European Journal of AI, Computing & Informatics

Vol. 1 No. 4 2025

Article **Open Access**

Causal Attribution and Evaluation Based on Large-Scale Advertising Data

Jing Xie [1,*]

[1] Steinhardt School of Culture, Education, and Human Development, New York University, New York, NY 10003, USA

* Correspondence: Jing Xie, Steinhardt School of Culture, Education, and Human Development, New York University, New York, NY 10003, USA

Received: 18 October 2025
Revised: 24 October 2025
Accepted: 16 November 2025
Published: 21 November 2025

Abstract: As the reach of digital advertising expands and multi-touchpoint data volumes surge exponentially, traditional advertising effectiveness evaluation methods based on correlation analysis can no longer accurately and effectively reflect the impact level from advertising touchpoints to behavioural conversion. This study constructs a causal attribution model and advertising effectiveness evaluation system based on causal principles, suitable for large-scale advertising data. Through latent outcome frameworks and causal effect estimation techniques, the model identifies the trajectory of advertising touchpoint influence. It further constructs an indicator system for causal effectiveness, primarily comprising average causal effects, behavioural uplift rates, and causal return on investment, establishing a systematic framework from causal modelling to effectiveness evaluation. The research achieves precise quantitative interpretation of advertising effectiveness, providing scientific grounds for formulating rational advertising allocation methods and enabling intelligent decision-making.

Keywords: large-scale advertising data; causal attribution; effectiveness evaluation; causal inference model

1. Introduction

Amidst the rapid evolution of the digital advertising landscape, campaigns now exhibit multi-touchpoint, highly interactive, and data-intensive characteristics. Gauging the actual impact of different touchpoints and their role in conversion remains a core task in advertising research. Traditional attribution methods, however, primarily rely on experience or correlation, failing to uncover the underlying causal relationships behind advertising effects. Concurrently, existing evaluation systems lack explicit, consistent logic, potentially leading to biased assessments of advertising effectiveness. Consequently, drawing upon causal inference theory offers a novel research direction for processing advertising data. From this perspective, this study aims to establish a framework for causal attribution and effectiveness evaluation of large-scale advertising data, thereby enabling the scientific quantification and interpretation of advertising outcomes [1].

2. Theoretical Foundations and Research Framework of Causal Attribution

2.1. Theoretical Foundations of Causal Attribution Research

The core of causal attribution lies in revealing genuine causal relationships between variables, rather than mere statistical correlations. In advertising data analysis, both ad

exposure and conversion behaviour are frequently confounded by factors such as user characteristics, timing, and external environments. Relying solely on correlation metrics fails to establish the genuine impact of advertising. Causal inference theory provides a scientific pathway for identifying such "intervention-outcome" relationships. Its fundamental concept originates from the Potential Outcome Framework, which posits that each individual possesses two potential outcomes: one under the influence of an advertisement and one without [2].

$$ATE = E[Y(1) - Y(0)] \quad (1)$$

In the equation, Y (1) denotes the outcome of an individual after receiving the advertising intervention, while Y (0) represents the outcome without intervention. The expected difference between these two values constitutes the Average Treatment Effect (ATE).

To enhance the reliability of estimation, research typically introduces the Conditional Average Treatment Effect (CATE):

$$CATE(x) = E[Y(1) - Y(0)|X = x] \quad (2)$$

Here, X denotes the set of confounding variables employed to control for user-level differences. In practical applications, methods such as Propensity Score Matching (PSM), Inverse Probability Weighting (IPW), and Double Robust Estimation (DR) are utilised to correct for confounding bias, approximating the reconstruction of counterfactual scenarios from observed data. Causal graph models further provide a structured visual representation, enabling the clear presentation of causal pathways between ad exposure, user characteristics, and conversion outcomes. This offers theoretical underpinnings for subsequent model construction [3].

