Multi-Agent Reinforcement Learning for Labor Supply Optimization in Digital Platforms under Self-Set and Platform-Assigned Goals

Anil Parekh
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
anilp@buffalo.edu

Colin L. Ramos
Department of Computer Science, George Mason University, Fairfax, VA, USA.
colin.ramos@gmu.edu

Arthur M. Lyons
School of Computing, Clemson University, Clemson, SC, USA.
helloarthur@clemson.edu

Hudson Pichards
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
richards214@uc.edu

Abstract

Digital labor platforms increasingly mediate the supply of contingent work by matching independent workers with task demand through algorithmic allocation systems. These platforms face a fundamental tension between allowing workers to self-set their own labor goals and imposing platform-assigned targets to optimize aggregate output. This paper develops a multi-agent reinforcement learning framework to model and analyze labor supply optimization under both goal structures. We conceptualize each worker as an independent learning agent with private utility functions, while the platform acts as a centralized or decentralized goal-setting mechanism that shapes the reward environment. Through a system-level discussion, we examine how self-set goals foster worker autonomy and intrinsic motivation but may lead to suboptimal system-wide coordination, whereas platform-assigned goals can align individual behavior with global efficiency but risk fairness violations and worker disengagement. The architecture we propose integrates hierarchical reinforcement learning with meta-governance layers that dynamically adjust goal assignment based on worker state, platform congestion, and long-term sustainability metrics. We explore structural trade-offs between exploration and exploitation at both the agent and platform levels, and we draw parallels with other large-scale socio-technical systems such as ride-hailing fleets and crowd work markets. Governance implications are analyzed through the lenses of distributive justice, transparency, and algorithmic accountability. Furthermore, we discuss deployment challenges including computational scalability, communication overhead, and the need for robust mechanisms against adversarial worker behavior. Policy recommendations are offered for platform regulators and designers seeking to balance productivity with worker well-being. The paper concludes that hybrid goal structures, where platforms offer voluntary default targets while preserving opt-in self-set options, provide a promising middle ground that can be optimized using multi-agent reinforcement learning with appropriate fairness constraints.

Keywords

multi-agent reinforcement learning, labor supply optimization, digital platforms, goal setting, platform governance, fairness, socio-technical systems.

1. Introduction

Digital labor platforms, ranging from ride-hailing services to micro-task marketplaces, have become central to the modern economy by enabling flexible, on-demand work arrangements [1]. These platforms rely on algorithmic systems to match workers with tasks, set prices, and influence the temporal and spatial distribution of labor supply. A critical design decision is how workers' labor supply decisions are governed: should workers autonomously decide how many hours to work and when, or should the platform prescribe targets to achieve system-level efficiencies? Both approaches have been studied in isolation, yet their interaction with the learning dynamics of workers and the platform remains poorly understood.

Recent advances in multi-agent reinforcement learning (MARL) provide a powerful lens for modeling such systems [2]. Each worker can be viewed as an agent that learns a policy over time to maximize its own cumulative reward, while the platform's goal-setting mechanism shapes the reward function and thereby influences emergent behavior. When workers set their own goals, they operate in a decentralized manner, each optimizing a personal objective that may or may not align with platform objectives. When the platform assigns goals, it effectively becomes a central controller that alters the reward structure, potentially leading to more coordinated outcomes. This paper investigates the structural trade-offs inherent in these two regimes, focusing on architecture, governance, infrastructure, fairness, and sustainability.

We adopt a system-level perspective, drawing on insights from large-scale socio-technical infrastructure studies [3] and reinforcement learning theory [4]. Our contribution is to synthesize existing fragments from algorithmic management, behavioral economics, and multi-agent systems into a coherent analytical framework. We also discuss the deployment of MARL in real-world platform environments, highlighting computational and policy challenges. The remainder of the paper is organized as follows. Section 2 reviews related work on platform labor dynamics, goal-setting theory, and multi-agent reinforcement learning. Section 3 presents a system architecture for MARL-based labor supply optimization. Section 4 compares the dynamics of self-set versus platform-assigned goals. Section 5 addresses governance, fairness, and policy implications. Section 6 discusses deployment and infrastructure considerations. Section 7 examines sustainability and robustness. Section 8 concludes with forward-looking perspectives.

2. Background and Related Work

The study of labor supply on digital platforms has roots in both operations research and behavioral science. Early models treated workers as rational utility maximizers responding to wage rates and time constraints [5]. However, empirical work reveals that worker decisions are influenced by non-monetary factors, including autonomy, schedule flexibility, and social comparison [6]. Goal-setting theory, originally developed in organizational psychology, demonstrates that specific and challenging goals enhance performance, especially when individuals have high goal commitment [7]. In the platform context, this has been extended to examine how self-set versus assigned goals affect worker effort and retention [8].

