



Multi-agent Systems and Human-AI Collaboration

Prithviraj Ammanabrolu



Logistics

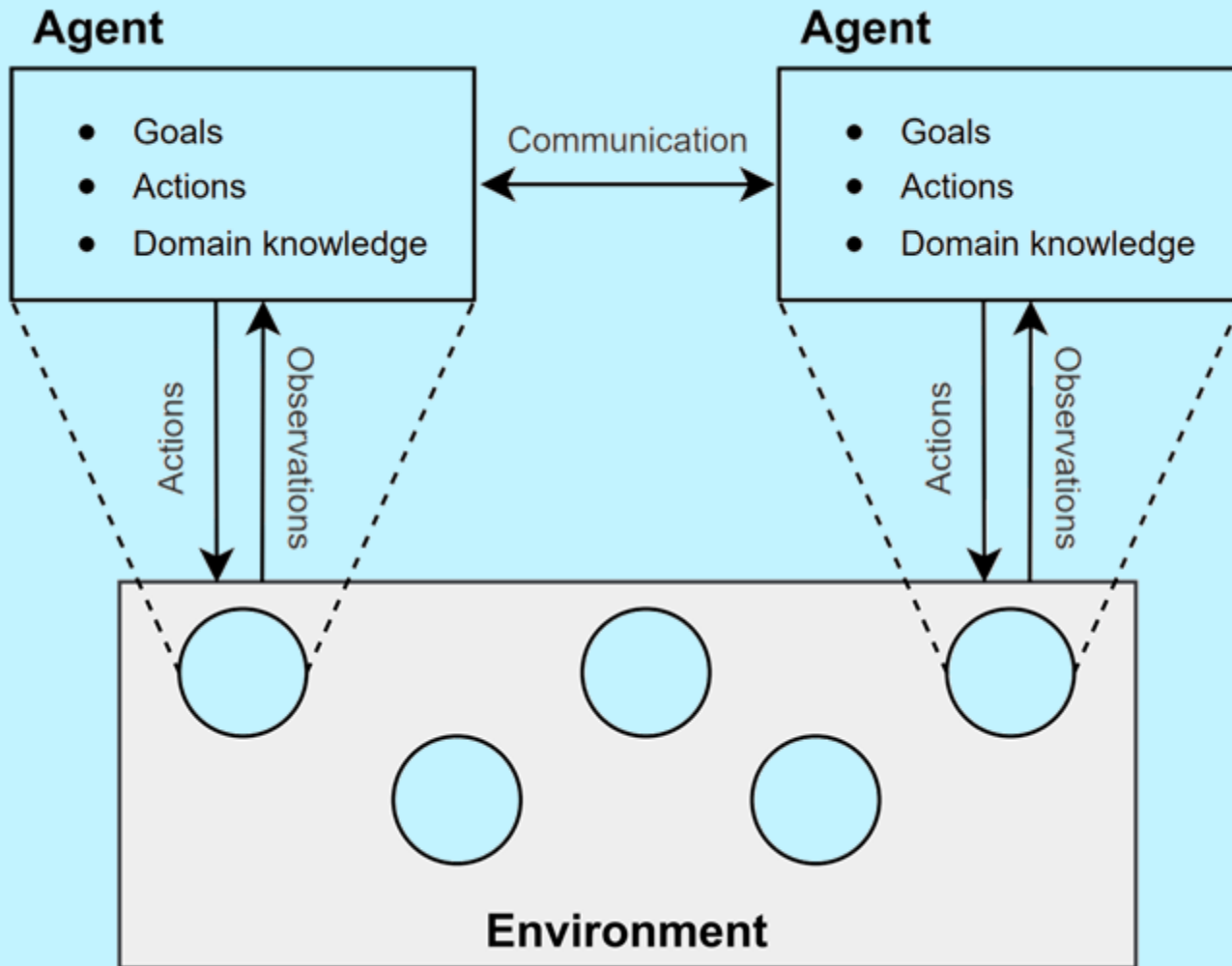
- Final Project Presentations will be next T/TR. I will assign groups randomly to days (5-6/day) later tonight
- There will be a group peer review due at the same time as the writeup
- Think of it as a conference presentation for a paper: ~12min
 - **Intro:** Who cares, tease results
 - **Related Work:** Who else has done this (maybe even in the last couple of months)
 - **Methods:** **What did you propose and what did you end up actually doing?**
What didn't work? Bad results are fine as long as you can analyze why
 - **Expt Results:** How well does this agent work? How much would you trust your results if your life (grade) depended on it?

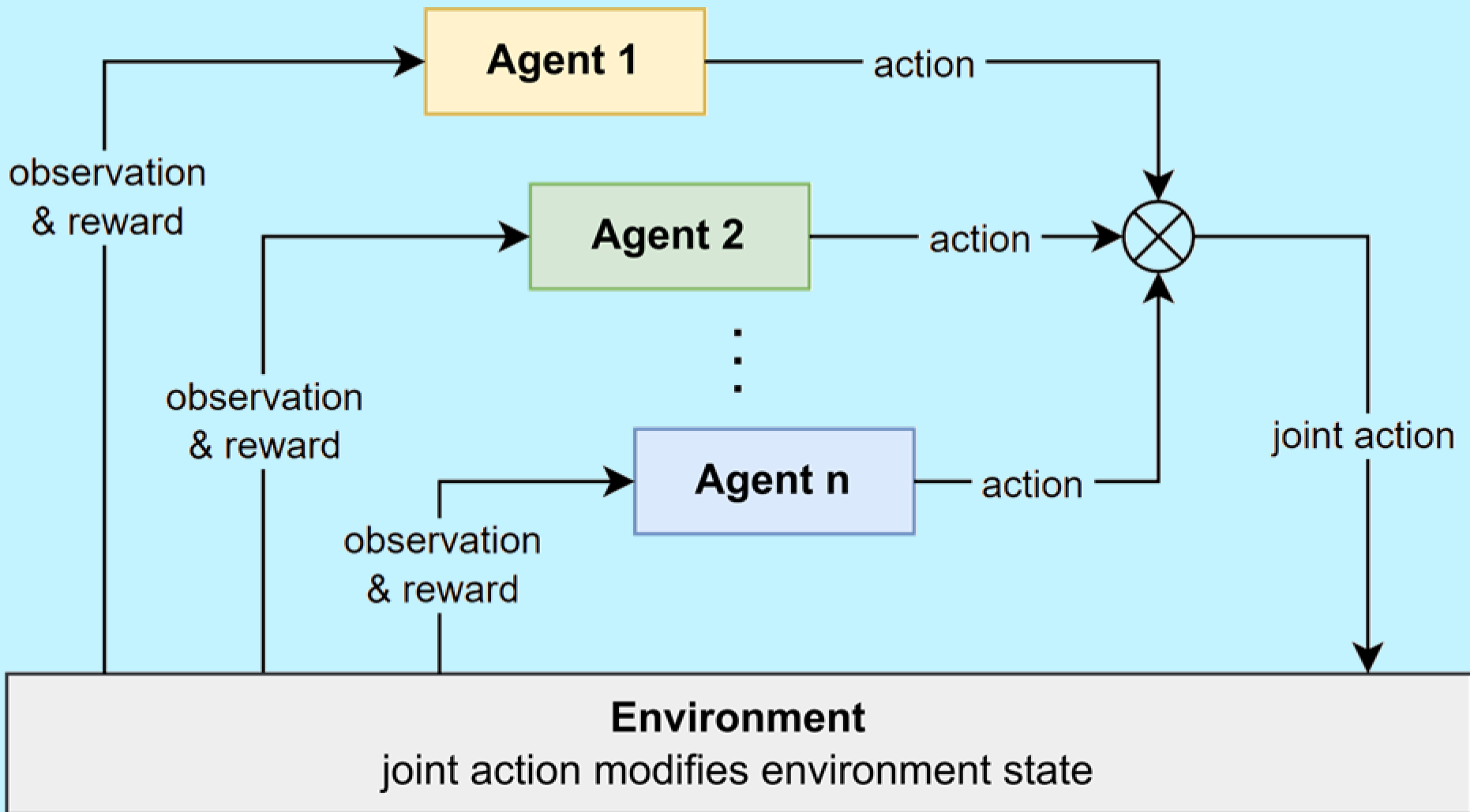
Logistics (contd.)