2.2. Research Framework for Advertising Causal Attribution

The research objective of advertising causal attribution is to identify the true contribution of advertising touchpoints to conversion behaviour. Within multi-touchpoint interaction environments, traditional "last-click" or "first-click" attribution methods struggle to reflect the cumulative and interactive effects between advertisements. The introduction of causal inference methods enables researchers to calculate the marginal contribution of each advertising touchpoint to conversion while controlling for confounding variables, thereby enhancing the scientific rigour and interpretability of attribution [4].

The comprehensive framework for advertising causal attribution comprises four stages: data input, causal modelling, effect estimation, and outcome evaluation. The data input stage involves collecting and feature engineering logs of ad impressions, clicks, and conversions. The causal modelling stage identifies variable relationships based on causal graph structures. The effect estimation stage calculates the treatment effects of advertising touchpoints. The outcome evaluation stage translates these estimates into quantifiable metrics such as average causal effects and conversion rate lift [5].

Figure 1 illustrates the comprehensive research framework for advertising causal attribution. This framework emphasises a systematic approach to causal inference within advertising data analysis: grounded in data, centred on models, and oriented towards evaluation, forming a complete logical closed loop.

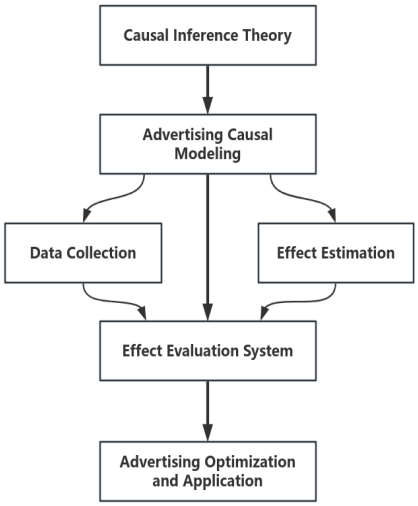

Figure 1. Overall Framework Diagram for Advertising Attribution Based on Causal Inference.

The establishment of this framework not only shifts advertising effectiveness analysis from empirical judgement to causal explanation, but also provides the theoretical basis and methodological foundation for subsequent advertising optimisation and intelligent placement [6].

3. Construction of an Advertising Attribution Model Based on Causal Inference

3.1. Model Setup and Variable Definition

In advertising causal analysis, to identify the genuine causal relationship between ad exposure and conversion behaviour, one must clarify the data generation mechanism and its identification conditions. Following the principles of causal inference, advertising effectiveness itself is often influenced by factors such as user characteristics, events, and context. Failure to control for these factors will lead to systematic bias in predictive outcomes. To ensure model identifiability, it must be assumed that, given the covariates, the advertising treatment is independent of the potential outcome, i.e.:

$$\big(Y(1), Y(0)\big) \perp T|X \tag{3}$$

Here, Y (1) and Y (0) denote the potential outcomes under conditions of advertisement exposure and non-exposure respectively. T represents the advertisement treatment status (whether exposed), while X denotes the set of covariates encompassing user characteristics, platform type, time, and environmental factors. ⊥ signifies conditional independence. This assumption ensures that, with covariates X fully controlled, the advertisement treatment assignment remains unaffected by potential outcomes, thereby eliminating confounding bias [7].

Simultaneously, to ensure that both exposed and unexposed samples exist across different feature combinations, the overlap condition must also be satisfied:

$$0 < P(T = 1|X) < 1 \tag{4}$$

Here, P(T=1|X) denotes the conditional probability of an advertisement exposure given the covariate X, which must lie between 0 and 1 to ensure comparability across all sample categories.

If the advertising causal model simultaneously satisfies the conditional independence and overlapping assumptions, counterfactual scenarios can be modelled from observational data. This enables estimation of the causal effect of advertising touchpoints, providing insights and justification for subsequent advertising effectiveness estimation and model validation.

3.2. Causal Effect Estimation Methods

After satisfying conditional independence and overlapping, the key lies in recovering comparable outcomes for "with/without advertising exposure" from observational data. Common strategies include Propensity Score Matching (PSM) and Inverse Probability Weighting (IPW), which respectively mitigate systematic bias from confounding through "matching" and "weighting".

