MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: November 25, 2024
https://doi.org/10.31893/multiscience.2025302

Evaluating the impact of marketing mix by using machine learning for digital marketing

Mithun S. Ullal[a] ✉ | Ramona Birau[b] | Virgil Popescu[c]

[a]Department of Commerce, MAHE Manipal DOC, Manipal Academy of Higher Education, Manipal, India.
[b]University of Craiova, "Eugeniu Carada" Doctoral School of Economic Sciences, Craiova, Romania.
[c]University of Craiova, Faculty of Economics and Business Administration, Craiova, Romania.

Abstract This paper evaluates the causal impact of digital marketing on customers' purchase decisions by leveraging advances in machine learning. We analyze buyer behavior under various digital marketing conditions and identify key performance drivers. Our findings emphasize the pivotal role of content in digital marketing. We assess how customized content, multiple offers, and customer decisions influence buying behavior via the AIDA model. Our methodology reveals optimal combinations of content and stimuli across different stages of the AIDA model. Clearance sales significantly enhance buying decisions and conversion rates, particularly in scenarios with high engagement but low conversion. We provide insights into how marketing policies can be optimized for higher sales through targeted content strategies. The study gives an accurate measure of heterogenous treatment effects, helping companies with valuable insights. This helps us understand the importance of segmentations for assessing causal effects and improving sales. The impact of the design elements and marketing interventions is understood by observational and experimental studies. The study used heterogeneous effects of digital ad treatments by using machine learning techniques to identify endogenous treatment responses. Individual ad impact through the AIDA model was studied, measuring the effect of treatment components. The results show that the effect of digital marketing is dependent on customer engagement. Practitioners can gather valuable insights through differential effects based on engagement levels. This will help them refine targeting according to semantic and promotional signals. Contextual signals can be applied to differential promotional strategies addressing existing deficiencies in targeting.

Keywords: machine learning, AIDA model, digital marketing, customers, buying decisions, digital ads

1. Introduction

Marketing involves intricate decisions, particularly in the digital realm, where combinations of words and images are used to target customers. The composite effects of digital promotions are often measured, yet the literature (Gordon et al., 2022) reveals heterogeneous effects with varying outcomes. The source of this heterogeneity remains unclear, primarily due to selection bias and the high-dimensional nature of customer responses. Our study aims to uncover the heterogeneous effects of individual content in digital marketing.

We focus on promotional content headlines and employ causal approximations to evaluate each digital marketing promotion.. Additionally, we examine the impact of semantic choices on buyer behavior, emphasizing the role of content. By leveraging causal machine learning, we seek to understand the heterogeneity in digital marketing treatment components and buyers' diverse responses to marketing campaigns.

Our contributions include demonstrating the financially productive impacts of content choices within the AIDA model framework. We show that machine learning can provide precise causal estimates of the mechanisms driving digital marketing success, using data from over 112 online marketing campaigns and 78,245 customer interactions. This approach helps identify how content influences results and the basis of differential effects.

The effectiveness of these mechanisms varies across the stages of the AIDA model and is influenced by customer engagement levels. We propose increasing profitability by exploiting the heterogeneity in digital marketing campaigns. Our framework predicts heterogeneous treatment effects via the machine learning approach of Chernozhukov et al. (2018), which manages the confounding effects of targeted treatments by predicting propensity scores and outcome equations simultaneously.Following Ellickson et al. (2023), we project orthogonalized signals onto covariates to identify each campaign's demographic and treatment mechanisms. The key is to determine the mechanisms driving heterogeneity. We estimate the unbiased average treatment effects, comparing unique digital marketing campaigns to standard programs. We subsequently calculate heterogeneity in these effects to improve the targeting of remaining advertisements.

Our methodology dissects the composite impact of digital advertisements into distinct mechanisms, identifying heterogeneous responses. We find that digital advertisements vary in their content, necessitating component-level

 Received: September 22, 2024 | Accepted: November 7, 2024

examination. Our approach reveals that combinations of keywords and stimuli produce varying results across the AIDA model stages. Volume discounts have distinct effects on purchase amounts when conditioned on consumer engagement.By studying complex interventions through experimental digital advertisements, we determine the causal effects of their treatment mechanisms. Given the targeting bias inherent in digital ads, we employ statistical modifications to manage this bias.

Personalized targeting aligns with classic unconfounded assignment conditions, and numerous econometric methods exist for identifying treatment effects.Using machine learning, we match policy functions by marketers to allocate treatments, classifying base-level customer responses conditional on observed covariates. The high-dimensional nature of compound treatments, involving promotional, contextual, and semantic choices, justifies our use of machine learning.We study the distinct effects of each mechanism, understanding the results of digital advertising campaigns by combining complex treatments. By orthogonalizing signals, we reduce bias in model selection. Focusing solely on digital ads minimizes challenges arising from intricate delivery systems (Johnson, 2020), reducing activity and delivery biases to zero. We track individual consumer responses to digital ad campaigns across the AIDA model, from clicking on the ad to purchasing.

2. Literature review

Numerous studies have explored the use of machine learning to measure heterogeneous causal effects in randomized control trials. Hitsch and Misra (2018) utilized a k-nearest neighbor approach, projecting treatment impacts onto pretreatment covariates to develop optimal targeting policies in randomized trials. Athey and Wager (2019) examined tails on observational data, whereas Imai and Strauss (2011) studied policy choices in this context.

The causal forest approach, which was selected by Wager and Athey (2018) and furthered by Athey and Imbens (2016), advanced the understanding of heterogeneous treatment effects. Yoganarasimhan et al. (2020) designed personalized trials on a subscription basis, using multiple estimators to measure effectiveness. Semenova and Chernozhukov (2021) provided a more specific output by forecasting linear treatment effect functions.

Ellickson et al. (2023) focused on complex and compound treatments in email marketing. Chernozhukov et al. (2017) developed methodologies for predicting features of heterogeneous effects and inference approaches. Imai and Ratkovic (2013) measured heterogeneous treatment effects via single variable selection and determined the most minor absolute shrinkage.

Bonfrer and Dreze (2009) studied customer responses to digital ads through a bivariate hazards model. Grimmer et al. (2017) examined heterogeneous treatment effects via ensemble methods. Ansari and Mela (2003) reported that customization of digital ads increases sales, whereas Kumar et al. (2014) reported the time required for customers to reach the action stage in digital marketing.

Sahni et al. (2017) used a propensity score approach to analyze digital marketing, revealing increased customer spending when exposed to digital ads. Zhang et al. (2017) identified optimal digital campaign frequency levels. Sahni et al. (2018) conducted randomized field experiments to determine how personal details affect customer buying decisions on digital ads. Our findings on engagement are consistent with those of Ellickson et al. (2023).

An interesting point highlighted by Ascarza (2018) is that customers most likely to switch brands are not the primary targets of loyalty programs. On the other hand, Shetty et al. (2023a) investigated the effect of Artificial Intelligence methods in the case of Indian banking sector. Moreover, Shetty et al. (2023b) examined the implications of digital financial inclusion in the emerging economy of India.

Link to Current Study: Our research builds on these findings by focusing on the heterogeneous effects of digital marketing content, using advanced machine learning techniques to provide precise causal estimates. By examining over 112 online marketing campaigns and 78,245 customer interactions, we aim to identify the optimal combinations of content and stimuli across different stages of the AIDA model. This study contributes to the literature by offering a detailed analysis of how digital marketing strategies can be tailored to enhance customer engagement and sales outcomes.

We contacted more than 124 retailers who use digital marketing in Indian cities. Ninety-two retailers agreed to participate in the study, providing us with 20 digital marketing campaigns over six months. This covered over 78,245 customers. To quantify the results, each customer was exposed to the digital ads at least once. The vector of the company's characteristics is updated every time a customer is exposed to digital ads. We use analytics to track customers' responses to digital ads, updating the variables involved in customer engagement. This helps us design future ads with better content. We classify the treatment components of each digital ad to address our objective of understanding how semantic signals and promotions affect customers' purchase decisions.

2.1. Digital Ads

We carefully note the semantics of each digital ad and record the response from prospects according to the AIDA model. We track whether customers respond to the ad, make a purchase, and the quantity purchased. The elements studied in each ad include the type of product sold, promotional strategy, semantics involved, and number of words. These details are summarized in a table 1.

We include a dummy variable to control for the effect of merchandise categories on customers. We create three merchandise groups: a merchandise group for a specific type of merchandise, a clearance group for merchandise offered as a clearance sale, and a baseline group for merchandise that does not fit into the other two categories.

We also classify promotional strategies into price and nonprice categories, following the methodology of Ellickson et al. (2023). We note discounts and rebates for each ad, treating discount variables as accurate percentages and rebate variables as dummy variables. This research also includes the effect of free products offered with a sale.

