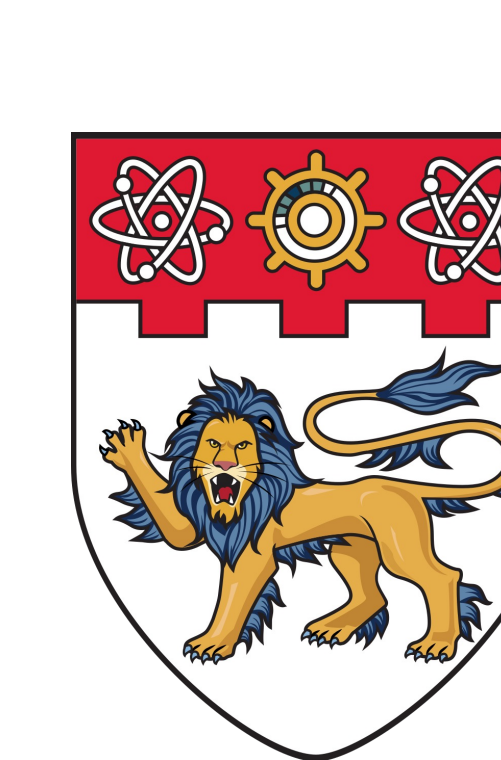


EngiAgent: Fully Connected Coordination of LLM Agents for Solving Open-ended Engineering Problems with Feasible Solutions

Xiyuan Zhou^{1*}, Ruixi Zou^{2*}, Xinlei Wang^{3,4*}, Yuheng Cheng², Yan Xu^{1†}, Junhua Zhao^{2,5†}, Jinjin Gu^{3†}

¹ NTU ² CUHK-Shenzhen ³ INSAIT ⁴ USYD ⁵ AIRS *Equal contribution †Corresponding author



1 Motivation & Contributions

Motivation

- Engineering problem solving requires **feasible solutions**, not just fluent reasoning or correct-looking formulations.
- Prior LLM systems often fail by ignoring real-world constraints.

Feasibility is the core requirement in engineering.

Common failure modes of prior methods

Error Type

Fancy but Vague Modeling

We propose an RL framework with game theory and NNs to simulate battery-PV interactions.

Sounds fancy but lacks concrete equations and variables.

Altering Key Data

Assume the community demand is 60 kWh, then PV plus the battery can fully meet the load without using the grid.

Solvable, but changes the original problem.

Violating Physical Laws

Discharge 60 kWh from the battery to cover the shortfall.

Exceeds battery capacity, violating physical limits.

Over-constraining the model

Set grid purchase to zero, requiring all demand to be met by PV and the battery.

PV + battery < demand, so no feasible solution.

Contributions

- We identify **feasibility as a core requirement** in engineering problem solving and build a benchmark across four engineering domains.
- We propose **EngiAgent**, a fully connected multi-agent system that coordinates analysis, modeling, verification, solving, and evaluation to produce feasible solutions.
- EngiAgent achieves substantially **higher feasibility** than prior LLM-based methods across diverse engineering tasks.

2 Benchmark & Evaluation Metrics

Our benchmark includes 53 peer-reviewed engineering tasks across four categories, each grounded in original-paper data and real-world constraints to evaluate feasibility as the criterion.



