



Skill Set Optimization: Reinforcing Language Model Behavior via Transferable Skills

UCI

Ai2

Kolby Nottingham* \diamond Bodhi Majumder \diamond Bhavana Dalvi Mishra
Sameer Singh \diamond Peter Clark \diamond Roy Fox

<https://allenai.github.io/ss0/>
*knotting@uci.edu

Motivation

Without costly training, LLM actors have no method to continually adapt and learn

Task: Measure the melting temp of chocolate.	State: You see chocolate and a stove.
LLM Actor	move chocolate to stove
State: You see a stove with chocolate.	
LLM Actor	activate stove
State: You see a stove with melted chocolate.	
LLM Actor	use thermometer on chocolate
State: The thermometer reads 40 C.	

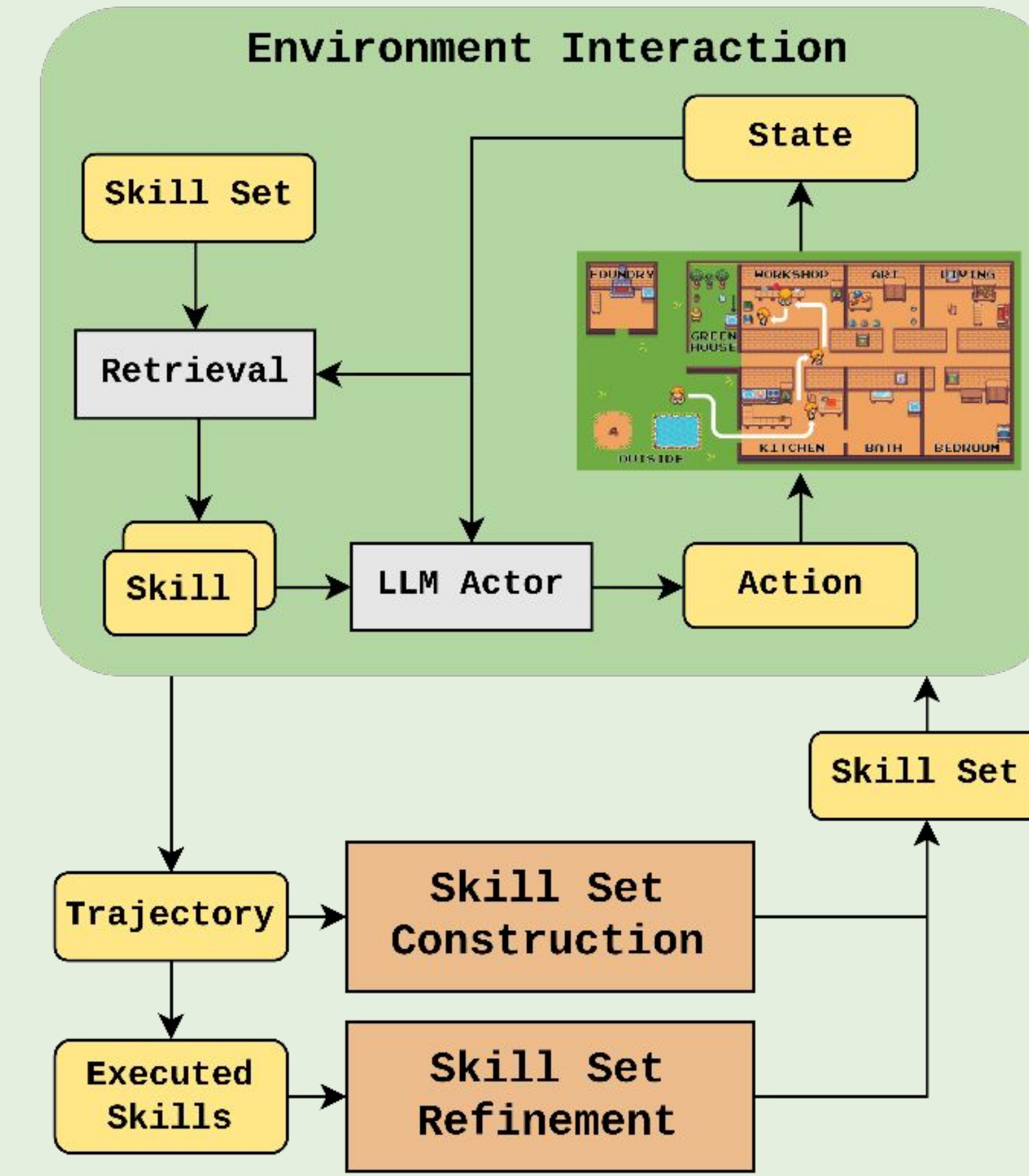
SSO summarizes successful experience into retrievable abstract skills

