Think, Speak, Decide: Language-Augmented Multi-Agent Reinforcement Learning for Economic Decision-Making

Heyang Ma[1,2,3,4*], **Qirui Mi**[1,5*], **Qipeng Yang**[6,2,3], **Zijun Fan**[6,2,3], **Bo Li**[7], **Haifeng Zhang**[1,2,5†]
[1]Institute of Automation, Chinese Academy of Sciences
[2]Nanjing Artificial Intelligence Research of IA
[3]University of Chinese Academy of Sciences, Nanjing
[4]University of International Business and Economics
[5]School of Artificial Intelligence, Chinese Academy of Sciences
[6]Nanjing University of Posts and Telecommunications
[7]School of Economics, Peking University
haifeng.zhang@ia.ac.cn

Abstract

Economic decision-making depends not only on structured signals—such as prices and taxes—but also on unstructured language, including peer dialogue and media narratives. While multi-agent reinforcement learning (MARL) has shown promise in optimizing economic decisions, it struggles with the semantic ambiguity and contextual richness of language. We propose **LAMP** (**L**anguage-**A**ugmented **M**ulti-Agent **P**olicy), the framework to integrate language into economic decision-making, narrowing the gap to real-world settings. LAMP follows a **Think–Speak–Decide** pipeline: **(1) Think** interprets numerical observations to extract short-term shocks and long-term trends, caching high-value reasoning trajectories. **(2) Speak** crafts and exchanges strategic messages based on the reasoning, updating beliefs by parsing peer communications. **(3) Decide** fuses numerical data, reasoning, and reflections into a MARL policy to optimize language-augmented decision-making. Experiments in economic simulation show that LAMP outperforms both MARL and LLM-only baselines in cumulative return (**+63.5%, +34.0%**), robustness (**+18.8%, +59.4%**), and interpretability. These results demonstrate the potential of language-augmented policies to deliver more effective and robust economic strategies.

Code — https://github.com/hey0223/LAMP

1 Introduction

Real-world economic settings are rich in multi-agent interactions and decision-making challenges, spanning labor markets, firm pricing, and government policy design. Solving these economic decision-making problems can yield explanatory insights into economic phenomena and prescriptive guidance for policy and strategy design (Tversky and Kahneman 1974; Varian and Varian 1992). However, their characteristics—dynamic interactions, long-term incentives, and uncertainty-make them substantially more challenging than conventional fixed-rule benchmarks with fully specified dynamics (Charpentier, Elie, and Remlinger 2023; Mi et al. 2024). Recent advances in artificial intelligence (AI), particularly reinforcement learning (RL), have been applied to model and optimize economic decision-making processes, with applications spanning household savings (Shi 2022a,b; Atashbar and Shi 2023), market pricing (Danassis et al. 2023), and tax policy (Zheng et al. 2022; Mi et al. 2024, 2025c). These studies provide evidence that RL can effectively address dynamic, multi-agent economic problems.

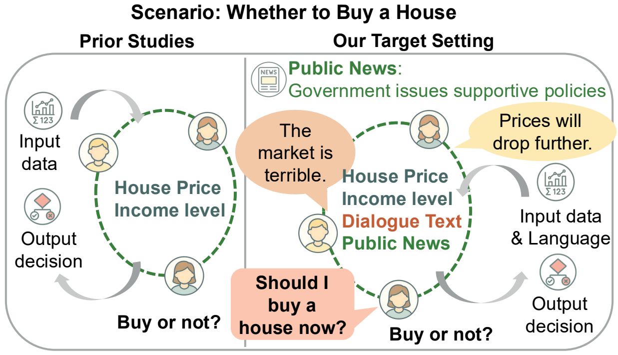

Figure 1. Comparison of prior studies and our target: Unstructured language signals, alongside structured numerical data, are critical to economic decision-making.

However, economic decision-making relies not only on numerical signals but also on language-based information, such as peer dialogue and media narratives (Luketina et al. 2019). The above-mentioned RL-based studies largely ignore the impact of language. Standard MARL algorithms typically assume clean, structured communication protocols (Zhu, Dastani, and Wang 2024), whereas real-world economic decisions involve noisy, semantically rich, and sometimes deceptive natural language. Large language mod-

---

[*] These authors contributed equally.

[†] Corresponding author: haifeng.zhang@ia.ac.cn

els (LLMs) offer powerful tools to process such language. Recent work in policy evaluation (Li et al. 2024; Hao and Xie 2025), trading (Xiao et al. 2024), and simulation (Mi et al. 2025b) demonstrates LLMs' potential for language-aware economic modeling. However, most employ LLMs to generate actions or simulate behaviors, without systematically optimizing agents' policies. This remains insufficient for solving complex economic problems or producing robust, actionable policy insights. We therefore focus on the key question: **In complex multi-agent economic environments, how can agents interpret and leverage natural-language information to support optimal decisions?**

To address this, we propose **LAMP** (**L**anguage-**A**ugmented **M**ulti-**A**gent **P**olicy Learning), which integrates LLM-driven reasoning and reflection over both numerical observations and textual signals to support optimal decision-making. LAMP follows a unified **Think–Speak–Decide** pipeline: (1) **Think**: Agents receive environment observations and generate both short-term shock analysis and long-term trend reasoning. High-reward reasoning trajectories are stored in an experience pool for retrieval. The long-term reasoning is also passed to the Speak module to inform message generation. (2) **Speak**: Guided by the Think module, each agent formulates multiple candidate public messages. A lightweight attention-based scorer selects one for broadcast. Other agents parse the message via the LLM, updating their beliefs, trust, and reflective states. These updated reflections are then passed to the Decide module. (3) **Decide**: The policy network integrates numerical observations, *Think*'s reasoning outputs, and *Speak*'s reflections into the RL policy. Under centralized training with a shared critic, agents learn strategies capable of processing reasoning and reflection signals to produce robust, language-aware economic decisions. We evaluate LAMP in TaxAI and show that it outperforms MARL and LLM-only baselines in both returns and shock robustness. Its reasoning traces explain language-guided choices, aiding insight and policy. **Our contributions are threefold:**

1. **Framework**: We propose LAMP, a language-augmented MARL framework that models the role of natural language in economic decision-making, bringing it closer to real-world contexts.
2. **Mechanism**: We introduce the *Think–Speak–Decide* pipeline, explicitly structuring how agents reason over trends, exchange and interpret strategic messages, and integrate these insights into policy optimization.
3. **Empirical Results**: LAMP surpasses MARL and LLM-only baselines in language-guided decision performance, while providing interpretable reasoning trajectories for transparent policy analysis.

