



NEC-AIST  
AI Cooperative  
Research Laboratory



# Utility-Aware Task Decomposition and Exchange across LLM Agents

Shunta Kimura<sup>1,2</sup>, Takuya Hiraoka<sup>1,3</sup>, Ryota Higa<sup>1,3</sup>, Yoshimasa Tsuruoka<sup>1,2</sup>, Katsuhide Fujita<sup>1,4</sup>

<sup>1</sup>National Institute of Advanced Industrial Science and Technology <sup>2</sup>The University of Tokyo

<sup>3</sup>NEC Corporation <sup>4</sup>Tokyo University of Agriculture and Technology

# Agenda

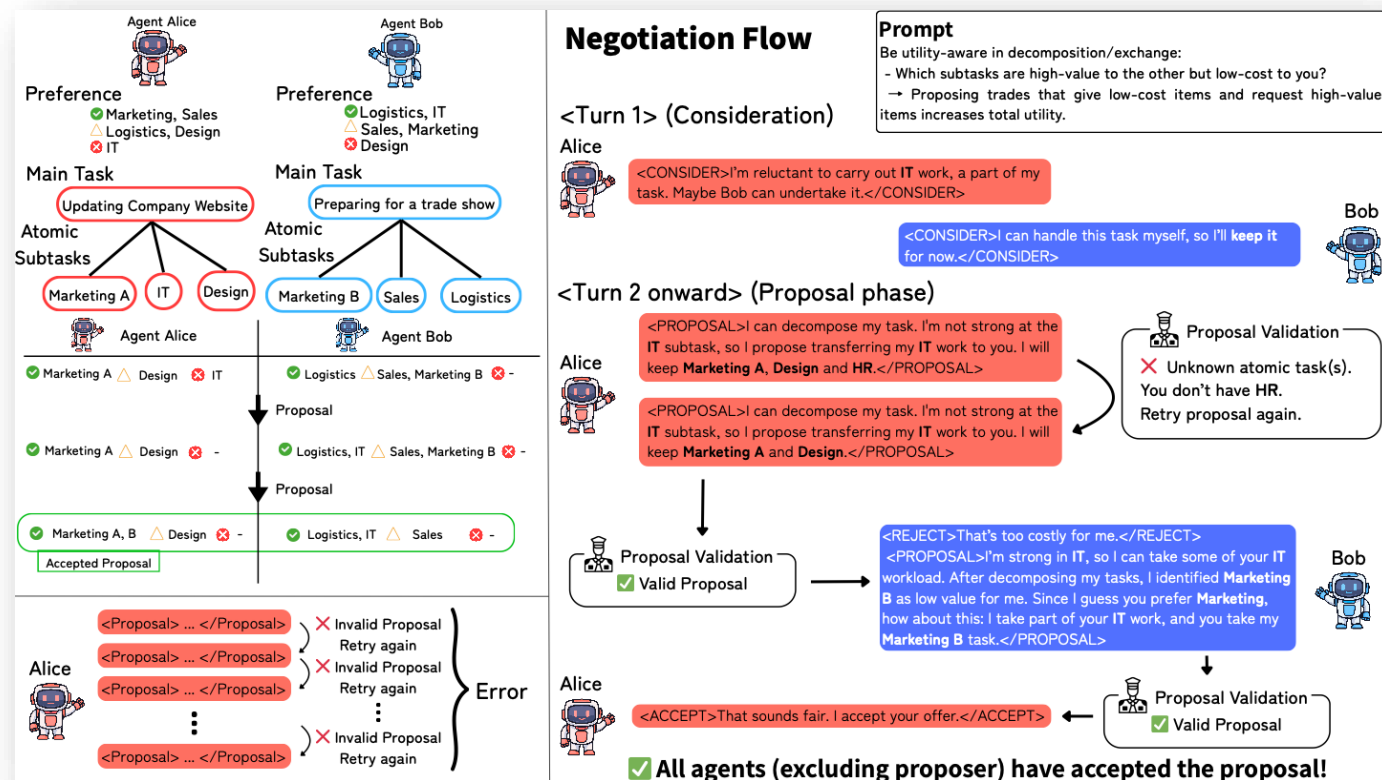
1. Introduction
2. Proposed Framework
3. Experimental Setup
4. Result
5. Analysis
6. Conclusion

# Introduction

- Previous research on LLM-based multi-agent systems assumes the presence of either:
  - **a central coordinator** (Abdelnabi et al. 2024)
  - or
  - **a shared utility function** (Gandhi, Sadigh, and Goodman 2023; Duan et al. 2024)
- In real-world, neither is typically available.