The unaltered context alone would fail to generalize knowledge to a new task

Task: Measure the melting temp of water.	State: You see a stove and a freezer with ice.
LLM Actor	use thermometer X
Melt substance: 1. Move substance to stove 2. Activate stove	
LLM Actor	move ice to stove ✓

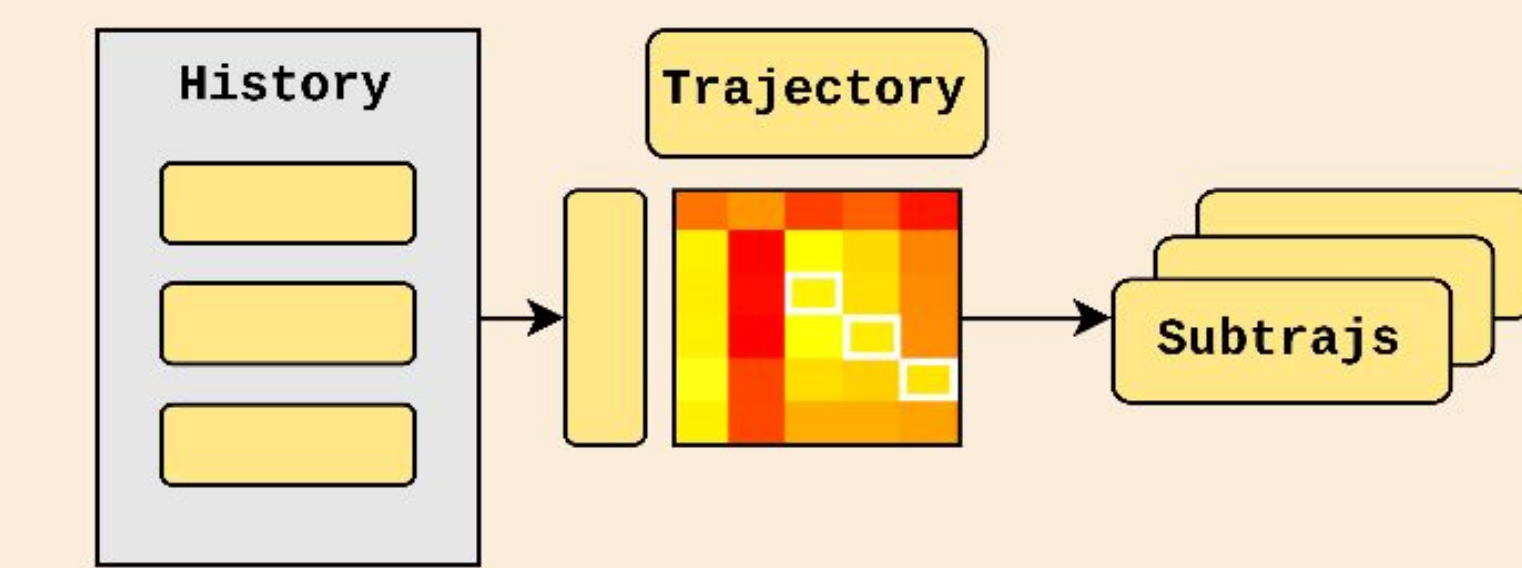
Adding learned skills in-context enables policy improvement

Method



Skill Set Optimization occurs after each episode in two steps:

