A garden of Oddities: 0-Shot, IMOL, XAg

GEI-SID
Víctor Giménez

Guide to understand the class:

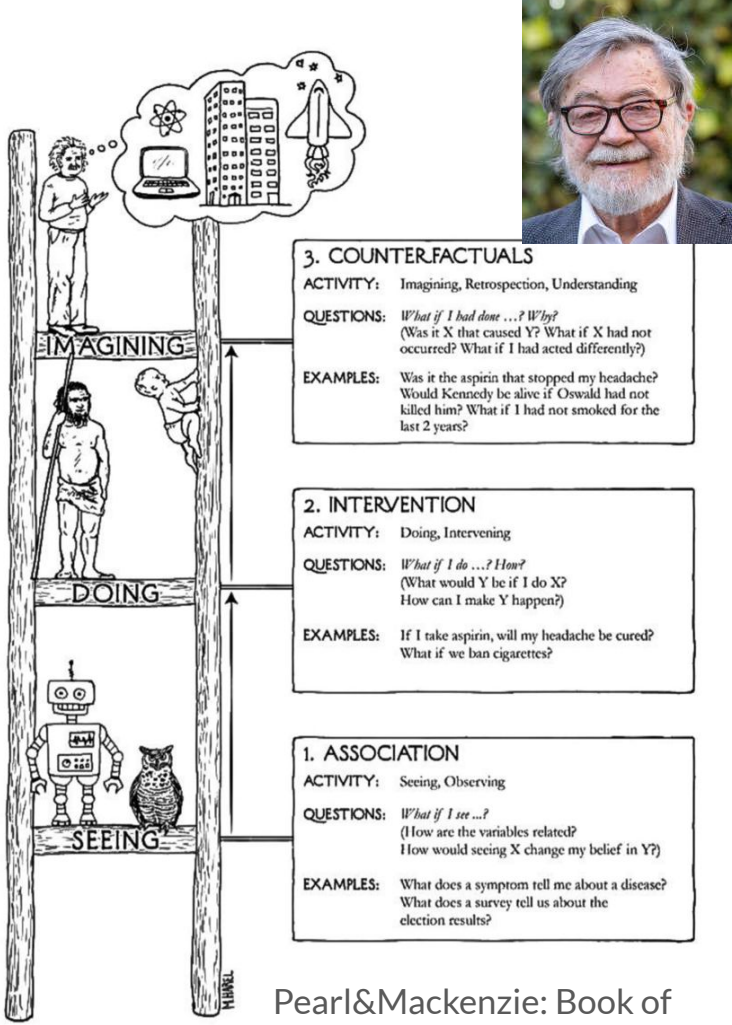
- Don’t focus on particular techniques
- Get the ‘gist’ of intuitions and what worked and didn’t
- Be amazed at how weird people can think
- Understand the diversity and disconnect between many disciplines
- Remember inspiration comes from any source
- Keep the big picture of what AI can and ***should*** do

AGI attempts & Oddities

Why is ML not AGI?

- World of Vampires
- Cognitive loads & natural selection
- Social selection & myths

- RL/Agent's advantage: 1 rung up



Pearl&Mackenzie: Book of Why

Why is RL not AGI?

Single objective

Discrete State-Action paradigm

Fixed, unchangeable utility

Adaptable modes?

How does nature fix this problem?

Why are living beings not single-utility?

What’s the driving force utility behind cognition?

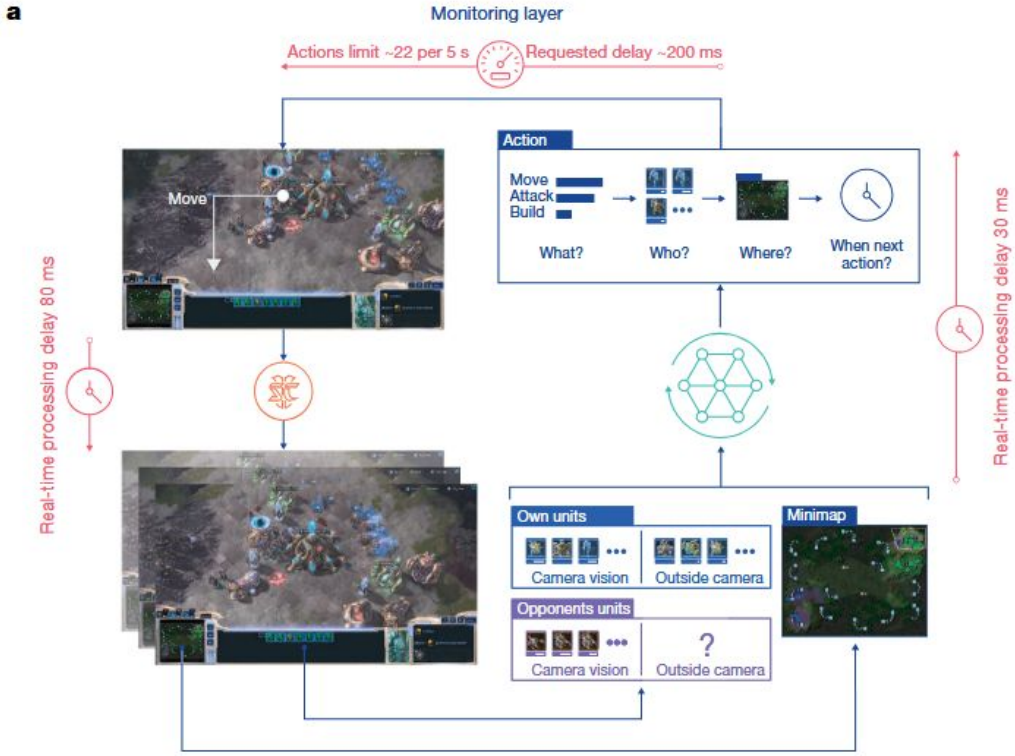
AlphaStar

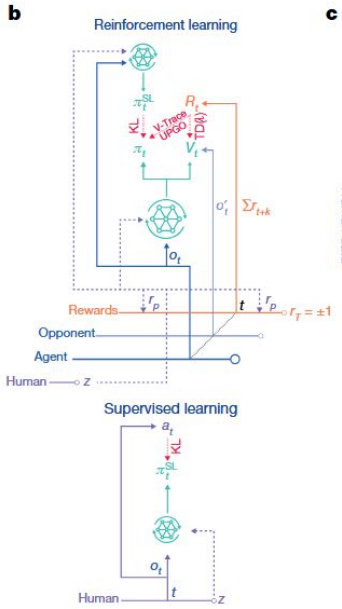
Grandmaster Level in Starcraft II
(Deepmind; Vinyals et al.)

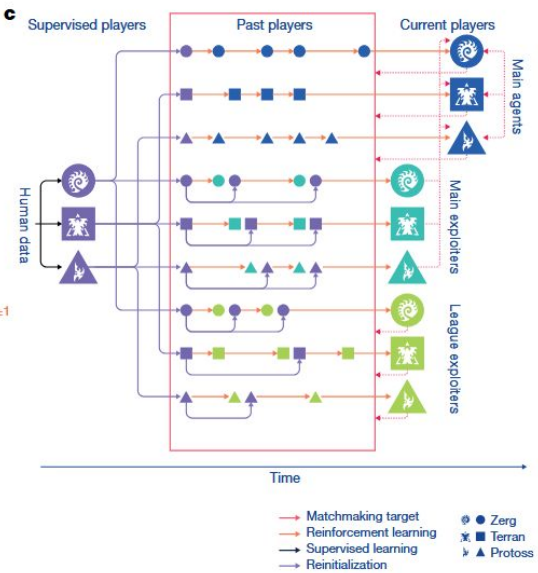
Motivation: Foundational models' success

Oriol Vinyals, UPC
Author of AlphaStar

Interesting Ideas







- Start from Supervision
- Per-subtask training
- Multi-action prediction (with time-staggering + Auto-regression)
- Speciation & exploiters

Extracurricular: Starcraft II presented by Vinyals

AlphaStar Conclusion

Very cool, many interesting paradigm-breaking ideas

Game-theoretic side solved by the ‘league’ and forcing exploitations

- Think of game-theory in behavioural evolution, nature-inspired

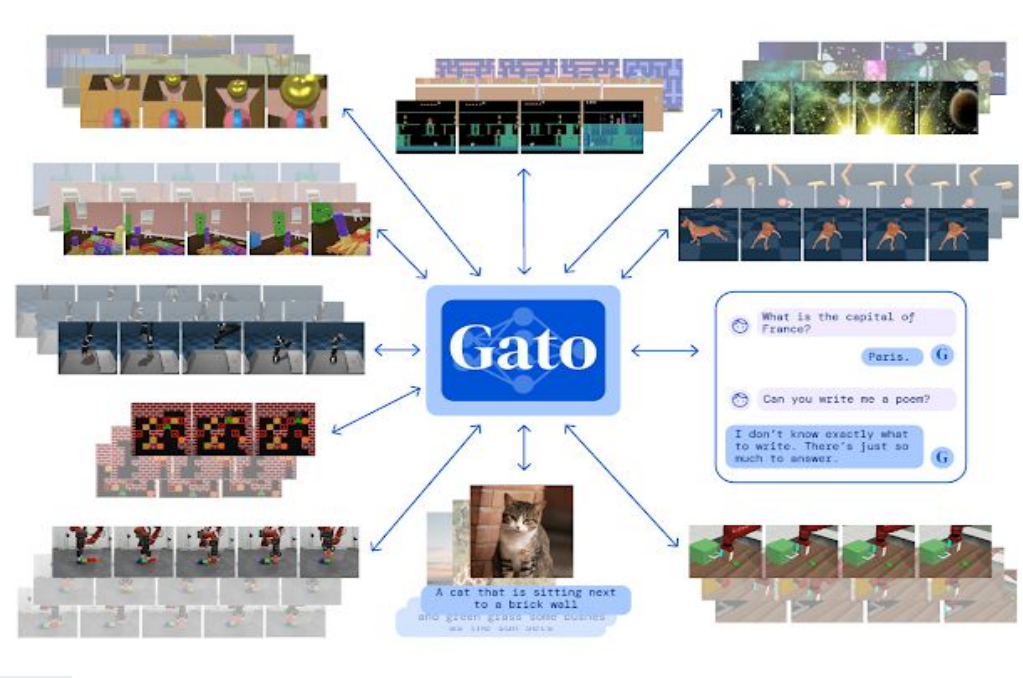
Not even experts train RL from scratch: starting from supervision is a good idea to develop ‘skills’ to later learn ‘strategy’

