Multi-Agents Conversations

The Next Frontier of AI Programming

Multi-Agents Conversations

The Next Frontier of AI Programming

5 October 2024 - BiDigital Biella

3155v2 [cs.AI] 3 Oct 2023

AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation

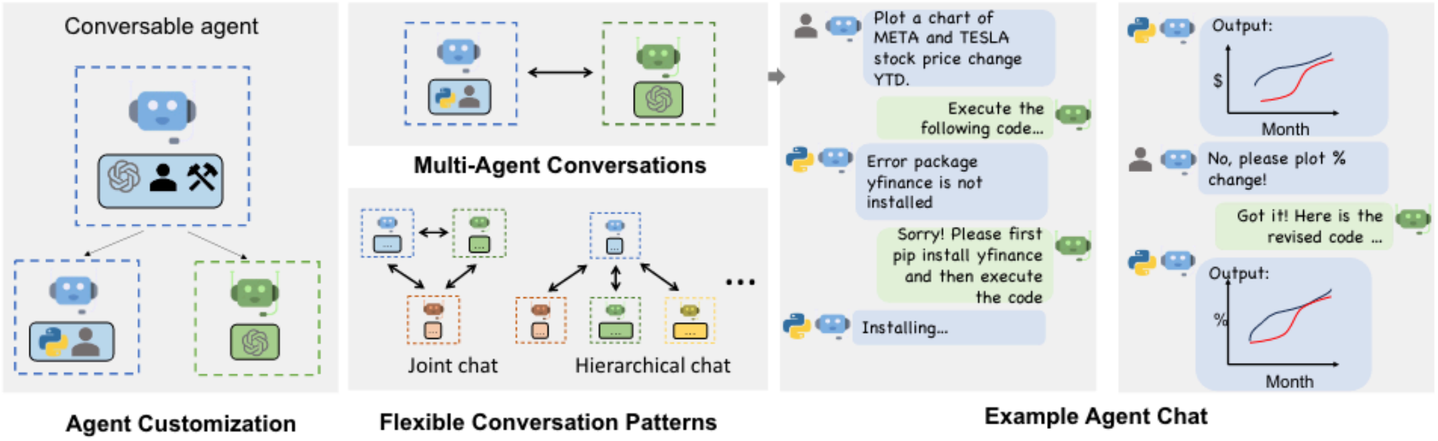
Qingyun Wu[†], Gagan Bansal[*], Jieyu Zhang[±], Yiran Wu[†], Beibin Li[*]

Erkang Zhu[*], Li Jiang[*], Xiaoyun Zhang[*], Shaokun Zhang[†], Jiale Liu[∓]

Ahmed Awadallah[*], Ryen W. White[*], Doug Burger[*], Chi Wang[*1]

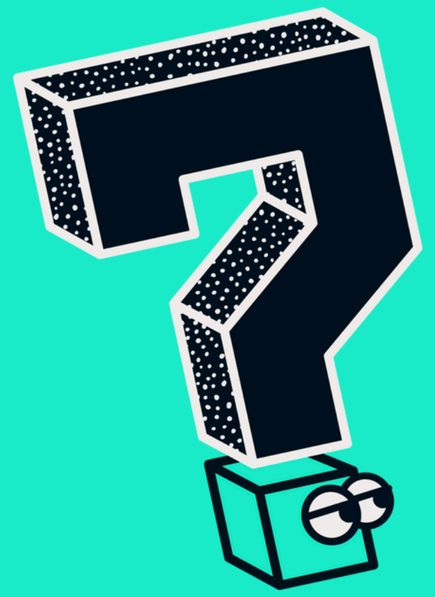
[*]Microsoft Research, [†]Pennsylvania State University

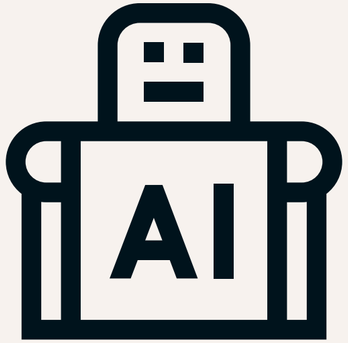
[±]University of Washington, [∓]Xidian University



AutoGen is an open-source framework that allows developers to build LLM applications via **multiple agents** that can **converse** with each other to accomplish tasks.

What is an AI Agent?

