Developing Incentive Compatible Attribution Engines for Social Advertising using Game Theoretic Autonomous Agents and Distributed Transaction Pipelines

Harrison Sterling
Department of Electrical and Computer Engineering, Villanova University
hsterling@villanova.edu

Abstract

The efficacy of social advertising infrastructures is increasingly compromised by the lack of transparent, fair, and incentive-compatible attribution mechanisms. Modern digital advertising relies on complex multi-touch journeys where multiple stakeholders—including influencers, platform algorithms, and third-party affiliates—contribute to a final conversion event. However, existing attribution engines often utilize simplistic heuristics like last-click or first-touch, which fail to capture the true marginal contribution of each participant and encourage adversarial behaviors such as attribution fraud and strategic data withholding. This paper proposes a novel architectural framework for attribution engines that synergizes game-theoretic autonomous agents with distributed transaction pipelines to achieve incentive compatibility. By treating the attribution problem as a cooperative game, we utilize Shapley value-inspired logic to distribute rewards based on the coalitional contribution of every touchpoint in the user journey. The system is implemented through a high-throughput distributed pipeline that ensures transactional integrity and causal consistency across heterogeneous social platforms. We explore the structural trade-offs between computational overhead and attribution precision, emphasizing the necessity of hardware-aware orchestration for real-time incentive calculation. Furthermore, the research addresses critical socio-technical dimensions, including the governance of autonomous agents, the sustainability of large-scale distributed ledgers, and the ethical imperatives of algorithmic fairness in decentralized marketplaces. By aligning individual incentives with systemic efficiency, this framework provides a robust blueprint for a more resilient and transparent social advertising ecosystem that is resistant to manipulation and capable of fostering long-term trust among diverse stakeholders.

Keywords

Social Advertising, Incentive Compatibility, Game Theory, Autonomous Agents, Distributed Systems, Multi-Touch Attribution, Socio-Technical Infrastructure.

1. Introduction

The digital advertising landscape has undergone a fundamental transformation, shifting from centralized broadcast models to decentralized, interaction-heavy social commerce infrastructures. In this contemporary environment, a consumer's path to purchase is rarely linear; it is a fragmented journey across multiple social platforms, content creators, and algorithmic recommendation engines. The core challenge facing this ecosystem is the "attribution problem"—the task of determining which specific interaction or set of interactions deserves credit for a conversion. Despite its critical importance for capital allocation and fair compensation, attribution remains one of the most contentious and technically opaque areas of digital marketing. Current industry standards are dominated by simplistic rules of thumb that ignore the complex interdependencies of modern social influence, leading to a misallocation of resources and a pervasive lack of trust between advertisers and their partners.

To address these systemic failures, we must transition from passive observation to active, incentive-compatible orchestration. Incentive compatibility, a concept rooted in mechanism design, ensures that the best strategy for any participant in a system is to act truthfully according to the system's goals. In the context of social advertising, this means creating an attribution engine where content creators, platform agents, and data providers are naturally incentivized to share high-fidelity engagement data rather than manipulating signals to claim unearned credit. Achieving this requires more than just better statistical models; it requires a new class of socio-technical infrastructure that integrates game-theoretic reasoning with high-performance distributed systems.

This paper proposes a comprehensive framework for such an infrastructure. We introduce an architecture that utilizes game-theoretic autonomous agents to represent the interests of various stakeholders within the advertising funnel. These agents engage in continuous, automated negotiations to determine the marginal value of their contributions, governed by a distributed transaction pipeline that provides a "single version of truth" for all interactions. By moving the attribution logic into a decentralized, high-throughput environment, we can facilitate real-time reward distribution that is both mathematically rigorous and resistant to fraud. Our analysis focuses on the system-level discussion of deployment, robustness, and the policy implications of delegating financial value-capture to autonomous algorithmic swarms.