- Architecture matters more than weights

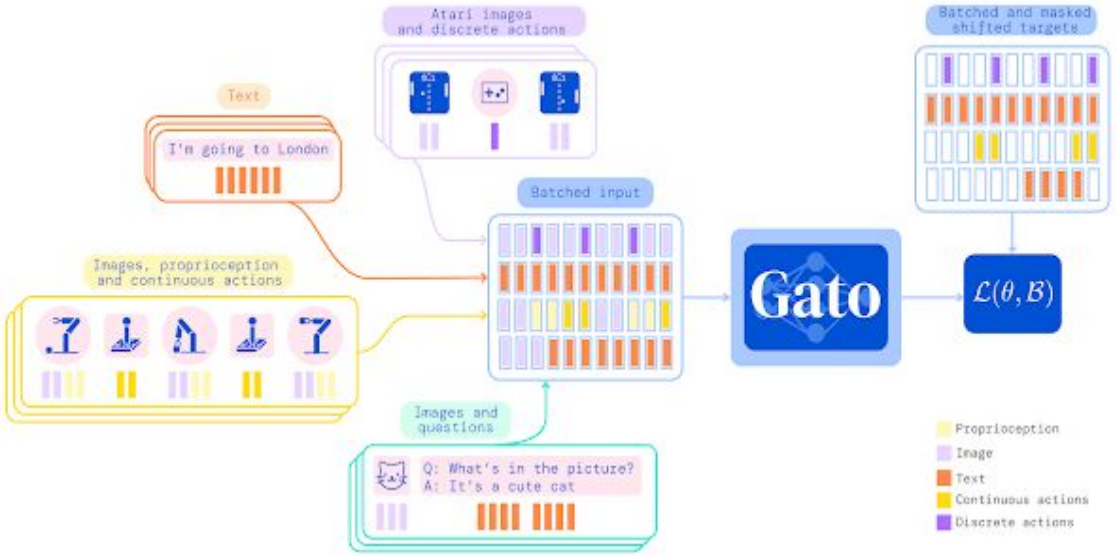
Attempt: GATO

A Generalist Agent
(Deepmind; Reed et al.)

Motivation: Foundational models' success



Attempt #1: GATO



Observation:

- DL can train anything
- Things trained in DL have some ‘knowledge’ of the world
- Reusing a model > training a model from 0 (Transfer Learning)

Train a model in as many ‘domains’ as possible

GATO Conclusion

Model was too big, too hard to train

Advantage was only on computational efficiency (if you needed to learn all those tasks). It is ‘easier’ to learn all together, but much harder than learning a few standalone.

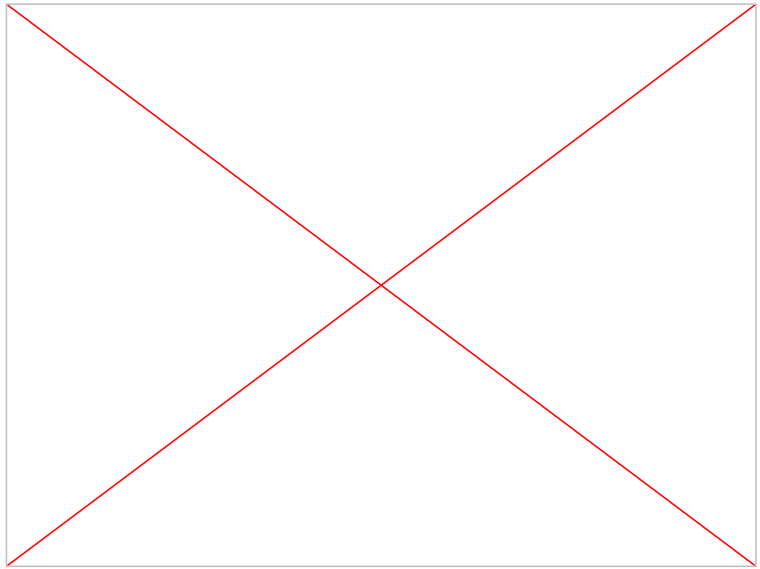
Even authors were skeptic of the claims that it was AGI

Project is more or less abandoned

Attempt: MineDOJO

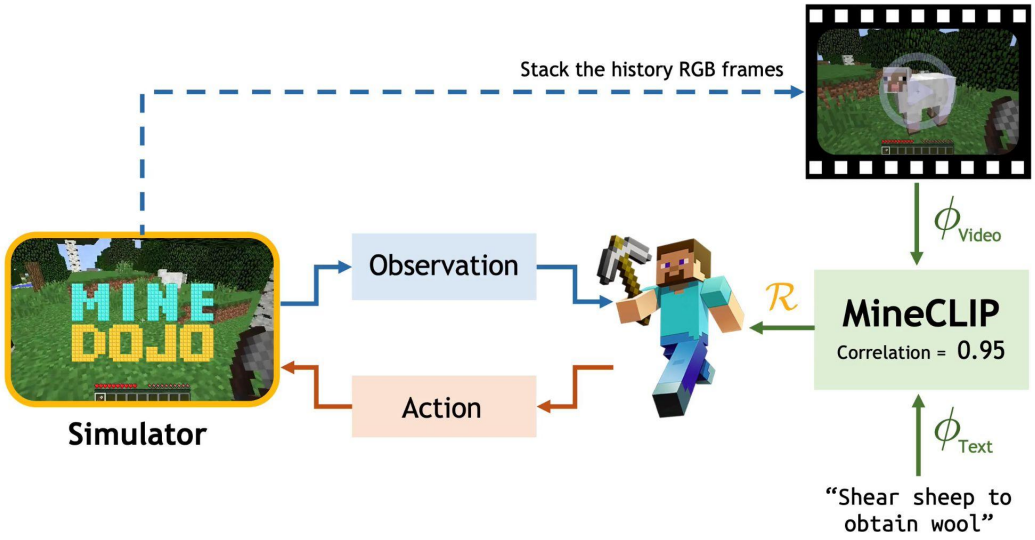
MineDOJO
(Nvidia, Linxi et al.)

Motivation: if getting a dense reward was easy, learning will be too



Linxi is better known as Jimmy Fan

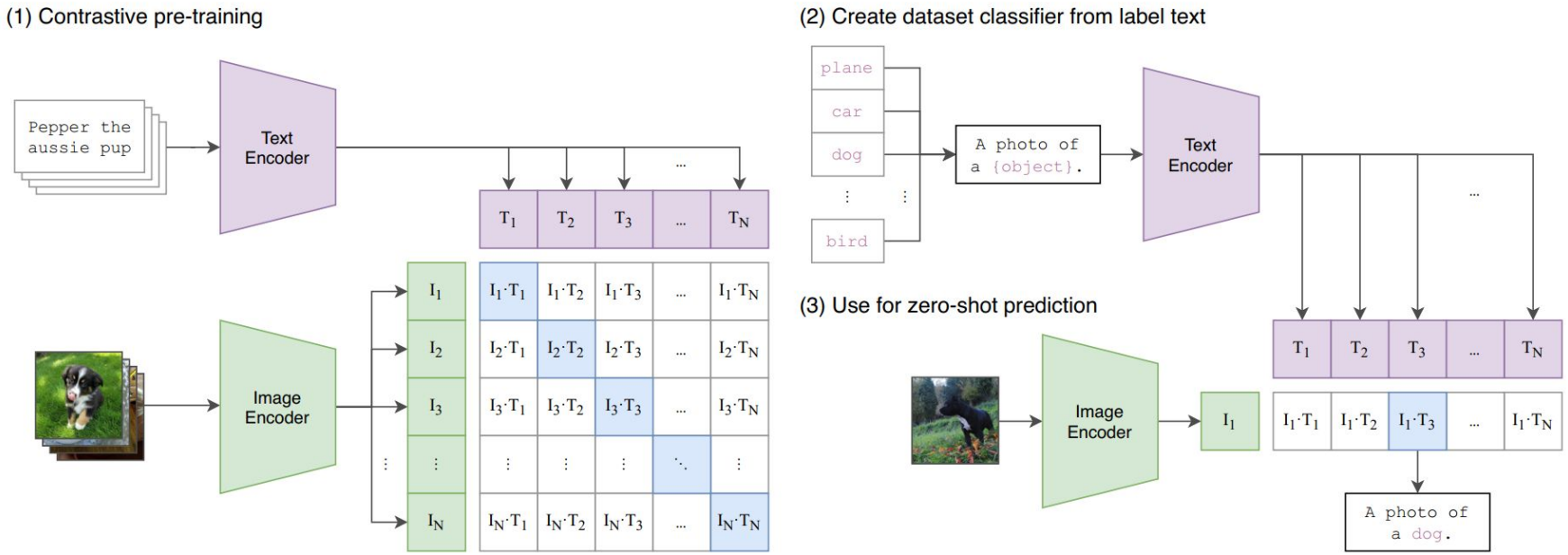
Internal Working



- Captioning video idea: "If a video of an agent is captioned doing the task, the agent is doing good"
- Goal-Video relatedness score -> Reward

What's a CLIP? Why are youtubers important?