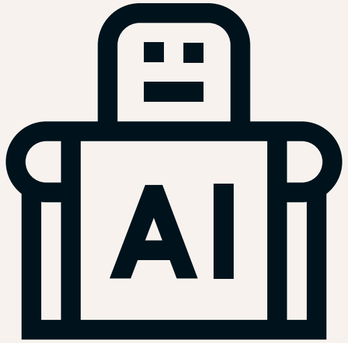




Planning

Memory

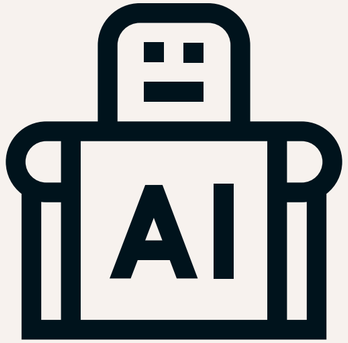
Tools



Planning

Memory

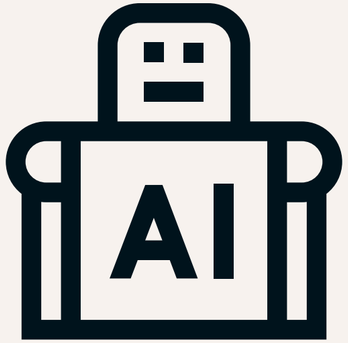
Tools



Planning

Memory

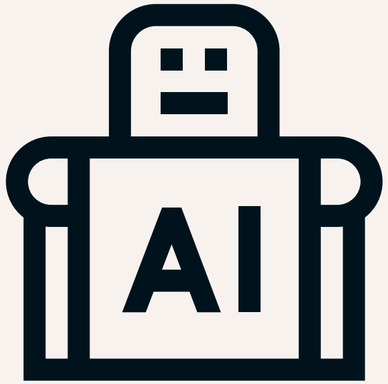
Tools



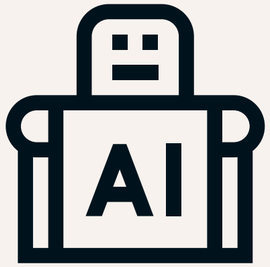
Planning

Memory

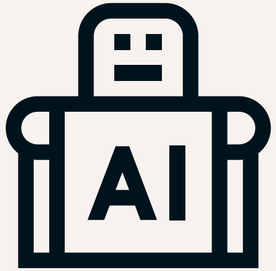
Tools



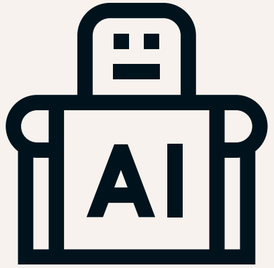
Group Chat Manager



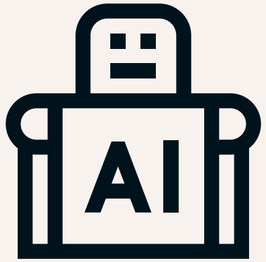
Coder



Software Engineer (Critic)



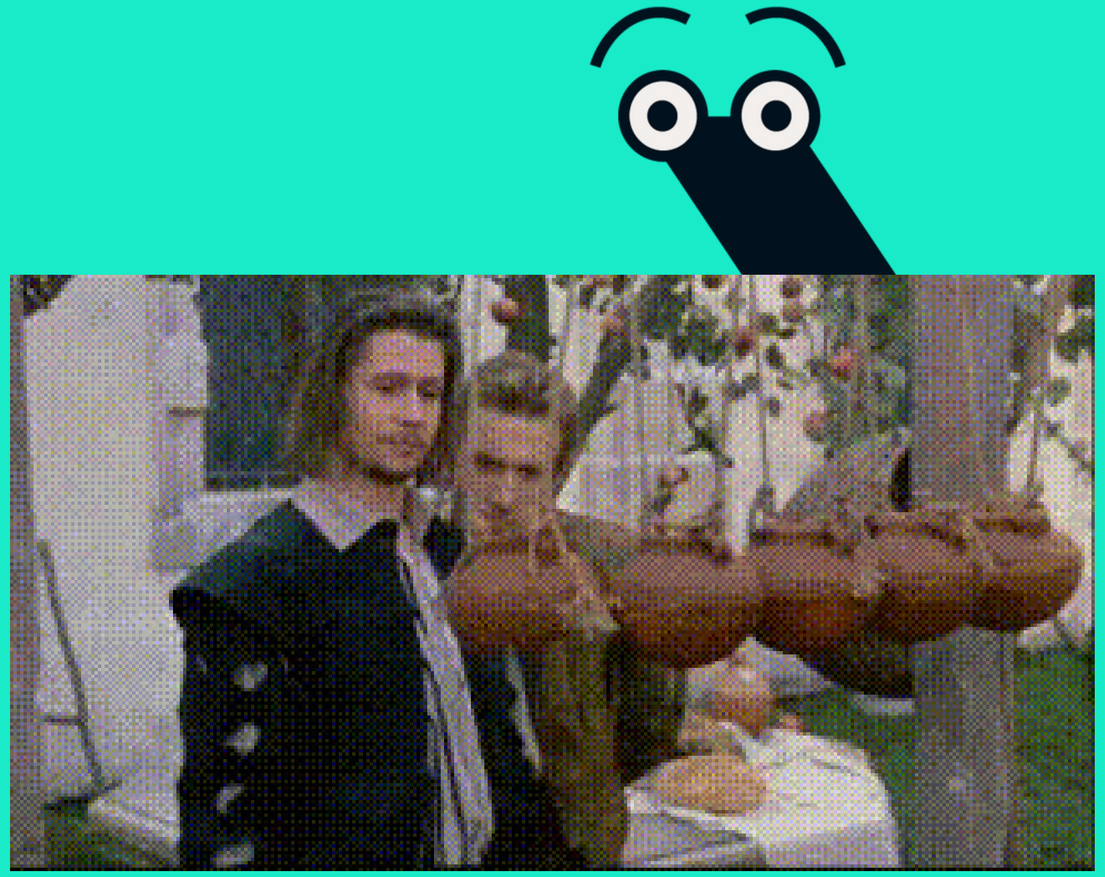
DataViz Expert (Critic)

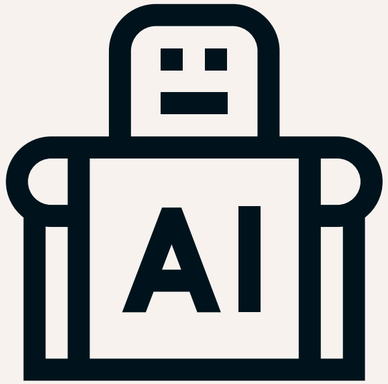


User Proxy (execude code)

