SwarmAgentic: Towards Fully Automated Agentic System Generation via Swarm Intelligence

Yao Zhang[1,3], Chenyang Lin [2], Shijie Tang[1], Haokun Chen[1], Shijie Zhou[2], Yunpu Ma[1,3], Volker Tresp[1,3]

[1]LMU Munich [2]Technical University of Munich [3]Munich Center for Machine Learning

About Me

Open to research collaborations & job opportunities

TL;DR: SwarmAgentic is the **first fully automated framework** that **constructs agentic systems from scratch** and jointly **optimizes their functionality and collaboration** through **language-driven swarm intelligence**.

Challenges in Agentic System Generation

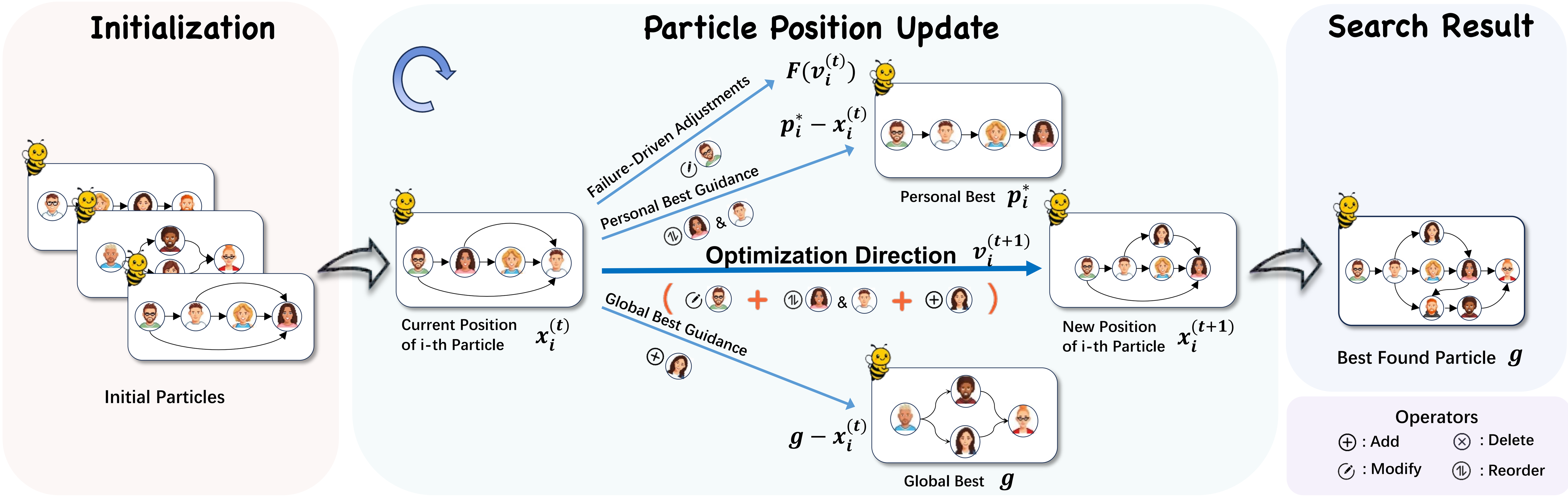
- **Template dependence.** Existing frameworks rely on seed agents or fixed templates, preventing from-scratch agentic system generation.
- **Partial optimization.** Existing frameworks improve either agent functionality or collaboration but lack joint structural optimization.
- **Lack of failure-aware optimization.** Existing systems cannot identify execution flaws or adapt their configurations based on failure feedback.
- **Discrete, non-differentiable search space.** The textual nature of agentic systems prohibits gradient updates, demanding interpretable, population-based exploration.

Key Contributions

- **Fully automated agentic system generation:** autonomously constructs complete agentic systems from scratch, requiring no predefined agents, templates, or human intervention.
- **Language-driven swarm optimization:** reformulates Particle Swarm Optimization (PSO) into a symbolic and interpretable optimization process that evolves both agent functionality and collaboration strategies.
- **Failure-aware self-optimization:** enables targeted, feedback-driven refinement through LLM-based flaw detection and iterative updates, allowing systems to adaptively correct errors and improve coordination.
- **Strong real-world performance:** achieves SOTA on six open-ended, multi-agent tasks, demonstrating robust generalization and scalability with only task descriptions and objective functions as input.

Overview of SwarmAgentic

SwarmAgentic reformulates Particle Swarm Optimization (PSO) in language space, where each particle $x_i^{(t)}$ represents a complete agentic system, including its functionality and collaboration strategies. The framework performs **textual transformations** instead of numeric updates, enabling interpretable, population-based search in discrete system space. p_i^* and g denote the personal-best and global-best systems that guide iterative refinement.