Propensity scores first measure the probability of receiving advertising exposure given covariates:

$$e(X) = P(T = 1|X) \tag{5}$$

Here, e(X) denotes the propensity score, P(.) represents the probability operator, T indicates the treatment indicator (1 signifies ad exposure, 0 signifies no exposure), and X denotes the set of covariates. Propensity score matching (PSM) based on e(X) pairs samples with similar propensity scores, rendering the exposed and unexposed groups more comparable in covariate distribution and thereby mitigating confounding bias [8].

IPW corrects for treatment allocation imbalance through re-weighting, with its estimator expressed as:

$$\hat{\tau}_{IPW} = \frac{1}{n}\sum_{i-1}^{n}\left[\frac{T_i Y_i}{\hat{e}(X)} - \frac{(1-T_i)Y_i}{1-\hat{e}(X_i)}\right] \tag{6}$$

Here, $\hat{\tau}_{IPW}$ denotes the estimated average causal effect, n represents the sample size, Y_i denotes the individual outcome (e.g., whether conversion occurred), T_i denotes the individual treatment indicator, and $\hat{e}(X_i)$ denotes the individual propensity score derived from data fitting. This formulation achieves inter-group balance by amplifying the sample weights of "rare treatment states", thereby approximating random allocation in the weighted sample.

For highly heterogeneous and non-linear scenarios, machine learning approaches can enhance estimation robustness. For instance, CausalForest employs decision tree-based methods to estimate heterogeneous effects across subgroups, while deep neural networks (such as DragonNet) can model non-linear and complex processes. Such approaches, when combined with PSM/IPW strategies, can enhance the effectiveness of matched weighting by employing more precise function approximators to estimate e(X) or construct outcome models.

To facilitate method selection based on data characteristics, Table 1 compares the core principles, advantages, limitations, and applicable scenarios of PSM, IPW, and machine learning-based causal inference methods [9].

Table 1. Comparison of Common Causal Effect Estimation Methods.

Method	Core Principle	Advantage	Limitation	Typical Application Scenario
PSM	Match similar samples to eliminate confounding	Simple and intuitive, strong interpretability	Difficult to match in high-dimensional settings	Number of features < 20, sample size 1k-5k
IPW	Balance distributions through sample weighting	Theoretically robust, easy to implement	Extreme weights lead to high variance	Sample size > 5k, balanced weight distribution
Machine Learning Methods	Model nonlinear relationships using tree models or deep networks	Capable of handling high-dimensional and heterogeneous effects	Computationally complex, weak interpretability	Number of features > 50, sample size > 10k

Different causal effect estimation methods vary in data dimensionality and complexity. Selecting an appropriate estimation approach based on data dimensions and sample characteristics allows for balancing model interpretability and computational efficiency, thereby enhancing the validity of advertising attribution results [10].

3.3. Model Identification and Validation Mechanisms

Following the construction of causal models, it is essential to ensure their identifiability and predictive efficacy to guarantee the accuracy of derived advertising benefits. The core of model identification lies in testing the validity of causal inference assumptions and employing robustness analyses to verify whether predictions are influenced by sample selection or strategy choices. Consequently, significance tests and repeated calculations are typically employed to assess stability.

In large-scale advertising data analysis, the Bootstrap method is widely used to construct confidence intervals, calculated as follows:

$$CI = \tau \pm z_{a/2} \times SE(\tau) \quad (7)$$

Here, CI denotes the confidence interval, τ represents the estimated average causal effect, SE is the standard error, and $z_{\alpha/2}$ is the critical value corresponding to the significance level. This method estimates variance through repeated sampling to gauge the stability of model parameters.

Furthermore, the validity of causal inferences can be verified through sensitivity analysis and bias diagnostics. For instance, altering confounding variables or excluding outliers may reveal whether predictions undergo significant shifts. Consistent inferences across different sample partitions and algorithmic settings indicate a robust causal relationship, signifying high reliability in advertising effectiveness.

