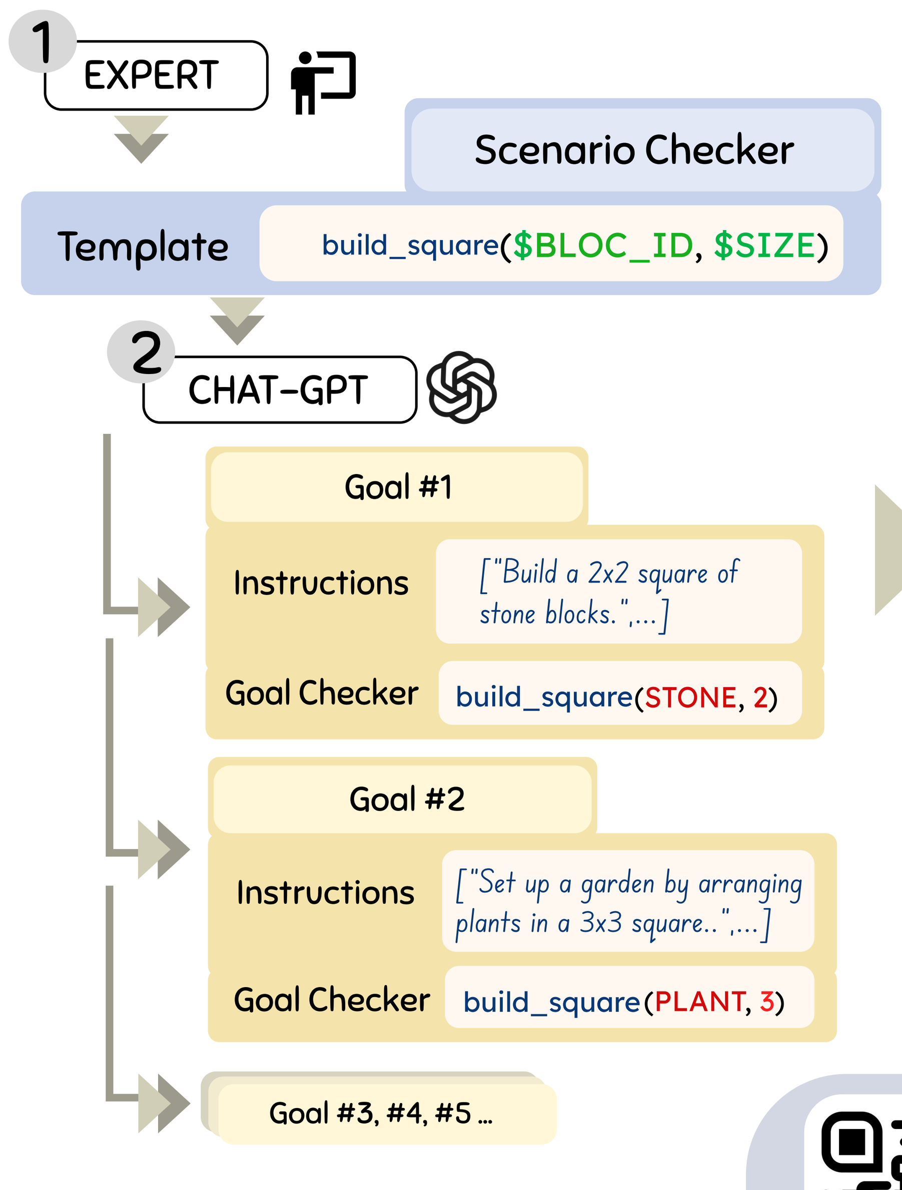
CrafText Benchmark: Advancing Instruction Following in Complex Multimodal Open-Ended World

