Agent Learning from Interactions:

Unifying Grounding and Large Language Models for Generalizable Agency

Agent Learning from Interactions:
Unifying Grounding and Large Language Models for Generalizable Agency

Keynote Speaker:
A/Prof. Wei Peng
Principal Research Fellow in AI

School of Engineering
RMIT University

Overview

- My Background
- Learning Pathway
- Learning from Interactions: Grounding Problem, Embodied Intelligence
- Towards more Generalizable LLM-based Agency: Agent Trajectory Learning
- Towards more Generalizable LLM-based Agency: Multi-agents Collective Decision-making

My Background

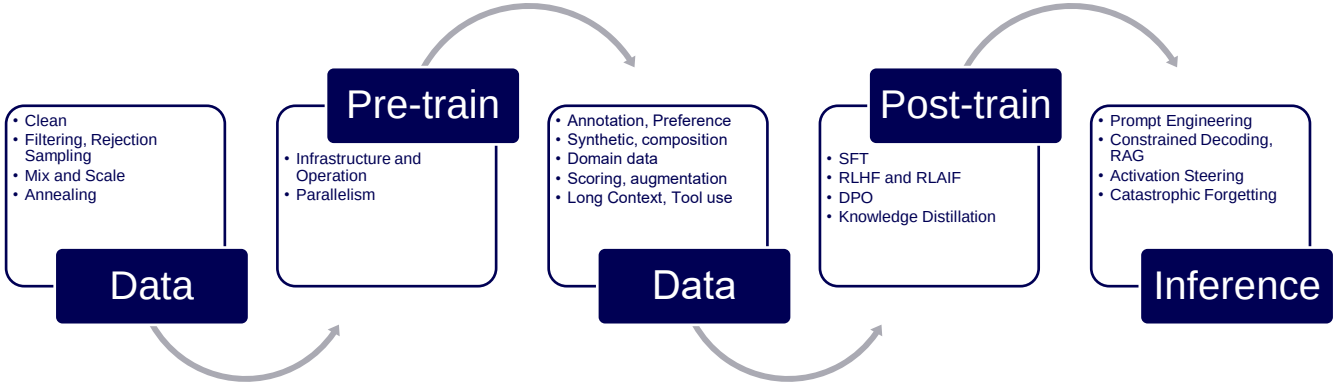
- 2025 – now: A/Professor, Principal Research Fellow in Industrial AI, School of Engineering, RMIT University
- 2019 – 2025: Research Director and AI Expert, BPIT, Huawei Technologies
- 2016 – 2019: Senior Lecturer, La Trobe University
- 2010 – 2016: Senior Data Scientist, AUSTRAC, Lenovo, Telstra
- 2008 – 2010: Research Fellow, RMIT University
- 2007 – 2008: Research Scientist, CSIRO ICT
- 2007: PhD from the University of Sydney in Cognitive Agents

Research Interests: *Large Language Models, Knowledge Representation, Cognitive Agents, and Embodied AI, along with their applications, 70+ papers, multiple patents*

Career Highlights: *research, innovation, and commercialization across both industry and academia*

Bifurcated Pathway to AGI (Data-driven)

1. Human knowledge build (Cybernetics, KBS, to Data and Computing Power, LLM/scale law/transient learning on features for tasks)

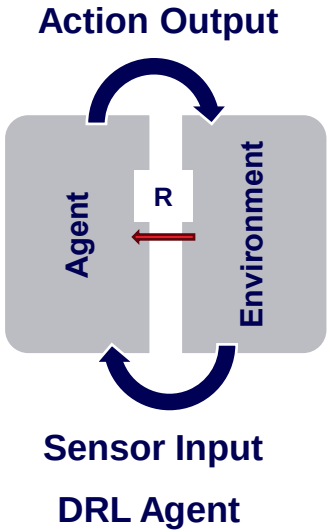


We try to scale model in development to host all human knowledge and capability by feeding mega data …

Bifurcated Pathway to AGI (Experience-driven)

2. Continual learning from experience

Rich Sutton's new path for AI : "… RL in AI, we don't have methods to learn continuously except for the linear case …" https://www.youtube.com/watch?v=NvfK1TkXmOQ

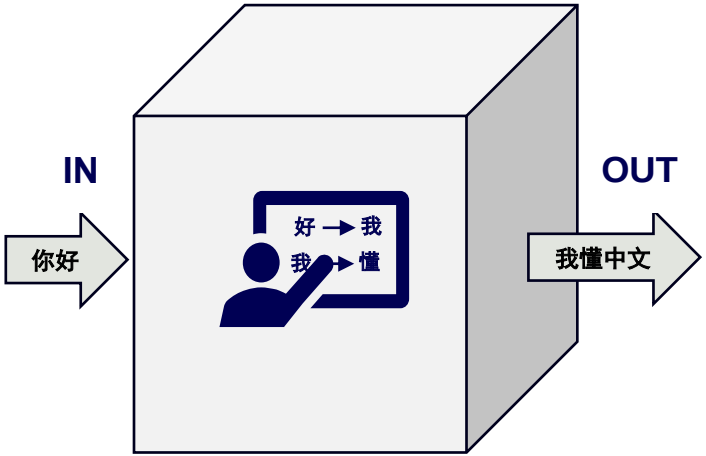


Learn from Interactions Challenges:

- Generalization: beyond the env. trained
- Inefficiency: learn from limited examples
- Catastrophic Forgetting: learn without forgetting priori knowledge
- Reward Misalignment: multi-objective, human values
- Lack World Model and Abstraction: reason on concept

Learning from Interactions (Grounding Problem)

- **Grounding:** intrinsic process of assigning meanings to symbols/words/vectors/concepts by referencing to real world experience (objects, events).
- **Symbolic grounding** (Harnad, 1990): how can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meaning in our heads?
- **Representation grounding** (Chalmers, 1992): how can a representation in a computational system possess true meaning?
- **Concept grounding** (Dorffner & Prem, 1993): design a cognitive model (connectionism) only interfacing with its environment using sensor and motor signals; any concept of the system develops through self-organization based on adaptive interaction with the environment (besides given meta-level representation like innate architecture) – is grounded in Harnard's sense.



