Proceedings of CONF-BPS 2026 Symposium: GenAI, Labour Markets, and the Economics of Human and Financial Capital
DOI: 10.54254/2754-1169/2026.LD31363

Algorithmic Transparency and the Personalization Paradox: A Structural Equation Review of AI-driven Social Media Advertising, User Trust, and Behavioral Outcomes

Pinxuan Yin

Faculty of Business and Management, Beijing Normal-Hong Kong Baptist University, Zhuhai, China
t330036180@mail.bnbu.edu.cn

Abstract. The black box design of the algorithmic choices made by AI recommendation systems with its extensive use in social media advertisements has come under significant focus regarding questions of transparency, privacy and user trust. Platforms also face the trust risk of personalized strategies in their pursuit of ad accuracy. This article is an analytical article that takes a structured review of the relationship between the characteristics of an algorithm output and the psychological variables of its users, including advertising experience, perception of privacy, trust, and behavioral intentions, using a quantitative framework analysis. Depending on the literature, it is found that personalized advertising influences users in two dimensions. On the one hand, recommendations will enhance advertising experience of users, promoting their higher short-term click-through rate and purchase intentions. In its turn, it happens to be the fact that over-personalization and high frequency usage can result in privacy related issues, as well as undermine the trust and the dedication of the users in the long term. In addition, the privacy image and the exploratory sale of the ad are also important factors that act as a mediator between the mentioned mechanisms. As per that statement, the following article proposes a generalizable theoretical framework in SEM that would provide a measurable framework in analyzing the two psychological impacts of personalized advertising.

Keywords: Algorithmic transparency, social media advertising, privacy concern, structural equation modeling.

1. Introduction

At present, social media have evolved to become a significant source of information that can be received and consumed by people. In that regard, recommendation algorithms that are performed by artificial intelligence have been thoroughly embedded into advertising framework. This technology has greatly improved the performance of ad clicks and conversions on the platform as advertisers are able to target their advertisement to the specific population. Nevertheless, non-existent transparency is another obvious disadvantage of algorithmic recommendations. Although there is a transparency

DOI: 10.54254/2754-1169/2026.LD31363

and explainability of algorithms that is being increasingly discussed, all the industry remains lacking an agreeable and practical technical standard.

Individualized suggestions on the side of social media advertising have the potential to enhance instantaneous involvement as well as buying actions. But, they do not necessarily add user trust or satisfaction which is a personalization paradox [1]. Over personalization can cause the user to experience violation of privacy, which in turn prompts them to be in a negative mood where they are defensive [2]. The ability to effectively personalize and at the same time reduce the amount of psychological pressure is one of the most important challenges.

Though, previous research has covered technical optimization and user psychology individually, there is yet to find an overall framework to connect the features of the algorithm with user perception and behavior. On transparency, there is still much speculation and still has not been related with real outputs of algorithms or user information.

Based on the aforementioned research gaps, this article aims to integrate the three research threads of technology, psychology, and governance to provide a systematic review of personalized social media advertising and propose an operational analytical framework. Specifically, this study first reviews the main technical paths for algorithm transparency and interpretability. These paths include local interpretation, global proxy models, counterfactual inference, and external behavior auditing to demonstrate the various implementation methods of "algorithms that can be understood and verified" in advertising recommendation systems. Secondly, this article compares the consideration of the user psychology and behavior. By resolving the "dual path" mechanism of personalization between experience enhancement and privacy threat, examine how it has a concurrent influence on the users short and long term behavioral reactions as the result of advertising experience, privacy issues, and trust aspects (ability, integrity, goodwill). Thirdly, the paper is based on structural equation modeling (SEM) as an unified framework to give a methodological overview of the method of measuring issues, path structures, and estimation strategies of latent issues and suggests a template model of psychological mechanisms of behavioral outcomes that can be used repeatedly. Lastly, this paper uses the integration framework to Tik Tok. It is a platform that has placed a lot of importance on the use of algorithm driven content and advertising to illustrate the interpretability and empirical worth of the model in a business context. As the logic of recommendation on TikTok, the controversies surrounding transparency, and psychological responses of users are analyzed, the present study points out the reason why the platform represents best the personalized dual path mechanism.