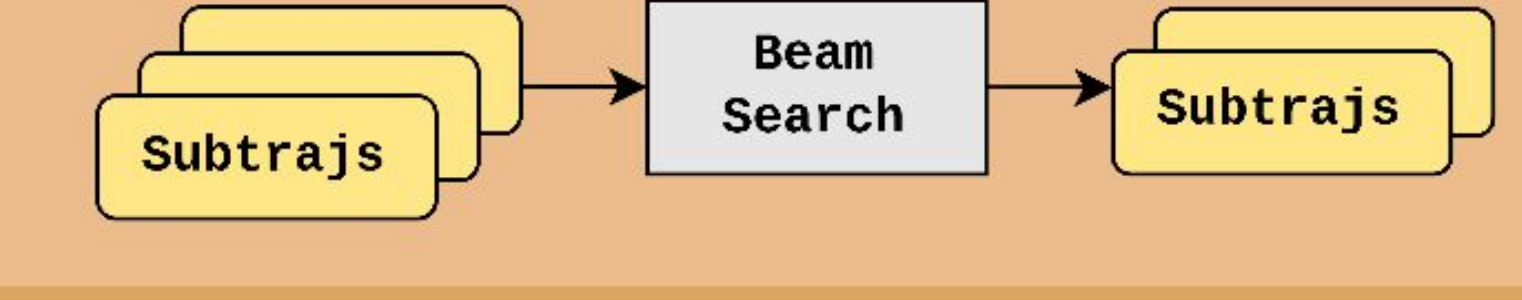
Extract



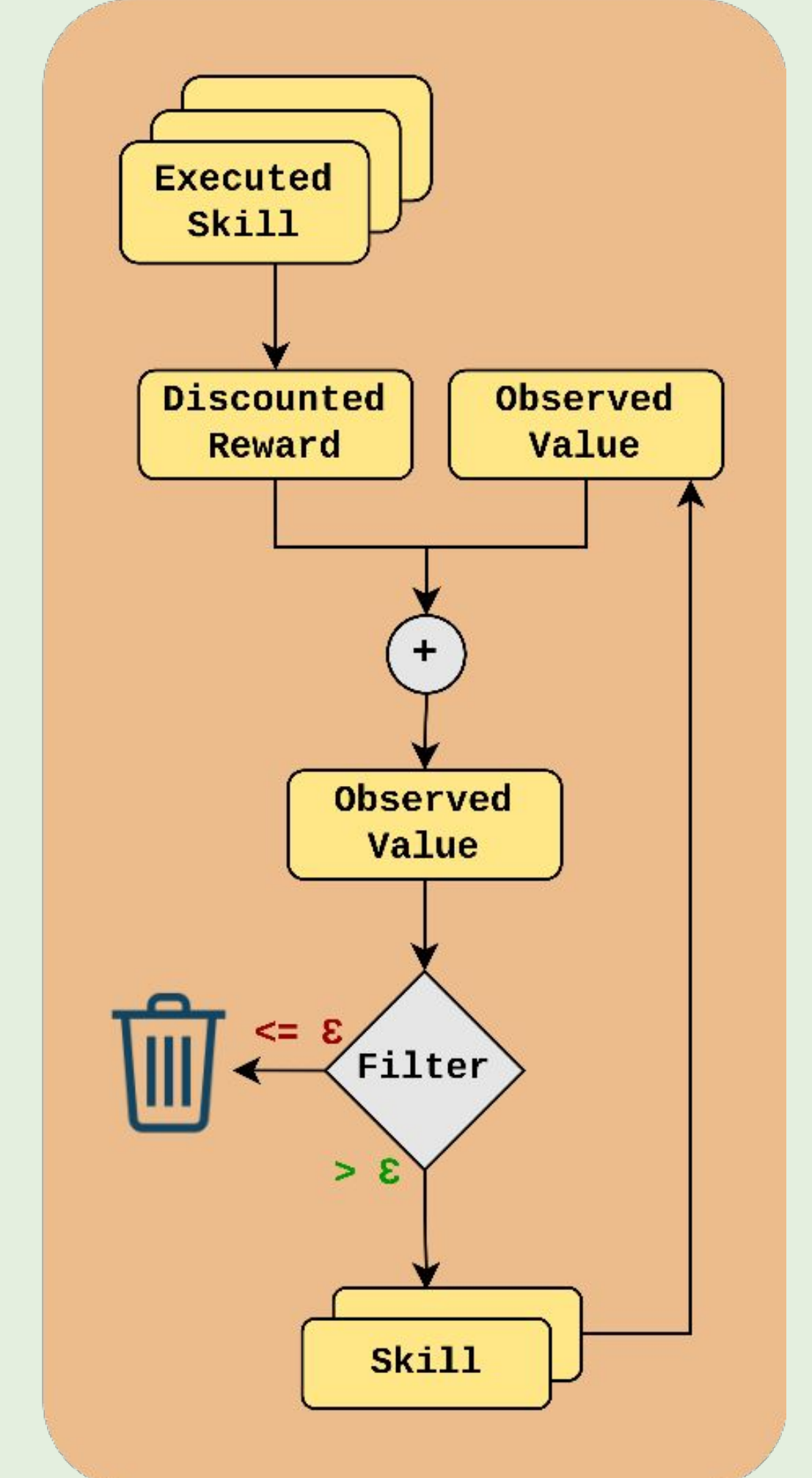
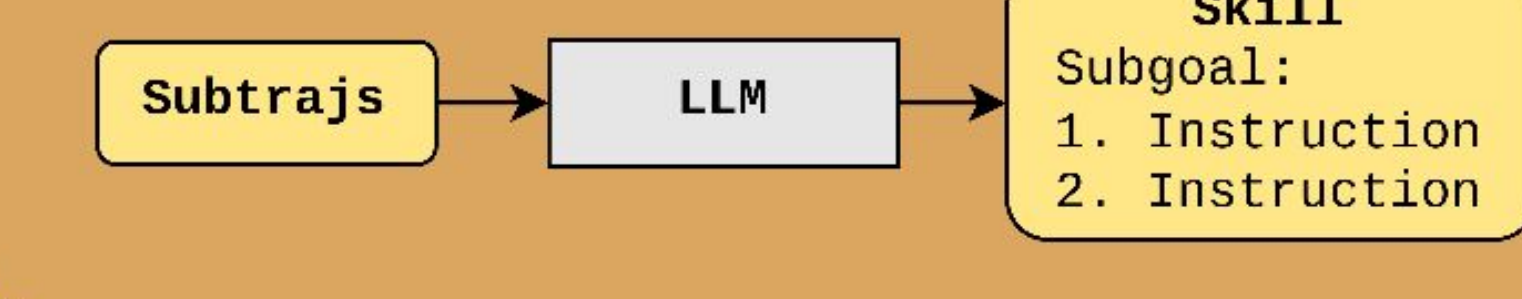
Score