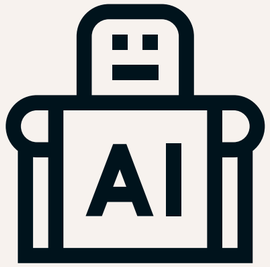
(Dangerous)

Demo Time

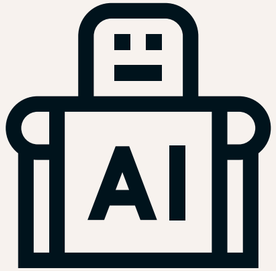




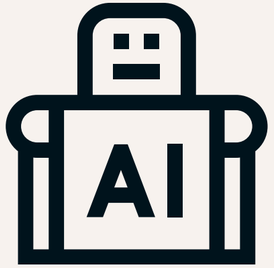
Group Chat Manager



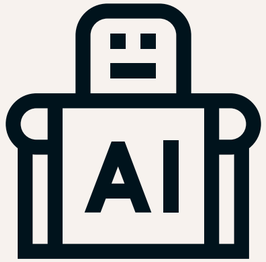
Coder



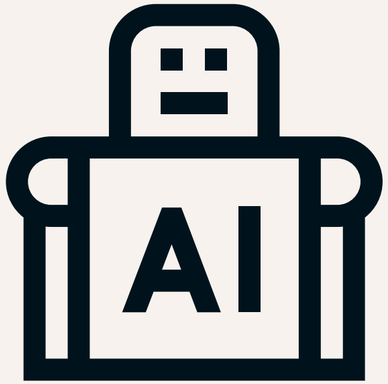
Software Engineer (Critic)



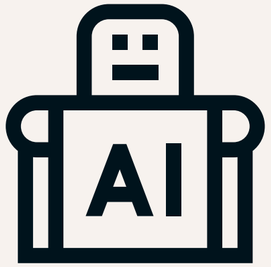
DataViz Expert (Critic)



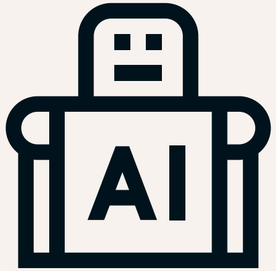
User Proxy (execude code)



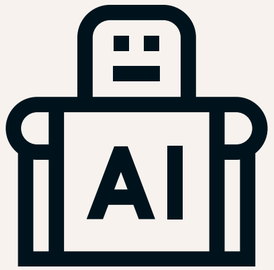
Group Chat Manager



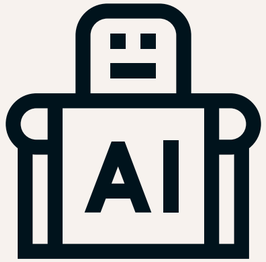
Coder



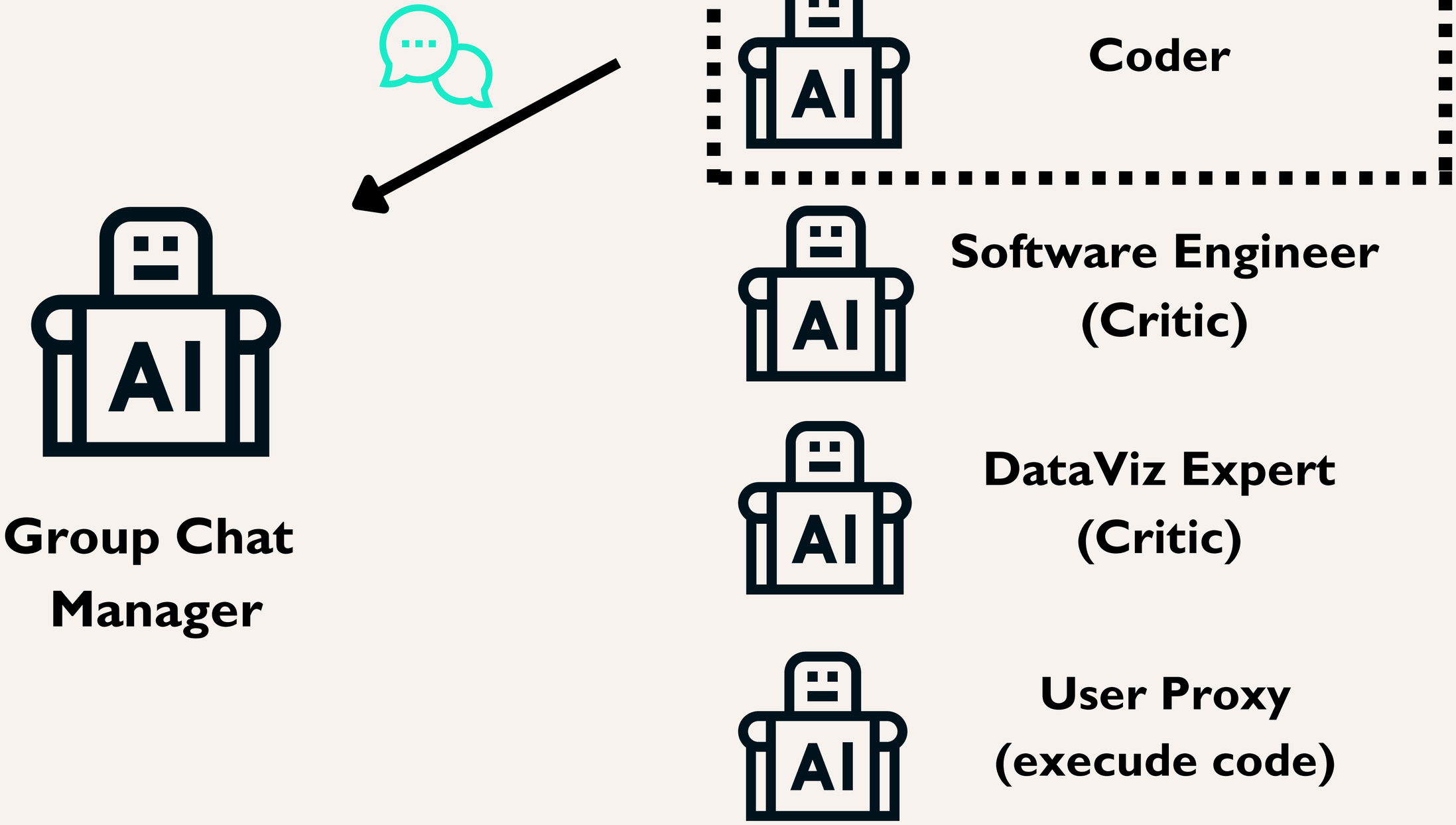
Software Engineer (Critic)