Searle's Chinese Room

Learning from Interactions (Embodied Intelligence)

- Rodney Brooks' Intelligence without Representation (Brooks, 1991): no traditional representation, intelligence from sensor motor interaction with the environment, behavior-based model

Take Aways: Learn from situated sensor motor coordination to generate complex behaviors

Challenges: Limited memory, planning, reasoning capability, simplistic world model

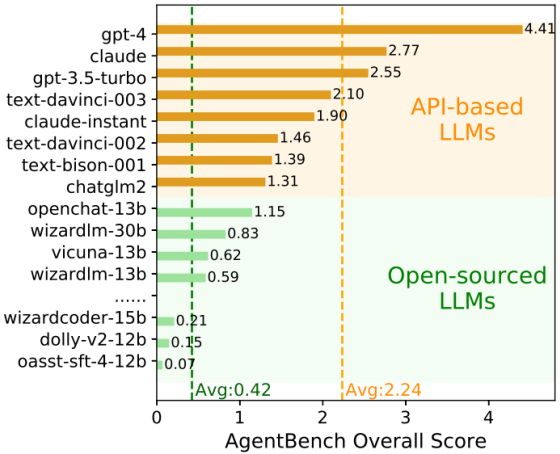
Key Insight: ground disembodied intelligence (i.e., LLM) in interactions to develop concepts (levels of abstractions) in self-organized manner

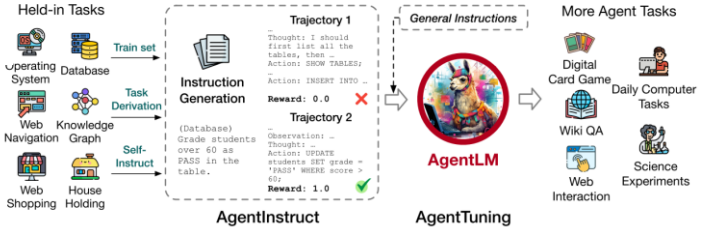
Agent Trajectory Learning

- **Agent Learning from Interaction:**

AgentBank: Towards Generalized LLM Agents via Fine-Tuning on 50000+ Interaction Trajectories, *In Findings of the Association for Computational Linguistics*, pages 2124–2141, Association for Computational Linguistics (citation 13)

Expert Trajectories





Agents solve only simple tasks, hard to generalize

AgentTuning (Zeng et al., 2024): only keep successful trajectories for training

Skill Dim.	Task	Action Space	Tool	#Inst.	Avg. Turns	Action Annotation
Reasoning	HotpotQA (Yang et al., 2018)	Continuous	Search	4273	3.1	Explore
	StrategyQA (Geva et al., 2021)	Continuous	Search	1267	3.6	Explore
	TriviaQA (Joshi et al., 2017)	Continuous	Search	4134	2.5	Explore
Math	GSM8K (Cobbe et al., 2021)	Continuous	Calculator	7471	4.5	Reformat
	MathQA (Amini et al., 2019)	Continuous	Python	4000	2.0	Explore
	MATH (Hendrycks et al., 2021)	Continuous	Python, Wiki	2312	2.5	Explore
Programming	IC-SQL (Yang et al., 2023)	Continuous	MySQL	4540	4.8	Explore+Answer Force
	APPS (Hendrycks et al., 2021)	Continuous	Python	4408	1.0	Reformat
	HumanEval (Chen et al., 2021)	Continuous	Python	134	2.7	Explore+Answer Force
	MBPP (Austin et al., 2021)	Continuous	Python	608	2.2	Explore+Answer Force
Web	Mind2Web (Deng et al., 2023)	Discrete	-	7770	1.0	Reformat
	WebArena (Zhou et al., 2023)	Discrete	-	657	1.0	Reformat
	WebShop (Yao et al., 2022a)	Discrete	-	5315	3.4	Explore & Reformat
Embodied	ALFWorld (Shridhar et al., 2020b)	Discrete	-	3554	10.1	Reformat
	RoomR (Weihs et al., 2021)	Discrete	-	300	30.2	Search+Reformat
	IQA (Gordon et al., 2018)	Discrete	-	1627	28.4	Search+Reformat
	Total (AGENTBANK)	-	-	51287	3.9	-

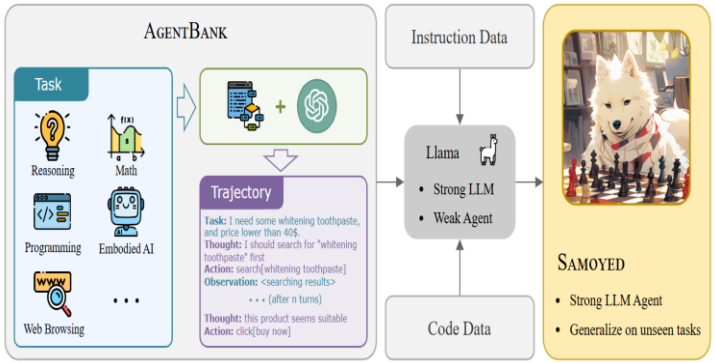
LLMs suffer even for SOTA commercial tools AgentBench (Liu et al., 2023)