Sample



Generate



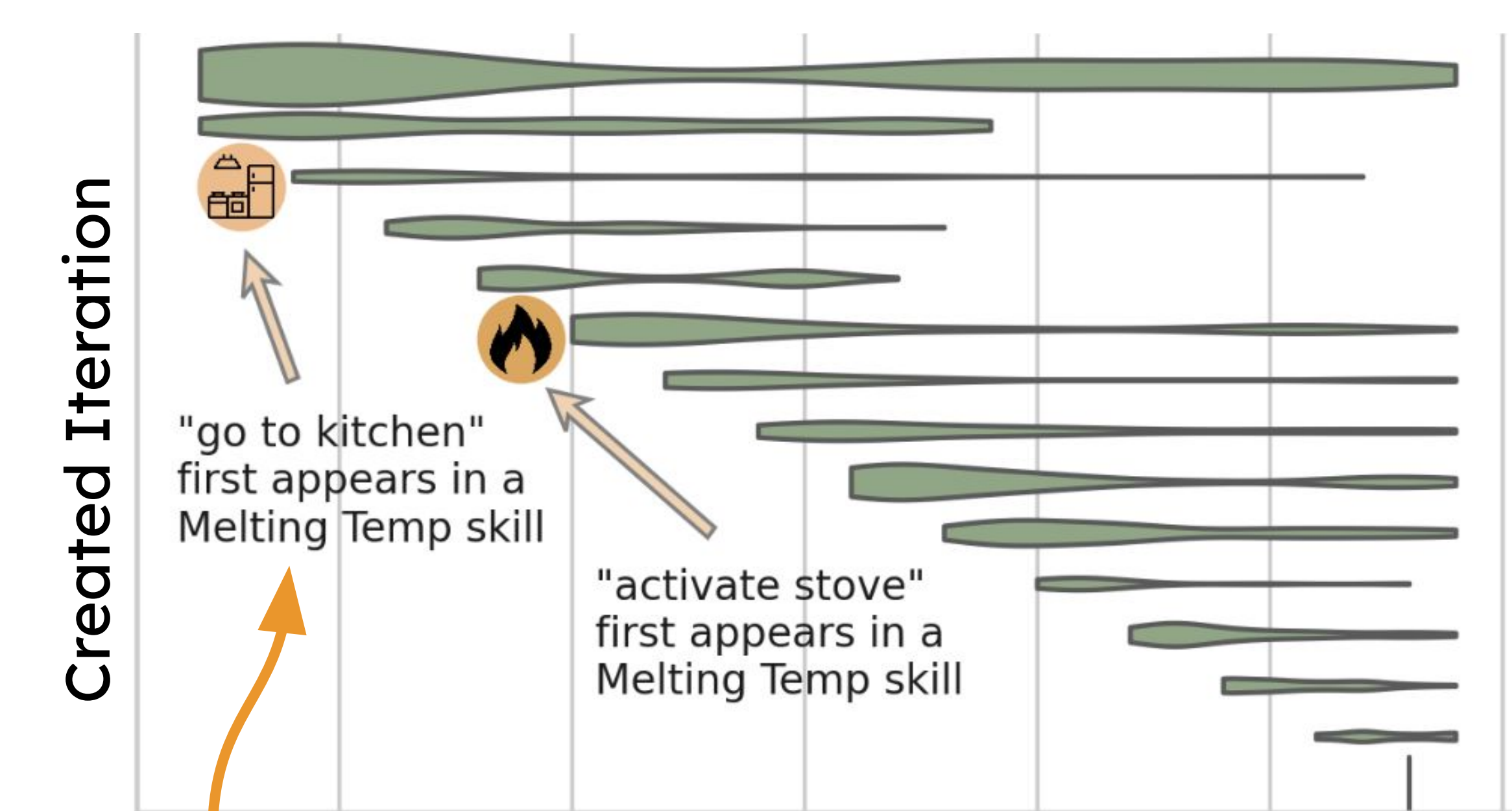
