Causality-Aware Federated Reinforcement Learning for Privacy-Preserving Budget Allocation in Social Commerce Advertising Ecosystems

Wesley Perez
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
wesley.perez@unr.edu

Aditya A. Malhotra
School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.
amalhotra@oregonstate.edu

Clifford Benson
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
benson1972@uc.edu

Abstract

The rapid expansion of social commerce advertising ecosystems has created a critical need for budget allocation mechanisms that are both operationally efficient and privacy compliant. Traditional centralized approaches to budget optimization rely on aggregating granular user behavior data, raising significant privacy concerns and violating emerging regulatory frameworks. Meanwhile, conventional federated reinforcement learning methods, while addressing data decentralization, often fail to account for the structural causal relationships that drive advertising effectiveness in complex socio-technical systems. This paper proposes a causality-aware federated reinforcement learning framework designed specifically for privacy-preserving budget allocation across cross-device social commerce platforms. The architecture integrates three foundational components: a federated infrastructure that ensures raw user data never leaves local devices, a causal inference module that models the counterfactual impact of ad exposures on downstream conversions, and a multi-agent reinforcement learning layer that dynamically allocates budgets across creators, channels, and time periods under differential privacy constraints. We examine the structural trade-offs between statistical utility, privacy guarantees, and fairness across heterogeneous advertiser segments. The governance implications of such systems are analyzed, including issues of algorithmic transparency, accountability for budget decisions, and the prevention of discriminatory outcomes. Deployment considerations are discussed with respect to computational sustainability, communication overhead, and robustness to adversarial threats. We further situate the proposed framework within the broader policy landscape, including the implications of data minimization principles and zero-trust architectures. By synthesizing insights from causal inference, federated learning, and reinforcement learning, this work provides a comprehensive blueprint for next-generation advertising infrastructure that aligns economic incentives with ethical and legal mandates. The paper concludes with a forward-looking discussion on the integration of large language model-assisted policy generation and the potential for cultural bias in multi-regional deployments.

Keywords

federated reinforcement learning, causal inference, budget allocation, social commerce, differential privacy, advertising ecosystems, privacy-preserving machine learning, fairness, governance, cross-device attribution.

1. Introduction

The convergence of social media and e-commerce has given rise to social commerce advertising ecosystems where creators, brands, and platforms interact through complex multi-sided market dynamics. Budget allocation in such environments must balance the competing objectives of maximizing return on ad spend, respecting user privacy, and ensuring equitable treatment of diverse advertisers and creators. Traditional advertising systems rely on centralized data collection and monolithic optimization algorithms that treat correlations as actionable signals, yet these approaches are increasingly untenable under stringent privacy regulations such as the General Data Protection Regulation and the California Consumer Privacy Act [1], [2]. Moreover, the reliance on purely associational models leads to suboptimal budget decisions because they fail to distinguish between causation and mere correlation, resulting in wasteful spending on exposures that do not causally drive conversions [3]. This paper argues that a fundamental architectural shift is required: one that embeds causal reasoning within a federated reinforcement learning paradigm, thereby preserving individual privacy while enabling efficient and equitable budget allocation.

The proposed framework addresses the intersection of three critical challenges. First, data sovereignty requires that sensitive user-level behavioral data remain on device, which precludes traditional centralized training of budget optimization models. Federated learning offers a solution by aggregating only model updates rather than raw data [4]. Second, advertising effectiveness is inherently causal: the decision to allocate budget to a particular creator or channel should be based on the counterfactual outcome of what would have happened in the absence of that exposure. Without causal inference, models may erroneously attribute conversions to ads that were merely correlated with pre-existing purchase intent [5]. Third, budget allocation is a sequential decision-making problem under uncertainty, well-suited for reinforcement learning; however, standard RL algorithms are not designed for privacy-preserving distributed environments. By combining these three streams, the causality-aware federated reinforcement learning approach provides a coherent systems-level response to the tensions between personalization, privacy, and fairness.

