Designing Fun: LLM-based Game Level Generation via Agent Gameplay Feedback

You-Zhe Xie [1] Yu-Hsuan Li [1] Cheng-Chih Tsai [1]

[1]National Yang Ming Chiao Tung University

Abstract

We use GRPO to train a large language model for MiniGrid level generation with a fun-aligned reward. The reward encourages skill-discriminative challenges, meaningful object interactions, and diverse level layouts, producing levels that are not only playable but also fun and behaviorally rich.

Problem - Game Level Generation

- **Problem 1:** Level design is labor-intensive..
- **Problem 2:** There is no generative method to generate **fun** level.

=> How to generate fun game levels? We take **Minigrid** as game environment

Motivation - LLMs is good but not enough

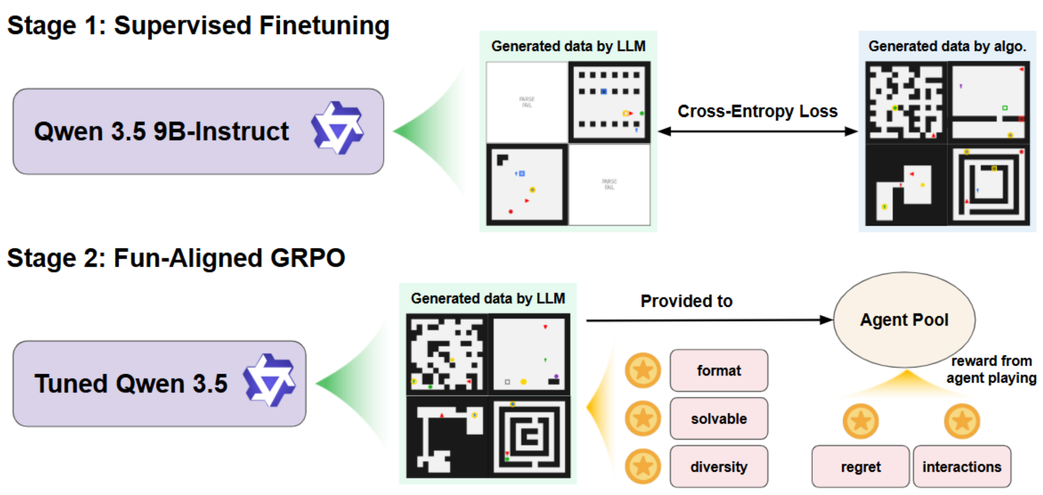
👍: LLMs have rich game design knowledge priors

🥲: LLMs don't know how to generate legal, fun levels

Large Language Models and Games: A Survey and Roadmap

Roberto Gallotta[1] *Graduate Student Member, IEEE*, Graham Todd[2] *Graduate Student Member, IEEE*, Marvin Zammit[1] *Graduate Student Member, IEEE*, Sam Earle[2] *Graduate Student Member, IEEE*, Antonios Liapis[1] *Member, IEEE* Julian Togelius[2] *Senior Member, IEEE*, and Georgios N. Yannakakis[1], *Fellow, IEEE*

Study methodology



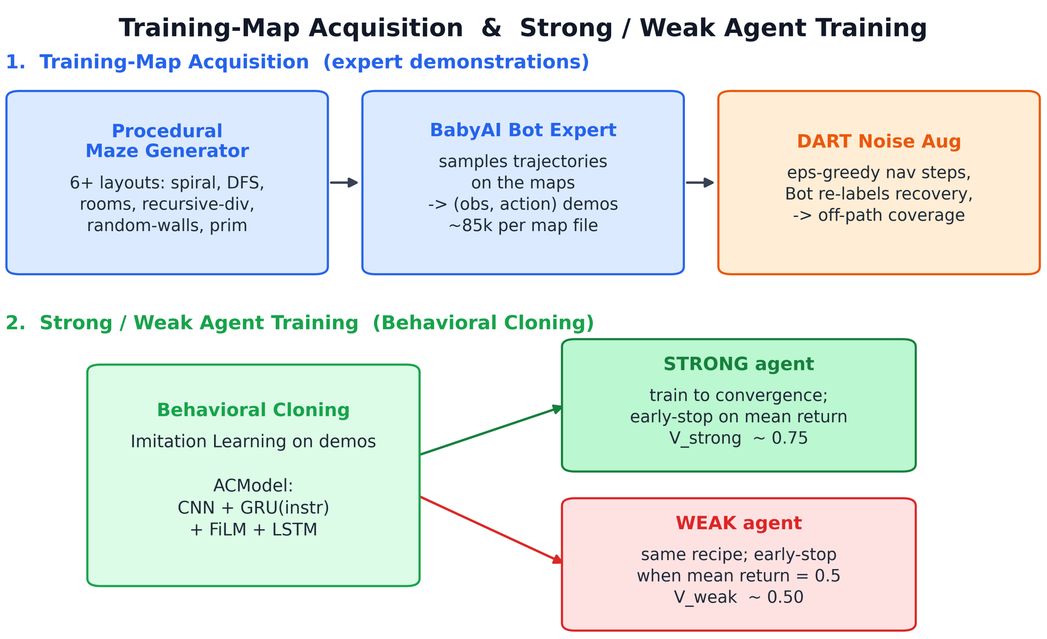
Stage 1: 11 maze-generation algorithms (random DFS, Prim's, recursive division, etc.) are used to construct supervised fine-tuning (SFT) data for Qwen.

