From Debate to Equilibrium:

Belief-Driven Multi-Agent LLM Reasoning via Bayesian Nash Equilibrium

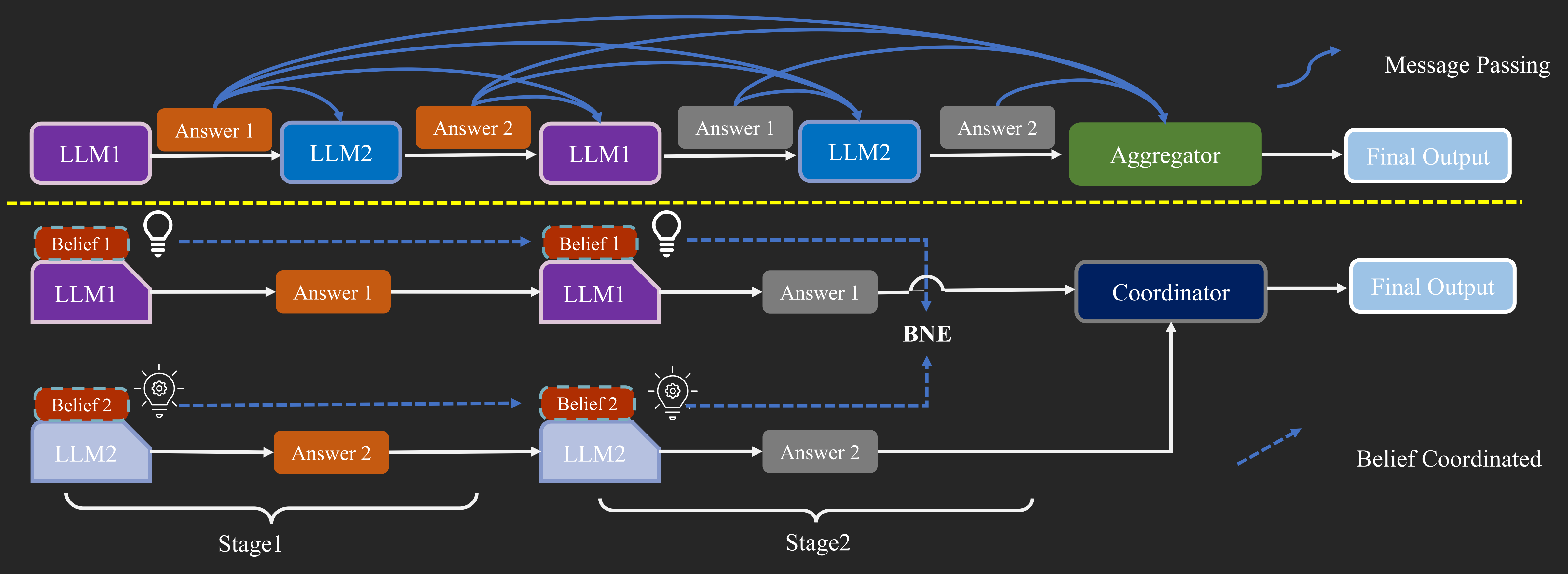
Yi Xie*, Zhanke Zhou*, Chentao Cao, Qiyu Niu, Tongliang Liu, Bo Han

ICML International Conference On Machine Learning

復旦大學 FUDAN UNIVERSITY 1905 · THE UNIVERSITY OF SYDNEY · 香港浸會大學 HONG KONG BAPTIST UNIVERSITY

Motivation: Limitations of Existing Multi-Agent Frameworks

- **Prohibitive Communication Costs (❌):** MAD relies on explicit message passing, incurring substantial token.
- **No Convergence Guarantees (❌):** lack theoretical assurances of converging to a stable, effective solution.
- **Scalability Challenges (❌):** The information exchange often exceeds LLM context limits, impeding scalability.



Our Solution: ECON (Efficient Coordination via Nash Equilibrium)