2. Forms and technical paths of algorithm transparency

2.1. The multi-dimensional connotation of algorithm transparency

The notion of algorithm transparency is not just one technical quality, but a network of both technical and institutional structures that operate to explainability, auditable and accountable at various levels [3]. According to the variations in the implementation scope and initial goals, the main three types of transparency exist in scholarly settings, i.e., local transparency, global transparency, and functional transparency.

Local transparency concentrates on describing the results of a single prediction. It reveals the internal reasoning that the model employs in order to generate a certain output. The global transparency is directed to expose the existing general decision-making trends and main traits of the model. In functional transparency, there is a focus on checking the logic of operation of algorithms based on an external behavioral viewpoint, which serves to assess their feasibility and possible bias.

DOI: 10.54254/2754-1169/2026.LD31363

2.2. LIME and SHAP local interpretation methods

Two representative in the concept of interpretable artificial intelligence, LIME (Local Interpretable Model Ignore Explanations) and SHAP (Shapley Additive exPlans), have become fundamental technical supports of improving the transparency of an algorithm on a local level. Each of the two approaches has its peculiarities and usefulness. LIME approximates the local contribution of every feature to the prediction outcome with the help of linear approximation of the data points surrounding a single sample [4]. The advantage of this approach is that it is model independent and therefore the users can know why they are being recommended some form of ads. Conversely, SHAP method relies on the Shapley value concept of game theory which is capable of estimating the mean marginal contribution of each feature to the model outcome of all possible combinations of features [5].

These two techniques have been used broadly in the interpretation scenarios of user interest prediction and advertising ranking logic in the area of advertisement recommendation. Indicatively, researchers apply SHAP method to comprehend the significance of features in the recommendation systems. This approach can be used to identify the features of user behavior that account for the largest percentage of the prediction of click-through rate [6].

However, local interpretation methods still have significant limitations in practical applications. Its high computational cost, limited real-time performance, and the possibility of causing users' "false understanding" problems. This type of method can only reflect local approximate results and is difficult to reveal the causal structure of the algorithm as a whole, which can easily lead users to mistakenly believe that model decisions are based on stable logic.

2.3. Global agents and causal methods

In order to know the general rules of decision making of the algorithm, a global surrogate model must be built. This kind of model is both visually and verifiably allowed to visualise the overall logic by the training of understandable simplified models, like decision trees and logistic regression, but preserving the fundamental predictive patterns of the original model [7].

Additionally, the method of causal and counterfactual explanations that has occurred in the recent years have also given a deeper approach in theory which can be used in studying transparency. This approach constructs hypothetical situations through which it makes an inference. To take an example, will advertising recommendations remain the same without the geographic location information? Through this sort of counterfactual analysis, the model output can be tested to determine whether it is causally stable [8]. It has an advantage in that it can expose the possible dependence among variables and provide mechanistic answers regarding the outcome of the input changing.

Nevertheless, several shortcomings exist to the use of such methods in the real platforms, including limited data access, heavy reliance on a sample, and inability to observe assumptions. Namely, the characteristics of training data in social media situations are very often high-dimensional and fail to meet the requirements of independent and identically distributed features. This implies that in construction of a stable and reliable counterfactual model you generally require extra assumptions and have to be validated externally.

DOI: 10.54254/2754-1169/2026.LD31363

2.4. External transparency and behavioral modeling audit

Besides the internal methods of explanation, recently, the academic community suggested application of external auditing as a technique to understand the true functioning process of the recommendation algorithms [9]. This approach implies that researchers attempt to derive by using black box experiments on how the algorithm ranks content. They accomplish this through simulation of user behavior, measurement of recommended content and statistical modeling, without even seeing the source code of the platform.

The differentiated performance and operation mechanism of algorithmic recommendation systems have attracted the attention of the academic community. For example, Bär et al. performed a behavioral audit with the large-scale metapolitical data on advertising [10]. They found that there was a big difference between the target audience of the advertising and the real audience. As evidenced in the study, the algorithm shows varying push patterns on the different groups of users. In their paper, the researchers of Gausen et al. analyzed the mechanisms of weight allocation of different content planning signals in the agent-based modeling framework of recommendation algorithms [11]. Its model demonstrates the mechanism of the trade-offs between various signals on the platform, including the intensity of interaction, stay time and the popularity of the content.