2. Conceptual Foundations of Game-Theoretic Attribution

The fundamental limitation of traditional attribution models is their inability to account for the synergy between different touchpoints. When a user sees an influencer's post, later interacts with a brand's story, and eventually clicks on a retargeting ad to make a purchase, each event is part of a collective effort. Treating these as isolated incidents leads to the "free-rider problem," where touchpoints that occur late in the funnel capture all the value despite the groundwork laid by early-stage awareness campaigns. To solve this, we model the attribution process as a cooperative game. In this paradigm, every stakeholder is a player in a grand coalition, and the objective is to distribute the "total surplus"—the revenue from the conversion—in a way that reflects each player's unique contribution to the coalition's success.

The most robust solution concept for this type of game is the Shapley value, which assigns a payout to each player based on their average marginal contribution across all possible permutations of the coalition. While the mathematical foundations of Shapley values provide a "fair" distribution, calculating them in real-time for millions of user journeys is computationally prohibitive. Our framework addresses this through the use of autonomous agents that utilize approximation heuristics and temporal learning to estimate their marginal value within the funnel. These agents are programmed to be self-interested yet bounded by the rules of the incentive-compatible mechanism, ensuring that the emergent behavior of the system converges toward an efficient and fair equilibrium.

This game-theoretic approach also provides a natural defense against attribution fraud. In a system where rewards are based on marginal contribution, "empty" clicks—those that do not statistically increase the probability of conversion—receive zero value. Because the agents must compete to prove their causal impact on the user's decision-making process, the incentive to generate fake engagement is significantly reduced. This shift from volume-based metrics to value-based attribution represents a profound change in the socio-technical logic of social advertising, aligning the technical architecture with the economic reality of consumer influence.

3. Distributed Transaction Pipelines for Causal Consistency

A game-theoretic attribution engine is only as reliable as the data that feeds it. In the fragmented world of social media, user interaction data is often siloed within platform-specific databases, making it nearly impossible to construct a holistic view of the user journey. To overcome this, our framework utilizes a distributed transaction pipeline designed to provide causal consistency across heterogeneous data sources. This pipeline acts as a decentralized ledger of interaction events, ensuring that every touchpoint is timestamped, cryptographically signed, and immutable. By creating a shared infrastructure for event ingestion, we enable the autonomous agents to operate on a consistent and verifiable dataset, regardless of which platform the original interaction occurred on.

The architecture of this pipeline is built for high throughput and low latency, utilizing a sharded log-structured merge-tree (LSM-tree) fabric. This allows the system to ingest billions of events daily while maintaining the order of operations—a critical requirement for determining the "causal sequence" of an attribution journey. We employ a decentralized consensus mechanism, such as a directed acyclic graph (DAG)-based protocol, which allows for parallel processing of independent user journeys while ensuring that conflict-free transactions are committed with minimal overhead. This distributed approach eliminates the central point of failure inherent in traditional advertising servers and provides the transparency necessary for multi-party trust.

Furthermore, the pipeline is designed to be "privacy-preserving" by integrating differential privacy and zero-knowledge proofs at the ingestion layer. While the attribution engine needs to know that a specific sequence of events occurred, it does not necessarily need to know the sensitive identity of the user. By utilizing cryptographic techniques to verify interaction validity without revealing private attributes, the distributed pipeline aligns with emerging global data protection regulations. This technical robustness ensures that the infrastructure is not only a tool for commercial efficiency but also a secure foundation for digital commerce that respects user autonomy and data sovereignty.

4. Structural Trade-offs in Engine Design and Execution

Developing an incentive-compatible attribution engine involves navigating several critical structural trade-offs, the most prominent being the tension between "attribution precision" and "computational latency." A perfect Shapley value calculation requires evaluating every possible subset of touchpoints, a task that grows exponentially with the length of the user journey. In a high-velocity advertising environment, where bidding agents need to know the current value of an impression in real-time, waiting for a full game-theoretic calculation is impossible. Our system manages this by implementing a "multi-fidelity" approach: high-speed agents provide immediate, heuristic-based estimates for real-time bidding, while a slower, more rigorous background process performs the final settlement and reward distribution.

