NEURAL INFORMATION PROCESSING SYSTEMS

Language Grounded Multi-agent Reinforcement Learning with Human-interpretable Communication

Huao Li, Hossein Nourkhiz Mahjoub, Behdad Chalaki, Vaishnav Tadiparthi, Kwonjoon Lee, Ehsan Moradi Pari, Michael Lewis, Katia Sycara

Honda Research Institute US

Carnegie Mellon University

University of Pittsburgh

Background

Ad-hoc human-agent Teamwork

- Collaborate with **unseen** humans without pre-coordination
- Communicate in human-interpretable language

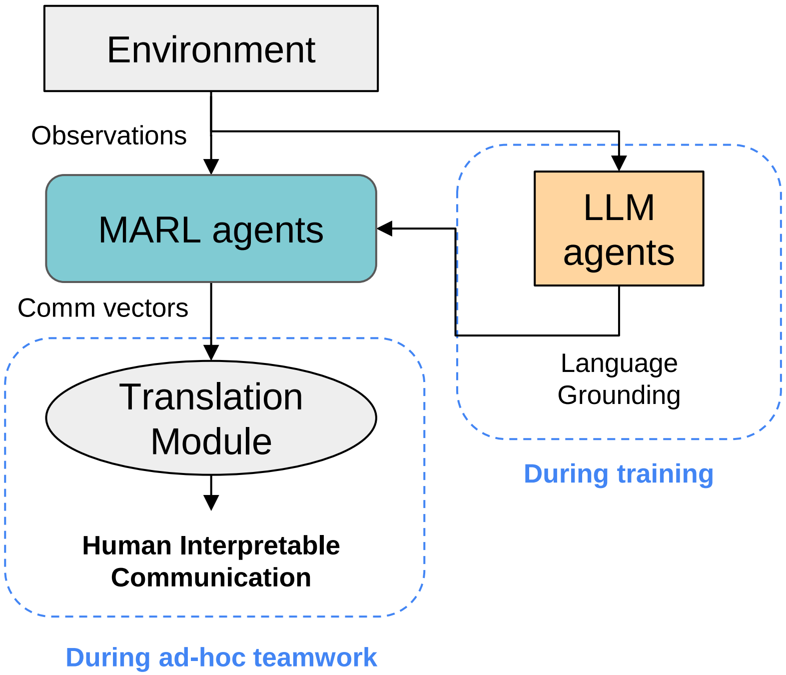
Multi-agent Reinforcement Learning with Communication (MARL-comm)

- Optimal task performance
- Not human **interpretable**

Embodied agents based on Large Language Models (LLMs)

- Common sense reasoning and human-like communication
- Suboptimal performance due to **hallucinations**

LangGround



Environment
Observations
MARL agents
LLM agents
Language Grounding
Comm vectors
Translation Module
Human Interpretable Communication
During training
During ad-hoc teamwork

LangGround

Reinforcement Learning objective

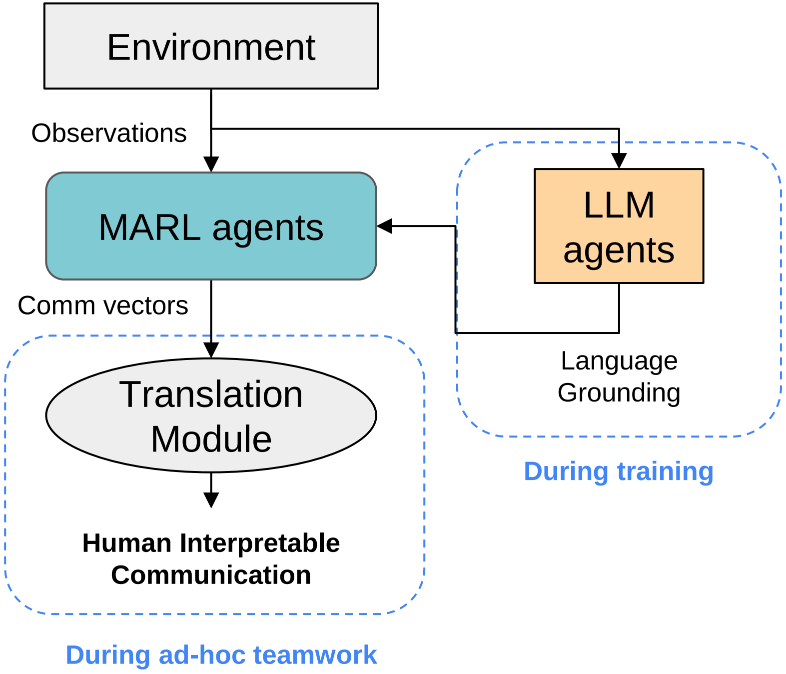
$$\max_{\pi^i:\mathcal{S}\to\mathcal{A}\times\mathcal{C}} \mathbb{E}\Big[\sum_{t\in T}\sum_{i\in\mathcal{I}} \gamma^t \mathcal{R}(o_t^i, a_t^i) \,|\, a_t^i \sim \pi^i, o_t^i \sim \mathcal{O}\Big]$$

