*

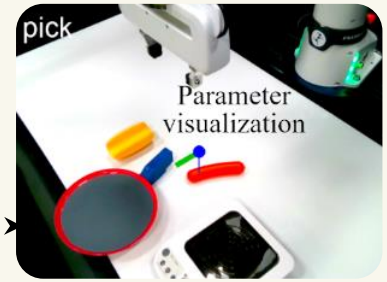
x

Parameter Selection

Human Feedback Estimation

H_θ *human critic*

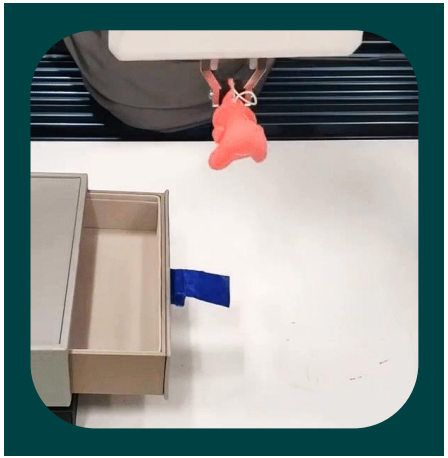
$H_\theta(s,a,x)$



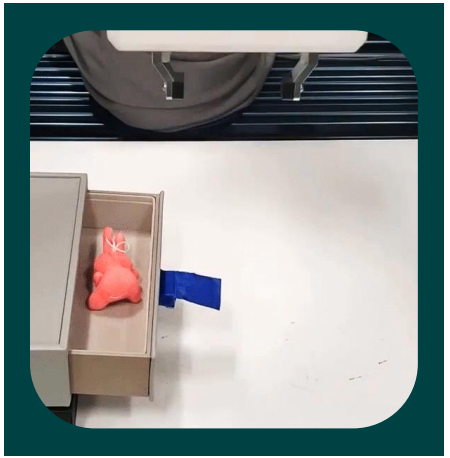
Human Evaluation

Method: Primitive Skills

pick (x, y, z)



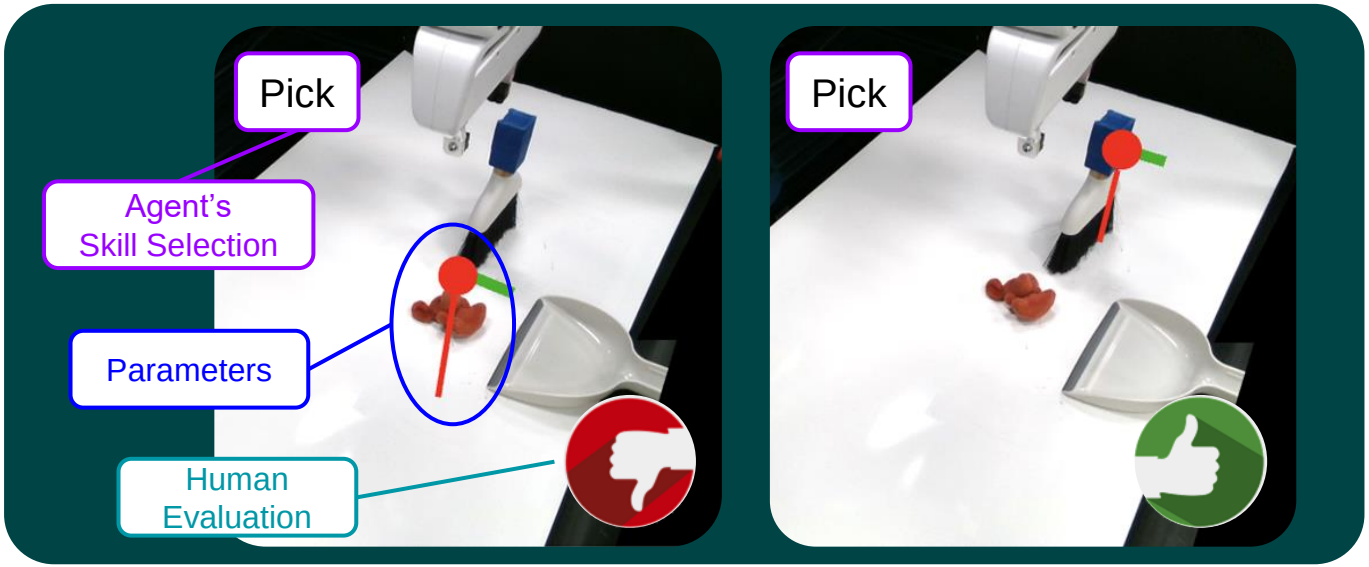
place (x, y, z)



push (x, y, z, δ)

Skills designed with operational space controller (OSC) and deployed on Franka arm

