From AI Implicit Associations to Market Outcomes: Evidence from Lab and Field Experiments

Debbie (Yafei) Liang
Haas School of Business
University of California, Berkeley
debbieliang-123@berkeley.edu

Mingduo Zhao
Sauder School of Business
University of British Columbia
mingduo@berkeley.edu

Abstract. Generative AI increasingly helps firms create the product information and advertising that consumers use to make purchase decisions. As firms delegate choices about what to emphasize and whom to depict to these models, demographic associations embedded in AI's outputs may distort consumer demand. We trace this pathway from AI-generated content to market response and test interventions in content generation and disclosure. In a multimodal benchmark adapted from the Implicit Association Test (IAT), our focal vision–language model selects negative descriptors 32.9 percentage points more often for Black than White faces. In an incentive-compatible lab marketplace experiment, holding reviews fixed, our focal model produces less favorable summaries under a Black rather than matched White seller cue. Randomizing these texts independently of the displayed seller photo reduces consumer WTP by 21%. Direct Preference Optimization (DPO) fine-tuning nearly closes both the sentiment and valuation gaps. In a TikTok field experiment, an AI video pipeline predominantly pairs the premium product with male characters. This default is commercially suboptimal: randomized comparisons holding all ad content fixed except character gender show that the matched female version generates 49% more store visits and 42% more purchases per two-second completion. Across both settings, disclosures also alter consumer responses. Together, the studies show how AI demographic associations enter market content, affect demand, and can be addressed through generation and disclosure design.

Keywords: Generative AI, algorithmic bias, advertising effectiveness, information disclosure, field experiment

1 Introduction

Generative AI increasingly creates, organizes, and presents market information. By choosing what to emphasize and whom to depict, these systems shape the signals consumers use to make purchase decisions. This delegation lowers production and search costs but gives models discretion over economically meaningful representations. Although contemporary models generally deny discrimination when asked directly, implicit demographic associations can persist in generated output (Caliskan et al. 2017, Bai et al. 2025). When those associations enter market content, they can change demand without any change in the underlying product.

We call this an *AI implicit association*: a systematic link between a social group and an evaluative or commercial attribute in generated output. Several features of model training and generation may contribute to these associations. Human-produced training data can represent demographic groups unevenly and repeatedly associate them with particular traits, roles, or contexts (Bender et al. 2021, Caliskan et al. 2017). Autoregressive models learn statistical patterns in these data to predict successive tokens, allowing recurring associations to influence descriptions even without an explicit request for demographic judgments. When prompts leave details unspecified, models may fill those gaps with dominant or stereotypical representations (Lahoti et al. 2023). Both initial training and later refinement can shape these patterns. Our study examines how the resulting associations enter generated market information and affect consumer responses, without isolating which training stage produced them.

The multimodal setting is already commercially plausible. Yelp's AI assistant draws on reviews and photos to answer questions about businesses, while eBay's image-based tool converts uploaded product photos into listing content (Yelp 2025, eBay 2023). Such photos can depict merchants, customers, or other people associated with a product or service, allowing demographic cues to enter alongside commercially relevant information. Online marketplace experiments show that apparent seller race, conveyed through a hand in a product photograph, affects offers and auction prices (Doleac and Stein 2013, Ayres et al. 2015). When these images become inputs to generative AI, demographic cues may also influence how written product information is selected and framed. Review summaries provide a setting for examining this possibility. Reviews influence product demand (Chevalier and Mayzlin 2006), and condensing them requires choices about which facts to emphasize and how to present them. We therefore examine whether visual demographic cues change AI summaries of otherwise identical product reviews and whether the resulting content differences affect consumer valuation.

The same concern extends from AI-generated review summaries to video advertising, where demographic associations may shape whom models depict and how they represent products. Generative text-to-video tools

lower production costs and let smaller merchants create ads without large production teams. The Interactive Advertising Bureau (IAB), a U.S. advertising trade association, reports that about half of surveyed video-ad buyers already use generative AI to create or enhance ads (Interactive Advertising Bureau 2025). However, adoption does not ensure reliable output. In a separate survey of 125 U.S. advertising executives, more than 70% reported AI-related incidents, including hallucinations, bias, or off-brand content (Koch and Pellatt 2025). For video advertising, delegating production also gives models discretion over creative choices, including whom to depict and how to represent a product. Gender is economically relevant to these choices because it carries associations with status and economic value (Ridgeway 2001, Williams et al. 2010), and models may reproduce such patterns from online representations (Guilbeault et al. 2025). A depiction that reflects these learned associations may appear plausible, but its commercial effectiveness depends on consumer response. We therefore test whether demographic choices in AI-generated video advertisements align with consumer demand by comparing alternative character genders in otherwise identical ads in a social media field experiment.

