When Personalization Becomes Persuasion: Disclosure, Autonomy, and Conversion in Generative-AI Advertising

Md. Nabidul Haque
Correspondence: mdnabidulhaque51@gmail.com

Published 2025 · DOI: 10.61424/rjbe.v3i4.1082

ABSTRACT

Generative AI enables advertisers to customize content on a per-person basis with minimal additional expense. The ability to make things seem more relevant can also diminish the persuasive message and reduce consumer choice. This study investigates the influence of personalization intensity and disclosure of AI generated content on the perceived relevance, manipulation, autonomy, trust, and conversion intention. A simulation-based 2 × 2 online experiment was designed using 1,248 synthetic respondent records, half assigned to disclosure-present and half to disclosure-absent settings, with half of the records in each of these disclosure conditions being assigned to low-personalization and the other half to high-personalization settings. Multi-item measures were analyzed by means of factorial analysis of variance, robust regression and 5,000-sample bootstrap mediation. High personalization also led to greater manipulation ($M = 4.48$) and less autonomy ($M = 4.18$) without disclosure, and greater relevance ($M = 5.15$) with disclosure ($p < .001$). Disclosure raised conversion under high personalization from $M = 3.75$ to $M = 4.28$, interaction $F(1, 1244) = 17.80$, $p < .001$, partial $\eta^2 = .014$. Conditional indirect effect pathway was mediated through autonomy, manipulation and trust was $b = .48$, 95% CI [.38, .58]. The results suggest that disclosure can be a governance tool supporting the conversion process if it can regain informed choice, rather than simply marking content. Personalization design can be made transparent to match performance with consumer agency.

Keywords: generative AI; personalized advertising; disclosure; consumer autonomy; trust; conversion

1.Introduction

Generative AI has revolutionized ad personalization, moving beyond the idea of using pre-existing variations of a message to actually creating the copy, images, offers, and conversation answers on the fly. This change in economics and operational scope of targeting is driven by this development. AI systems can be used throughout the customer journey, allowing companies to personalize their advertising messages based on predictive information about consumer needs in large volumes (Davenport et al., 2020). Simultaneously, generative systems create governance risks as their outputs are not always easy to audit, not always clearly linked to the mechanisms that have long regulated advertising claims and may be opaque (De Bruyn et al., 2020). Advertising is becoming not only more personalised, but also more adaptive in form, timing and persuasive structure of its messages.

AI is typically framed as a suite of features and tools designed to be used in various marketing tasks and business goals. Strategic frameworks categorize AI into mechanical, analytical and intuitive forms, and claim that more sophisticated systems will take on more of an active role in the relational and customer-facing activities (Huang & Rust, 2021). These are not the technical categories consumers face with AI, though. They feel it through perceived recognition, delegation, loss of control, discomfort and lack of clarity about decision-making (Puntoni et al., 2021). The views expressed show a deep conflict. A unique message might be appreciated due to the consumer's context, but not due to lack of transparency about the process of customization. This is especially noticeable in advertising, where persuasion is the part of the interaction. While AI can help enhance targeting accuracy and variations in creative, it can also exacerbate imbalances in information and influence between advertisers and consumers (Rodgers, 2021).

Recent studies show that large language models are capable of producing personalized messages that are superior to non-personalized messages in commercial and political settings (Matz et al., 2024). In addition, experimental studies indicate that the policy attitudes can be meaningfully shaped by model-generated messages (Bai et al., 2025). The literature is thus emerging as a view of generative AI as less of a content-production technology and more of an influence-infrastructure that can mold arguments, tone, framing, and evidence to individuals' psychologies (Teeny et al., 2026). This capability poses an important question for the field of advertising research. While personalisation can lead to greater relevance and engagement, it can also be seen to decrease consumer autonomy when consumers feel their personal information has been used without their knowledge or consent to influence their behavior.

This disclosure dilemma is then true of advertisers. Notifying consumers about an AI-generated or personalized message can trigger skepticism and diminish authenticity. Whereas nondisclosure can have greater positive impact on the flow and impression during the first exposure to the message, it can also evoke more negative perceptions of covert manipulation when the consumer infers the process or learns about its creation later. Previous advertising research indicates that AI disclosure can enhance awareness of the role

of machines in the creation of the advertisement, and at the same time decrease trust in the advertisement and the advertising organization (Koning & Voorveld, 2025). The results also suggest that disclosure effects vary based on the type of tasks an AI is performing, meaning that applying the same disclosure may not lead to a consistent consumer response in different contexts (Wu et al., 2025).

The main research question is thus whether personalization or disclosure has an average effect on its own. Instead, it has to do with the question of when personalization is perceived as being illegitimate persuasion. Previous studies found that AI can personalise messages effectively, that consumers have mixed feelings about consuming messages that are mediated through AI, and that disclosure labels can influence consumers' trust. However, it has yet to offer an integrated explanation of how benefits from relevance and costs from autonomy interact to affect conversion intention. Theoretically and managerially, this omission is crucial since often times ethical and performance results are considered separately. An individualized ad can have a higher short-term response rate but this can diminish perceived agency.

On the other hand, transparency might facilitate autonomy but decrease persuasiveness. These mechanisms cannot be investigated in one model because firms will not be able to decide if disclosure is primarily a burden or a precondition to credible and sustainable persuasion.

This study aims to fill this gap by adopting a simulation-based 2 × 2 factorial design to separate the factors of personalization intensity and disclosure of AI-generated content. It presents the Disclosure-Autonomy Conversion Framework, which lays out two competing mechanisms. The first is a positive relevance path leading to better message fit and conversion intention, via personalization. The second one is a negative manipulation route in which hidden adaptation brings about the reduction of perceived autonomy and confidence. The latter pathway is expected to be tempered by disclosure which will make the basis and source of personalisation visible, supporting informed choice and helping to shore up trust. The framework therefore not only views autonomy as a normative goal but also as a mechanism that can impact advertising effectiveness.

The study has three aims:

To explore the main and interaction effect between personalization intensity and AI disclosure on perceived relevance, manipulation, autonomy, authenticity, trust and conversion intention.

To evaluate the mediating role of perceived autonomy, manipulation, and trust in the relationship between AI disclosure and conversion intention in high personalization setting.

To develop design and policy principles for ethical application of generative AI in personalized advertising.