By tracking these features, we aim to determine how different combinations of keywords, promotional strategies, and semantic signals influence customer behavior and purchase decisions. Our approach helps us design more effective digital marketing campaigns tailored to various stages of the AIDA model.

Table 1 Heterogenous treatment components for product type.

Heterogenous treatment frequency		Product Type		
		Product	Clearance	Others
Price	Discount	22	10	12
	Rebate	2	2	10
Non-Price	Free	10	1	8
	Shipping free	6	2	6
	Shipping partly free	2	1	4
	Return free	4	2	5
Nonprice semantic choices	Personalized	0	2	6
	Interesting	2	1	3
	Additional	3	1	8
	Limited	3	1	7
	Call	15	0	2
	Sale	2	0	10
	Scratch card	4	1	8

This gives us distinct promotions, including discounts on shipping, entirely free shipping, and returns. These are categorized under nonprice semantic factors, which have been shown to increase the chances of buying (Sahni et al., 2018). One semantic factor includes personal characteristics added to the digital ad. Another involves hidden information that can be accessed only when the customer notices and reads the ad.

We record the words in each ad that may encourage the customer to buy. These words might denote a discount or represent an offer tailored to a customer. Some ads include unique characteristics, which we measure via a dummy variable. Additionally, we include details about sales in the ad and note the number of ads containing a code that can be used to avail offers. We also count the number of characteristics in the ad to capture its overall footprint.

We observe variations in customers' reading rates in response to these treatment components, indicating that some ads are more effective than others. Furthermore, some components perform differently across various stages of the AIDA model. This analysis helps us understand which elements of digital ads are most effective in driving customer engagement and purchases.

We observed that ads with no shipping charges receive a moderate response rate in the attention stage of the AIDA model. However, these ads lead to higher sales than all other mechanisms we examined. Notably, this analysis did not consider the characteristics of the customers exposed to the ads or the specific content within the ads. This approach was chosen to confound the descriptive patterns and promote a robust method for causal inference.

2.2. Consumer Variables

We recorded a list of variables corresponding to each ad, which were used for targeting and analysis to ensure a heterogeneous response from the prospects to various ad contents. These covariates, measured before treatment, include demographics such as age, income, and date of birth. Additionally, we used recency, frequency, and monetary (RFM) metrics, which measure the customer's average money, the number of purchases in the past year, and engagement variables.

The average amount spent by the customer, both online and offline, and the total number of purchases in the last year were included. Engagement was measured by the customer’s response to the digital ad, such as signing up or making additional purchases after viewing the ad, along with the retailer's department. We also tracked the time the customer received the ad, clicked on it, opened it, and made a purchase, using these timestamps to measure the number of days between each action.

The customer variables for each ad are shown in Table 3. These variables differ at the customer level for each ad. The variation in recency, frequency, and monetary variables illustrates the segmentation used by the retailers. Interestingly, there is no variation in age and income, indicating that retailers did not use these factors when designing digital ads.

4. The framework

We develop our framework on the basis of the Neyman–Rubin potential outcome framework and the work of Heckman and Vytlacil (2007). The treatment states' results for each ω are shown as Y(e,ω), where ω ∈ Ω. Treatment T represents digital ads with element e. The results are binary indicators of whether customers click on the ad and if the customer purchased, including the value spent. The results for ω are shown as $Y(e,\omega)_{e\in T}$. Digital ads have a vector of design elements showing the semantics and offers in the ads. Each treatment condition is a (Table 2).

Table 2 The performance of heterogeneous treatment componentes.

Type	Treatment components	Number of customers	Marginal response		Purchase	
Product Type	Product	78245	Ad click rate	Purchase rate	Amount	The average Purchase amount is $
	Clearance	1239	0.231	0.01	0.1	10
Price	Discount	3627	0.184	0.02	0.1	12
	Rebate	2893	0.342	0.04	0.2	11
Nonprice	Free	27381	0.423	0.03	0.3	16
	Shipping free	2883	0.283	0.01	0.2	12
	Shipping partly free	4038	0.382	0.02	0.2	12
	Return free	2938	0.324	0.02	0.1	21
Noninformative semantic choice	Personalized	9201	0.229	0.8	0.5	18
	Interesting	2910	0.282	0.02	0.2	22
	Additional	3424	0.192	0.03	0.2	24
	Limited	2848	0.128	0.02	0.1	16
	Call	4839	0.116	0.05	0.1	12
	Sale	5828	0.176	0.02	0.2	16
	Scratch card	1827	0.272	0.01	0.1	17

Table 3 Pretreatment covariates of all ads.

Customer	Pretreatment covariates	Summary statistics	
	Variable	Mean	Std deviation
Recency	Read	78.32	132.91
	Click	276.45	662.12
	Purchase	587.56	819.78
Frequency	No. of items	5.86	6.22
	Different product purchased	5.23	5.22
Monetary	Offline	12.62	41.21
	Online	61.89	51.24
Habitual	Catalog	0.22	0.32
	Customized	0.37	0.46
	Offline purchase	0.52	0.58
	Online purchase	0.94	0.88
Demographic	Date of birth	7.22	4.54
	Age	4.22	1.82
	Income	5.68	2.46

sum of all mechanism states in the Heckman and Vytlacil representation, shown as $e = (e_1, e_2, \dots e_c)$ for mechanism m.

Here, the unbundling of mechanisms of complex treatment is performed only in low-dimensional settings. We find the effects of individual contributions of each mechanism and consider heterogeneous responses to overall treatment or its included mechanisms. The personal treatment effect for ω that is associated with the results of treatment e is given by Y(e',ω) - Y(e,ω), e≠e', which includes two elements e, e∈T. This is used to construct all unique pairs e, e∈T, which is made of an unobservable primitive measure of causal effect. We use an assignment mechanism α: Ω → T to assign these treatments to individuals, allocating treatment to ω.

We consolidate all the mechanisms for ω. We consider the result of the digital ad and assume D(e,ω)= 1 when ω in e is in targeting policy p. The observed outcome Y(ω) can then be defined as:

$$Y(\omega) = \Sigma_{e\in T} D(e,\omega)\ Y(e,\omega)_r$$

This is a fundamental problem in causal inference where D(e,ω)=1 makes Y(e,ω) measurable, but Y(e',ω) for e≠e' cannot be measured (Holland 1986).

In our research, we send different ads to different prospects, where the individual ads can be identified under certain conditions. The solution to unknown individual treatment effects is to impose restrictions on the assignment method, focusing on more aggregate estimates and conditional average treatment effects (CATEs). Comparing treatment t to treatment $t_1$yields:

$$ATE(t,t_1) = E(Y(t,\omega) - Y(t_1, \omega)) = \equiv \tau tt_1.$$

When we condition on pretreatment covariates X related to components ω, this is defined as:

$$CATE(t,t_1) = E(Y(t,\omega) - Y(t_1, \omega)) \mid X = x) \equiv \tau_{tt1}(x)$$

The treatment component (s_1, $s_{2\ 2 \ldots} s_c$) is our next application, where the divergence is among a set of treatments that vary by addition and rejection of a given mechanism. The marginal average treatment effect is the pairwise contrast between addition and rejection (Grimmer et al., 2017). This combines the remaining treatment mechanisms and pretreatment covariates. The marginal impact of demographic factors is also characterized by a similar method. Econometric identification of conditional and average treatment effects on the assignment mechanism D(e,ω) is known. We chose the nonrandom assignment, which was loosely correlated with the unconfounded assignment. This mechanism is randomly assigned without the knowledge of assignment regulation. The actual result does not affect the assignment to treatment but may depend on ω for X covariates. Unconfounded assignment is a natural assumption that happens automatically because of the algorithm's use. The natural estimates here are conditional average treatment effects or marginal conditional average treatment effects, which account for response heterogeneity among the targeted segments. The unconfounded assignment can be stated in the identification problem as follows:

$$Y(e,\omega) \perp\perp D \mid F \text{ (Unconfounded assignment)}$$

Conditional or average treatment effects are identified under positive conditions without interference across units (Rubin 1980).

This is explained as follows: $0<Pr(D = 1| F)<1$, F

is a set of observable features of ω required to satisfy the unconfounded condition. Positivity for all individuals in the target population with the probability of being assigned to every treatment requires sufficient overlap, also referred to as common support or positivity in different studies. This ensures that the probability of being assigned to each treatment is greater than 0 for all individuals in the target population. The lack of sufficient overlap of demographics among groups assigned different treatments indicates a lack of positivity. Positivity cannot target marketing settings if the company's directions are deterministic. Companies must include a degree of residual randomization in every campaign.

During the study, prospects received multiple digital advertisements, allowing us to assess the potential carryover effects of each treatment. To ensure that the impact of a digital ad is independent of prior ads and temporal factors, we rigorously tested for carryover effects by [specific method used]. One hypothesis is that the high frequency of ads and the diversity of products might influence the effectiveness of individual ads. This hypothesis will be tested by [brief explanation of the testing method].