Evaluation Metrics

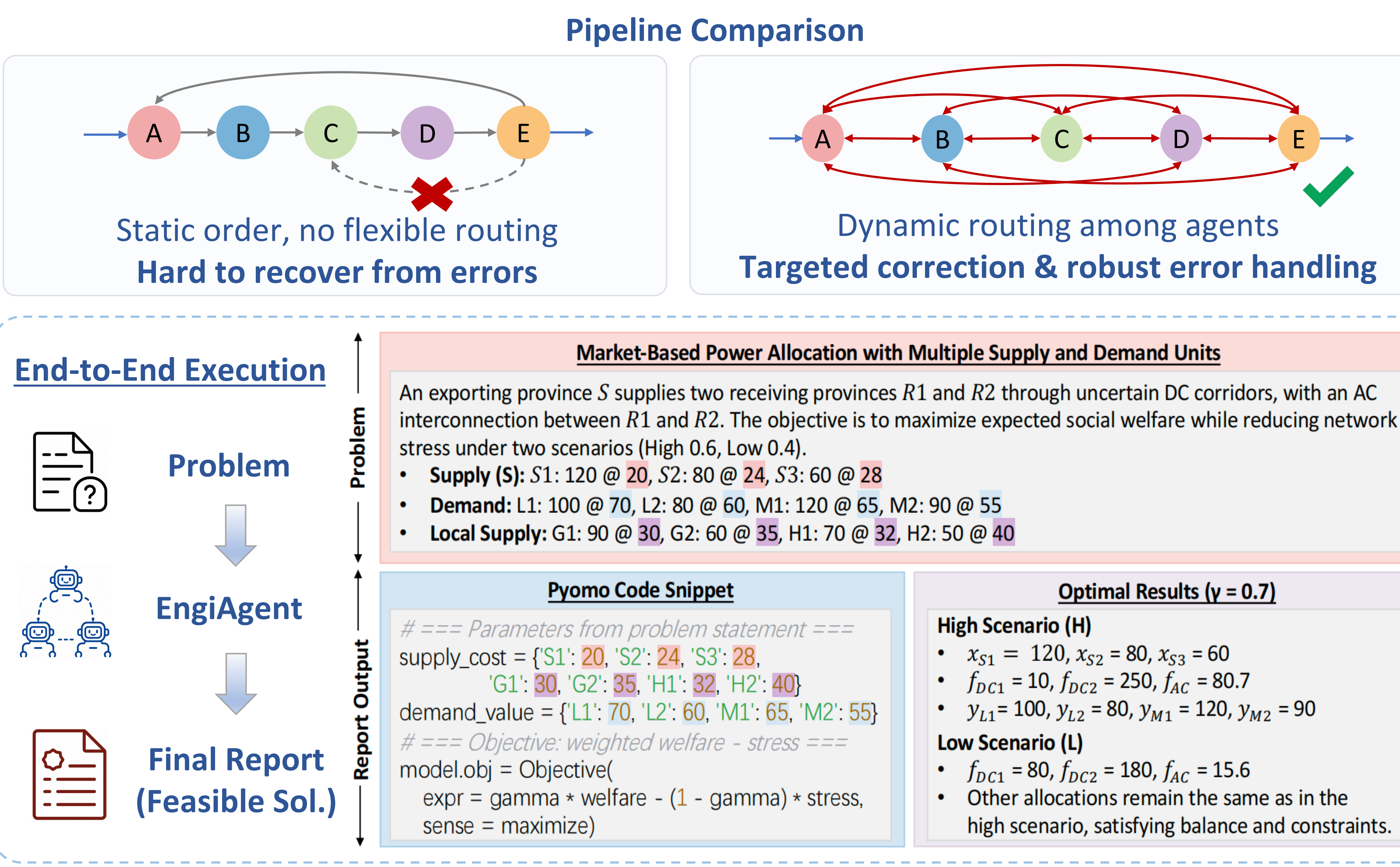
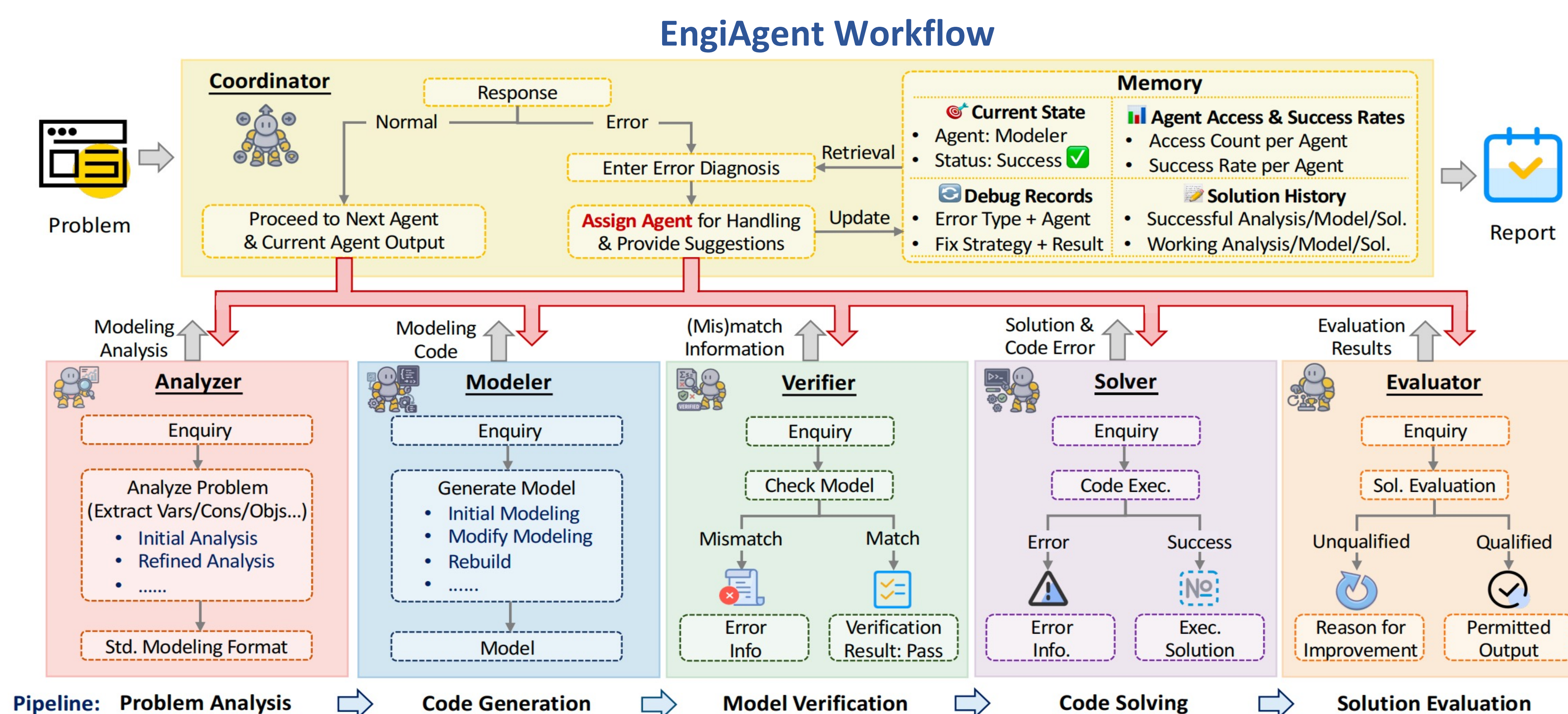
Text-based quality metrics