Parallel to these behavioral studies, algorithmic management systems have emerged, where platforms use data-driven methods to nudge workers toward desired behaviors [9].

Reinforcement learning offers a natural formalism for modeling such adaptive policies. Single-agent reinforcement learning has been applied to dynamic pricing and task allocation [10], but the multi-agent setting is more realistic because workers interact non-cooperatively through shared resources such as task demand and payment pools [11]. MARL algorithms, such as independent Q-learning, centralized training with decentralized execution (CTDE), and value decomposition networks, have been deployed in simulated ride-hailing environments [12]. However, most prior work assumes a fixed goal structure imposed by the platform, ignoring the possibility that workers may independently set their own goals.

The distinction between self-set and platform-assigned goals has been examined in a recent field experiment on a large gig platform, which found that self-set goals increased worker output and satisfaction compared to no goals, but that platform-assigned goals could induce higher output when workers had low baseline productivity [17]. That study, however, did not consider the learning dynamics over multiple interactions. We extend this line of inquiry by embedding goal structures within a MARL framework, allowing us to analyze long-term equilibrium properties and system-level trade-offs.

3. System Architecture and Multi-Agent Formulation

We consider a platform populated by N workers, each modeled as an independent reinforcement learning agent. The state of each worker includes its own history of earnings, hours worked, fatigue level, and current goal (if any). The platform observes aggregated states and can assign goals, adjust reward multipliers, or provide information that influences worker value functions. The overall system can be formalized as a stochastic game [13]. Workers select actions (e.g., accept a task, reject a task, or stop working) and receive rewards composed of monetary payments, goal completion bonuses, and intrinsic satisfaction from meeting self-imposed targets.

The architecture consists of three layers. The first layer is the agent-level policy, where each worker learns a behavioral policy using a deep Q-network or actor-critic method. Workers may share a common reward structure if goals are platform-assigned, or they may have heterogeneous reward functions if they self-set goals. The second layer is the goal-setting mechanism, which operates either as a fixed feed-forward rule (e.g., assigning weekly earnings targets based on historical median) or as a learned meta-policy. The meta-policy is trained using offline MARL to maximize platform-level objectives such as total completed tasks, worker retention, and fairness metrics. The third layer is the governance module, which monitors outcome distributions and adjusts the goal-setting mechanism to ensure equity and robustness against distributional shift.

One critical architectural choice is the degree of centralization. In a fully decentralized approach, workers set goals independently without any platform intervention, leading to a Nash equilibrium where each worker's policy is a best response to the policies of others [14]. This equilibrium may be inefficient because of externalities: for example, if all workers increase supply simultaneously, task prices drop, reducing individual incentives. In a centralized approach, the platform assigns goals to all workers, equivalent to a centralized controller that solves a large-scale optimization problem. However, full centralization suffers from high computational complexity and reduced worker autonomy. A middle ground is hierarchical reinforcement learning, where the platform sets high-level goals (e.g., weekly earnings targets) while workers retain low-level decisions (e.g., which specific tasks to accept). This approach aligns with the concept of "steering not rowing" in platform governance [15].

4. Comparative Dynamics of Self-Set and Platform-Assigned Goals

When workers self-set goals, they can draw on intrinsic motivation and personal preferences. Empirical evidence suggests that self-set goals are often more ambitious than externally imposed ones when workers have high self-efficacy [7]. In the MARL context, self-set goals mean each worker defines its own reward function, typically a linear combination of earnings and effort cost. The learning dynamics then converge to a set of policies that may exhibit excessive variability, as workers with optimistic goals may work longer hours during peak demand, creating congestion, while workers with conservative goals may undersupply during off-peak periods. This decentralized equilibrium tends to produce a fat-tailed distribution of worker output, which can be beneficial for covering extreme demand surges but leads to inefficiencies in regular operation [16].

In contrast, platform-assigned goals homogenize the reward structure across workers. If the platform sets a uniform high target, all workers learn to work longer hours, which can increase total supply but may also cause worker burnout and high churn rates. If the platform differentiates goals based on worker characteristics (e.g., assigning higher goals to more productive workers), it can achieve better alignment with comparative advantage. However, such differentiation introduces fairness concerns: workers with lower baseline productivity may feel penalized [17]. The MARL setting reveals that assigned goals create a coordination signal: workers can anticipate that others will also increase supply, reducing the risk of overworking relative to peers. This coordination can stabilize supply patterns.