We controlled for recency, frequency, and monetary variables in our analysis via [specific models or methods]. This control is essential for accurately measuring treatment effects and ensuring that our estimates reflect the impact of the ads rather than confounding factors.

Our methodology incorporates machine learning techniques to address potential misspecifications in the results or propensity models. By employingspecific machine learning methods, we improve the robustness of our treatment effect estimates and increase the model's accuracy. Using these techniques helps correct potential biases and inaccuracies in the estimation process, providing more reliable insights into the effectiveness of digital marketing strategies.

5. Analysis: Machine learning estimation of average treatment effects

We estimate the digital advertisements' average treatment effects (ATE) and conditional average treatment effects (CATE). Our counterfactual off-policy evaluation approach informs pairwise estimates, which improves targeting precision. We obtain orthogonalized scores by fitting individual-level outcomes and use these scores as inputs in our analysis. Our machinelearning approach is based on the methodology developed by Robins and Rotnitzky (1995) and extended by Ellickson et al. (2023).

Our framework selects a digital ad as the baseline treatment and compares all other ads against this reference. We use an indicator function I to denote whether an observation was assigned to treatment (1) or control (0). The propensity score, which represents the probability of treatment assignment, is then measured as follows:

$$e(x) = E(I_i \mid Xi = x)$$

The expected result is then

$$\mu(i, x) = E(Y_i \mid Ii = i, Xi = x)$$

τ is a function of the joint distribution of (Ii,Xi,Yi) in multiple ways. The possibilities could be

$$\tau = E(\mu (1, X_i) - \mu (0, X_i))$$

$$\tau \ = E(\underline{Y_i I_i} - \underline{\ Y_{i(}1-I_i) \quad}$$

$$e(X_i) \quad 1 - e(X_i)$$

$$\tau = E \; \frac{(Y_i - \mu(1, X_i)) I_i}{e(X_i)} - \frac{(Y_i - \mu(0, X_i)) I_i \quad (1-I_i)}{1 - e(X_i)}$$

Eemploy a regression adjustment strategy to estimate treatment effects, where conditional results are calculated for the treatment and control conditions. This method uses a direct comparison of fitted values from outcome models, which approximates the population expectation. Additionally, we apply a propensity weighting approach, where the sample analog is computed via the fitted values from the propensity score model.

We combine these two methods to increase accuracy by leveraging their respective influence structures. However, a single machine learning approach can lead to regularization and overfitting, introducing bias. To address this issue, we use orthogonalization, which helps mitigate bias by adjusting for confounding factors.

In our third representation, the orthogonality score is calculated as follows:

$$\Psi(y,w,x) = \mu(1, X) - \mu(0, X) + \frac{w}{e_t(x)} (y - \mu(1, X)) + \frac{w}{e_c(x)} (y - \mu(0, X))$$

Here, $\Psi_i = \psi(Y_i, W_i, X_i)$, where $e_t(x)$ is a propensity for treatment control. Flexible approximations of $\mu(w,x)$ and $e_t(x)$ are based on machine learning approaches and then combined with the IF to calculate $\Psi_i = \psi(Y_i, W_i, X_i)$ for every observation.

To estimate average treatment effects, we calculate the average of Ψi, incorporating orthogonalization and cross-fitting to manage potential bias. Orthogonalization helps to adjust for confounding factors, whereas cross-fitting is used to prevent overfitting by dividing the data into training and validation sets.

For predicting the propensity and outcome models, we use the random forest approach, as outlined by Ellickson et al. (2023). This method is effective because of its automatic variable selection capabilities and its robustness in out-of-sample estimation, which helps mitigate overfitting. We utilize a comprehensive set of targeting variables from the digital ads to build our models.

Our methodology ensures accurate weight estimation because conditioning variables can be low-dimensional without detailed mapping knowledge. We predict binary propensity scores for each digital ad, focusing on the central population to account for variability in treatment effects.

Additionally, we follow the methodology of Hernán and Robins (2020) to construct stabilized propensity weights. This approach is chosen for its computational efficiency and ability to provide robust estimates of treatment effects by controlling for potential confounding factors.

4.2. Treatment components

To obtain unbiased estimates of the effects of different digital ads, we develop orthogonalized scores for treatment components, incorporating a set of pretreatment covariates. Orthogonalization helps to adjust for confounding factors, ensuring that these covariates do not skew our estimates of ad performance.

In our analysis, we define "treatment covariates" (TCs) as variables related to the characteristics of the digital ads and "low-dimensional subset" (LDS) as a reduced set of these most influential variables. The function g(S) represents the conditional expectation of the outcome given the treatment and covariates, summarizing the causal effects and the response surface for each treatment component.

The "target surface" refers to the treatment components and demographic variables we aim to analyze. The conditional expectation function g(S) helps us understand how treatments impact outcomes while accounting for demographic factors. This approach allows us to pinpoint why certain digital ads perform better than others do and to refine targeting strategies on the basis of these insights.

$$g(S) = E(Y(\eta_0) \mid S = s)$$

$Y\eta_0$ is the orthogonal score, which depends on the η_0 TC. It is now projected on the low-dimensional level via least squares.

$$\beta = \operatorname{argmin}_{b \in R^d} E(Y(\eta) - p(S')b)^2$$

The prespecified basis functions $p(S')$ with approximations can be defined to ease interpretability. Pairwise comparisons are projected onto demographic covariates, creating a group average treatment effect. It is a prediction of the $\tau tt_1(x)$ parameter defined with comparisons among pairs of treatments on the basis of pretreatment covariates. We develop complete scores for an entire vector of treatment components to target estimates on marginal average treatment effects via linear projection, similar to Ellickson et al., 2023.

5. Analysis

Unconfounded assignment and sufficient overlap are crucial assumptions for identifying pairwise average treatment effects in causal inference. Unconfounded assignment implies that the treatment assignment is independent of potential outcomes given the covariates. Moreover, sufficient overlap ensures a range of covariate values across treatment groups to allow for meaningful comparisons.

The "overlay condition," which refers to ensuring that the treatment effect can be estimated reliably across different levels of covariates, can be challenging to meet with stringent targeting. This is because precise targeting might limit the range of covariate values represented in the treatment groups, affecting the estimation of treatment effects.

To address this, we remove binary assignment rules and assess the deterministic assignment for all digital ads. By applying relatively less smoothing to the data and focusing on advertisements with randomization, we aim to eliminate observations where the actual assignment mechanism was nondeterministic. This approach helps maintain the integrity of the assignment process but may not ensure a complete covariate balance.

We test balance by computing the standard bias of each covariate population. This involves measuring the difference in covariate distributions between treatment groups to ensure that imbalances in covariates do not confound any observed treatment effects. Ensuring covariate balance is essential for accurately estimating treatment effects and robust causal inference.

This is shown as

$PSB_{ak} = |X_{ka} − X_{kp}|/σkp$

Here, Xkt is the propensity score of covariate k for digital ad a, and $e_a(X_i)$ is the propensity score for ads.

$T_i(t)$ indicates t, $X_{ka,}$ and $σ_{kp}$ are the unweighted mean and standard deviation for covariate k.

We assessed the degree of balance in our propensity score matching (PSM) analysis by categorizing propensity score bias (PSB) scores into high- and low-average groups. Our study revealed that the balance between treatment groups was acceptable, with no significant imbalances observed in the segment of large differences. Specifically, we did not identify any instances where the PSB scores indicated substantial discrepancies between treatment groups.

The PSB scores across covariates for each digital ad are detailed in the tables below, which show the distribution of these scores and the maximum values observed for each pretreatment covariate. These results help us ensure that any observed treatment effects are not confounded by imbalances in covariates, confirming the robustness of our balance assessment.

Some ads exhibited moderate effectiveness scores, with only one ad achieving a high score. Personalized content did not significantly improve inference, possibly due to residual variation observed when customers interacted with personalized ads. This variation may stem from the lack of residual randomization and the unrestricted nature of the ads. Despite these challenges, we maintained exchangeability among consumer types for the targeted component-level effects. We assessed the degree of balance in the targeting variables to understand how well component-level intersections capture the heterogeneity among consumers exposed to different ad types.

Equation (1) calculates the PSB value for every covariate t_1 and treatment mechanism a.

$X_{t1} = \underline{\Sigma_{i=1}{}^{n}S_i(a)\ X_{t1i}/e_a(X_i)}$

$\Sigma_{i=1}{}^{n}S_i(a)/e_a(X_i)$

is the propensity-weighted mean of covariate t_1. Ea (X_i) is the propensity score estimated.

$S_i(a)$ is the indicator of assignment a.

