A Multi-agent Reinforcement Learning Study of Evolution of Communication and Teaching under Libertarian and Utilitarian Governing Systems

Anonymous Author(s)
Affiliation
Address
email

Abstract

Laboratory experiments have shown that communication plays an important role in solving social dilemmas. Here, by extending the AI-Economist, a mixed motive multi-agent reinforcement learning environment, we intend to find an answer to the following descriptive question: which governing system does facilitate the emergence and evolution of communication and teaching among agents? To answer this question, the AI-Economist is extended by a voting mechanism to simulate three different governing systems across individualistic-collectivistic axis, from Full-Libertarian to Full-Utilitarian governing systems. In the original framework of the AI-Economist, agents are able to build houses individually by collecting material resources from their environment. Here, the AI-Economist is further extended to include communication with possible misalignment –a variant of signalling game –by letting agents to build houses together if they are able to name mutually complement material resources by the same letter. Moreover, another extension is made to the AI-Economist to include teaching with possible misalignment –again a variant of signalling game –by letting half the agents as teachers who know how to use mutually complement material resources to build houses but are not capable of building actual houses, and the other half as students who do not have this information but are able to actually build those houses if teachers teach them. The result shows that collectivistic environment such as Full-Utilitarian system is more favourable for the emergence of communication and teaching, or more precisely, evolution of language alignment. Finally, a discussion is provided to justify this result in the simulation environment of this paper.

1 Introduction

Multi-agent reinforcement learning (MARL) comprised of multiple decision making units each one is able to observe the state of and act upon environment to achieve its goal. An agent changes the state of the environment by taking an action and then receives a reward and a new observation. This loop continues until the agent achieves its goal or the time-step of the environment reaches to its maximum limit (Albrecht et al. (2023)). The AI-Economist is a two-level MARL framework (Zheng et al. (2022)), comprised of one single agent as a rational social planner who designs a particular mechanism or policy generally having a goal of optimising a particular kind of social welfare functions in the society. The other agents are a set of rational economic mobile agents who behave in response to the implemented mechanism or policy generally following their own self-interests. This framework has been used to model the tax-gaming behaviour of agents - optimising their labours, trading, and building, while the central social planner maximises productivity or equality in the society (Zheng et al. (2022)). More precisely, the agents in the Gather-Trade-Build environment of the AI-Economist make efforts to move, gather wood and stone from the environment, trade them with each other via double-auctions using coins as a mean of exchange, and finally –contingent on their build-skill –build houses to earn incomes. On the other hand, the social planner aims to find an optimised taxing schedule to increase productivity or equality in the society (Zheng et al. (2022)). The agenda of this framework is descriptive meaning that the goal is to simulate the actual behaviour of a population of humans. The number of mobile agents is between 2 and 10 which is a reasonable choice in MARL. The game is a mixed motive partially observable stochastic game with simultaneous cooperation and competition. As a result, there are two challenges of heterogeneous incentives and the difficulty of defining a suitable collective reward function (Du et al. (2023)). The agents share their weights in a centralised training and decentralised execution by having their own set of observations. The non-stationary of the multiple learning agents is partially overcome by curriculum learning and entropy regularisation, while the optimality of the final selected equilibrium is partially confirmed by letting the environment to go through a very large number of time-steps. Due to complexity of the environment, a two-level Proximal Policy Optimisation (PPO) gradient method as a deep reinforcement learning technique is used to solve the equations.

The main question of this paper is that which governing system is more favourable for the evolution of communication and teaching through language alignment. To this end, we extends the AI-Economist to model a governing system along individualistic-collectivistic axis –from Full-Libertarian to Full-Utilitarian governing systems –through a voting mechanism. Previously, in MARL literature, it has been shown that communication can enhance exploration, maximise reward, and diversify solutions in complex optimisation simulations (Du et al. (2023)). Also, human experiments show that the essence of communication even without any enforcement is more effective than the content of the communication in facilitating cooperation in social dilemmas. Several possible mechanisms have been proposed including communication might allow coordination among participants, thus they might develop trust relationships, and communication might express social pressure (Hertz et al. (2023)). Here, in this project, communication and teaching are modelled through two simple variants of signalling game, a game which was originally devised to simulate the emergence of convention (Karch (2023); Ohmer et al. (2022)). In game theoretic words, a convention is a system of arbitrary rules that enables two players to share meaningful information. In the Modified AI-Economist with Communication, it is assumed that the agents beside being able to build two types of houses individually, they are able to build the same types of houses with another agent if they communicate with each other and their communication is aligned. In the Modified AI-Economist with Teaching, the agents are divided to two groups, teachers and students. The teachers know how to combine two complementary natural resources to build a house type but they are not capable of, while the students do not know which complementary natural resources to use to build a house, but if a teacher tell them, they are capable of building. Thus if they communicate with each other and their communication is aligned, they would be able to build houses together. The final results show that collectivistic environment such as Full-Utilitarian system is more favourable for the emergence of communication and teaching, or more precisely, evolution of language alignment. One possible reason behind this observed phenomena is briefly discussed in the results section of the paper. Overall, these two extensions of the AI-Economist are another manifestation of the power of multi-agent reinforcement learning to model social and economical phenomena (Trott et al. (2021); Zheng et al. (2022); Zhang et al. (2022); Leibo et al. (2019, 2021); Johanson et al. (2022)).

2 The Modified AI-Economist with Communication/Teaching

For a complete description of the AI-Economist, please refer to Appendix (A). Here, three major modifications to the original framework of the AI-Economist are described.

First, two new resource materials –iron and soil –are added to the environment, and now with four building materials, the number of possible house types is diversified to two: a red house is exclusively can be built from wood and stone, and a blue house is exclusively can be built from iron and soil. Also, each one of these house types can be built via two means: individually by each single agent, or together with two agents. Moreover, there are two kinds of build skills for each agent: the required skill for building alone and the required skill for building together which is higher than the former. Furthermore, the required labour for building a house type alone or together is also different: the latter is higher than the former.