We propose AgentBank to host 50000+ interaction trajectories

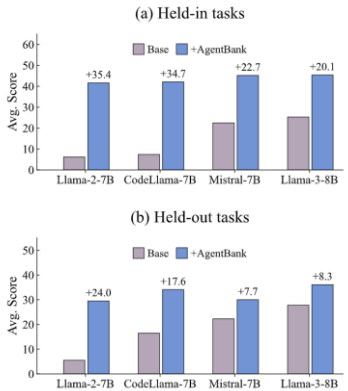
Trajectory Learning (cont.)

- **Agent Learning from Interaction:**

AgentBank: Towards Generalized LLM Agents via Fine-Tuning on 50000+ Interaction Trajectories, *In Findings of the Association for Computational Linguistics*, pages 2124–2141, Association for Computational Linguistics (citation 13)



Model	Held-in Tasks						Held-out Tasks					
	Reason	Math	Program	Web	Embodied	Avg.	Reason	Math	Program	Web	Embodied	Avg.
Closed-Source Model												
GPT-4	61.6	73.0	54.9	40.6	77.8	59.8	41.6	51.0	69.4	69.4	36.4	53.6
GPT-3.5-Turbo	41.0	41.5	51.2	42.0	10.5	40.2	32.0	32.0	54.8	66.7	21.2	41.3
7B Open-Source Model												
Llama-2-7B-Chat	4.0	7.5	2.5	13.9	0.0	6.2	4.0	8.0	7.0	0.4	7.8	5.5
Vicuna-7B	29.0	2.0	19.0	24.2	6.0	17.1	8.8	14.0	19.0	18.2	12.8	14.6
CodeLlama-7B	3.5	3.5	1.5	24.8	0.0	7.4	1.0	13.0	21.8	**41.3**	5.5	16.5
AgentLM-7B	29.5	10.0	12.0	**37.2**	63.4	26.7	19.2	13.0	50.5	13.5	13.3	21.9
Agent-FLAN-7B	31.0	10.5	13.1	35.4	**65.3**	27.3	22.2	11.0	53.1	17.9	14.1	23.7
SAMOYED-7B	**48.0**	**30.5**	**41.6**	36.4	61.2	**41.6**	**32.0**	**18.0**	**59.2**	24.2	**14.2**	**29.5**
13B Open-Source Model												
Llama-2-13B-Chat	12.5	10.5	8.2	11.2	0.0	9.4	9.6	11.0	33.0	17.6	7.3	15.7
Vicuna-13B	25.5	6.5	30.4	34.2	2.2	21.7	24.8	17.0	37.0	34.2	14.8	25.6
CodeLlama-13B	13.5	18.5	5.1	15.3	0.0	11.7	6.4	16.0	11.1	**46.5**	5.5	17.1
AgentLM-13B	38.0	13.5	22.8	38.1	52.2	30.8	20.8	13.0	46.6	21.6	14.6	23.3
SAMOYED-13B	**54.5**	**38.5**	**55.4**	**40.9**	**72.4**	**50.1**	**35.0**	**23.0**	**62.4**	38.9	**18.4**	**35.5**

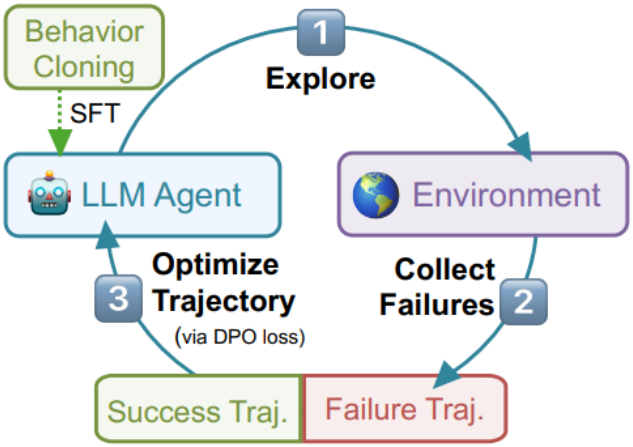


Organize trajectories into multi-turn dialogues, mix general domain instructions and codes, utilize failure trajectories and propose the exploration-based trajectory optimization (ETO) method to learn the task-solving process, leading to significant performance gains.

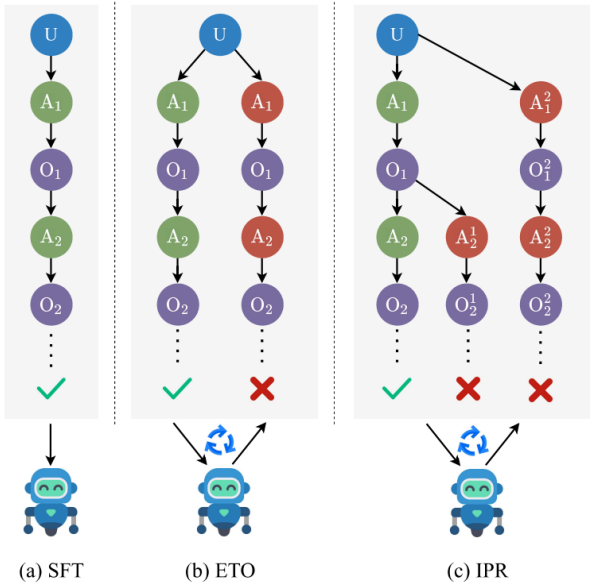
Trajectory Learning (cont.)