# Introduction

## Framework for utility-aware task decomposition and exchange



# Proposed Framework

1. Utility-aware prompt
2. Per-turn proposal validation

# Proposed Framework

## 1. Utility-aware prompt

1. Utility-aware prompt
2. Per-turn proposal validation

### Negotiation Flow

<Turn 1> (Consideration)

#### Prompt

Be utility-aware in decomposition/exchange:

- Which subtasks are high-value to the other but low-cost to you?  
→ Proposing trades that give low-cost items and request high-value items increases total utility.

Alice



<CONSIDER>I'm reluctant to carry out **IT** work, a part of my task. Maybe Bob can undertake it.  
</CONSIDER>

<CONSIDER>I can handle this task myself, so I'll **keep it** for now.</CONSIDER>

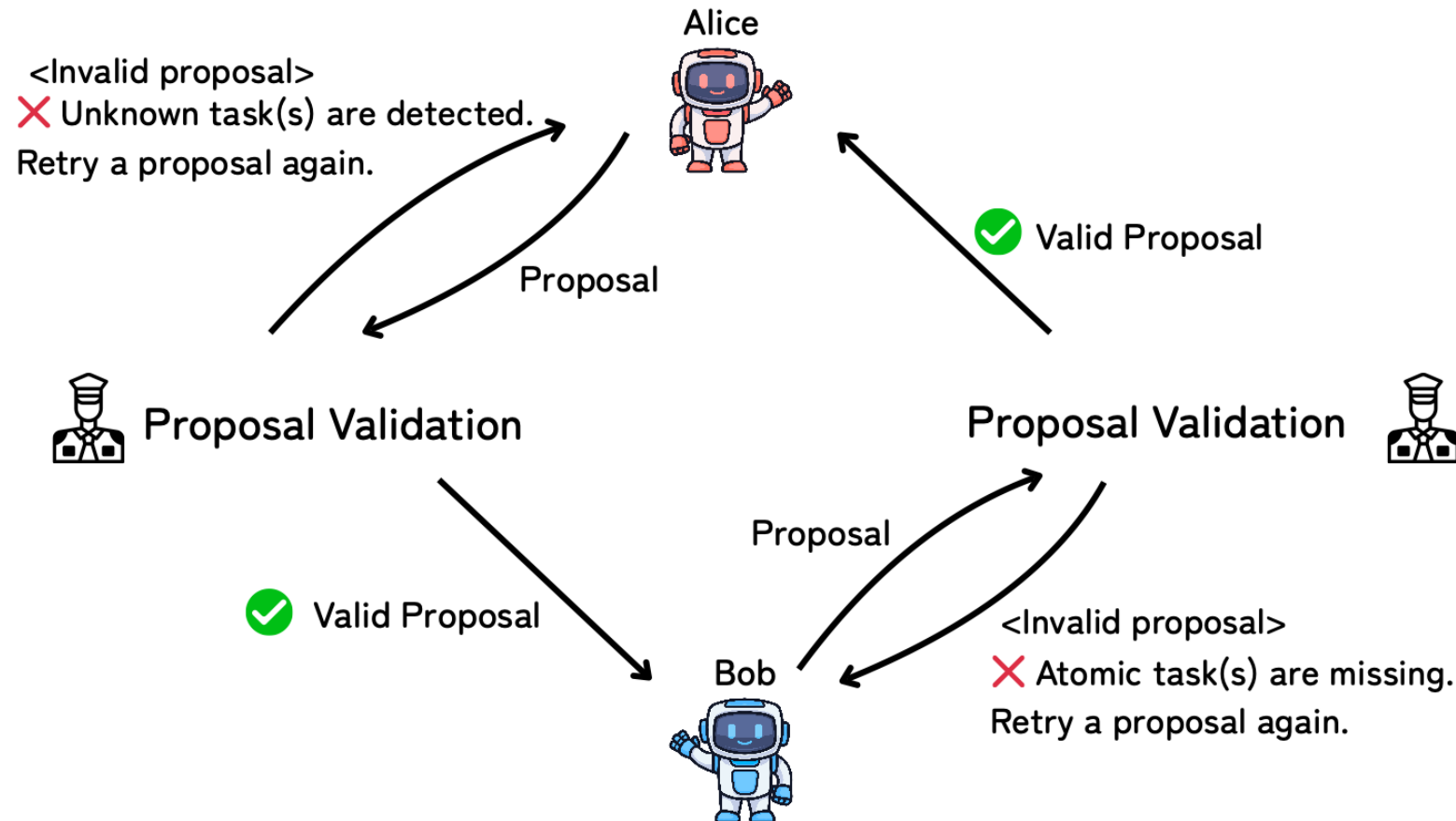
Bob



# Proposed Framework

## 2. Per-turn Proposal Validation

1. Utility-aware prompt
2. Per-turn proposal validation



# Proposed Framework

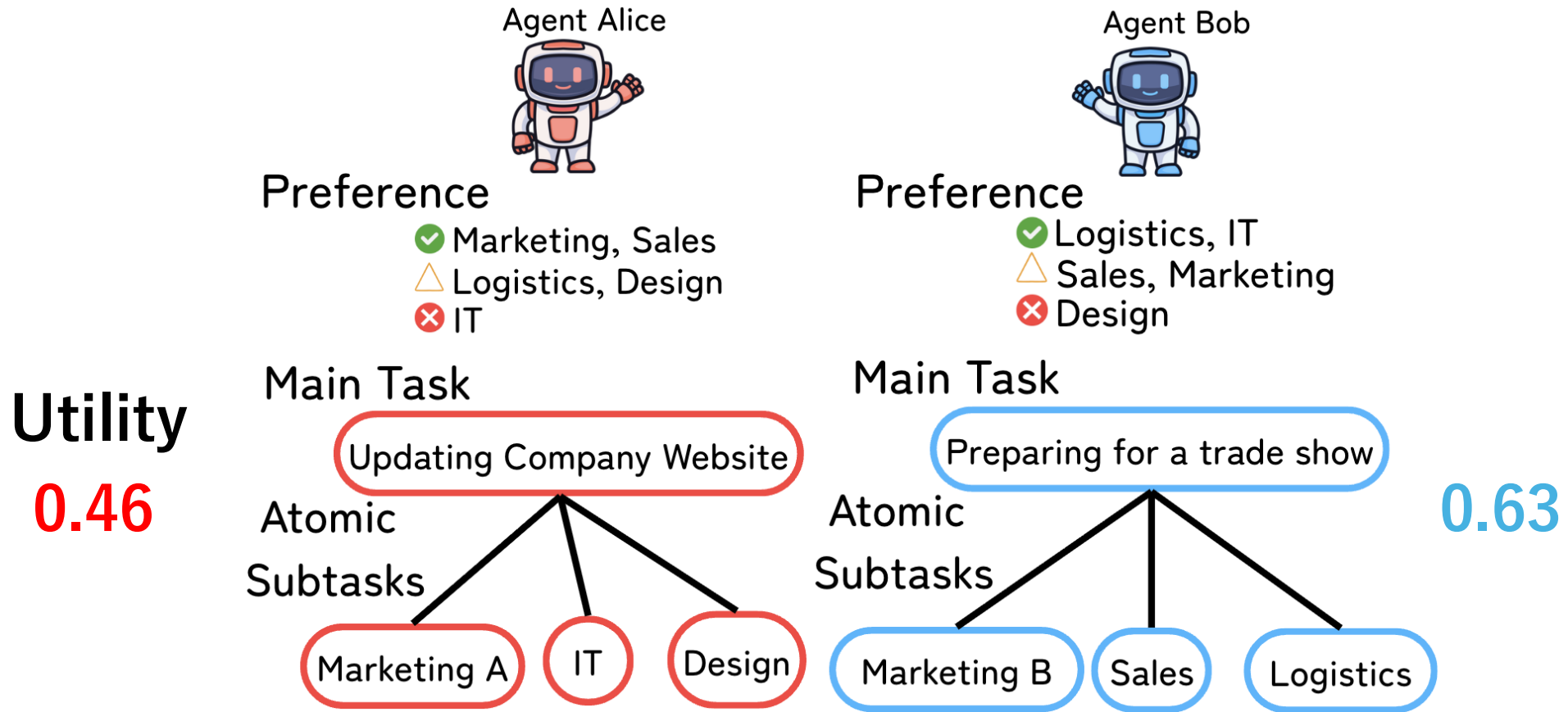
1. Utility-aware prompt
2. Per-turn proposal validation

Task:

Decomposition and Exchange  
for mutually satisfactory  
outcomes.



# Proposed Framework



# Proposed Framework

## Utility-aware prompt

### Negotiation Flow

<Turn 1> (Consideration)

#### Prompt

Be utility-aware in decomposition/exchange:

- Which subtasks are high-value to the other but low-cost to you?  
→ Proposing trades that give low-cost items and request high-value items increases total utility.

Alice



<CONSIDER>I'm reluctant to carry out IT work, a part of my task. Maybe Bob can undertake it.  
</CONSIDER>

Bob



<CONSIDER>I can handle this task myself, so I'll **keep it** for now.</CONSIDER>

# Proposed Framework

**Initial Utility**

**0.46**    **0.63**

<Turn 2 onward> (Proposal phase)

Alice



<PROPOSAL>I can decompose my task. I'm not strong at the IT subtask, so I propose transferring my IT work to you. I will keep **Marketing A**, **Design** and **HR**.</PROPOSAL>

<PROPOSAL>I can decompose my task. I'm not strong at the IT subtask, so I propose transferring my IT work to you. I will keep **Marketing A** and **Design**.</PROPOSAL>



Proposal Validation

✗ Unknown atomic task(s).  
You don't have **HR**.  
Retry proposal again.

