# Multi-Agent Collaboration via Cross-Team Orchestration

Zhuoyun Du[1]*, Chen Qian[2]*, Wei Liu[3], Zihao Xie[4], Yifei Wang[4], Yufan Dang[4], Weize Chen[4], Cheng Yang[5], YeTian[6], Huantang Xiong[6], Lei Han[6]

[1]Zhejiang University & [2]Shanghai Jiao Tong University & [3]King's College London & [4]Tsinghua University & [5]BUPT & [6]Tencent Robotics X

ACL 2025 VIENNA JULY 27 - AUGUST 1
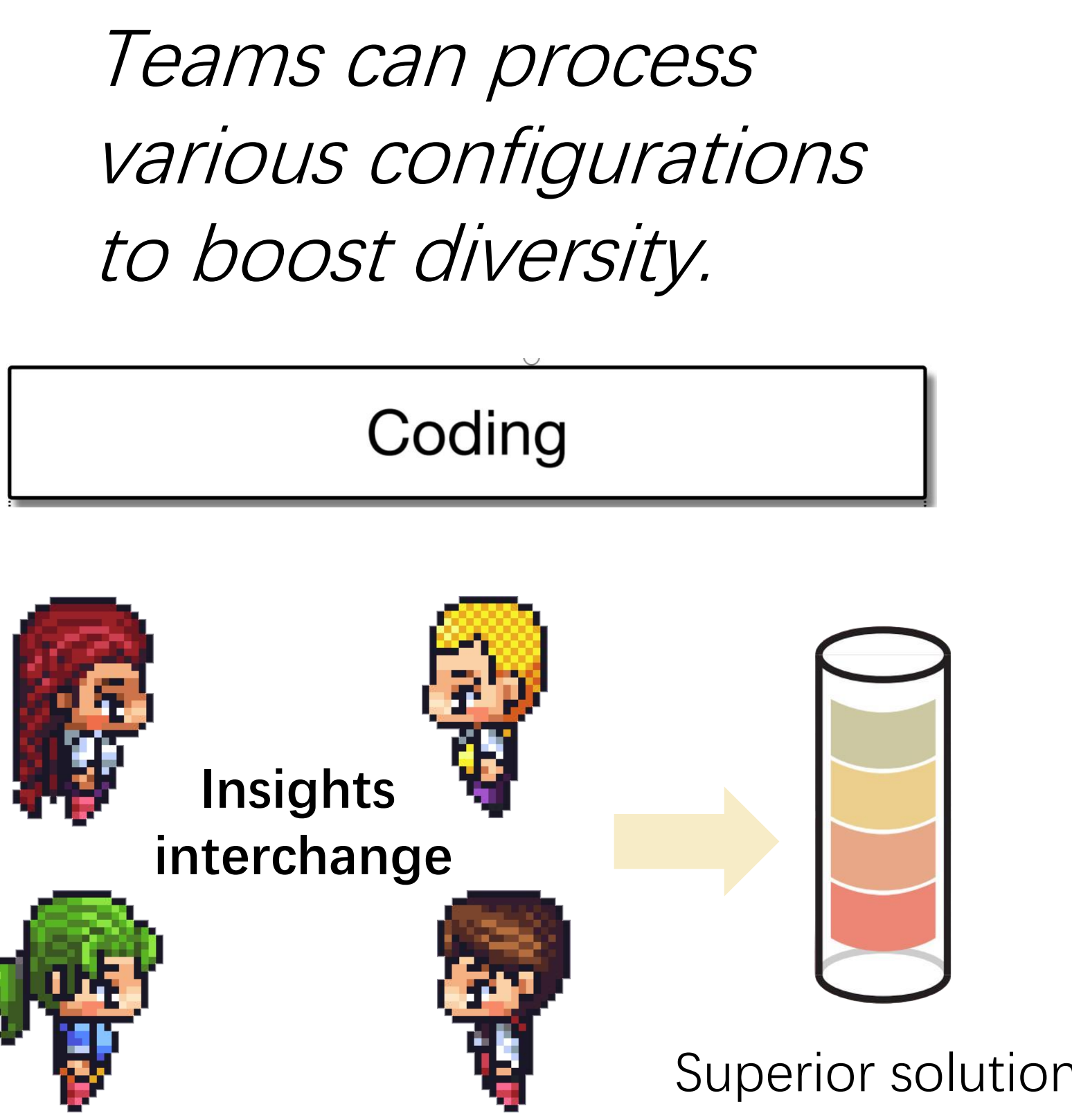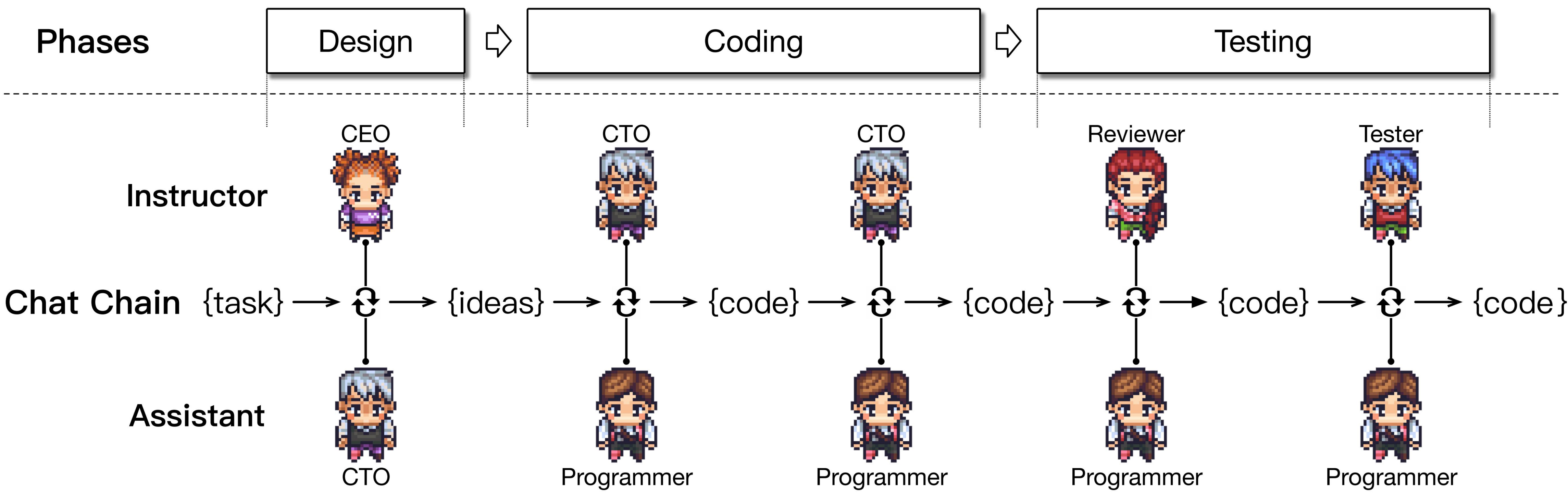
Legend: Agent, Content, Feature, Collaborate
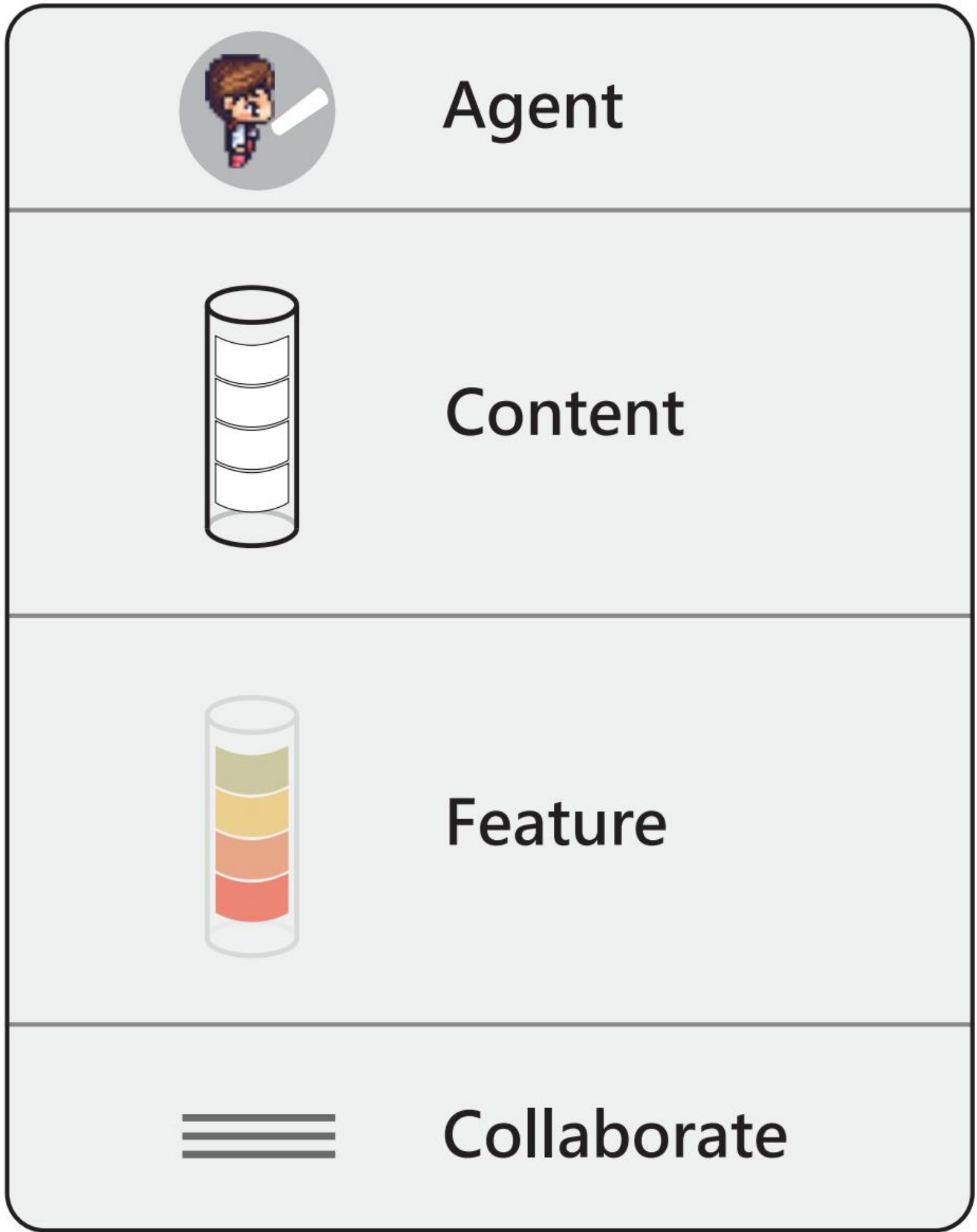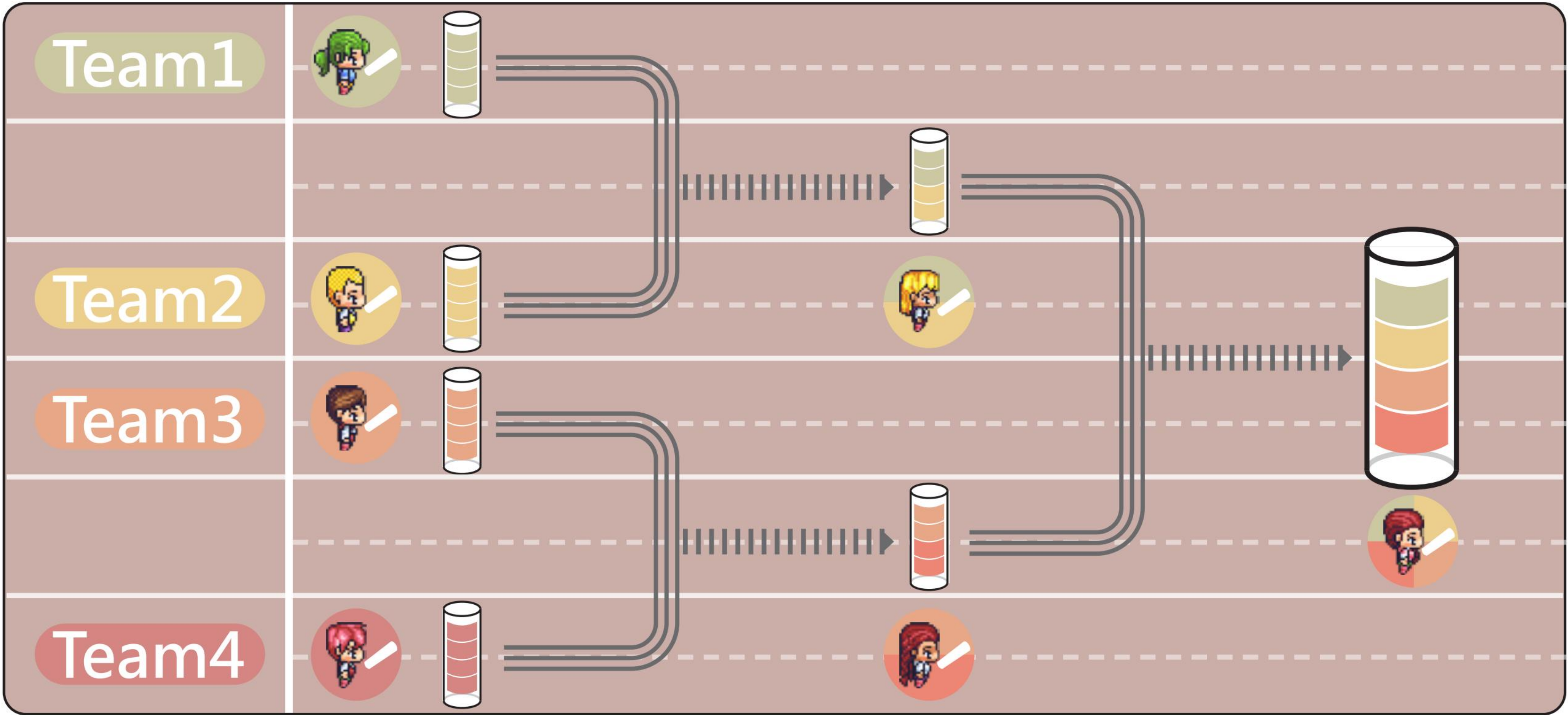
Motivation