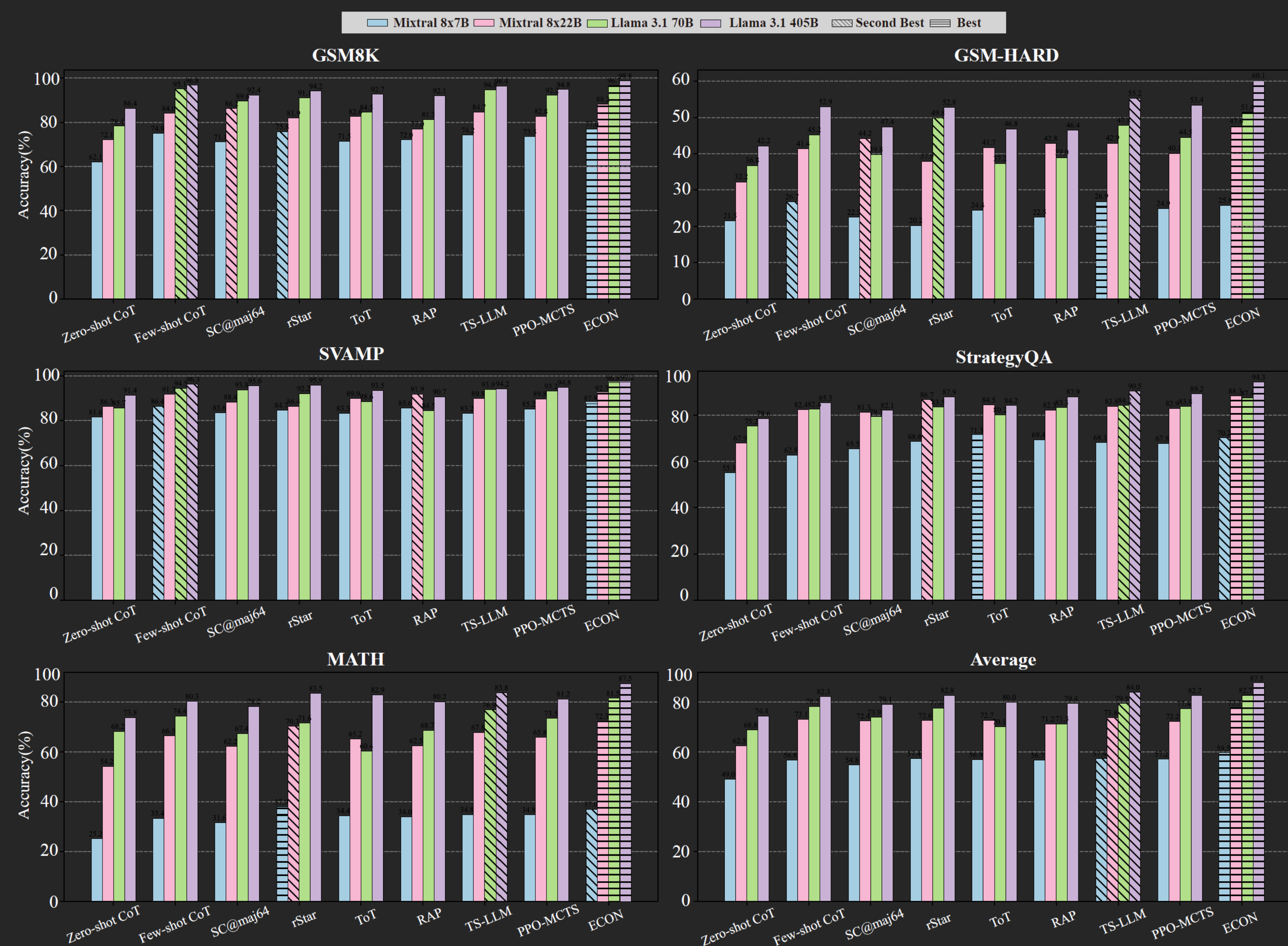
- **Implicit Belief-Driven Coordination (✅):** Replaces costly message passing with belief-based coordination.
- **Guaranteed Convergence to Equilibrium (✅):** Establishes a rigorous Bayesian Nash Equilibrium (BNE) framework.
- **Hierarchical & Scalable Architecture (✅):** Enables effective coordination in large ensembles via a local-to-global.

Theoretical guarantee:

We prove that ECON converges efficiently to a Bayesian Nash Equilibrium (BNE) by deriving a sublinear Bayesian regret bound. The analysis proceeds as follows:

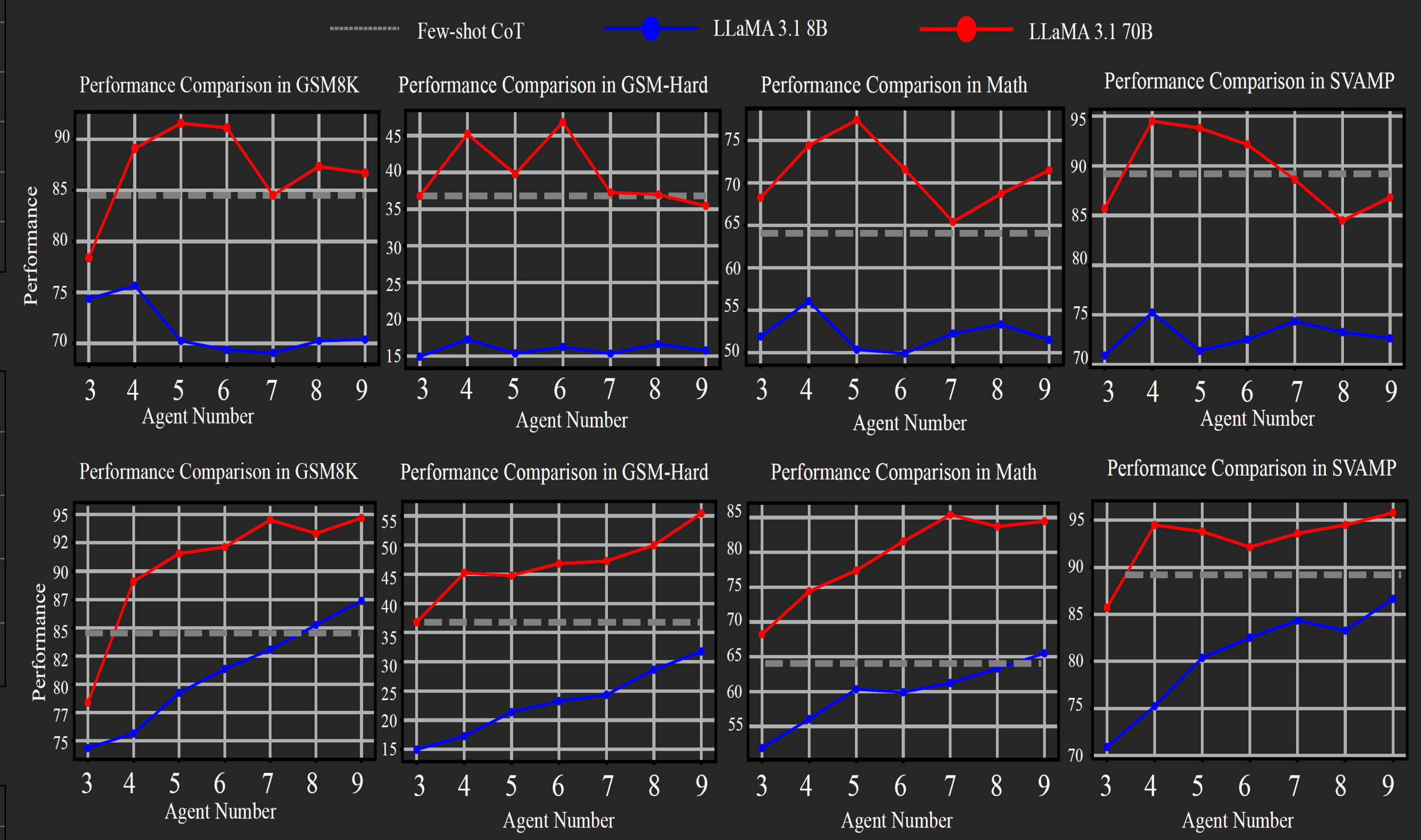
1. BNE Existence: We first establish the existence of a BNE strategy profile π^* in our multi-agent LLM framework, grounded in Glicksberg's Fixed Point Theorem.
2. Performance Difference Lemma: We apply the value difference between the optimal policy π^* and the current policy π^t to the advantage function: $V^*(s) - V^{\pi^t}(s) = \frac{1}{1-\gamma}\mathbb{E}_{s'\sim d_{\pi^*}, a\sim\pi^*}\left[A^{\pi^t}(s', a)\right]$
3. Sublinear Regret Bound: By bounding the advantage with the Q-function estimation error (ε_t) and policy suboptimality (δ_t), both of which decrease at a rate of $\mathcal{O}(1/\sqrt{t})$, we derive the final regret bound for ECON: $\leq \mathcal{O}\left(\frac{N\sqrt{T}}{1-\gamma}\right)$

Reasoning Benchmark: Across five diverse reasoning datasets, ECON outperforms a wide range of strong baselines, from self-consistency to other multi-agent frameworks.