Method: Evaluation without Execution



Next Subgoal: Pick up the Broom

Robot Proposes Action

Human Evaluates

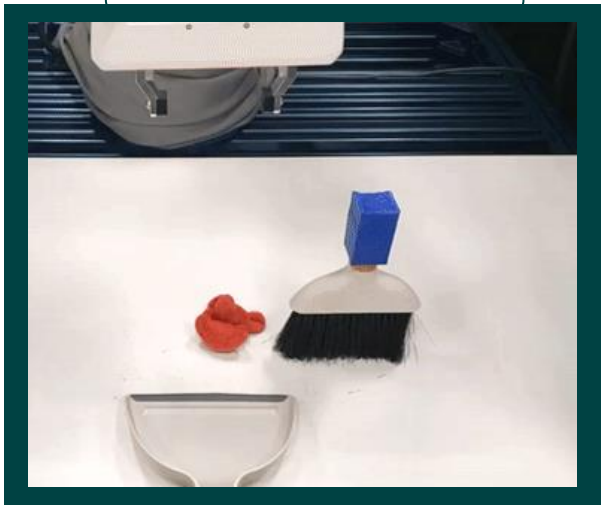
When human is confident

Robot Executes Action

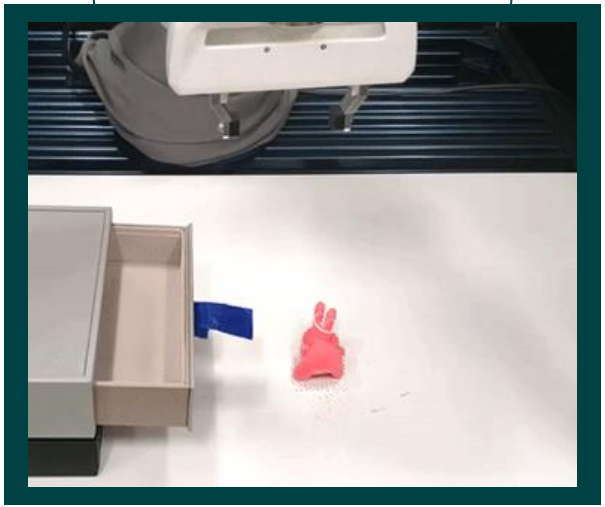
Sample Rollouts for Long-Horizon Tasks

Cooking Hotdog

Sweeping



Putting Toy Away



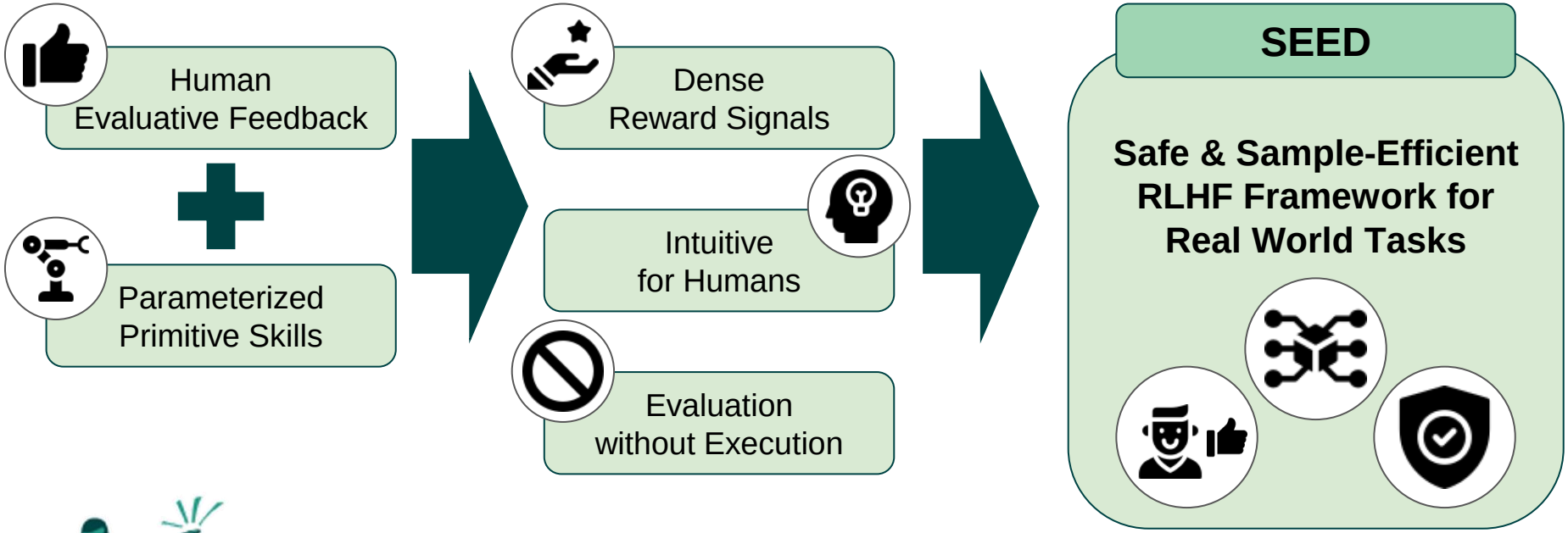
Learned policy can recover from errors!

Result Highlights

Compared to baselines:

Sample Efficiency	**10x faster** training
Training Outcome	**9x** higher success rate
Safety	**3-7x** fewer safety violations

Summary


Human
Evaluative Feedback
Parameterized
Primitive Skills
Dense
Reward Signals
Intuitive
for Humans
Evaluation
without Execution
SEED
Safe & Sample-Efficient
RLHF Framework for
Real World Tasks

Thank You

Visit Our Poster

10:00-11:30 am, Hall E

Contact Info

ayanoh@stanford.edu, mjhwang@stanford.edu

Website