Jimmy Fan: Play minecraft at work, you're generating curated AI data!



MineDojo Conclusion

Quite cool, got people very excited. Training a task was never easier

A bit spastic even when trained by Nvidia burning compute

Does not do AGI, each task is a new policy

Requires a full CLIP of your environment: does not work for most

Quickly ‘beaten’ by the paper-down-the-line: Voyager (by Jimmy Fan)

Zero-shot & IMOL

Can an agent beat a task on its 1st try?

Can we build agents that are motivated to ‘learn’ and not ‘learn to do x’?

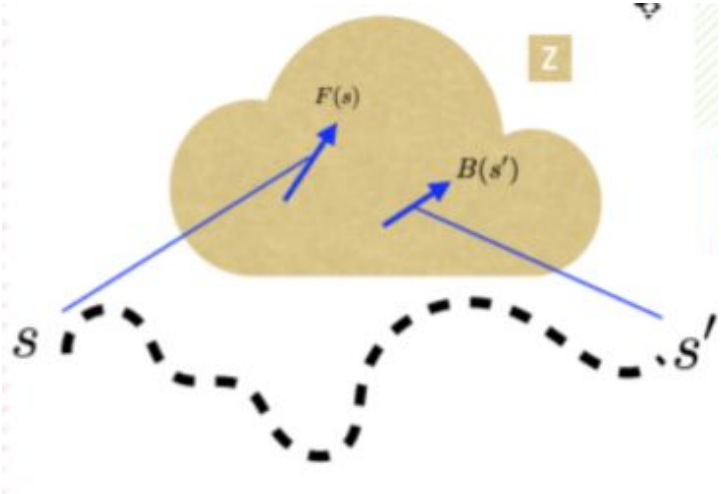
Few/Zero-Shot; Offline RL

- Assume you get a robot. You want it to learn to do stuff, but you don't know what yet. When you do, you want it to learn 'fast'.
- You let the robot 'play', do whatever, no task. You give it many many hours
- You get an idea of a task, and want to **train the robot without it interacting with the environment: that's too costly**
- You get the robot to try to solve the task in the environment, and is must succeed in few shots, or immediately even.

Forward-Backward

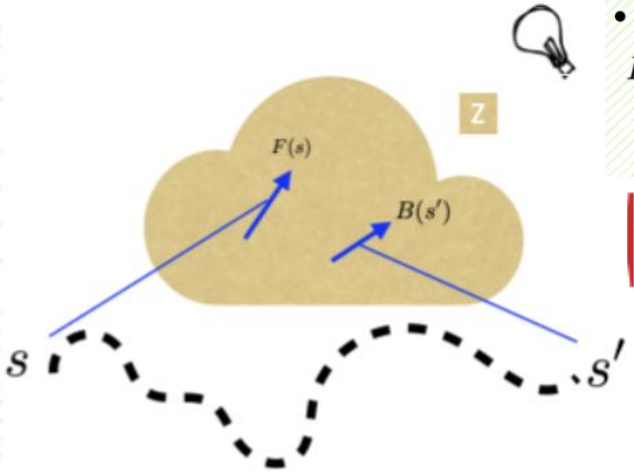
[Does 0-shot RL exist?]
(Touati)

Motivation: I can learn where and how to get to states before knowing to which I have to go



How it works

- **Idea**: train **two representations *F* and *B***. *F*(*s*) represents the future of state *s*, and *B*(*s*) represents the past of *s*.
- **Training criterion**: if it is easy to reach *s'* from *s* (in many steps), then $F(s)^\top B(s')$ is large.
- **Thm**: when this training loss is 0, *F* and *B* encode all optimal policies for all tasks.



- **Unsupervised training criterion**: for all s, a, s', z,

$$F(s,a,z)^\top B(s') \approx \sum_{t=0}^{\infty} \gamma^t Pr(s_t = s' \mid s, a, \pi_z)$$

$$\pi_z(s) = \arg\max_a F(s,a,z)^\top z$$

- Once rewards are accessible, compute $z_R = \mathbb{E}[r(s)B(s)]$
- For instance, $z_R = B(s)$ if the reward is located at s

$$\pi_{z_R}(s) = \arg\max_a F(s,a,z_R)^\top z_R$$

FB Conclusions

Offline RL and few-shot RL are small communities

RL was worked on for decades before starting serious usage (manufacturing, industry, etc.)

Offline RL, despite usefulness, does not yet have ‘a good but low-danger’ economic area. Autonomous Driving is too high-stakes

Robotics may change that

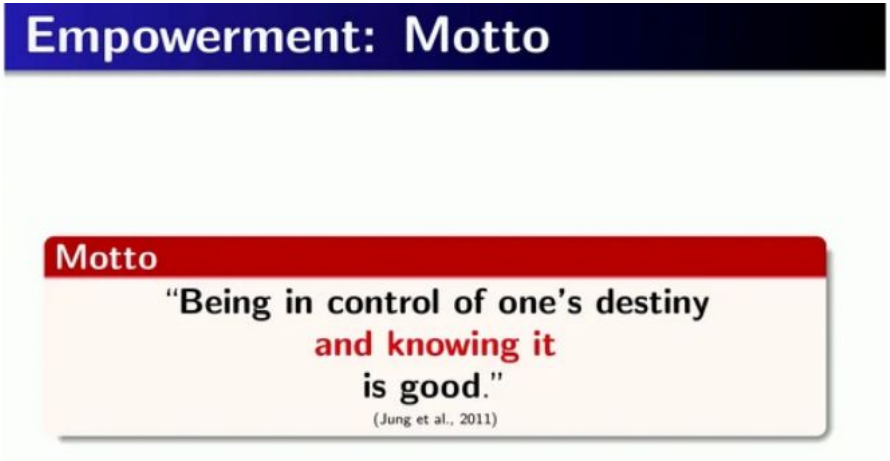
Note: paradigm change, representing **declaratively** instead of **imperatively**

Empowerment

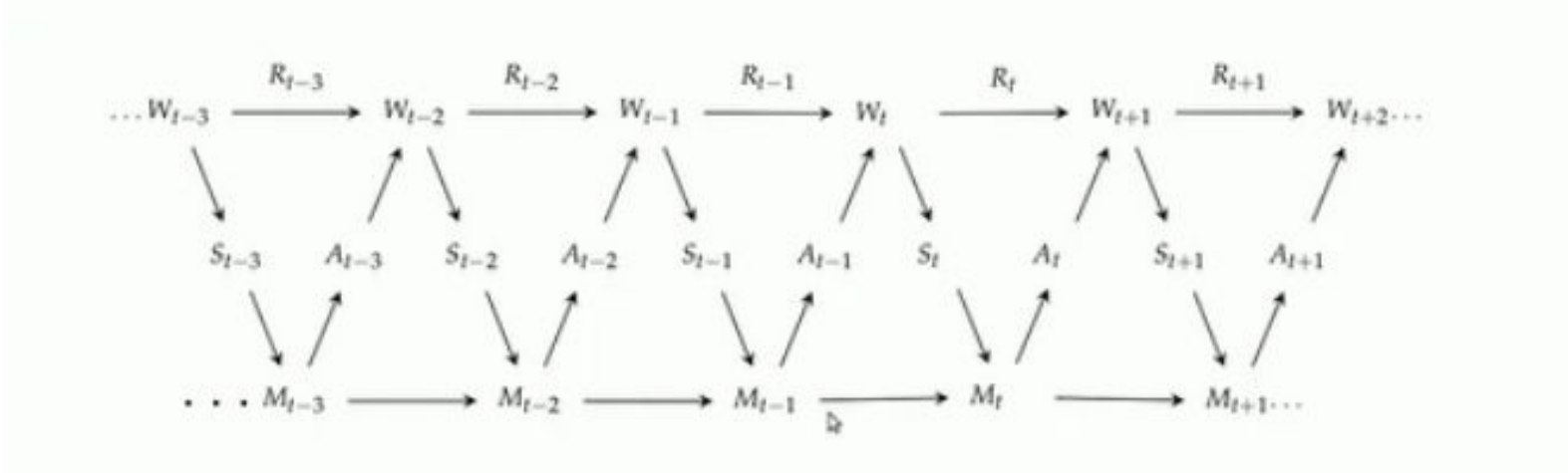
Empowerment (Salge, **Polani)**

Motivation: What's the source of curiosity and survival in humans?