These experiments demonstrate that even without the direct access to model permissions inside the platform when the researchers are still able to recognize the operation properties of algorithms by recreating the behavioral patterns between the inputs and outputs. This approach achieves unveiling the transparency of the algorithm on the functional level. Moreover, an external audit is another promising means of conducting academic research, which may allow finding observable signs and support model hypotheses.

3. The mechanism of personalization on user cognition, trust, and behavior

3.1. Algorithmic psychological model of cognition and attribution

When users receive personalized advertisements, they will tend to theorize in their minds why the platform decides to recommend the advertising content of the project to them. This is an academic term that defines this process of thinking judging by personal experience as an algorithmic mental model [12]. Different clues will be used by the user to construct their knowledge of the logic that the algorithm operates based on which specific content the ad showed, the length of time taken to push, and the level of association invoked to the ad and their previous behavior. However, such psychological thoughts are not always complete and even contain significant cognitive biases. For example, some users may attribute ad recommendations to the system recognizing their own topics of interest, while others may feel that the platform is monitoring their privacy.

Research has shown that algorithm transparency is not a single dimension. It includes multiple aspects such as performance transparency and process transparency. These different forms of transparency will respectively affect users' perception and trust in the algorithm [13]. Moderate transparency can enhance users' trust in algorithm capabilities. However, if the information is too technical or overly exposed to commercial motives, it may actually trigger users' suspicion and defensive psychology. It can be seen that transparency has a double-edged sword effect, and its actual effect depends not only on the type of information disclosed, but also closely related to the user's existing cognitive foundation.

DOI: 10.54254/2754-1169/2026.LD31363

3.2. The three-dimensional structure of trust: ability, integrity and goodwill

Trust is a multidimensional psychological state. It is made up of three components, which include ability, integrity and goodwill [14]. These three categories of trust are applied in advertising algorithms, and it means that users trust the algorithm to make relevant predictions, believe that the platform does not break the rules, and look forward to the platform not abusing their information.

Understanding the user trust structure, three dimensional trust must be separated, to minimise the simplification of transparency effect to one direction, as much as possible. The empirical evidence is that the transparency of the algorithm primarily contributes to the establishment of confidence in skills. This will imply that it will make users assured that the system is capable of effectively matching their interests and preferences [15]. However, in some particular cases, it can undermine integrity and altruism of the user. Particularly in cases where the information published by the algorithm causes users to realize that they have some bi-business motive. As an example, in the case a platform publicly states that it uses algorithms to cut advertisements placement depending on the browsing history and location of a user, users can be in doubt of the fact that the platform does care about their interests.

3.3. Attitude towards privacy and emotional reaction

Individualized recommendations are based on data gathering and profiling of the user, which usually transcends the privacy of the user. In cases where the advertisements become overly aligned with the interests of the users particularly in cases where the modified information is sensitive like location and health, one of the users may have the impression that their privacy has been compromised. More deadly, it even has the ability to give the users the feeling of being stalked [1]. This not only causes a decrease in the favorability of users of the platform, it can also result in defensive actions like blocking of advertisements, fewer permissions over personal data or less frequent use of the platform. According to Boerman et al., consumers who are informed that they are the victims of cross-platform tracking via advertisements respond with reduced levels of trust ($p < 0.001$) and greater levels of discomfort [2]. As it may be observed, one of the mediating variables is the factor of privacy perception, which gives rise to negative outcomes of personalized recommendations. The extreme customization of adverts and the users not only captivate them, but also make them anxious because of their extreme precision, users fear their personal information is being misused.

3.4. Advertising experience and short-term behavioral response

The most significant psychological channel that enhances personalization to produce positive effects in the short-term is advertising experience, as compared to the negative influence of privacy issues. This is a notion comprising of emotional appeal, informativeness and creativity [16]. In case the advertising content is strongly suited to the interests of users, rendered more simply and somewhat entertaining, users are likely to create positive emotions, and these emotions cause them to initiate clicking or purchasing. A study connected with the structural equation modeling [17] proves that personalization may greatly increase in the purchase intention of the users by the optimization of the advertisement experience along the indirect paths (indirect effect is significant, $p<0.001$). This temporary positive impact however, cannot be maintained in the long run. Once users become aware of the excessive push of the advertisement or when the advertisement content is redundant, a decline in the pleasure in their experience will ensue very fast. The depreciation of such experience will in turn undermine their credence and general goodwill about the platform.