Second, the agents are equipped with a voting mechanism and each one of them will have twenty four extra actions to rank the four material resources considering twenty four ranking possibilities. Additionally, all four materials are placed, planted, or extracted randomly in a uniform environment. The agents are initially placed randomly in the environment too. Also, in the modified version of the AI-Economist, the social planner is able to observe the complete public map of the environment (Appendix (B) Figs. 2 and 5). Based on this voting mechanism, across individualistic-collectivistic axis, three different governing systems are introduced. In the Full-Libertarian system, the social planner determines the tax rates considering a particular social welfare function –such as inverse income weighted utility or the multiplication of equality and productivity –and the policy network of the agents now produces an action ranking four different resources. Then, the agents can invest individually their taxes on planting or extraction rates of each one of the four material resources considering how they rank them. In the Semi-Libertarian/Utilitarian system, the tax rates are optimised by the social planner again considering a particular social welfare function. Moreover, the policy network of the agents, as before, produces an action ranking the four material resources. However, the social planner in this case uses the Borda vote counting method to rank the four material types based on the votes of all agents. Then the social planner invests the collected taxes on planting or extraction of the four resources based on the counted votes of all agents. Finally, in the Full-Utilitarian system, the social planner simultaneously optimises the tax rates and the ranking order of all four materials considering again a suitable social welfare function. Then, it invests the collected taxes accordingly on the planting or extraction rates of all four materials.

Third, communication and teaching capabilities are added to the environment through two variants of signalling game with possible misalignment. When agents decide to build each one of two house types together, they need to communicate the kind of required materials that they have to build that specific house type via four alphabetic letters: [a, b, c, d]. [a, b] refer to the materials required for building a red house type such as wood and stone, while [c, d] refer to the materials required for building a blue house type such as iron and soil. In the Modified AI-Economist with Communication, they are six agents and their language is maximally misaligned in the beginning of each simulation ([a, b, c, d], [d, a, b, c], [c, d, a, b], [b, c, d, a], [d, c, b, a], and [b, a, d, c]). Here, if the agents both use the same kind of letters for both required material resources, they are able to build a house and obtain a large reward. Otherwise, they correct one misaligned letter and obtain a small reward. This process continues until the language of all agents are aligned together, or an episode is ended. In the Modified AI-Economist with Teaching, three agents are teachers and three agents are students. The language of teachers is fixed across training to [a, b, c, d], while again initially the language of students are set to be maximally misaligned ([d, a, b, c], [c, d, a, b], and [b, c, d, a]). Here, a teacher decides to teach building a house type to a student so they both would be able to build a house and share a large reward. In this case, the language of teachers is a reference language which the language of students is compared against. If a pair of teacher and student use the same set of two letters for the required two material resources, they would be able to build that kind of house type. Otherwise, one letter of the language of the student is modified to match the language of the teacher and they obtain a small reward. This process continues until the language of all students are matched to the language of the teachers, or an episode is ended. Finally, the language alignment of agents are calculated across an episode as the average letter alignment of two sequence of language letters among all possible pairs of agents.

3 Results

Fig. 1 panel (A) shows the language alignment across three governing systems –Full-Libertarian, Semi-Libertarian/Utilitarian, and Full-Utilitarian –for the Modified AI-Economist with Communication. This figure shows that the rate of language alignment under a collectivistic governing system such as Full-Utilitarian government is higher that the rate of language alignment under an individualistic governing system such as Full-Libertarian government. However, full-alignment never happens under all three governing systems. Moreover, Fig. 1 panel (B) shows the language alignment across three governing systems –Full-Libertarian, Semi-Libertarian/Utilitarian, and Full-Utilitarian –for the Modified AI-Economist with Teaching. This figure again shows that the rate of language alignment under a collectivistic governing system such as Full-Utilitarian government is higher that the rate of language alignment under an individualistic governing system such as Full-Libertarian government. In this case, full-alignment happens under all three governing systems, but happens faster under more collectivistic government.

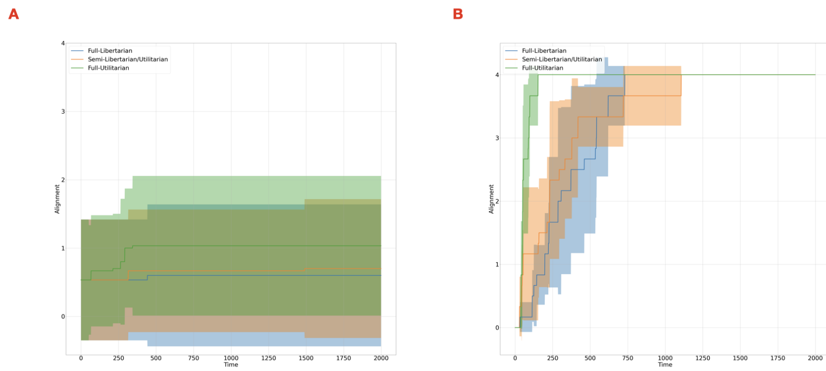

Figure 1: The language alignment across an episode for three governing systems of the Modified AI-Economist with Communication (A) and with Teaching (B). As it is clear from panel (A), a collectivistic government such as Full-Utilitarian governing system is more favourable for the evolution of language alignment compared to an individualistic government such as Full-Libertarian governing system. However, full alignment –which is equal to 4 –does not happen under any of these governing systems, thus none of the agents can build houses together. Moreover, panel (B) shows that a collectivistic government such as Full-Utilitarian governing system is more favourable for the evolution of language alignment compared to an individualistic government such as Full-Libertarian governing system. In this case, full alignment –which is equal to 4 –happens under all three governing systems, thus the agents can build houses together under all these governing systems.