Language Learning objective

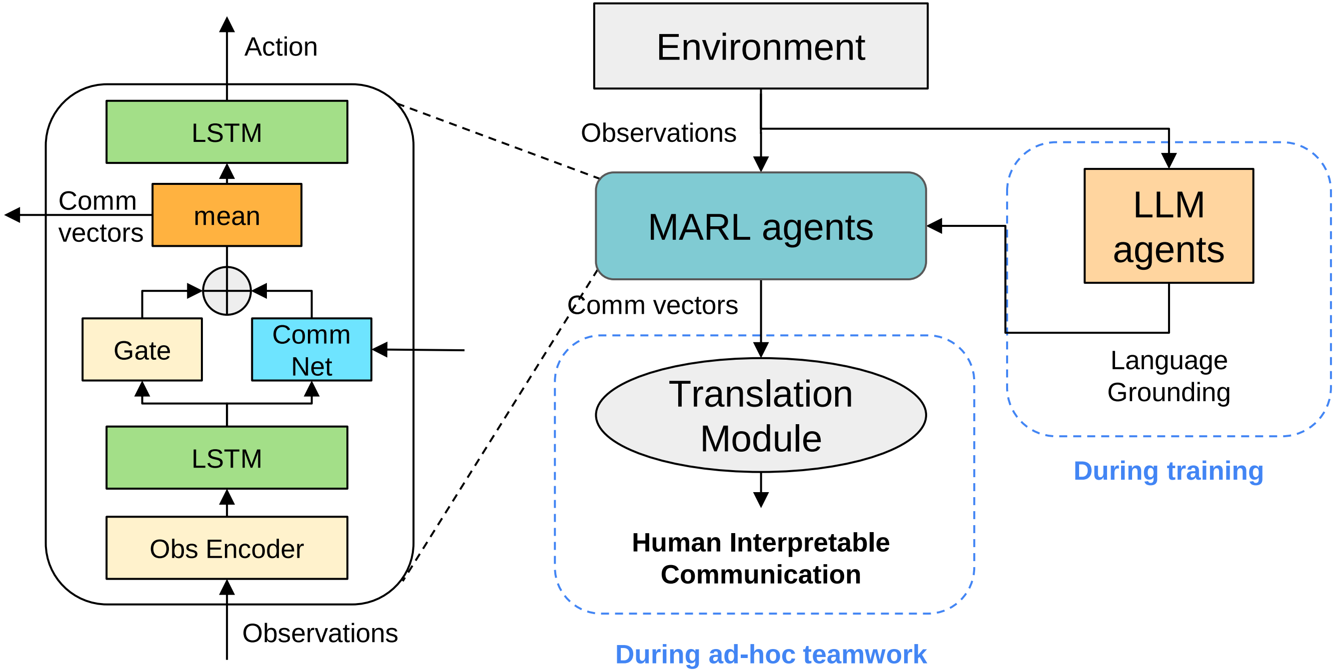
- L: O ➜ C
- Mimic target language L* ∈ L