- **Agent Learning from Interaction:**

Watch Every Step! LLM Agent Learning via Iterative Step-Level Process Refinement, In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024),* pages 1556-1572, Association for Computational Linguistics (citation 20)



ETO (Song et al., 2024)



- **Agents start to learn from Interactions and explorations: from SFT on trajectories to ETO (SAMOYED)**

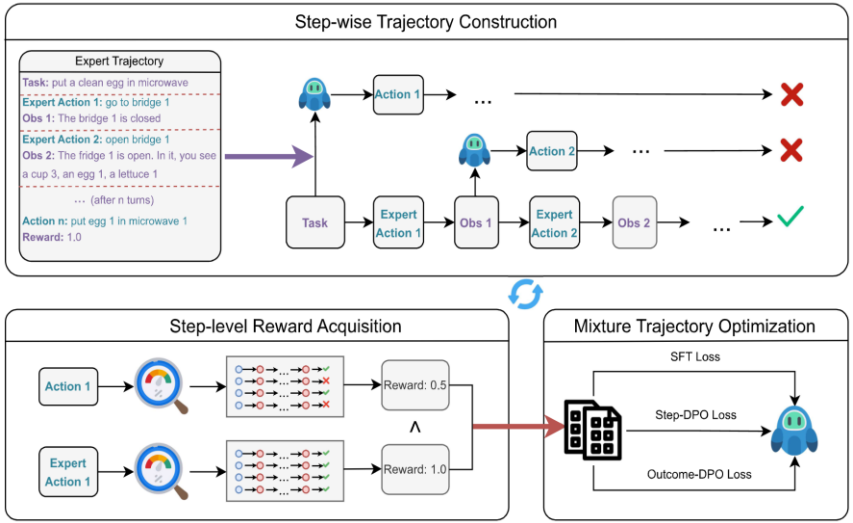
Treat an entire trajectory as single entity during training and prioritize the final reward of a trajectory over the process, thus overlooking exploitable information throughout interaction process.

We need to consider step level optimization

Trajectory Learning (cont.)

- **Agent Learning from Interaction:**

Watch Every Step! LLM Agent Learning via Iterative Step-Level Process Refinement, In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024),* pages 1556-1572, Association for Computational Linguistics (citation 20)



Design Step-level Process Refinement: Step-level Reward Acquisition and Iterative Agent Optimization.

Monte Carlo method to estimate rewards via sampling N trajectories to construct step award.

$$\{e^{(i)}|i=1,...,N\} = MC^{\pi_s}(e_{t-1};N),$$

$$r_s(s_t,a_t) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} r_o(u,e^{(i)}), & \text{for } t<n \\ r_o(u,e_n), & \text{for } t=n \end{cases}$$

$$\mathcal{L} = \mathcal{L}_{\text{o-DPO}} + \mathcal{L}_{\text{s-DPO}} + \mathcal{L}_{\text{SFT}}$$

- Outcome-DPO Loss

$$\mathcal{L}_{\text{o-DPO}} = -\mathbb{E}_{(u,e_n^w,e_m^l)\sim\mathcal{D}_t}\left[\log\sigma(\beta\log\frac{\pi_\theta(e_n^w|u)}{\pi_{ref}(e_n^w|u)} - \beta\log\frac{\pi_\theta(e_m^l|u)}{\pi_{ref}(e_m^l|u)})\right],$$

- Step-DPO Loss

$$\mathcal{L}_{\text{s-DPO}} = -\mathbb{E}_{(e_{t-1},e_{t:n}^w,e_{t:m}^l)\sim\mathcal{D}_s}\left[\log\sigma(\beta\log\frac{\pi_\theta(e_{t:n}^w|e_{t-1})}{\pi_{ref}(e_{t:n}^w|e_{t-1})} - \beta\log\frac{\pi_\theta(e_{t:m}^l|e_{t-1})}{\pi_{ref}(e_{t:m}^l|e_{t-1})})\right],$$

- Supervised Loss

$$\mathcal{L}_{\text{SFT}} = -\mathbb{E}_{(u,e_n^w,e_m^l)\sim\mathcal{D}_t}\left[\log\pi_\theta(e_n^w|u)\right],$$

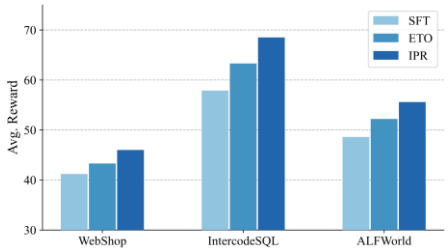
Trajectory Learning (cont.)