While there are limitations in the current study, it is worthwhile to ponder about the reasons behind more favourable conditions of Full-Utilitarian government for the evolution of language alignment compared to the Full-Libertarian government in the way that they are modelled in this work. One speculation is that basically among Full-Libertarian, Semi-Libertarian/Utilitarian, and Full-Utilitarian governments as they are modelled in this study, the Full-Utilitarian governing system lacks any individual voting mechanism. If we consider individual voting as a mean of communication and coordination among agents, this lack of individual voting in the Full-Utilitarian government might be the reason behind fast emergence and evolution of language alignment under this type of governing system. Basically, in the current modelling, the agents under Full-Utilitarian society without any means of communication or coordination are motivated to incur high costs and simultaneously gain large rewards by selecting the action to build houses together compared to the societies which have a voting mechanism as a mean of communication and coordination.

4 Current Limitations

There are at least two limitations to the current tentative study. The first limitation comes from the fact that for each set of input parameters of the Modified AI-Economist with Communication or Teaching, only one simulation has been run to generate one set of results. Then a pair of similar runs have been pooled together to have average results across different conditions (Appendix (B) Fig. 7). Thus it is reasonable to increase the number of simulations per condition. The second limitation is the number of episodes each training has been run which is equal to 5000 (this is still a reasonable choice in MARL experiments (Albrecht et al. (2023))). The only way to make sure that the final results are optimal is to let the training of MARL runs for very large number of time-steps if computational resources are available. Figs. 8 and 9 in Appendix (B) show that even with this choice almost all simulations have been converged, while the training of the Modified AI-Economist with Teaching is more stable than the training of the Modified AI-Economist with Communication.

References

Albrecht, S. V., Christianos, F., and Schäfer, L. (2023). Multi-agent reinforcement learning foundations and modern approaches. *Book*.

Camerer, C. F. (2003). Behavioral game theory: Experiments in strategic interaction. *Book*.

Du, Y., Leibo, J. Z., Islam, U., Willis, R., and Sunehag, P. (2023). A review of cooperation in multi-agent learning. *ArXiv*.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*.

Greene, J. D. (2013). The cognitive neuroscience of moral judgment and decision making. *Preprint*.

Hertz, U., Koster, R., Janssen, M. A., and Leibo, J. Z. (2023). Beyond the matrix: Using multi-agent-reinforcement learning and behavioral experiments to study social-ecological systems. *ArXiv*.

Hughes, E., Leibo, J. Z., Phillips, M., Tuyls, K., Edgar Dueñez-Guzman, A. G. C., Dunning, I., Zhu, T., McKee, K., Koster, R., Roff, H., and Graepel, T. (2018). Inequity aversion improves cooperation in intertemporal social dilemmas. *NeurIPS*.

Janssen, M. A., Holahan, R., Lee, A., and Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science*.

Johanson, M. B., Hughes, E., Timbers, F., and Leibo, J. Z. (2022). Emergent bartering behaviour in multi-agent reinforcement learning. *DeepMind*.

Karch, T. (2023). Towards social autotelic artificial agents. *PhD thesis of the graduate school of mathematics and computer Science at University of Bordeaux*.

Leibo, J. Z., Duenez-Guzman, E., Vezhnevets, A. S., John P. Agapiou, P. S. R. K., Matyas, J., Beattie, C., Mordatch, I., and Graepel, T. (2021). Scalable evaluation of multi-agent reinforcement learning with melting pot. *ICML*.

Leibo, J. Z., Hughes, E., Lanctot, M., and Graepel, T. (2019). Autocurricula and the emergence of innovation from social interaction: A manifesto for multi-agent intelligence research. *DeepMind*.

Ohmer, X., Marino, M., Franke, M., and Konig, P. (2022). Mutual influence between language and perception in multi-agent communication games. *PLOS Computational Biology*.

Trott, A., Srinivasa, S., van der Wal, D., Haneuse, S., and Zheng, S. (2021). Building a foundation for data-driven, interpretable, and robust policy design using the ai economist. *ArXiv*.

Zhang, T., Williams, A., Phade, S., Srinivasa, S., Zhang, Y., Gupta, P., Bengio, Y., and Zheng, S. (2022). Ai for global climate cooperation: Modeling global climate negotiations, agreements, and long-term cooperation in rice-n. *Salesforce*.

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*.

5 Appendix A: The AI-Economist

Here is a detailed description of the original AI-Economist (Zheng et al. (2022)):