Isolated workflows Current team like MASs can only execute phases sequentially according to its predefined config, which may lead to repetitive errors and lack of creativity.

Isolated workflows We enables different teams to independently propose various task-oriented solutions as insights and then communicate for insights.

Method

Our multi-agent cross team orchestration framework, know as Croto, is struct into three integral components: intra-team interactions, greedy aggregation and hierarchy partitioning (Shown in Figure above) that orchestrate teams to jointly propose various solutions and interact with their insights for superior solutions generation.
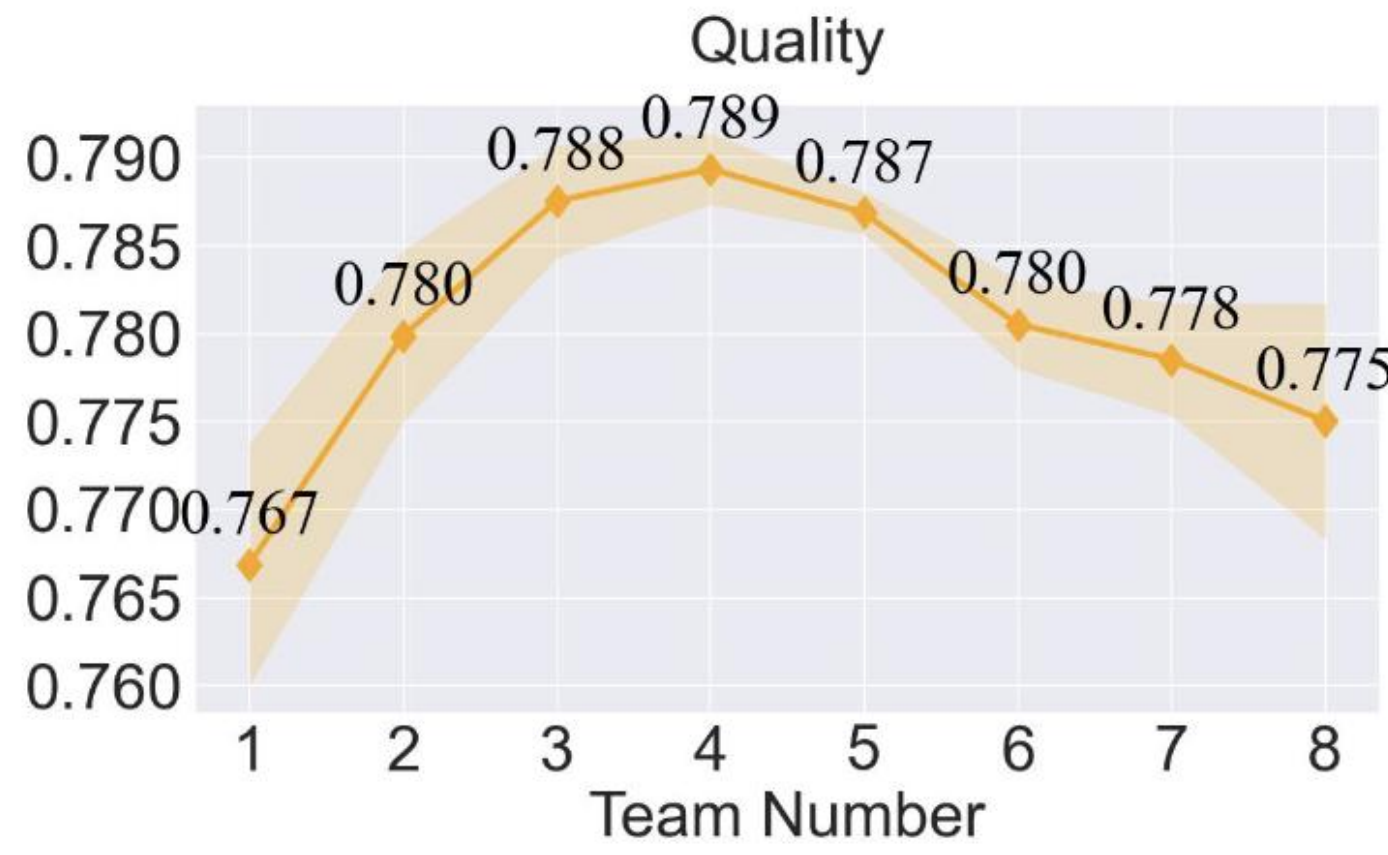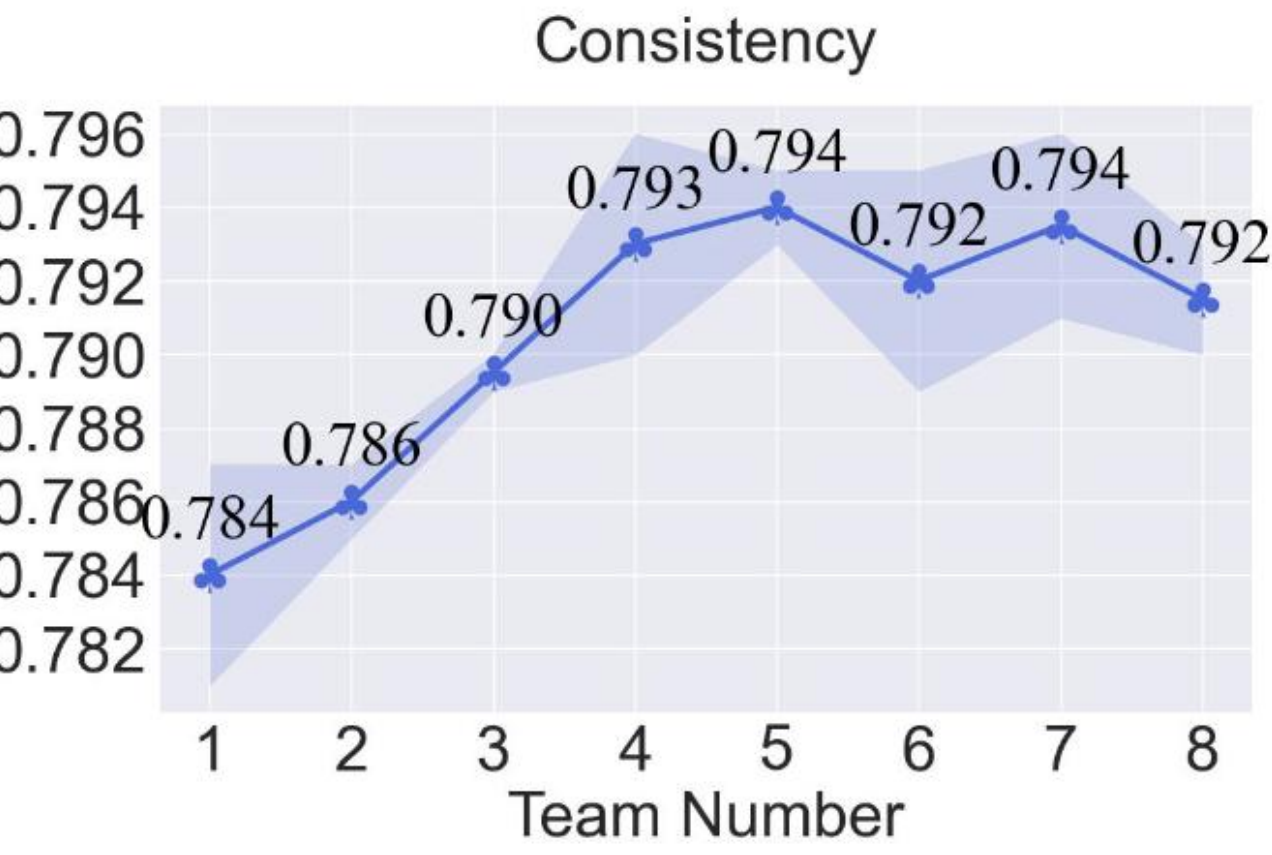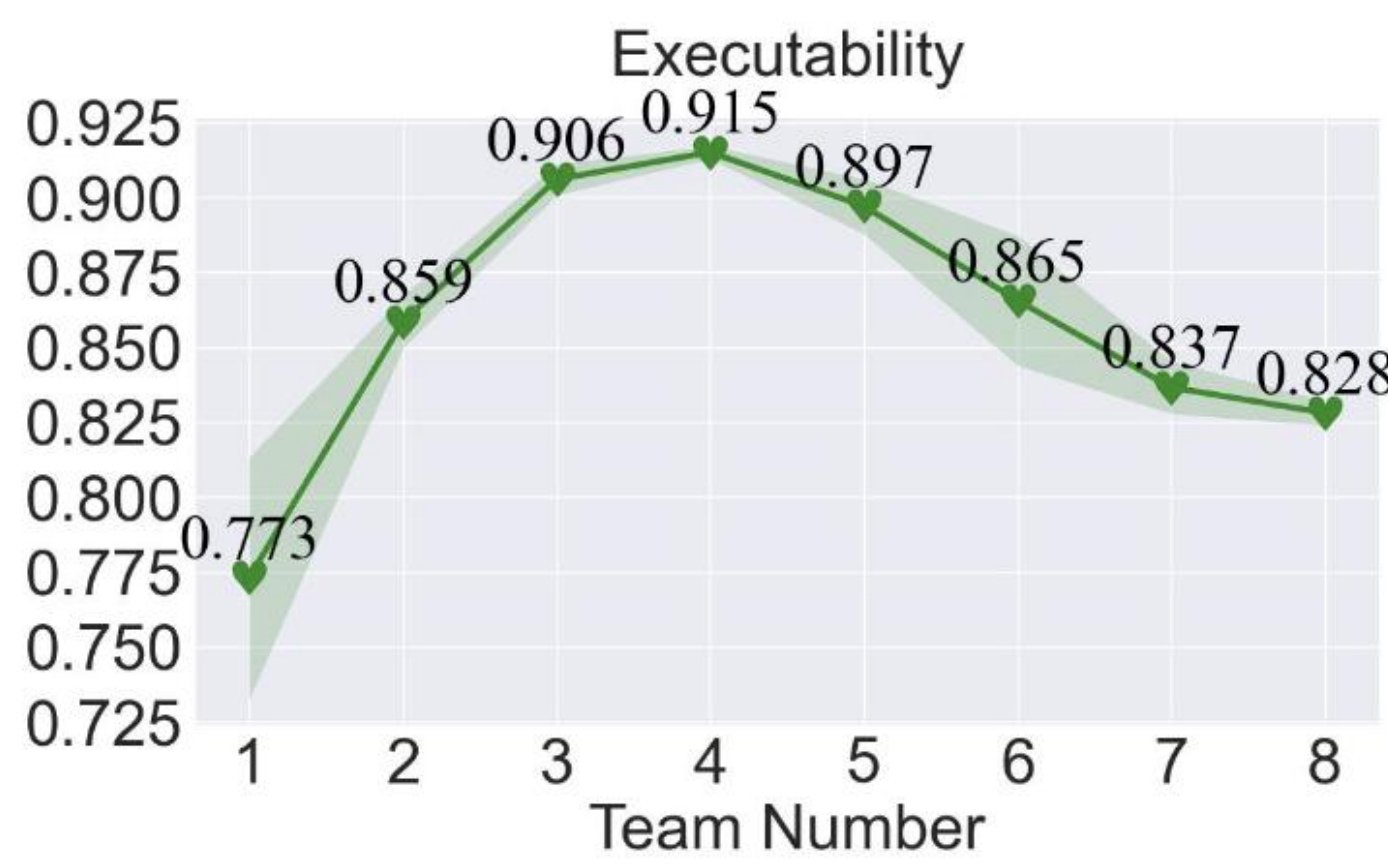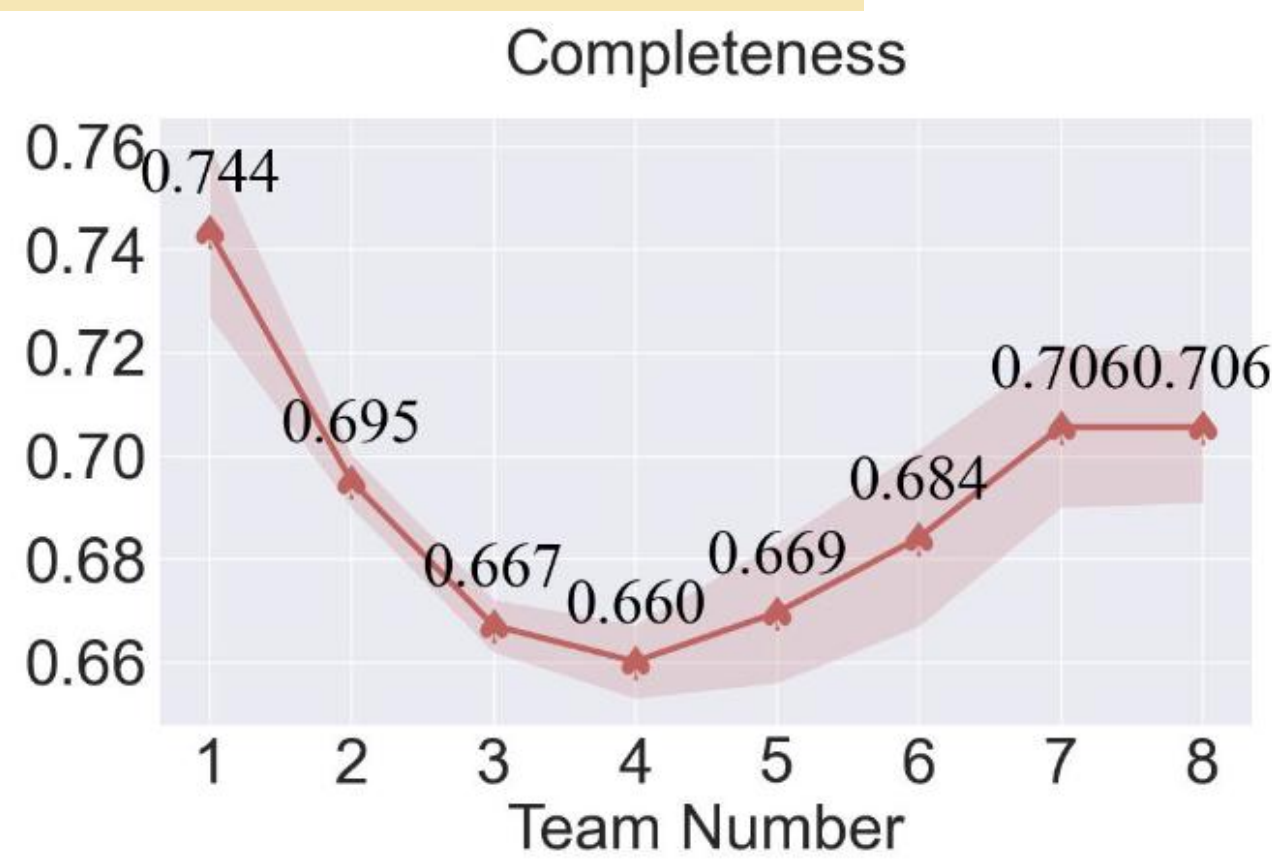
Teams can process various configurations to boost diversity.