Optimization Process

- **Initialization:** Generate diverse candidate agentic systems from task descriptions using temperature-controlled sampling.
- **Flaw Identification:** LLM diagnoses agent- and structure-level deficiencies against the objective.
- **Failure-Aware Velocity Update:** Combine failure-driven, personal-best, and global-best guidance to generate targeted optimization directions.
- **Position Update:** Apply interpretable edits (Add / Delete / Modify / Reorder) to refine agents and collaborations.
- **Iteration:** Repeat updates until convergence; retain the best-performing system as the final solution.

Experimental Results

Performance on TravelPlanner

Method	Delivery Rate	Commonsense Micro	Commonsense Macro	Hard Constraint Micro	Hard Constraint Macro	Final
Direct	100.0 / 100.0	57.3 / 79.4	3.9 / 15.8	11.0 / 27.5	3.3 / 16.1	0.0 / 2.2
CoT (Wei et al., 2022)	100.0 / 100.0	61.0 / 76.7	2.8 / 11.7	10.0 / 22.4	3.3 / 12.8	0.0 / 2.2
Self-Refine (Madaan et al., 2024)	100.0 / 98.9	56.0 / 75.3	1.7 / 7.2	3.1 / 12.4	1.1 / 7.2	0.0 / 1.1
SPP (Wang et al., 2023)	99.4 / 96.7	54.6 / 70.6	1.7 / 5.6	3.8 / 11.4	1.1 / 7.8	0.0 / 0.6
EvoAgent (Yuan et al., 2024)	100.0 / 100.0	64.2 / 81.5	7.8 / 21.1	11.0 / 31.4	4.4 / 18.9	1.1 / 7.2
ADAS (Hu et al., 2024)	100.0 / 100.0	70.9 / 88.5	6.1 / 34.4	17.4 / 50.2	9.4 / 27.8	1.1 / 8.9
SwarmAgentic	100.0 / 100.0	**70.9 / 92.9**	**12.8 / 56.1**	**21.0 / 66.7**	**9.4 / 52.8**	**3.3 / 32.2**

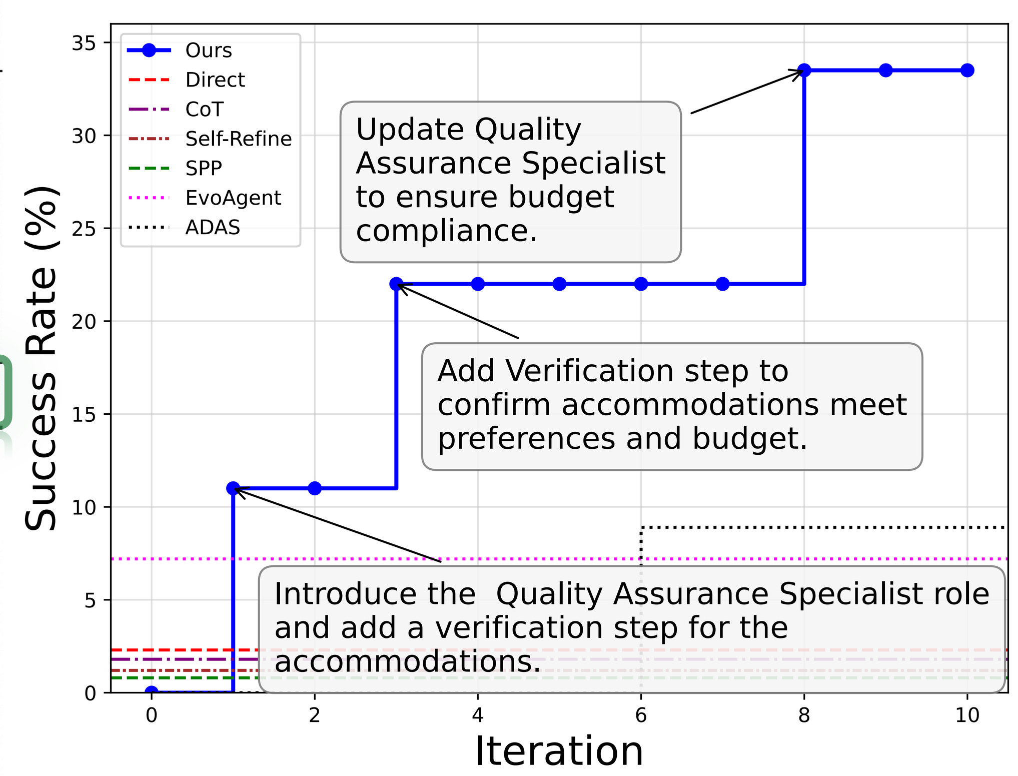
Performance on Reasoning and Planning Tasks