Jointly optimize RL and SL loss

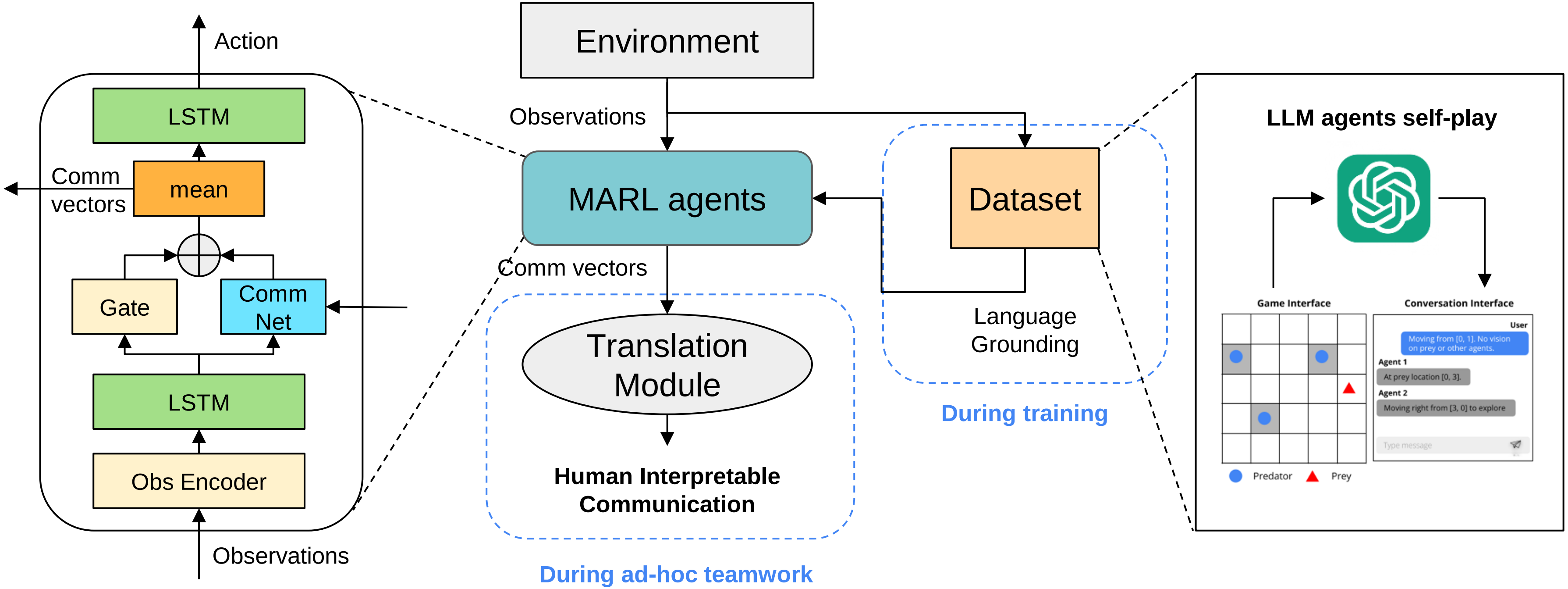
$$L = L_{RL} + \lambda L_{sup}$$



LangGround


Action
LSTM
Comm vectors
mean
Gate
Comm Net
LSTM
Obs Encoder
Observations
Environment
Observations
MARL agents
LLM agents
Comm vectors
Language Grounding
Translation Module
Human Interpretable Communication
During training
During ad-hoc teamwork