2 Related Work

RL for Economic Decision-Making. Artificial intelligence provides a powerful computational tool for solving complex economic decision-making problems. Early work includes Bayesian structural time series for policy causal inference (Brodersen et al. 2015) and heuristic search for tax design (Małecka-Ziembińska and Ziembiński 2020), but these approaches struggle with real-world complexity. Reinforcement learning (RL) now supports a broad macroeconomic research agenda, including tax policy design such as TaxAI (Mi et al. 2024), monetary rule learning (Chen et al. 2023), trade bargaining (Sch 2021), learning-based heterogeneous-agent macroeconomic modeling (Kuriksha 2021), and large-population policy learning (Zhao et al. 2025; Mi et al. 2025c). At the microeconomic level, RL has modeled household consumption–saving behavior (Shi 2022a,b), responses to income shocks (Atashbar and Shi 2023), and emergent barter and exchange (Johanson et al. 2022; Ozhamaratli and Barucca 2022). While these studies show RL's effectiveness in economic decision-making, they largely ignore language signals—policy debates, media reports, public opinion—thereby oversimplifying real-world settings.

Algorithm 1: Language-Augmented Multi-agent Policy

```
for episode $e = 1, 2, \dots$ do
    Reset environment; clear short experience
    for $t = 0$ to $T$ do
        Determine news type: $type \leftarrow long, short, none$
        Generate news: $\mathcal{R}_t \leftarrow \textbf{Think}(\cdot)_{type}$
        for all agents $i$ do
            Clear the current step's experience $\mathcal{H}^i_{k,t}$
            if $t$ is long-term checkpoint then
                Retrieve $\mathcal{H}^i_{k,t}$ from $\mathcal{H}^{\text{long}}$ and $\mathcal{H}^{\text{short}}_{t,i}$
            Generate economic status and reasoning:
            $\mathcal{L}_{\text{reason}}(\mathcal{R}_t, O^{h,i}_t, \mathcal{H}^i_{k,t})$
            if $t$ is long-term checkpoint then
                Generate statement: $v^i_t \leftarrow \textbf{Speak}(O^{h,i}_t, \mathcal{R}_t)$
                Self-reflection and update belief and trust:
                $\left(w^{i\to j}_t,\ \tau^{i\to j}_t,\ \alpha^i_t\right) \leftarrow \mathcal{L}_{\text{reflect}}(\cdot)$
            Generate action: $a^i_t \leftarrow \mu_{\theta_i}(o^i_t, E_{\text{text}}(v^i_t, \mathcal{R}_t))$
        Execute $a_t$; observe $(r_t, x_{t+1})$; store in $\mathcal{D}$
        Update $Q_\phi$, $\{\theta_i\}$ from $\mathcal{D}$
        Harvest top trajectories $\rightarrow$ short experience $\mathcal{H}^{\text{short}}_{t,i}$
    Harvest top trajectories $\rightarrow$ long experience $\mathcal{H}^{\text{long}}$
```

LLMs for Economic Research. LLMs excel at processing language signals, and recent studies have explored their applications in economics. *Homo Silicus* models human fairness and risk aversion (Horton 2023). *Generative Agents* simulate sandbox societies (Park et al. 2023). *EconAgent* uses LLM agents to evaluate fiscal and monetary policies (Li et al. 2024). Other studies extend LLM agents to population behavior simulation (Mi et al. 2025a), long-term financial planning (Douglas and Verstyuk 2024), and market trading (Xiao et al. 2024; Yu et al. 2024). General platform *EconGym* (Mi et al. 2025b) benchmarks LLM agents in diverse economic scenarios. While these studies demonstrate the versatility of LLMs in economics, most remain focused on direct action generation or simulation, leaving open questions about their role in optimizing economic policies.

Integration of MARL and LLMs. We focus on combining MARL's strength in policy optimization for multi-agent settings with LLMs' capacity to process language signals.

Recent work explores this direction: *FAMA* aligns LLM knowledge for multi-agent coordination (Slumbers et al. 2024); *LAMARL* uses LLM-generated priors for policy and reward design (Zhu et al. 2025); *MAPoRL* co-trains LLMs to enhance cooperation (Park et al. 2025); and *CORY* fine-tunes duplicated LLM agents in cooperative settings (Ma et al. 2024). Economic decision-making is typically dynamic, non-cooperative, and long-horizon. Agents must interpret diverse numerical signals alongside semantically rich and potentially noisy language inputs, rendering prior MARL–LLM methods inadequate for such settings.

3 Language-Augmented Multi-Agent Policy

This section first presents a mathematical formulation of the language-augmented multi-agent decision-making problem in economic environments (Section 3.1) and then details our proposed LAMP framework (Section 3.2).

3.1 Problem Formulation

We formulate the economic decision-making problem with language involvement. Building on the economic modeling in TaxAI (Mi et al. 2024), we incorporate language by augmenting each household's observation as

$$m_t^i = \mathcal{E}\big(\mathcal{L}(a_t^i, e_t^i, O_t^g)\big)$$

Here, $\mathcal{L}$ denotes a large language model producing a textual message from inputs, and $\mathcal{E}$ denotes an embedding model that maps this text into $\mathbb{R}^n$. For inputs, all agents share a global observation O_t^g. The government observes $O_t^g = \{W_t,\, \bar{a}_t^{r,p},\, \bar{i}_t^{r,p},\, \bar{e}_t^{r,p}\}$, where W_t denotes the wage, and the remaining terms are group-level averages of assets, income, and efficiency. Each household i observes the same O_t^g and, in addition, its private asset a_t^i and efficiency e_t^i.