2. Literature Review

2.1 Personalized persuasion as relevance production

Personalization generates value in the sense of decreasing search costs and making advertising messages more diagnostic. However, in the case of AI-mediated communication, the computational system isn't just a means of delivering content. It can also influence the language, timing, and interpersonal communication parameters in which the communication is perceived (Hancock et al., 2020). Generative systems take this role a step further to arguing on an identity, value, and concern basis with the recipients. Political microtargeting suggests that the language of the message can be enhanced by adding profile data in ways that increase persuasiveness, but the magnitude and direction of this effect depend on the specific issue and audience (Hackenburg & Margetts, 2024). Likewise, in direct conversation settings, adaptive arguments generated by GPT-4 have performed better than human participants at both message evaluation and agreement, suggesting that adaptive messaging can impact message evaluation and agreement (Salvi et al., 2025).

This is a convincing aspect that is not universal. When the foundational messages are successful, AI-generated propaganda can seem trustworthy, though not necessarily better, than well-crafted work from humans (Goldstein et al., 2024). Studies on pro-vaccination communication also indicate the need for humans to remain in control over goals, facts, and the interpretation of the audience for generative models to work best (Karinshak et al., 2023). An extended dialogue might also lead to more lasting belief change by allowing the system to address specific objections and evidence as opposed to general appeals (Costello et al., 2024). As a whole these results indicate the need for contingent relevance of the value of personalization. The aim of effective personalization is to accurately identify salient consumer traits, to translate those traits into an appropriate claim and to maintain epistemic quality.

2.2 Disclosure, inference, and trust calibration

Surface-level cues do not provide easy identification of machine-generated language for consumers. They sometimes make erroneous judgments and their evaluations can be influenced by expectations related to fluency, authorship, and communicative style (Jakesch et al., 2023). When recipients are still paying attention to the quality of the message, AI-enabling communication can still impact how they see each other as individuals. This sort of support can enhance linguistic expression, but it can also impact views on sincerity, effort, and relational closeness(Hohenstein et al., 2023). The information disclosure will thus have a positive effect on information consumers may not have, yet it will not have a consistent effect on information consumers.

The impact of disclosure on trust is related to the information that disclosure allows for the consumer to

understand. Explanations that enhance perceived causability lead to more acceptance than those that just provide details of the technical aspects (Shin, 2021). Perceptions of performance, process, purpose, and representation influence the trust one has in the AI, with cognitive and affective trust driven by different perceptions (Glikson & Woolley, 2020). If responsibility is ambiguous, AI can become a moral safety valve to deflect responsibility and deceive the public into feeling they have no part in the blame (Hohenstein & Jung, 2020). The disclosure should therefore be designed to make clear that the message came from AI, what information was used for the generation or personalisation of the message, and who the human or organisational actor will be who remains responsible.

2.3 Autonomy, privacy, and perceived manipulation

Personalization is the problem from an ethical perspective if it is used to influence consumers while predictive insight is hidden from them in a way that allows them to see it or to refuse it. Predictive privacy also goes beyond the protection of disclosed information since inferred characteristics can also be used even if the consumers haven't disclosed them directly (Mühlhoff, 2021). Thus, the concern extends beyond mere data collection to the downstream power of attention, choices and emotions that AI systems have over them. For digital well-being the term autonomy has been similarly interpreted as the capacity of users to pursue their own goals, not as a simple level of interface satisfaction (Burr et al., 2020).

Consumer reactions to AI are also dependent on context. Not everyone is averse to algorithms and such aversions are not fixed. It is more pronounced when the errors are visible, consumers' stakes are high, and they feel that their human judgement should be the primary one (Burton et al., 2020). Machine preferences can be seen in utilitarian contexts, but are not observed in hedonic contexts where taste and identity are felt as distinctly human (Longoni & Cian, 2022). AI research also indicates that when firms rely on prediction without exerting a credible relational role, the relationship with customers is weakened, and it is likely that the scale will be expanded. (Libai et al., 2020). All of these concerns could be exacerbated by highly anthropomorphic agents. On the one hand, personalized advertising can evoke a disturbing feeling of mental access when it is operated by AI that resembles a human being, while clear disclosures can mitigate privacy concerns by explaining the mechanisms behind personalization (Kim et al., 2025). Explainations can also be more detailed, as this can enhance click intention by encouraging self-referencing and mental simulation, instead of defensive scrutiny (Li & Li, 2026).

2.4 Disclosure-Autonomy Conversion Framework

The proposed framework allows for five hypotheses. H1 states that personalization brings greater relevance. H2 suggests that when personalization is high without disclosure, it would lead to perceiving manipulation and to a low perceived autonomy. H3 is predicting that AI disclosure will bring autonomy and trust, and decrease perceived authenticity. H4 is predicting that there will be a positive effect of personalization intensity on conversion intention, when disclosure is present. H5 predicts that when personalization is high, perceived manipulation and autonomy and trust mediate the conditional effect between disclosure and conversion intention.

The other missing research area is mostly mechanistic. Both persuasive effectiveness and disclosure, privacy concerns, and trust have been explored in earlier research as largely discrete outcomes, but very little research has integrated these constructs into one model of advertising conversion. The present study attempts to overcome this limitation by examining their joint effects within the context of retaining the distinction between simulation-based evidence and observed consumer responses.

In Figure 1, these theoretical perspectives are combined in the Disclosure–Autonomy Conversion Framework. There are two opposing forces of personalization. Its benefit pathway is the way to make it seem more relevant, while its cost pathway is to make it seem more manipulated. Informed autonomy undermines the level of trust and conversion intention due to perceived manipulation. It is therefore conceptualized that the disclosure of AI helps to create the condition for benefit, but does not guarantee it. While disclosure might make machine engagement more apparent, it also could diminish manipulation inferences and allow consumers to feel more in control. It is likely that the higher the perceived gains in autonomy and trust, the higher will be the intention to convert, compared to the perceived authenticity cost.