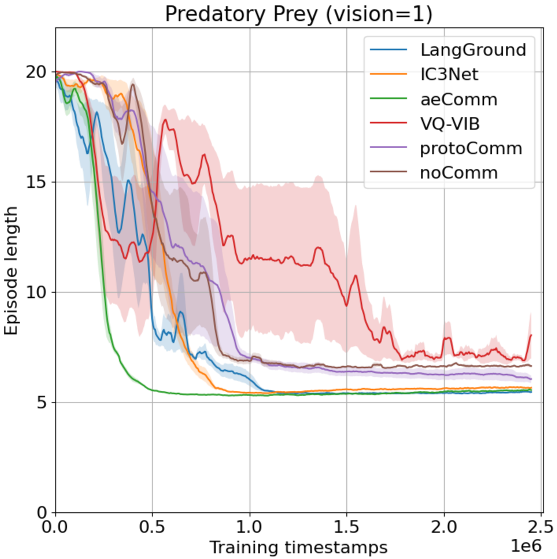
LangGround

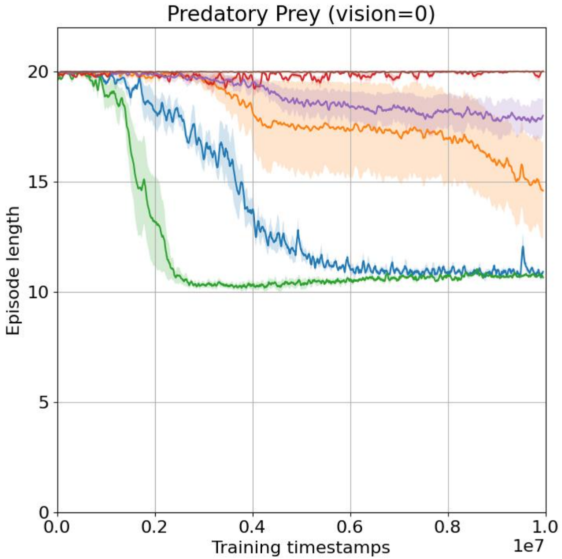

Action
LSTM
Comm
vectors
mean
Gate
Comm
Net
LSTM
Obs Encoder
Observations
Environment
Observations
MARL agents
Comm vectors
Translation
Module
Human Interpretable
Communication
During ad-hoc teamwork
Dataset
Language
Grounding
During training
LLM agents self-play
Game Interface
Conversation Interface
User
Moving from [0, 1]. No vision on prey or other agents.
Agent 1
At prey location [0, 3].
Agent 2
Moving right from [3, 0] to explore
Type message
Predator
Prey

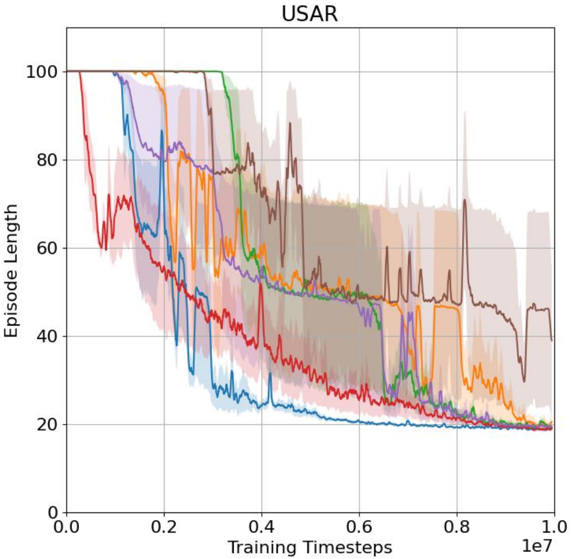
Task Performance

Takeaways

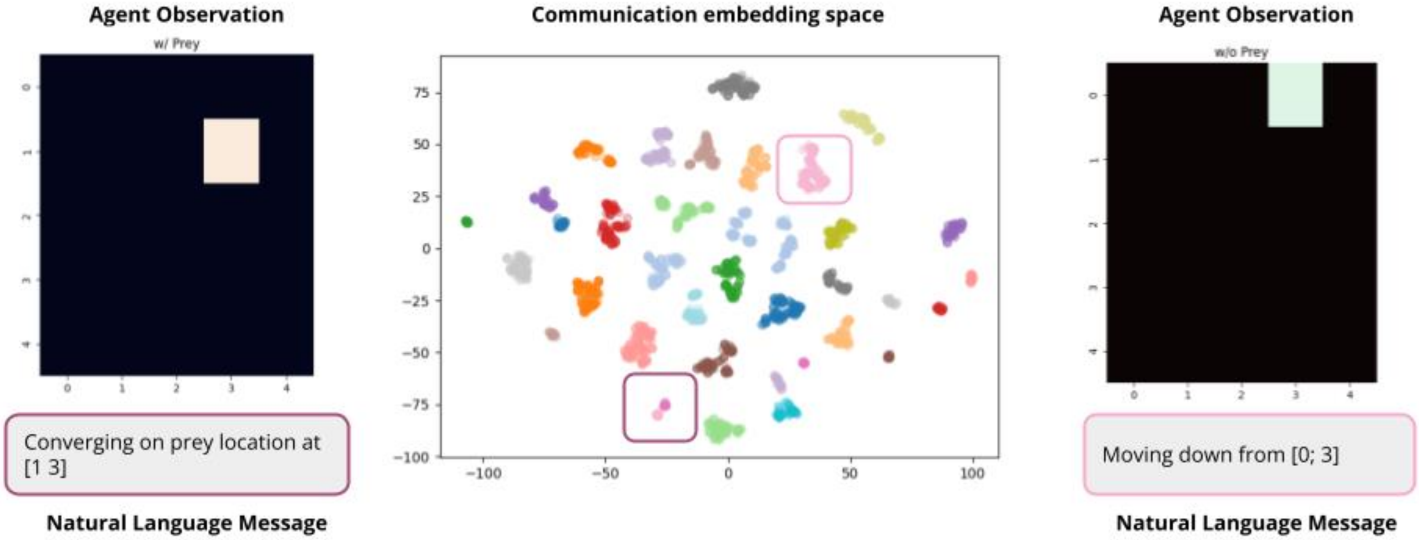
- On par final task performance with SOTA methods
- Converge faster in complicated task environments