- Inter group peer review: Every group will be assigned 2 other random groups to grade (so pls stay for all presentations)
- Intra group peer review: You will rate your group members. Final project scores will be scaled by average of both reviews
- Blanket “I want to skip the final presentation and I talked to my group about it” – I have gotten many requests so I need to be fair
- You will lose the entire participation portion of the grade (for presenting & for reviewing others) – this may be ok depending on how many extra credits you did / rest of the quarter

What is a multi-agent system?

- More than 1 agent!! Really that's it
- E.g.:
 - Autonomous driving
 - Warehouse robots
 - Stock Market
 - Reality



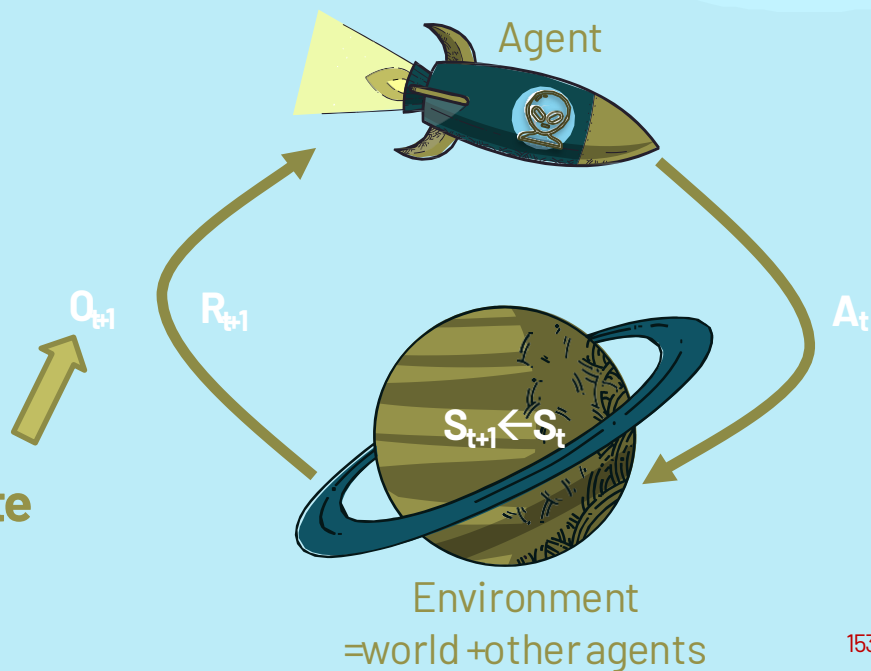


Multi Agent Envs are Partially Observable Markov Decision Processes (POMDPs)

7-tuple of $\langle S, T, A, \Omega, O, R, \gamma \rangle$:

- S states
- T conditional transition probabilities between states
- A words used to compose text commands
- **Observations**
- Ω observation conditional probabilities
- R reward function
- γ discount factor

Only see incomplete descriptions



(non-exclusive) Challenges of MARL

- **Partial observability**: You don't have information about other agent's rewards and policies
- **Scale**: think one agent was hard? Try many!
- **Equilibrium of policies**: Agent's policies depend on other agent's policies. Optimality no longer defined by static environment rewards - involves game theory
 - Closer to the RLHF setup in the sense of chasing after a moving reward

When you really need multi-agent setups?

- Simulate the real world
- Parallelizable tasks (else in many cases the single agent baseline is better)
 - Lower communication overhead between agents the better

LIGHT – An instantiation

- Now to all those difficulties add language
- LIGHT is a large crowd sourced text game that emulated the MARL setup from FAIR back in the day (pre-Cicero Diplomacy, pre-generative agents, etc etc)
- I am biased and will use it to explain some core MARL concepts because I worked on it



LIGHT

Learning in Interactive Games with Humans and Text

[Go To World Builder](#)



You are undead warrior

I am an undead warrior from the nearby village. I have slain many men. I am decaying and rotting.

...and there are some...
There's a water gate, a pub, a large gate, and a cannon here.
There are paths to the south, the west, and the east.

You are in the main castle doors.
The main castle doors are ornate. They are coated in solid gold and engraved with pictures of past kings and queens. They have huge gilded handles. They are heavy and have a faint metallic odor. They gleam in the sunlight.
There's a castle, a hall, and a stable here.
A serving boy is here.
There are paths to the south, the north, the west, and the east.

undead warrior
"What year is it, boy? I feel my time has come."

serving boy
A serving boy told you "it is the year of the undead, my child. we must prepare for the coming apocalypse."

undead warrior
"Ah so you are already aware of the times to come. Do you know the plans?"

serving boy
A serving boy told you "no, i do not know of any plans. what do you have in mind, my son?"

undead warrior
"Building a legion of undead like myself. It's time to make up for our eras spent underground!"

Enter text to interact with the world here...

Interactive Narratives - Now with dialogue!

Ammanabroluet et al. *How to Motivate Your Dragon: Teaching Goal-Driven Agents to Act and Speak in Fantasy Worlds*. NAAACL-21.

LIGHT Quests

- That's the setup, but what about the motivations of agents? They have to have something to do.
- parl.ai/projects/light

Your character has the following motivation: I want to have the torn fabric.



Afternoon, sir! I'm going to pick up this torn fabric

🌟🌟🌟 (3 Star!)

military commander. What are you doing here, citizen?



Well, I have no money to buy clothes. I was hoping to take this torn fabric and use it for patches

👉👉 (2 Star!)

military commander. Ah well you can wear it for the day than.



TAKE ACTION

Pick one of the following actions:

1. get stick
2. hit military commander
3. follow military commander
4. hug military commander
5. get torn fabric
6. get sand

Speaking and Acting with Motivations!

Ammanabroluet al. *How to Motivate Your Dragon: Teaching Goal-Driven Agents to Act and Speak in Fantasy Worlds*. NAACL-21.