The pretreatment factor weighted means are summarized in the table below and categorized by treatment mechanism. (*) * indicates the extent to which balance was affected. The results show that only a small fraction of cases present a mean above the low level, indicating strong balance and overlap across treatments. Our analysis of carryover effects reveals no evidence that prior ads impact the current results. Specifically, focal results for baseline ads demonstrate no significant influence of previous ads on current outcomes. This confirms that our framework meets the requirements for sequential exchangeability, effectively predicting across a full set of recency, frequency, monetary, and other engagement variables in the dataset (Table 4).

6. Empirical results and findings

We measure pairwise average treatment effects and project orthogonalized scores onto treatment mechanisms and pretreatment covariates. Our model effectively provides valid and precise average treatment effects by leveraging individual treatment effect signals to assess the mechanism-specific impact of ads on consumer behavior through orthogonal projection. By comparing two ads, we highlight the importance of targeting and demonstrate the heterogeneity in consumer responses to different ad combinations. The recovered pairwise average treatment effects revealed significant variability, emphasizing the need for accurate targeting strategies. We also evaluate how the characteristics of the ads affect their performance at various stages of the AIDA model. This assessment provides insights into optimizing targeting policies, illustrating how companies can increase profits by refining their ad strategies (Table 5).

Table 4 Pretreatment covariate weighted means of ad.

Ad	Pretreatment covariates													
	Offline purchase average	Online purchase average	DOB	Age	Income	No. of the different products purchased	Catalog	Choice	Purchase offline	Purchase online	Total purchase	Read	Click	Purchase
1	15.42	52.56	8.12	4.22	5.42	4.26	0.48	0.35	0.56	1.04	7.22	72.92.	312.89	422.82
2	14.26	57.82	8.45	4.34	5.43	4.64	0.12	0.36	0.51	1.05	7.34	63.28	328.68	438.28
3	16.87	61.62	8.32	4.23	5.65	4.32	0.32	0.28	0.56	1.02	7.21	87.92	367.28	438.29
4	14.52	57.78	8.12	4.67	5.78	4.38	0.54	0.38	0.52	1.07	7.56	71.28	364.28	487.59
5	16.38	56,92	8.25	4.45	5.21	4.84	0.67	0.25	0.58	1.03	7.76	72.87	328.28	486.26
6	15.26	62.28	8.67	4.65	5.64	4.37	0.45	0.36	0.57	1.08	7.45	67.89	364.86	497.92
7	14.29	52.78	8.65	4.34	5.57	4.46	0.35	0.46	0.56	1.05	7.79	65.78	332.29	428.65
8	16.65	61.98	8.76	4.67	5.37	5.62	0.34	0.25	0.59	1.08	7.23	74.88	355.89	429.86
9	15.38	56.87	8.86	4.23	5.78	4.87	0.35	0.48	0.52	1.02	7.39	69.92	376.87	498.12
10	14.38	54.29	8.67	4.64	5.45	4.59	0.53	0.26	0.56	1.07	7.48	82.78	389.82	484.28
11	14.65	62.98	8.34	4.53	5.78	5.35	0.23	0.48	0.47	1.03	7.36	81.67	392.12	419.28
12	15.39	59.92	8.56	4.36	5.35	4.22	0.43	0.27	0.59	1.08	7.38	76.46	328.36	469.23
13	16.86	58.65	8.34	4.23	5.78	4.87	0.23	0.26	0.54	1.04	7.36	72.68	345.27	465.28
14	14.49	59.91	8.67	4.45	5.57	5.68	0.23	0.17	0.58	1.09	7.73	80.12	376.87	438.28
15	16.46	51.62	8.24	4.37	5.43	4.38	0.43	0.37	0.52	1.03	7.73	75.26	364.87	462.29
16	14.52	57.78	8.67	4.78	5.59	4.79	0.23	0.28	0.54	1.05	7.74	78.28	329.28	478.32
17	15.89	56.85	8.87	4.35	5.28	5.82	0.12	0.37	0.57	1.06	7.83	74.27	354.28	488.28
18	16.72	59.92	8.45	4.62	5.38	4.49	0.43	0.27	0.58	1.07	7.63	75.29	344.23	446.48
19	14.68	60.12	8.36	4.72	5.74	4.18	0.12	0.49	0.52	1.03	7.73	71.56	328.38	498.24
20	16.98	61.12	8.34	4.53	5.37	4.38	0.34	0.37	0.51	1.08	7.82	81.26	334.37	487.65

Table 5 Propensity Score Bais of covariates.

Ad	PSB score
1	0.12
2	0.22
3	0.23
4	0.14
5	0.12
6	0.06
7	0.12
8	0.45
9	0.23
10	0.43
11	0.23
12	0.37
13	0.27
14	0.35
15	0.27
16	0.28
17	0.38
18	0.48
19	0.27
20	0.25

6.1. Motivation

Our machine learning approach provides practical and accurate average treatment effects by enabling precise targeting. To illustrate its application, we compare two ads offering a 25% discount on products but targeting different consumer segments. This comparison helps isolate the unbiased effect of the discount, as it is the only variable differing between the ads. The table below presents a detailed comparison between the two ads, including the selection of pretreatment targeting variables. This approach allows us to assess the impact of the discount and targeting strategies on the effectiveness of the ads.

Our study of ad conversions reveals that ads offering a 50% discount generally perform better than those offering a 25% discount, particularly across various stages of the AIDA model. We tested pretreatment covariates and found that the 25% discount ads were primarily directed at already engaged customers. It is crucial to interpret ad results carefully, as relying solely on conditional means could misrepresent how ads influence conversion. We applied our model to pairs of ads via two random forest algorithms to assess unbiased pairwise average treatment effects. These algorithms estimate the treatment propensity and calculate score functions. We then computed the difference in average treatment effects between the 25% and 50% discount ads at different AIDA model stages. The analysis revealed that the 50% discount ad consistently outperformed the 25% discount ad across all stages, with robust tstatistics supporting the validity of these results. This underscores the effectiveness of our methodology in providing accurate and meaningful insights into ad performance (Table 6).

Table 6 Maximum PSB scores of ads for pretreatment covariates.

Customer	Pretreatment covariates	Maximum PSB scores
Recency	Read	0.22
	Click	0.12
	Purchase	0.38
Frequency	No. of items	0.48
	Different product purchased	0.18
Monetary	Offline	0.36
	Online	0.54
Habitual	Catalog	0.19
	Customized	0.25
	Offline purchase	0.23
	Online purchase	0.42
Demographic	Date of birth	0.25
	Age	0.22
	Income	0.32

6.2. All ad's average treatment effects

We assess ads' pairwise average treatment effects by comparing each ad to a fixed baseline ad across the entire population. Using a pairwise average treatment effect approach, our framework identifies the best-performing ads by evaluating them against this baseline. The results are illustrated in Figure 1, which shows that ads with positive responses outperform the control ads, which is consistent with the findings of Ellickson et al. (2023), although their study was limited to email ads. Orthogonal forecasts help identify features of ads that lead to differential performance outcomes. We observed heterogeneity in the initial stages of the AIDA model due to various influencing factors—figures 2 and 3 present findings related to purchase incidence and values. Our analysis reveals that the baseline ads generally perform better, with only eight ads showing higher performance. The effects observed are notably large, suggesting significant differences compared with the conversion outcomes typically seen in today's large-scale ads.

The lifts in treatment effects for various ads are significantly different from those of the baseline ad, with variations observed from the open stage to the purchase amount. Our analysis shows that ads differ in their effectiveness in generating clicks and actual purchases. Specifically, while some ads exhibit positive average treatment effects compared with the baseline, others show adverse effects. Additionally, certain ads successfully increase engagement but do not translate this engagement into sales. For example, some ads may drive higher click-through rates but fail to convert clicks into purchases, highlighting the need for a nuanced understanding of ad performance beyond mere engagement metrics. This underscores the importance of evaluating engagement and sales outcomes when assessing ad effectiveness.

The analysis of average treatment effects by ordering allows us to measure the impact of specific ad features on performance. We gain insights into which features drive ad effectiveness by examining these effects. Our findings indicate that ads in the clearance category perform better than others do, and personalized ads show increased engagement. Combining certain ad features can enhance engagement, resulting in greater average treatment effects.

6.3. Orthogonal score predictions: Causal heterogeneity

Orthogonal score predictions reveal that the variability in ad effects is due mainly to differences in content and framing. To address this heterogeneity, we calculate the marginal effects of a complete set of ad features, which includes all compound treatments. This approach helps highlight the limitations of not accounting for heterogeneity and improves our understanding of how various ad features contribute to performance.