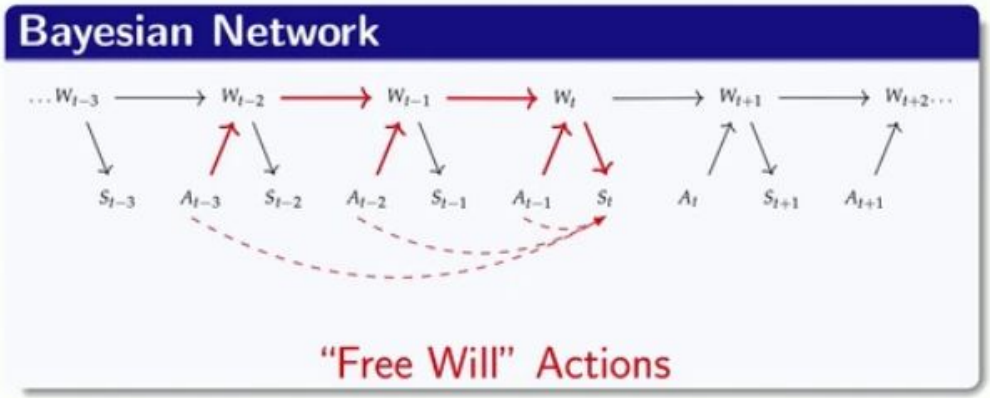
NeurIPS Presentation



Causal, Markovian World and Agent



Empowerment Idea: I could change the world



- 'k'-steps Empowerment: I check my future actions and see how much they determine the kth future state. How many futures?
  - I **could** change the world.
- Pick actions so that I get to states that maximise empowerment.
  - It only works if I can **tell apart** the futures (I see them)

Biological implications

- Money, Politics, Mobility
- Being alive, helplessness, ability to change future
  - Empowerment is ‘Strategy’, not ‘Tactics’. You can only play tactically if you know what will happen. Politicians talk Strategy
- Maximise options, even if you’ll take only one.
- “A chicken is just an egg’s way of making another egg” ~S. Butler
  - Being an animated animal is what gives an egg the ‘option’ to create an egg; but from a reductionist view: it’s just a means.
  - The Egg increases empowerment via becoming a chicken

Empowerment Conclusions

There’s a community working on it

Still in low TRL (tech readiness level): a mathematician-in-a-basement’s model without good enough efficiency or use-case to apply in real world

Agentic:
AI in worlds, powered by LLMs

Idea: The revolution will not be supervised

- Idea of foundational models works^{tm}, but not directly training.
- Apparent intelligence may come from knowing common sense
- LLMs appear^{tm} to have common sense
- Building an architecture for **behaviour** around an LLM may multiply the LLM's capabilities

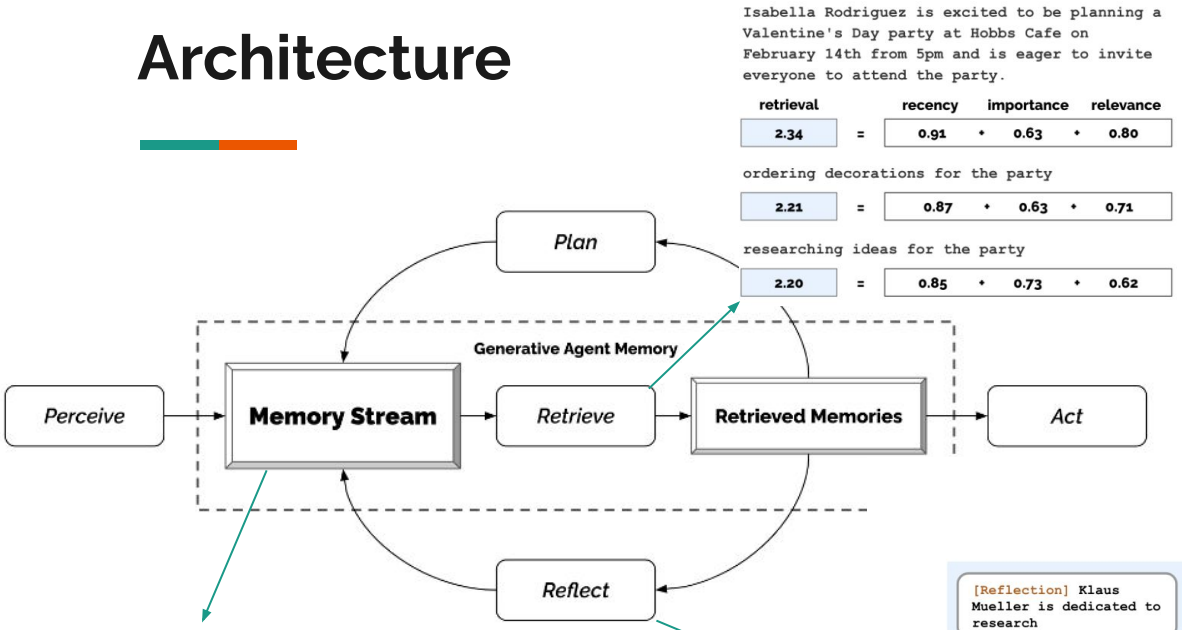
- **Bonus**: If the LLM can program… why not let it act through self-programming?

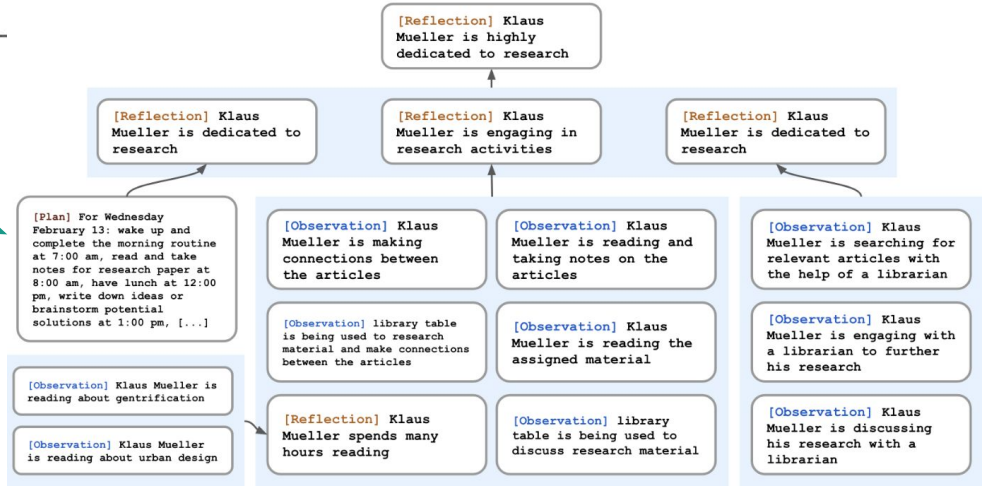
Generative Agents

[Interactive Simulacra of Human Behavior](#) (Park et al. Stanford)

Non-utilitarian paradigm.
Can we make human-like behaviour by giving LLMs a body and a way to structure memory?

Architecture



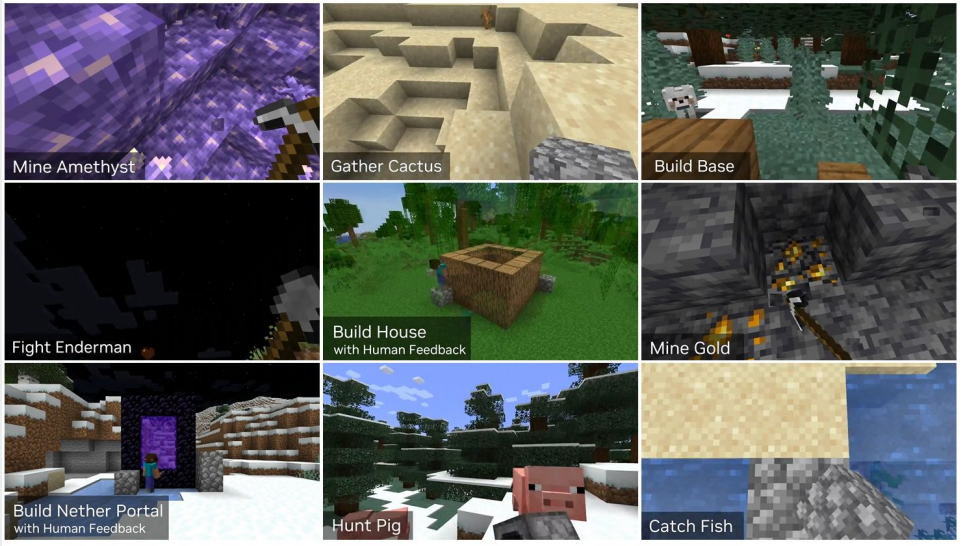


Description: John Lin is a pharmacy shopkeeper at the Willow Market and Pharmacy who loves to help people. He is always looking for ways …