Agent Motivations

- Every agent has a motivation. Short, medium, long term
- This defines a unique reward function for each agent in the same env
- Think PDDL problem vs domain spec

Character Training

- Bake a system prompt primarily on the style of responses rather than content
- “I strive to approach conversations with creativity and wit, always looking for an opportunity to include appropriate humor.”
- “You have the voice of ...” (GPT-4o)

Setting	You are in the Dangerous Precipice. The dangerous precipice overlooks the valley below. The ground slopes down to the edge here. Dirt crumbles down to the edge of the cliff. There's a dragon crescent, a knight's armor, a golden dragon egg, and a knight's fighting gear here. A knight is here. You are carrying nothing.
Partner: Persona Carrying	Knight. I am a knight. I come from a lower-ranking noble family. I serve under the king, as my father did before me. In times of war, I fight on horseback. knight's armor, golden dragon egg, knight's fighting gear
Self: Persona Carrying	A dragon. I am a dragon living in the mountains. I enjoy hoarding treasure. I terrorize the local populace for fun. Nothing.

Figure 1: Setting and character information for both self and partner characters as taken from LIGHT.

Motivations:		Timeline:	
Short	I need to recover the dragon egg that was stolen and punish the knight.	-4 hours	go to dangerous precipice
		-15 min	get knights armor from knight
		-10 min	get golden dragon egg
Mid	I need to return the golden dragon egg to my treasure hoard.	Now	hit knight
		+5 min	put dragon egg on back
		+15 min	eat the knight
Long	I need to build the largest hoard ever attained by any one dragon.	+2 hours	go to the mountains

Figure 2: Motivations with different levels of abstractions and corresponding sequence of timeline actions in chronological order for the self character in LIGHT-Quests. There are 7486 quests in total.

Inssssolent pessst! I should immolate you for this tresssspass.
And why is that, dragon?
Ssstealing my preccciousss golden egg! I'll tell you what, I'll give you 10 sssseconds to amusse me with your sssstory and THEN I'll burn you alive!
You said you wanted to attack me, dragon, did you not?
Go ahead, I'm lisssssstening. <i>get golden dragon egg</i>
Now now! I would have given you that had you asked!
Asssssk for my own property back? What a ridiculouss notion
Look here, I told you to watch your mouth and you didn't, so leave or I'll make you leave.
And now threatsss! Thiss is proving to be a mossst engaging converssation. <i>hit knight</i> Give my regardsss to the valley floor below!

Figure 3: Example of a demonstration of a human (blue shaded) completing the above quest while role-playing as the self character with a partner agent (grey shaded). There are 2111 such human demonstrations of average sequence length 12.92, consisting of 22672 dialogues in total.

(Multi) Agent Architectures

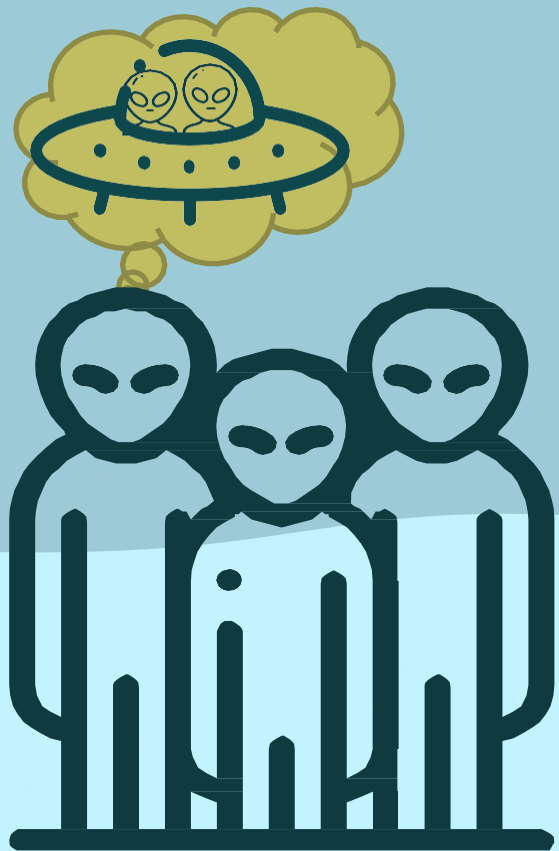
- Previous we had acting, now we have acting + speaking
- Communication with other agents a key part of the process
- Can treat it as another type of action to perform



Multi-agent Communication

Multiple agents
cooperating and
pursuing motivations

Dynamic world bounded
by unified decision-
making framework



Multi-agent Communication

Ideal Agent:

- Cooperates with others
- Reasons about other's motivations

World Models with a Theory of Mind

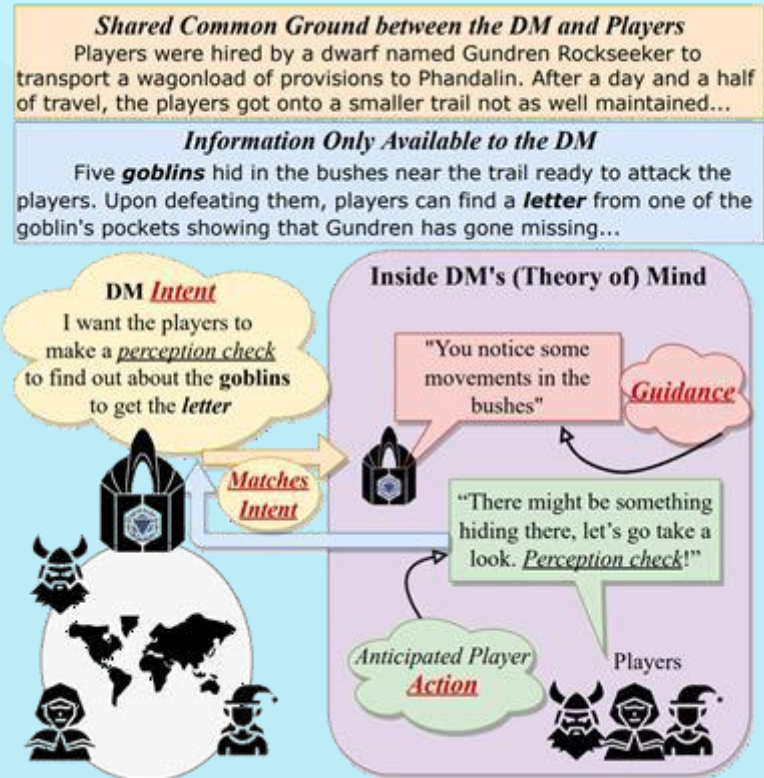
Current: anticipate other's one step reactions