Result

Higher quality in softwafre development

diminishing return

Method	Paradigm	Completeness	Executability	Consistency	Quality
GPT-Engineer		$0.502^\dagger$	$0.358^\dagger$	$0.768^\dagger$	$0.543^\dagger$
MetaGPT		$0.483^\dagger$	$0.415^\dagger$	$0.739^\dagger$	$0.545^\dagger$
ChatDev		$\underline{0.744}^\dagger$	$0.813^\dagger$	$\underline{0.781}^\dagger$	$\underline{0.779}^\dagger$
AgentVerse		$0.650^\dagger$	$\underline{0.850}^\dagger$	$0.776^\dagger$	$0.759^\dagger$
GPTSwarm		**0.800**	$0.550^\dagger$	$0.779^\dagger$	$0.710^\dagger$
Croto		$\underline{0.795}$	**0.928**	**0.796**	**0.840**

Great Generalizability in story generation

Mechanism	Paradigm	Grammar and Fluency	Context Relevance	Logic Consistency	Quality
Single-Agent		$2.150^\dagger$	$2.005^\dagger$	$2.425^\dagger$	$2.193^\dagger$
Single-Team Execution		$2.250^\dagger$	$2.325^\dagger$	$2.500^\dagger$	$2.358^\dagger$
2-team Croto		2.725	2.800	3.000	2.842
3-team Croto		2.967	2.767	2.967	2.900
4-team Croto		2.967	2.850	2.908	2.908
5-team Croto		2.980	2.880	2.960	2.940
6-team Croto		2.983	2.900	2.983	2.956
7-team Croto		$\underline{3.000}$	3.171	$\underline{3.014}$	3.062
8-team Croto		$\underline{3.000}^\dagger$	$\underline{3.250}^\dagger$	$3.000^\dagger$	$\underline{3.083}^\dagger$
8-team Croto + Prune		**3.625**	**3.750**	**3.250**	**3.642**

Effectively alleviate by pruning.

Tempetature variety boosts quality

Mechanism	Completeness	Executability	Consistency	Quality
8-team Croto	$\underline{0.706}^\dagger$	$\underline{0.828}^\dagger$	$\underline{0.792}^\dagger$	$\underline{0.775}^\dagger$
+ Prune	**0.795**	**0.928**	**0.796**	**0.840**
Δ compared to Vanilla	+0.089	+0.100	+0.004	+0.065
4-team Croto	0.660	$0.915^\dagger$	**0.793**	$\underline{0.789}^\dagger$
(0.1 0.1 0.1 0.1)	**0.700**	$0.794^\dagger$	0.791	$0.762^\dagger$
(0.4 0.4 0.4 0.4)	0.583	$0.773^\dagger$	$\underline{0.792}$	$0.716^\dagger$
(0.2 0.4 0.6 0.8)	0.575	$0.875^\dagger$	0.790	$0.747^\dagger$
(0.2 0.2 0.4 0.4)	$\underline{0.670}$	**0.925**	0.790	**0.795**
Δ compared to Vanilla	+0.010	+0.010	+0.003	+0.006

Completeness
Team Number values: 1: 0.744, 2: 0.695, 3: 0.667, 4: 0.660, 5: 0.669, 6: 0.684, 7: 0.706, 8: 0.706

Executability
Team Number values: 1: 0.773, 2: 0.859, 3: 0.906, 4: 0.915, 5: 0.897, 6: 0.865, 7: 0.837, 8: 0.828

Consistency
Team Number values: 1: 0.784, 2: 0.786, 3: 0.790, 4: 0.793, 5: 0.794, 6: 0.792, 7: 0.794, 8: 0.792

Quality
Team Number values: 1: 0.767, 2: 0.780, 3: 0.788, 4: 0.789, 5: 0.787, 6: 0.780, 7: 0.778, 8: 0.775

Phases: Design → Coding → Testing

Instructor: CEO, CTO, CTO, Reviewer, Tester
Chat Chain: {task} → {ideas} → {code} → {code} → {code} → {code}
Assistant: CTO, Programmer, Programmer, Programmer, Programmer

Coding: Insights interchange → Superior solution

Contact us!