DOI: 10.54254/2754-1169/2026.LD31363

3.5. Exposure frequency and social media fatigue

Exposure frequency is another significant outside process in influencing personalized advertising experience of the users against the advertising content itself. It has been previously demonstrated that the user in circumstances where there is an overload of information input tends to experience cognitive overload resulting in psychological exhaustion. Bright et al. discovered that excessive information on social media contributes immensely to fatigue on the part of the users and avoidance behavior [18]. Even though the study was not specifically aimed at advertising situations, this notion of overload mechanism that the study brought to light has been utilized extensively in future research on advertising to explain the resistance to high-frequency advertising in users. Structural equation modeling studies also provide further information that the frequency of the exposure itself is moderating in the relationship between the advertising experience and trust. The promoting effect of the relevance of content and pleasure on user trust will be highly undermined when the frequency or intrusion on the advertisement is too high [19]. It means that the too high attempt by the platforms to achieve exposure efficiency can cause overcrowding of information. This is not only undermining a good impression on advertising experience to trust, but it also hurts the relationship between the users and the platform in the long run.

3.6. Dual-path psychological mechanism model

The psychological process of customized advertisement could be outlined as a two-path process, based on the literature above. The first direction is the positive one: perceived personalization tends to enhance the advertising experience (made one feel more relevant or enjoyable). This is capable of making the users more confident with the site and the ads, after which behavioral intentions can be expanded like clicking, buying, or further interaction. The negative one is the second: perceived personalization can also pose a privacy risk, decrease the trust of the users, and cause such avoidance behavior as ignoring, blocking, or leaving. Meanwhile, exposure frequency influences the two paths. This means that moderate factors reinforce positive path but high-frequency repetition has the effect of reinforcing privacy anxiety and fatigue and therefore the bad effects are more. The Dual path model brings to light the two psychological effects that personalized advertisement has on the users.

4. SEM-based integrative framework for personalization studies

4.1. The reason for adopting the structural equation model

In examining the connection between algorithm characteristics, the variables of psychology, and user action in personalized advertisement, the conventional regression analysis is ineffective in most cases as it does not account for numerous mediating and moderating variables. Structural Equation Modeling (SEM) has an added benefit of estimating these effects either in the same run and transforming abstract psychological constructs to latent variable [20]. In the work, SEM is applied to combine the indicators of the algorithm output, like its accuracy of prediction and the rate of exposure, with variables of users in this situation, trust, privacy concerns, and advertising experience, allowing the complex analysis of the factors connecting the characteristic of an algorithm and user reactions.

SEM also gives an opportunity to study how perceived personalization is effective but affecting an advertising experience and the perception of privacy, which in turn affects the attitudes to trust

DOI: 10.54254/2754-1169/2026.LD31363

and behavioral intentions and moderate factors (frequency of exposure is used as examples). The approach offers a strict quantitative model of recording the reaction of users to personalized ads in social media.

4.2. Variable design and measurement index suggestions

Based on the integration of the aforementioned literature, the SEM framework proposed in this study includes five core latent variables and one technical variable (see Table 1). Each latent construct is measured using three to four observed items on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Subjective perceptual variables are obtained through questionnaires, while selected algorithm output indicators are derived from platform or experimental data.

Table 1. Variable design and measurement indicators in the SEM framework

Potential variables	Sample questions	Data source
Perceived Personalization	“The advertisement is highly matched with my interests”; “I think the advertisement is tailor-made for me”	Questionnaire + algorithm output(relevance_score, CTR_pred)
Advertising Experience	“The advertisement makes me happy”; “The advertising content is friendly and attractive”	Questionnaire
Privacy Concern, PrivCon	“I am concerned that the platform may misuse my personal information”; “I cannot control how the data is used”	Questionnaire
Trust	“I trust the advertisements recommended by this platform”; “I think this platform protects my privacy”	Questionnaire
Behavioral Intention	“I'm willing to click on similar advertisements”; “I may buy related products because of it”	Questionnaire
Exposure Frequency	Average number of advertisement appearances; time spent on the feed (time_on_feed)	Platform or experimental data

Among them, the output variables of the algorithm (such as relence_store, CTR_pred, exposure_comunt, time_on_feed) have dual application value. They can be included in SEM as independent exogenous observation variables or together as measurement indicators for the latent variable of "perceived personalization". The theoretical basis for doing so is that these indicators reflect the objective evaluation of the algorithm's personalized matching degree from a technical perspective. And there is a synchronous correlation between the evaluation results and the subjective perception of users [21].