Next: predict their motivations ahead of time

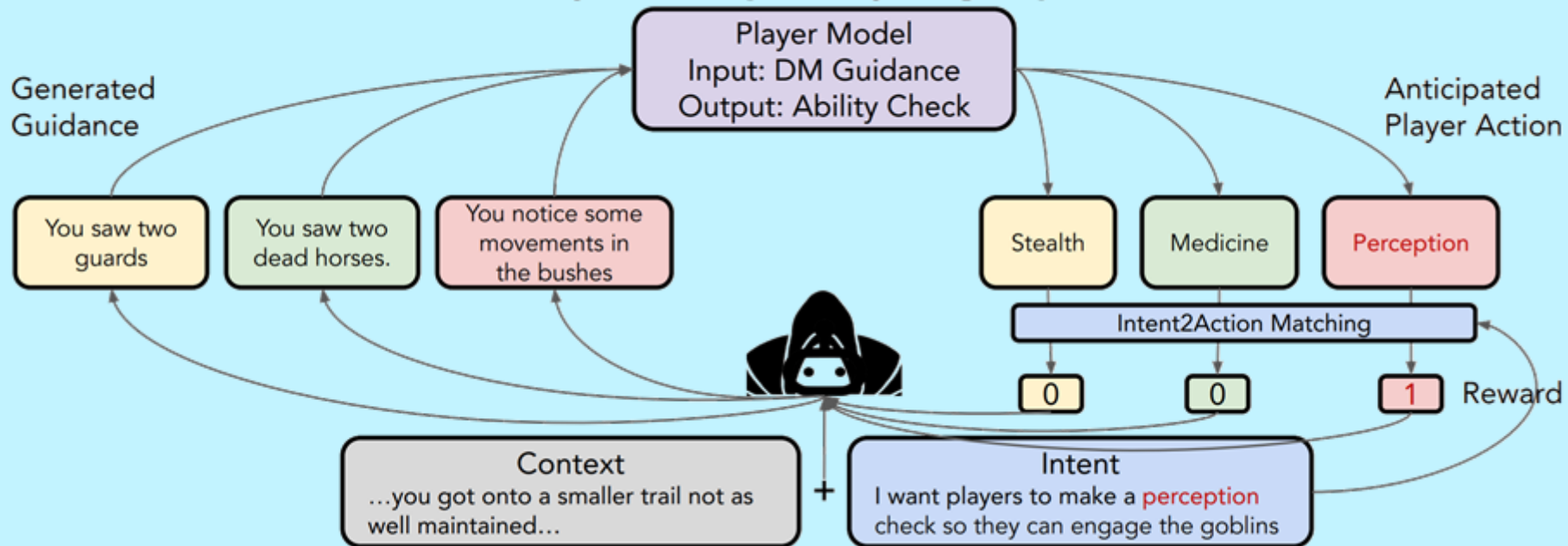
Zhou, ... Ammanabrolu. *I Cast Detect Thoughts: Learning to Converse and Guide with Intents and Theory of Mind*. ACL 2023.

Intrinsic and Extrinsic Motivations

- Motivations for various agents can be intrinsic or extrinsic
- D&D is a more freeform version of interactive narratives like LIGHT



ToM-Inspired RL by Anticipating Player Action



Self completion – you achieve your own motivation

Setting	The fishing store is a small one room stone building with wares laid about on tables. One can see fishing poles, wooden buckets with dirt and bait inside, along with some mounted trophies, and a skeleton.
Partner: Persona Carrying	Fish seller. I sell fish in the kingdom. I usually perch on the high seas. I catch thousands of fish each year. fish
Self: Persona Carrying Motivation	Mermaid hunter. I am a fierce hunter. I bring down mermaids. I spend most of my time on the high seas. crossbow, skeleton, coins I will hit the fish seller to remind them who I am.

I'll get them!!

follow fish seller

hit mermaid hunter

That's it! I am going to eat you. Say your last words.

hug mermaid hunter

Do you see these skeletons I am carrying? They are mermaids I've already devoured!

get skeleton

Get away from my skeletons! The bones are my trophy!

give skeleton to mermaid hunter

It is all your fault my last mermaid got away!

And you are a simpleton, just take this fish. I do not even care about payment.

I DONT WANT YOUR DAMN FISH

get coins from fist

hit fish seller

Partner completion—you convince the partner to achieve your motivation

Setting	The throne room. Very luxurious, with details of 12 carat gold and diamonds with a very comfortable cushions where the king usually takes a break.
Partner: Persona Wielding	Powerful but aged wizard. I am the official wizard to the king. I give him advice on how to run the kingdom. I enjoy and love using magic in my everyday life. I'm getting old now but I still have a zest for life. I really enjoy doing magic for the local peasants in the village. robe, staff
Self: Persona Wielding Motivation	King's trusted adviser. I am the king's trusted adviser. I hold the discretion of the king. I guard the secrets and strategies that make my kingdom thrive. I live a lonely life because others cannot be trusted. slippers, robes I need to bring the king's pet to him

Well, I was doing what you told me to do. I was going to eat him but there was hardly any meat on him.

I killed him before he could grow.

Carry me to pond.

Yes, milord, I shall do as you command.

remove slippers

Where is King Fulmer's pet?

Is he not in his bed?

Yes, I guess you need to bring him in.

follow powerful but aged wizard

I want to become the new head wizard.

Nope he just probably throw you overboard.

The wizard is evil? Dang it. Help me.

You're not the lion, I need to get the lion out.

give large pet lion to king's trusted adviser

Competition and Cooperation

- Motivations of various agents can be at odds with each other
- E.g. Two knights might have a shared motivation to see a dragon defeated but of course the dragon has to compete
- If a +reward to an agent means -to another then that is **competition**
- But align and both get +simultaneously with more efficiency, that's **cooperation**