Stage 2: Multiple rewards capturing aspects of "fun" (skill discrimination, diversity, and interaction richness) are optimized via GRPO because of sparseness of rewards.

Fun-aligned reward:

- **Format:** LLM output is parseable as a valid MiniGrid environment.
- **Solvability:** Generated level is confirmed solvable by BFS.
- **Regret:** Performance gap between a strong and a weak agent; rewards skill-discriminative levels.
- **Interaction:** Number of object interactions the agent performs in an episode.
- **Diversity:** Mean pairwise Hamming distance across a group of generated levels.

Map collection and Agent training

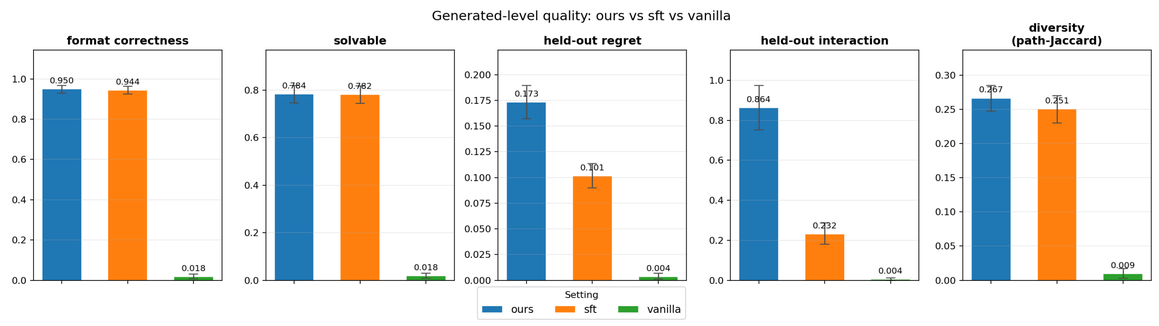


Score / validation: roll the agent on 1000 val maps; mean return = 1 - 0.9 x steps / 256 (0 if unsolved). Expert (Bot) upper bound ~ 0.81.

Results and discussion

We evaluate generated levels from both structural and agent-behavior perspectives.

- **Format correctness:** outputs of LLM can be parsed into valid MiniGrid levels.
- **Solvability:** generated levels with a BFS path from start to goal.
- **Held-out regret:** return gap between held-out strong and weak agents trained on different random-generated dataset; higher means more skill-discriminative.
- **Held-out interactions:** effective object interactions by held-out agents.
- **Path diversity:** Jaccard distance between 5 held-out agent routes; higher means more diversity.



Our GRPO-trained model improves gameplay-oriented quality over both vanilla and SFT baselines. It preserves high format correctness and solvability while achieving substantially higher held-out regret, object interactions, and path diversity, indicating more skill-discriminative and behaviorally rich levels.

Ablation

Table 1. Ablation study on methods excluding one of rewards

model	format correctness ↑	solvable ↑	held-out regret ↑	held-out interactions ↑	diversity (path-Jaccard) ↑
ours	95.0%	79.4%	**0.173**	0.864	0.324
w/o regret reward	96.6%	**88.2%**	0.152	**1.106**	0.291
w/o interaction reward	**97.6%**	81.2%	0.152	0.252	**0.360**
w/o diversity reward	95.4%	87.8%	0.156	1.024	0.309

Ablations show that each reward component shapes a distinct aspect of level quality: regret improves skill discrimination, interaction reward promotes object use, and diversity reward supports varied paths. The full reward achieves the best overall balance and the highest held-out regret, suggesting that combining rewards encourages richer exploration and more skill-discriminative levels.

Conclusions

- We propose a two-stage SFT + GRPO framework that trains an LLM to generate fun game levels.
- We decompose "fun" into three complementary rewards — skill discrimination (regret), interaction richness, and layout diversity — each grounded in actual agent gameplay and structural property.
- Our model substantially outperforms baselines on all gameplay-oriented metrics, and ablations confirm each reward captures a distinct aspect of fun.

Appendix

Project Page

Github