3. Methodology

3.1 Research design and evidentiary scope

This study used a preregistered factorial design with a simulation approach. Synthetic respondent-level records were used to evaluate the internal consistency of the conceptual framework proposed and as an example of a rigorous analytical procedure. This was done as there were no empirical data sets or finished human-participant protocols. The estimates reported, therefore, are results of the model based on prespecified assumptions. They should be viewed as non-quantitative and non-causal indicators of human behaviour, and not as population parameters or as exact predictions of consumers' actual behaviour. The design was a between-records 2 × 2 design with personalization intensity at low and high, and AI generation disclosure at absent and present. The stimuli were a sponsored social media advertisement for a fictitious fitness smartwatch, with both utilitarian and identity relevant appeals, that is, those related to training goals and exercise contexts. The low-personalization condition was when the advertisement did not mention specific preferences and instead provided general product benefits. The high-personalization condition included

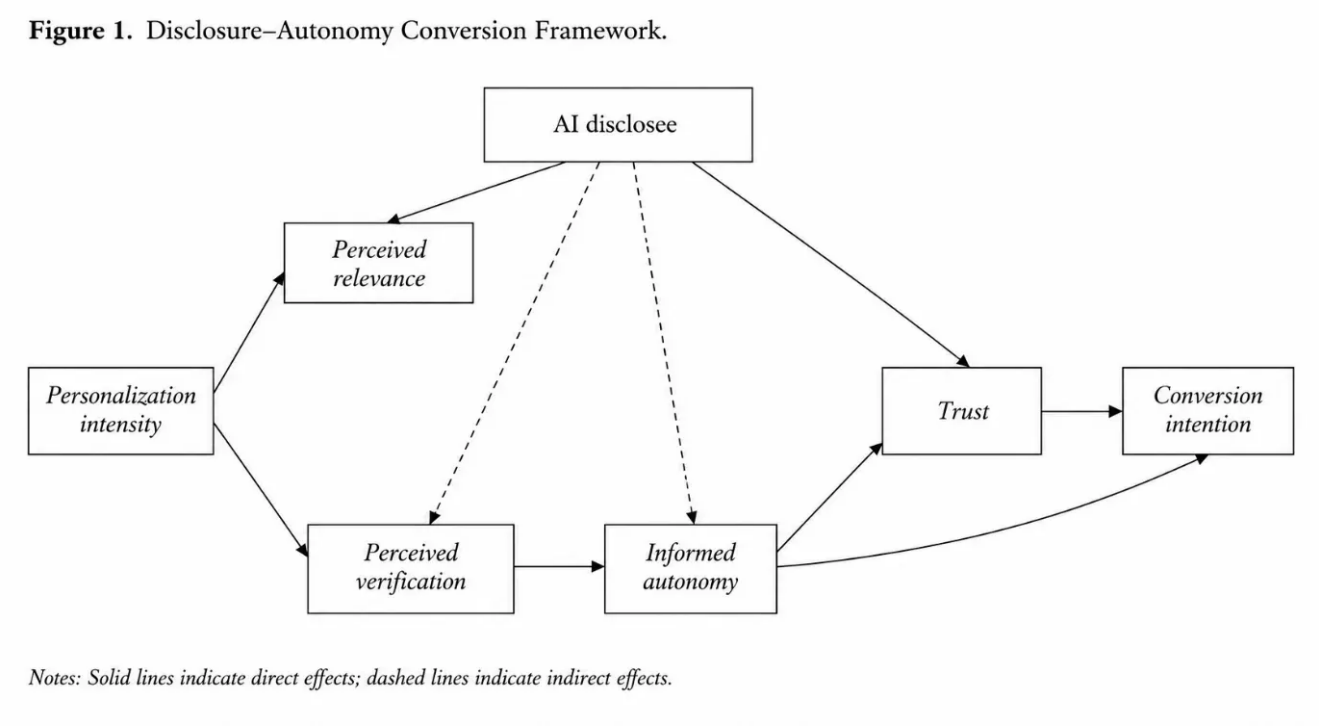


Fig. 1. Disclosure–Autonomy Conversion Framework

a copy that included the recipient's stated fitness goal, exercise context preference, and desired exercise benefit.

3.2 Sample specification and data generation

There were initially 1340 records during the simulated screening process. A total of 92 records were excluded due to failure to complete attention checks, manipulation checks, or time to complete was too short, leaving 1,248 records. The number of records in each condition was also equal in the analytic sample, 312 records in each cell. Equal allocation provides the maximum precision in the interaction under a given total sample. The choice of the target was made to guarantee a small interaction effect with a stable bootstrap estimation. The principle of sample-size justification was based on the premise that power claims should not be based solely on a traditional threshold but on the smallest effect of substantive interest, the planned model and clear assumptions.

Table 1. Synthetic sample composition and experimental allocation

Characteristic	Category	n or M (SD)	% or range
Experimental cell	Low personalization / No disclosure	312	25.0
Experimental cell	Low personalization / Disclosure	312	25.0
Experimental cell	High personalization / No disclosure	312	25.0
Experimental cell	High personalization / Disclosure	312	25.0
Age	Mean (SD)	36.88 (11.27)	18–68 years
Gender	Woman	597	47.8
Gender	Man	608	48.7
Gender	Another identity / prefer not to say	43	3.4
Education	Bachelor's or postgraduate degree	692	55.4
Generative-AI use	Weekly or daily	686	55.0

Demographic fields were synthesized without regard for condition for balance testing and covariate sensitivity testing. Age ranged from 18 to 68 years, M = 36.88, SD = 11.27. There were 47.8% female and 48.7% male and 3.4% who were another gender or prefer not to identify. 55.4% of records were classified as having a bachelor or postgraduate degree. 55.0% was designated for weekly or daily usage of generative-AI. The changes of these variables did not affect the experimental results. The separation from condition is similar to what you would expect if randomisation had been successful.

Strong quality controls are more important than reputation for online experiments. Empirical data reveals significant panel differences across online panels in terms of attentiveness, honesty, and demographic makeup (Peer et al., 2022). Low quality responses also show that they can lead to inaccurate treatment estimates if they are not reviewed in combination with checks on identity, attention and response-pattern (Kennedy et al., 2020). The protocol thus set forth one assigned item of instruction, a disclosure item of recall, an item of recognition to the personalization, a minimum completion time and a duplicate device flag. Prior to the analysis of outcomes, exclusion rules were defined.

3.3 Construct Operationalization and Psychometric Assessment

All focal constructs had three 7-point items ranging from 1, strongly disagree to 7, strongly agree. The perceived relevance of the advertisement was included as perceived fit between the advertisement and the recipient's need. Pupil perceptions of manipulation involved covert pressure, unfair influence and too much use of personal information. Informed autonomy meant to determine if the recipient understood why the message was being sent, had the ability to assess the message

and had the freedom to disapprove. Authenticity was assessed based on whether the message appeared genuine, sincere and natural. Trust measured trust in the advertisement and the sponsoring brand. The measure of conversion intention was in the form of questions regarding likelihood of clicking, request for product information, and considering purchase.