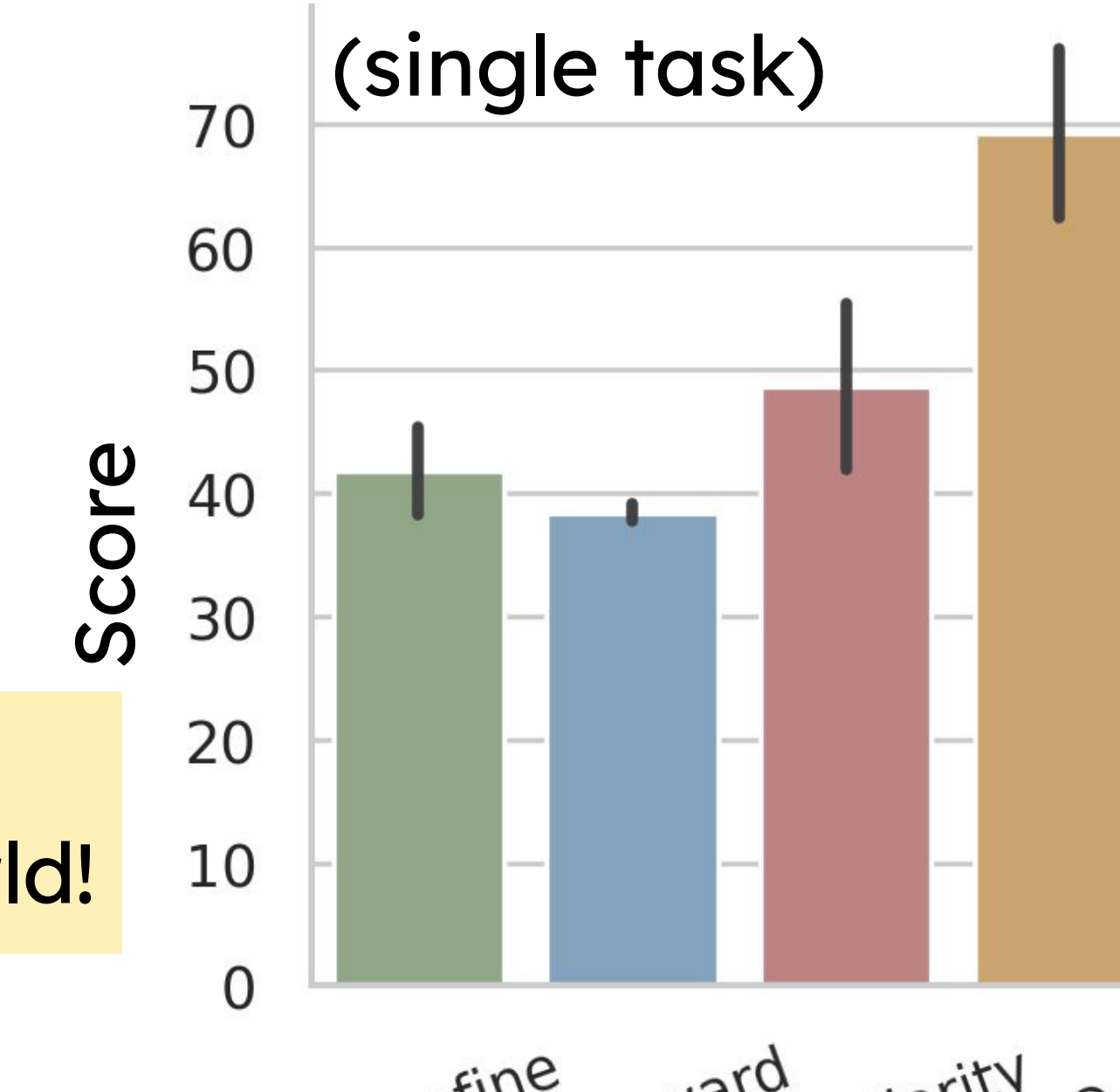
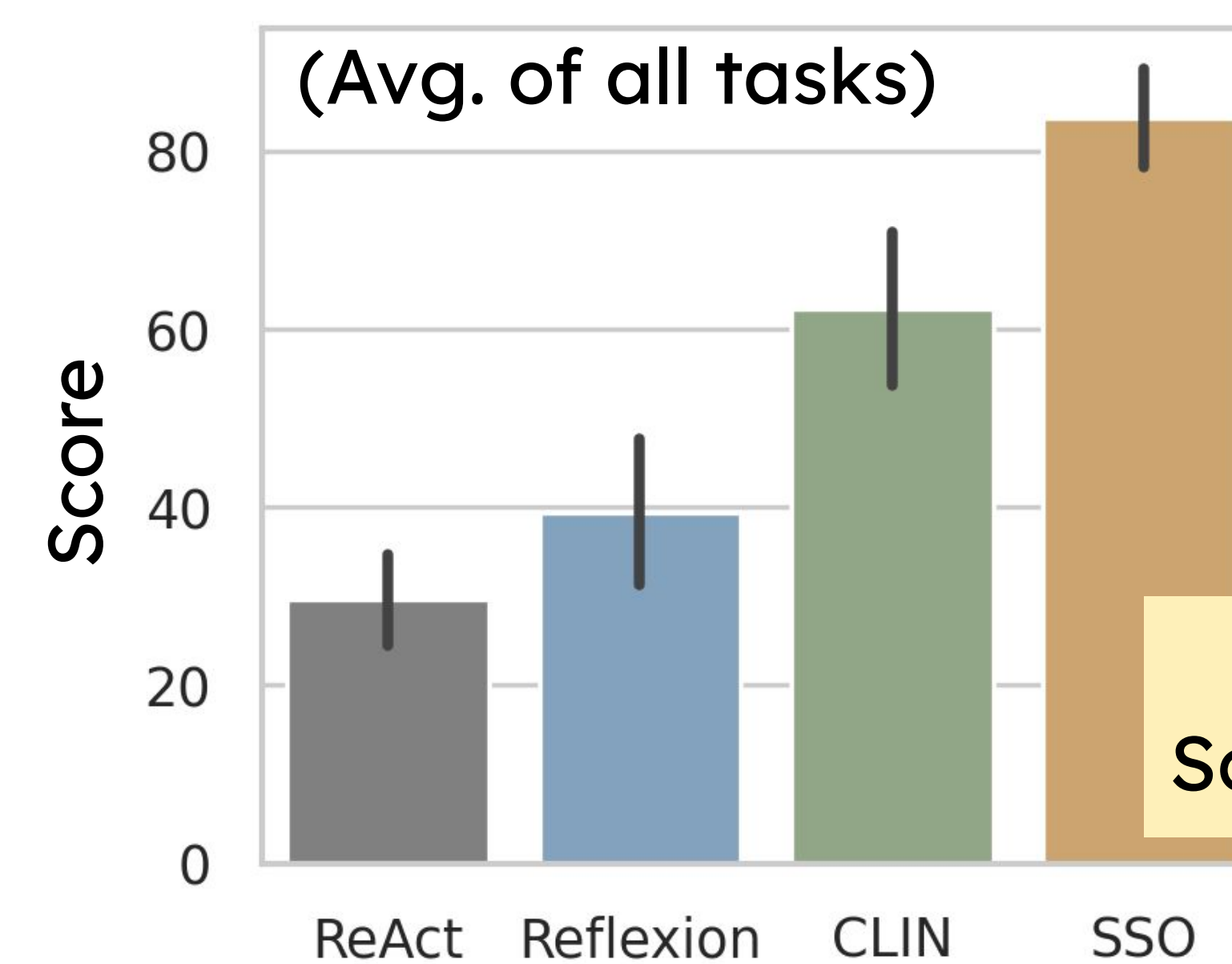
1. Construction of subgoals and instructions