3.4. Systematic Framework for Causal Attribution Modelling

To achieve systematic identification and quantitative assessment of advertising causality, an integrated technical framework spanning data collection, model estimation, and result interpretation must be established. Grounded in causal inference theory, this framework synthesises advertising exposure, click-through, and conversion data to identify genuine impact through modelling and effect estimation. The system comprises four interconnected stages: data input, causal modelling, effect estimation, and result output. These stages form a closed-loop structure with logical continuity, enabling both estimability and testability of causal effects. As illustrated in Figure 2, this framework constructs a complete causal analysis chain from data input to result output, embodying the holistic logic of advertising causal attribution.

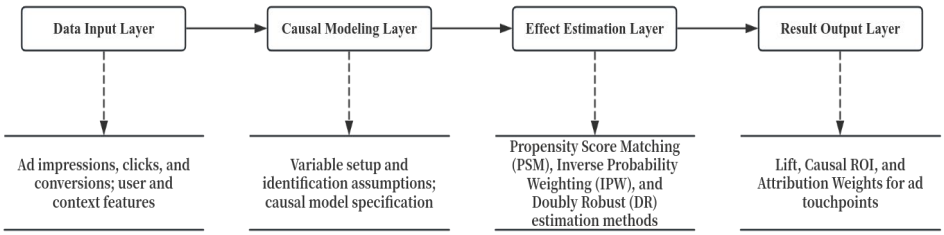

Figure 2. System Framework Diagram of the Advertising Causal Attribution Model.

This framework enables the systematic identification of causal effects in advertising from data collection to outcome delivery, demonstrating the feasibility and practical value of causal inference methods in advertising analysis. It provides methodological support for subsequent effectiveness evaluation systems.

4. Building an Advertising Effectiveness Evaluation System Based on Causal Inference

4.1. Designing the Effectiveness Evaluation Metrics System

In advertising causal analysis, constructing an effectiveness evaluation system is crucial for measuring model value and validating causal inference results. Traditional advertising evaluations predominantly rely on correlation metrics such as click-through rates and impressions, which fail to reveal the true incremental impact of advertising touchpoints. An effectiveness evaluation system grounded in causal inference should centre on "causal effects", reflecting the genuine influence of advertising through dimensions such as incremental contribution, return on investment, and touchpoint synergy.

The commonly used metric for measuring the genuine uplift effect of advertising is Incremental Lift, calculated as follows:

$$Lift = \frac{E[Y|T=1]-E[Y|T=O]}{E[Y|T=0]} \quad (8)$$

Here, E[Y|T=1] denotes the average conversion rate for the ad-exposed group, while E[Y|T=0] represents the average conversion rate for the unexposed group. The Lift value reflects the relative incremental conversions attributable to ad exposure.

To further systematise advertising effectiveness analysis, a causal evaluation framework incorporating multi-dimensional metrics may be established, as shown in Table 2. This framework serves to uniformly measure the economic value and contribution of advertising touchpoints.

Table 2. Advertising Causal Effect Evaluation Indicator System.

Indicator Category	Indicator Name	Calculation Objective	Interpretive Meaning
Causal Effect Indicator	Conversion Lift	Measure the incremental effect of ad exposure	Reflects the true improvement in conversion
Revenue Effect Indicator	Causal ROI	Calculate the revenue return generated by advertising	Measures the input-output ratio
Structural Effect Indicator	Contribution Weight	Allocate the impact of each touchpoint on conversion	Reflects the collaborative value of touchpoints

This indicator system achieves a transition from single-metric conversion measurement to causal effect analysis, providing a benchmark for subsequent measurement methodologies and computational workflows.

4.2. Measurement Methods and Computational Workflow for Causal Evaluation

Within the causal inference framework, the core of advertising effectiveness evaluation lies in translating model estimates into quantifiable economic and behavioural metrics. By measuring the incremental revenue generated from advertising exposure, refined and traceable advertising decisions can be achieved. The evaluation process generally comprises four stages: data preparation, effect estimation, indicator calculation, and result output.