- **Agent Learning from Interaction:**

Watch Every Step! LLM Agent Learning via Iterative Step-Level Process Refinement, In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024),* pages 1556-1572, Association for Computational Linguistics (citation 20)

Paradigm	Models	WebShop	InterCodeSQL	ALFWorld		Average
				Seen	Unseen	
Prompt-based	GPT-4 (Achiam et al., 2023)	63.2	38.5	42.9	38.1	45.7
	GPT-3.5-Turbo (Ouyang et al., 2022)	62.4	37.8	7.9	10.5	29.7
	Llama-2-7B (Touvron et al., 2023)	17.9	4.0	0.0	0.0	5.5
Outcome Refinement	Llama-2-7B + SFT (Chen et al., 2023)	60.2	54.9	60.0	67.2	60.6
	Llama-2-7B + PPO (Schulman et al., 2017)	64.2	52.4	22.1	29.1	42.0
	Llama-2-7B + RFT (Yuan et al., 2023)	63.6	56.3	62.9	66.4	62.3
	Llama-2-7B + ETO (Song et al., 2024)	67.4	57.2	68.6	72.4	66.4
Process Refinement	Llama-2-7B + Step-PPO	64.0	60.2	65.7	69.4	64.8
	Llama-2-7B + IPR (ours)	**71.3**	**61.3**	**70.3**	**74.7**	**69.4**

Training Scheme	WebShop	InterCodeSQL	ALFWorld
w/o o-DPO	70.2	59.3	72.4
w/o s-DPO	66.4	58.0	70.2
w/o SFT	61.8	31.7	64.9
Iteration=1	63.6	56.6	68.7
Iteration=2	63.7	58.2	70.2
Iteration=3	68.2	59.2	**74.7**
Iteration=4	**71.3**	**61.3**	73.5
Iteration=5	68.1	57.9	71.4



Conclusion:

- Agent learns from interaction via trajectory with step awards
- Learning from failure actions
- Automated process reward acquisition
- Step level process supervision via mixture trajectory optimization
- Enhanced performance on three benchmarks
- Generalizable on unseen hold out

Limitation:

- Overfitting with limited data (need to leverage AgentBank data)
- MC method constrained by sample size
- Consider GPT 4 to label process supervision data

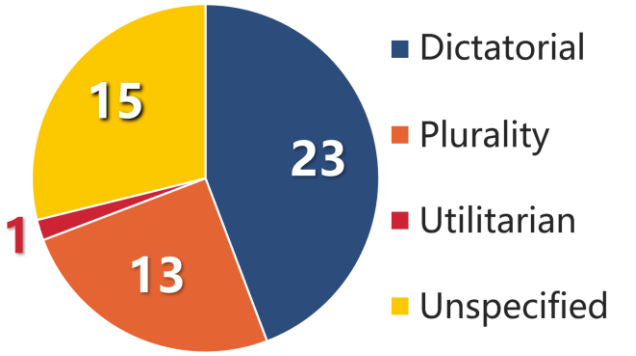
Multi-Agents Collective Decision-making

- **Multi-agents Collective Decision Making (CDM):**

An Electoral Approach to Diversify LLM-based Multi-Agent Collective Decision-Making, *in Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024),* pages 2712–2727, Association for Computational Linguistics, https://aclanthology.org/2024.emnlp-main.158/

CDM Method	Systems and Frameworks	Note
Dictatorial	Xiong et al. (2023)	Assigned role
	Wu et al. (2023)	Assigned role
	Hao et al. (2023)	Assigned role
	Liu et al. (2023b)	Assigned role
	Li et al. (2023a)	Assigned role
	Zhang et al. (2023a)	Assigned role
	Nair et al. (2023)	Assigned role
	Talebirad and Nadiri (2023)	Assigned role
	Liang et al. (2023)	Assigned role
	Tang et al. (2023)	Assigned role
	Qian et al. (2023)	Assigned role
	Sun et al. (2023)	Assigned role
	Chen et al. (2023a)	Assigned role
	Jinxin et al. (2023)	Assigned role
	Li et al. (2023b)	Assigned role
	Fang et al. (2024)	Assigned role
	Tang et al. (2024)	Assigned role
	Hang et al. (2024)	Assigned role
	D'Arcy et al. (2024)	Assigned role
	Hua et al. (2024)	Assigned role
	Wang et al. (2024b)	Assigned role
	Li et al. (2023f)	Assigned role
	Chen et al. (2023b)	Oligarchy
No CDM or Unspecified	He et al. (2023)	Decentralized team
	Li et al. (2023c)	Decentralized team
	Nakajima (2023)	Decentralized team
	Ni and Buehler (2024)	Human judgement
	Ghafarollahi and Buehler (2024)	Human judgement
	Wang et al. (2023a)	Linear workflow
	Ding et al. (2023)	Linear workflow
	Hong et al. (2023)	Linear workflow
	Rasheed et al. (2024)	Linear workflow
	Wei et al. (2024)	Linear workflow
	Liu et al. (2023a)	Scenario simulation
	Park et al. (2023)	Scenario simulation
	Ghaffarzadegan et al. (2023)	Scenario simulation
	Hua et al. (2023)	Scenario simulation
	Zhang et al. (2024)	Scenario simulation
Plurality	Du et al. (2023)	Consensus
	Wang et al. (2023d)	Consensus
	Chen et al. (2023d)	Consensus
	Chen et al. (2023c)	Consensus
	Li et al. (2023d)	Consensus
	Shi et al. (2023)	Game rule
	Stepputtis et al. (2023)	Game rule
	Xu et al. (2023a)	Game rule
	Chan et al. (2023)	Relative majority
	Xu et al. (2023b)	Relative majority
	Zhang et al. (2023b)	Relative majority
	Li et al. (2024)	Relative majority
	Hamilton (2023)	Scenario simulation
Utilitarian	Jarrett et al. (2023)	

Table 4: Full list of 52 surveyed LLM-based multi-agent collaboration works.