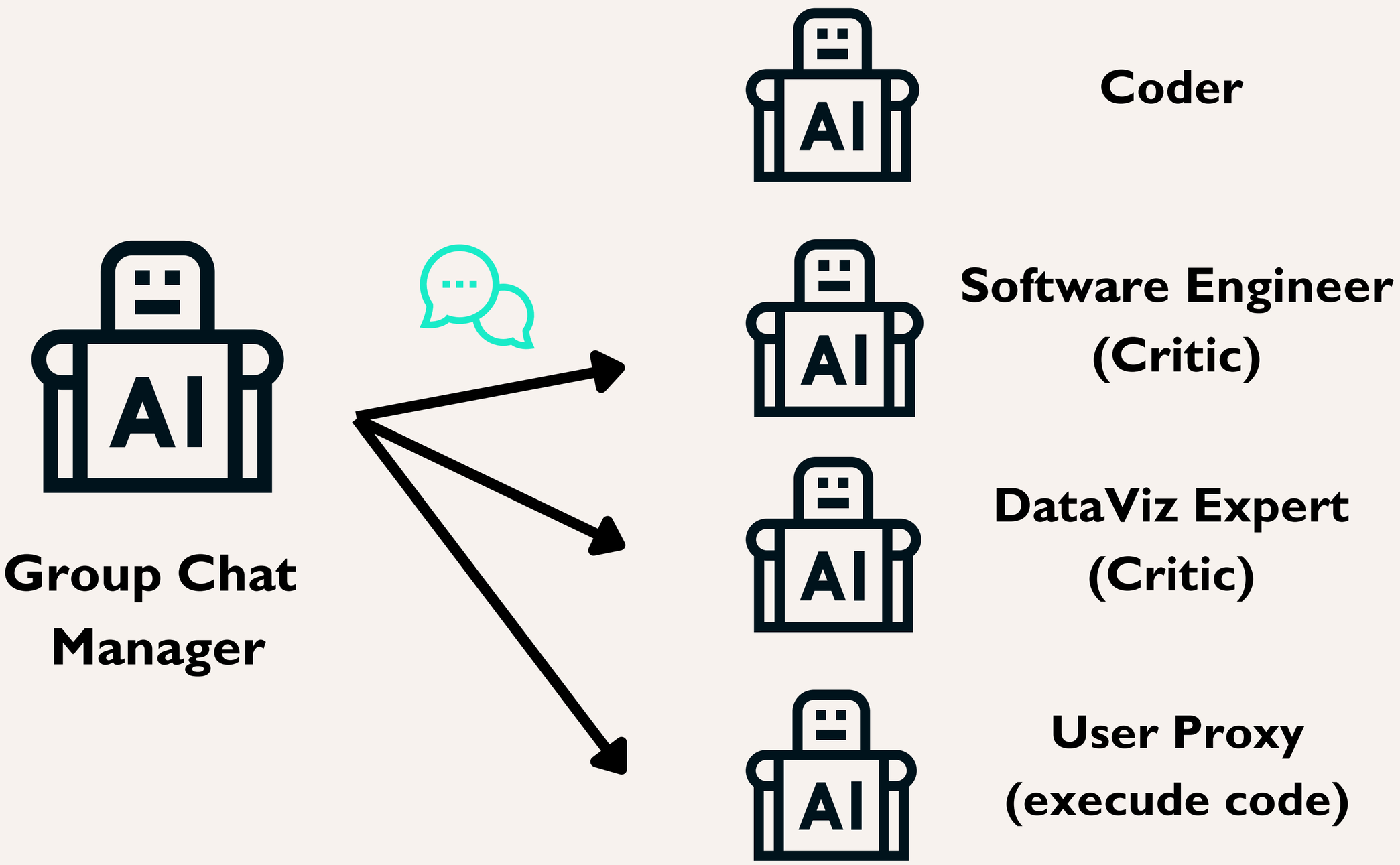


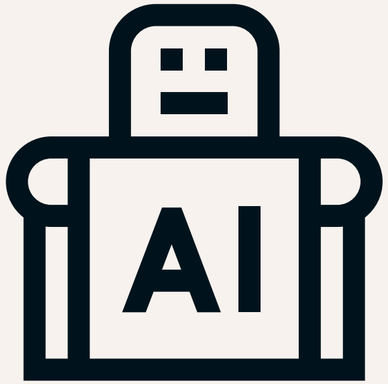
DataViz Expert (Critic)