A particularly interesting regime is a hybrid structure where the platform offers default goals that workers can adjust. In MARL terms, this corresponds to a conditional reward function: the platform provides a bonus for meeting the default goal, but workers may voluntarily adopt a higher self-set goal for an additional bonus. This allows the platform to nudge behavior without removing autonomy. The meta-policy can learn optimal default goals by balancing between worker satisfaction and system throughput. Simulation studies in ride-hailing contexts indicate that hybrid structures outperform both pure self-set and pure assigned regimes in terms of total welfare and fairness metrics [18].

5. Governance, Fairness, and Policy Implications

Governance of goal-setting mechanisms must address distributive justice: how are the benefits of increased platform efficiency shared between workers and the platform? Platform-assigned goals, if set too high, can exacerbate income inequality by rewarding only high-output workers while low-output workers are penalized or drop out [19]. The MARL framework allows us to incorporate fairness constraints directly into the learning objective, for example by minimizing the variance of worker returns or ensuring a minimum income floor. These constraints can be implemented as additional terms in the platform's reward function or as constraints in the goal-setting meta-policy.

Transparency is another crucial governance principle. Workers should understand how their goals are determined and what behaviors lead to goal changes. In a MARL system, the goal-setting mechanism may be a black-box neural network, which undermines trust [20]. To mitigate this, platforms can use interpretable models for goal assignment, such as linear thresholds or decision trees, at the cost of some performance. Alternatively, they can employ post-hoc explanation techniques to provide workers with rationales. Regulatory frameworks, such as the European Union's proposed AI Act, may require that algorithmic management systems be transparent and subject to human oversight [21].

Policy implications extend to competition law and labor rights. If multiple platforms in the same market each use MARL to assign goals, the resulting dynamics could lead to a race to the bottom in terms of worker conditions. Regulators might need to cap the aggressiveness of goal-setting, akin to maximum working hours regulations. Our analysis suggests that MARL-based platforms could be required to submit their goal-setting policies for audit, much like financial algorithms are audited for market manipulation.

6. Deployment and Infrastructure Considerations

Deploying MARL for labor supply optimization on a large scale requires significant computational infrastructure. Each worker agent may need to run a local policy network on a mobile device, while the platform maintains a centralized meta-policy that processes aggregated data [22]. Communication bandwidth and latency become important: goal updates must be delivered in real-time to influence worker decisions before the next interaction. Federated learning can be used to train agent-level policies without transmitting raw worker data, preserving privacy.

Scalability is a major challenge. With millions of workers, the action space and state space explode. Value decomposition methods like QMIX or VDN reduce the complexity by factorizing the joint Q-function into individual utilities [23]. However, these approaches assume monotonic mixing, which may not hold when goal heterogeneity is high. Alternatively, mean-field MARL approximates the interactions of many agents by a single aggregate signal [24]. This is particularly suited for labor markets where individual workers are small relative to the total population.

Robustness to adversarial behavior is essential. Workers may learn to game the goal-setting system by artificially lowering their own productivity early in a goal period to receive lower future targets, a phenomenon known as ratcheting. The platform can counteract this by using reference-free reward functions or by randomizing goal assignments. Additionally, the platform must guard against coordinated manipulation by groups of workers, which could be modeled as coalitional game theory [25]. Infrastructure for anomaly detection and countermeasures should be integrated into the governance module.

7. Sustainability and Robustness

Long-term sustainability of the platform-worker relationship depends on maintaining adequate worker engagement without causing burnout. Under platform-assigned goals, the MARL policy may exploit workers by setting ever-higher targets, leading to a decay in worker base over time. To prevent this, the platform's reward function should include a term for worker retention, e.g., discounted future participation probability. Multi-objective reinforcement learning can be used to optimize a trade-off between short-term throughput and long-term workforce health [26].

Robustness to environmental changes is another critical dimension. Demand shocks, such as seasonal variations or external events, require the goal-setting meta-policy to adapt quickly. Transfer learning techniques can pre-train the meta-policy on historical data and then fine-tune online [27]. Workers' policies also need to adapt; if they have been over-specialized to a particular goal pattern, sudden shifts may cause suboptimal behavior. Introducing noise or periodic exploration in goal assignment can maintain flexible agent policies.

Finally, fairness over time must be considered. A system that treats all workers equally in the aggregate may still produce transient unfairness, for example by assigning high goals to a

randomly chosen subset early in the day. Equality of opportunity over temporal windows can be enforced by using a sliding window fairness metric in the meta-policy's loss function. The multi-agent setting complicates this because agents' actions affect each other's opportunities, creating a need for coordinated fairness constraints across agents.