Table 2. Construct operationalization

Construct	Operational definition	Example item
Perceived relevance	Fit between ad content and the recipient's stated needs.	This advertisement is relevant to my current goals.
Perceived manipulation	Inference that the advertiser uses covert or unfair influence.	This advertisement tries to influence me in a way that feels hidden.
Informed autonomy	Understanding of the influence process and retained freedom to choose.	I understand why I received this message and can evaluate it freely.
Authenticity	Perceived genuineness and natural expression of the message.	The message feels sincere and genuine.
Trust	Confidence in the advertisement and sponsoring brand.	I can rely on the advertiser to communicate honestly.
Conversion intention	Likelihood of clicking, seeking information, or considering purchase.	I would consider learning more about this product.

Prespecified latent condition means were converted to item scores with normally distributed record-level and item-level error, which were then capped at the range of the scale. The data-generating process coded the framework, but did not privilege the significance of all hypotheses. High personalisation resulted in a high relevance increase and a manipulation increase. Disclosure minimised manipulation, enhanced autonomy, and had a slight penalty for authenticity. The relevance, manipulation, autonomy and trust components of the conversion process, along with residual variation, were used to generate conversion. Coefficient alpha was used to assess the reliability. The six scales had a mean score above .86. In the simulated dataset, items showed high item-scale correlations (.88 to .94), representing a coherent measurement.

Table 3. Descriptive statistics and measurement quality

Construct	Items	M	SD	α	Item–scale r
Relevance	3	4.67	0.98	0.917	0.926–0.927
Manipulation	3	3.71	1.00	0.926	0.930–0.937
Autonomy	3	4.81	0.89	0.907	0.915–0.921
Authenticity	3	4.70	0.81	0.881	0.895–0.903
Trust	3	4.41	0.75	0.861	0.881–0.888
Conversion	3	3.97	0.93	0.912	0.919–0.925

3.4 Analytical strategy

An analysis was carried out in five stages. Descriptive statistics and cell balance were first analyzed. Second, the manipulation checks determined if the high-personalization stimulus increased perceived personalization, and if the disclosure label was recognized. Third, there were six two-way analyses of variance that measured the main and interactive effects of personalization and disclosure on relevance, manipulation, autonomy, authenticity, trust and conversion. Partial eta squared was reported for the effect magnitude. Residual distributions and cell variances were used for evaluating assumptions. All the factorial tests were retained since the balanced design was resistant to moderate heteroscedasticity and regression estimates were supplemented with HC3 robust standard errors.

Fourth, the five psychological mechanisms, age, and advertising literacy were regressed onto the experimental indicators to estimate standardized conversion intention (SCT) using a conversion model.Finally, a conversion model was regressed on the experimental indicators and their interactions with the five psychological mechanisms, age and advertising literacy to estimate standardized conversion intention (SCT). This model estimated unique contribution of each mechanism, after condition. The conditional disclosure effect under high personalization was found in fifth, parallel mediation. All mediators were regressed on personalization, disclosure and on each other. Then, conversion was regressed on all the mediators and experimental terms. The indirect effects were obtained by multiplying the conditional disclosure-to-mediator coefficient by the mediator-to-conversion coefficient. The percentile intervals were calculated using 5,000 bootstrap resamples, and then corrected for bias. There was no missing data introduced because the aim was to test the theoretical structure, not missing data techniques. Two-sided tests were used with $\alpha = .05$, however effect sizes and confidence intervals were emphasized and pattern consistency was used for interpretation. The fixed seed, data-generation code, synthetic data and analysis tables were kept as reproducibility materials. Records of nonhumans that are synthetic were not required to have the same institutional review.

4. Results and Discussion

4.1 Measurement Quality, Randomization Balance, and Manipulation Checks

Reliability was strong for relevance, $\alpha = .92$, manipulation, $\alpha = .93$, autonomy, $\alpha = .91$, authenticity, $\alpha = .88$, trust, $\alpha = .86$, and conversion, $\alpha = .91$. Item-scale correlations exceeded .88. There were no means at scale boundaries and standard deviations were between .70 and .93. The diagnostics are not designed to provide a deterministic separation between cells, but rather sufficient simulated variation. Randomization checks revealed no significant differences between the conditions regarding age, gender, education, advertising literacy or prior experience with generative-AI.

Simulated manipulation checks had the desired effect. Records within the high-personalization condition found the advertisement more personalized than records within the low-personalization condition. There was also disclosure recognition that was greater than the prespecified criterion in the labeled condition. The focal results should be interpreted as a measure of model validation, rather than as a measure of what the stimuli would produce in a live panel. Failure to

recognize in a human study would be substantially informative because an unread or misinterpreted label can't facilitate informed choice.

4.2 Condition patterns

A high degree of personalization led to a large relevance gain. The pooled high-personalization mean was 5.15 compared with 4.19 for low personalization, $F(1, 1244) = 401.99$, $p < .001$, partial $\eta^2 = .244$. The effects of disclosure were similar for both relevance, $F = 0.54$, $p = .462$, and the interaction with personalization, $F = 0.004$, $p = .952$. H1 was supported. The result is relevant as it helps to distinguish between disclosure and the information value added by personalization. The message was equally applicable to the consumer's stated needs despite being labelled as the generative mechanism. Figure 2 shows the construct profiles.