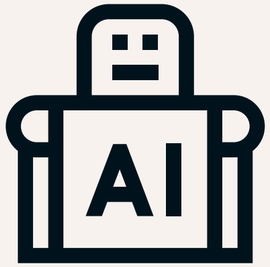
User Proxy (execude code)


Coder
Software Engineer
(Critic)
DataViz Expert
(Critic)
User Proxy
(execude code)
Group Chat
Manager

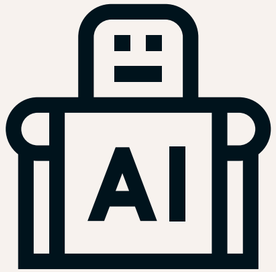

AI
Group Chat Manager
AI
Coder
AI
Software Engineer (Critic)
AI
DataViz Expert (Critic)
AI
User Proxy (execude code)



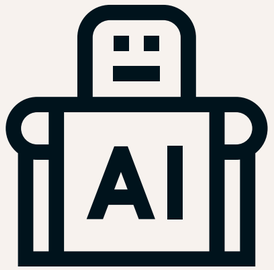
Group Chat Manager



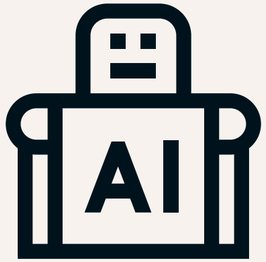
Coder



Software Engineer (Critic)

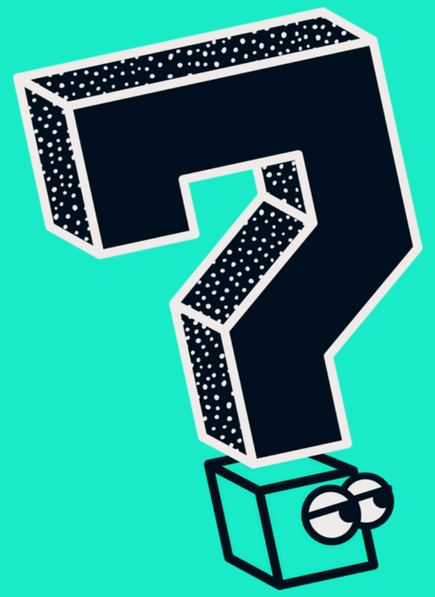


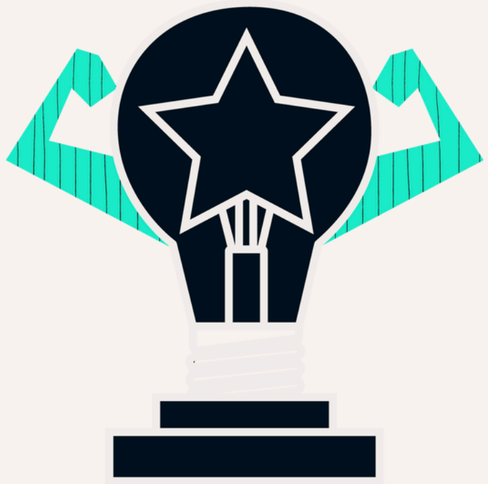
DataViz Expert (Critic)



User Proxy (execude code)

Why Multi-Agents?





Empirical studies demonstrate the **effectiveness** of the framework in many example applications

:2312.13010v3 [cs.CL] 24 May 2024

AgentCoder: Multi-Agent Code Generation with Effective Testing and Self-optimisation