1. The AI-Economist is a two-level deep RL framework for policy design in which agents and a social planner co-adapt. In particular, the AI-Economist uses structured curriculum learning to stabilise the challenging two-level, co-adaptive learning problem. This framework has been validated in the domain of taxation. In two-level spatiotemporal economies, the AI-Economist has substantially improved both utilitarian social welfare and the trade-off between equality and productivity over baselines. It was successful to do this despite emergent tax-gaming strategies, accounting the emergent labor specialisation, agent interactions, and behavioural changes.
2. Stabilising the training process in two-level RL is difficult. To overcome, the training procedure in the AI-Economist has two important features - curriculum learning and entropy-based regularisation. Both of them encourage the agents and the social planner to co-adopt gradually and not stopping exploration too early during training and getting trapped in local minima. Furthermore, the AI-Economist is a game of imperfect (the agents and the social planner do not have access to the perfect state of the world) but complete (the agents and the social planner know the exact rules of the game) information.
3. The Gather-Trade-Build economy of the AI-Economist is a two-dimensional spatiotemporal economy with agents who move, gather resources (stone and wood), trade them, and build houses. Each agent has a varied house build-skill which sets how much income an agent receives from building a house. Build-skill is distributed according to a Pareto distribution. As a result, the utility-maximising agents learn to specialise their behaviours based on their build-skill. Agents with low build-skill become gatherers: they earn income by gathering and selling resources. Agents with high build-skill become builders: they learn that it is more profitable to buy resources and then build houses.
4. The Open-Quadrant environment of the Gather-Trade-Build economy has four regions delineated by impassable water with passageways connecting each quadrant. Quadrants contain different combinations of resources: both stone and wood, only stone, only wood, or nothing. Agents can freely access all quadrants, if not blocked by objects or other agents. This scenario uses a fixed set of build-skill based on a clipped Pareto distribution and determine each agent's starting location based on its assigned build-skill. The Open-Quadrant scenario assigns agents to a particular corner of the map, with similarly skilled agents being placed in the same starting quadrant (Agents in the lowest build-skill quartile start in the wood quadrant; those in the second quartile start in the stone quadrant; those in the third quartile start in the quadrant with both resources; and agents in the highest build-skill quartile start in the empty quadrant).
5. The state of the world is represented as an $n_h \times n_w \times n_c$ tensor, where n_h and n_w are the size of the world and n_c is the number of unique entities that may occupy a cell, and the value of a given element indicates which entity is occupying the associated location. The action space of the agents includes four movement actions: up, down, left, and right. Agents are restricted from moving onto cells that are occupied by another agent, a water tile, or another agent's house. Stone and wood stochastically spawn on special resource regeneration cells. Agents can gather these resources by moving to populated resource cells. After harvesting, resource cells remain empty until new resources spawn. By default, agents collect one resource unit, with the possibility of a bonus unit also being collected, the probability of which is determined by the agent's gather-skill. Resources and coins are accounted for in each agent's endowment x, which represents how many coins, stone, and wood each agent owns.
6. Agent's observations include the state of their own endowment (wood, stone, and coin), their own build-skill level, and a view of the world state tensor within an egocentric spatial window. The experiment use a world of 25 by 25 for 4-agent and 40 by 40 for 10-agent environments, where agent spatial observations have size 11 by 11 and are padded as needed when the observation window extends beyond the world grid. The planner observations include each agent's endowment but not build-skill level. The planner does not observe the spatial state of the world.

7. Agents can buy and sell resources from one another through a continuous double-auction. Agents can submit asks (the number of coins they are willing to accept) or bids (how much they are willing to pay) in exchange for one unit of wood or stone. The action space of the agents includes 44 actions for trading, representing the combination of 11 price levels (0, ..., 10 coins), 2 directions (bids and asks), and 2 resources (wood and stone). Each trade action maps to a single order (i.e., bid three coins for one wood, ask for five coins in exchange for one stone, etc.). Once an order is submitted, it remains open until either it is matched (in which case a trade occurs) or it expires (after 50 time steps). Agents are restricted from having more than five open orders for each resource and are restricted from placing orders that they cannot complete (they cannot bid with more coins than they have and cannot submit asks for resources that they do not have). A bid/ask pair forms a valid trade if they are for the same resource and the bid price matches or exceeds the ask price. When a new order is received, it is compared against complementary orders to identify potential valid trades. When a single bid (ask) could be paired with multiple existing asks (bids), priority is given to the ask (bid) with the lowest (highest) price; in the event of ties, priority then is given to the earliest order and then at random. Once a match is identified, the trade is executed using the price of whichever order was placed first. For example, if the market receives a new bid that offers eight coins for one stone and the market has two open asks offering one stone for three coins and one stone for seven coins, received in that order, the market would pair the bid with the first ask and a trade would be executed for one stone at a price of three coins. The bidder loses three coins and gains one stone; the asker loses one stone and gains three coins. Once a bid and ask are paired and the trade is executed, both orders are removed. The state of the market is captured by the number of outstanding bids and asks at each price level for each resource. Agents observe these counts for both their own bids/asks and the cumulative bids/asks of other agents. The planner observes the cumulative bids/asks of all agents. In addition, both the agents and the planner observe historical information from the market: the average trading price for each resource, as well as the number of trades at each price level.

8. Agents can choose to spend one unit of wood and one unit of stone to build a house, and this places a house tile at the agent's current location and earns the agent some number of coins. Agents are restricted from building on source cells as well as locations where a house already exists. The number of coins earned per house is identical to an agent's build-skill, a numeric value between 10 and 30. Hence, agents can earn between 10 and 30 coins per house built. Build-skill is heterogeneous across agents and does not change during an episode. Each agent's action space includes one action for building. Over the course of an episode of 1000 time steps, agents accumulate labor cost, which reflects the amount of effort associated with their actions. Each type of action (moving, gathering, trading, and building) is associated with a specific labor cost. All agents experience the same labor costs.

9. Simulations are run in episodes of 1000 time steps, which is subdivided into 10 tax periods or tax years, each lasting 100 time steps. Taxation is implemented using income brackets and bracket tax rates. All taxation is anonymous: Tax rates and brackets do not depend on the identity of taxpayers. On the first time step of each tax year, the marginal tax rates are set that will be used to collect taxes when the tax year ends. For taxes controlled by the deep neural network of the social planner, the action space of the planner is divided into 7 action subspaces, one for each tax bracket: $(0, 0.05, 0.10, ..., 1.0)^7$. Each subspace denotes the set of discretised marginal tax rates available to the planner. Discretisation of tax rates only applies to deep learning networks, enabling standard techniques for RL with discrete actions. The income bracket cutoffs are fixed. Each agent observes the current tax rates, indicators of the temporal progress of the current tax year, and the set of sorted and anonymised incomes the agents reported in the previous tax year. In addition to this global tax information, each agent also observes the marginal rate at the level of income it has earned within the current tax year so far. The planner also observes this global tax information, as well as the non-anonymised incomes and marginal tax rate (at these incomes) of each agent in the previous tax year.