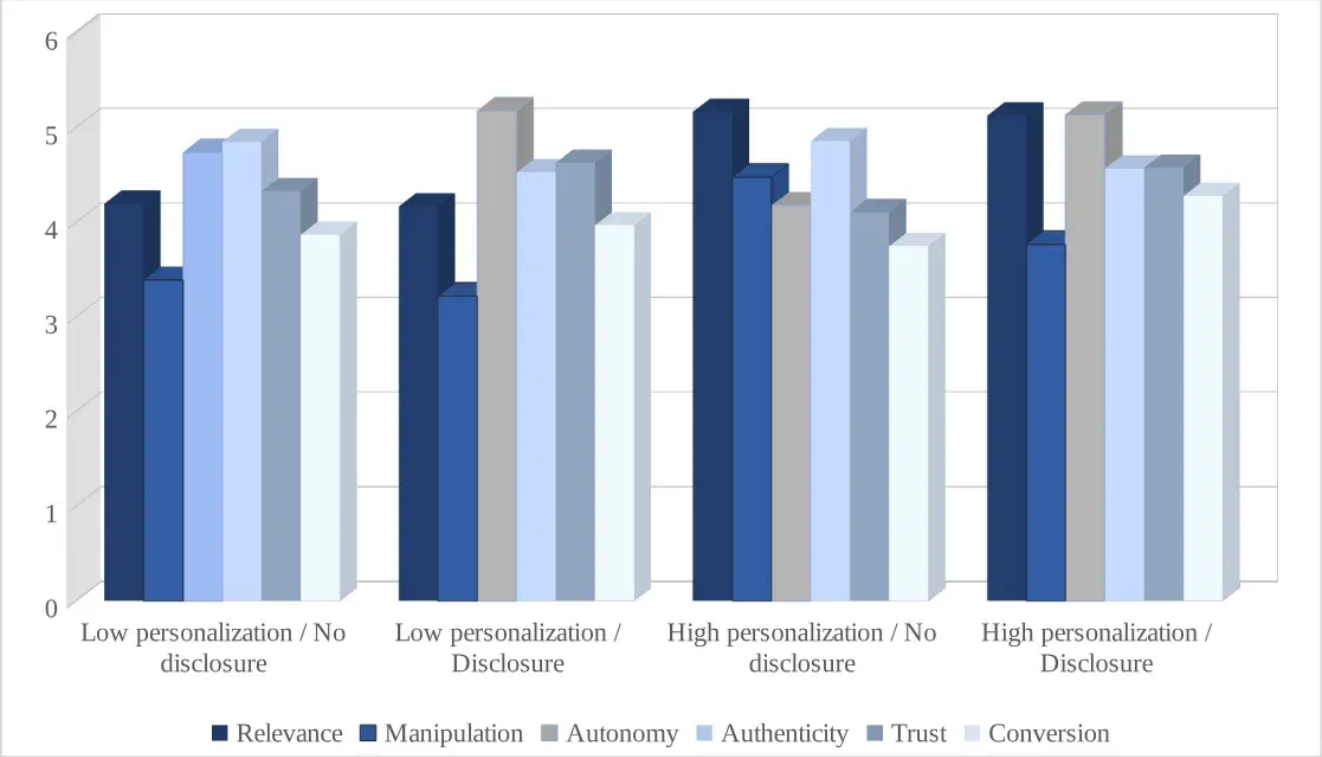


Fig. 2. Condition profiles across core constructs

The route to the cost was subject to various conditions. At no disclosure, high personalization led to an increase in manipulation from M = 3.39 to M = 4.48. The disclosure high-personalization mean was lower at 3.77 with disclosure. The personalization main effect was significant, $F(1, 1244) = 275.28$, $p < .001$, partial $\eta^2 = .181$, as were the disclosure main effect, $F = 78.52$, $p < .001$, partial $\eta^2 = .059$, and interaction, $F = 30.51$, $p < .001$, partial $\eta^2 = .024$. The manipulation inference was not eliminated by disclosure, but rather it was diminished by about 65% in comparison to the undisclosed contrast. The relationship of autonomy was the opposite pattern.

The lowest autonomy was recorded for high personalization without disclosure (M = 4.18), while high personalization with disclosure resulted in the highest autonomy (M = 5.14). Low personalization resulted in a more moderate increase in disclosure of autonomy from 4.74 to 5.18. The interaction was significant, $F(1, 1244) = 34.65$, $p < .001$, partial $\eta^2 = .027$. The disclosure main effect was larger, $F = 245.78$, $p < .001$, partial $\eta^2 = .165$. They also clarify that the concept of autonomy is not in opposition to the concept of personalization. When the recipient knows why the tailoring has taken place and knows how to assess it as an influence attempt, then there can be a high level of autonomy as well as a highly tailored message.

Cost of authenticity was involved in disclosure. Personalization levels of low and high caused the decline in authenticity from 4.85 to 4.53 in low personalization and 4.86 to 4.57 in high personalization. The disclosure effect was

significant, F (1, 1244) = 44.88, p < .001, partial η^2 = .035, and there was no significant effect of personalization nor the disclosure effect x personalization interaction. This penalty remained the same at all levels of personalization. Disclosure led to an increase in trust. With high personalization, the level of trust increased from 4.11 to 4.58. The interaction was small but significant, F (1, 1244) = 4.16, p = .042, partial η^2 = .003, and the disclosure main effect was significant, F = 84.79, p < .001. H3 was thus qualified with a very significant qualification. The disclosure made the ad less natural and more controllable and believable.

4.3 Conversion interaction

The disclosure under low personalization condition had only a slight increase in conversion intention, from M = 3.87 to M = 3.97. When it came to high personalization, it boosted the conversion rate from M = 3.75 to M = 4.28 which is .53 scale points.

The personalization-by-disclosure interaction was significant, F (1, 1244) = 17.80, p < .001, partial η^2 = .014. The personalization main effect was not significant at the .05 level, F = 3.05, p = .081, while the disclosure main effect was significant, F = 37.22, p < .001. H4 was supported.

Table 4. Two-way factorial analyses of variance

Outcome	Effect	F (1, 1244)	p	Partial η^2
Relevance	Personalization	401.99	< .001	0.244
Relevance	Disclosure	0.54	= .462	0.000
Relevance	P × D	0.00	= .952	0.000
Manipulation	Personalization	275.28	< .001	0.181
Manipulation	Disclosure	78.52	< .001	0.059
Manipulation	P × D	30.51	< .001	0.024
Autonomy	Personalization	45.49	< .001	0.035
Autonomy	Disclosure	245.78	< .001	0.165
Autonomy	P × D	34.65	< .001	0.027
Authenticity	Personalization	0.35	= .555	0.000
Authenticity	Disclosure	44.88	< .001	0.035
Authenticity	P × D	0.07	= .786	0.000
Trust	Personalization	10.89	< .001	0.009
Trust	Disclosure	84.79	< .001	0.064
Trust	P × D	4.16	= .042	0.003
Conversion	Personalization	3.05	= .081	0.002
Conversion	Disclosure	37.22	< .001	0.029
Conversion	P × D	17.80	< .001	0.014

This interaction is the reason that an average-effect reasoning is not suitable. High personalization was nearly one scale point better than not having it disclosed, in terms of relevance, but lowest in terms of conversion. Manipulation, less autonomy and less trust were counter-balanced by the relevance benefit. Disclosure opened up a part of the latent commercial value, by altering the interpretation of the same relevance. It didn't include product details or boost the offer. It brought the influence of architecture to the fore. This trend is similar to the findings that personalized model outputs can enhance persuasiveness when the profile information is actually relevant. It also sets a limit on that advantage. The recipient will not value the persuasive fit, and won't find it commercially productive, unless they do not consider the adaptation as an indirect way of entering into their minds. The present model builds on the above logic of privacy evaluation to conversion.