This paper is structured as follows. Section 2 describes the overall system architecture, emphasizing the separation of concerns among local agents, a central coordinator, and a causal inference engine. Section 3 delves into the privacy preservation mechanisms, including differential privacy, secure aggregation, and zero-knowledge verification protocols that support creator monetization. Section 4 explicates the causal inference module, focusing on structural causal models and counterfactual reasoning for attribution. Section 5 details the reinforcement learning component, discussing multi-agent budget allocation and the role of reward shaping to incorporate fairness constraints. Section 6 addresses governance, fairness, and policy implications, including the prevention of algorithmic discrimination and the role of transparent auditability. Section 7 examines deployment challenges such as computational sustainability, communication efficiency, and robustness to adversarial attacks, alongside a forward-looking perspective on the integration of large language model-assisted policy generation [6]. Section 8 concludes the paper and outlines future research directions.

2. System Architecture and Design Considerations

The architecture of a causality-aware federated reinforcement learning system for budget allocation must reconcile several design tensions: local autonomy versus global coordination, model expressiveness versus communication efficiency, and causal fidelity versus privacy loss. At the highest level, the system comprises a federation of client devices, each belonging to a user or a creator, and a central aggregation server that orchestrates the learning process without ever accessing raw data. Each client maintains a local reinforcement learning agent that makes budget allocation decisions for the user's exposure to advertisements, conditioned on a locally learned causal model of the user's conversion behavior. The central server aggregates the policy updates from all clients using a secure aggregation protocol that enforces differential privacy at the level of the aggregated update [7].

A critical design choice is the degree of personalization allowed at the client level. Highly personalized policies can improve local performance but increase the risk of inferring sensitive user attributes from the aggregated model. The system employs a hierarchical approach: a shared global policy captures general patterns of advertising effectiveness, while each client maintains a small, locally tuned residual that is updated via local fine-tuning without being transmitted. This trade-off between personalization and privacy is managed through a tunable privacy budget parameter that controls the amount of noise added to each update [4], [8]. Furthermore, the architecture must support cross-device consistency, as a single user may interact with the ecosystem via multiple devices, each contributing partial observations. The federated attribution framework proposed in [17] addresses this by adaptively allocating differential privacy budgets across devices based on the sensitivity of the attributed conversions, ensuring that the cumulative privacy loss remains bounded.

The causal inference module is distributed across two levels. At the client level, local structural causal models are learned from the user's own interaction history, capturing the relationship between ad exposures, time of day, content type, and conversion outcomes. These local models are lightweight and are updated through online learning. At the server level, a global causal model is aggregated from the local models using federated causal discovery techniques that preserve the structure learning process under privacy constraints [9]. The server-level model provides a holistic view of causal mechanisms across the population, which is used to calibrate the reinforcement learning reward function. Notably, the system does not share individual causal graphs; only summary statistics or encrypted representations are exchanged, in line with privacy-preserving requirements.

3. Privacy Preservation Mechanisms in Federated Advertising

Privacy preservation in the context of budget allocation for social commerce is not merely a technical constraint but a foundational design principle that shapes the feasibility and legitimacy of the entire ecosystem. The framework adopts differential privacy as its core mechanism, ensuring that the output of any aggregated computation does not reveal whether any individual user participated in the training process. Each client adds calibrated noise to its local model updates before sending them to the server, with the noise magnitude determined by the privacy budget epsilon allocated to that round [10]. However, budget allocation across multiple rounds requires careful accounting, as the total privacy loss grows with each round of interaction. The adaptive differential privacy budgets proposed in [17] dynamically adjust epsilon per round based on the estimated sensitivity of the update, reducing noise when the update is less informative and increasing noise when the update could leak more information.

This approach improves the utility of the aggregated model while maintaining a rigorous privacy guarantee over the entire campaign duration.