10. The payable tax for income z is computed as follows:

$$T(z) = \sum_{j=1}^{B} \tau_j \cdot ((b_{j+1} - b_j)\mathbf{1}[z > b_{j+1}] + (z - b_j)\mathbf{1}[b_j < z \leq b_{j+1}]) \tag{1}$$

where B is the number of brackets, and τ_j and b_j are marginal tax rates and income boundaries of the brackets, respectively.

11. An agent's pretax income z_i for a given tax year is defined simply as the change in its coin endowment C_i since the start of the year. Accordingly, taxes are collected at the end of each tax year by subtracting $T(z_i)$ from C_i. Taxes are used to redistribute wealth: the total tax revenue is evenly redistributed back to the agents. In total, at the end of each tax year, the coin endowment for agent i changes according to $\triangle C_i = -T(z_i) + \frac{1}{N}\sum_{j=1}^{N} T(z_j)$, where N is the number of agents. Through this mechanism, agents may gain coin when they receive more through redistribution than they pay in taxes. Following optimal taxation theory, agent utilities depend positively on accumulated coin $C_{i,t}$, which only depends on post-tax income $\tilde{z} = z - T(z)$. In contrast, the utility for agent i depends negatively on accumulated labor $L_{i,t} = \sum_{k=0}^{t} l_{i,k}$ at time step t. The utility for an agent i is:

$$u_{i,t} = \frac{C_{i,t}^{1-\eta} - 1}{1-\eta} - L_{i,t}, \eta > 0 \tag{2}$$

12. Agents learn behaviours that maximise their expected total discounted utility for an episode. It is found that build-skill is a substantial determinant of behaviour; agents' gather-skill empirically does not affect optimal behaviours in our settings. All of the experiments use a fixed set of build-skills, which, along with labor costs, are roughly calibrated so that (i) agents need to be strategic in how they choose to earn income and (ii) the shape of the resulting income distribution roughly matches that of the 2018 U.S. economy with trained optimal agent behaviours.

13. RL provides a flexible way to simultaneously optimise and model the behavioural effects of tax policies. RL is instantiated at two levels, that is, for two types of actors: training agent behavioural policy models and a taxation policy model for the social planner. Each actor's behavioural policy is trained using deep RL, which learns the weights θ_i of a neural network $\pi(a_{i,t}|o_{i,t};\theta_i)$ that maps an actor's observations to actions. Network weights are trained to maximise the expected total discounted reward of the output actions. Specifically, for an agent i using a behavioural policy $\pi_i(a_t|o_t;\theta_i)$, the RL training objective is (omitting the tax policy π_p):

$$\max_{\pi_i} E_{a_1\sim\pi_1,\ldots,a_N\sim\pi_N,s'\sim P}[\sum_{t=0}^{H}\gamma^t r_t] \tag{3}$$

where $s^{'}$ is the next state and P denotes the dynamics of the environment. The objective for the planner policy π_p is similar. Standard model-free policy gradient methods update the policy weights θ_i using (variations of):

$$\triangle\theta_{\mathbf{i}} \propto E_{a_1\sim\pi_1,\ldots,a_N\sim\pi_N,s'\sim P}[\sum_{t=0}^{H}\gamma^t r_t \nabla_{\theta_i}\log\pi_i(a_{i,t}|o_{i,t};\theta_{\mathbf{i}})] \tag{4}$$

14. In this work, the proximal policy gradients (PPO) is used to train all actors (both agents and planner). To improve learning efficiency, a single-agent policy network $\pi(a_{i,t}|o_{i,t};\theta)$ is trained whose weights are shared by all agents, that is, $\theta_i = \theta$. This network is still able to embed diverse, agent-specific behaviours by conditioning on agent-specific observations.

15. At each time step t, each agent observes the following: its nearby spatial surroundings; its current endowment (stone, wood, and coin); private characteristics, such as its building skill; the state of the markets for trading resources; and a description of the current tax rates. These observations form the inputs to the policy network, which uses a combination of convolutional, fully connected, and recurrent layers to represent spatial, non-spatial, and historical information, respectively. For recurrent components, each agent maintains its own hidden state. The policy network for the social planner follows a similar construction but differs somewhat in the information it observes. Specifically, at each time step, the planner policy observes the following: the current inventories of each agent, the state of the resource markets, and a description of the current tax rates. The planner cannot directly observe private information such as an agent's skill level.

16. Rational economic agents train their policy π_i to optimise their total discounted utility over time while experiencing tax rates τ set by the planner's policy π_p. The agent training objective is:

$$\forall i : \max_{\pi_i} E_{\tau \sim \pi_p, a_i \sim \pi_i, \mathbf{a_{-i}} \sim \pi_{-\mathbf{i}}, s' \sim P}[\sum_{t=1}^{H} \gamma^t r_{i,t} + u_{i,0}], r_{i,t} = u_{i,t} - u_{i,t-1} \tag{5}$$

where the instantaneous reward $r_{i,t}$ is the marginal utility for agent i at time step t. Bold-faced quantities denote vectors, and the subscript $-i$ denotes quantities for all agents except for i.

17. For an agent population with monetary endowments $\mathbf{C_t} = (C_{1,t}, ..., C_{N,t})$, the equality $eq(\mathbf{C_t})$ is defined as:

$$eq(\mathbf{C}_t) = 1 - \frac{N}{N-1} gini(\mathbf{C}_t), 0 \leq eq(\mathbf{C}_t) \leq 1 \tag{6}$$

where the Gini index is defined as:

$$gini(\mathbf{C}_t) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} |C_{i,t} - C_{j,t}|}{2N \sum_{i=1}^{N} C_{i,t}}, 0 \leq gini(\mathbf{C}_t) \leq \frac{N-1}{N} \tag{7}$$