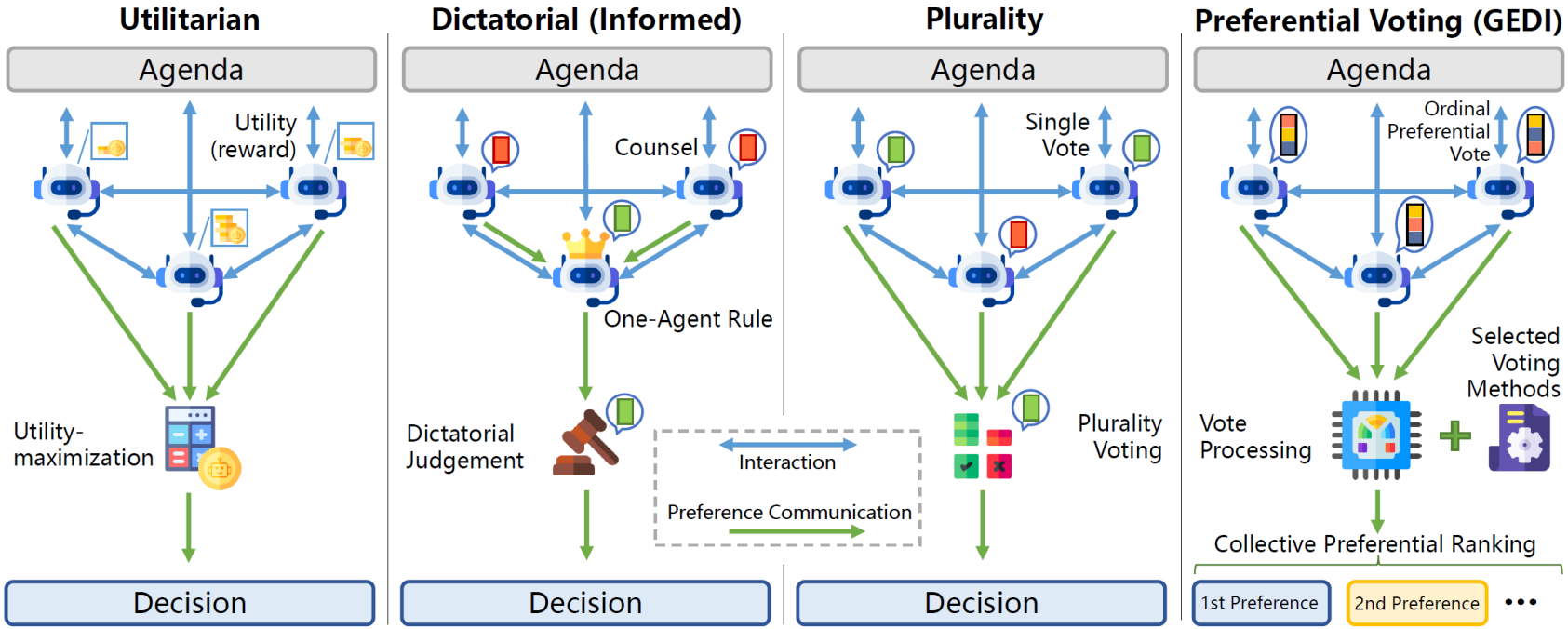
52 multi-agent collaboration frameworks: lack of diversity in Collective Decision-making (CDM)

CDM Method	Major -ity	Mono -tonic	Consis -tency	IIA	Cond -orcet	Ballot type
Dictatorial (Blind)	✗	✓	✓	✓	✗	Ranking
Range Voting	✗	✓	✓	✓	✗	Scores
Plurality	✓	✓	✓	✗	✗	Single*
Borda Count	✗	✓	✓	✗	✗	Ranking
IRV	✓	✗	✗	✗	✗	Ranking
Ranked Pairs	✓	✓	✗	✗	✓	Ranking

Table 1: Criteria compliance of some typical CDM methods. *Range Voting* can be viewed as a special *utilitarian* method. **IIA** denotes *Independence from Irrelevant Alternatives*. *Single ballots can be derived from ranking ones. Find some examples in Appendix D.

