

Sung Cho, Maxwell Delorenzo, Gyubin Han

Optimizing Hyperparameter Tuning Methods for *Evolutionary Strategies*

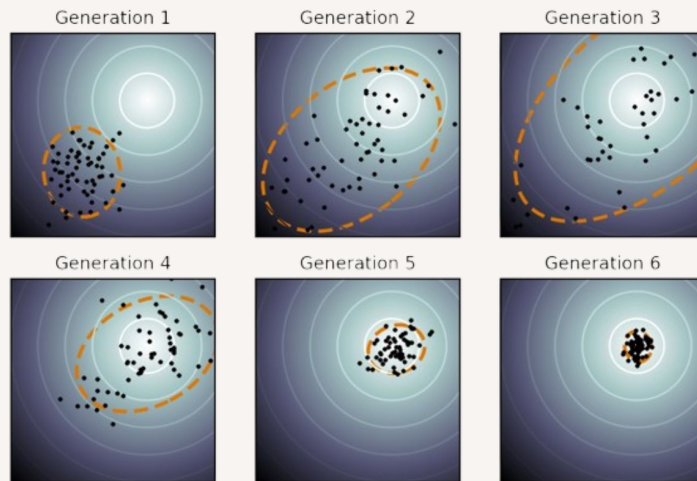
TABLE OF CONTENTS

1. Problem Statement
2. Technical Approach
3. Results
4. Limitations
5. Next Steps

01 Problem Statement

Problem Statement

- **Objective**
 - Explore tuning strategies for evolutionary strategies
 - Develop prompt level policies
- **Motivation**
 - Benefits of Zeroth Order Optimization
- **Constraints**
 - Evaluation Budget
 - Time Constraints
- **Success Criteria**
 - Correctness
 - Comparability
 - Qualitative Optimization Behavior
- **Possible Limitations**



02

Technical Approach

Technical Approach: Algorithms

The RL Objective: $\max_{\theta} \mathbb{E}_{\tau \sim \pi_{\theta}} [R(\tau)]$

Policy Gradient

$$a \sim \pi_{\theta}(a), \quad a \in \{0, 1\}^K$$

- Define a Bernoulli policy over prompt modules

$$\theta \leftarrow \theta + \alpha(R - b) \nabla_{\theta} \log \pi_{\theta}(a)$$

- Update parameters using REINFORCE

Evolutionary Strategies

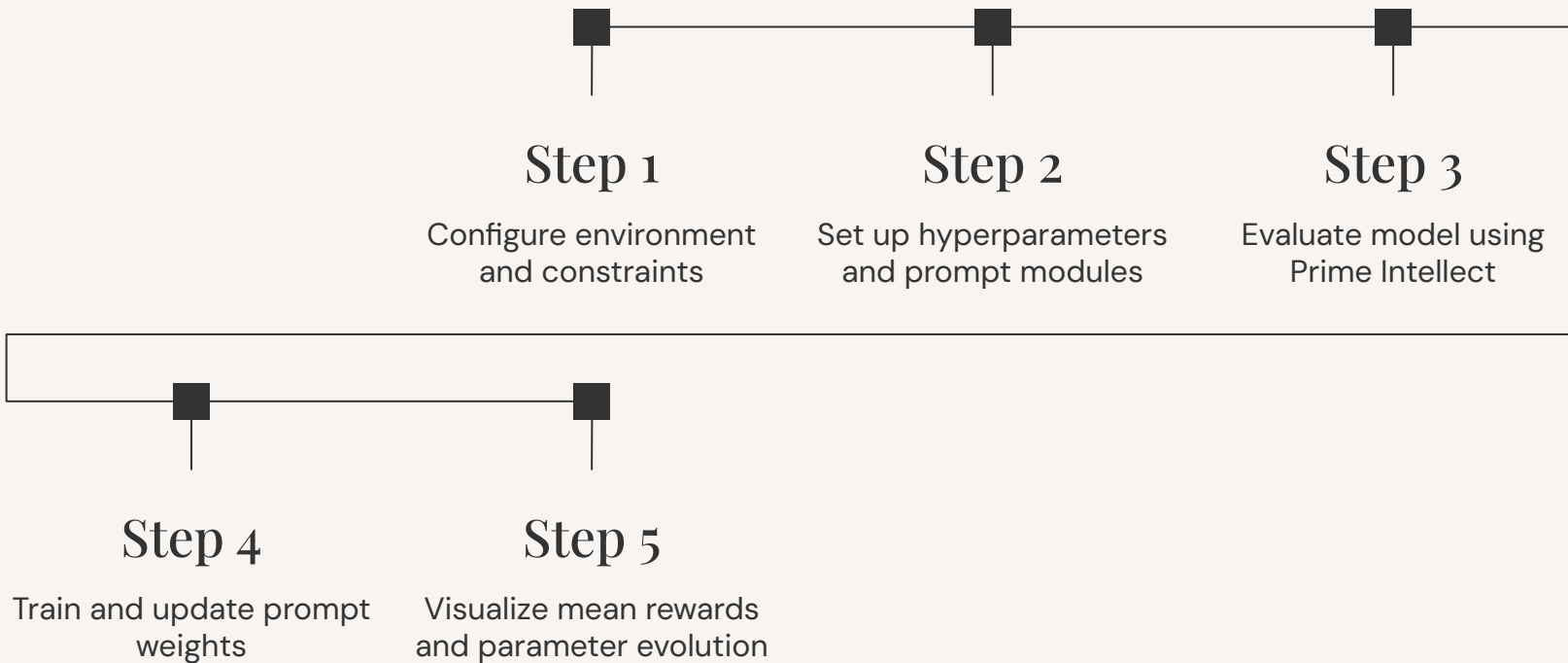
$$\epsilon_i \sim \mathcal{N}(0, I), \quad R_i = R(\theta + \sigma \epsilon_i)$$