2. Refinement by pruning poor performing skills

Results

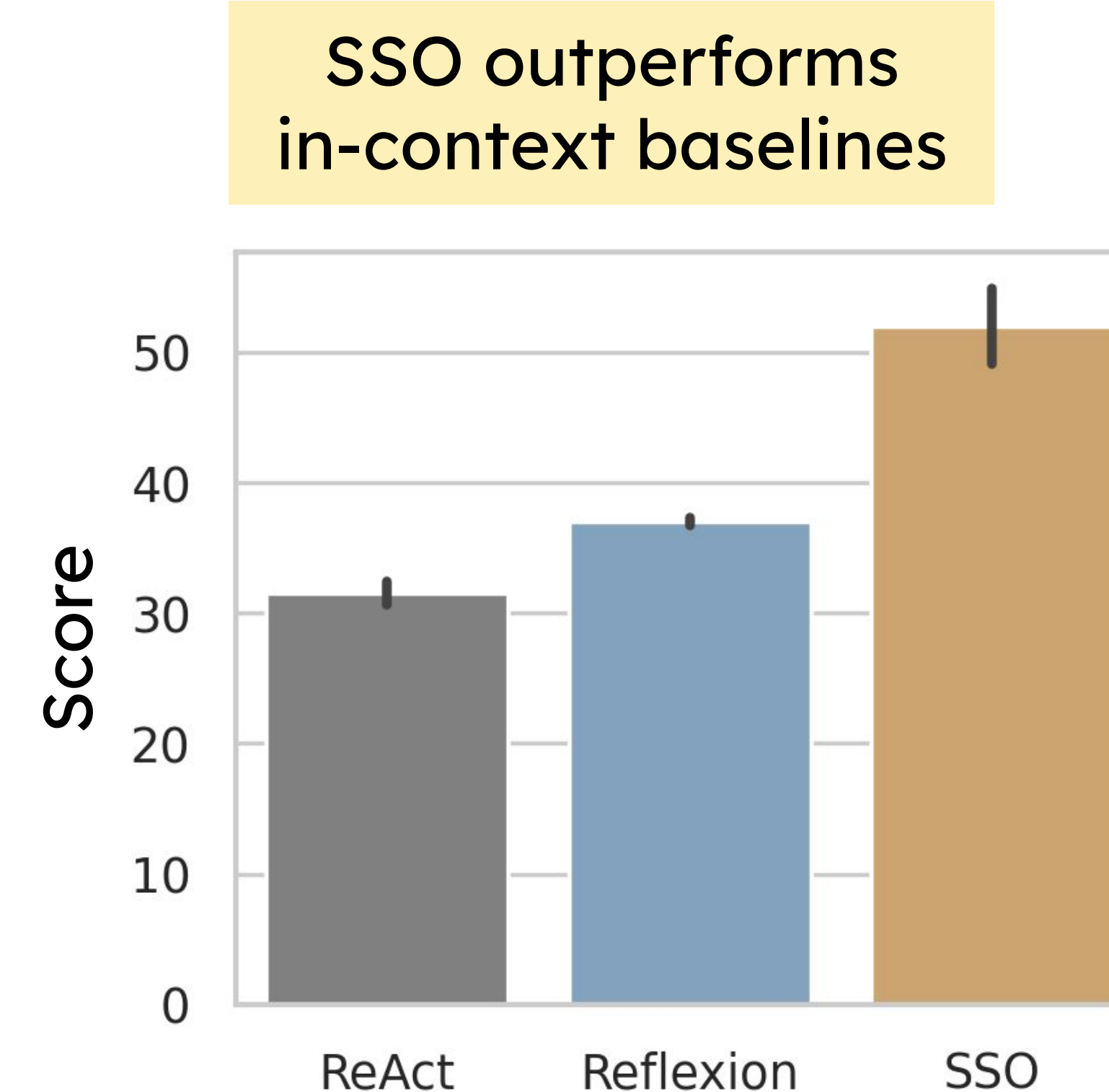
ScienceWorld

Text-based game for simple science experiments and questions. Tests 18 different tasks.