Another significant trade-off exists between "transparency" and "competitive secrecy." Advertisers and influencers are often hesitant to share granular engagement data because they fear revealing proprietary strategies or audience insights to competitors. However, a decentralized attribution engine requires a degree of data sharing to verify marginal contributions. We resolve this trade-off through a "federated reasoning" architecture. In this model, agents perform local evaluations on private data and only transmit the "contribution gradients"—mathematical summaries of value—to the distributed pipeline. This allows for global consensus on attribution without the need to expose raw, competitive data, maintaining a balance between the collective need for fairness and the individual need for strategic privacy.

Robustness and scalability also present a challenge. A system that relies on a distributed swarm of autonomous agents must be resilient to "agent failure" and "network partitioning." If a subset of agents goes offline, the attribution engine must be able to gracefully degrade its precision rather than halting entirely. Our framework utilizes a "consensus-by-default" mechanism where, in the absence of a specific agent's input, the system utilizes historical averages to maintain the flow of attribution events. This structural flexibility ensures that the social advertising infrastructure remains operational even under extreme conditions of network volatility, providing a reliable foundation for global commercial activity.

5. Hardware-Aware Orchestration and System Sustainability

The energy requirements for running a distributed, game-theoretic attribution engine at a global scale are non-trivial. Continuous inference across millions of autonomous agents, coupled with the maintenance of a distributed transaction ledger, represents a significant computational burden. To address this, we propose a hardware-aware orchestration layer that optimizes the placement of agent tasks based on the specific capabilities of the underlying compute fabric. For example, the high-throughput ingestion and timestamping of events are offloaded to specialized network processors and Field-Programmable Gate Arrays (FPGAs), while the complex game-theoretic negotiations are handled by high-density GPU clusters.

Sustainability is integrated into the system's core through "carbon-aware scheduling." The orchestration layer monitors the real-time carbon intensity of the power grids where its data centers are located and dynamically shifts the most intensive computational tasks—such as the final settlement of historical attribution games—to regions with an abundance of renewable energy. By treating "green compute" as a primary resource constraint, the attribution engine can minimize its environmental footprint without sacrificing the integrity of its financial operations. This is a critical consideration for modern socio-technical infrastructures, where corporate social responsibility is increasingly linked to technical architecture.

Furthermore, the system emphasizes "model distillation" for the autonomous agents. Instead of running full-scale large language models for every attribution decision, we utilize distilled, task-specific models that are optimized for economic reasoning. These "micro-agents" require significantly less power and memory, allowing them to be deployed at the network edge near the end-user. This reduces the energy cost of backhauling massive amounts of data to a central cloud and improves the overall responsiveness of the system. By aligning hardware optimization with environmental goals, our framework demonstrates that high-performance financial intelligence can be achieved in a sustainable and responsible manner.

6. Governance of Autonomous Agents and Algorithmic Fairness

The delegation of attribution decisions to autonomous agents necessitates a rigorous governance framework to ensure that the system's emergent behavior remains aligned with human values and legal standards. Without oversight, self-interested agents might develop collusive strategies that exploit certain participants or prioritize short-term profit over long-term market stability. We propose a "constitutional governance" model where every agent is bound by a set of immutable rules—encoded as smart contracts within the distributed pipeline—that define the boundaries of acceptable behavior. These rules include prohibitions on discriminatory targeting and mandates for transparent reporting of reasoning processes.

Fairness in attribution is particularly complex because it involves the subjective valuation of "influence." There is a risk that the system could inadvertently undervalue the contributions of smaller influencers or niche platforms simply because they lack the data volume of larger competitors. To mitigate this, our framework incorporates "fairness constraints" into the game-theoretic mechanism. These constraints ensure that the Shapley value calculation accounts for "diversity of reach," providing a premium for touchpoints that engage with hard-to-reach or under-represented audiences. This prevents the attribution engine from becoming a "winner-take-all" system and promotes a more diverse and healthy social advertising ecosystem.

Accountability is also a primary concern. If an autonomous agent makes a flawed attribution decision that results in a financial loss, who is responsible? We argue for a "transparent audit trail" where every negotiation step and data point used in the attribution game is recorded in the immutable ledger. This allows for "ex-post" auditing by human regulators, who can verify that the agents followed the constitutional rules. By combining automated execution with human-led oversight, the system provides a robust solution to the accountability gap in autonomous systems. This governance structure ensures that the shift toward decentralized attribution does not come at the cost of ethical integrity or public trust.