Using orthogonalized scores allows us to measure the effects of different ad features on engagement and purchasing behavior across various stages of the AIDA model, independent of earlier stages. Table 10 presents the individual promotion mechanisms and their marginal effects. Ads offering freebies generally increase engagement, as supported by Chatterjee and McGinnis (2010). However, adding free shipping significantly increases engagement throughout all stages of the AIDA model. The results indicate that distinguishing between treatment types is crucial, as some ads perform well in isolation but perform poorly when combined with other mechanisms. For instance, gifts lead to the highest level of engagement, whereas sales drive more significant purchases. Free shipping impacts purchase amounts across all nonprice promotions, affecting AIDA's opening and buying stages (Table 7) (Table 8).

Table 7 Heterogenous treatment components weighted mean pretreatment covariates.

Type	Treatment components	Offline purchase average	Online purchase average	DOB	Age	Income	No. of the different products purchased	Catalog	Choice	Purchase offline	Purchase online	Total purchase	Read	Click	Purchase
		Pretreatment covariates													
Product Type	Product	12.81	53.92	8.12	4.24	5.42	4.34	0.14	0.43	0.62	1.12	7.25	81.29	322.91	482.91
	Clearance	12.82	54.82	8.12	4.22	5.38	4.48	0.15	0.44	0.64	1.14	7.42	73.12	341.12	491.29
Price	Discount	12.36	52.38	8.22	4.24	5.72	4.82	0.12	0.48	0.62	1.15	7.46	88.12	368.81	449.82
	Rebate	12.65	51.28	8.36	4.45	5.82	4.81	0.14	0.42	0.61	1.09	7.48	77.72	372.91	481.42
Nonprice	Free	12.49	58.28	8.11	4.23	5.46	4.91	0.16	0.41	0.65	1.08	7.46	71.97	341.51	491.92
	Shipping free	12.82	52.28	8.14	4.61	5.81	4.89	0.18	0.49	0.64	1.12	7.78	72.12	371.92	499.91
	Shipping partly free	12.73	57.38	8.28	4.32	5.62	4.92	0.12	0.51	0.62	1.17	7.81	71.81	344.24	439.61
	Return free	12.91	53.34	8.38	4.81	5.42	5.23	0.18	0.49	0.61	1.19	7.97	77.81	361.82	491.81
Noninformative semantic choice	Customized	12.76	58.74	8.78	4.78	5.82	4.65	0.16	0.41	0.62	1.12	7.36	71.82	387.91	499.22
	Interesting	12.54	51.19	8.29	4.46	5.58	4.87	0.28	0.39	0.63	1.28	7.51	87.81	391.92	481.21
	Additional	12.49	54.28	8.24	4.48	5.82	5.62	0.29	0.41	0.61	1.12	7.48	87.72	398.24	428.21
	Limited	12.38	56.29	8.47	4.29	5.42	4.72	0.12	0.49	0.68	1.16	7.82	81.12	348.49	471.24
	Call	12.78	55.28	8.92	4.78	5.82	4.91	0.11	0.49	0.62	1.16	7.61	77.81	361.82	471.29
	Sale	11.91	54.28	8.27	4.82	5.75	5.48	0.14	0.42	0.61	1.18	7.89	85.22	381.54	481.92
	Scratch card	12.84	58.26	8.48	4.97	5.64	4.92	0.18	0.48	0.67	1.14	7.81	79.81	372.91	471.92
Population parameters	Mean	12.65	59.29	8.38	4.78	5.62	4.91	0.18	0.47	0.64	1.15	7.92	79.81	342.81	481.31
	Standard deviation	12.72	54.12	4.12	1.81	2.48	6.26	0.49	0.57	0.68	1.17	7.82	79.71	522	798

Table 8 Consumer response and pretreatment covariates of discounts before endogenous targeting.

Customer	Pretreatment covariates	50% discount		25% discount	
	Variable	Mean	Std deviation	Mean	Std deviation
Response	Click rate	0.0813	0.4192	0.212	0.4125
	Purchase	0.0008	0.0329	0.0032	0.0562
	Amount	0.0506	3.2298	0.2267	4.6782
Recency	Read	302.19	381.12	52.62	71.28
	Click	584.12	812.24	81.12	85.46
	Purchase	1018.78	935.24	202.16	142.38
Frequency	No. of items	1.02	2.82	6.22	5.12
	Different product purchased	2.05	3.12	7.64	5.43
Monetary	Offline	7.28	26.45	14.59	35.29
	Online	22.34	37.92	54.58	44.58
Habitual	Catalog	0.09	0.31	0.18	0.38
	Customized	0.21	0.52	0.32	0.52
	Offline purchase	0.42	0.51	0.61	0.72
	Online purchase	81.29	51.67	1.02	0.02
Demographic	Date of birth	7.42	5.12	7.38	4.24
	Age	4.24	1.24	4.28	2.19
	Income	5.28	2.39	5.24	2.12

This study extends the work of Ellickson et al. (2023) by identifying the mechanisms driving sales and addressing causal heterogeneity. Like Sahni et al. (2018), personalized ads generally enhance engagement, but contests have a limited effect on clicks and purchases, although they increase purchase amounts. Ads detailing features improve click rates but do not significantly affect sales, highlighting the reversals in average treatment effects. Clearance promotions perform better than product promotions do, negatively impacting engagement in the awareness stage but showing positive effects in later stages (Table 9).

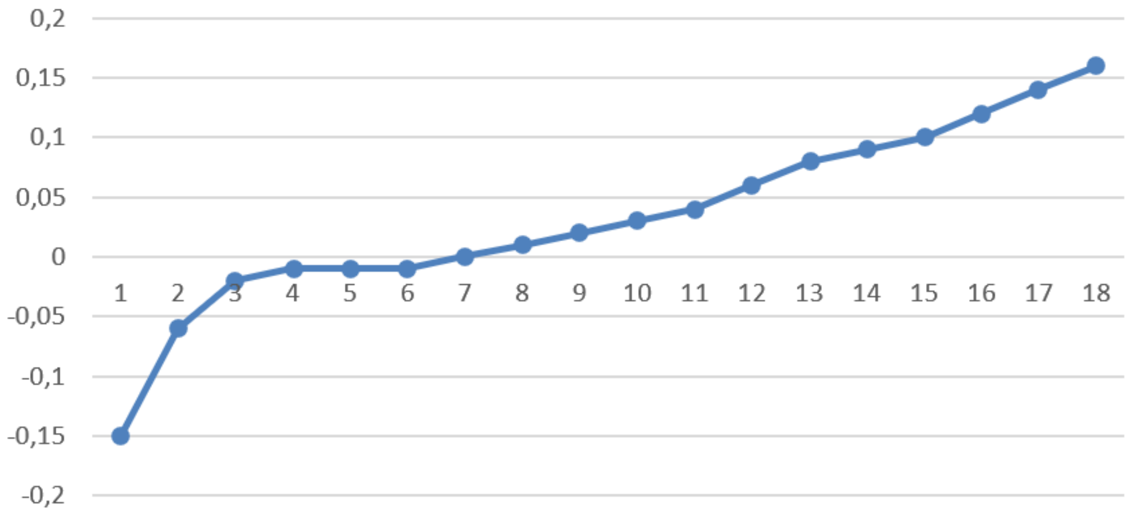

Figure 1 Average treatment effects compared with those at baseline and in the awareness stage.

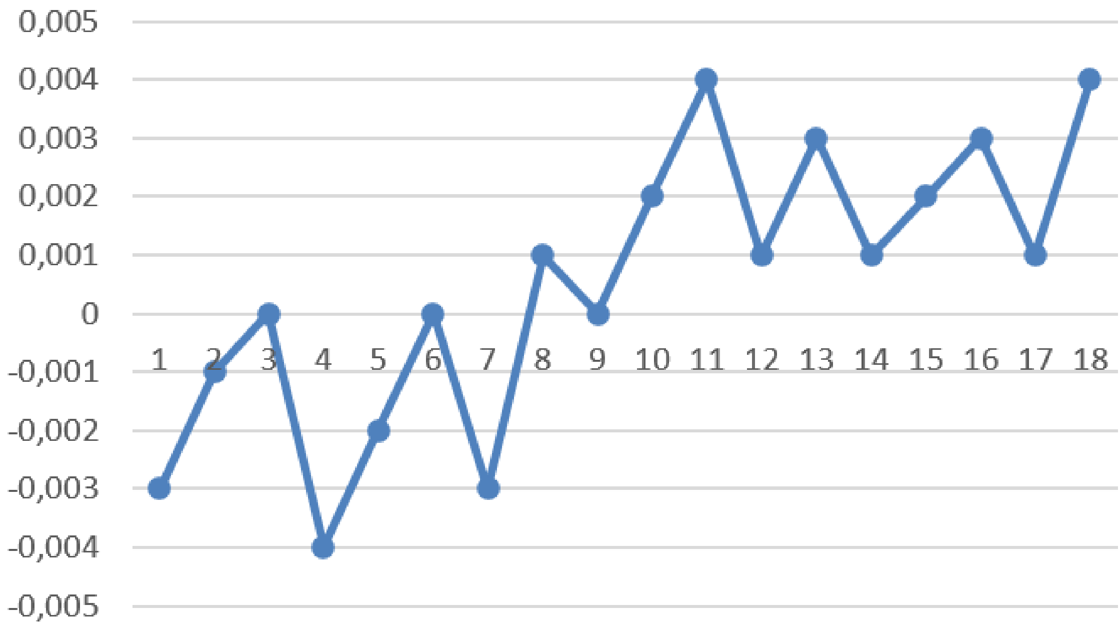

Figure 2 Average treatment effects compared with those at baseline and in the interest and desire stages.