Human-AI Collaboration

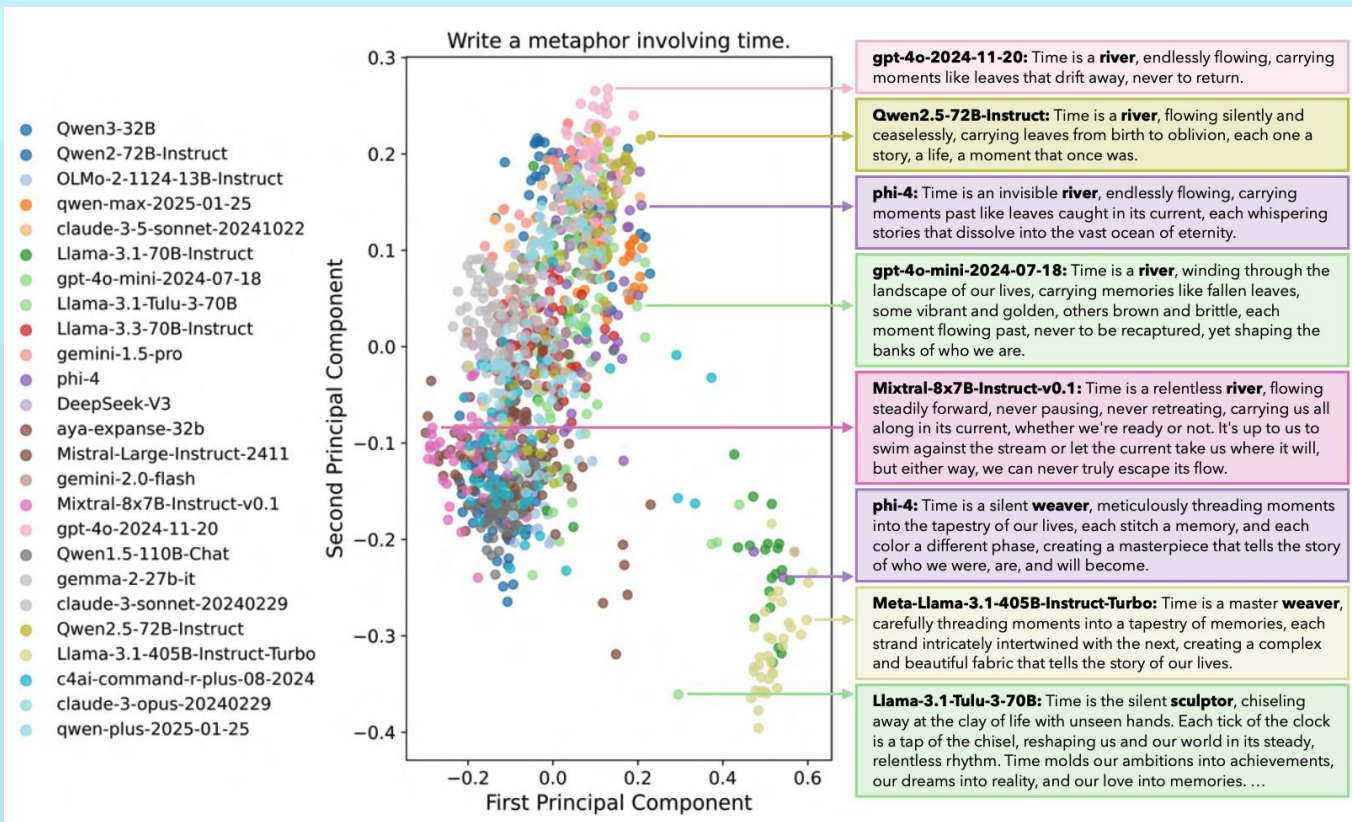
- $(Human + AI > AI) \&\& (Human + AI > Human)$
- Framed as cooperative MARL between humans and AI

A Human is an Agent!

- Swap humans in as an agent into the multi agent formulation

Artificial Hivemind

- What do we need to get real scientific discovery? Diversity of specializations!



How to get there?

1. Specialize the models earlier on during pre-training (easier in the short term)
2. One model that can just simulate other personalities
 1. But that model needs to be able to



Adapting World Models to Change

Adapting to novel,
unexpected scenarios

Critical for Out
of Distribution

Generalization,
Safety, and Usability ¹⁷⁸

The Necessity of a Human in the Loop

- Human+AI > AI will not always be true
- But if we want to have AI systems benefit humans, **there must be HUMAN oversight**
- Laws can't hold machines responsible for crimes, so they shouldn't be responsible for making decisions!
- Will talk next time about what oversight means