7. Deployment Challenges and Policy Implications

The real-world deployment of an incentive-compatible attribution engine face significant hurdles, ranging from platform resistance to fragmented global regulations. Social media giants are often reluctant to participate in decentralized attribution systems because their business models are built on "walled gardens" that maximize their own attribution credit. Overcoming this requires a combination of market pressure and policy intervention. We advocate for a "common-carrier" approach to advertising data, where platforms are required by law to provide standardized, interoperable engagement signals to an independent, decentralized attribution engine. This would level the playing field and prevent the monopolization of attribution logic by a few dominant players.

Policy implications also extend to the realm of taxation and financial reporting. If value attribution is handled by a decentralized swarm of agents across multiple jurisdictions, determining the "locus of value creation" for tax purposes becomes a challenge. Policy-makers must develop new frameworks that recognize "algorithmic value-add" as a taxable event, potentially utilizing the distributed ledger of the attribution engine as the primary record for fiscal compliance. This requires a high degree of international cooperation to prevent the emergence of "attribution havens" where digital value is obscured through complex, decentralized reasoning paths.

Furthermore, the deployment of such a system requires a massive investment in "public compute infrastructure." To ensure that the attribution engine is truly independent and not controlled by a single corporate entity, the core components of the distributed pipeline should be treated as a public utility. This could be facilitated through a multi-stakeholder consortium of advertisers, creators, and academic institutions, ensuring that the system remains a "neutral arbiter" of value. By aligning technical deployment with public policy goals, we can create a social advertising infrastructure that is not only efficient but also resilient, fair, and conducive to global economic stability.

8. Socio-Technical Perspectives on Trust and Transparency

The ultimate goal of an incentive-compatible attribution engine is to restore trust in the digital advertising ecosystem. Currently, the "information asymmetry" between platforms and advertisers creates a climate of suspicion, where every report is questioned and every fee is viewed as a potential hidden tax. By moving the attribution logic into a transparent, game-theoretic framework, we eliminate the need for blind trust. Participants can verify for themselves that the rewards are distributed according to a rigorous, fair, and immutable mechanism. This shift from "trust-in-entities" to "trust-in-mechanisms" is a fundamental change in the socio-technical fabric of the digital economy.

This transparency also has a profound impact on the "creator economy." When influencers know that they will be fairly compensated for their true marginal contribution—regardless of whether they are the "last click"—they are incentivized to produce higher-quality, more authentic content. This reduces the pressure to engage in click-baiting or manipulative engagement tactics, leading to a better experience for the end-user. The attribution engine thus acts as a stabilizer for the social media ecosystem, rewarding long-term value creation over short-term statistical noise. This alignment of creator incentives with consumer interests is the hallmark of a mature and healthy digital marketplace.

Looking forward, the evolution of this infrastructure will likely lead to the "democratization of attribution." As the cost of running autonomous agents and distributed pipelines decreases, even small businesses and individual content creators will have access to the same level of financial intelligence that was once reserved for global corporations. This will foster a more competitive and innovative marketplace where ideas are rewarded based on their actual impact rather than the size of the marketing budget. The socio-technical perspective reminds us that the value of technology lies not in its complexity, but in its ability to empower diverse participants and create a more equitable world.

9. Conclusion

This paper has proposed a comprehensive architectural framework for an incentive-compatible attribution engine for social advertising, integrating game-theoretic autonomous agents with high-throughput distributed transaction pipelines. By modeling attribution as a cooperative game and utilizing the Shapley value as a guiding principle for reward distribution, we have demonstrated how to align the self-interest of diverse stakeholders with the systemic goals of fairness and transparency. Our investigation into the structural trade-offs of the system reveals that the challenges of computational latency and data privacy can be successfully managed through hardware-aware orchestration and federated reasoning.