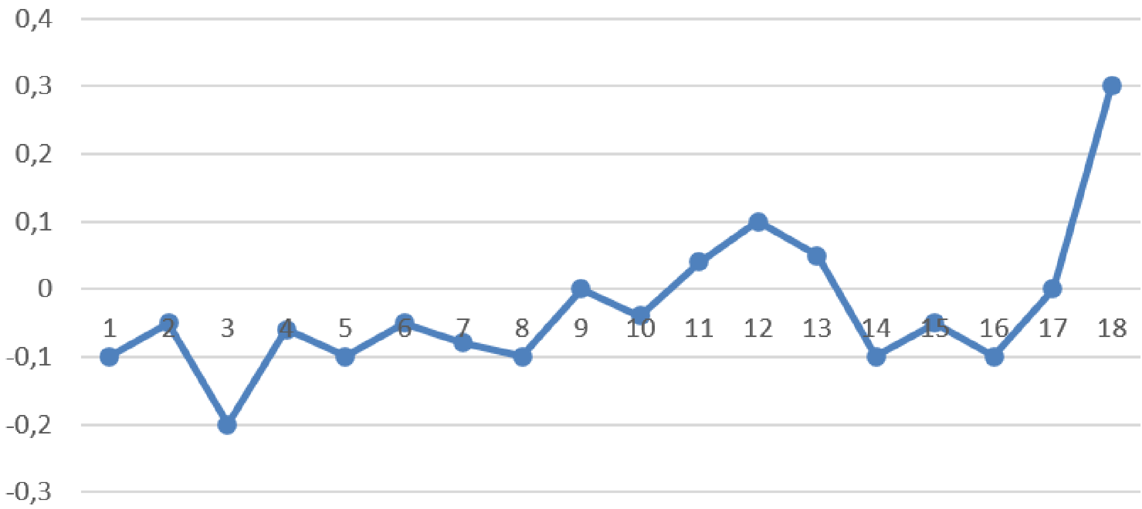

Figure 3 Average treatment effects compared with those at baseline and in the action stage treatment effects compared with those at baseline and in the action stage.

Table 9 Average treatment effects based on the AIDA model for ads with one common treatment componente.

Ad	Product	Clearance	Discount	Free shipping	Customized	Additional	Call	DOB	Code	Click	Purchase	Amount
5	2	0	0.5	0	1	2	2	0	0	-0.0324	-0.0014	-0.0752
10	0	2	0.6	0	0	2	2	1	2	0.0532	0.0032	0.0848
15	0	2	0.5	0	0	2	2	0	2	0.0436	0.0042	0.0386
20	2	0	0.6	1	0	1	0	0	1	-0.0326	-0.0009	-0.0021

Table 10 The effect of heterogeneous treatment features open rates, purchase rates, and amounts.

Type	Variables	Click		Purchase		Amount	
	Intercept	-0.0522	0.0021	-0.0021	0.0004	-0.0287	0.0276
	Characters	-0.0009	0.0002	0.000	0.000	-0.0022	0.0005
Product Type	Product	-0.0112	0.0009	0.0004	0.0002	-0.0028	0.01
	Clearance	0.0692	0.0018	0.0051	0.0005	0.1881	0.0132
Price	Discount	-0.0524	0.0036	0.0028	0.0007	0.0991	0.0381
	Discount Product	0.0392	0.0032	-0.0042	0.0006	-0.1129	0.0423
	Discount Clearance	0.4822	0.0086	0.0181	0.0014	0.3212	0.0923
	Rebate	0.0398	0.0021	0.0032	0.0003	0.1811	0.0291
Nonprice	Free	0.0492	0.0022	0.0037	0.0002	0.0622	0.0218
	Shipping partly free	-0.0342	0.0021	-0.00011	0.0003	-0.1328	0.0262
	Shipping free	0.0312	0.0018	0.0006	0.0004	0.0812	0.0191
	Return free	-0.0191	0.0018	-0.0008	0.0004	-0.0492	0.0196
Noninformative semantic choice	Customized	0.0991	0.0012	0.0051	0.0003	0.1342	0.0128
	Interesting	0.0162	0.0020	0.004	0.0005	0.1422	0.0268
	Additional	-0.0010	0.0015	0.0014	0.0003	0.0128	0.0162
	Limited	0.0062	0.0008	-0.0030	0.0002	-0.052	0.0082
	Call	0.0132	0.0007	-0.0010	0.0002	0.0912	0.0132
	Sale	-0.0224	0.0012	0.0016	0.0001	0.0914	0.00136
	Scratch card	0.0172	0.0009	0.0012	0.0002	0.0312	0.0098
	Observation	10292		10292		10292	
	R^2	0.012		0.010		0.006	

The nuanced influence of discounts reveals that while they generally reduce engagement in the awareness stage, they have a positive effect when associated with clearance items. In contrast, discounts linked to product categories negatively affect purchase amounts, although they positively influence engagement across all AIDA stages. Rebates boost engagement according to the AIDA model. The disaggregated effects suggest that incorrect grouping of components can lead to misleading results, although aggregate effects align with existing knowledge. Our framework can aid researchers and marketers in evaluating the impact of complex advertising interventions.

6.4. Orthogonal score

In this analysis, we measure heterogeneity in treatment effects by focusing on variations in consumer responses. We utilize a low-dimensional approach to segment customers on the basis of past buying behavior and pretreatment covariates, identifying optimal customer groups. A summary of these findings is presented in Table 11. Our analysis, following the approach outlined by Gopalakrishnan and Park (2021), reveals that one group, termed the "engaged group," demonstrates greater engagement than the "disengaged group." This differentiation in engagement levels is crucial for effectively tailoring and targeting market promotions.

To recover heterogeneous treatment effects, we employ the methodology developed by Ascarza (2018), which assesses how differences in engagement influence the performance of digital ads. By grouping individuals on the basis of their engagement scores concerning the treatment components of each ad, we gain insights into how varying levels of engagement impact promotional effectiveness. This approach allows us to better understand and target consumer segments, optimizing the impact of digital advertising strategies.

The results of the promotional offers and semantic choices are detailed in Tables 12 and 13, with responses analyzed on the basis of customer engagement levels. The analysis reveals that the impact of promotional offers is more pronounced in the high-engagement (HE) group than in the low-engagement (LE) group. Specifically, in the HE group, the effects are notably positive across the AIDA model, especially for discount clearance offers. Conversely, in the LE group, the positive impact is primarily observed at the Awareness stage of the AIDA model.

Table 11 Pretreatment covariates for heterogeneous customers.

		Highly involved customer		Low involved customer	
Customer	Pretreatment covariates	Mean	SD	Mean	SD
Recency	Click rate	41.28	72.82	115.58	172.78
	Purchase	144.69	302.87	481.18	752.24
	Amount	186.23	188.68	902.17	915.36
Frequency	Read	6.68	7.02	1.02	1.19
	Click	8.03	2.12	2.03	2.19
Monetary	Purchase	22.31	41.18	6.12	18.39
	No. of items	72.35	51.22	18.38	14.26
Habitual	Different product purchased	0.18	0.41	0.14	0.34
	Offline	0.61	0.62	0.45	0.54
	Online	1.14	0.24	0.88	0.51
	Catalog	0.18	0.41	0.21	0.42
Demographic	Date of birth	7.22	4.65	7.18	5.07
	Age	4.38	2.64	4.68	2.61
	Income	5.67	2.45	4.91	2.64

When comparing the policy analysis effects to those of the HE groups, it is evident that the effects are less pronounced, suggesting that the level of engagement significantly influences the strength of promotional responses. Additionally, the contrast groups provide insights into overall engagement across multiple components, demonstrating variability in the effects of different treatment components.

Table 12 Heterogeneous treatment components for high-engagement customers.