Kenneth Arrow's Social Choice Theory

Diversifying CDM in LLM MAS



Key Findings

Base Model	Rand.	Score	Dictatorial-based			Ordinal Ranking					
MMLU	Rand.	Range Voting	Blind Dicta.	Informed Dicta.	Mis-Informed Dicta.	Plurality	Bucklin	Borda Count	IRV	Minimax	Ranked Pairs
mistral-7b	24.8	51.8 (-4.6)	56.4	55.9 (-0.5)	36.1 (-20.3)	56.8 (+0.4)	57.1 (+0.7)	56.9 (+0.5)	56.9 (+0.5)	57.0 (+0.6)	57.0 (+0.6)
llama-3-8b	25.0	37.7 (-7.3)	45.0	36.5 (-8.5)	32.2 (-12.8)	45.9 (+0.9)	46.4 (+1.4)	46.3 (+1.3)	45.7 (+0.7)	45.9 (+0.9)	46.0 (+1.0)
glm-4-9b	25.2	61.3 (-0.4)	61.7	54.3 (-7.4)	**53.0 (-8.7)**	**64.6 (+2.9)**	**64.5 (+2.8)**	64.1 (+2.4)	**64.9 (+3.2)**	**64.4 (+2.7)**	**64.6 (+2.9)**
llama-3-70b	25.3	74.9 (+1.6)	73.3	70.1 (-3.2)	62.6 (-10.7)	73.9 (+0.6)	73.8 (+0.5)	73.7 (+0.4)	73.9 (+0.6)	73.9 (+0.6)	73.9 (+0.6)
qwen-2-72b	25.1	69.2 (-0.5)	69.7	69.7 (±0.0)	39.5 (-30.2)	70.0 (+0.3)	69.9 (+0.2)	70.0 (+0.3)	69.9 (+0.2)	69.9 (+0.2)	69.9 (+0.2)
qwen-1.5-110b	25.0	71.3 (-1.5)	72.8	73.0 (+0.2)	46.3 (-26.5)	72.9 (+0.1)	72.9 (+0.1)	72.7 (-0.1)	72.9 (+0.1)	72.9 (+0.1)	72.9 (+0.1)
gpt-3.5	24.9	63.0 (+2.2)	60.8	**64.7 (+3.9)**	36.9 (-23.9)	**65.9 (+5.1)**	**65.5 (+4.7)**	**65.6 (+4.8)**	**65.6 (+4.8)**	**65.6 (+4.8)**	**65.6 (+4.8)**
gpt-4	25.0	**80.7 (+5.1)**	75.6	**82.1 (+6.5)**	**70.9 (-4.7)**	**82.5 (+6.9)**	**81.9 (+6.3)**	**81.9 (+6.3)**	**81.9 (+6.3)**	**81.9 (+6.3)**	**81.9 (+6.3)**
MMLU-Pro											
mistral-7b	9.6	20.9 (-9.0)	29.9	27.7 (-2.2)	15.6 (-14.3)	31.7 (+1.8)	30.7 (+0.8)	31.4 (+1.5)	31.2 (+1.3)	31.7 (+1.8)	31.7 (+1.8)
llama-3-8b	9.7	18.9 (-2.4)*	21.3	**23.8 (+2.5)**	**19.3 (-2.0)**	22.2 (+0.9)	**23.8 (+2.5)**	**24.5 (+3.2)**	22.6 (+1.3)	23.0 (+1.7)	23.4 (+2.1)
glm-4-9b	9.6	26.2 (-5.7)*	31.9	28.2 (-3.7)	**23.9 (-8.0)**	**36.4 (+4.5)**	**35.9 (+4.0)**	**34.8 (+2.9)**	**36.7 (+4.8)**	**35.6 (+3.7)**	**36.2 (+4.3)**
llama-3-70b	10.3	**46.7 (+3.5)**	43.2	44.6 (+1.4)	24.6 (-18.6)	42.8 (-0.4)	43.5 (+0.3)	43.6 (+0.4)	43.0 (-0.2)	43.2 (±0.0)	43.5 (+0.3)
qwen-2-72b	10.4	35.1 (-1.7)	36.8	37.4 (+0.6)	19.5 (-17.3)	37.2 (+0.4)	36.7 (-0.1)	36.7 (-0.1)	37.2 (+0.4)	37.3 (+0.5)	37.2 (+0.4)
qwen-1.5-110b	10.1	45.7 (+0.9)	44.8	42.8 (-2.0)	16.6 (-28.2)	44.7 (-0.4)	44.9 (+0.1)	44.6 (-0.2)	45.1 (+0.3)	45.0 (+0.2)	44.8 (±0.0)
gpt-3.5	9.9	**28.5 (+2.6)**	25.9	27.1 (+1.2)	13.0 (-12.9)	26.5 (+0.6)	27.0 (+1.1)	**28.5 (+2.6)**	26.5 (+0.6)	26.7 (+0.8)	27.2 (+1.3)
gpt-4	9.9	46.4 (-0.5)	46.9	46.9 (±0.0)	34.6 (-12.3)	47.3 (+0.4)	47.5 (+0.6)	47.7 (+0.8)	47.5 (+0.6)	47.8 (+0.9)	47.7 (+0.8)
ARC-Challenge											
mistral-7b	24.9	53.1 (-17.9)	71.0	70.3 (-0.7)	47.7 (-23.3)	71.7 (+0.7)	71.7 (+0.7)	71.6 (+0.6)	71.7 (+0.7)	71.7 (+0.7)	71.6 (+0.6)
llama-3-8b	25.2	44.4 (-21.8)	66.2	52.8 (-13.4)	41.1 (-25.1)	**71.3 (+5.1)**	**70.0 (+3.8)**	**70.0 (+3.8)**	**71.6 (+5.4)**	**71.3 (+5.1)**	**71.3 (+5.1)**
glm-4-9b	24.8	69.9 (-9.7)*	79.3	80.1 (+0.8)	65.1 (-14.2)	**82.7 (+3.4)**	**82.3 (+3.0)**	**82.0 (+2.7)**	**82.8 (+3.5)**	**83.0 (+3.7)**	**82.7 (+3.4)**
llama-3-70b	25.3	88.9 (+1.1)	87.8	87.9 (+0.1)	**80.8 (-7.0)**	88.5 (+0.7)	88.4 (+0.6)	88.1 (+0.3)	88.5 (+0.7)	88.4 (+0.6)	88.4 (+0.6)
qwen-2-72b	24.8	84.7 (-1.1)	85.8	86.0 (+0.2)	36.7 (-49.1)	86.3 (+0.5)	86.2 (+1.3)	85.8 (±0.0)	86.3 (+0.5)	86.3 (+0.5)	86.2 (+0.4)
qwen-1.5-110b	24.7	87.0 (-0.7)	87.7	88.3 (+0.6)	53.4 (-34.3)	88.1 (+0.4)	88.1 (+0.4)	88.0 (+0.3)	88.1 (+0.4)	88.1 (+0.4)	88.1 (+0.4)
gpt-3.5	25.2	78.1 (+1.2)	76.9	77.0 (+0.1)	29.9 (-47.0)	78.2 (+1.3)	77.9 (+1.0)	78.2 (+1.3)	78.1 (+1.2)	77.9 (+1.0)	77.9 (+1.0)
gpt-4	25.0	92.9 (+0.4)	92.5	92.8 (+0.3)	**87.3 (-5.2)**	92.9 (+0.4)	92.7 (+0.2)	92.8 (+0.3)	92.8 (+0.3)	92.8 (+0.3)	92.9 (+0.4)