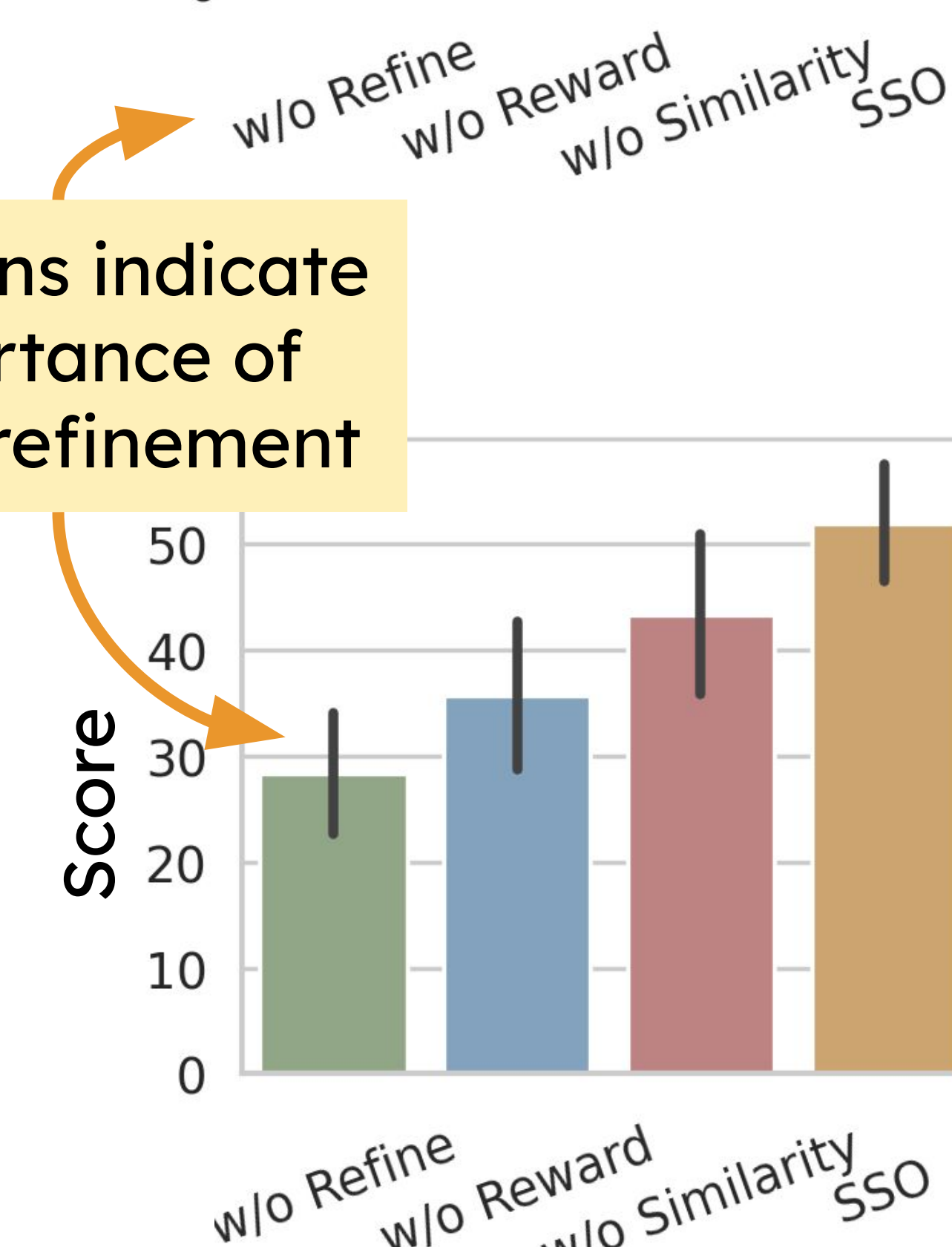


NetHack

Grid-based game that tests spatial reasoning and object interaction. Agent uses the potion to cross the lava to the stairs.



Ablations indicate importance of online refinement



Skills are learned in order