18. The productivity is defined as the sum of all incomes:

$$prod(\mathbf{C}_t) = \sum_i C_{i,t} \tag{8}$$

The economy is closed: subsidies are always redistributed evenly among agents, and no tax money leaves the system. Hence, the sum of pretax and post-tax incomes is the same. The planner trains its policy π_p to optimise social welfare:

$$\max_{\pi_p} E_{\tau \sim \pi_p, \mathbf{a} \sim \pi, s' \sim P}[\sum_{t=1}^{H} \gamma^t r_{p,t} + swf_0], r_{p,t} = swf_t - swf_{t-1} \tag{9}$$

19. The utilitarian social welfare objective is the family of linear-weighted sums of agent utilities, defined for weights $\omega_i \geq 0$:

$$swf_t = \sum_{i=1}^{N} \omega_i \cdot \mathbf{u}_{i,t} \tag{10}$$

Inverse-income is used as the weights: $\omega_i \propto \frac{1}{C_i}$, normalised to sum to one. An objective function is defined that optimises a trade-off between equality and productivity, defined as the product of equality and productivity:

$$swf_t = eq(\mathbf{C}_t) \cdot prod(\mathbf{C}_t) \tag{11}$$

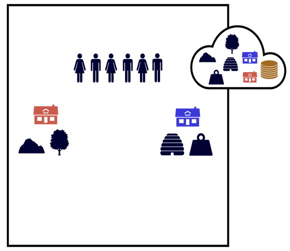

Figure 2: A schematic figure showing the environment of the Modified AI-Economist with Communication/Teaching used in this paper. In all simulations of this paper, there are 6 agents in the environment which simultaneously cooperate and compete to gather and trade four natural resources, using them to build houses alone or together –via communication or teaching –and earn incomes, and at the end of each tax period, pay their taxes to the central planner. The central planner optimises its own reward function which could be a combination of equality and productivity in the society, and returns an equal division of the total collected taxes to the mobile agents.

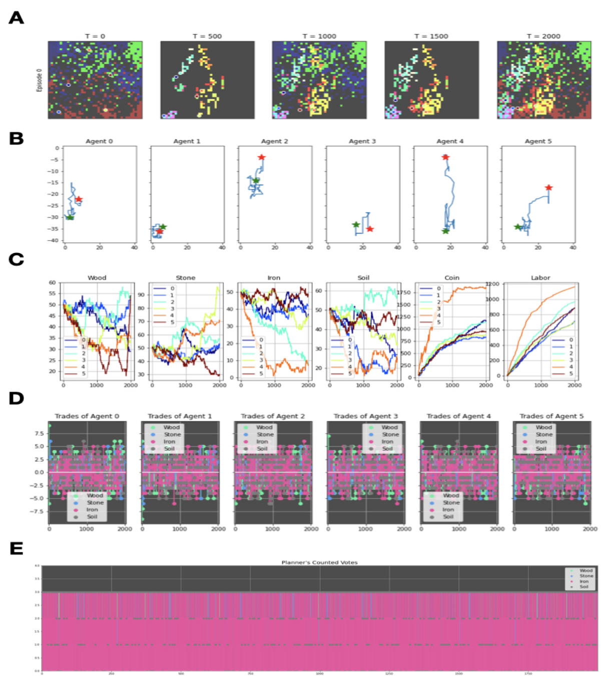

Figure 3: Sample plots obtained from running the Modified AI-Economist with Communication under Semi-Libertarian/Utilitarian governing system with equality times productivity as the objective function of the central planner. (A) The environment across five time-points of an episode, (B) the movement of the agents across an episode, (C) the budgets of four resources plus coin and labor of the agents across an episode, (D) the trades of four resources of the agents across an episode, (E) the counted votes of the agents across an episode.

6 Appendix B: Supplemental Figures

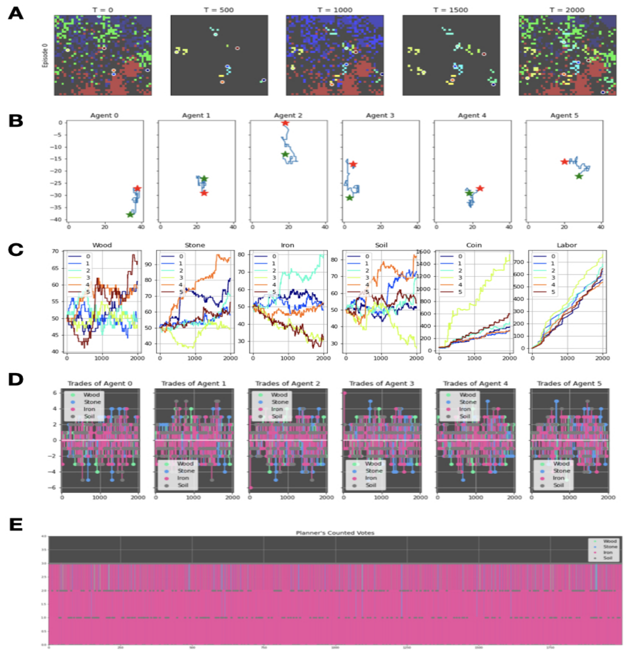

Figure 4: Sample plots obtained from running the Modified AI-Economist with Teaching under Semi-Libertarian/Utilitarian governing system with equality times productivity as the objective function of the central planner. (A) The environment across five time-points of an episode, (B) the movement of the agents across an episode, (C) the budgets of four resources plus coin and labor of the agents across an episode, (D) the trades of four resources of the agents across an episode, (E) the counted votes of the agents across an episode.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly reflect the contribution of this paper.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: In the current limitations section, we briefly described the limitations of the current work.