Dong Huang
University of Hong Kong
dhuang@cs.hku.hk

Jie M.Zhang
jie.zhang@kcl.ac.uk
jie.zhang@kcl.ac.uk

Michael Luck
University of Sussex
Michael.Luck@sussex.ac.uk

Qingwen BU
Shanghai Jiao Tong University
qwbu01@sjtu.edu.cn

Yuhao Qing
University of Hong Kong
yhqing@cs.hku.hk

Heming Cui
University of Hong Kong
heming@cs.hku.hk

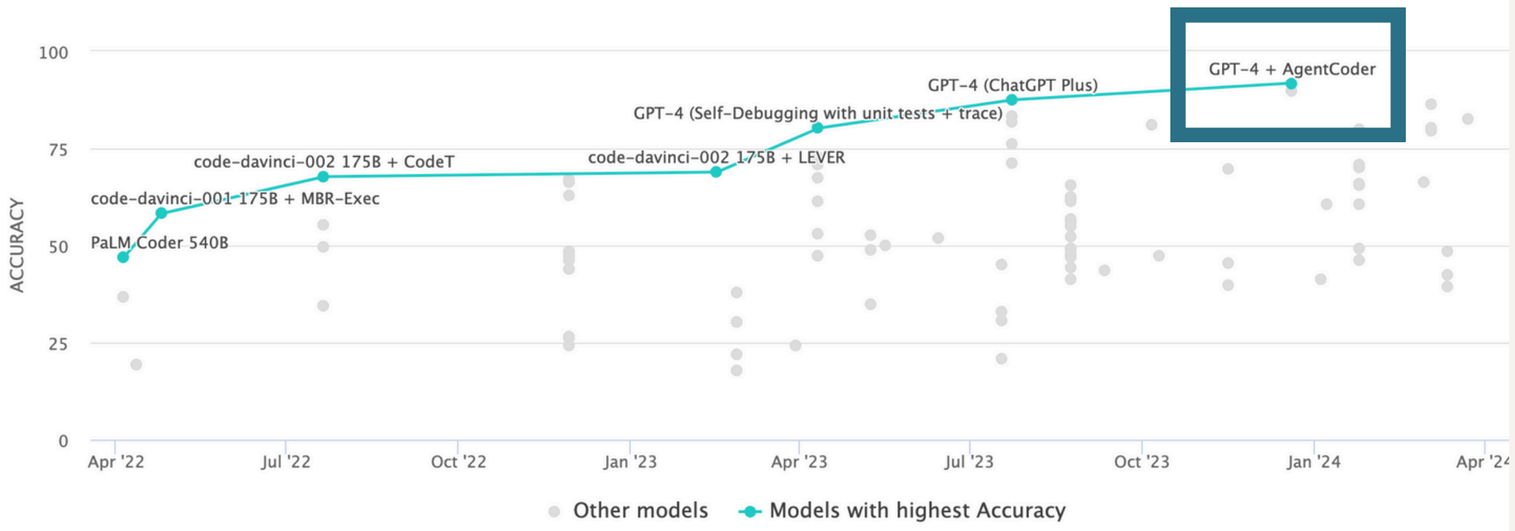
Abstract

Advances in natural language processing (NLP) have been significantly boosted by the development of transformer-based large language models (LLMs). These models have revolutionized NLP tasks, particularly in code generation, aiding developers in creating software with enhanced efficiency. Despite their advances, challenges remain in balancing code snippet generation with effective test cases. To address these issues, this paper introduces AgentCoder, a novel code generation solution comprising a multi-agent framework with a specialized test designer agents in addition to the programmer agent and the test executor agent. During the coding procedure, the test designer agent generates effective test cases for the generated code, and the test executor agent runs the code with the test cases and writes feedback to the programmer agent for it to refine the code. This collaborative system enhances code generation efficiency with less cost, outperforming both single-agent models and earlier multi-agent strategies, demonstrated by our extensive experiments on 14 LLMs and 16 baseline approaches. For example, AgentCoder (GPT-4) achieves 96.3% and 91.8% pass@1 in HumanEval and MBPP datasets with an overall token overhead of 56.9K and 66.3K, while state-of-the-art obtains only 90.2% and 78.9% pass@1 with an overall token overhead of 138.2K and 206.5K.

Code Generation on MBPP

Leaderboard Dataset

View Accuracy by



Filter: checked untagged

Edit L

Rank	Model	Accuracy	Paper	Code	Result	Year
1	**GPT-4 + AgentCoder**	91.8	AgentCoder: Multi-Agent-based Code Generation with Iterative Testing and Optimisation			2023
2	**GPT-3.5 Turbo** (ChatGPT)	89.9	AgentCoder: Multi-Agent-based Code Generation with Iterative Testing and Optimisation			2023
3	**GPT-4** (ChatGPT Plus)	87.5	How Does Naming Affect LLMs on Code Analysis Tasks?			2023
4	**Claude 3 Opus**	86.4	The Claude 3 Model Family: Opus, Sonnet, Haiku			2024

Austin, Jacob, et al. **"Program synthesis with large language models."** arXiv preprint arXiv:2108.07732 (2021).

1	**LDB** (O1-mini, based on seed programs from Reflexion)	99.39	×	Debug like a Human: A Large Language Model Debugger via Verifying Runtime Execution Step-by-step		2024	
2	**LDB** (O1-preview, based on seed programs from Reflexion)	99.39	×	Debug like a Human: A Large Language Model Debugger via Verifying Runtime Execution Step-by-step		2024	
3	**LDB** (GPT-4o, based on seed programs from Reflexion)	98.2	×	Debug like a Human: A Large Language Model Debugger via Verifying Runtime Execution Step-by-step		2024	
4	**LDB** (GPT-4, based on seed programs from Reflexion)	96.9	×	Debug like a Human: A Large Language Model Debugger via Verifying Runtime Execution Step-by-step		2024	
5	**AgentCoder** (GPT-4)	96.3	×	AgentCoder: Multi-Agent-based Code Generation with Iterative Testing and Optimisation		2023	
6	**LDB** (GPT-3.5, based on seed programs from Reflexion)	95.1	×	Debug like a Human: A Large Language Model Debugger via Verifying Runtime Execution Step-by-step		2024	
7	**MapCoder** (GPT-4)	93.9	×	MapCoder: Multi-Agent Code Generation for Competitive Problem Solving		2024	
8	**Language Agent Tree Search** (GPT-4)	92.7	×	Language Agent Tree Search Unifies Reasoning Acting and Planning in Language Models		2023	
9	**Claude 3.5 Sonnet** (0-shot)	92.0	×	Claude 3.5 Sonnet Model Card Addendum		2024	myth
10	**FractalResearch : Pioneer-SWO** (GPT-4T)	91.6	×	AgentCoder: Multi-Agent-based Code Generation with Iterative Testing and Optimisation		2023	

Chen, Mark, et al. **"Evaluating large language models trained on code."** arXiv preprint arXiv:2107.03374 (2021).