- Information Extraction
- Domain-specific Reasoning
- Multi-objective Decision-making
- Uncertainty Handling

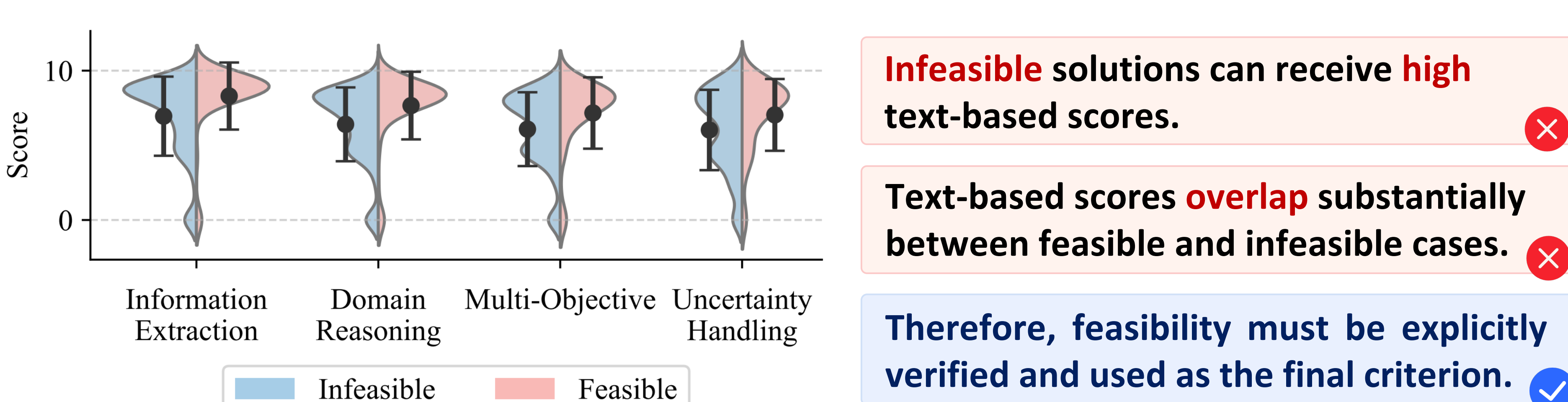
Solution-level metrics

- Numerical Solution Rate
- Feasible Solution Rate

3 EngiAgent Framework



5 Why Text-Based Judgments Are Not Enough

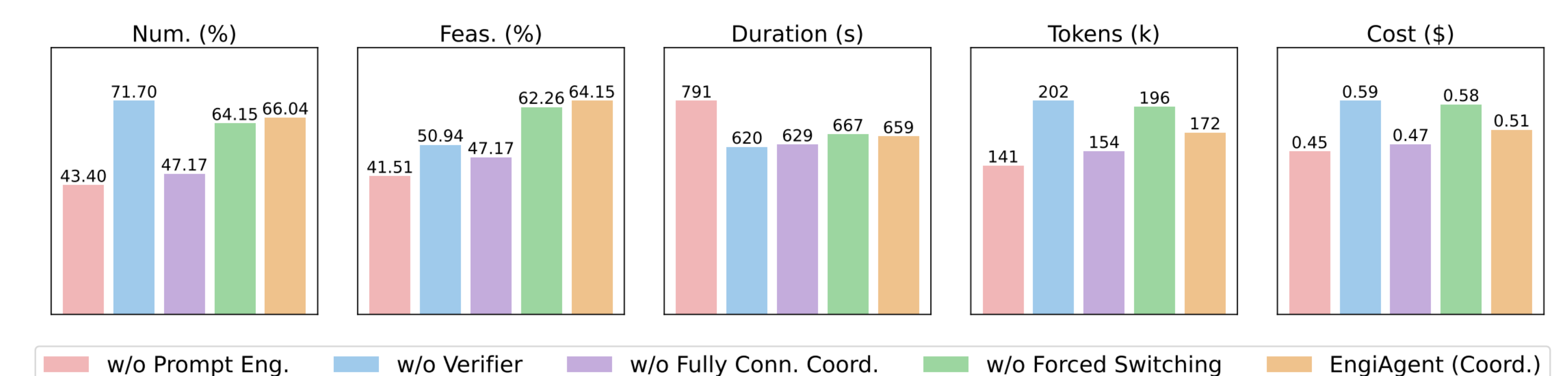


4 Main Results

Methods	Num. ↑		Total					Feasible				
		Feas. ↑	IE ↑	DR ↑	MO ↑	UH ↑	Avg. ↑	IE ↑	DR ↑	MO ↑	UH ↑	Avg. ↑
GPT-4o												
Zero-shot	22.64%	5.66%	5.66	5.42	4.47	3.33	4.72	5.67	5.10	1.33	4.00	4.03
ResearchAgent	15.09%	3.77%	5.36	5.25	5.10	5.17	5.22	5.50	5.50	6.00	5.50	5.63
DS-Agent	62.26%	5.66%	5.91	4.83	4.85	4.79	5.10	6.00	5.75	5.38	6.25	5.85
MM-Agent	13.21%	7.55%	6.89	7.21	6.48	7.98	7.14	8.25	8.38	7.88	7.50	8.00
EngiAgent (Fixed)	47.17%	47.17%	8.30	7.22	6.67	7.14	7.33	8.84	8.12	7.40	7.56	7.98
EngiAgent (Coord.)	66.04%	64.15%	8.67	7.74	7.05	7.41	7.72	8.86	8.13	7.53	7.94	8.12
Gemini-2.5 Flash												
Zero-shot	0.00%	0.00%	6.80	5.67	5.53	3.55	5.39	/	/	/	/	/
ResearchAgent	1.89%	0.00%	4.56	4.41	4.34	4.36	4.42	/	/	/	/	/
DS-Agent	60.38%	0.00%	7.17	6.75	6.71	5.85	6.62	/	/	/	/	/