   Guidelines:

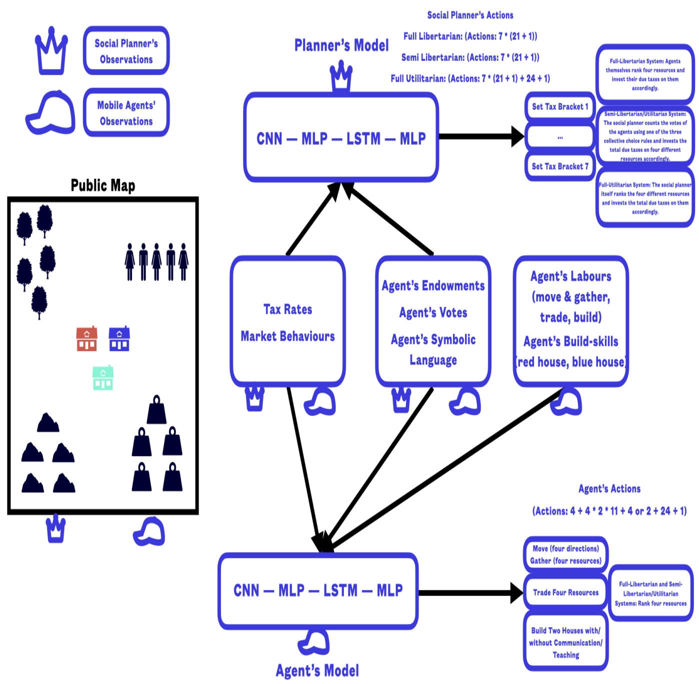

Figure 5: Observation and action spaces for economic agents and the social planner. The agents and the planner observe different subsets of the world state. Agents observe their spatial neighbourhood, market prices, tax rates, inventories, votes, symbolic language, labor, and skill level. Agents can decide to move (and therefore gather if moving onto a resource), buy, sell, build, vote, or communicate. There are maximum 121 (communication) and 119 (teaching) unique actions available to the agents. The planner observes the public spatial map, market prices, tax rates, agent inventories, votes, and symbolic language. The social planner in both environments decides how to set tax rates, choosing one of 22 settings for each of the 7 tax brackets. MLP: multi-layer perceptron, LSTM: long short-term memory, CNN: convolutional neural network. This figure should be compared to Fig. 9 of the original AI-Economist paper (Zheng et al. (2022)).

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover

A

```
# ===== SCENARIO CLASS =====
  'scenario_name' : 'uniform_scenario_for_vote_and_invest',
  # ===== COMPONENTS =====
  'components' : [
    # (1) Communicating/Teaching and building houses
    {'CommunicateBuild'/'TeachBuild': {
      'payment' :                          10,
      'payment_max_skill_multiplier' :     np.array([5, 5, 10, 10])/np.array([10, 10])
      'skill_dist' :                       'pareto',
      'build_labor' :                      np.array([10.0, 10.0, 20.0, 20.0])/np.array([20.0, 20.0])
    }},
    # (2) Trading collectible resources
    {'ContinuousDoubleAuction' : {
      'max_bid_ask' :           10,
      'order_labor' :           0.25,
      'order_duration' :        50,
      'max_num_orders' :        5,
    }},
    # (3) Moving and gathering resources
    {'Gather' : {
      'move_labor' :        1,
      'collect_labor' :     1,
      'skill_dist' :        'none',
    }},
    # (4) Voting and investing resources
    {'AgentVotesPlannerInvestsResources',
     'AgentVotesAgentInvestsResources',
     'PlannerVotesPlannerInvestsResources' : {
      'vote_count_method' :     'Borda',
      'institution_type' :      'Inclusive',

      # Similar to components/redistribution/PeriodicBracketTax
      'disable_taxes' :             False,
      'tax_model' :                 'model_wrapper',
      'period' :                    100,
      'rate_min' :                  0.0,
      'rate_max' :                  1.0,
      'rate_disc' :                 0.05,
      'n_brackets' :                7,
      'top_bracket_cutoff' :        100,
      'usd_scaling' :               1000.0,
      'bracket_spacing' :           'us-federal',
      'fixed_bracket_rates' :       None,
      'pareto_weight_type' :        'inverse_income',
      'saez_fixed_elas' :           None,
      'tax_annealing_schedule' :    None,
    }},
  ,]
```

B

```
# ===== STANDARD ARGUMENTS ======
  'n_agents' :                          6,
  'world_size' :                        [40, 40],
  'episode_length' :                    2000,
  'multi_action_mode_agents' :          False,
  'multi_action_mode_planner' :         True,
  'flatten_observations' :              True,
  'flatten_masks' :                     True,
  'allow_observation_scaling' :         True,
  'dense_log_frequency' :               20,
  'world_dense_log_frequency' :         50,
  'collate_agent_step_and_reset_data' : False,
  'seed' :                              None,
```

C

```
# ===== SCENARIO CLASS ARGUMENTS =====
  # Similar to scenarios/simple_wood_and_stone_and_iron_and_soil/dynamic_layout/Uniform
  'planner_gets_spatial_info' :         True,
  'full_observability' :                False,
  'mobile_agent_observation_range' :    5,
  'starting_wood_coverage' :            0.10
  'wood_regen_halfwidth' :              0.50
  'wood_regen_weight' :                 0.10,
  'wood_max_health' :                   1,
  'wood_clumpiness' :                   0.35,
  'starting_stone_coverage' :           0.10,
  'stone_regen_halfwidth' :             0.50,
  'stone_regen_weight' :                0.10,
  'stone_max_health' :                  1,
  'stone_clumpiness' :                  0.35,
  'starting_iron_coverage' :            0.10,
  'iron_regen_halfwidth' :              0.50,
  'iron_regen_weight' :                 0.10,
  'iron_max_health' :                   1,
  'iron_clumpiness' :                   0.35,
  'starting_soil_coverage' :            0.10,
  'soil_regen_halfwidth' :              0.50,
  'soil_regen_weight' :                 0.10,
  'soil_max_health' :                   1,
  'soil_clumpiness' :                   0.35,
  'gradient_steepness' :                1,
  'checker_source_blocks' :             False,
  'starting_agent_coin' :               50,
  'energy_warmup_constant' :            0,
  'energy_warmup_method' :              'decay',
  'planner_reward_type' :               'coin_eq_times_productivity',
                                        'inv_income_weighted_utility',
  'mixing_weight_gini_vs_coin' :        0.0,
  'mixing_weight_maximin_vs_coin' :     0.0,
```

Figure 6: (A)(B)(C) Different features and input parameters of the Modified AI-Economist with Communication and with Teaching which are tested and their aggregated plots are brought and discussed in the main text. The orange texts indicate various parts of the input structure. The green texts show the alternative parameters which are tested in this paper.