In terms of model structure, algorithm metrics can be integrated in two ways. On the one hand, it can be set as an exogenous variable that directly affects psychological variables such as advertising experience or privacy concerns. This pathway can be used to examine the effects of technical input on psychological responses, thereby identifying cross level mechanisms of action. On the other hand, it can serve as a formative indicator for personalized perception, and together with subjective scales, form a mixed measurement model. This approach helps to enhance the external validity and empirical basis of latent variables.

By combining subjective questionnaire and algorithm log data, SEM can more accurately depict the chain of "technology input psychological mechanism behavioral output". This makes the model

DOI: 10.54254/2754-1169/2026.LD31363

not only have theoretical explanatory power in psychology, but also meet the operational requirements of algorithm research.

4.3. Structural equation model and example equation expression

Based on theoretical assumptions, the SEM path structure proposed in this article is as follows:

$$AdExp = \beta_1 \cdot PersPer + \zeta_1$$

$$PrivCon = \beta_2 \cdot PersPer + \beta_3 \cdot AdExp + \zeta_2$$

$$Trust = \beta_4 \cdot AdExp - \beta_5 \cdot PrivCon + \beta_6 \cdot (AdExp \times Exposure) + \zeta_3$$

$$BehInt = \beta_7 \cdot Trust + \zeta_4$$

Among them, ζ_i represents the residual term of each endogenous variable.

This model includes three main paths. First, perceived personalization can raise users' trust in the platform by improving the advertising experience, forming a positive mediating path. Second, perceived personalization may also cause privacy concerns, which weaken trust and create a negative mediating path. At the same time, exposure frequency plays a regulatory role in the path of "advertising experience trust". It means that when the exposure frequency is high, the positive impact of advertising experience on trust is significantly weakened ($\beta_6<0$). This structure presents both the dual track effect of psychological mechanisms and the technical dimension of algorithmic features, consistent with the previous dual path model.

4.4. Estimate strategy and model inspection process

To guarantee the reliability of the model's statistical outcomes, a standardized estimation and testing procedure ought to be adopted in the research. Firstly, the sample size should be controlled between 200 and 400. If the actual sample size is limited, partial least squares structural equation modeling can be used for parameter estimation to improve the reliability of the results. Second, reliability and validity are tested using Cronbach's α (>0.70) and AVE (>0.50), and model fit is checked with CFI/TLI (≥0.90) and RMSEA/SRMR (≤0.08). The mediating effect is tested using Bootstrap (about 5000 times) with a 95% confidence interval. In addition, multiple SEM models are compared to see differences in paths under different exposure conditions or user groups. This process ensures the model's estimates are reliable at both technical and psychological levels.

5. Case analysis: TikTok's AI personalized advertising system

TikTok is one of the fastest-growing content platforms globally, and its AI recommendation and advertising system is highly representative. Different from traditional advertising platforms, TikTok embeds advertisements directly into the "FOR YOU" information stream displayed in full screen, achieving almost seamless integration between commercial content and ordinary videos. This design not only improves the algorithm's ability to capture user preferences in real time, but also increases the psychological impact that a "high personalization, low transparency" model can create. Therefore, TikTok has become a typical case for testing the psychological mechanism of personalized advertising.

5.1. Path I: algorithmic inputs and direct psychological effects

TikTok is founded on more down-to-the-nine behavioral indicators, such as rate of completion, watch time, and repeat views, to continue to tune the user preference models. Audit studies demonstrate that such behavioral indicators are the key to the achievement of personalization [22].

DOI: 10.54254/2754-1169/2026.LD31363

These indicators in the former channel are the exogenous algorithmic input to the channel that produces the exposure setting and user psychological effects directly.

