Exploring Generalisation in NLP and RL via Baby-AI

1. Introduction

Language-guided reinforcement learning (RL) sits at the intersection of natural language processing and decision-making. Unlike traditional RL, where agents act on low-dimensional states, text- or instruction-based environments require agents to interpret instructions, ground them into actions, and execute multi-step plans. This makes such environments powerful testbeds for studying **generalization, compositionality, and reasoning** in AI.

Recent progress in pretrained language models (LMs) has improved zero-shot instruction following, but when paired with RL agents, these models often **fail to generalize systematically**. For example, an agent trained on “pick up the red ball” and “pick up the green box” usually cannot solve “pick up the red box.” This gap between *pattern memorization* and *systematic generalization* remains one of the central open challenges.

2. Problem Statement

Instruction-following RL agents in BabyAI-like environments (simple gridworlds with language instructions) are heavily limited by their lack of **compositional generalization**. Existing approaches often succeed in narrow training distributions but collapse on unseen combinations of known primitives. The problem can be phrased as:

- *How can RL agents trained on a finite set of instructions generalize to novel compositions of those instructions without explicit retraining?*

Even for LLMs if you only ever asked it to pick up red balls and green boxes during training/fine-tuning, and then suddenly say *“pick up red box”*, the LLM-as-agent doesn’t automatically transfer the concept algebra the way humans do. It often hallucinates or just fails unless explicitly fine-tuned on those compositions.

Gap: While LLMs provide linguistic prior knowledge, and BabyAI benchmarks have exposed generalization issues, there is no widely accepted method that achieves *robust compositional generalization* in instruction-following RL.

3. What we can work on

I propose to work on a **language-grounded RL architecture** to tackle compositional generalization in BabyAI-like environments. Our main goal would be to generalise the tasks. And there are many generalisation tests like one mentioned in the example above or another

can be say we trained in 2 different environments in one it learnt to pick up the key and unlock the door, in another to pick a ball… Now it should work in a room where you need to open a door and pick a ball.

4. Expected Results & Contributions

- Demonstrate improved **systematic generalization** over existing RL and language-conditioned baselines.

- Release code and evaluation splits to the community as a benchmark for future work.

5. More Plans

ARC AGI benchmarks are somewhat based on simmilar concept of generalisation by llms. If we succeeded in the above tasks and if we got some innovative methods we can move-on to working with ARC AGI Benchmark(Although it does not seem to be so easy to reach there).

6. Relevant Papers I Found

https://arxiv.org/pdf/2203.04806 (One-Shot Learning from a Demonstration with Hierarchical Latent Language)

https://openreview.net/pdf?id=Y87Ri-GNHYu (ASK YOUR HUMANS: USING HUMAN INSTRUCTIONS TO IMPROVE GENERALIZATION IN REINFORCEMENT LEARNING)