	Validation (#180)						Test (#1,000)					
	Delivery Rate	Commonsense Pass Rate		Hard Constraint Pass Rate		Final Pass Rate	Delivery Rate	Commonsense Pass Rate		Hard Constraint Pass Rate		Final Pass Rate
		Micro	Macro	Micro	Macro			Micro	Macro	Micro	Macro	
Greedy Search	100	74.4	0	60.8	37.8	0	100	72.0	0	52.4	31.8	0
Two-stage												
Mixtral-8x7B-MoE	49.4	30.0	0	1.2	0.6	0	51.2	32.2	0.2	0.7	0.4	0
Gemini Pro	28.9	18.9	0	0.5	0.6	0	39.1	24.9	0	0.6	0.1	0
GPT-3.5-Turbo	86.7	54.0	0	0	0	0	91.8	57.9	0	0.5	0.6	0
GPT-4-Turbo	89.4	61.1	2.8	15.2	10.6	0.6	93.1	63.3	2.0	10.5	5.5	0.6
Debate (GPT-4) @3round	95.2	67.3	6.7	22.7	13.1	2.3	97.8	72.4	11.3	17.4	12.1	3.7
ECON (GPT-4)	**100**	**71.4**	**15.6**	**32.1**	**25.7**	**7.2**	**100**	**82.1**	**26.6**	**32.4**	**17.6**	**9.3**
Sole-planning												
Direct$_{\text{GPT-3.5-Turbo}}$	100	60.2	4.4	11.0	2.8	0	100	59.5	2.7	9.5	4.4	0.6
CoT$_{\text{GPT-3.5-Turbo}}$	100	66.3	3.3	11.9	5.0	0	100	64.4	2.3	9.8	3.8	0.4
ReAct$_{\text{GPT-3.5-Turbo}}$	82.2	47.6	3.9	11.4	6.7	0.6	81.6	45.9	2.5	10.7	3.1	0.7
Reflexion$_{\text{GPT-3.5Turbo}}$	93.9	53.8	2.8	11.0	2.8	0	92.1	52.1	2.2	9.9	3.8	0.6
Direct$_{\text{Mixtral-8x7B-MoE}}$	100	68.1	5.0	3.3	1.1	0	99.3	67.0	3.7	3.9	1.6	0.7
Direct$_{\text{Gemini Pro}}$	93.9	65.0	8.3	9.3	4.4	0.6	93.7	64.7	7.9	10.6	4.7	2.1
Direct$_{\text{GPT-4-Turbo}}$	**100**	80.4	17.2	47.1	22.2	4.4	**100**	80.6	15.2	44.3	23.1	4.4
Debate (GPT-4)@3round	97.7	78.9	15.6	43.3	20.6	6.7	98.2	79.5	18.8	41.7	22.9	7.1
ECON (GPT-4)	**100**	**83.3**	**22.2**	**51.7**	**27.8**	**12.9**	**100**	**84.2**	**23.5**	**49.8**	**28.7**	**15.2**

Effective Scalability through Hierarchical Coordination: A single coordinator struggles with excessive agents (blue line). ECON forms a Global Nash Equilibrium from multiple Local Equilibria, enables performance to scale effectively (red line) from 3 to 9 Execution LLMs.



Consumption Analysis: ECON achieves SOTA accuracy while being significantly more token-efficient. On MATH, it uses 7x fewer tokens than Self-Consistency yet yields 14% higher performance.

Dataset	Inference Strategy	LLaMA3.1 70B		Mixtral 8x7b		Mixtral 8x22b	
		Token Usage	Performance	Token Usage	Performance	Token Usage	Performance
MATH	Multi-Agent Debate (3 rounds)	2154.87	71.58	1462.12	31.28	5345.56	67.41
	RAP	2653.27	68.71	1737.73	33.99	6668.55	62.53
	ECON (with detailed strategy)	3270.06	72.38	2150.23	26.18	8054.03	68.23
	Self Consistency (64 rounds)	11917.00	67.39	8066.21	31.58	29616.13	62.21
	ECON	1629.79	81.47	1128.23	35.02	4270.86	72.29
GSM8K	Multi-Agent Debate (3 rounds)	1391.57	86.32	1463.40	70.19	5714.05	81.95
	RAP	1907.86	81.33	1248.66	72.03	6517.77	76.97
	ECON (with detailed strategy)	2772.24	85.17	1188.13	65.37	9341.60	81.46
	Self Consistency (64 rounds)	9574.25	89.56	6601.34	71.08	24671.91	86.24
	ECON	1131.65	92.70	1284.98	76.97	4715.31	88.20
GSM-Hard	Multi-Agent Debate (3 rounds)	3030.73	41.98	1478.14	20.04	9250.78	45.21
	RAP	1768.72	38.97	1036.11	22.47	6464.52	42.79
	ECON (with detailed strategy)	3662.64	44.12	2239.07	18.52	11464.98	41.04
	Self Consistency (64 rounds)	16724.69	39.76	11668.19	22.47	74544.25	44.19
	ECON	1518.76	51.43	1271.53	25.76	7101.62	47.58

Planning Benchmark: On the highly complex TravelPlanner benchmark, ECON achieves a Final Pass Rate of 9.3%, more than doubling MAD (3.7%).

paper code slides