We then model the economic decision-making problem as a partially observable Markov game $\mathcal{M} = \langle N, S, O, A, P, R, \delta\rangle$, where $N = \{g, 1, \ldots, N_h\}$, $\delta \in [0, 1)$, and P is the transition kernel induced by $A = A^g \times A^{h,1} \times \cdots \times A^{h,N_h}$. At each step, the government's action is $A_t^g = \{\tau_t,\ \xi_t,\ \tau_{a,t},\ \xi_{a,t},\ r_t^G\}$, where τ_t and ξ_t parameterize the marginal income-tax schedule, $\tau_{a,t}$ and $\xi_{a,t}$ analogously parameterize the marginal asset-tax schedule, and r_t^G denotes the expenditure-to-output ratio. Each household i selects a savings rate and labor supply $h_t^i \in [0, h_{\max}]$: $A_t^{h,i} = \{p_t^i, h_t^i\}$.

The government policy π_g and household policies π_i map their observations to actions. The household's objective is to maximize lifetime utility from consumption and leisure, with consumption increasing utility and labor hours reducing it:

$$\max\ \mathbb{E}_0 \sum_{t=0}^{T_N} \beta^t \left(\frac{c_t^{1-\eta}}{1-\eta} - \frac{h_t^{1+\gamma}}{1+\gamma} \right)$$

$$\text{s.t.} \quad (1+\tau_s)c_t + a_{t+1} = i_t - T(i_t) + a_t - T^a(a_t)$$

where c_t and h_t are consumption and labor, β is the discount factor, η is the relative risk aversion coefficient, and γ is the inverse Frisch elasticity.

The government's objective is GDP growth; the government remains as in TaxAI, full details are provided in the extended version.

Symbol	**Description**
Economic Variables	
N_h	Number of households
O_t^g	Government observation (wage, group averages)
$a_t^i,\ e_t^i$	Asset, efficiency of household i
$c_t,\ h_t$	Consumption, labor
$\beta,\ \eta,\ \gamma$	Discount, risk aversion, Frisch elasticity
$Y_t,\ G_t,\ B_t,\ T_t$	GDP, spending, debt, tax
Framework Variables	
$\mathcal{X}_t$	Global indicators (Gini, welfare, GDP)
$\mathcal{L}, \mathcal{E}$	Language model, Embedding model
$\sigma,\ L_i$	Shock threshold, long-term step size
$\mathcal{R}_t^s,\ \mathcal{R}_{L_i}^l$	Short-/long-term news
$\mathcal{H}^s,\ \mathcal{H}^l$	Short-/long-term experience
$\psi_t^i,\ V_t$	Reasoning, public statements
$m_t^i,\ x_t$	Embedding, fused state

Table 1. Key symbols in the economic problem and LAMP.

3.2 LAMP Framework

To address the above problem, we propose the LAMP framework (see Pseudocode 1), which comprises three modules:

Think *Think* translates global numerical signals into shared news, providing both short- and long-term economic interpretations to guide agents' reasoning and dialogue. At fixed checkpoints L_i, it issues **long-term news** capturing structural trends. Whenever a key indicator $\mathcal{X}_t = \big(G_w,\, \mathcal{W},\, Y\big)$ —wealth Gini G_w, social welfare $\mathcal{W}$, or per-capita GDP Y—changes by more than a threshold σ, it broadcasts a **short-term shock**. Then the news type is:

$$\text{type}(t) = \begin{cases} \text{long}, & t \in \{L_1, \ldots, L_n\}, \\ \text{short}, & \max\limits_j \big|\mathcal{X}_{j,t} - \mathcal{X}_{j,t-1}\big| > \sigma, \\ \text{none}, & \text{otherwise}. \end{cases}$$

This design ensures agents receive timely, context-rich updates—similar to how real-world economic actors rely on news outlets—rather than raw numerical data.

A shared LLM-driven news service synthesizes appropriate texts $\mathcal{R}_t^{\text{short}}$ or $\mathcal{R}_{L_i}^{\text{long}}$ and disseminates them to all agents. Short-term news is generated as:

$$\mathcal{R}_t^{\text{short}} = \mathcal{L}_{\text{S}}\Big(O_t^g,\, O_{t-1}^g,\, \mathcal{R}_{L_k}^{\text{long}}\Big), \quad L_k < t < L_{k+1}$$

incorporating the current and previous global observations, as well as the most recent long-term news. Long-term news is generated over a two-step observation window:

$$\mathcal{R}_{L_i}^{\text{long}} = \mathcal{L}_{\text{L}}\big(O_{L_i-1:L_i}^g\big), \quad i = 1, 2, \ldots, n$$

Upon receiving short-term news, each agent infers its economic status $\kappa_t^i \in \{0, 1, 2\}$ (good / neutral / poor) and produces a private reasoning ψ_t^i. Long-term news additionally triggers the Experience Pool and Speak module for deeper reasoning. After each short-term reasoning phase, agent i ranks candidate reasoning trajectories by reward and stores its top k_1 reasoning trajectories into a *short-term* buffer:

$$\mathcal{H}_{t,i}^{\text{short}} = \text{Top}_{k_1}(\mathcal{T}_i)$$

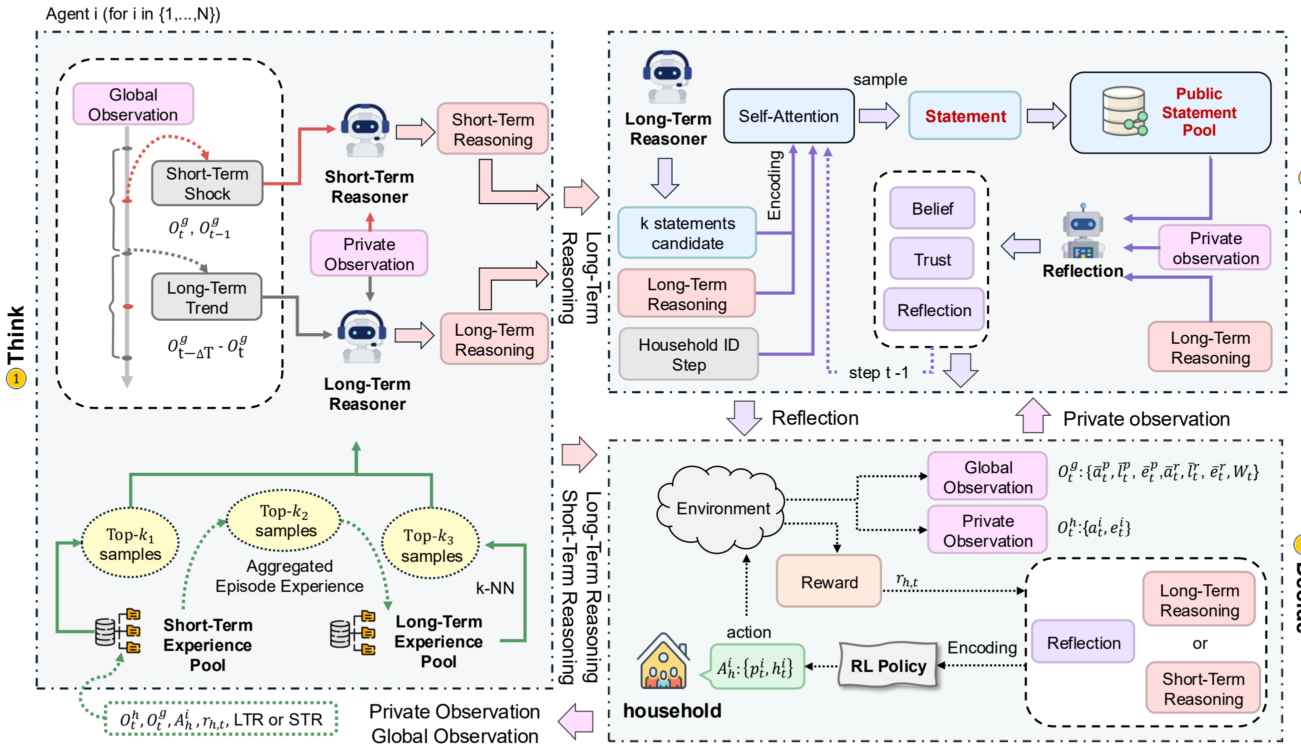

Figure 2. Workflow of LAMP: (a) Dual-path Think module extracts long-term trends and short-term shocks into compact reasoning embeddings; (b) Speak module applies self-attention to sample and broadcast a single message and performs a reflection step to update beliefs; (c) Decide module's policy network concatenates numeric observations with language and reflection embeddings to select actions.

At each long-term checkpoint k, the system collects the top k_2 trajectories (by reward) across all agents and appends them to the *long-term* FAISS index:

$$\mathcal{H}_k^{\text{long}} = \mathcal{H}_{k-1}^{\text{long}} \cup \text{Top}_{k_2}\Big(\bigcup_{i=1}^{N_h} \mathcal{T}_i\Big)$$

Before the next long-term reasoning step, agent i retrieves the k_3 nearest neighbors from $\mathcal{H}_k^{\text{long}}$ using FAISS, where similarity is computed against a query embedding derived from its current observation $O_t^{h,i}$, and merges them with its current $\mathcal{H}_{t,i}^{\text{short}}$. This combined set of past high-reward insights is then used as contextual prompts for the LLM:

$$\mathcal{H}_{k,t}^i = \text{kNN}_{k_3}(\mathcal{H}_k^{\text{long}}) \cup \mathcal{H}_{t,i}^{\text{short}}$$

allowing the agent to remember and reuse successful strategies in similar future scenarios.

Speak Building on the news from *Think* and each agent's private reasoning, *Speak* produces a concise strategic statement per agent, broadcasts it to peers, and returns language-based peer assessments for the next reasoning step.

Inspired by (Xu et al. 2023), the LLM generates three candidate statements for agent i; a self-attention selector $\mathcal{S}$ scores them to form a distribution $p_t^{i,\cdot}$, from which one statement is sampled and broadcast to all agents. Let V_t denote the multiset of broadcast statements. After broadcasting and receiving messages V_t, each agent i uses a Reflection Module $\mathcal{L}_{reflect}$ to interpret the content. This produces an assessment of each peer j, including an estimated wealth tier ($w_t^{i\mapsto j} \in \{\text{low, mid, high}\}$) and a numeric belief confidence $\tau_t^{i\mapsto j} \in [0, 10]$. The evaluator also generates a brief self-reflection α_t^i summarizing agent i's own situation:

$$(w_t^{i\mapsto j},\, \tau_t^{i\mapsto j},\, \alpha_t^i) = \mathcal{L}_{\text{reflect}}(O_t^{h,i}, V_t, \psi_t^i)$$

These peer assessments are fed back to $\mathcal{S}$ and the LLM policy to guide the next round of reasoning and candidate selection, closing a loop that links language reasoning, dialogue, and adaptive coordination.

Decide Consuming language embeddings from *Think* and *Speak* together with numeric observations, *Decide* compresses language vectors and maps the enriched state to actions under centralized training with decentralized execution (CTDE). All texts (private reasoning and reflection) are encoded by a text encoder $\mathcal{E}_{\text{text}}$, pooled into a fixed-length vector h_t^i, and passed through a small projection $P : \mathbb{R}^D \to \mathbb{R}^d$ for dimensionality reduction and feature alignment:

$$\tilde{m}_t^i = \frac{P(h_t^i)}{\|P(h_t^i)\|_2} \in \mathbb{R}^d.$$

Unless otherwise noted, gradients do not flow into $\mathcal{E}_{\text{text}}$ (the encoder is frozen for stability). At time t, the global observa-