- Define prompt parameter and apply Gaussian perturbations

$$\nabla_{\theta} J \approx \frac{1}{N} \sum_i (R_i - b) \epsilon_i.$$

- Calculate a gradient estimate

Technical Approach: Code Pipeline

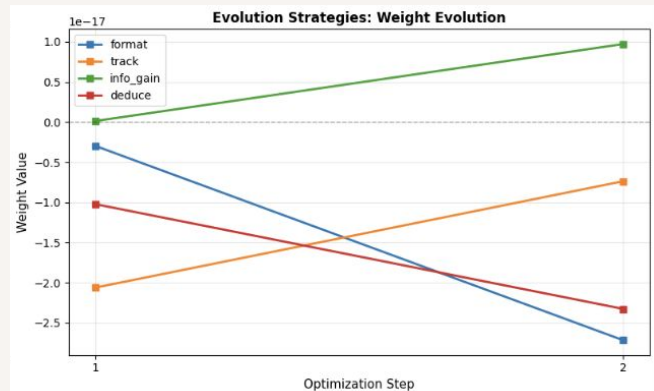
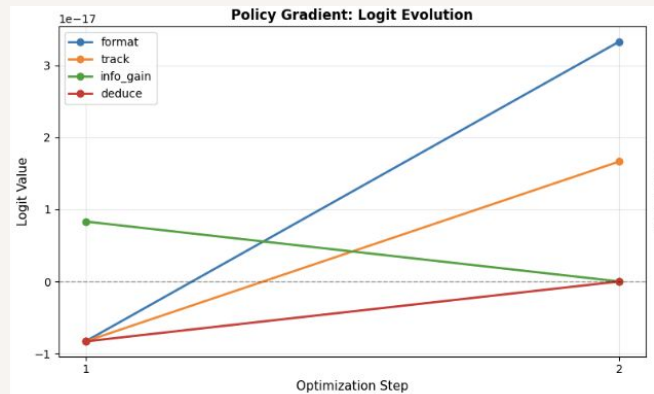


03

Results

Results

- **Initial Results**
 - PG and ES implementation successfully train model
 - Prompt based policies show effectiveness in training
- **What we have learned so far**
 - PG and ES behave qualitatively differently even in small-scale



04

Limitations

Limitations

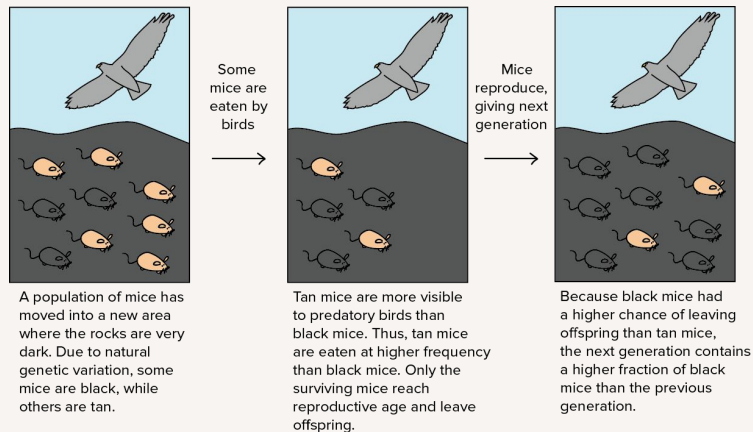
- **Prime Intellect**
 - Reliance on pre-made environments limits direct control
 - Prime Intellect environment lack features to directly change model weights
- **Project-wide**
 - Constraints on model choice by pricing and availability
 - Large scale training is difficult due to evaluation costs

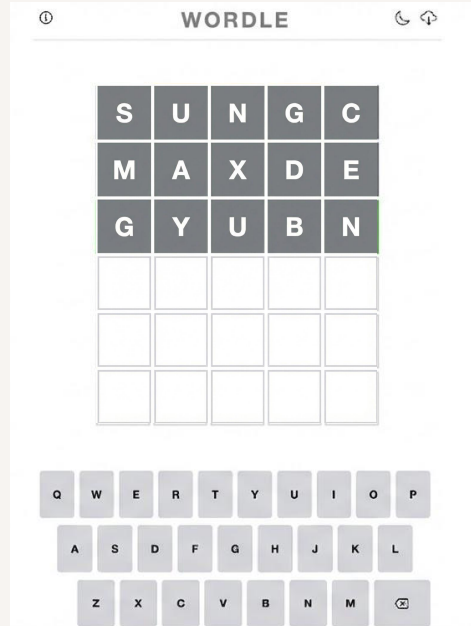
05

Next Steps

Next Steps

- **Immediate Next Steps**
 - Implement and understand model-free random search algorithms
 - [Paper](#)
- **Next Steps for the Project**
 - Run further experiments on PG vs ES for different environments
 - Plot reward trajectories and variance over step size





THANK YOU

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)