Number	Experiment	Governing System	Reward Function
1	Communication	Full-Libertarian	Equality * Productivity
2	Communication	Full-Libertarian	Inverse Income Weighted Utility
3	Communication	Semi-Libertarian/Utilitarian	Equality * Productivity
4	Communication	Semi-Libertarian/Utilitarian	Inverse Income Weighted Utility
5	Communication	Full-Utilitarian	Equality * Productivity
6	Communication	Full-Utilitarian	Inverse Income Weighted Utility
7	Teaching	Full-Libertarian	Equality * Productivity
8	Teaching	Full-Libertarian	Inverse Income Weighted Utility
9	Teaching	Semi-Libertarian/Utilitarian	Equality * Productivity
10	Teaching	Semi-Libertarian/Utilitarian	Inverse Income Weighted Utility
11	Teaching	Full-Utilitarian	Equality * Productivity
12	Teaching	Full-Utilitarian	Inverse Income Weighted Utility

Figure 7: A figure showing all different runs of the Modified AI-Economist with Communication and with Teaching with different values as input parameters. The *Reward Function* refers to the reward function of the central planner. To generate the plots in the main text, the generated results of a pair of consecutive simulations belonging to one kind of experiment and one governing system are pooled together.

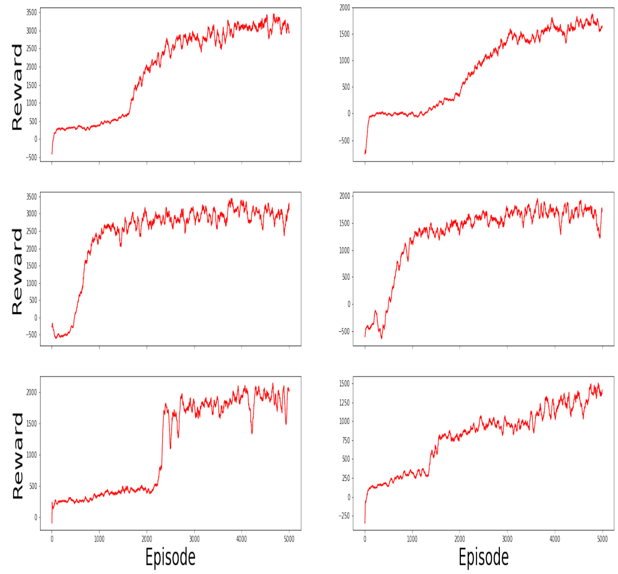

Figure 8: Average episode reward across training - 5000 episodes - for all runs of the Modified AI-Economist with Communication. The plots of the 6 runs 1-6 of Fig. 7 are brought in the order from left-to-right and top-to-bottom. It is worthwhile to mention that the training of two-level RL is particularly unstable, but it seems that almost all the simulations have been converged, but they are less stable than the plots of Fig. 9.

limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: There is not an theoretical result in this paper.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

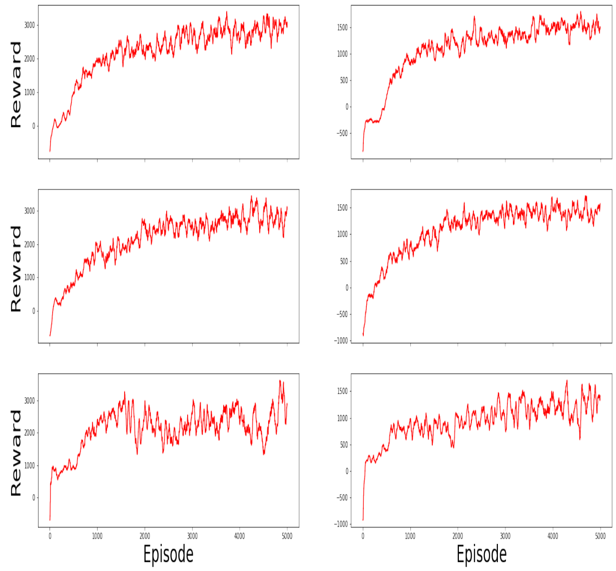

Figure 9: Average episode reward across training - 5000 episodes - for all runs of the Modified AI-Economist with Teaching. The plots of 6 runs 7-12 of Fig. 7 are brought in the order from left-to-right and top-to-bottom. It is worthwhile to mention that the training of two-level RL is particularly unstable, but it seems that almost all the simulations have been converged, and they are more stable than the plots of Fig. 8.

Answer: [Yes]

Justification: In the main text and the appendices of the paper all the required information to reproduce the results of this paper are brought. The code of the paper is also available upon request.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

(b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: There is not any data used in this study, and the code to reproduce all the results of this paper is available upon request.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The code to reproduce all the results of this paper is available upon request.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: In this paper, for each set of parameters, one experiment has been done once.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [No]

   Justification: This point is addressed in the accompanying github repository which will be available upon request.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in this paper conform, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: The results of this paper is tentative and must be interpreted cautiously. Thus there is not any associated potential positive or negative impacts with this paper.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper does not use these kinds of data or models.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The original code of the AI-Economist has been properly credited.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: There is not any new asset introduced in this paper.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: There is not any research with human subjects in this paper.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: There is not any research with human subjects in this paper.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.