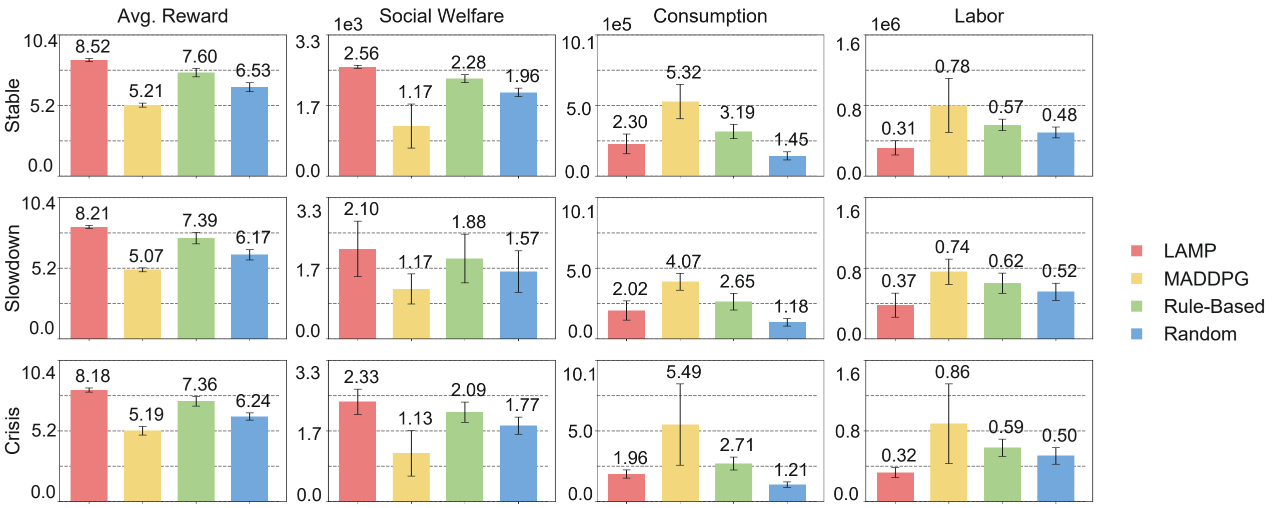

Figure 3. Across three economic environments, LAMP outperforms non-language baselines (Random, rule-based, MADDPG) with higher social welfare and consumption, lower welfare variance, and similar labor usage.

tion is concatenated with household language embeddings to form $x_t = \left(O_t^g, m_t^{1:N_h}\right)$, which, together with the joint action a_t, is stored in the replay buffer $\mathcal{D}$. We adopt a standard MADDPG framework (Lowe et al. 2017), where a centralized critic minimizes Bellman error, and decentralized actors update their policies by maximizing the expected Q-value via deterministic policy gradients. Full optimization details and loss formulations are provided in the extended version.

4 Experiments

Our experiments address two key questions:

1. **How effective is LAMP?** (§ 4.2) We compare LAMP with non-language and LLM-based baselines across 3 economic scenarios to evaluate its performance.
2. **What drives LAMP's gains?** (§ 4.3) We remove core modules of LAMP to assess their contribution to performance and stability.

4.1 Experimental Setup

Environment All experiments are conducted in **TaxAI** (Mi et al. 2024), a dynamic economic simulator. It models complex economic interactions between heterogeneous households and a government, and is calibrated with real-world data—making it a realistic and challenging testbed for economic decision-making.

Evaluation Metrics We evaluate LAMP and baselines with five metrics: (1) **Average Household Reward** — mean reward per step across households; (2) **Social Welfare**: sum of utilities across all households over the horizon; (3) **Total Consumption**: aggregate consumption of households; (4) **Total Labor**: aggregate labor supply in an economy; and (5) **Years**: number of simulated years before collapse (max 300, higher indicates greater stability). *Total Consumption* and *Total Labor* do not directly measure policy performance, but help analyze policy preferences.

Baselines. We benchmark LAMP against two baseline categories with identical training budgets and horizons. All LLM-based baselines use the same backbone (**Qwen2.5-72B-Instruct-INT4**) and prompts. We compare different language models in the extended version.

(1) Conventional Baselines: Random: Agents select actions uniformly at random. **Rule-Based**: Economic method based on the utility–production model (details in extended version). **MADDPG**: Multi-Agent Deep Deterministic Policy Gradient (Lowe et al. 2017). We also compare different MARL algorithms in the extended version.

(2) LLM-based Baselines: Only-LLM: Directly query an LLM to generate actions. **CoT / ReAct / Reflexion**: LLM reasoning methods using CoT (Wei et al. 2022), ReAct (Yao et al. 2023), or Reflexion (Shinn et al. 2023).

4.2 How effective is LAMP?

We evaluate LAMP and baselines under three settings:

- **Economic Stability (S1):** Matches training conditions, representing a stable macroeconomic scenario.
- **Economic Slowdown (S2):** Introduces a moderate shift, simulating reduced growth and mild market stress.
- **Crisis Shock (S3):** Applies a large, coupled shift, modeling severe economic shocks for robustness evaluation.

Detailed setup is provided in the extended version.

Quantifying Gains over LLM-based Baselines. LAMP also outperforms language-integrated baselines, demonstrating the advantage of combining MARL with language-guided policy optimization. In **S1**, using the same backbone and prompt budget, LAMP surpasses the strongest language baseline (ReAct) with **+14.8%** higher welfare and **+14.5%** higher reward, while reducing consumption and labor. Under distribution shifts, the advantage remains: in **S2** and **S3**, welfare gains are **+1.0%** and **+10.4%**, reward gains are **+18.1%** and **+16.0%**, with corresponding reductions in