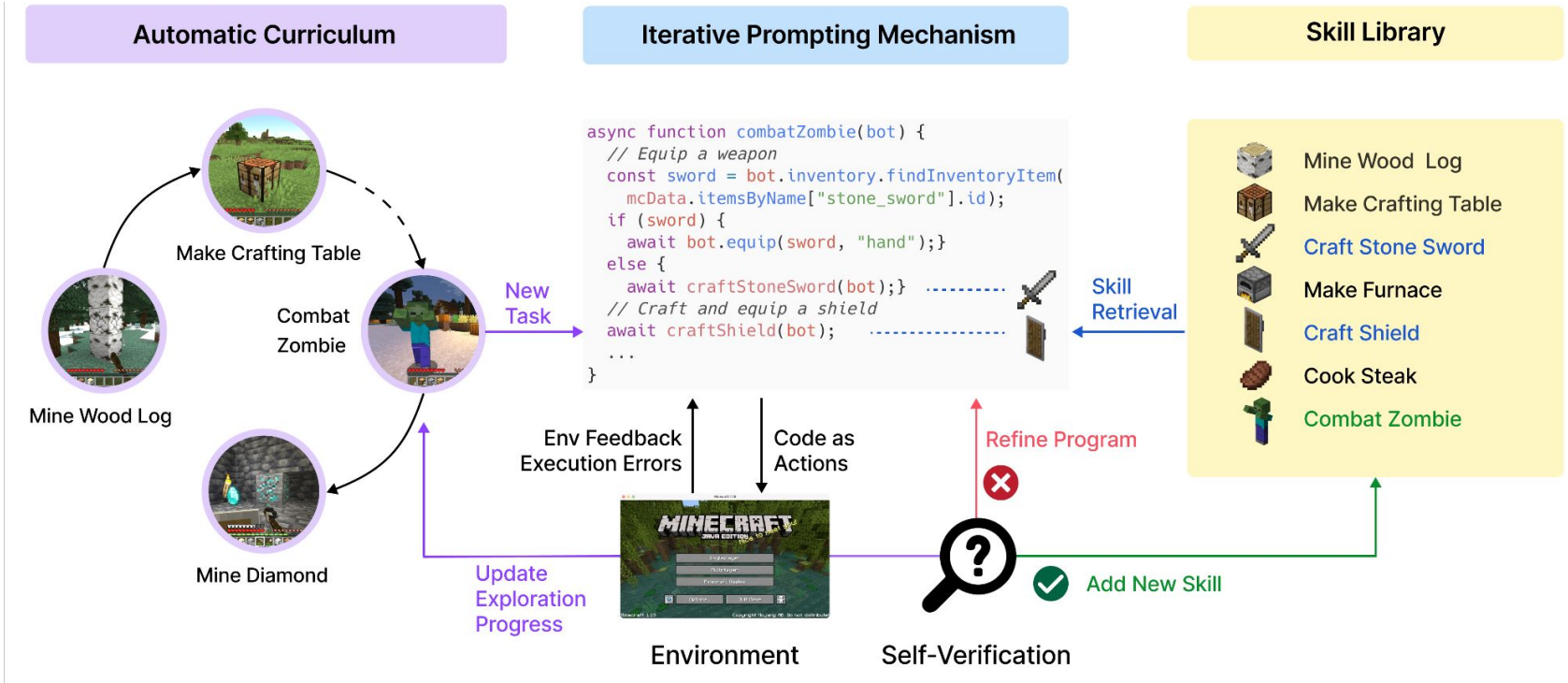
Voyager

[Open-Ended Embodied Agent with LLMs](#) (Jimmy Fan)

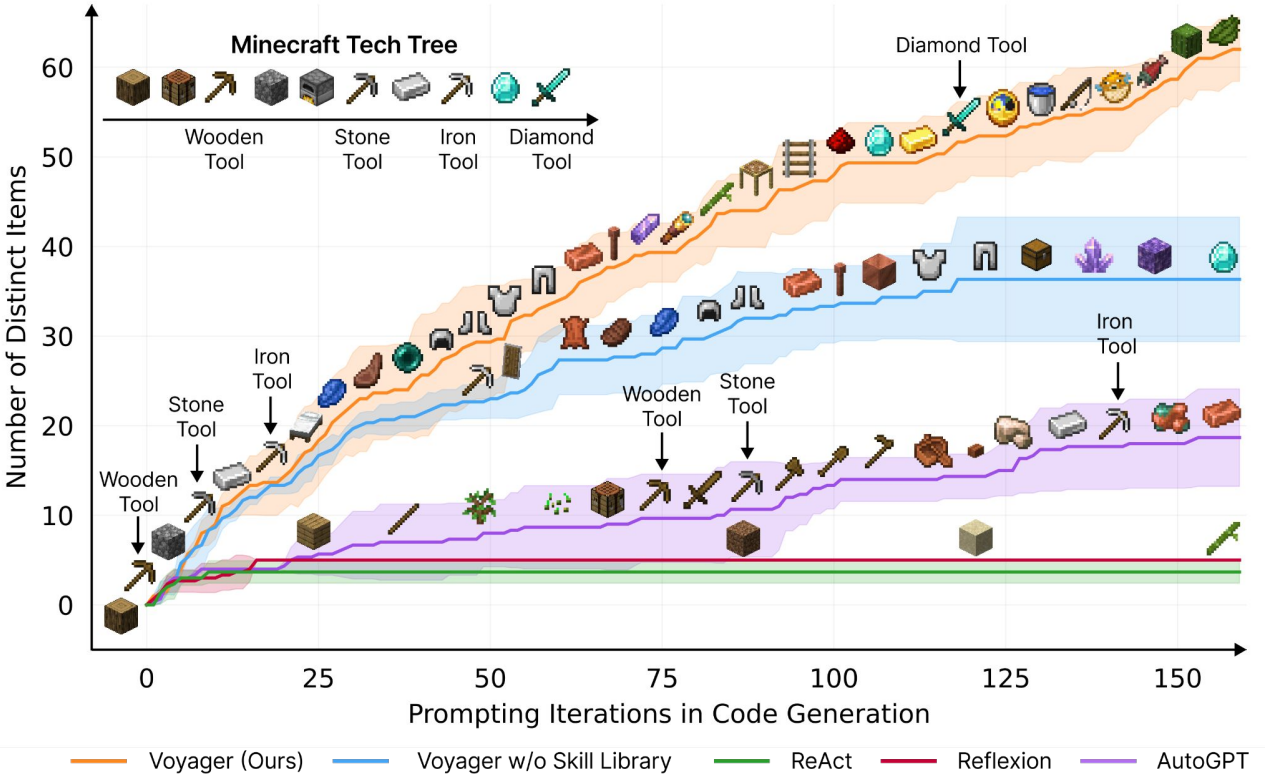
Motivation: If LLMs can do generative agents by ‘acting’ constrainedly, what if we let them play with an API and code?



Split in modules: desire, coder & ‘skills’



How good is it?



Voyager Conclusions

Declarative vs Procedural knowledge: representing procedural knowledge in a declarative way (skill-execution vs skill-description)

Curiosity! Relevance of a **curriculum-maker**

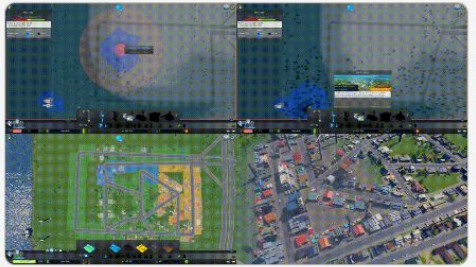
Declarative knowledge consultation & wiki-crawl

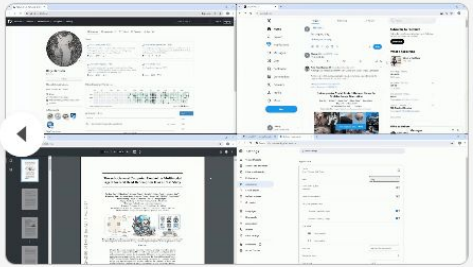
Intervention & Feedback

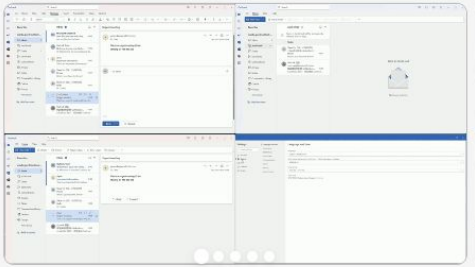
CRADLE

Empowering Agents toward Generalised Computer Control
(Nanyang University)

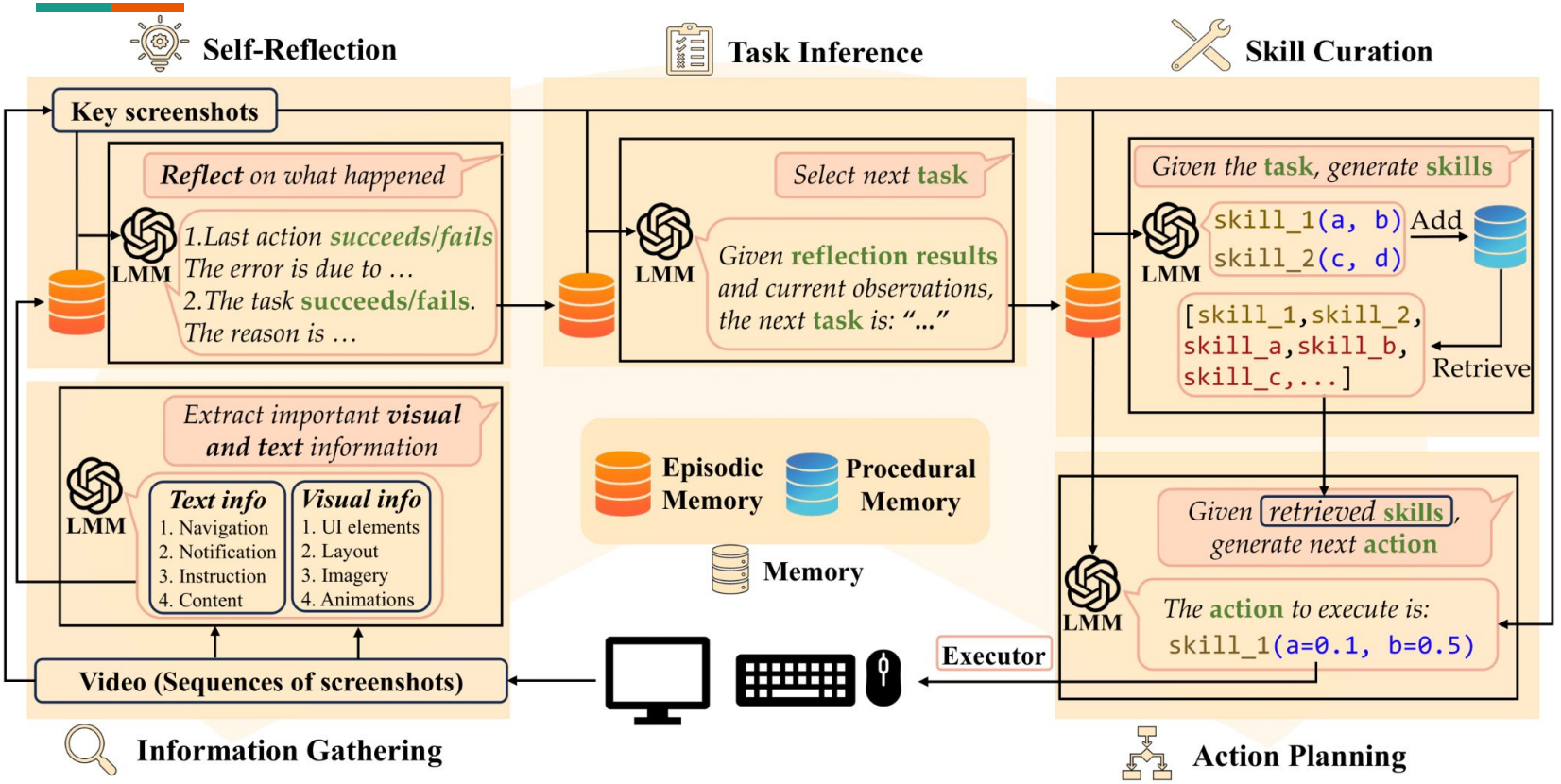
Motivation: Let's just give it control of mouse and keyboard, no API







Yet another architecture



Does it work?