**0.65** ↓ **0.47**



Proposal Validation

✓ Valid Proposal



# Proposed Framework

**Initial Utility**

**0.46**    **0.63**

**0.65** ↓ **0.47**



Proposal Validation  
✓ Valid Proposal



<REJECT>That's too costly for me.</REJECT>  
<PROPOSAL>I'm strong in **IT**, so I can take some of your **IT** workload. After decomposing my tasks, I identified **Marketing B** as low value for me. Since I guess you prefer **Marketing**, how about this: I take part of your **IT** work, and you take my **Marketing B** task.  
</PROPOSAL>

Bob



Alice



<ACCEPT>That sounds fair. I accept your offer.  
</ACCEPT>

**0.58** ↓ **0.72**



Proposal Validation  
✓ Valid Proposal



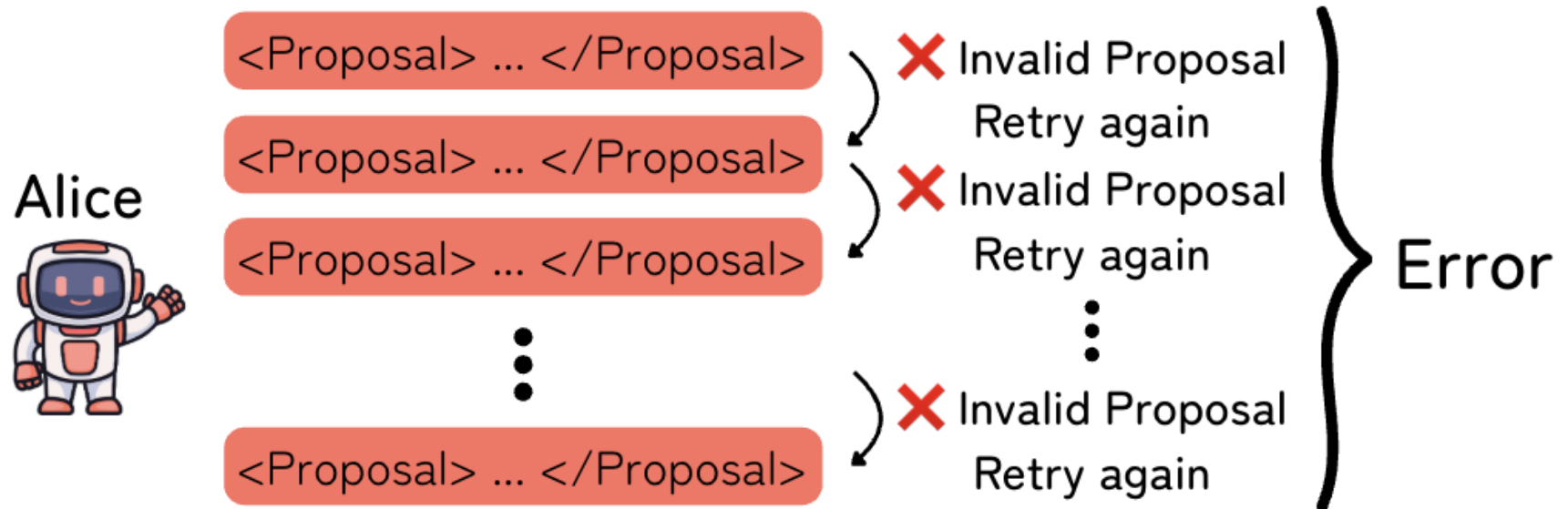
# Proposed Framework

- Termination and Outcome

- **Agreement**

- **Disagreement**

- **Error**



# Experimental Setup

- o4-mini; GPT-4.1-mini; GPT-4.1-nano
- Baseline:
  - **Removal** of the utility-aware prompt
  - **No** per-turn validation