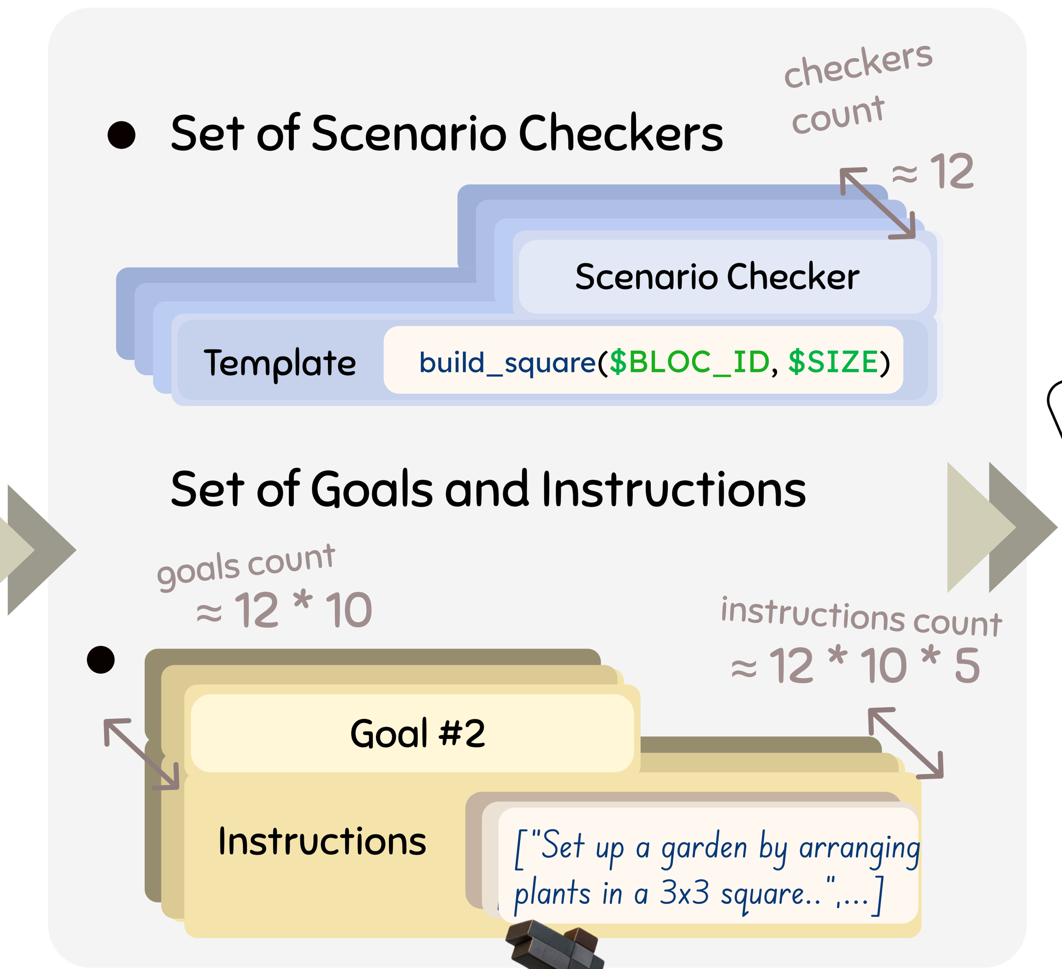
Zoya Volovikova [1,2], Gregory Gorbov[2,3], Petr Kuderov[1,2], Aleksandr Panov[1,2,3], Alexey Skrynnik[1,2]

[1] AIRI, Moscow, Russia, [2] MIPT, Moscow, Russia, [3] FRC CSC RAS, Moscow, Russia