Agentic AI is now one of the hot topics in industry

Millions of € on it. Our lab just (2025) got about 500K to work on this for a couple of years on a single project. We have at least 4 proposals on it, one valued in 80M.

I’ve directed 2 theses already working with this idea.

Copilot, embodied agents, assistive technologies, administrative handlers…

Discussion: Will they replace humans?

The world we find ourselves living in...

A video game company made a bot the CEO, and its stock climbed

Article (Business insider)

Article (The Independent)

(not a small business: NetDragon Websoft)

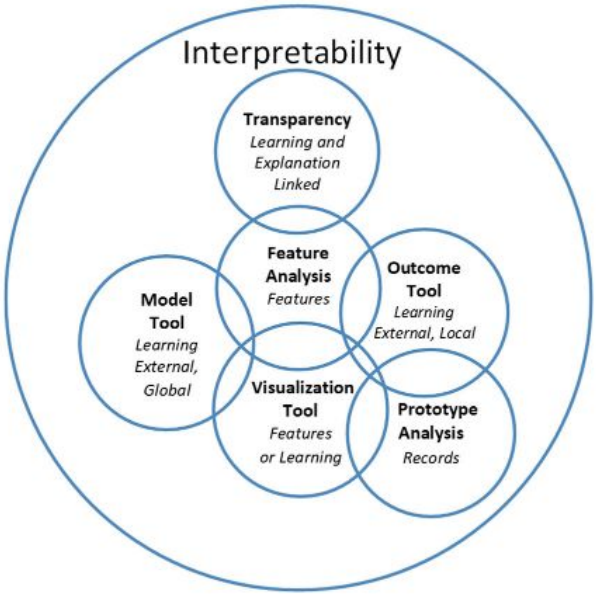
Other environments to tackle?

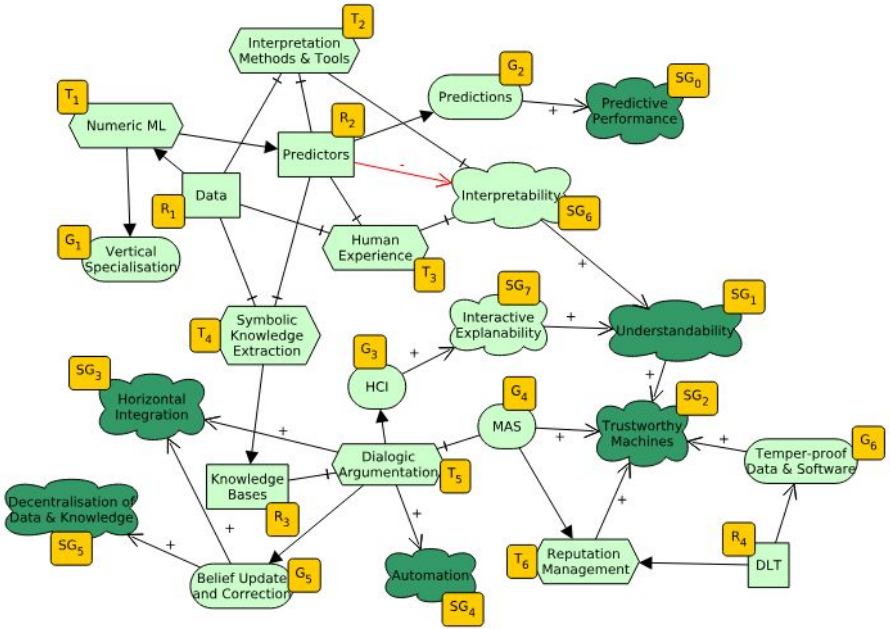
Develop AI Agents for System Engineering in *Factorio*

tain trade-offs and ensuring proactive adaptability. **This position paper advocates for training and evaluating AI agents' system engineering abilities through automation-oriented sandbox games—particularly *Factorio*.** By directing research efforts in this direction, we can equip AI

XAg: how to make sense of your agent

A big chaos in theory





What is actually done:

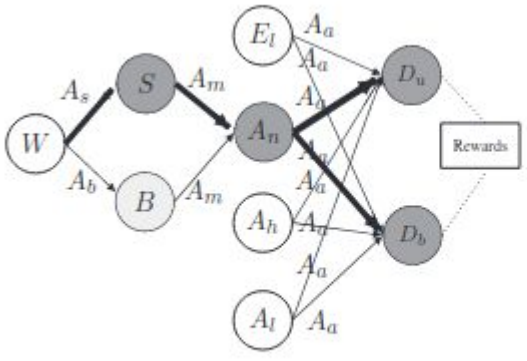
Base:

- What of S made you pick A?
- How would S need to change to pick A

Symbolic:

- More interesting stuff

Causal people: This →
(Madumal, Sonenberg, Miller)



State variables:
W - Worker number
S - Supply depot number
B - barracks number
E - enemay location
A_n - Ally unit number
A_h - Ally unit health
A_l - Ally unit location
D_u - Destoryed units
D_b - Destroyed buildings
Actions:
A_s - build supply depot
A_b - build barracks
A_m - train offensive unit
A_a - attack

Figure 1: Action influence graph of a Starcraft II agent

Winikoff/Dignum's Why Bad Coffee

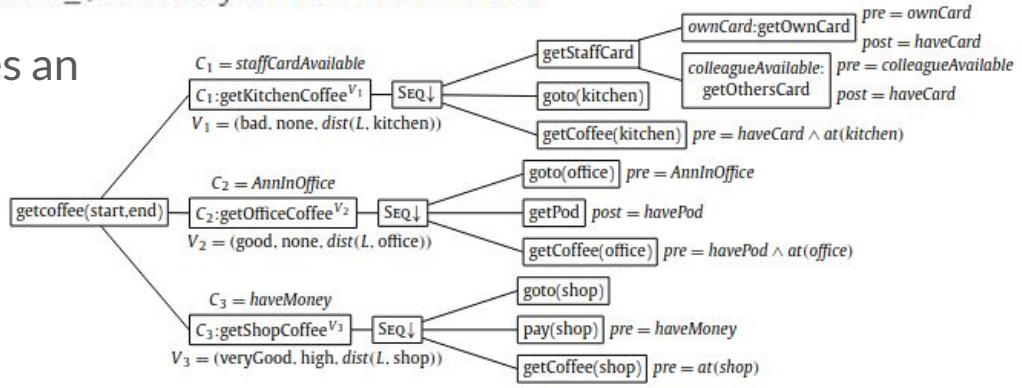
When Jo is low on money or busy he prefers the free office coffee which is good quality, but when he has time and money, he prefers the very good coffee from the shop.

$$(\text{veryGood}, \text{high}, x) <_{bank\_balance < \$400 \vee busy} (\text{good}, \text{none}, x)$$

$$(\text{good}, \text{none}, x) <_{bank\_balance \geq \$400 \wedge \neg busy} (\text{veryGood}, \text{high}, x)$$

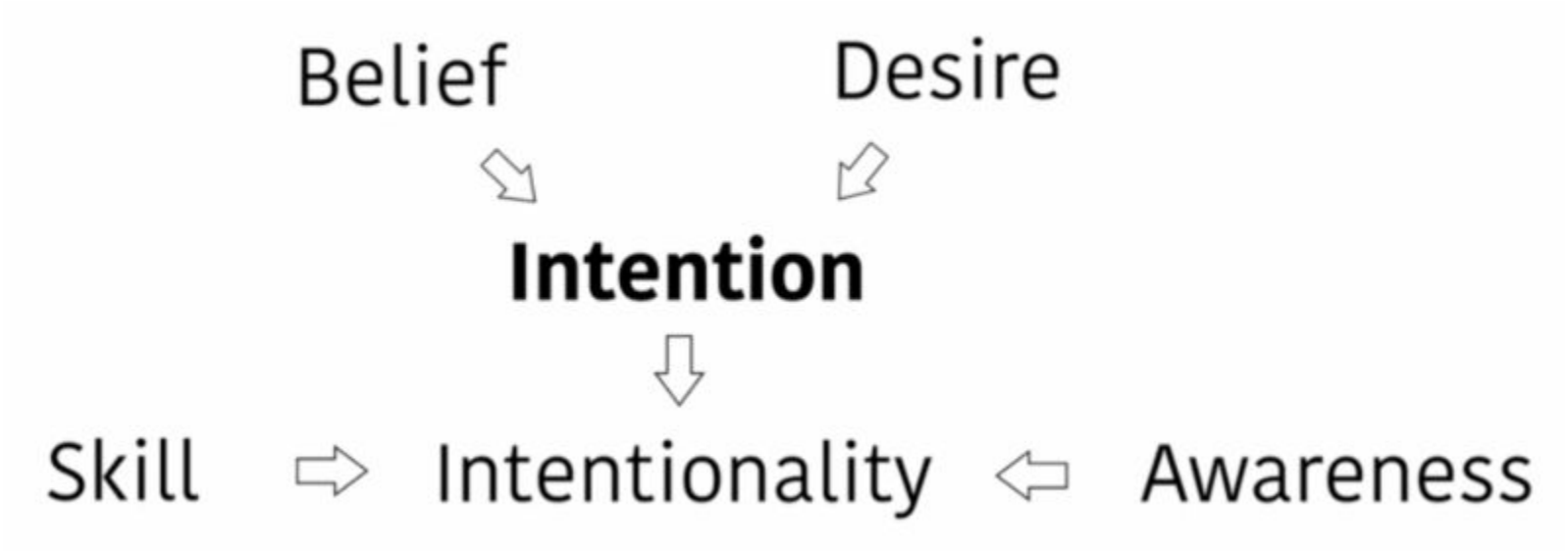
Effort on user-studies & what makes an explanation good.