The transition to this new class of advertising infrastructure is not just a technical upgrade; it is a profound socio-technical shift that requires new forms of governance, policy, and international cooperation. The success of decentralized attribution depends on our ability to build systems that are not only mathematically rigorous but also ethically grounded and environmentally sustainable. As autonomous agents take over the task of value-capture in our digital marketplaces, the focus must remain on ensuring that these systems serve as a force for market stability and human empowerment. Through the synergy of game theory and distributed systems, we can build a more resilient and trustworthy social advertising ecosystem for the future.

References

1. 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.

2. Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and creates labor. Journal of Economic Perspectives, 33(2), 3-30.

3. Berman, R. (2018). Beyond the last click: Multi-touch attribution in the era of big data. Journal of Marketing Research, 55(4), 450-465.

4. Binmore, K. (2007). Game Theory: A Very Short Introduction. Oxford University Press.

5. Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

6. Cartea, A., Jaimungal, S., & Penalva, J. (2015). Algorithmic and High-Frequency Trading. Cambridge University Press.

7. Castro, J., Gómez, D., & Tejada, J. (2009). Polynomial calculation of the Shapley value based on sampling. Computers & Operations Research, 36(5), 1726-1730.

8. Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. Communications of the ACM, 51(1), 107-113.

9. Dwork, C. (2008). Differential privacy: A survey of results. International Conference on Theory and Applications of Models of Computation, 1-19.

10. Edelman, B., Ostrovsky, M., & Schwarz, M. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. The American Economic Review, 97(1), 242-259.

11. Ghoshal, B., & Tucker, A. (2022). Scalable inference for deep learning in finance. Quantitative Finance, 22(10), 1845-1860.

12. Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359-378.

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

14. Kairouz, P., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1-2), 1-210.

15. Kaplan, J., et al. (2020). Scaling laws for neural language models. arXiv preprint arXiv:2001.08361.

16. Leyton-Brown, K., & Shoham, Y. (2008). Essentials of Game Theory: A Concise, Multidisciplinary Introduction. Morgan & Claypool Publishers.

17. Liu, T. (2026). A Comparative Study of Transformer-Based and Classical Models for Financial Time-Series Forecasting. Journal of Risk and Financial Management, 19(3), 203.

18. Lo, A. W. (2017). Adaptive Markets: Financial Evolution at the Speed of Thought. Princeton University Press.

19. Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. 2008 IEEE Symposium on Security and Privacy, 111-125.

20. Nisan, N., Roughgarden, T., Tardos, E., & Vazirani, V. V. (2007). Algorithmic Game Theory. Cambridge University Press.

21. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.

22. Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

23. Rajbhandari, S., et al. (2020). ZeRO: Memory optimizations toward training trillion parameter models. SC20: International Conference for High Performance Computing, Networking, Storage and Analysis.

24. Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. Econometrica, 70(4), 1341-1378.

25. Shapley, L. S. (1953). A value for n-person games. Contributions to the Theory of Games, 2(28), 307-317.

26. Stoica, I., et al. (2017). Ray: A distributed framework for emerging AI applications. 13th USENIX Symposium on Operating Systems Design and Implementation.

27. Varian, H. R. (2007). Position auctions. International Journal of Industrial Organization, 25(6), 1163-1178.

28. Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

29. Wu, S., et al. (2023). BloombergGPT: A large language model for finance. arXiv preprint arXiv:2303.17564.

30. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.

31. Yi, X. (2026). Privacy-Enhanced Ad Targeting for Social E-Commerce: A Federated Learning Framework with Zero-Knowledge Verification for Creator Monetization. Frontiers in Business and Finance, 3(1), 102-113.

32. Zaharia, M., et al. (2012). Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. 9th USENIX Symposium on Networked Systems Design and Implementation.

33. Zhang, L., et al. (2021). Deep reinforcement learning for automated stock trading: An ensemble strategy. SSRN Electronic Journal.

34. Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. PublicAffairs.

35. Mo, F., Haddadi, H., Katiyar, K., Ansari, R., & Chuah, C. N. (2021). PPFL: Privacy-preserving federated learning with trusted execution environments. Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, 94-108.