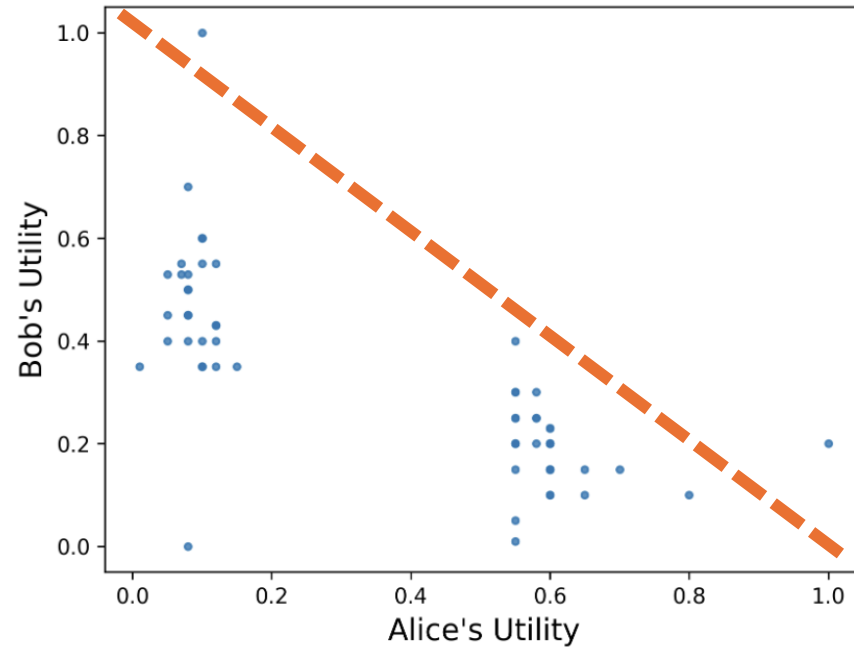
# Utility Generation

- LLM generated the utility table.
- Provided task description and preference

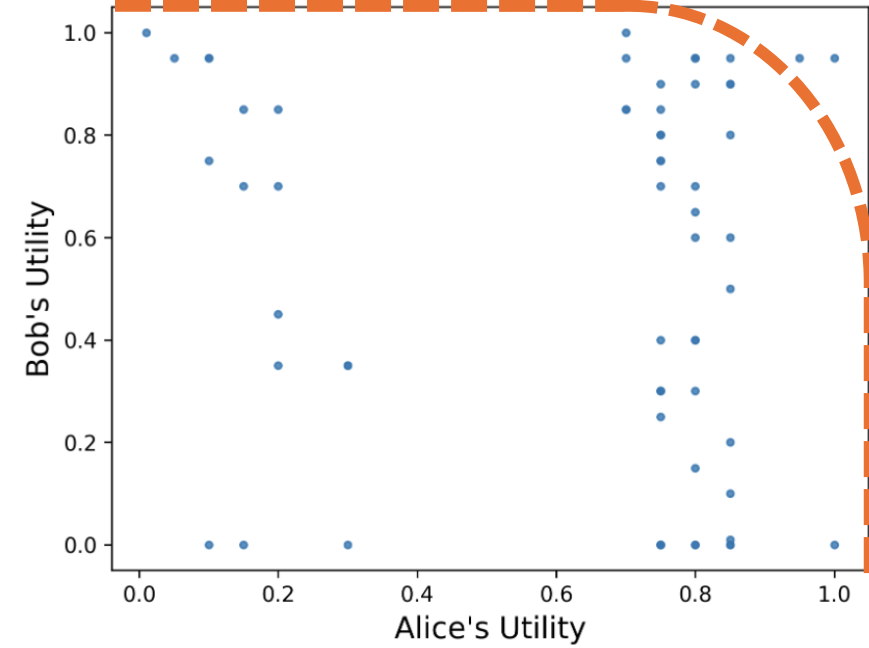
```
{  
    "Alice": "You are good at ...",  
    "Bob": "You are a marketing  
            professional ..."  
}
```

# Scenario

- Competitive and cooperative scenarios.



(a) Competitive scenario



(b) Cooperative scenario



# Result

(a) Cooperative

Model	Setting	Agreement	Error	Turns	Alice	Bob	SW	$L_2$ to Pareto
o4-mini	Proposed	$100 \pm 0\%$	$0 \pm 0\%$	$3.0 \pm 1.4$	$0.78 \pm 0.07$	$0.74 \pm 0.15$	$1.52 \pm 0.20$	$0.27 \pm 0.15$
	Baseline	$83.3 \pm 15.2\%$	$40 \pm 15\%$	$4.8 \pm 4.0$	$0.81 \pm 0.10$	$0.79 \pm 0.12$	$1.60 \pm 0.22$	$0.21 \pm 0.15$
GPT-4.1-mini	Proposed	$100 \pm 0\%$	$80 \pm 13\%$	$2.0 \pm 0.0$	$0.75 \pm 0.00$	$0.70 \pm 0.00$	$1.45 \pm 0.00$	$0.32 \pm 0.0$
	Baseline	$28.6 \pm 17.1\%$	$30 \pm 14\%$	$7.7 \pm 3.9$	$0.75 \pm 0.00$	$0.70 \pm 0.00$	$1.45 \pm 0.00$	$0.32 \pm 0.0$
GPT-4.1-nano	Proposed	$0 \pm 0\%$	$100 \pm 0\%$	—	—	—	—	—
	Baseline	$0 \pm 0\%$	$50 \pm 16\%$	$10 \pm 0.0$	$0.75 \pm 0.00$	$0.70 \pm 0.00$	$1.45 \pm 0.00$	$0.32 \pm 0.0$