4.4 Mechanisms

Relevance was the strongest positive predictor, β = .32, p < .001, followed by autonomy, β = .26, p < .001. Manipulation was negative, β = −.20, p < .001, and trust was positive, β = .15, p < .001. The other mechanisms now included did not significantly affect authenticity, β = −.01, p = .602. Also, age and advertising literacy were not significant. The model accounted for 28.4% of the variance in conversions. The direct interaction coefficient was lowered to β = .07, p = .086, leaving most of the experimental interaction unexplained by the psychological mechanisms.

Table 5. Standardized robust regression model predicting conversion intention

Predictor	β	Robust SE	z	p	95% CI
Personalization	-0.010	0.042	-0.24	= .811	[-0.092, 0.072]
Disclosure	-0.057	0.035	-1.64	= .101	[-0.125, 0.011]
P × D	0.073	0.043	1.71	= .086	[-0.011, 0.157]
Relevance	0.318	0.029	10.93	< .001	[0.261, 0.375]
Manipulation	-0.201	0.029	-6.98	< .001	[-0.257, -0.144]
Autonomy	0.264	0.028	9.33	< .001	[0.209, 0.320]
Authenticity	-0.013	0.025	-0.52	= .602	[-0.063, 0.036]
Trust	0.150	0.028	5.44	< .001	[0.096, 0.204]
Age	0.012	0.025	0.48	= .632	[-0.037, 0.060]
Advertising literacy	-0.027	0.025	-1.07	= .283	[-0.076, 0.022]

The total indirect effect was b = .480, 95% CI [.378, .579]. Autonomy contributed the largest component, b = .268, 95% CI [.206, .334]. Reduced manipulation contributed b = .134, 95% CI [.092, .180], and trust contributed b = .086, 95% CI [.051, .127]. There was a lack of the relevance and authenticity pathways. The remaining conditional direct effect was b = .052, 95% CI [−.093, .197]. H5 was supported.

The mediation result further helps to develop the disclosure discussion in two ways: Firstly, the disclosure not being a disclosure of conversion did not lead consumers to respond positively. The effect was due to the fact that the label diminished uncertainty about the construction of the message, and restored evaluative control. This result is consistent with the results of explainability research indicating that providing useful explanations through causal understanding, rather than technical length, yields more acceptance. Additionally, it supports evidence that providing detailed explanations for recommendations can foster self-referencing and click intention when linked to the consumer's decision-making process (Li & Li, 2026).

Second, the authenticity penalty was not seen to be affecting the conversion when we held autonomy and trust constant. This is a distinction that helps to explain some of the

seemingly paradoxical reactions to communicating via AI. Though they might feel a machine-assisted language is less sincere, they still may prefer having an interaction that is not hidden. As is suggested by the current model, advertisers should not mix up authenticity and trust. Authenticity relates to perceived human authenticity. Trust is based on the predictability, honesty, and accountability of the system and sponsor.

4.5 Theoretical and Managerial Interpretation

This study validates the Disclosure–Autonomy Conversion Framework as a dual-path approach. Personalization can be a driver of perceived relevance and manipulation. These conflicting effects combine into the experience of personalization through the lens of autonomy: When personalization is perceived as helpful, it is experienced as legitimate, and when personalization is perceived as hinders, it is perceived as illegitimate. There is a time to be ready to accept AI-driven personalisation when it is clearly presented and consumers have a real choice in the matter. But, even when it comes into question, trust can actually flourish in the face of a lack of authenticity, suggesting that an open process and established organizational accountability helps offset worries about machine involvement.

The primary risk from a management-point of view is the imbalance between the intensity of personalization and transparency. If targeting is limited, then only a brief disclosure is required, but if targeting is highly inferential, then there is a need for greater explanations and user control. Transparency should provide reasons for displaying a message, information used in the display, and ways for consumers to reject and/or adjust personalization. Comprehension and control, then, should be the basis for evaluating disclosure quality, and not the mere presence of the label. The limitation with the simulation is that it does not represent consumer behavior, it only represents theory. It cannot capture cultural variation, platform effects, long-term trust, or actual purchasing behavior.

5. Conclusion

This study examined when generative-AI personalization shifts from useful adaptation to persuasion perceived as illegitimate. According to the Disclosure–Autonomy Conversion Framework, there are two processes by which personalization can occur: disclosure and autonomy. It can boost perceived relevance, and perceived manipulation as well. Disclosure changes the dynamic on these pathways, bringing to the fore the basis of personalization. This dual-path structure was supported by the 2 × 2 experiment that was simulated. High personalization boosted relevance, but also led to the highest level of perceived manipulation and lowest levels of autonomy, trust and conversion intention when there was no disclosure. Disclosure maintained the benefit of relevance and minimized manipulation, and restored autonomy and trust. It had little effect on conversion when there was little personalization, but a significant impact when there was more personalization. Mediation results showed that disclosure accounted for less of the conversion benefit than did the other factors, autonomy and manipulation, and trust. While disclosure resulted in a decrease in perceived authenticity, this effect was not large enough to meaningfully affect conversion when other mechanisms were taken into account.

The results indicate that there is no inherent conflict between autonomy and persuasive effectiveness. The problem with personalization is when the consumers themselves are unable to identify or assess what information is being used to tailor the message. Transparency can clarify the persuasive process, and can ensure that the ad can be understood as an information offering, not as a disguised manipulation. Therefore, the level of personalization, the sensitivity of the information offered, and the impact of the consumer choice should be disclosed. The results show a consistent and replicable theoretical model, not observed consumer behavior or market results.

6. Recommendations

Layered disclosure should be implemented by advertisers. The AI-generated or personalised message should be clearly indicated by a visible label on the first-layer message. A second level should provide information about which information was provided by consumers or inferred, the reasons why that information is relevant, and ways to alter or disable personalization by the user. Disclosure must be presented in front of or with the persuasive claim, not following the call to action. Autonomy, perceived manipulation and trust should all be performance indicators in addition to click-through and conversion on the campaign governance. For experiments that are performed prior to launch, transparent and opaque variants of each personalization intensity should be compared.