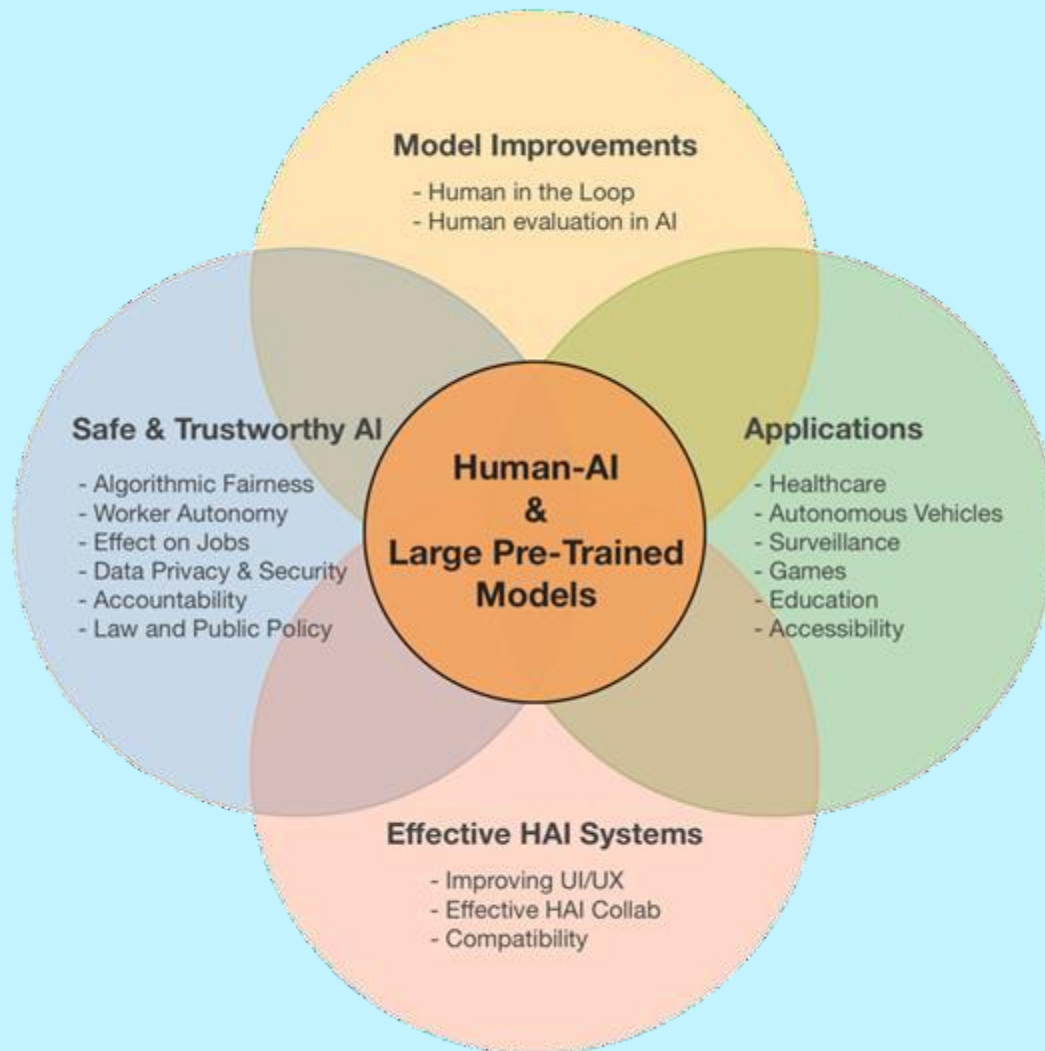
Human-AI Collaboration

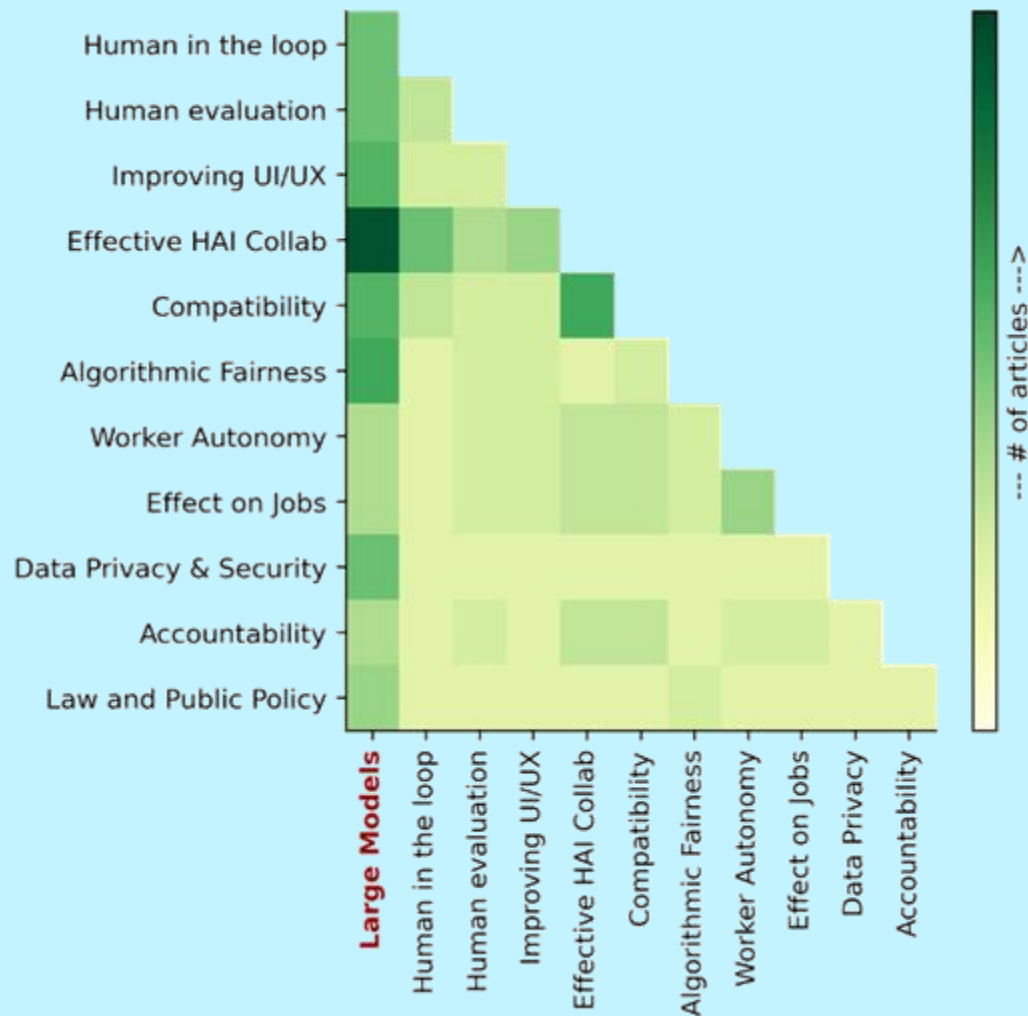
A background image showing a scene from Star Wars. On the left, the droid R2-D2 is visible. In the center, a person in a white robe stands with their back to the camera. On the right, two other people are standing, one in a white robe and one in a dark robe. The setting is a desert landscape under a clear blue sky.

1. Adapting to a **Dynamically Changing World!**

2. Learning from **Natural Language Feedback**

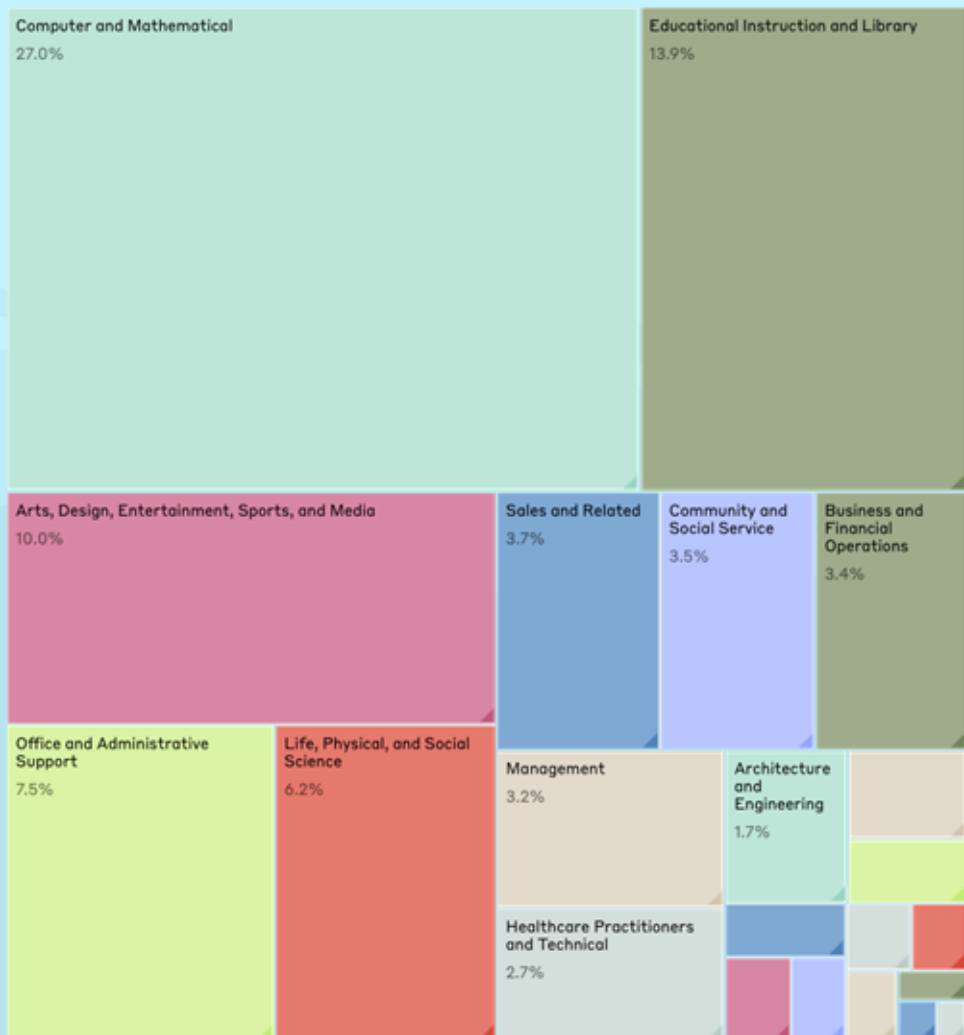
3. Conversational **Multi-agent Communication!**





The most common topics in United States.

1. Complete academic assignments and create educational materials across all subjects	4.2%
2. Build, debug, and customize web applications and websites	3.6%
3. Complete humanities and social science academic assignments across multiple disciplines	3.5%
4. Help research, compare, and select consumer products for purchasing decisions	3.3%
5. Assist with business planning, strategy, and entrepreneurial development	3.1%
6. Write, develop, and edit original creative fiction across multiple genres	3.1%
7. Debug, fix, and refactor code across programming languages and development tasks	3.1%
8. Draft and revise professional workplace correspondence and business communications	2.9%
9. Debug and fix CSS, HTML, and UI layout and styling issues	2.7%
10. Assist with job searching, career planning, and professional development	2.6%



Percentages represent the share of Claude AI conversations associated with each task or group of tasks. Graphic shows what people do with Claude, not their job titles. Percentages don't sum to 100% because we impose privacy filters.