During the data preparation stage, the system extracts exposure status, conversion outcomes, and relevant covariates from advertising display, click, and conversion logs, followed by feature cleaning and standardisation. During effect estimation, the causal model established in the preceding chapter calculates the treatment effect value for each advertising touchpoint, forming an individualised causal effect distribution. In the metric calculation phase, these estimates are transformed into causally evaluated metrics with

clear economic significance. The most representative metric is Causal Return on Investment (Causal ROI), calculated as follows:

$$Causal\ ROI = \frac{Incremental\ \ Re\,uenue}{Aduerti\,sin\,g\ \ Cost} \tag{9}$$

Here, Incremental Reuenue denotes the incremental revenue generated by ad impressions, while Aduertising Cost represents the corresponding advertising expenditure. This metric measures the genuine economic return on advertising investment, reflecting the overall efficiency of advertising performance.

The output phase integrates causal effect estimates with economic indicators to generate results, thereby converting behavioural causal effects into economic benefit metrics. This facilitates the transition of advertising evaluation from statistical to economic impact assessment, providing quantifiable reference points for campaign optimisation.

4.3. Significance and Robustness Testing in Causal Evaluation

To ensure precision in advertising causal evaluation outcomes, model outputs must undergo significance testing and robustness verification. Significance testing primarily determines whether advertising impacts possess statistical significance, whilst robustness testing examines the stability of predictive outcomes across diverse sample sets and algorithmic configurations. These tests serve as crucial diagnostic tools for identifying model errors and verifying the reliability of causal inference.

In causal inference practice, the standardised mean difference (SMD) is a commonly employed measure of balance. Its calculation formula is:

$$SMD_k = \frac{\bar{X}_{1k}\bar{X}_{0k}}{s_{p,k}}, \quad s_{p,k} = \sqrt{\frac{8_{1k}^2+8_{0k}^2}{2}} \tag{10}$$

Here, $\bar{X}_{1k}$ and $\bar{X}_{0k}$ denote the mean values of covariate X_k in the exposed and unexposed groups respectively, 8_{1k}^2 and 8_{0k}^2 represent the corresponding variances, and $s_{p,k}$ is the pooled standard deviation. The absolute value of SMD serves to measure the balance of covariate matching; typically, when $|SMDk| < 0.1$, the balance is considered satisfactory.

Beyond balance tests, sensitivity analyses and sample re-evaluation may also be employed to assess the robustness of estimated outcomes. Should the model yield consistent results across different sample partitions, algorithms, and adjustment variables, the causal evaluation findings may be deemed robust and reliable, indicating that the advertising effect possesses both statistical significance and interpretative validity.

4.4. Practical Significance and Deployment Potential of the Causal Evaluation Framework

The causal inference-based advertising effectiveness evaluation system demonstrates strong operational feasibility. It quantifies the contribution of advertising touchpoints, providing quantitative grounds for allocating advertising resources and adjusting placement strategies to enhance return on investment and budget efficiency. At the application level, this framework can be integrated into advertising management and performance monitoring platforms to enable real-time monitoring, refreshing, and updating of causal model changes. Its modular architecture facilitates integration with diverse channel information systems, enabling cross-platform, multi-perspective advertising effectiveness comparisons and causal analysis. Through continuous refinement of metric systems and algorithmic rules, the causal evaluation model can establish a universal analytical paradigm applicable to digital advertising, content delivery, and user behaviour research.

Conclusion: This study employs causal inference to uncover genuine causal relationships between advertising exposure and consumer conversion, demonstrating its validity and robustness within large-scale data environments. Causal evaluation results demonstrate that advertising effects can be modelled and validated through quantifiable, interpretable inference frameworks, establishing a closed-loop pathway from theoretical

derivation to practical measurement. Methodologically, this research refines the technical framework for advertising causality analysis. Application-wise, it validates the feasibility and scalability of causal inference within advertising evaluation, charting new research directions for causally-driven advertising optimisation, cross-platform assessment, and intelligent placement decisions.

5. Conclusion

This study proposes a rigorous and scalable framework for evaluating advertising effectiveness based on causal inference. By replacing traditional correlation-based indicators with causally interpretable metrics, the framework establishes a complete analytical pathway that spans metric system construction, treatment-effect estimation, robustness verification, and economic value translation. The results demonstrate that causal inference is not only capable of identifying the true incremental contribution of advertising exposure but also provides a theoretically grounded and operationally feasible basis for budget allocation and campaign optimization.