As the logic of recommendation on the TikTok platform is not evident, users can only make an informed guess on what influences the system because it impacts them. This smog may go to an extent of making the users feel that there is some form of manipulation of the algorithms to make them read as commercial [23]. This is further aggravated by the aspect of high-frequency exposure which destroys the experience and the confidences in advertising and shows the direct input of technical intensity to the psychology of user.

5.2. Path II: perceived personalization, psychological mediation, and behavioral outcomes

The personalization paradox describes how stronger personalization does not necessarily enhance trust [1], and TikTok represents a typical real-world case. By inferring user interests and emotional states, the platform pushes highly targeted advertising, particularly around sensitive topics. The perceived personalization is reinforced by this practice, and the privacy and the feeling of being perceived as being looked through are heightened.

Under the circumstances of the dark side of the route as proposed in this paper, the perceived improvements of the personalization extend or increase the privacy and reduce user trust, decreasing the interaction in the long term. In the meantime, the immersion advertising systems of TikTok complement the advertising experience during a short time and prefer immediate communication [16]. These findings conform to the positive mediation effect that has been discovered in this paper in which perceived personalization increases advertising experience and ability based trust that spawns short term behavioral reactions.

5.3. SEM mapping: technical variables to psychological variables

TikTok's highly quantifiable metrics make it highly suitable for SEM empirical analysis. The key measurements correspond to the following mapping between SEM latent variables and observable TikTok indicators (see Table 2).

Table 2. Mapping between TikTok technical indicators and SEM variables

SEM variable	Corresponding TikTok observable indicators
Perceived Personalization	similarity_score; behavioral match; interest-vector alignment
Ad Experience	entertainment; informativeness; non-intrusiveness
Privacy Concern	perceived surveillance; emotion inference concern
Trust	platform trust; integrity trust
Behavioral Intention	click; engagement; app install; purchase
Moderator (Exposure)	ad fill_rate; repetition count; impression frequency

In order to verify the applicability of the aforementioned theoretical framework in real platforms, this paper maps the system features of TikTok to the core latent variables of SEM. At the technical level, log indicators such as interest vectors and similarity score can serve as objective proxy variables for perceived personalization. The entertainment and informational aspects of immersive short video advertising correspond to the dimension of "advertising experience". And emotional inference based on micro behavior reinforces the psychological variable of 'privacy concerns'.

DOI: 10.54254/2754-1169/2026.LD31363

Users' judgments on platform capabilities, integrity, and data processing norms constitute 'trust', while clicking, interaction, and installation behavior can be observed as' behavioral intentions'. In addition, TikTok's high ad fill rate and repeated content exposure make "exposure frequency" a key moderating variable. The increase of this variable will weaken the positive effect of advertising experience on user trust. Through this mapping, the SEM-based chain of action of “technology input–psychological mechanism–behavioral outcomes” can be empirically tested, with TikTok serving as a concrete application context.

6. Conclusion

Artificial intelligence driven social media advertising has become a core component of the digital economy. However, as much as it enhances the efficiency and commercial worth of information that is a match, it also presents new concerns related to privacy protection, transparency on the algorithms, and trust to the users. The paper methodically evaluates available research at three tiers, namely, algorithm transparency, user psychological processes, and model quantification techniques, and develops a framework of integrated analysis that relies on structural equation modeling. The purpose of this framework is to unveil the full process of individualized advertising in a systematic manner through which the process starts with the technical input and finally ends with behavioral outcomes.

The transparency is done technically by using local interpretation, global proxy models, counterfactual reasoning and behavioral auditing. These techniques are complementary in the sense of interpretability, system analysis, and independent verification and offer the technical foundations of a fairly transparent algorithmic control model. Psychologically, this article is a summary of the dual process of personalized advertising: advertisement experience increases trust and temporary behavior; whereas privacy concerns created by the inference of the data result in decreased trust in the long run. The frequency of exposure is an important moderating variable, and it has an impact on the strength of these two pathways and is a crucial factor in trust stability. At the methodological level, the structural equation modeling approach that has been suggested in this paper allows considering jointly algorithm log indicators and psychological latent variables through the way of a quantitative analysis. In this framework, technical indicators, including relevance score, click through rate prediction, and exposure frequency are combined with psychological indicators, including advertising experience, privacy concerns, and trust, into one analytical system. The framework enhances the explanatory value of personalized advertisement studies and offers a transposable analytical file of platform auditing, system detection, and cross-platform comparative investigations.