Method	Natural Plan (NP) Trip Planning	Natural Plan (NP) Meeting Planning	Natural Plan (NP) Calendar Scheduling	Creative Writing (CW)	MGSM
Direct	7.3 / 3.7	19.0 / 45.0	19.9 / 43.0	5.0 / 6.3	28.1 / 87.3
CoT (Wei et al., 2022)	9.0 / 1.0	19.0 / 50.0	20.0 / 60.0	5.3 / 7.0	28.7 / 81.0
Self-Refine (Madaan et al., 2024)	4.4 / 4.4	12.0 / 41.0	13.0 / 63.0	5.2 / 6.2	30.5 / 86.4
SPP (Wang et al., 2023)	5.0 / 1.3	4.0 / 33.0	22.0 / 44.0	5.9 / 7.6	55.2 / 84.9
EvoAgent (Yuan et al., 2024)	5.6 / 1.9	4.0 / 38.0	21.6 / 52.0	6.1 / 7.1	57.3 / 87.0
ADAS (Hu et al., 2024)	1.9 / 3.1	11.0 / 43.0	21.0 / 66.0	6.2 / 7.3	29.0 / 87.0
SwarmAgentic	**13.1 / 13.1**	**23.0 / 56.0**	**28.0 / 82.0**	**8.2 / 8.5**	**65.6 / 88.4**

Results Summary: SwarmAgentic achieves **+261.8%** on TravelPlanner and leads on NP/CW/MGSM, showing that **fully automated, language-driven optimization** generalizes across open-ended and structured tasks.

Case Study

Search Trajectory on TravelPlanner



SwarmAgentic iteratively refines agent roles and collaboration: QA specialist → **11%**, verification → **22%**, budget compliance → **33%**, surpassing all baselines and illustrating interpretable, **failure-aware evolution**.

Ablation Study

Methods	Score	Δ
Direct	6.2	0%
Different Iteration Count		
$\text{SwarmAgentic}_{(3,1)}$	5.9	-4.8%
$\text{SwarmAgentic}_{(3,5)}$	6.4	+3.2%
$\text{SwarmAgentic}_{(3,10)}$	7.0	+12.9%
Different Particle Count		
$\text{SwarmAgentic}_{(1,5)}$	6.3	+1.6%
$\text{SwarmAgentic}_{(3,5)}$	6.7	+8.1%
$\text{SwarmAgentic}_{(5,5)}$	6.9	+11.3%
Different Design Settings		
$\text{SwarmAgentic}_{(5,10)}$ *w/o Collab. Struc. Reconfig.*	6.7	+8.1%
$\text{SwarmAgentic}_{(5,10)}$ *w/o Agent-Level Adapt.*	7.3	+17.7%
$\text{SwarmAgentic}_{(5,10)}$ *w/o Failure-Driven Adjust.*	8.4	+35.5%
$\text{SwarmAgentic}_{(5,10)}$	8.8	+41.9%

- Removing any **key module** significantly **reduces performance**, confirming their necessity for self-optimization.
- Increasing **iterations** or **particle count** steadily **improves results**, validating the benefit of **iterative refinement and diverse exploration.**

Cross-Model Transferability

Method	GPT-4o	Claude-3.5-sonnet	DeepSeek-V3	Gemini-1.5	Gemini-1.5*
Direct	6.3	5.6	6.4	5.4	-
CoT (Wei et al., 2022)	7.0	5.7	5.9	5.8	-
Self-Refine (Madaan et al., 2024)	6.2	5.8	6.1	5.4	-
SPP(Wang et al., 2023)	7.6	8.0	8.3	7.1	-
EvoAgent(Yuan et al., 2024)	7.1	7.9	8.8	6.8	-
ADAS(Hu et al., 2024)	7.3	7.9	7.8	7.1	6.6
SwarmAgentic	**8.5**	**8.3**	**9.0**	**7.5**	**7.8**

- **Cross-model transfer:** Agentic systems discovered with GPT-4o transfer effectively to other LLMs, outperforming all baselines and showing strong generalization.
- **Model-specific adaptability:** Direct optimization on target models (e.g., Gemini-1.5*) yields additional gains, confirming both robust transfer and adaptive optimization.

arXiv Paper

Code

30th ANNIVERSARY EMNLP 2025 Suzhou, China November 4-9 中国苏州 11月4日-9日