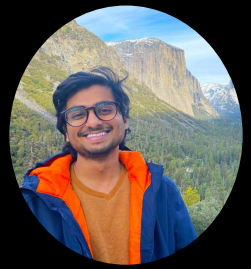
REvolve: Reward Evolution with Large Language Models using Human Feedback



RISHI HAZRA*

ALKIS SYGKOUNAS*

ANDREAS PERSSON

AMY LOUTFI

PEDRO ZUIDBERG DOS MARTIRES

* equal contribution

WASP WALLENBERG AUTONOMOUS SYSTEMS PROGRAM

ÖREBRO UNIVERSITET

REvolve: Reward Evolution

- *We know more than we can tell* (Polanyi's Paradox)

- **What** is REvolve?
  - LLM-based Reward Design Framework for Reinforcement Learning

- **Why** do we need it?
  - Reward Design is hard (think Autonomous Driving).
  - "*Good behavior*" is subjective.
  - REvolve automates *reward generation* and refines them with *behavior verification* by humans.

```python
def reward_fn(speed, collision, min_pos, distance):

    reward_components = {}

    # Speed reward component
    speed_reward = np.exp(-speed_temp * speed_difference)
    reward_components['speed_reward'] = speed_reward

    # Collision reward component (high penalty for collision)
    collision_reward = -100 if collision else 0
    reward_components['collision_reward'] = collision_reward

    # encouraging safe distance maintenance
    distance_reward = np.exp(-distance_temp * max(0, 20 - distance))
    reward_components['distance_reward'] = distance_reward

    ...

    # Calculate total reward
    total_reward = sum(reward_components.values())
    total_reward = np.clip(total_reward, -1, 1)

    return total_reward
```

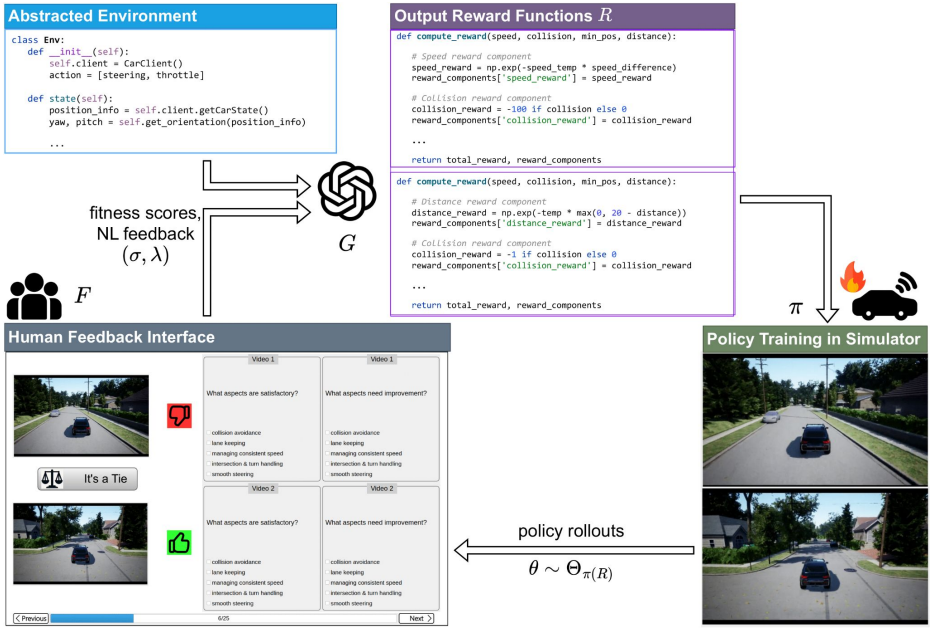
REvolve Framework

REvolve: Reward Evolution with Large Language Models using Human Feedback, Hazra and Sygkounas et al, ICLR 2025

Contributions:

- Evolutionary Algorithms for Reward Design
- Human as fitness functions.
- Elo rating system for ranking