Type	Variables	Click	Std Error	Purchase	Std Error	Amount	Std Error
	Intercept	-0.0721	0.0036	-0.0052	0.0006	-0.3652	0.0421
	Characters	-0.0021	0.0002	0.0001	0.0001	-0.0041	0.0008
Product Type	Product	-0.0426	0.0014	0.0002	0.0003	-0.0156	0.0181
	Clearance	0.1298	0.0029	0.0099	0.0006	0.3398	0.0291
Price	Discount	-0.0761	0.0048	0.0029	0.0009	0.1265	0.0532
	Discount Product	0.0328	0.0056	-0.0059	0.0008	-0.1227	0.0658
	Discount Clearance	0.7864	0.0142	0.0491	0.0028	0.5768	0.1886
	Rebate	0.0482	0.0033	0.0041	0.0006	0.2289	0.0452
Nonprice	Free	0.0733	0.0029	0.0036	0.0004	0.0985	0.0312
	Shipping partly free	0.0041	0.0034	-0.0008	0.0006	-0.1449	0.0492
	Shipping free	0.0519	0.0021	0.0009	0.0005	0.1449	0.0381
	Return free	-0.0132	0.0024	-0.00012	0.0005	-0.0991	0.0291
Noninformative semantic choice	Customized	0.1245	0.0019	0.0072	0.0005	0.3287	0.0288
	Interesting	0.0188	0.0031	0.0039	0.0008	0.2214	0.0428
	Additional	-0.0081	0.0024	0.0018	0.0005	0.0512	0.0292
	Limited	-0.0118	0.0018	-0.0019	0.0002	-0.0292	0.0192
	Call	0.0154	0.0012	-0.0008	0.0004	0.1498	0.0212
	Sale	-0.0265	0.0016	0.0024	0.0003	-0.0647	0.0141
	Scratch card	0.0345	0.0014	0.0014	0.0003	0.1452	0.0281
	Observation	10292		10292		10292	
	R^2	0.018		0.009		0.007	

The experiences of the high-engagement (HE) groups reveal increased responsiveness to multiple promotional offers, reflecting their higher interaction levels. This variability underscores the potential for targeted marketing strategies and highlights differential effects across the AIDA model stages. Ad personalization positively influences ad clicks and subsequent actions for both engagement groups. Notably, in the action stage, the HE group has higher purchase amounts than the low-engagement (LE) group does, which has lower purchase amounts.

One challenge is balancing engagement with sales among different prospect segments. While personalization is highly effective for the HE group, it is less prevalent in the LE group because of limited connections. In the nonprice category, the type of promotion significantly impacts engagement: gifts substantially affect the awareness stage for HE groups. Moreover, free shipping has the most significant effect on LE groups.

The differences in promotional effects across groups underscore the importance of contextual considerations. Off-policy assessments enable companies to understand both types of heterogeneity and refine their targeting strategies to increase sales.

6.5. Targeting

Debiased measurements are seamlessly integrated with our off-policy evaluation (OPE) methods, as detailed by Dudík et al. (2011; 2014). These OPE methods improve the reliability of parameter estimates by reweighting results to account for changes in assignment on the basis of estimated propensity and results models. This integration enhances the accuracy and significance of our findings.

Table 13 Heterogeneous treatment components for high-engagement customers.

Type	Variables	Click	Std Error	Purchase	Std Error	Amount	Std Error
	Intercept	-0.0282	0.0041	-0.0012	0.0008	-0.0846	0.0288
	Characters	-0.0006	0.000	0.000	0.0000	-0.0012	0.0009
Product Type	Product	-0.0005	0.0012	0.0009	0.0004	0.0118	0.0131
	Clearance	0.00328	0.0021	0.0018	0.0002	0.1143	0.0114
Price	Discount	-0.0336	0.0037	0.0018	0.0004	0.0867	0.0338
	Discount Product	0.0412	0.0041	-0.0054	0.0005	-0.0814	0.0386
	Discount Clearance	0.0691	0.0102	-0.0006	0.0012	0.0489	0.0691
	Rebate	0.0062	0.0029	0.0031	0.0003	0.0861	0.0235
Nonprice	Free	0.0213	0.0019	0.0021	0.0002	0.0224	0.0294
	Shipping partly free	-0.0529	0.0018	-0.0006	0.0004	-0.0853	0.0141
	Shipping free	0.0346	0.0018	0.0003	0.0001	0.0092	0.0182
	Return free	-0.0131	0.0018	-0.0006	0.0007	0.00241	0.0192
Noninformative semantic choice	Customized	0.0781	0.0018	0.0041	0.0002	-0.0812	0.0104
	Interesting	0.0142	0.0029	0.0031	0.0003	0.0652	0.0245
	Additional	0.0086	0.0021	0.0021	0.0006	-0.0318	0.0182
	Limited	0.0291	0.0009	-0.0031	0.0004	-0.0048	0.0099
	Call	0.0141	0.0009	0.0011	0.0001	0.0429	0.0112
	Sale	-0.0158	0.0016	0.0024	0.0003	-0.0647	0.0141
	Scratch card	0.0061	0.0012	0.0003	0.0001	-0.0058	0.0098
	Observation	10292		10292		10292	
	R^2	0.008		0.006		0.005	

Our analysis indicates that customers receiving a 25% discount demonstrate greater engagement than those receiving a 50% discount do. This suggests that disengaged customers are less responsive to discounts, whereas engaged customers are more responsive. Companies can refine their targeting strategies by adjusting discount levels and enhancing advertisements, especially for clearance promotions, to maximize sales.

Our framework supports off-policy evaluation by using estimates to evaluate different policy regimes. By applying distinct samples for forecasting, we enhance the precision of our effect estimates across the overall population. Targeting high-engagement groups with 50% discount offers leads to a significant increase in sales, illustrating the framework's effectiveness in handling high-dimensional treatment data.

This methodology delivers accurate measures of heterogeneous treatment effects, offering valuable insights for companies aiming to fine-tune their targeting strategies. This underscores the importance of market segmentation for assessing causal effects and boosting sales. This research highlights the necessity of observational and experimental studies to understand the impact of design elements and marketing interventions.

7. Discussions

This research study's findings offer significant insights which aligns with the empirical results of other research studies conducted by certain authors such as: Darke and Dahl (2003), Grewal et al. (1998) into how customer engagement levels modulate responses to promotional offers, a factor that marketers can leverage for more effective targeting strategies. The responsiveness of high-engagement (HE) groups to personalized and contextual promotions underscores the importance of engagement-based segmentation. HE consumers display greater interactivity with multiple promotional offers, which suggests they are more attuned to marketing stimuli. This heightened responsiveness across different stages of the AIDA (Attention, Interest, Desire, Action) model indicates that personalized strategies resonate more deeply with HE customers and result in greater action-stage conversions (Barry and Howard, 1990). The finding that HE group members make larger purchases in the action stage compared to their low-engagement (LE) counterparts demonstrates the potential of tailoring promotions to customer engagement levels.

However, these findings also reveal a central challenge: balancing engagement with purchase outcomes across different customer segments. While HE customers respond positively to ad personalization, generating higher engagement and purchase amounts, the same level of personalization does not produce a comparable effect for LE customers. LE groups are less connected to the brand and thus may require distinct strategies to nurture engagement and move toward action. Our results indicate that for LE groups, nonprice promotions such as free shipping significantly impact engagement, especially in the awareness stage. Gifts, on the other hand, significantly increase awareness in the HE group. This differential effect by group highlights the need for contextual promotion types, as varied incentives appeal differently across engagement levels.

The methodological advancements of this study, particularly the integration of off-policy evaluation (OPE) methods with debiased measurements, facilitate a robust understanding of heterogeneous treatment effects. By adjusting for propensity and result model biases, OPE techniques provide more accurate parameter estimates, enhancing the reliability of our findings. This process is especially relevant for assessing discount levels, where our results suggest that a 25% discount elicits greater engagement than a 50% discount, hinting at possible discount fatigue or perceived value judgments among disengaged customers. These findings provide actionable insights for companies aiming to optimize promotions: offering moderate discounts to engaged customers may be more effective than steep discounts, which may not add further value for LE segments.

Our analysis also highlights the utility of high-dimensional treatment data in refining targeting strategies. When targeting HE customers with 50% discounts, companies experience a significant boost in sales, illustrating the potential of segmentation and personalization for maximizing return on investment. By utilizing distinct samples and debiased estimates, this framework supports a refined targeting model that considers diverse customer characteristics, behaviors, and engagement levels. This granularity enables marketers to tailor promotional elements precisely to audience segments, increasing the likelihood of meaningful engagement and purchase actions.

These insights point to the broader applicability of market segmentation in advertising, allowing businesses to draw causal inferences from treatment effects across varied customer segments. To address deficiencies in prior targeting policies, our study emphasizes the importance of understanding engagement-based heterogeneity in consumer response to marketing interventions. Future research could explore observational and experimental studies on these design elements to further validate and enhance targeting strategies. Moreover, several research studies investigated the impact of Artificial Intelligence in the case of the emerging country such as India (Shetty et al., 2023a; Gaddi et al., 2024; Shetty et al., 2023b; Ullal et al., 2022).