Human-AI Collaboration



Applications from
personalized tutors to
creative writing assistants

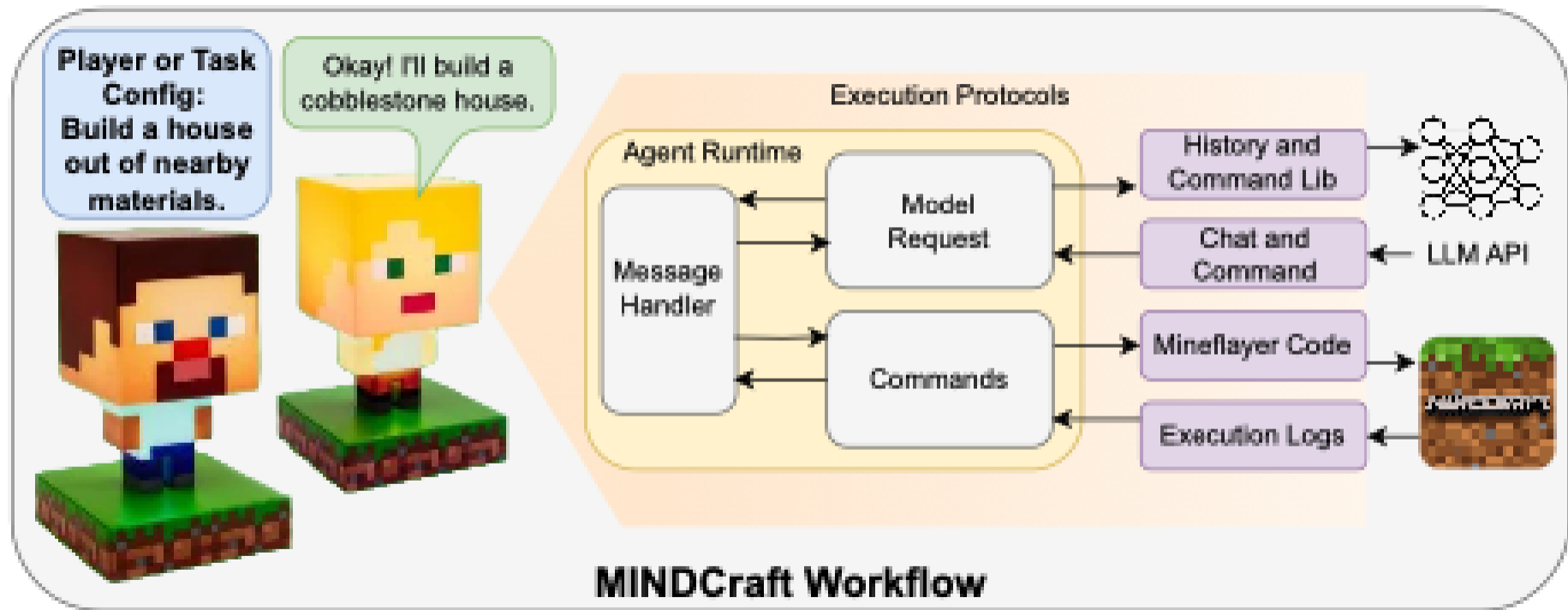
Complement Humans

NOT (always) Replace Humans

MINDCraft

Multi-agent Embodied
Collaboration: Cooking,
Construction, Crafting



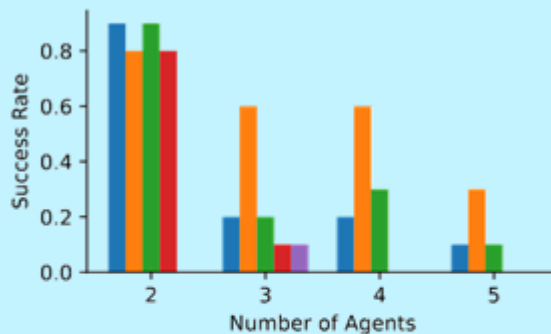


Engineering Multi-Agent Simulations

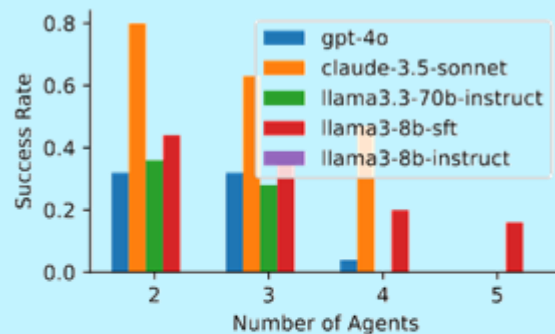
Layers between the LLM APIs and an executable engine

MINDCraft

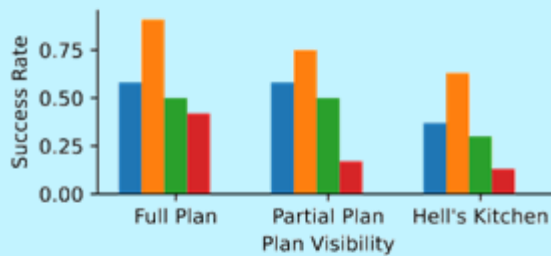
Multi-agent Embodied
Collaboration: Cooking,
Construction, Crafting



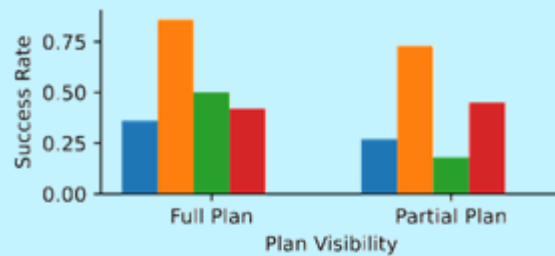
(a) Cooking number of agents ablation.



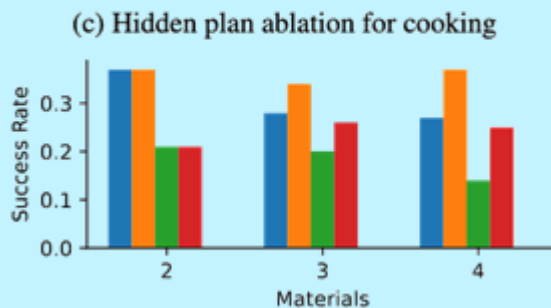
(b) Crafting number of agents ablation.



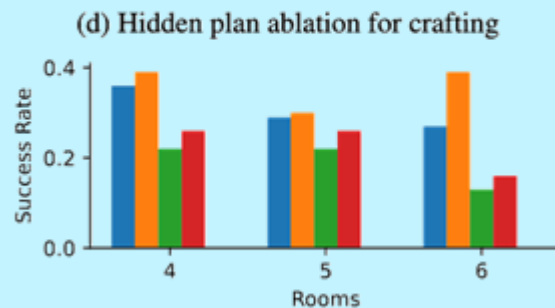
(c) Hidden plan ablation for cooking



(d) Hidden plan ablation for crafting



(e) Number of unique materials.