In general, personalized advertising is constructively based on the dynamic equilibrium between effectiveness and trust. Further studies can confirm its dual pathway mechanism in multicultural backgrounds, multi platform condition, and in the long term and integrate both controlled experiment and platform audit data to increase the external validity of the results. Personalized advertising is likely to keep on advancing in their more transparent, controllable and long term value oriented direction through constant experimentation.

References

[1] Aguirre E., Mahr D., Grewal D., de Ruyter K., Wetzels M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. Journal of Retailing, 91(1), 34–49.

DOI: 10.54254/2754-1169/2026.LD31363

[2] Boerman S. C., Kruikemeier S., Zuiderveen Borgesius F. J. (2021). Understanding the effects of personalization as a privacy calculus. Computers in Human Behavior, 123, 106868.
[3] Johnson D. G. (2021). Algorithmic accountability in the making. Social Philosophy and Policy, 38(1), 115–136.
[4] Ribeiro M. T., Singh S., Guestrin C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016) (pp. 1135–1144). New York: ACM.
[5] Lundberg S. M., Lee S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems, 30, 4765–4774.
[6] Zhang Y., Chen X. (2018). Explainable recommendation: A survey and new perspectives. Foundations and Trends in Information Retrieval, 14(1), 1–101.
[7] Molnar C. (2020). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). Leanpub.
[8] Wachter S., Mittelstadt B., Russell C. (2018). Counterfactual explanations without opening the black box. Harvard Journal of Law & Technology, 31(2), 842–887.
[9] Raji I. D., Xu P., Honigsberg C., Ho D. E. (2022). Outsider oversight: Designing a third-party audit ecosystem for AI governance. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES 2022) 557–571. New York: ACM.
[10] Bär D., Pierri F., De Francisci Morales G., Feuerriegel S. (2024). Systematic discrepancies in the delivery of political ads on Facebook and Instagram. PNAS Nexus, 3(7), 247.
[11] Gausen A., Guo C., Luk W. (2025). An approach to sociotechnical transparency of social media algorithms using agent-based modelling. AI and Ethics, 5(2), 1827–1845.
[12] Eslami M., Krishna-Kumaran S. R., Sandvig C., Karahalios K. (2018). Communicating algorithmic process in online behavioral advertising. In Proceedings of the CHI Conference on Human Factors in Computing Systems (pp. 1–13). New York: ACM.
[13] Zerilli J., Bhatt U., Weller A. (2022). How transparency modulates trust in artificial intelligence. Patterns, 3(4), 100455.
[14] Mayer R. C., Davis J. H., Schoorman F. D. (1995). An integrative model of organizational trust. Academy of Management Review, 20(3), 709–734.
[15] Shin D., Park Y. J. (2019). Role of algorithmic transparency in fostering trust and adoption of AI-powered systems. Computers in Human Behavior, 98, 138–156.
[16] Lee J., Hong I. B. (2016). Predicting positive user responses to social media advertising. International Journal of Information Management, 36(3), 360–373.
[17] Bleier A., Eisenbeiss M. (2015). Personalized online advertising effectiveness. Marketing Science, 34(5), 669–688.
[18] Bright L. F., Kleiser S. B., Grau S. L. (2015). Too much Facebook? Computers in Human Behavior, 44, 148–155.
[19] Kim H., Sundar S. S. (2020). Can interface features affect users’ reactions toward online ads? Journal of Computer-Mediated Communication, 25(6), 402–418.
[20] Kline R. B. (2016). Principles and practice of structural equation modeling (4th ed.). New York: Guilford Press.
[21] Tucker C. E. (2014). Social networks, personalized advertising, and privacy controls. Journal of Marketing Research, 51(5), 546–562.
[22] Vombatkere K., Mousavi S., Zannettou S., Roesner F., Gummadi K. P. (2024). TikTok and the art of personalization. In Proceedings of the ACM Web Conference 2024 (WWW 2024) (pp. 3789–3797). New York: ACM.
[23] Huang Y., Liu L. (2025). The impact of algorithm awareness on the acceptance of personalized social media content recommendation. Acta Psychologica, 259, 105383.