8. Conclusion

This paper has presented a multi-agent reinforcement learning perspective on labor supply optimization in digital platforms under self-set and platform-assigned goals. We argued that goal structures are not merely behavioral nudges but fundamentally alter the reward landscape in which workers learn, affecting system efficiency, fairness, and sustainability. Through a system-level architectural discussion, we compared decentralized, centralized, and hybrid goal-setting mechanisms, highlighting trade-offs between autonomy and coordination. We examined governance implications, emphasizing transparency, distributive justice, and regulatory oversight. Deployment challenges related to scalability, privacy, and adversarial robustness were addressed, and we proposed extensions for long-term sustainability.

Future research directions include empirical validation of the proposed MARL framework using real platform data, development of fair and interpretable meta-goal policies, and the integration of multi-platform competition dynamics. As digital labor platforms continue to expand, understanding the interplay between algorithmic goal-setting and worker learning will be essential for designing systems that are both productive and humane.

References

1. Horton, J. J. (2017). The effects of algorithmic labor market matching on worker earnings. Journal of Labor Economics, 35(S1), S319-S351.

2. Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multi-agent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, Part C, 38(2), 156-172.

3. Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 1603-1612.

4. Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.

5. Rogerson, R., Shimer, R., & Wright, R. (2005). Search-theoretic models of the labor market: A survey. Journal of Economic Literature, 43(4), 959-988.

6. Chen, M. K., Chevalier, J. A., Rossi, P. E., & Oehlsen, E. (2019). The value of flexible work: Evidence from Uber drivers. Journal of Political Economy, 127(6), 2735-2794.

7. Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American Psychologist, 57(9), 705-717.

8. Ang, S., & Slaughter, S. A. (2001). Work outcomes and job design for contract versus permanent information systems professionals on software development teams. MIS Quarterly, 25(3), 321-350.

9. Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. International Journal of Communication, 10, 3758-3784.

10. Li, Y. (2017). Deep reinforcement learning: An overview. arXiv preprint arXiv:1701.07274.

11. Shoham, Y., & Leyton-Brown, K. (2009). Multiagent systems: Algorithmic, game-theoretic, and logical foundations. Cambridge University Press.

12. Lin, K., Zhao, R., Xu, Z., & Zhou, J. (2018). Efficient large-scale fleet management via multi-agent deep reinforcement learning. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1774-1783.

13. Littman, M. L. (1994). Markov games as a framework for multi-agent reinforcement learning. Proceedings of the 11th International Conference on Machine Learning, 157-163.

14. Nash, J. (1951). Non-cooperative games. Annals of Mathematics, 54(2), 286-295.

15. Goldsmith, S., & Eggers, W. D. (2004). Governing by network: The new shape of the public sector. Brookings Institution Press.

16. Agrawal, A., Lacetera, N., & Lyons, E. (2016). Does information help or hinder job applicants from less developed countries in online markets? Journal of International Economics, 100, 129-141.

17. Min, X., Chi, W., Hu, X., & Ye, Q. (2024). Set a goal for yourself? A model and field experiment with gig workers. Production and Operations Management, 33(1), 205-224.

18. Tang, X., Qin, Z., Zhang, J., & Ye, Q. (2021). Dynamic goal setting in crowdsourcing: A multi-armed bandit approach. Manufacturing & Service Operations Management, 23(5), 1179-1196.

19. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 214-226.

20. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, 59-68.

21. European Commission. (2021). Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM(2021) 206 final.

22. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273-1282.

23. Rashid, T., Samvelyan, M., Schroeder, C., Farquhar, G., Foerster, J., & Whiteson, S. (2018). QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. Proceedings of the 35th International Conference on Machine Learning, 4295-4304.

24. Yang, Y., Luo, R., Li, M., Zhou, M., Zhang, W., & Wang, J. (2018). Mean field multi-agent reinforcement learning. Proceedings of the 35th International Conference on Machine Learning, 5571-5580.

25. Chalkiadakis, G., Elkind, E., & Wooldridge, M. (2011). Computational aspects of cooperative game theory. Morgan & Claypool Publishers.

26. Van Moffaert, K., & Nowé, A. (2014). Multi-objective reinforcement learning using sets of Pareto dominating policies. Journal of Machine Learning Research, 15(1), 3483-3512.

27. Taylor, M. E., & Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. Journal of Machine Learning Research, 10, 1633-1685.