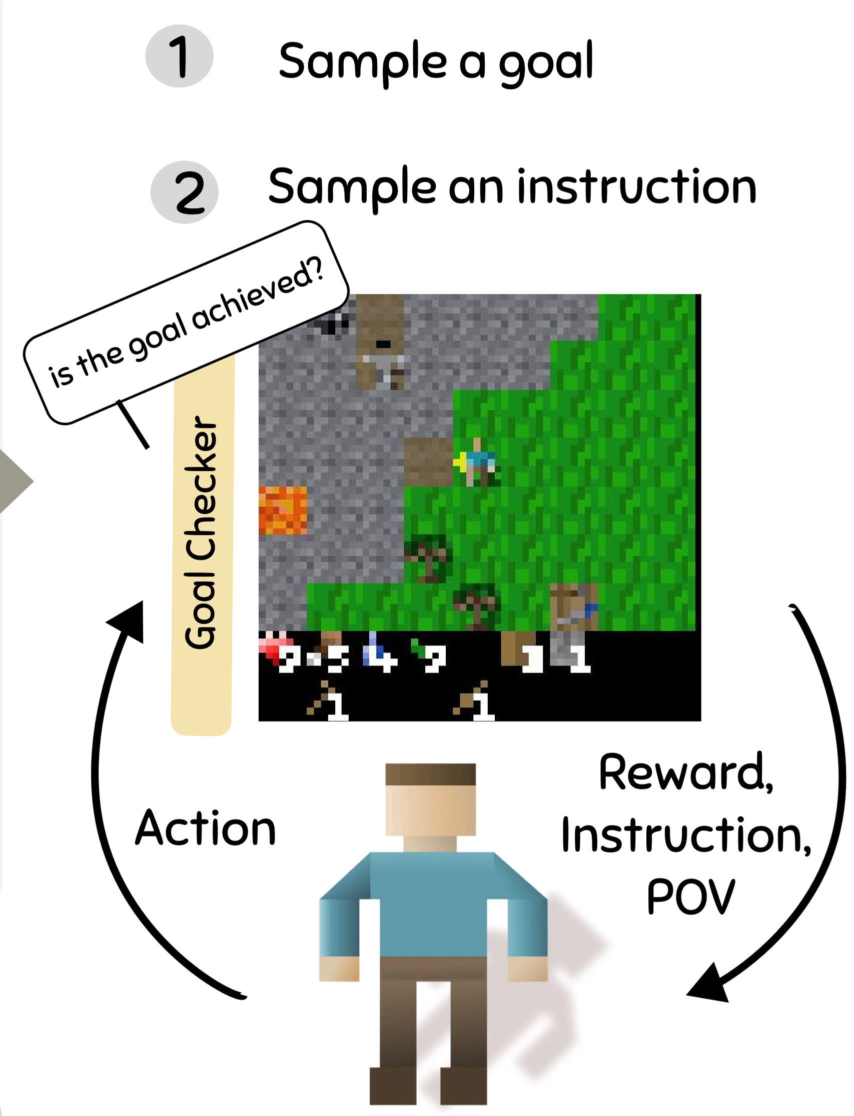
Gathering Pipeline



Dataset



Environment



Task Definition

Instruction Following tasks involve providing the agent with an observation $o = (I, g)$, where It is a visual input and g is an instruction that defines the goal $\tau(g)$. The task is to extract the goal $\tau(g)$ from the instruction g and select actions $a \in A$ to maximize g reward. The environment transitions via $P(s'|s, a)$. The optimal policy π^* maximizes the expected reward.

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{T} \gamma^t R(s_t, a_t, g) \mid o_0 \right].$$

*At the same time, the environment is stochastic and dynamic.

Dataset Overview

Train set | Paraphrased | New objects

Task Categories

Achievements

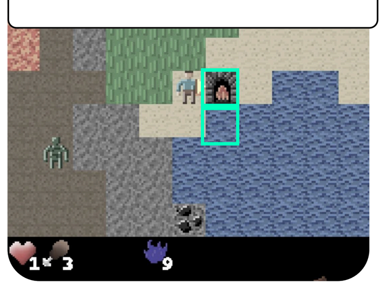
Find a cave or a mine with iron ore blocks.

Conditional

Before collecting a stone, place a furnace.

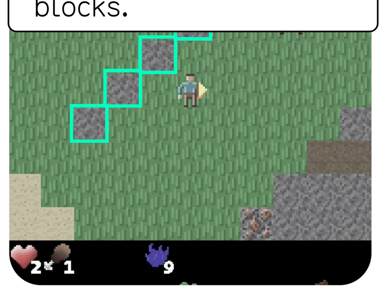
Localization

Place furnaces north of the lake.



Build

Construct a diagonal line using 4 stone blocks.



Experiments

Baselines

PPO-T – PPO with BERT-based instruction embeddings for improved language understanding.

PPO-T+ – PPO-T model using ChatGPT-generated multi-step plans instead of single-step instructions.

Dynalang – Model-based RL with integrated language processing in the dreaming phase of learning.

FiLM – Feature-wise Linear Modulation layers in the actor-critic network.

Baseline Comparison

Instruction type	Algorithm	Conditional	Build	Localization	Achievements	Total
Train Set	PPO-T	0.15	0.25	0.33	0.55	0.40
	PPO-T+	0.17	0.24	0.30	0.70	0.45
	Dynalang	0.00	0.12	0.15	0.17	0.15
	FiLM	0.07	0.38	0.29	0.76	0.43
Paraphrased	PPO-T	0.12	0.13	0.35	0.50	0.36
	PPO-T+	0.16	0.17	0.30	0.48	0.35
	Dynalang	0.00	0.09	0.13	0.10	0.05
	FiLM	0.10	0.20	0.30	0.53	0.35
New objects	PPO-T	0.12	0.13	0.17	0.34	0.22
	PPO-T+	0.20	0.17	0.19	0.43	0.28
	Dynalang	0.00	0.09	0.09	0.14	0.10
	FiLM	0.17	0.20	0.19	0.38	0.26

Results

— Handling complex instructions in dynamic environments remains a significant challenge for generalization.

All baselines demonstrate limited training performance. Dynalang achieves only a 0.15 success rate (SR), while PPO-T (0.40 SR), PPO-T+ (0.45), and FiLM (0.43) perform moderately better using BERT-based instruction encoding.

— Existing methods show limited robustness to linguistic variation.

All models show a performance drop on the Paraphrased test set. PPO-T+ is most affected (-0.10 SR).

— Transforming the instruction into a plan helps to generalize to unseen goals.

PPO-T+ achieves the highest SR on the New Objects test set (0.28), outperforming FiLM (0.26), PPO-T (0.22), and Dynalang (0.10). This suggests PPO-T+ generalizes better to novel goals by effectively decomposing instructions into reusable subtasks. FiLM shows competitive results, likely due to its flexible text-visual integration via FiLM layers

We aim for greatest challenges

1. Solving Instruction-Following Tasks in Dynamic Environments
2. Multi-Step Desidion Making with Implicit Preconditions
3. Linguistic Variation and Paraphrasing
4. Generalization to novel goals
5. Balancing Environment Exploration and Instruction Following