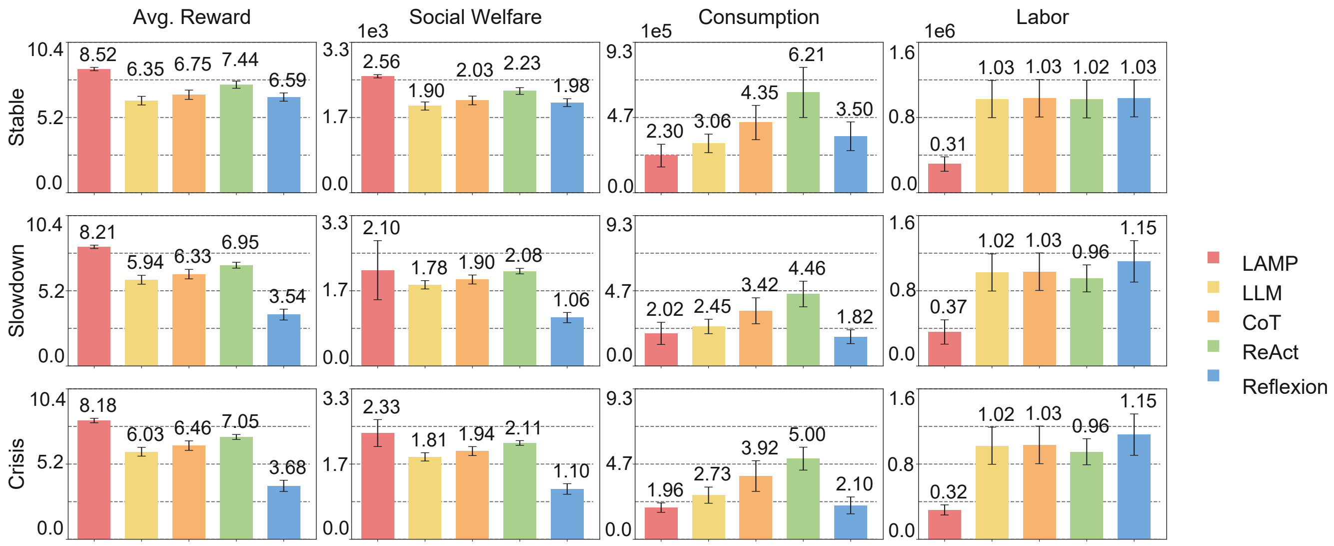

Figure 4. LAMP vs. other language-based agents (Only-LLM, CoT, ReAct, Reflexion) on the same metrics across the three economic environments. LAMP outperforms all these LLM-driven baselines, obtaining higher social welfare and consumption and generally lower welfare variance in each environment.

Category	Algorithms	Avg. Reward (↑)	Social Welfare (↑)	Consumption (-)	Labor (-)
Ours	**LAMP**	**8.52 ± 0.13**	**2.56e+03 ± 3.77e+01**	2.30e+05 ± 7.52e+04	3.13e+05 ± 8.46e+04
Conventional	**MADDPG**	5.21 ± 0.16	1.17e+03 ± 5.51e+02	5.32e+05 ± 1.31e+05	7.82e+05 ± 3.20e+05
	Rule-Based	7.60 ± 0.33	2.28e+03 ± 9.99e+01	3.19e+05 ± 5.46e+04	5.68e+05 ± 6.73e+04
	Random	6.53 ± 0.35	1.96e+03 ± 1.06e+02	1.45e+05 ± 3.10e+04	4.84e+05 ± 6.41e+04
LLM-based	**LLM-Only**	6.35 ± 0.32	1.90e+03 ± 9.56e+01	3.06e+05 ± 6.14e+04	1.03e+06 ± 2.18e+05
	CoT	6.75 ± 0.34	2.03e+03 ± 1.03e+02	4.35e+05 ± 1.14e+05	1.03e+06 ± 2.19e+05
	ReAct	7.44 ± 0.26	2.23e+03 ± 7.92e+01	6.21e+05 ± 1.66e+05	1.02e+06 ± 2.21e+05
	Reflexion	6.59 ± 0.31	1.98e+03 ± 9.16e+01	3.50e+05 ± 9.51e+04	1.03e+06 ± 2.16e+05

Table 2. Comparison of LAMP with conventional and LLM-based baselines in the real-data–calibrated environment (S1: Economic Stability). Results for S2 and S3 appear in the extended version. Values are mean ± SD. Notation: (↑) higher is better; (–) non-monotonic. Consumption and Labor jointly shape household utility with non-monotonic effects.

consumption and labor. These results confirm that LAMP's language-guided coordination improves both stability and efficiency, even in stressed economic conditions.

Isolating Language Effects. LAMP consistently outperforms non-language baselines, demonstrating the benefit of language integration in economic decision-making. In **S1**, LAMP achieves the highest *Social Welfare* and *Average Household Reward*. Compared to the strongest non-language baseline (Rule-Based), welfare improves by **+12.3%** and reward by **+12.1%**; relative to numeric MARL (MADDPG), gains reach **+118.8%** and **+63.5%**, respectively. Efficiency gains are evident from lower *Consumption* and *Labor*. Versus Rule-Based, LAMP uses –27.9% consumption and –44.9% labor (vs. MADDPG: –56.8% and –60.0%), suggesting that higher welfare stems from efficiency rather than brute-force spending or overwork. Under **S2** and **S3**, LAMP consistently outperforms the baselines.

Analysis and Insights. We share **interesting findings** from experiments, supported by LLM outputs:

(1) Economic decision-making involves many interdependent variables that change frequently, with causal links often unclear. Purely data-driven MARL starts from scratch, fitting policies without explicit understanding of these variables, making optimal policy search slow and uncertain.

(2) LAMP addresses this by using LLM reasoning at each step to extract concise, high-value insights, which are then passed to the MARL component. These structured signals—hard for pure data-driven methods to obtain—are readily produced by LLMs. Representative examples illustrate the LLM's clear interpretation of economic variables and targeted reasoning that enhance decision-making. More examples are shown in the extended version.