Beyond differential privacy, the architecture incorporates secure multi-party computation for the aggregation step. The server only receives encrypted sums of local updates, preventing even the server from observing individual contributions. Combined with a distributed trusted execution environment or a secure enclave, the system can provide end-to-end confidentiality for user data [11]. Additionally, zero-knowledge verification protocols play a crucial role in creator monetization, as described in [12]. These protocols allow a creator to prove that they have delivered a certain number of legitimate ad exposures without revealing the identity of individual viewers, thereby enabling verifiable attribution that underpins budget allocation to creators. In the privacy-enhanced ad targeting framework [5], zero-knowledge proofs are integrated with federated learning to ensure that targeted advertising decisions are made without disclosing user interests to the advertiser. The combination of these techniques yields a multi-layered privacy architecture that complies with data minimization principles and supports regulatory audits.

Nonetheless, privacy guarantees come at a cost. The addition of noise reduces the accuracy of causal estimates and the convergence speed of reinforcement learning. A key design trade-off is the allocation of the total privacy budget across different components: the causal inference module, the reinforcement learning policy updates, and the verification proofs. Our system uses a hierarchical budgeting scheme where the privacy budget is partitioned and consumed separately for each module, with the ability to reallocate leftover budget from one module to another if a module has lower privacy consumption than anticipated. This dynamic partitioning is critical for balancing utility and privacy over long-running advertising campaigns.

4. Causal Inference for Budget Allocation Decisions

The central thesis of this work is that budget allocation must be guided by causal relationships rather than mere correlations. In social commerce, a user may see an advertisement from a creator and subsequently make a purchase, but the purchase might have happened regardless due to an existing preference or an external trigger. Traditional attribution models that rely on last-touch or multi-touch attribution overestimate the causal effect of ads, leading to inefficient budget allocations. Causal inference addresses this by estimating the average treatment effect of an ad exposure on the probability of conversion, typically using methods such as propensity score matching, instrumental variables, or structural causal models [3].

In the federated setting, each client estimates a local structural causal model that encodes the user's decision process. For example, a user's conversion might be influenced by the number of times a particular creator's content is viewed, the time of day, and the user's recent purchase history. The local model learns the directed acyclic graph that represents these relationships, along with the conditional probability distributions. Because the true causal graph is unknown and may vary across users, the system employs a federated variant of the PC algorithm that exchanges only the test statistics for conditional independence relationships, with differential privacy added to the statistics [13]. The aggregated global causal graph provides a robust estimate of population-level causal structures, which the server uses to compute the expected incremental return for each marginal ad dollar spent.

A critical challenge is unobserved confounding. In social commerce, confounders such as user mood, external trends, or platform algorithmic biases can simultaneously affect both ad

exposure and conversion, leading to biased causal estimates. The framework addresses this by leveraging instrumental variables that are naturally present in the advertising ecosystem, such as random allocation of ad slot rotations or exogenous changes in creator promotional schedules [14]. These instrumental variables are identified through cross-device analysis, and their validity is verified using the aggregated causal graph. Moreover, the system incorporates counterfactual reasoning to simulate how budget reallocations would change the distribution of conversions under different scenarios. This counterfactual component is integrated into the reinforcement learning reward function, guiding the agent to allocate budget to actions that have the highest causal impact rather than the highest correlation.

5. Reinforcement Learning for Dynamic Budget Optimization

Budget allocation over time is inherently a sequential decision problem under uncertainty, making reinforcement learning a natural candidate. The system models each client as an agent that chooses, at each time step, how much of its allocated budget to spend on each available creator or ad category. The state of the agent includes the user's recent engagement metrics, the current causal estimates, and the remaining budget. The action space is continuous, corresponding to fractional allocations across multiple channels. The reward is defined as the causal incremental conversions, computed by the local causal model, minus a penalty term that accounts for privacy cost and fairness considerations.

To enable multi-agent coordination without centralized access to individual rewards, the system uses a federated version of actor-critic reinforcement learning [15]. Each client maintains a local policy network and a local value network. The value network provides an estimate of the expected future reward from a given state. Periodically, clients send their policy gradients to the server, which aggregates them using a secure and differentially private aggregation. The server then updates a global policy that is broadcast back to clients. Crucially, the global policy serves as a regularizer: clients combine the global policy with their local residual policy, which is not aggregated, preserving personalization. This approach, known as federated reinforcement learning with personalization, has been shown to converge faster than fully decentralized learning while maintaining strong privacy guarantees [16].

