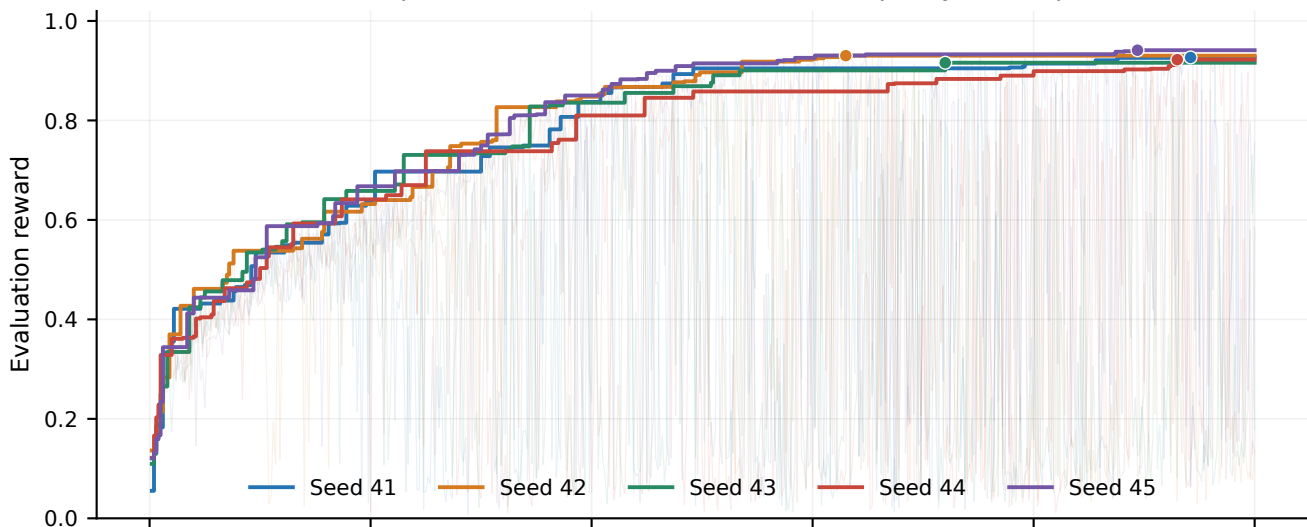


Re100 exploration is volatile while policy discovery progresses monotonically

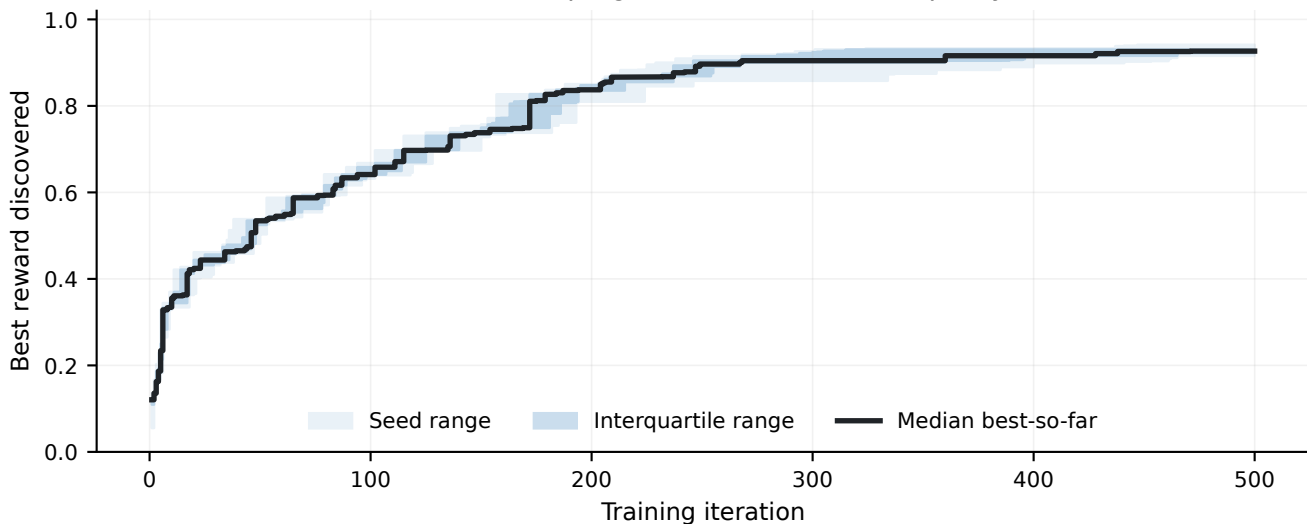
a

Raw exploration outcomes and best-so-far policy envelopes



b

Across-seed progress of the incumbent policy



Thin lines are per-iteration evaluations; stair-step curves retain the best checkpoint found up to each training iteration.