(b) Competitive

Model	Setting	Agreement	Error	Turns	Alice	Bob	SW	$L_2$ to Pareto
o4-mini	Proposed	$100 \pm 0\%$	$10 \pm 9\%$	$3.2 \pm 2.0$	$0.62 \pm 0.14$	$0.26 \pm 0.06$	$0.88 \pm 0.13$	$0.022 \pm 0.013$
	Baseline	$60.0 \pm 21.9\%$	$50 \pm 16\%$	$6.8 \pm 4.0$	$0.58 \pm 0.00$	$0.25 \pm 0.00$	$0.83 \pm 0.00$	$0.028 \pm 0.0$
GPT-4.1-mini	Proposed	$0 \pm 0\%$	$100 \pm 0\%$	—	—	—	—	—
	Baseline	$0 \pm 0\%$	$0 \pm 0\%$	$10 \pm 0.0$	$0.58 \pm 0.00$	$0.25 \pm 0.00$	$0.83 \pm 0.00$	$0.028 \pm 0.0$
GPT-4.1-nano	Proposed	$0 \pm 0\%$	$100 \pm 0\%$	—	—	—	—	—
	Baseline	$0 \pm 0\%$	$40 \pm 15\%$	$10 \pm 0.0$	$0.58 \pm 0.00$	$0.25 \pm 0.00$	$0.83 \pm 0.00$	$0.028 \pm 0.0$

- High Error rate in non-reasoning models.

# Result (o4-mini)

- Cooperative

(c) Initial utilities				
Scenario	Alice	Bob	SW	$L_2$ to Pareto
Cooperative	0.75	0.70	1.45	0.32
Competitive	0.58	0.25	0.83	0.028

Setting	Agreement	Error	Turns	Alice	Bob	SW	$L_2$ to Pareto
Proposed	$100 \pm 0\%$	$0 \pm 0\%$	$3.0 \pm 1.4$	$0.78 \pm 0.07$	$0.74 \pm 0.15$	$1.52 \pm 0.20$	$0.27 \pm 0.15$
Baseline	$83.3 \pm 15.2\%$	$40 \pm 15\%$	$4.8 \pm 4.0$	$0.81 \pm 0.10$	$0.79 \pm 0.12$	$1.60 \pm 0.22$	$0.21 \pm 0.15$

- The **baseline** performs **better** in terms of distance from the Pareto front.

# Result (o4-mini)

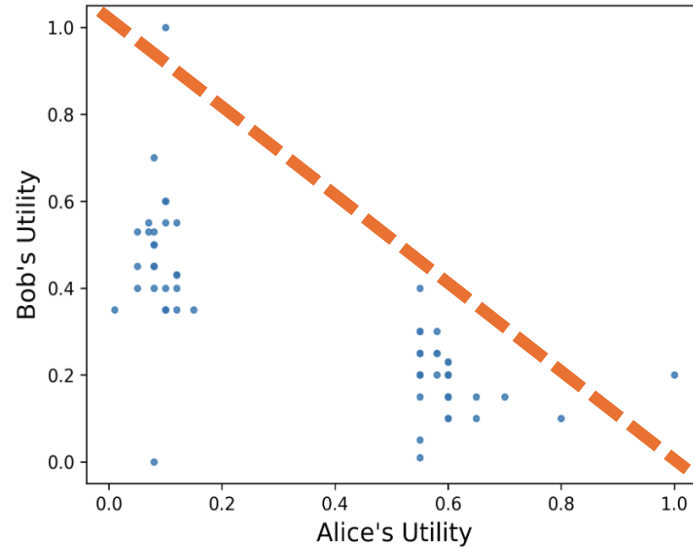
- Competitive

(c) Initial utilities				
Scenario	Alice	Bob	SW	$L_2$ to Pareto
Cooperative	0.75	0.70	1.45	0.32
Competitive	0.58	0.25	0.83	0.028