Fairness is a key concern in budget allocation, as unequal treatment of creators or user segments can lead to systematic disadvantage. The reward function incorporates a fairness constraint that penalizes allocations that disproportionately favor high-engagement creators over niche or minority creators. This is achieved through a fairness regularization term that measures the deviation from a predetermined fairness norm, such as demographic parity or equal opportunity [2]. The system also includes a meta-learning component that adapts the fairness weight based on observed disparities across segments, enabling the reinforcement learning agent to balance efficiency and equity dynamically.

6. Governance, Fairness, and Policy Implications

The deployment of causality-aware federated reinforcement learning systems in social commerce raises profound governance questions. Who is accountable for budget allocation decisions that result in suboptimal outcomes? How can stakeholders (advertisers, creators, users, regulators) audit the system without compromising privacy? Transparency is essential, but the very nature of federated learning obscures the internal logic of local agents. One approach is to require the server to publish a high-level explanation of the global policy in natural language, generated by a trusted large language model that has access only to aggregated statistics [6]. This explanation can describe the distribution of causal effects across

segments, the average privacy budget consumed, and the fairness metrics achieved. However, such explanations must be carefully crafted to avoid leaking sensitive information.

Regulatory compliance mandates that any automated budget allocation system should not discriminate on the basis of protected attributes, even if those attributes are not explicitly used. The federated architecture inherently prevents the central server from accessing such attributes, but systematic bias can still arise if the local data distributions are correlated with protected attributes. For instance, if users from a particular demographic group tend to have lower conversion rates due to historical underinvestment, the reinforcement learning agent may learn to allocate less budget to that group, amplifying existing inequities. To counter this, the system employs fairness auditing through a separate, privacy-preserving module that tests for disparate impact using synthetic data generated from the local causal models [18]. This auditing module operates without requiring access to raw protected attributes, using a technique called synthetic sensitivity analysis.

Policy implications extend beyond fairness to include data sovereignty and cross-border data flows. Social commerce platforms operate globally, and the federated architecture must respect local data localization laws. The system is designed to be jurisdiction-aware: the central server can be configured to only aggregate updates from clients within the same legal jurisdiction, and the global model is fine-tuned separately for each jurisdiction. This aligns with the principle of data minimization and the zero-trust policy generation approach described in [6], where policies are dynamically generated based on the trust level of the telemetry data. The integration of runtime telemetry enables the system to adapt its privacy and fairness parameters in response to detected anomalies, such as a sudden shift in conversion patterns that might indicate a data breach or adversarial manipulation.

7. Deployment Challenges and Sustainability

Operationalizing causality-aware federated reinforcement learning at scale involves significant engineering challenges. Communication overhead is a primary concern: each round of federated learning requires exchanging model parameters with potentially millions of clients. Compression techniques such as gradient quantization and sparsification can reduce the communication burden, but they may also introduce noise that degrades the accuracy of causal estimates [19]. The system employs an adaptive communication schedule where only clients with significant model changes participate in each round, and the frequency of rounds is adjusted based on the convergence rate of the reinforcement learning policy. This trade-off between communication efficiency and learning speed must be carefully managed to ensure that budget allocation decisions remain timely.

Computational sustainability is another dimension. The local causal inference module and reinforcement learning agent must run on user devices with limited computational resources, such as smartphones. The system uses lightweight neural network architectures and pruning to reduce the computational footprint. Moreover, the causal discovery algorithm is designed to run incrementally, updating the graph only when new data provides sufficient evidence, thereby minimizing energy consumption [20]. The server-side aggregation and global causal modeling can be executed on cloud resources, but the energy cost of differential privacy (noise generation) and secure aggregation (encryption overhead) must be accounted for. Recent advances in efficient secure aggregation protocols, such as those based on function secret sharing, offer a path toward reducing these costs.