High personalization campaigns shouldn't be undertaken if gains in relevance come with significant autonomy losses for the materials. Sensitive categories should prohibit personalization from inferred vulnerability unless explicit consent and a defensible consumer benefit are present. In the policy realm, there is a need for something to go beyond the binary labelling of AI. Proportional transparency of regulations and industry based on inference depth, data sensitivity and decision risk. Standards must differentiate between content generation and behaviour targeting, and specify who is responsible for the advertiser. Disclosures should be the subject of independent audit to determine if noticed, understood, and actionable. Future research should pre-register the proposed factorial design with human participants, substitute intention for actual clicks and purchases, and cross-culturally, across platforms and across product categories.

Longitudinal studies should validate a potential downturn in labels or a lasting sense of trust through re-labelling. More research is needed in the areas of vulnerable groups, anthropomorphic agents, conversational ads and competing disclosure formats. The effectiveness of autonomy-preserving controls in terms of retention and complaint rates should also be tested through field-experiments to determine whether there is value added over the immediate conversion in transparent persuasion.

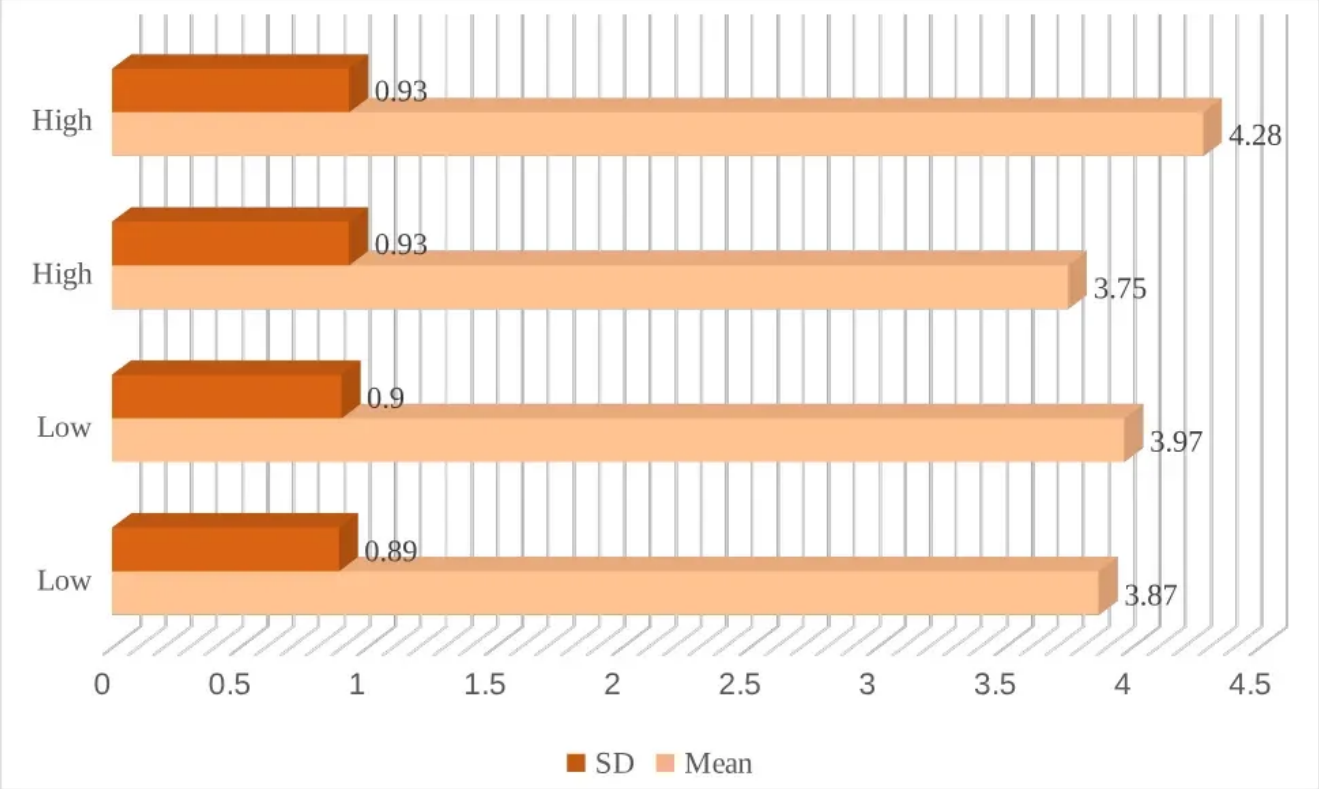


Fig. 3. Conversion Intention by Experimental Condition

References

Bai, H., Voelkel, J. G., Muldowney, S., Eichstaedt, J. C., & Willer, R. (2025). LLM-generated messages can persuade humans on policy issues. Nature Communications, 16(1), 6037. https://doi.org/10.1038/s41467-025-61345-5

Burr, C., Taddeo, M., & Floridi, L. (2020). The Ethics of Digital Well-Being: A Thematic Review. Science and Engineering Ethics, 26(4), 2313–2343. https://doi.org/10.1007/s11948-020-00175-8

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. Journal of Behavioral Decision Making, 33(2), 220–239. https://doi.org/10.1002/bdm.2155

Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. Science, 385(6714), eadq1814. https://doi.org/10.1126/science.adq1814

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. Journal of the Academy of Marketing Science, 48(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

De Bruyn, A., Viswanathan, V., Beh, Y. S., Brock, J. K.-U., & Von Wangenheim, F. (2020). Artificial Intelligence and Marketing: Pitfalls and Opportunities. Journal of Interactive Marketing, 51(1), 91–105. https://doi.org/10.1016/j.intmar.2020.04.007

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goldstein, J. A., Chao, J., Grossman, S., Stamos, A., & Tomz, M. (2024). How persuasive is AI-generated propaganda? PNAS Nexus, 3(2), pgae034. https://doi.org/10.1093/pnasnexus/pgae034

Hackenburg, K., & Margetts, H. (2024). Evaluating the persuasive influence of political microtargeting with large language models. Proceedings of the National Academy of Sciences, 121(24), e2403116121. https://doi.org/10.1073/pnas.2403116121