Overall, the research offers a comprehensive framework that marries theoretical constructs from the AIDA model with advanced machine learning techniques. By recognizing the nuanced impact of digital ads on different consumer groups, this approach provides valuable insights that help companies refine their marketing interventions to increase both engagement and sales.

8. Conclusions

Our framework effectively estimates the heterogeneous effects of various digital ad treatments by leveraging advanced machinelearning techniques to address endogenous targeting decisions and responses. In our study, we assess the impact of individual digital ads via the AIDA model, calculating the effects of treatment components despite potential endogeneity. We derive robust findings by incorporating semantic, contextual, and promotional signals.

The results reveal that the impact of these components is conditional on customer engagement, which is consistent with the literature. Differential effects on the basis of engagement levels provide valuable insights for practitioners, enabling them to refine targeting strategies according to varying semantic and promotional signals. This approach improves sales by applying contextual signals to different promotional strategies, addressing previous deficiencies in targeting policies.

Our framework, which combines theoretical constructs with off-policy evaluation methods, demonstrates how enhanced targeting can increase sales.

Ethical considerations

This research was conducted in adherence to ethical standards for studies involving human participants. All data were collected in a manner that ensured the privacy and confidentiality of the participants. No personal identifying information was collected, and all analyses were performed on anonymized data to protect participant identities. Given the study's observational nature, there were minimal risks to participants, and any potential for harm was mitigated through careful data management practices. The study received approval from the appropriate institutional review board (IRB), ensuring compliance with ethical guidelines.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this paper. No affiliations, financial arrangements, or personal relationships influenced the research, analysis, or conclusions drawn in this study.

Funding

No external funding was received for this study. The research was carried out independently without financial.

References

Ansari, A., & Mela, C. F. (2003). E-customization. *Journal of Marketing Research, 40*(2), 131–145.

Ascarza, E. (2018). Retention futility: Targeting high-risk customers might be ineffective. *Journal of Marketing Research, 55*(1), 80–98.

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7353–7360.

Athey, S., & Imbens, G. W. (2019). Economists should know about machine learning methods. *Annual Review of Economics, 11*, 685–725.

Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. *Observational Studies, 5*(2), 37–51.

Barry, T. E., & Howard, D. J. (1990). A review and critique of the hierarchy of effects in advertising. *International Journal of Advertising, 9*(2), 121–135.

Bonfrer, A., & Dreze, X. (2009). Real-time evaluation of email campaign performance. *Marketing Science, 28*(2), 251–263.

Box, G., & Wilson, K. (1951). On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society. Series B (Methodological), 13*(1), 1–45.

Chatterjee, P., & McGinnis, J. (2010). Customized online promotions: The moderating effect of promotion type on deal value, perceived fairness, and purchase intent. *Journal of Applied Business Research, 26*(4).

Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2017). Generic machine learning inference on heterogeneous treatment effects in randomized experiments. Technical report, National Bureau of Economic Research, Cambridge, MA.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econometrics Journal, 21*(1), C1–C68.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Academic Press.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton Mifflin.

Cox, D. R. (1958). *Planning of experiments.* Wiley.

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimating average treatment effects. *Biometrika, 96*(1), 187–199.

Darke, P. R., & Dahl, D. W. (2003). Fairness and discounts: The subjective value of a bargain. *Journal of Consumer Psychology, 13*(3), 328–338.

Diamond, W. D., & Johnson, R. R. (1990). The framing of sales promotions: An approach to classification. In M. E. Goldberg, G. Gorn, & R. W. Pollay (Eds.), *Advances in Consumer Research* (Vol. 17, pp. 494–500). Association for Consumer Research.

Dudík, M., Langford, J., & Li, L. (2011). Doubly robust policy evaluation and learning. *Proceedings of the 28th International Conference on Machine Learning*, 1097–1104.

Dudík, M., Erhan, D., Langford, J., & Li, L. (2014). Doubly robust policy evaluation and optimization. *Statistical Science, 29*(4), 485–511.

Ellickson, P. B., Kar, W., & Reeder III, J. C. (2023). Estimating marketing component effects: Double machine learning from targeted digital promotions. *Marketing Science, 42*(4), 704–728.

Gaddi, A., Kulkarni, P., Shetty, S. K., Birau, R., Popescu, V., & Hiremath, G. S. (2024). Exploring evolving H.R. and recruitment strategies in the age of technology advancements based on artificial intelligence. *Multidisciplinary Science Journal, 7*(2), Article 2025043. https://doi.org/10.31893/multiscience.2025043

Gopalakrishnan, A., & Park, Y.-H. (2021). The impact of coupons on the visit-to-purchase funnel. *Marketing Science, 40*(1), 48–61.

Gordon, B. R., Moakler, R., & Zettelmeyer, F. (2022). Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. *Preprint.*

Gordon, B. R., Zettelmeyer, F., Bhargava, N., & Chapsky, D. (2019). A comparison of approaches to advertising measurement: Evidence from extensive field experiments at Facebook. *Marketing Science, 38*(2), 193–225.

Grimmer, J., Messing, S., & Westwood, S. J. (2017). Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Political Analysis, 25*(4), 413–434.

Grewal, D., Monroe, K. B., & Krishnan, R. (1998). The effects of price-comparison advertising on buyers' perceptions of acquisition value, transaction value, and behavioral intentions. *Journal of Marketing, 62*(2), 46–59.

Heckman, J. J., & Vytlacil, E. J. (2007). Econometric evaluation of social programs, Part I: Causal models, structural models, and econometric policy evaluation. *Handbook of Econometrics* (Vol. 6, pp. 4779–4874).

Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health, 60*(7), 578–586.

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if.* Chapman & Hall/CRC.

Hitsch, G. J., & Misra, S. (2018). Heterogeneous treatment effects and optimal targeting policy evaluation. *Preprint.*

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81*(396), 945–960.

Imai, K., & Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *Annals of Applied Statistics, 7*(1), 443–470.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalised boosted models. *Statistics in Medicine, 32*(19), 3388–3414.

Neyman, J. (1923). Statistical problems in agricultural experimentation. *Journal of the Royal Statistical Society, 2*(2), 107–180.

Park, C. H., Park, Y.-H., & Schweidel, D. A. (2018). The effects of mobile promotions on customer purchase dynamics. *International Journal of Research in Marketing, 35*(3), 453–470.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association, 53*(284), 873–880.

Robins, J. M., & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association, 90*(429), 122–129.

Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics, 6*(1), 34–58.

Rubin, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test. *Journal of the American Statistical Association, 75*(371), 591–593.

Sahni, N. S., Wheeler, S. C., & Chintagunta, P. (2018). Personalization in email marketing: The role of noninformative advertising content. *Marketing Science, 37*(2), 236–258.

Sahni, N. S., Zou, D., & Chintagunta, P. K. (2017). Do targeted discount offers serve as advertising? Evidence from 70 field experiments. *Management Science, 63*(8), 2688–2705.

Semenova, V., & Chernozhukov, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *Econometrics Journal, 24*(2), 264–289.

Shetty, S. K., Spulbar, C., Birau, R., & Filip, R. D. (2023a). Impact of artificial intelligence in banking sector with reference to private banks in India. *Physics AUC, 32*, 59–75.

Shetty, S. K., Birau, R., Ninulescu, P. V., Filip, R. D., & Lupu (Filip), G. A. M. (2023b). Acceptance and adoption of digital financial inclusion by the rural population of India: A case study. *Revista de Științe Politice. Revue des Sciences Politiques, 80*, 162–177.

Sinha, I., & Smith, M. F. (2000). Consumers' perceptions of promotional framing of price. *Psychology & Marketing, 17*(3), 257–275.

Ullal, M. S., Nayak, P. M., Dais, R. T., Spulbar, C., & Birau, R. (2022). Investigating the nexus between artificial intelligence and machine learning technologies in the case of Indian services industry. *Business: Theory and Practice, 23*(2), 323–333. https://doi.org/10.3846/btp.2022.15366

Van der Laan, M. J., & Rose, S. (2011). *Targeted learning: Causal inference for observational and experimental data.* Springer Science & Business Media.

Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology, 6*(1), 1–21.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association, 113*(523), 1228–1242.

Yi, Y., & Yoo, J. (2011). The long-term effects of sales promotions on brand attitude across monetary and non-monetary promotions. *Psychology & Marketing, 28*(9), 879–896.

Yoganarasimhan, H., Barzegary, E., & Pani, A. (2020). Design and evaluation of personalized free trials. Technical report, University of Washington, Seattle.

Zeelenberg, M., & Putten, V. M. (2005). The dark side of discounts: An inaction inertia perspective on the post-promotion dip. *Psychology & Marketing, 22*(8), 611–622.

Zhang, X., Kumar, V., & Cosguner, K. (2017). Dynamically managing a profitable email marketing program. *Journal of Marketing Research, 54*(6), 851–866.