minimal human involvement



Related Works

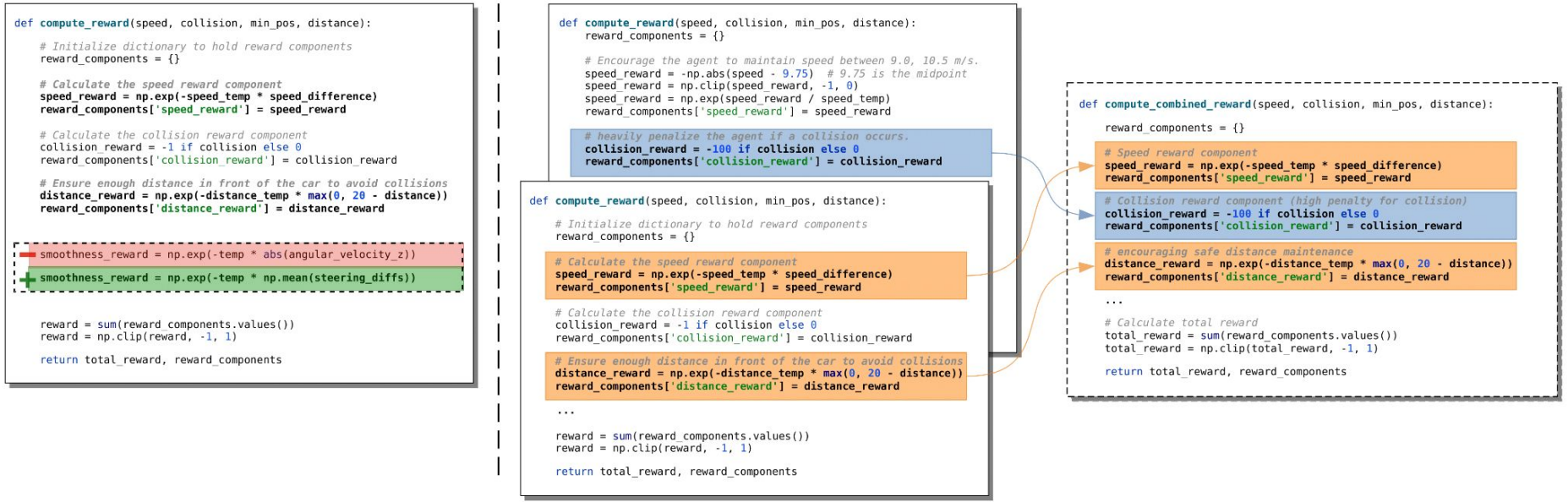
REvolve: Reward Evolution with Large Language Models using Human Feedback, Hazra and Sygkounas et al, ICLR 2025

	Expert	Eureka [1]	REvolve
Reward Generation	Manual	Automated	Automated
Fitness Function	Not required	Required	Not required
Search Strategy	Trial & Error	Greedy Iterative	Evolutionary
Feedback Source	-	Automated	Human / Automated

[1] Eureka: Human-Level Reward Design via Coding Large Language Models, Ma et al, ICLR 2024

LLM Operators

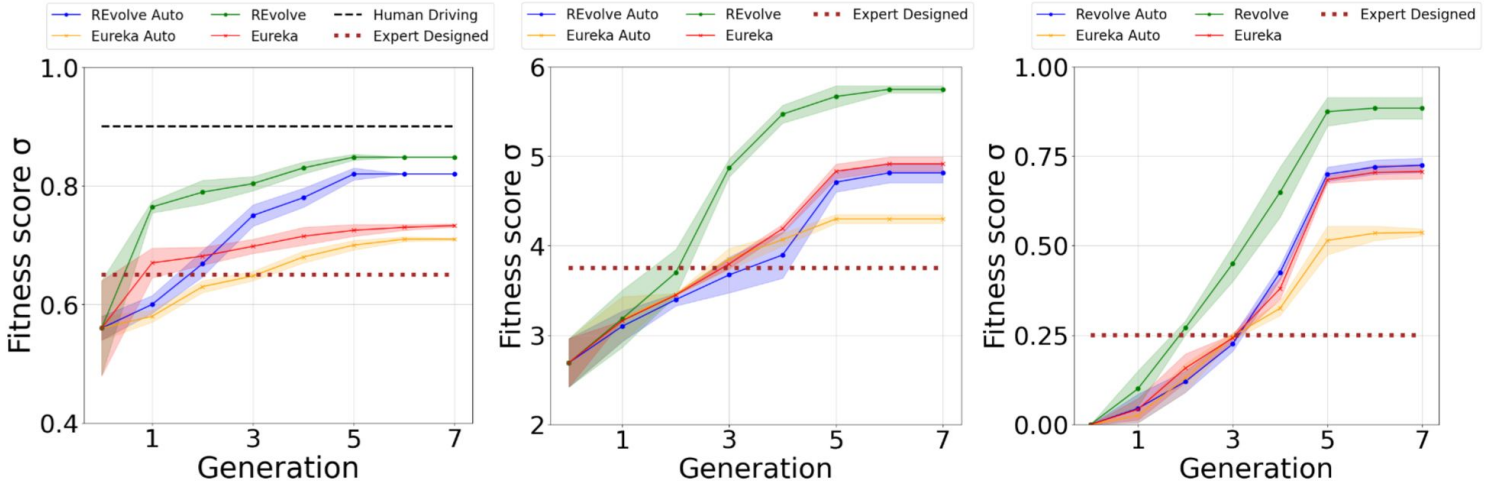
REvolve: Reward Evolution with Large Language Models using Human Feedback, Hazra and Sygkounas et al, ICLR 2025



Mutation **Crossover**

Results: Designed Fitness Function

REvolve: Reward Evolution with Large Language Models using Human Feedback, Hazra and Sygkounas et al, ICLR 2025