MM-Agent	9.43%	3.77%	6.39	5.72	5.98	6.40	6.12	8.00	8.00	8.25	6.00	7.56
EngiAgent (Fixed)	45.28%	39.62%	7.30	6.93	5.82	5.21	6.32	8.60	6.71	6.14	5.36	6.70
EngiAgent (Coord.)	52.83%	50.94%	8.30	6.89	6.30	6.06	6.89	8.80	7.23	6.50	6.81	7.34
DeepSeek-V3-671B												
Zero-shot	0.00%	0.00%	6.20	5.60	5.09	3.99	5.22	6.50	6.25	6.50	6.50	6.44
ResearchAgent	9.43%	7.55%	5.12	4.89	4.84	4.98	4.96	6.50	5.98	4.92	4.98	5.60
DS-Agent	77.36%	28.30%	7.39	6.88	6.77	6.01	6.76	8.22	7.64	7.38	7.12	7.59
MM-Agent	11.32%	5.66%	8.98	8.84	7.64	7.01	8.12	9.00	8.50	8.00	7.85	8.34
EngiAgent (Fixed)	73.58%	67.92%	8.08	7.28	7.20	6.98	7.39	8.38	7.45	7.53	7.33	7.67
EngiAgent (Coord.)	79.25%	75.47%	7.85	7.42	7.08	6.59	7.24	8.05	8.16	7.26	6.52	7.50

- Average **~7x improvement** over prior SOTA methods.
- The fully connected coordinator improves feasibility by **more than 10 percentage points** on average versus the fixed pipeline.
- Narrow the gap between numerical outputs and feasible solutions.

6 Ablation Study



- The **verifier** and the **fully connected coordinator** are essential for feasibility.
- Structured prompts guide the **entire** modeling-and-solving process.