Works exclusively on a particular case of BDI agent.

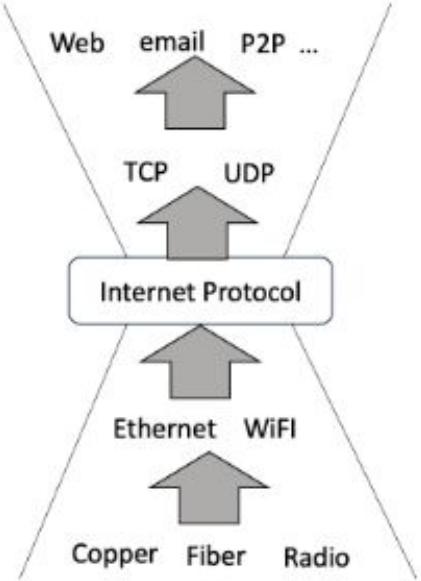


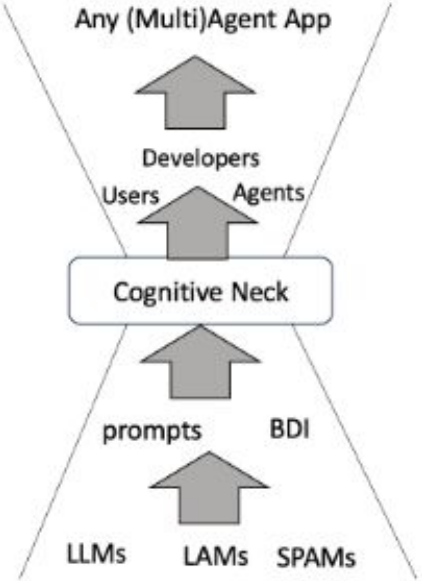
Paper

A proposal from folk-psychology



Ricci's Cognitive bottleneck



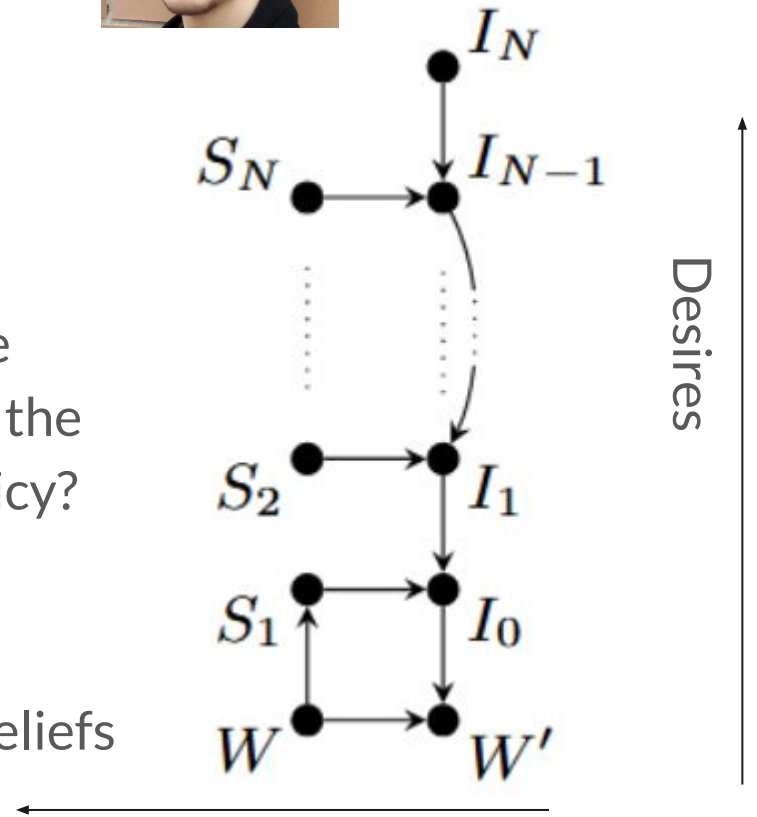


Me
(back before I lost hair)

Ladder model

Web (Paper pending proceedings)

Main intuition: is the ‘state’ (S1) the sole cause of an action (I0)? No, there is also the policy (I1). What are the causes of a policy?



Example of levels

n	REINFORCE	Q-learning	BDI	FB Representation	Voyager
1	Percepts+Reward	Percepts+Reward	Percepts	Percepts+Reward	Percepts + Errors + API (Mindflayer)
	Policy $(a \sim P^{\pi}(a\|s))$	Policy $(argmax_a Q(s,a))$	Plan	$argmax_a max_z$ $F(s,a,z)B(z,s')$	Program/skill
2	Empirical $v = Q(s,a)$, $\nabla_\theta log \pi_\theta(s,a)v$	Estimated Q-function $(Q(s,a))$	World model (*e.g.* PDDL domain file), desires	Successor Functions (F,B), desires/rewards of states	Available skills, Possible tasks, *Large Language Model* (LLM)[1], Feedback
	Policy training algorithm	Action-sampling policy generator	Means-ends reasoning to solve the goal	FB explorer (off-line); FB exploiter	Skill generator/corrector to solve a task
3		$\varepsilon = P(Q(s,a) < Rand\, a)$	Desire prioritisations	Given current goal	Task list priorisation[2], directive prompt
		Explore/exploit mechanism	Deliberation (goal selection)	Goal selector	Automatic curriculum planner loop
4			Values over desire prioritisation (when used)		
			Value reasoner (*e.g.* water tanks)		

Managing XAI with Intention in mind?

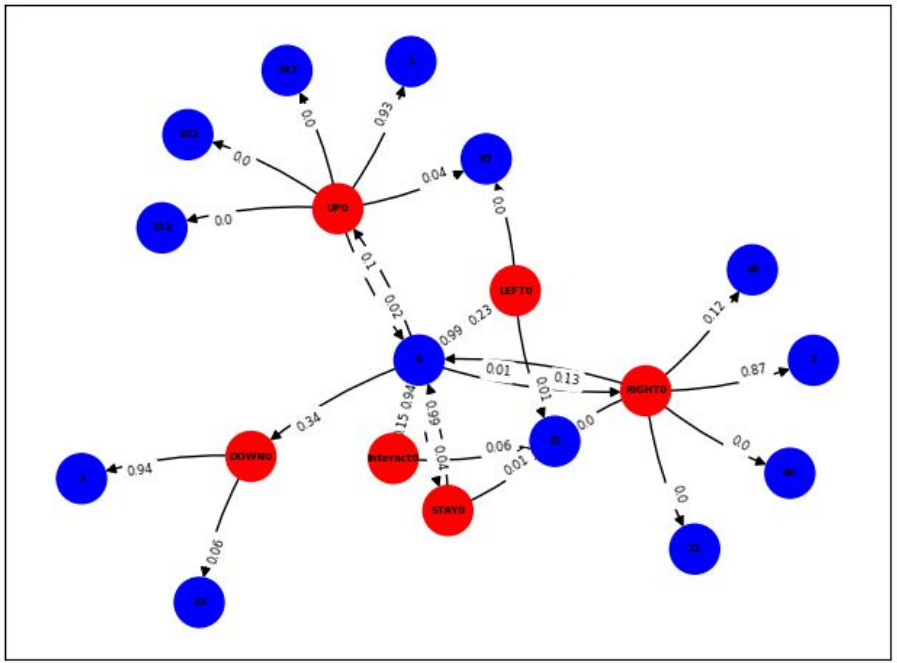
Intention: Belief + Desire

Attributing intention… to others? How does the human do it?

Target:

- observational (can only look at agent and environment)
- agent-agnostic (unknown architecture)
- reliable (should be truthful or have a notion of truthfulness)
- XAI method (provides answer to behaviour)

Policy Graphs (PG)



Intention-aware Policy Graphs (IPG)