4.3 What drives LAMP's gains?

Speak Module: Strategy Communication & Opponent Modeling. The *Speak* module enables agents to exchange

Ablation Setting	Avg. Reward (↑)	Social Welfare (↑)	Consumption (-)	Labor (-)	Years (↑)
LAMP (Ours)	**8.52**	**2.56e+03**	2.30e+05	3.13e+05	**3.00e+02**
w/o Speak	8.42 (−1%)	2.53e+03 (−1%)	3.24e+05 (+41%)	5.36e+05 (+71%)	**3.00e+02** (+0%)
w/o Experience Pool	8.45 (−1%)	1.25e+03 (−51%)	5.12e+05 (+122%)	4.50e+05 (+44%)	1.50e+02 (−50%)
w/o Long-Term	5.31 (−38%)	1.15e+03 (−55%)	2.27e+05 (−2%)	4.10e+05 (+31%)	2.19e+02 (−27%)
w/o Short-Term	8.18 (−4%)	1.67e+03 (−35%)	3.51e+05 (+53%)	5.25e+05 (+68%)	2.08e+02 (−30%)
w/o Timing Scheduler	**8.52** (−0%)	1.19e+03 (−53%)	3.48e+05 (+51%)	5.70e+05 (+82%)	1.41e+02 (−53%)

Table 3. Ablation under the baseline economy. Percentages denote change vs. LAMP (Ours). Notation: (↑) higher is better; (–) non-monotonic. Consumption and Labor jointly shape household utility with non-monotonic effects.

strategic messages and infer others' states, providing the coordination essential for high performance. Removing it causes a 1.2% welfare drop alongside sharp increases in labor and consumption. This indicates that, without strategic communication, agents compensate through brute-force effort. With Speak enabled, comparable or higher welfare is achieved with far less input. Representative outputs show the mechanism: after detecting widening inequality and low wages, the LLM revises beliefs toward demand fragility and restraint, then recommends disciplined actions such as moderating labor expansion and investing in human capital, thereby reducing overshooting and volatility.

Experience Pool: Enhancing Stability and Efficiency. The experience pool substantially improves efficiency and stability. Removing it **cuts social welfare by 50.9**% and average household reward by 0.8%, while labor rises 43.6% and consumption surges 122.4%. The unexpected jump in consumption suggests that, without stored successful trajectories, agents overshoot spending and output, oscillating in search of workable strategies. Stability also deteriorates, with **50.2% fewer simulated years** sustained before failure. Beyond performance, the pool improves interpretability by preserving reasoning traces as an auditable knowledge base explaining why certain strategies are followed.

Reasoning Paths: Trend Tracking & Shock Response. *Long-term reasoning* is essential for capturing structural trends. Removing it **drops average household reward by 37.7%** and reduces stable years from 300 to 219. Without long-term reasoning, agents become myopic, reacting only to immediate stimuli and producing unstable policies.

Short-term reasoning supports rapid adjustment to shocks. Disabling it has a moderate effect on final returns (**–3.99% reward**) but significantly harms efficiency: labor rises 67.7%, consumption 52.7%, and stable years fall from 300 to 208.

Trigger timing is critical. Random triggers keep welfare similar but collapse efficiency: labor increases 81.9%, consumption 51.2%, and stable years decrease from **300** to **141**. This shows aligning reasoning with actual needs reduces turbulence and sustains consistent performance.

We observe an adaptive policy shift in LLM outputs: upon detecting rising inequality—top 10% volatility widening and bottom 50% declining—the LLM revised its earlier "work more" stance. It recommended slightly reducing work hours, increasing savings, delaying non-essential spending, and investing in skills for long-term stability, while publicly supporting progressive taxation and minimum wages.

Representative LLM Reasoning and Experience

`Short-term`

Reasoning: "... The family's personal productivity (0.7741) and wealth (0.0957) place them in a vulnerable position. Given the volatility and risk of instability, the economic status is rated as 'Bad'."

Economic status: `0`

`Long-term`

Statement: "We should advocate for policies that promote fair wage growth and equitable wealth distribution to stabilize the broader economic environment and ..."

Reasoning: "The family should avoid overwork and instead focus on savings, education, and..."

Reflection: "...highlight the importance of balancing increased labor time... Investing in education and advocating for fairness improves resilience and security."

Economic status: `1`

Belief: `[0, 1, 0, 0, 0, 1, 1, 1, 0, 2]`

Trust: `[8, 9, 9, 8, 9, 8, 9, 9, 8, 10]`

`Experience`

ID=Household1, Reward=0.95, Personal productivity(e): 1.846, Personal wealth: 0.196, savings ratio:-0.947, working time ratio:-0.963, Reasoning: "..."

5 Conclusion

This paper introduced the **Language-Augmented Multi-Agent Policy (LAMP)** framework, offering a new approach to complex economic decision-making. LAMP leverages LLM reasoning and reflection over language signals—such as peer dialogue and public news—alongside numerical data to inform optimal policies. The framework follows a *Think–Speak–Decide* pipeline: agents extract short-term shocks and long-term trends, communicate strategic insights, and execute language-informed policies. **Experiments demonstrate LAMP's strong performance and reveal interesting insights**: LLM reasoning and reflection dynamically distill key information from numerous, volatile economic variables, enabling agents to make efficient decisions.

This contrasts with fully data-driven methods that search for optimal solutions from scratch—a process particularly challenging in economics. We hope this work offers novel methods and insights for AI in economic decision-making.

Acknowledgments

We sincerely thank Prof. Hao Huang from the University of International Business and Economics for his valuable guidance and insightful suggestions during the early stage of this work.

This work was supported in part by the National Natural Science Foundation of China under the Original Exploration Program (Grant No. 72450002).

References

Atashbar, T.; and Shi, R. A. 2023. *AI and macroeconomic modeling: Deep reinforcement learning in an RBC model*. International Monetary Fund.

Brodersen, K. H.; Gallusser, F.; Koehler, J.; Remy, N.; and Scott, S. L. 2015. Inferring causal impact using Bayesian structural time-series models.

Charpentier, A.; Elie, R.; and Remlinger, C. 2023. Reinforcement learning in economics and finance. *Computational Economics*, 62(1): 425–462.

Chen, M.; Joseph, A.; Kumhof, M.; Pan, X.; and Zhou, X. 2023. Deep Reinforcement Learning in a Monetary Model. arxiv:2104.09368.

Danassis, P.; Filos-Ratsikas, A.; Chen, H.; Tambe, M.; and Faltings, B. 2023. AI-driven Prices for Externalities and Sustainability in Production Markets. arxiv:2106.06060.

Douglas, M. R.; and Verstyuk, S. 2024. Consumption and Savings with Large Language Model Agents. *Available at SSRN 4909749*.

Hao, Y.; and Xie, D. 2025. A Multi-LLM-Agent-Based Framework for Economic and Public Policy Analysis. *arXiv preprint arXiv:2502.16879*.