Alignment



Agent Observation
w/ Prey
Converging on prey location at [1 3]
Natural Language Message
Communication embedding space
Agent Observation
w/o Prey
Moving down from [0; 3]
Natural Language Message

Zero-shot generalization

Methods

- Remove certain prey locations during training
- Compre LangGround agent's comm vector in novel states with LLM agent's communication

Takeaways

- LangGround is not memorizing, but aligning agent's comm space with embedding space of human language

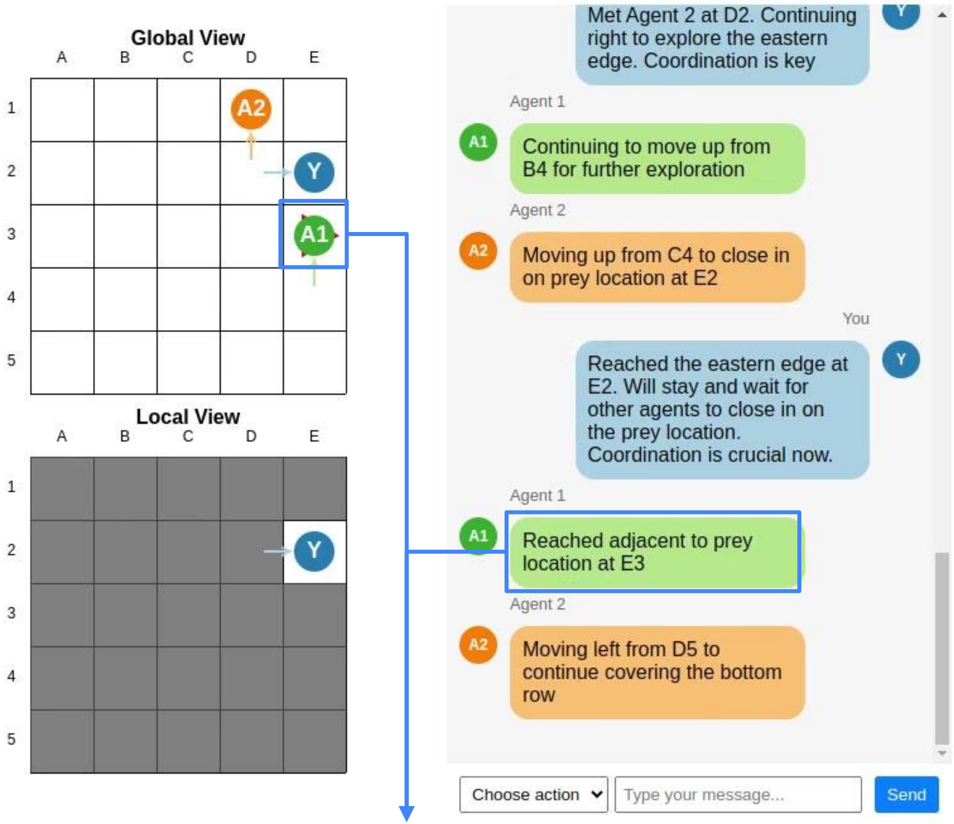
Table 2: Zero-shot communication in pp_{v0}

Prey Loc	Cos sim	Bleu score	Example message
(1,1)	0.81	0.41	Moving up to converge on prey location at (1,0) for capture
(1,3)	0.81	0.27	Converging on prey location at (1,3)
(3,1)	0.82	0.51	Moving up towards prey location at (3,1)
(3,3)	0.78	0.72	Converging on prey location at (3,3)