From a methodological perspective, this research strengthens the technical foundation of causal advertising analytics by formalising uplift-based metrics, introducing Causal ROI for economic interpretation, and incorporating significance and robustness tests to ensure model reliability. From an application standpoint, the proposed evaluation system can be embedded into real-world advertising management platforms, enabling cross-channel causal comparison, real-time effect monitoring, and intelligent decision support. Its modular and extensible architecture ensures deployability across diverse advertising ecosystems.

Nevertheless, the framework still faces several potential limitations, including the dependence on high-quality observational data, sensitivity to unobserved confounders, and computational complexity in large-scale multi-touchpoint scenarios. Future research should explore advanced causal modelling approaches-such as double machine learning, causal forests, and structural deep learning-to enhance the accuracy and interpretability of advertising effect estimation. Additionally, expanding the system to address cross-device attribution, long-term brand effects, and dynamic treatment regimes represents promising directions for further investigation.

Overall, this study establishes an end-to-end causal evaluation paradigm with both theoretical rigor and practical value, providing a scientific and trustworthy foundation for advertising optimisation, cross-platform assessment, and the development of intelligent, causality-driven advertising systems.

References

1. P. Gujar, S. Panyam, and V. Pissaye, "Elevating Digital Advertising by Streamlining Agency Client Collaboration through Cloud,". doi: 10.14445/22312803/ijctt-v72i5p110
2. S. Almahmoud, B. Hammo, B. Al-Shboul, and N. Obeid, "A hybrid approach for identifying non-human traffic in online digital advertising," *Multimedia Tools and Applications*, vol. 81, no. 2, pp. 1685-1718, 2022. doi: 10.1007/s11042-021-11533-4
3. N. Martin, and M. K. Mayan, "Optimization of SMART production inventory model with E-logistics and digital advertising costs parameters together with advertising errors," In *AIP Conference Proceedings*, November, 2022, p. 320010. doi: 10.1063/5.0108503
4. E. O. Ajike, J. A. Aderimiki, A. G. Bamidele, and N. Idowu, "Customer purchase decisions of clothing amongst students in Nigerian private universities: The effect of digital advertising," *Innovative Marketing*, vol. 20, no. 3, p. 209, 2024.
5. A. N. Odoh, "The Role of AI-driven Digital Advertising in Sustainable Real Estate Marketing in Nigeria," In *8th International Academic Conference on Research in Social Sciences.*, 2024.
6. J. Porter, "Commentary: inefficiencies in digital advertising markets: evidence from the field," *Journal of Marketing*, vol. 85, no. 1, pp. 30-34, 2021. doi: 10.1177/0022242920970133
7. W. M. Lim, S. Gupta, A. Aggarwal, J. Paul, and P. Sadhna, "How do digital natives perceive and react toward online advertising? Implications for SMEs," *Journal of Strategic Marketing*, vol. 32, no. 8, pp. 1071-1105, 2024.

8. G. de Oliveira Collet, F. de Morais Ferreira, D. F. Ceron, M. de Lourdes Calvo Fracasso, and G. C. Santin, "Influence of digital health literacy on online health-related behaviors influenced by internet advertising," *BMC Public Health*, vol. 24, no. 1, p. 1949, 2024. doi: 10.1186/s12889-024-19506-6
9. M. Pittman, A. Oeldorf-Hirsch, and A. Brannan, "Green advertising on social media: Brand authenticity mediates the effect of different appeals on purchase intent and digital engagement," *Journal of Current Issues & Research in Advertising*, vol. 43, no. 1, pp. 106-121, 2022. doi: 10.1080/10641734.2021.1964655
10. B. R. Chadagonda, "Optimizing Marketing Spend with Causal Attribution: Moving Beyond Correlation to Incremental Impact," *Journal Of Engineering And Computer Sciences*, vol. 4, no. 8, pp. 150-157, 2025.