Horton, J. J. 2023. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research.

Johanson, M. B.; Hughes, E.; Timbers, F.; and Leibo, J. Z. 2022. Emergent Bartering Behaviour in Multi-Agent Reinforcement Learning. arxiv:2205.06760.

Kuriksha, A. 2021. An Economy of Neural Networks: Learning from Heterogeneous Experiences. arxiv:2110.11582.

Li, N.; Gao, C.; Li, M.; Li, Y.; and Liao, Q. 2024. EconAgent: Large Language Model-Empowered Agents for Simulating Macroeconomic Activities. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 15523–15536.

Lowe, R.; Wu, Y. I.; Tamar, A.; Harb, J.; Pieter Abbeel, O.; and Mordatch, I. 2017. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30.

Luketina, J.; Nardelli, N.; Farquhar, G.; Foerster, J.; Andreas, J.; Grefenstette, E.; Whiteson, S.; and Rocktäschel, T. 2019. A survey of reinforcement learning informed by natural language. *arXiv preprint arXiv:1906.03926*.

Ma, H.; Hu, T.; Pu, Z.; Liu, B.; Ai, X.; Liang, Y.; and Chen, M. 2024. Coevolving with the Other You: Fine-Tuning LLM with Sequential Cooperative Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems (NeurIPS) 2024*.

Małecka-Ziembińska, E.; and Ziembiński, R. 2020. Application of genetic algorithm to optimal income taxation. *Journal of Risk and Financial Management*, 13(11): 251.

Mi, Q.; Xia, S.; Song, Y.; Zhang, H.; Zhu, S.; and Wang, J. 2024. TaxAI: A Dynamic Economic Simulator and Benchmark for Multi-agent Reinforcement Learning. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, 1390–1399.

Mi, Q.; Yang, M.; Yu, X.; Zhao, Z.; Deng, C.; An, B.; Zhang, H.; Chen, X.; and Wang, J. 2025a. MF-LLM: Simulating Population Decision Dynamics via a Mean-Field Large Language Model Framework. In *Advances in Neural Information Processing Systems (NeurIPS) 2025*.

Mi, Q.; Yang, Q.; Fan, Z.; Fan, W.; Ma, H.; Ma, C.; Xia, S.; An, B.; Wang, J.; and Zhang, H. 2025b. EconGym: A Scalable AI Testbed with Diverse Economic Tasks. In *Advances in Neural Information Processing Systems (NeurIPS) 2025*.

Mi, Q.; Zhao, Z.; Ma, C.; Xia, S.; Song, Y.; Yang, M.; Wang, J.; and Zhang, H. 2025c. Learning Macroeconomic Policies through Dynamic Stackelberg Mean-Field Games. In *28th European Conference on Artificial Intelligence (ECAI) 2025*.

Ozhamaratli, F.; and Barucca, P. 2022. Deep Reinforcement Learning for Optimal Investment and Saving Strategy Selection in Heterogeneous Profiles: Intelligent Agents Working towards Retirement. arxiv:2206.05835.

Park, C.; Han, S.; Guo, X.; Ozdaglar, A. E.; Zhang, K.; and Kim, J.-K. 2025. Maporl: Multi-agent post-co-training for collaborative large language models with reinforcement learning. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 30215–30248.

Park, J. S.; O'Brien, J.; Cai, C. J.; Morris, M. R.; Liang, P.; and Bernstein, M. S. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22.

Sch, A. A. O. 2021. Intelligence in the Economy: Emergent Behaviour in International Trade Modelling with Reinforcement Learning.

Shi, R. A. 2022a. Can an AI agent hit a moving target? arXiv:2110.02474.

Shi, R. A. 2022b. Learning from zero: how to make consumption-saving decisions in a stochastic environment with an AI algorithm. arXiv:2105.10099.

Shinn, N.; Cassano, F.; Gopinath, A.; Narasimhan, K.; and Yao, S. 2023. Reflexion: Language agents with verbal reinforcement learning. *Advances in neural information processing systems*, 36: 8634–8652.

Slumbers, O.; Mguni, D. H.; Shao, K.; and Wang, J. 2024. Leveraging Large Language Models for Optimised Coordination in Textual Multi-Agent Reinforcement Learning.

Tversky, A.; and Kahneman, D. 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157): 1124–1131.

Varian, H. R.; and Varian, H. R. 1992. *Microeconomic analysis*, volume 3. Norton New York.

Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q. V.; Zhou, D.; et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837.

Xiao, Y.; Sun, E.; Luo, D.; and Wang, W. 2024. TradingAgents: Multi-Agents LLM Financial Trading Framework. *arXiv preprint arXiv:2412.20138*.

Xu, Z.; Yu, C.; Fang, F.; Wang, Y.; and Wu, Y. 2023. Language agents with reinforcement learning for strategic play in the werewolf game. *arXiv preprint arXiv:2310.18940*.

Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K. R.; and Cao, Y. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. In *The Eleventh International Conference on Learning Representations*.

Yu, Y.; Yao, Z.; Li, H.; Deng, Z.; Jiang, Y.; Cao, Y.; Chen, Z.; Suchow, J.; Cui, Z.; Liu, R.; et al. 2024. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. *Advances in Neural Information Processing Systems*, 37: 137010–137045.

Zhao, Z.; Ma, C.; Mi, Q.; Yang, N.; Yan, X.; Yang, M.; Zhang, H.; Wang, J.; and Yang, Y. 2025. Mean Field Correlated Imitation Learning. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, 2364–2372.

Zheng, S.; Trott, A.; Srinivasa, S.; Parkes, D. C.; and Socher, R. 2022. The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Science Advances*, 8(18): eabk2607.

Zhu, C.; Dastani, M.; and Wang, S. 2024. A survey of multi-agent deep reinforcement learning with communication. *Autonomous Agents and Multi-Agent Systems*, 38(1): 4.

Zhu, G.; Zhou, R.; Ji, W.; and Zhao, S. 2025. LAMARL: LLM-Aided Multi-Agent Reinforcement Learning for Cooperative Policy Generation. *IEEE Robotics and Automation Letters*, 10(7): 7476–7483.