Ad-hoc Teamwork

Demo: LLM agents + LangGround

- **LangGround** agents can accurately share task-related information with LLMs in natural language

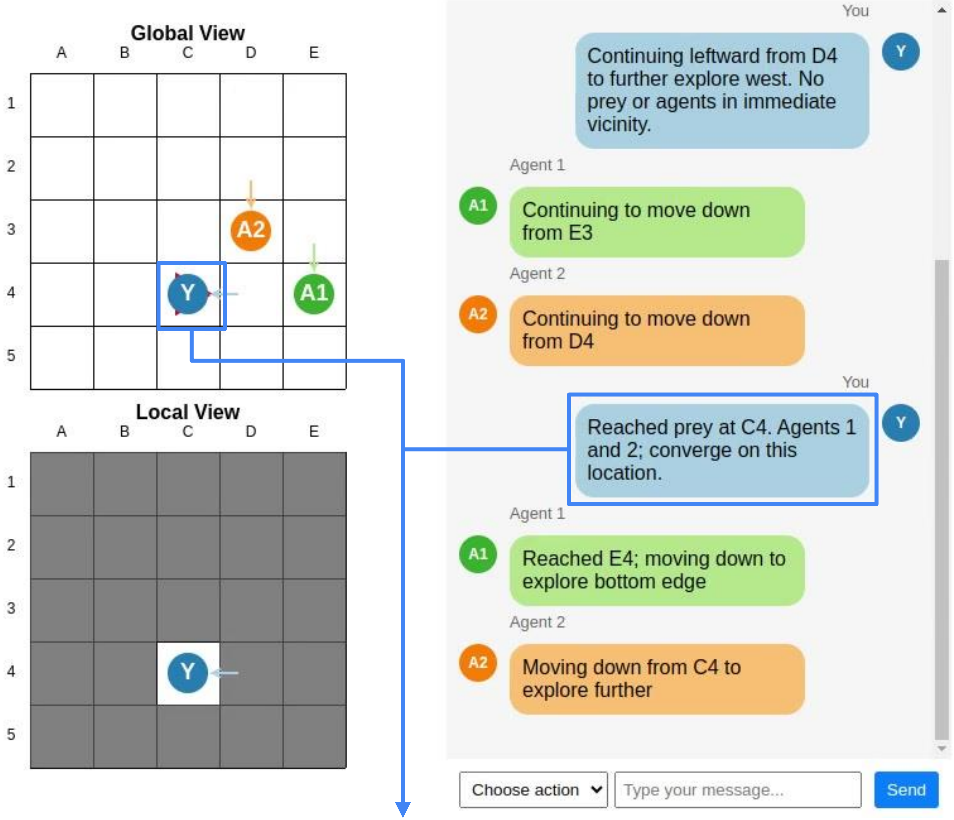


Agent 1 finds the prey, and accurately reports its location

Ad-hoc Teamwork

Demo: LLM agents + LangGround

- **LangGround** agents can accurately share task-related information with humans in natural language
- **LangGround** agents are able to understand novel messages generated by LLMs and behave accordingly



LLM finds the prey first and communicates this information

Ad-hoc Teamwork

Demo: LLM agents + LangGround

- **LangGround** agents can accurately share task-related information with humans in natural language
- **LangGround** agents are able to understand novel messages generated by humans and behave accordingly
- **LangGround** agents perform better than other methods in ad-hoc teamwork with unseen agents

Team composition	Predator Prey (vision = 1)	Predator Prey (vision = 0)
LangGround + LLMs	**8.5 steps**	**15.5 steps**
Autoencoder + LLMs	10.3 steps	17.5 steps
RL w/o Comm + LLMs	10.6 steps	20.0 steps

LangGround agents take less steps in completing the task

Takeaways

Summary

- We propose **LangGround**, a MARL pipeline to train agents with human-interpretable communication
- Align multi-agent communication space with human language by combining SL and RL
- Collect synthetic human data of team behaviors and communication from embodied LLM agents

Contributions

- Enhance the robustness of emergent communication learning via groundings provided by LLM agents
- Learn human interpretable communication protocols across diverse tasks
- Enable ad-hoc teamwork between MARL, LLM, and humans without pre-coordination