Table 2: Overall accuracy results on MMLU, MMLU-Pro and ARC-Challenge benchmarks. 'Rand.' and 'Dicta.' denote 'random' and 'dictatorial', respectively. The numbers in parentheses are relative to the *blind dictatorial* baselines. Performance gains are marked in red, and loss in blue. Notable cases are marked in **bold**. *Results marked with asterisk are calculated utilizing partial profiles (see Appendix C).

Robustness against Unreliable Agents

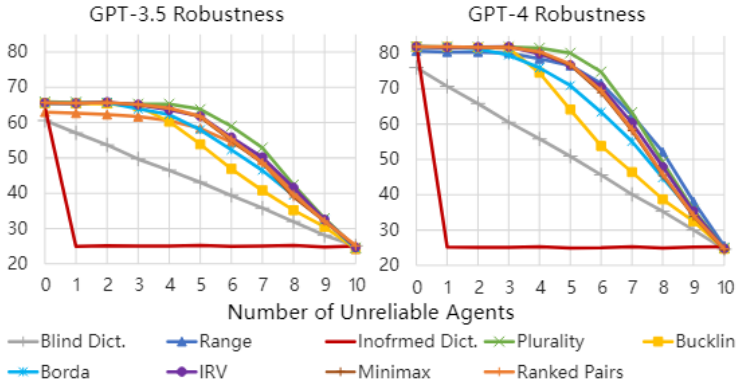


Figure 4: Accuracy impact of increasing number of unreliable agents built on `gpt-3.5` and `gpt-4`.

Limitation:

- MCQA is a limited scenario of CDM (preference over correctness)
- Limited CDM methods in GEDI, no compound of multiple voting strategies
- Voting Tax: computation cost of inter-agent communication is high

Major Literature

- John, Searle. 1980. Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3: 417-457.
- Stevan, Harnad. 1990. The Symbol Grounding Problem. *Physica D: Nonlinear Phenomena*, 42(1-3): 335–346.
- Rodney A. Brooks. 1991. Intelligence without Representation. *Artificial Intelligence,* 47:139- 159.
- David J. Chalmers. 1992. Subsymbolic Computation and the Chinese Room. In *Dinsmore, J. (ed.) The Symbolic and Connectionist Paradigms: Closing the Gap*. Hillsdale, NJ: Lawrence Erlbaum.
- Georg Dorffner and Erich Prem. 1993. Connectionism, Symbol Grounding, and Autonomous Agents. In P*roceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, pages 144-148. Hillsdale, NJ: Lawrence Erlbaum.
- Richard Sutton, 2024, New Path for AI: Approximately Correct Podcast, URL: https://www.youtube.com/watch?v=NvfK1TkXmOQ
- Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, et al. Agentbench: Evaluating llms as agents. ArXiv preprint, abs/2308.03688, 2023. URL https://arxiv.org/abs/2308.03688.
- Aohan Zeng, Mingdao Liu, Rui Lu, Bowen Wang, Xiao Liu, Yuxiao Dong, and Jie Tang. 2024. AgentTuning: Enabling Generalized Agent Abilities for LLMs. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3053–3077, Bangkok, Thailand. Association for Computational Linguistics.
- Yifan Song, Da Yin, Xiang Yue, Jie Huang, Sujian Li, and Bill Yuchen Lin. 2024. Trial and Error: Exploration-Based Trajectory Optimization of LLM Agents. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7584–7600, Bangkok, Thailand. Association for Computational Linguistics.
- Yifan Song, Weimin Xiong, Xiutian Zhao, Dawei Zhu, Wenhao Wu, Ke Wang, Cheng Li, Wei Peng, and Sujian Li. 2024. AgentBank: Towards Generalized LLM Agents via Fine-Tuning on 50000+ Interaction Trajectories. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 2124–2141, Miami, Florida, USA. Association for Computational Linguistics.
- Weimin Xiong, Yifan Song, Xiutian Zhao, Wenhao Wu, Xun Wang, Ke Wang, Cheng Li, Wei Peng, and Sujian Li. 2024. Watch Every Step! LLM Agent Learning via Iterative Step-level Process Refinement. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1556–1572, Miami, Florida, USA. Association for Computational Linguistics.
- Xiutian Zhao, Ke Wang, and Wei Peng. 2024. An Electoral Approach to Diversify LLM-based Multi-Agent Collective Decision-Making. *In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 2712–2727, Miami, Florida, USA. Association for Computational Linguistics.

Thanks

Questions & discussion welcome.

Acknowledgement of Country

RMIT University acknowledges the people of the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation on whose unceded lands we conduct the business of the University.

RMIT University respectfully acknowledges their Ancestors and Elders, past and present.

RMIT also acknowledges the Traditional Custodians and their Ancestors of the lands and waters across Australia where we conduct our business.

Artwork ‘Sentient’ by Hollie Johnson
Hollie is a Gunaikurnai and Monero Ngarigo woman from Gippsland who graduated from RMIT with a BA in Photography in 2016.