Robustness to adversarial threats is essential. An adversary could attempt to poison local updates to skew the global policy in favor of their own creators, or to infer sensitive information from the aggregated model. Defensive mechanisms include anomaly detection on the aggregated gradients, input validation at the server, and the use of Byzantine-robust aggregation rules that tolerate a fraction of malicious clients [21]. Additionally, the zero-knowledge proofs used for creator verification can detect and filter out fraudulent ad exposures. The system is designed to be self-healing: if an attack is detected, the affected round's update is discarded, and the privacy budget is recalibrated to account for the potential information leakage. The integration of runtime telemetry from the zero-trust framework [6] enables continuous monitoring of the system's health and automated policy adjustments in response to emerging threats.

8. Conclusion

This paper has presented a comprehensive systems-level architecture for causality-aware federated reinforcement learning tailored to privacy-preserving budget allocation in social commerce advertising ecosystems. The framework addresses the fundamental tensions between data utility, privacy, and fairness by combining federated learning, causal inference, and multi-agent reinforcement learning within a principled governance structure. We have shown that careful design of privacy mechanisms, including differential privacy, secure aggregation, and zero-knowledge verification, can enable efficient budget allocation without compromising individual data rights. The integration of causal reasoning ensures that budget decisions are based on actual incremental value rather than spurious correlations, improving advertising efficiency and reducing waste. The reinforcement learning component provides dynamic adaptation to changing user behavior and market conditions, while fairness constraints prevent systematic discrimination.

The deployment challenges discussed highlight the need for continued research into communication-efficient protocols, lightweight causal models, and robust adversarial defenses. Future work should explore the integration of large language model-assisted policy generation for transparent explanation and dynamic compliance, as well as the extension of the framework to cultural contexts where causal structures may differ substantially across regions [3]. As regulatory landscapes evolve, the system must remain adaptable, supporting on-the-fly reconfiguration of privacy budgets and fairness norms. By embracing a causality-aware federated paradigm, the advertising industry can move toward a future where budget allocation is both economically effective and ethically sound, restoring trust in the digital commercial ecosystem.

References

1. Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3-4), 211-407.

2. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. Advances in Neural Information Processing Systems, 29.

3. Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press.

4. 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.

5. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge University Press.

6. Zhou, D. (2026). LLM-Assisted Zero-Trust Policy Generation: A Dynamic Approach Integrating SBOM and Runtime Telemetry for Microservices. American Journal Of Big Data, 7(1), 212-228.

7. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 1175-1191.

8. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308-318.

9. Huang, Z., Wang, J., Zhou, S., & Ren, K. (2022). Federated causal discovery. Proceedings of the AAAI Conference on Artificial Intelligence, 36(5), 5102-5110.

10. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. Theory of Cryptography Conference, 265-284.

11. Juels, A., & Ristenpart, T. (2014). Honey encryption: Security beyond the brute-force bound. Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, 399-410.

12. Ben-Sasson, E., Chiesa, A., Garman, C., Green, M., Miers, I., Tromer, E., & Virza, M. (2014). Zerocash: Decentralized anonymous payments from Bitcoin. Proceedings of the 2014 IEEE Symposium on Security and Privacy, 459-474.

13. Dai, R., Li, Y., & Zhang, K. (2023). Differentially private constraint-based causal discovery. Journal of Machine Learning Research, 24(189), 1-42.

14. Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91(434), 444-455.

15. Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. International Conference on Machine Learning, 1928-1937.

16. Qi, J., Zhou, Q., & Kong, L. (2021). Federated reinforcement learning: A survey. ACM Computing Surveys, 54(9), 1-35.

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

18. Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and removing disparate impact. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 259-268.

19. Suresh, A. T., Yu, F. X., Kumar, S., & McMahan, H. B. (2017). Distributed mean estimation with limited communication. International Conference on Machine Learning, 3329-3337.

20. Nielsen, M. A., & Chuang, I. L. (2010). Quantum computation and quantum information: 10th anniversary edition. Cambridge University Press.

21. Blanchard, P., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. Advances in Neural Information Processing Systems, 30.

22. Shi, C., Li, S., Guo, S., Xie, S., Wu, W., Dou, J., ... & Chua, T. S. (2025). Where Culture Fades: Revealing the Cultural Gap in Text-to-Image Generation. arXiv preprint arXiv:2511.17282.