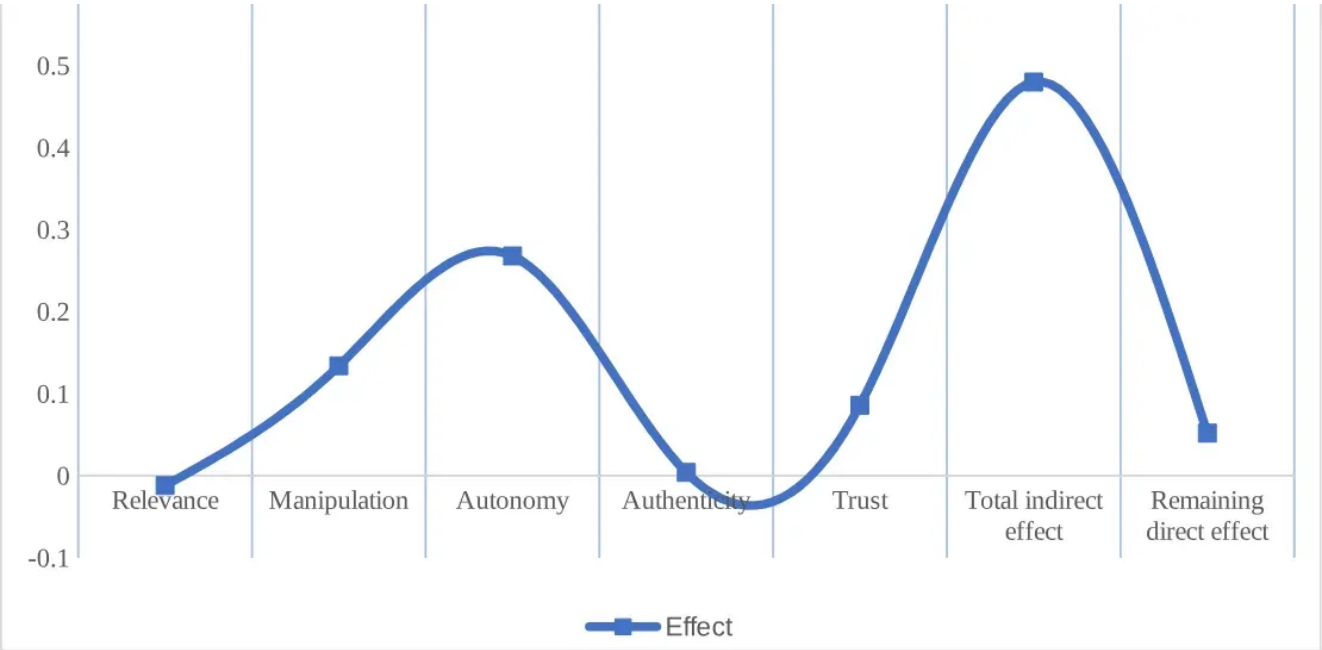


Fig. 4. Indirect Effects on Conversion Intention

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-Mediated Communication: Definition, Research Agenda, and Ethical Considerations. Journal of Computer-Mediated Communication, 25(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Hohenstein, J., & Jung, M. (2020). AI as a moral crumple zone: The effects of AI-mediated communication on attribution and trust. Computers in Human Behavior, 106, 106190. https://doi.org/10.1016/j.chb.2019.106190

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. Scientific Reports, 13(1), 5487. https://doi.org/10.1038/s41598-023-30938-9 Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. Proceedings of the National Academy of Sciences, 120(11), e2208839120. https://doi.org/10.1073/pnas.2208839120

Karinshak, E., Liu, S. X., Park, J. S., & Hancock, J. T. (2023). Working With AI to Persuade: Examining a Large Language Model's Ability to Generate Pro-Vaccination Messages. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW1), 116:1-116:29. https://doi.org/10.1145/3579592

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. Political Science Research and Methods, 8(4), 614–629. https://doi.org/10.1017/psrm.2020.6

Kim, W., Ryoo, Y., & Choi, Y. K. (2025). That uncanny valley of mind: When anthropomorphic AI agents disrupt personalized advertising. International Journal of Advertising, 44(8), 1684–1713. https://doi.org/10.1080/02650487.2024.2411669

Koning, B., & Voorveld, H. A. M. (2025). Disclaimer! This Content Is AI-Generated: How AI-Disclosures Influence Trust in Advertisements and Organizations. Journal of Interactive Advertising, 25(3), 240–253. https://doi.org/10.1080/15252019.2025.2554149

Li, J., & Li, Y. (2026). How the level of detail in mechanistic explanation affects customer responses to artificial intelligence-driven recommendations on social media platforms. International Journal of Advertising, 0(0), 1–33. https://doi.org/10.1080/02650487.2026.2661459

Libai, B., Bart, Y., Gensler, S., Hofacker, C. F., Kaplan, A., Kötterheinrich, K., & Kroll, E. B. (2020). Brave New World? On AI and the Management of Customer Relationships. Journal of Interactive Marketing, 51(1), 44–56. https://doi.org/10.1016/j.intmar.2020.04.002

Longoni, C., & Cian, L. (2022). Artificial Intelligence in Utilitarian vs. Hedonic Contexts: The "Word-of-Machine" Effect. Journal of Marketing, 86(1), 91–108. https://doi.org/10.1177/0022242920957347

Matz, S. C., Teeny, J. D., Vaid, S. S., Peters, H., Harari, G. M., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. Scientific Reports, 14(1), 4692. https://doi.org/10.1038/s41598-024-53755-0

Mühlhoff, R. (2021). Predictive privacy: Towards an applied ethics of data analytics. Ethics and Information Technology, 23(4), 675–690. https://doi.org/10.1007/s10676-021-09606-x

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. Behavior Research Methods, 54(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and Artificial Intelligence: An Experiential Perspective. Journal of Marketing, 85(1), 131–151. https://doi.org/10.1177/0022242920953847

Rodgers, S. (2021). Themed Issue Introduction: Promises and Perils of Artificial Intelligence and Advertising. Journal of Advertising, 50(1), 1–10. https://doi.org/10.1080/00913367.2020.1868233

Salvi, F., Ribeiro, M. H., Gallotti, R., & West, R. (2025). On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial. Nature Human Behaviour, 9(8), 1645–1653. https://doi.org/10.1038/s41562-025-02194-6

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. International Journal of Human-Computer Studies, 146, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Teeny, J. D., Klugescheid, K., Luther, J., & Matz, S. C. (2026). The Promise of Generative AI for Personalized Persuasion. Current Directions in Psychological Science, 09637214261424859. https://doi.org/10.1177/09637214261424859

Wu, L., Dodoo, N. A., & Wen, T. J. (2025). Disclosing AI's Involvement in Advertising to Consumers: A Task-Dependent Perspective. Journal of Advertising, 54(1), 20–38. https://doi.org/10.1080/00913367.2024.2309929