07.07924v4 [cs.SE] 19 Dec 2023

Communicative Agents for Software Development

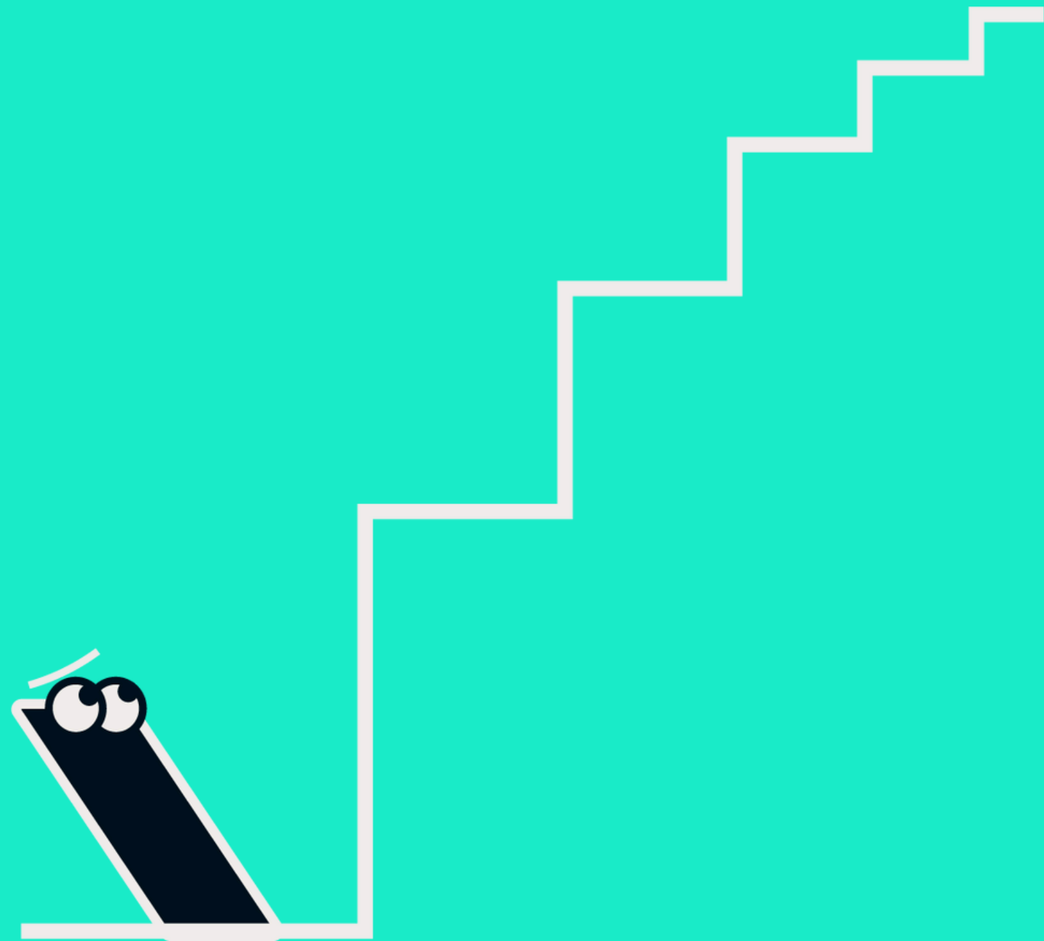
Chen Qian[♠] Xin Cong[♠] Wei Liu[♠] Cheng Yang[♣] Weize Chen[♠] Yusheng Su[♠]
Yufan Dang[♠] Jiahao Li[♦] Juyuan Xu[▲] Dahai Li[★] Zhiyuan Liu[♠✉] Maosong Sun[♠✉]
[♠]Tsinghua University [♣]Beijing University of Posts and Telecommunications
[♦]Dalian University of Technology [▲]Brown University [★]Modelbest Inc.
qianc62@gmail.com liuzy@tsinghua.edu.cn sms@tsinghua.edu.cn