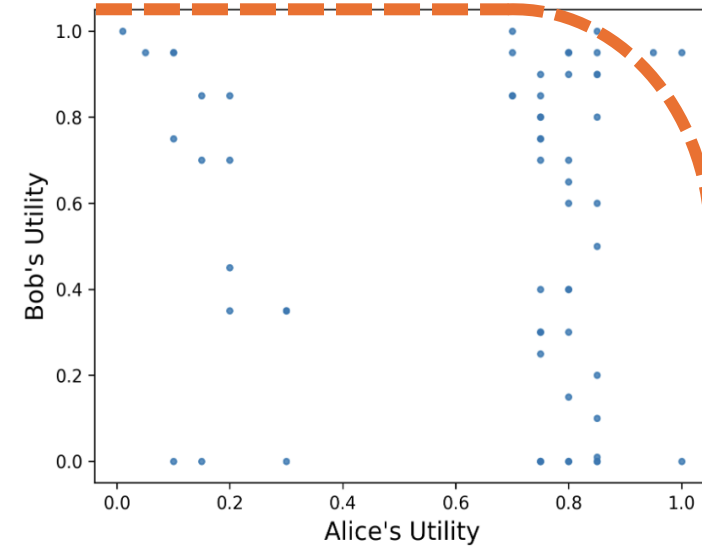
Setting	Agreement	Error	Turns	Alice	Bob	SW	$L_2$ to Pareto
Proposed	$100 \pm 0\%$	$10 \pm 9\%$	$3.2 \pm 2.0$	$0.62 \pm 0.14$	$0.26 \pm 0.06$	$0.88 \pm 0.13$	$0.022 \pm 0.013$
Baseline	$60.0 \pm 21.9\%$	$50 \pm 16\%$	$6.8 \pm 4.0$	$0.58 \pm 0.00$	$0.25 \pm 0.00$	$0.83 \pm 0.00$	$0.028 \pm 0.0$

- The proposed framework outperforms the baseline across **all** metrics.