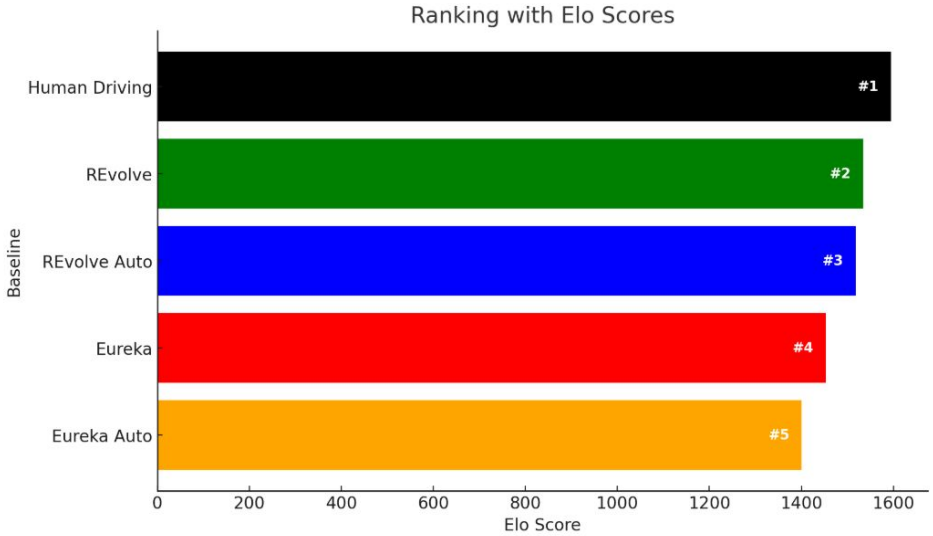


REvolve > REvolve Auto > Eureka > Expert Designed

[1] Eureka: Human-Level Reward Design via Coding Large Language Models, Ma et al, ICLR 2024

Results: Human Evaluation

REvolve: Reward Evolution with Large Language Models using Human Feedback, Hazra and Sygkounas et al, ICLR 2025



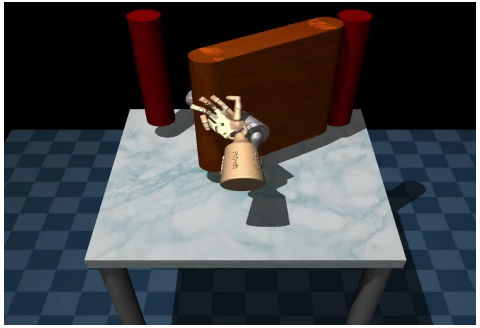
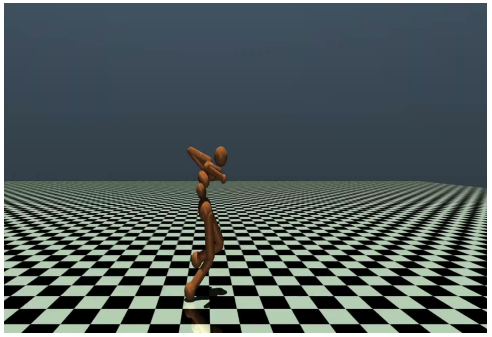
REvolve > REvolve Auto > Eureka > Expert Designed

[1] Eureka: Human-Level Reward Design via Coding Large Language Models, Ma et al, ICLR 2024

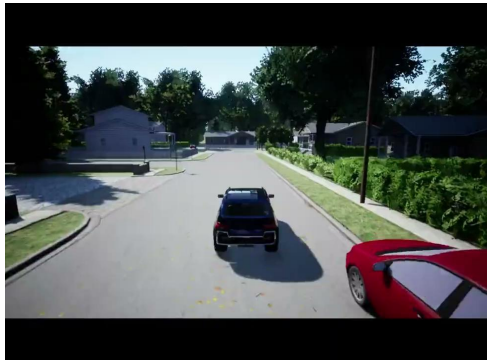
REvolve vs. Eureka

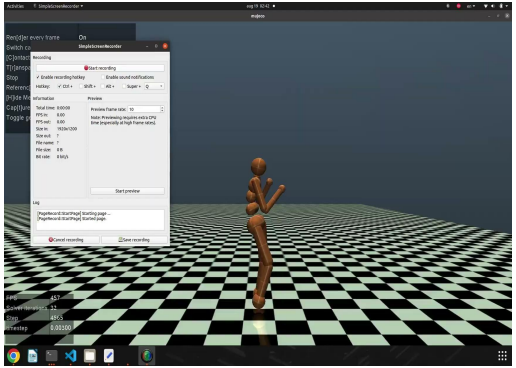
REvolve: Reward Evolution with Large Language Models using Human Feedback, Hazra and Sygkounas et al, ICLR 2025
Eureka: Human-Level Reward Design via Coding Large Language Models, Ma et al, ICLR 2024

REvolve

Eureka





REvolve: Reward Evolution with Large Language Models using Human Feedback

https://rishihazra.github.io/REvolve/