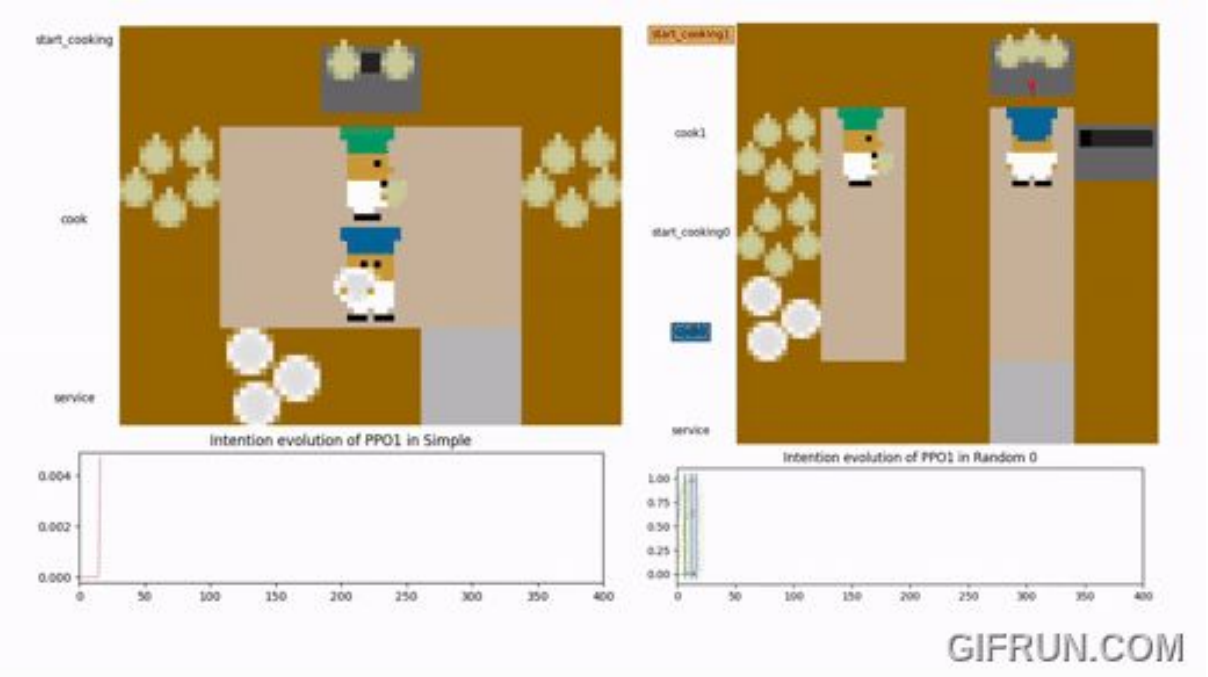
Add notion of intention:

Probability that agent will *bring about* something

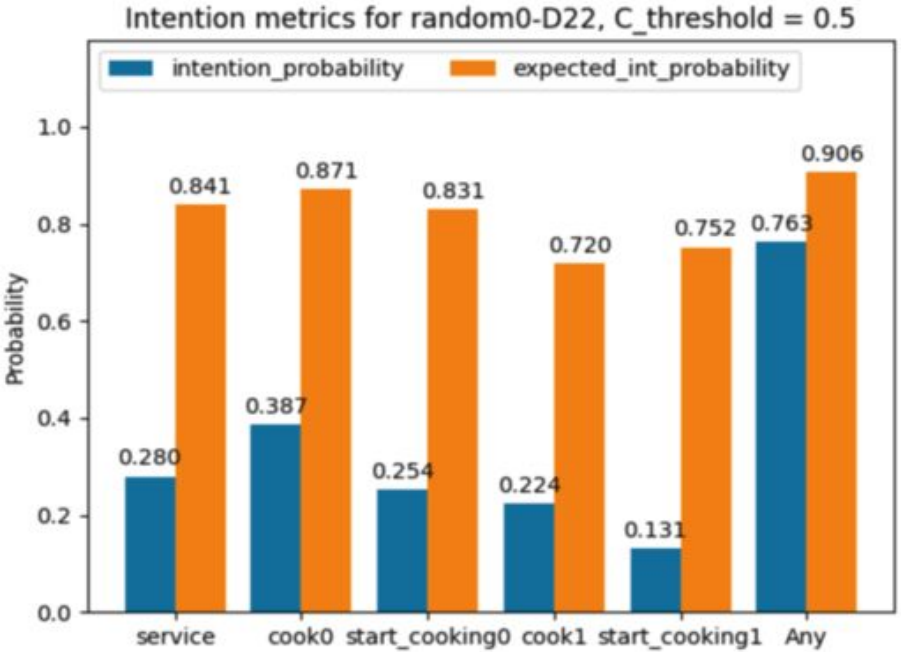
P(s'|a,s): World belief

P(a|s): agent actions

d: 'a hypothesised desirable transition/state'



Is it truthful? Is it interpretable?

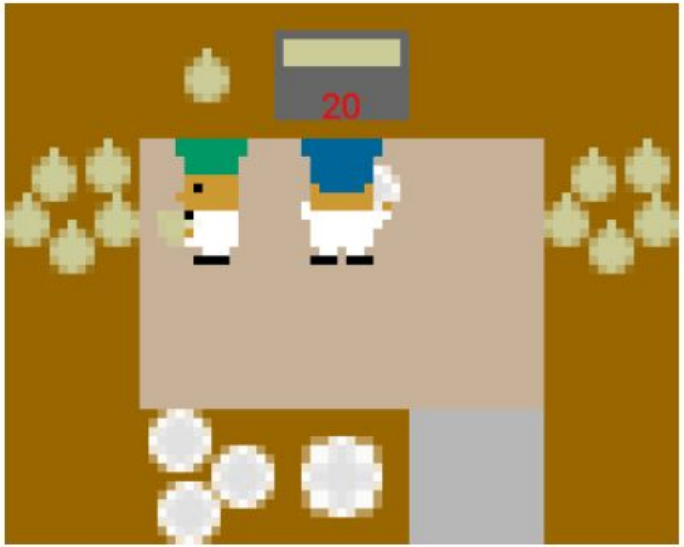


77% of the time
I understand

90% of the time
I'm right

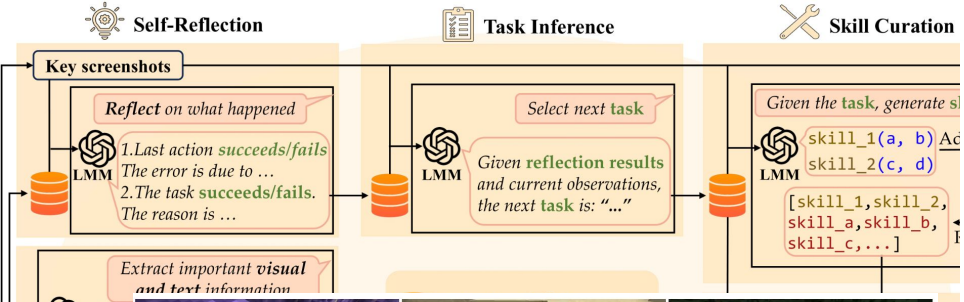
Proportional to rationality! Better agents = more explainable

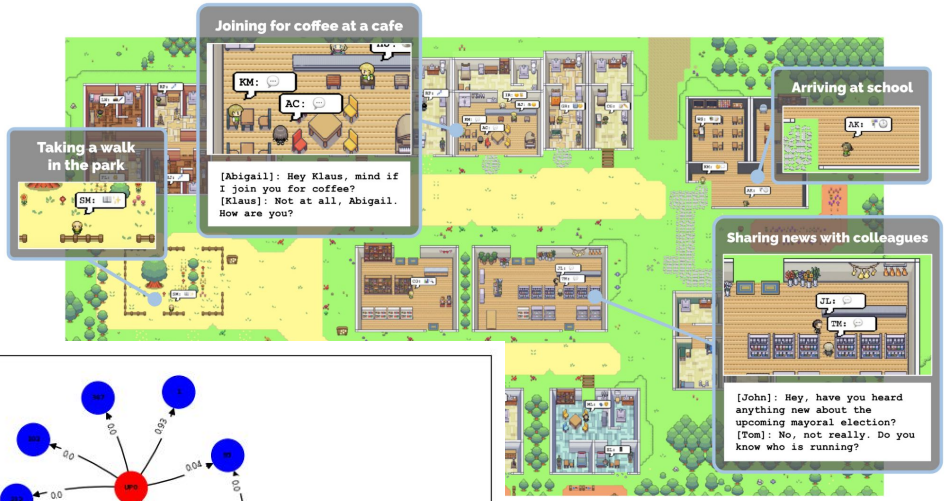
Basic XAI test-questions

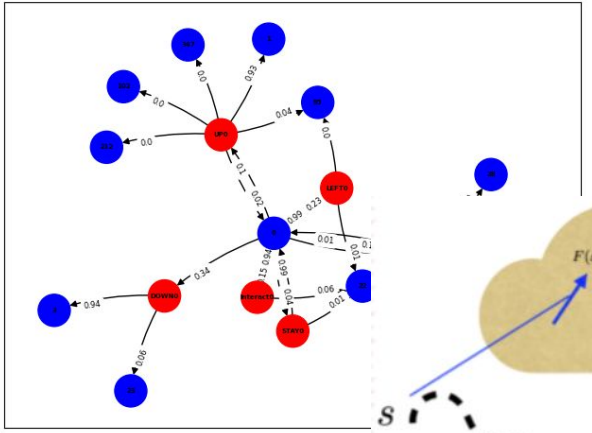


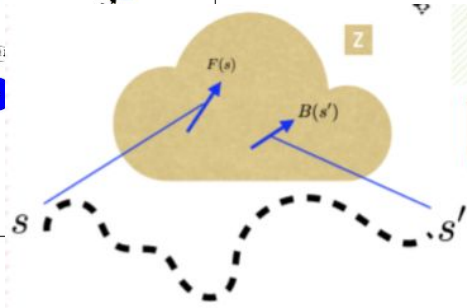
- What does it intend to do?
  Desire to **SERVE SOUP**: 0.625
- Why would it **INTERACT**?
  Intentions attributed for **SERVING SOUP**
  expected to increase by 0.05.
- How would it fulfill its desire?
  By performing the following chain of actions:

Interact(0.82)	Right(0.89)	Down(1.0)	Interact
held(S) pot_state (Pot$_0$;Empty)	item_pos(O;I) item_pos (Pot$_0$;←) item_pos (Service;↓) item_pos(S;↓)	item_pos(O;↑) item_pos (Service;I) item_pos(S;→) pot_state (Pot$_0$;Cooking)	
held(D) pot_state (Pot$_0$;Finished)	item_pos(O;→) item_pos (Pot$_0$;I) item_pos (Service;→) item_pos(S;→)	item_pos(O;I) item_pos (Service;↓) item_pos(S;↓) pot_state (POT$_0$;Empty)	









And that's about it for today