# Analysis



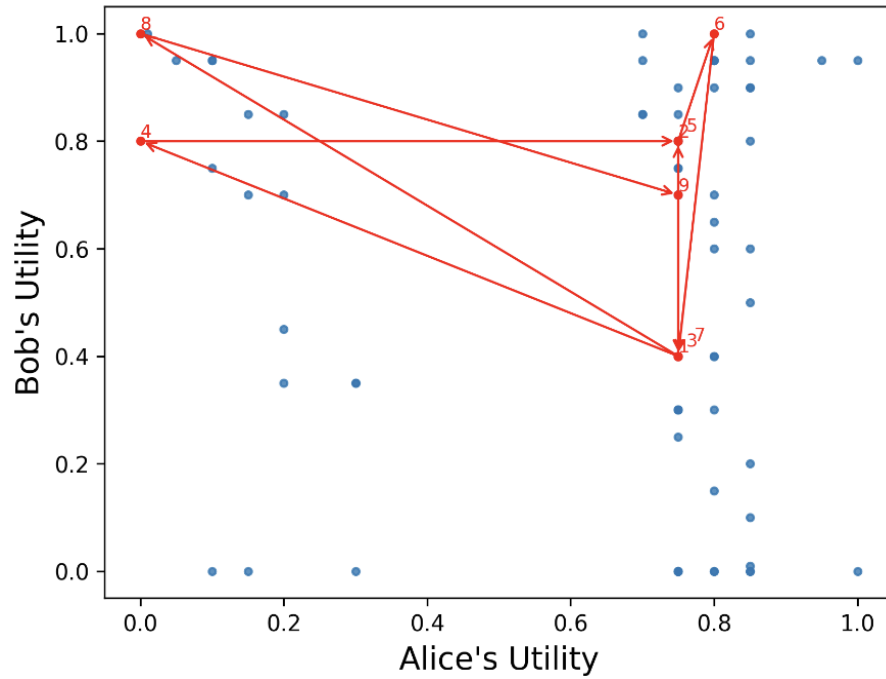
(a) Competitive scenario



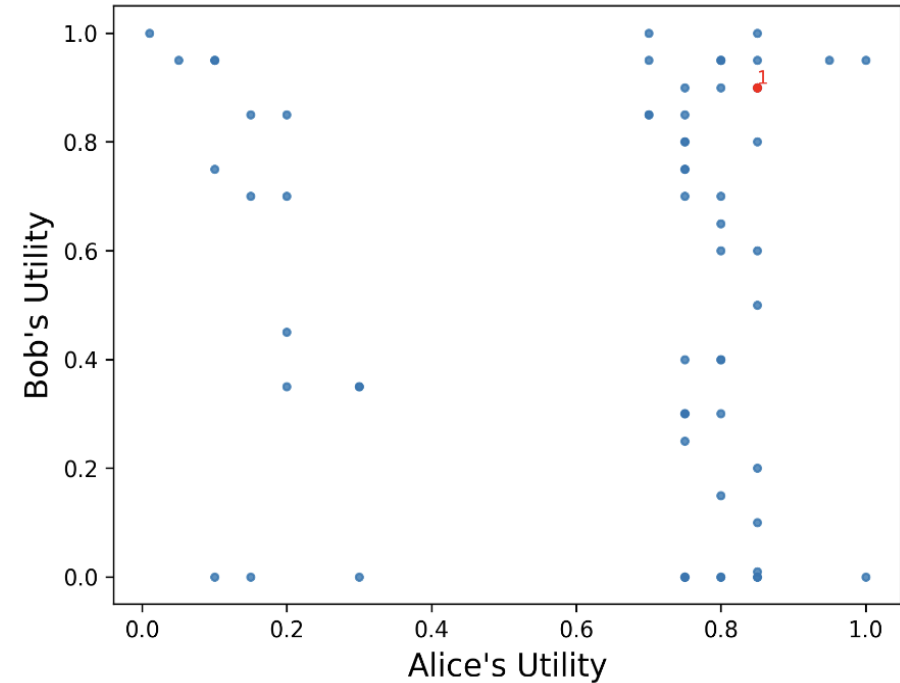
(b) Cooperative scenario

- Many “high-high” allocations in **cooperative** scenario

# Analysis (Cooperative)



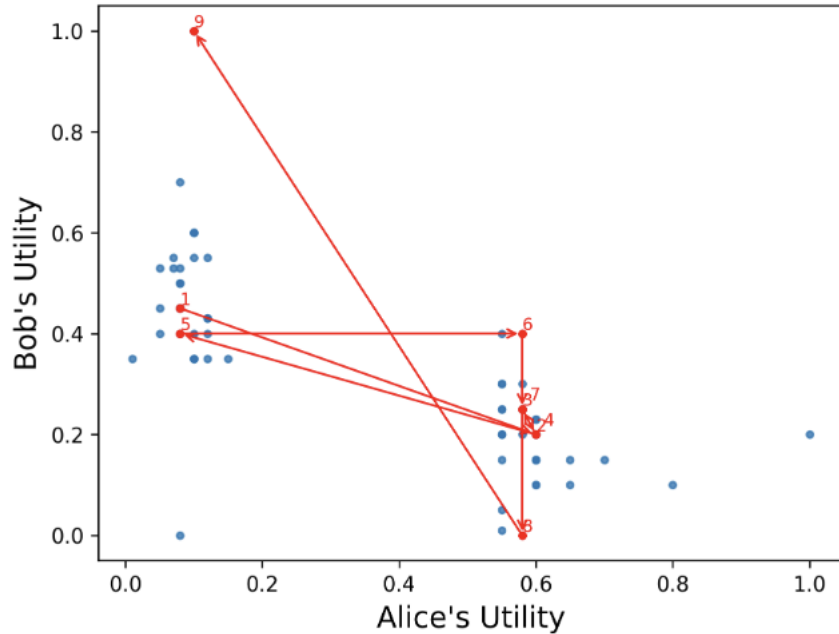
Baseline



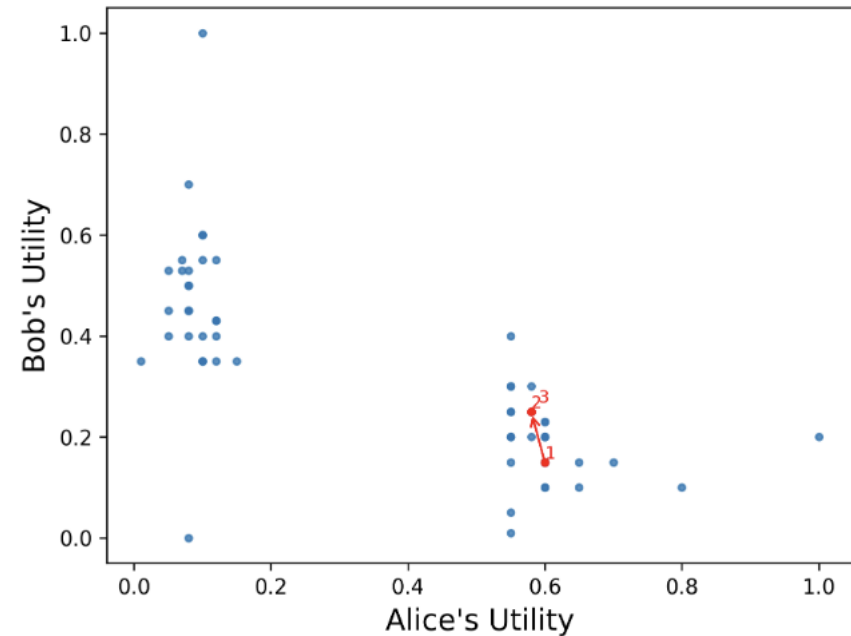
Proposed Framework

- Compromise in the proposed framework

# Analysis (Competitive)



Baseline



Proposed Framework

- Seeking the Pareto front in our framework.

# Conclusion

- Strong models under competitive settings showed fast and reliable convergence.
- We will continue the empirical analysis under a wider range of conditions.