Motivation	Research Domain & Goals		Work	Agents-Env. Interface	Agents Profiling		Agents Communication		Agents Capabilities Acquisition	
					Profiling methods	Profiles (examples)	Paradigms	Structure	Feedback from	Agents Adjustment
Problem Solving	Software development		[Qian *et al.*, 2023]	Sandbox	Pre-defined, Model-Generated	CTO, programmer	Cooperative	Layered	Environment, Agent interaction, Human	Memory, Self-Evolution
			[Hong *et al.*, 2023]	Sandbox	Pre-defined	Product Manager, Engineer	Cooperative	Layered, Shared Message Pool	Environment, Agent interaction, Human	Memory, Self-Evolution
			[Dong *et al.*, 2023b]	Sandbox	Pre-defined, Model-Generated	Analyst, coder	Cooperative	Layered	Environment, Agent interaction	Memory, Self-Evolution
	Embodied Agents	Multi-robot planning	[Chen *et al.*, 2023d]	Sandbox, Physical	Pre-defined	Robots	Cooperative	Centralized, Decentralized	Environment, Agent interaction	Memory
		Multi-robot collaboration	[Mandi *et al.*, 2023]	Sandbox, Physical	Pre-defined	Robots	Cooperative	Decentralized	Environment, Agent interaction	Memory
		Multi-Agents cooperation	[Zhang *et al.*, 2023c]	Sandbox	Pre-defined	Robots	Cooperative	Decentralized	Environment, Agent interaction	Memory
	Science Experiments	Optimization of MOF	[Zheng *et al.*, 2023]	Physical	Pre-defined	Strategy planers, literature collector, coder	Cooperative	Centralized	Environment, Human	Memory
	Science Debate	Improving Factuality	[Du *et al.*, 2023]	None	Pre-defined	Agents	Debate	Decentralized	Agent interaction	Memory
		Examining, Inter-Consistency	[Xiong *et al.*, 2023]	None	Pre-defined	Proponent, Opponent, Judge	Debate	Centralized, Decentralized	Agent interaction	Memory
		Evaluators for debates	[Chan *et al.*, 2023]	None	Pre-defined	Agents	Debate	Centralized, Decentralized	Agent interaction	Memory
		Multi-Agents for Medication	[Tang *et al.*, 2023]	None	Pre-defined	Cardiology, Surgery	Debate, Cooperative	Centralized, Decentralized	Agent interaction	Memory
World Simulation	Society	Modest Community (25 persons)	[Park *et al.*, 2023]	Sandbox	Model-generated	Pharmacy, shopkeeper	-	-	Environment, Agent interaction	Memory
		Online community (1000 persons)	[Park *et al.*, 2022]	None	Pre-defined, Model-generated	Camping, fishing	-	-	Agent interaction	Dynamic Generation
		Emotion propagation	[Gao *et al.*, 2023a]	None	Pre-defined, Model-generated	Real-world user	-	-	Agent interaction	Memory
		Real-time social interactions	[Kaiya *et al.*, 2023]	Sandbox	Pre-defined	Real-world user	-	-	Environment, Agent interaction	Memory
		Opinion dynamics	[Li *et al.*, 2023a]	None	Pre-defined	NIN, NINL, NIL	-	-	Agent interaction	Memory
	Gaming	WereWolf	[Xu *et al.*, 2023b] [Xu *et al.*, 2023c]	Sandbox	Pre-defined	Seer, werewolf, villager	Cooperative, Debate, Competitive	Decentralized	Environment, Agent interaction	Memory
		Avalon	[Light *et al.*, 2023a] [Wang *et al.*, 2023c]	Sandbox	Pre-defined	Servant, Merlin, Assassin	Cooperative, Debate, Competitive	Decentralized	Environment, Agent interaction	Memory
		Welfare Diplomacy	[Mukobi *et al.*, 2023]	Sandbox	Pre-defined	Countries	Cooperative, Competitive	Decentralized	Environment, Agent interaction	Memory
	Psychology	Human behavior Simulation	[Aher *et al.*, 2023]	Sandbox	Pre-defined	Humans	-	-	Agent interaction	Memory
		Collaboration Exploring	[Zhang *et al.*, 2023d]	None	Pre-defined	Agents	Cooperative, Debate	Decentralized	Agent interaction	Memory
	Economy	Macroeconomic simulation	[Li *et al.*, 2023e]	None	Pre-defined, Model-generated	Labor	Cooperative	Decentralized	Agent interaction	Memory
		Information Marketplaces	[Anonymous, 2023]	Sandbox	Pre-defined, Data-Derived	Buyer	Cooperative, Competitive	Decentralized	Environment, Agent interaction	Memory
		Improving financial trading	[Li *et al.*, 2023g]	Physical	Pre-defined	Trader	Debate	Decentralized	Environment, Agent interaction	Memory
		Economic theories	[Zhao *et al.*, 2023]	Sandbox	Pre-defined, Model-Generated	Restaurant, Customer	Competitive	Decentralized	Environment, Agent interaction	Memory, Self-Evolution
	Recommender Systems	Simulating user behaviors	[Zhang *et al.*, 2023a]	Sandbox	Data-Derived	Users from MovieLens-1M	-	-	Environment	Memory
		Simulating user-item interactions	[Zhang *et al.*, 2023e]	Sandbox	Pre-defined, Data-Derived	User Agents Item Agents	Cooperative	Decentralized	Environment, Agent interaction	Memory
	Policy Making	Public Administration	[Xiao *et al.*, 2023]	None	Pre-defined	Residents	Cooperative	Decentralized	Agent interaction	Memory
		War Simulation	[Hua *et al.*, 2023]	None	Pre-defined	Countries	Competitive	Decentralized	Agent interaction	Memory
	Disease	Human Behaviors to epidemics	[Ghaffarzadegan *et al.*, 2023]	Sandbox	Pre-defined, Model-Generated	Conformity traits	Cooperative	Decentralized	Environment, Agent interaction	Memory
		Public health	[Williams *et al.*, 2023]	Sandbox	Pre-defined, Model-Generated	Adults aged 18 to 64	Cooperative	Decentralized	Environment, Agent interaction	Memory, Dynamic Generation

Guo, Taicheng, et al. **"Large language model based multi-agents: A survey of progress and challenges."** arXiv preprint arXiv:2402.01680 (2024).

Every agents could use a **specialised LLM**, tailored for the specific task

Challenges



Scaling up LLM Multi-Agents systems **is costly** both in terms of resources and communication

Hallucinations of an agent can have **cascade effects** on all other agents

Useful Resources

Tools

AutoGen Documentation

CrewAI Documentation

Papers

Wu, Qingyun, et al. **"Autogen: Enabling next-gen llm applications via multi-agent conversation framework."** arXiv preprint arXiv:2308.08155 (2023).

Guo, Taicheng, et al. **"Large language model based multi-agents: A survey of progress and challenges."** arXiv preprint arXiv:2402.01680 (2024).

Qian, Chen, et al. **"Communicative agents for software development."** arXiv preprint arXiv:2307.07924 (2023).

Who Am I?

Fabio Lipreri

AI Engineer @ xtream

Thank you

Sellalab / BTREES