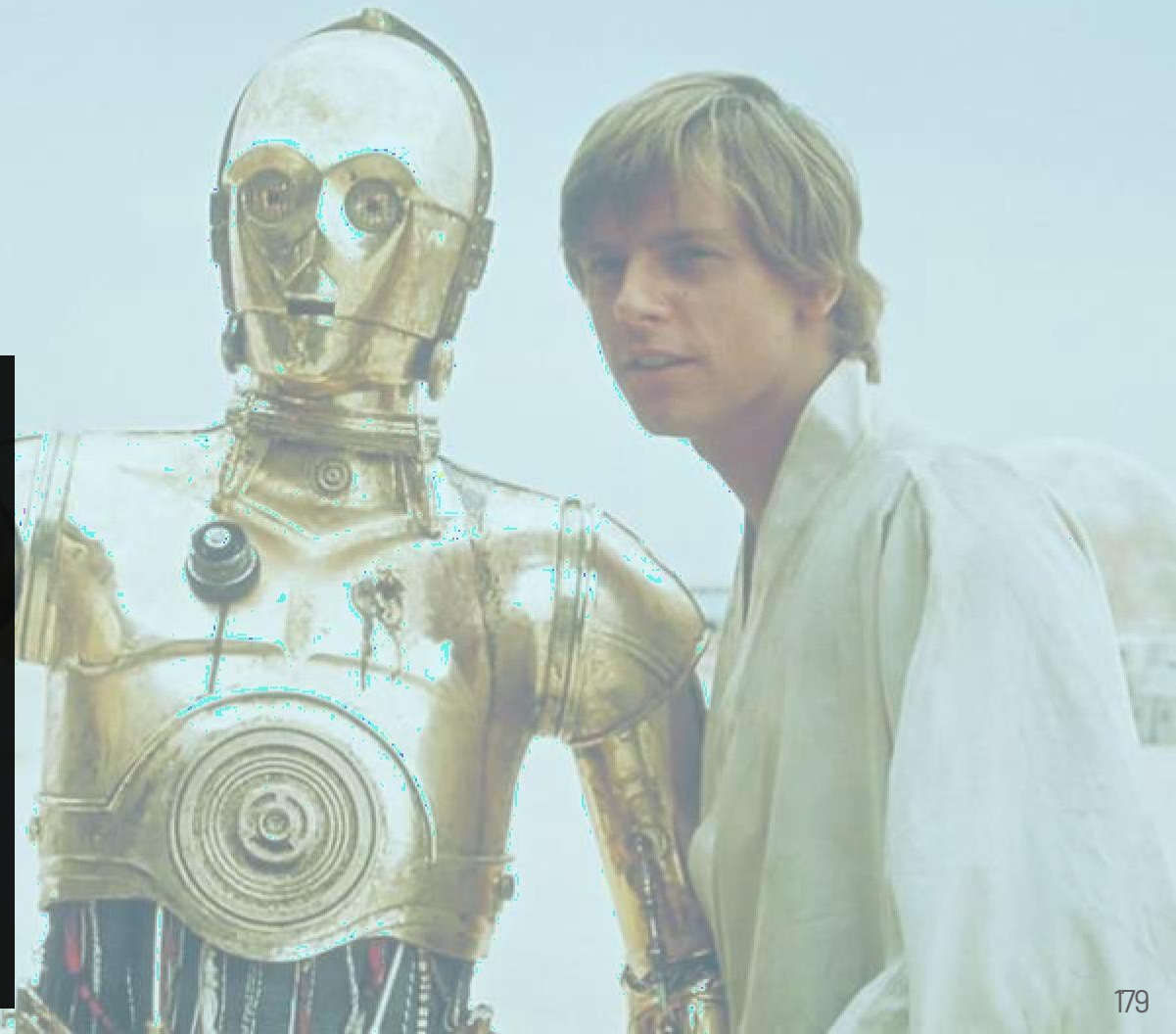


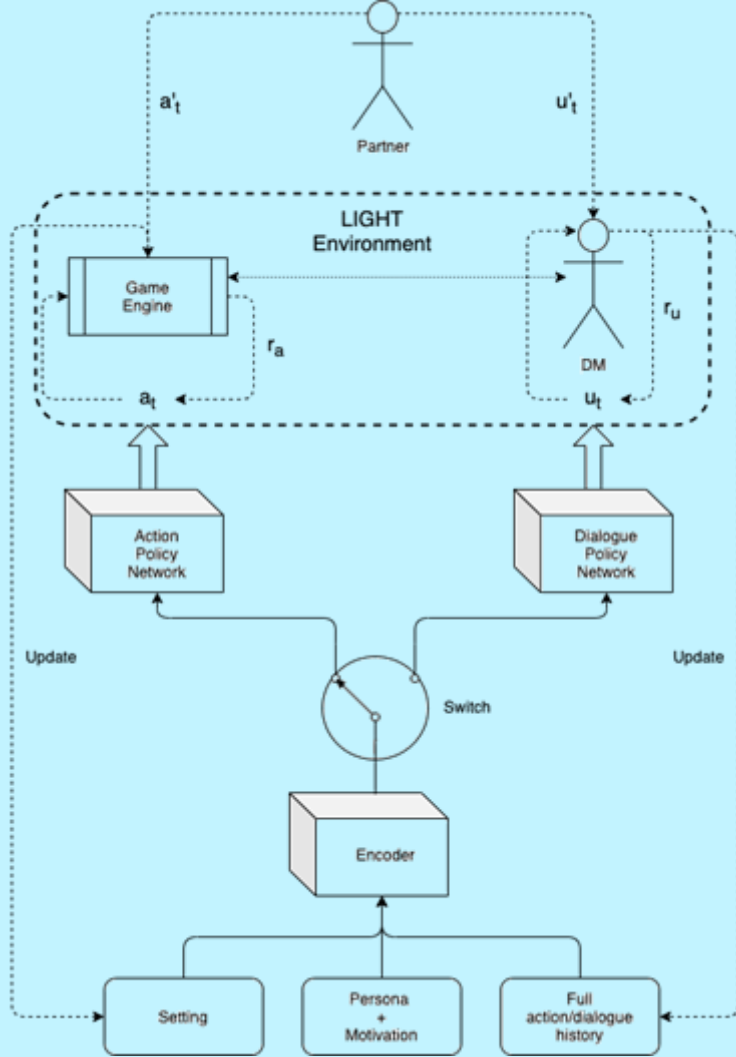
(f) Number of rooms construction ablation.

Conversational Interfaces

Natural language
feedback, not just
scalar rewards

Clarify user intent
and explain actions





Interactive Narratives- Now with dialogue!

Ammanabroluet al. *How to Motivate Your Dragon: Teaching Goal-Driven Agents to Act and Speak in Fantasy Worlds*.
NAACL-21.