Two results connect demographic associations in AI-generated content to consumer demand. First, an Implicit Association Test (IAT) benchmark reveals a pronounced association between Black faces and negative descriptors in our focal vision-language model. A paired online marketplace review-summary benchmark then shows that the model produces less favorable summaries of identical reviews under a Black rather than matched White seller cue. In the lab, these Black-cue summaries reduce willingness to pay (WTP) by 21% relative to White-cue summaries in the reference condition. Second, under gender-neutral prompts, an AI video pipeline more often depicts men in premium product advertisements. In a social media ads campaign, within each product condition, we randomly vary whether the promoted AI-generated advertisement features a male or female character, holding the product, script, and all other ad content fixed. For premium wine, the female-character version generates 49% more store visits and 42% more purchases per unit of delivered attention.

We examine two intervention channels that address the generation and interpretation of market information. Direct Preference Optimization (DPO) fine-tunes the model using preferred and rejected outputs (Rafailov et al. 2023) and narrows both the summary sentiment and valuation gaps. Disclosure addresses consumers' interpretation of generated information. Consumers cannot directly observe model quality and may be unaware of how the quality of generated information is heterogeneous across models and outputs. Disclosure can therefore shape how they interpret that information. This focus on transparency also connects our analysis to policy initiatives in the EU, China, and United States that address transparency around AI-generated content (California Legislature 2024, European Union 2024, National Radio and Television Administration

of China 2025, Tabassi 2023). In the lab, AI-labeled summaries receive higher WTP than human-labeled summaries when representativeness is undisclosed or the summaries are described as unrepresentative. In the field, visible AI disclosure paired with a placebo visual-design statement increases conversion relative to the hidden-marker control. Replacing the placebo visual-design statement with a demographic-association statement decreases conversion.

Our studies contribute to three related areas of research. First, we extend research on AI and algorithmic bias by connecting demographic associations in generated market information to consumer demand. Prior work documents implicit associations in learned representations (Caliskan et al. 2017, Bai et al. 2025) and demographic disparities in algorithmic prediction and allocation (Obermeyer et al. 2019, Lambrecht and Tucker 2019). We examine how these associations enter the content consumers receive and affect valuation and purchases. The lab isolates bias in review summaries by holding products and reviews fixed, while the field examines the commercial consequences of demographic depictions in ads. DPO provides a complementary test of whether mitigation in generated content translates into smaller valuation gaps.

Second, we contribute to research on advertising content and effectiveness by examining demographic choices delegated to generative AI. Advertising research links creative content to consumer engagement (Lee et al. 2018) and demographic representation to advertising effectiveness (Kim et al. 2024). Recent studies extend this literature to AI-generated personalized video advertisements (Kapoor and Kumar 2025) and the selection of effective, brand-aligned advertising visuals (Daviet and Nishimura 2026). We connect these strands by testing whether a model's demographic defaults align with realized consumer demand. Within each product condition, our field experiment randomizes character gender while holding the product and all other ad content fixed. Store visits and purchases then reveal the commercial consequences of reversing the AI model's demographic association.

Third, we contribute to research on disclosure by distinguishing source identity from descriptions of generated information. Economic research establishes that disclosing otherwise unobserved quality can influence consumer demand (Jin and Leslie 2003, Dranove and Jin 2010). With AI, identifying the source can itself change purchases (Luo et al. 2019), and consumer responses to algorithms vary with perceived task suitability (Castelo et al. 2019). What people learn about the model's reasoning also matters, as explanations revealing demographic penalties can reduce willingness to follow AI recommendations (Chan 2025). We extend this distinction to generated market content by varying source labels and statements about representativeness or demographic associations while holding the summary or video fixed. The results show that both source identity and content descriptions affect valuation and conversion.

The remainder of the paper is organized as follows. Section 2 develops the conceptual framework.

Section 3 presents the association benchmark, targeted post-training, and review-summary evaluation. Section 4 describes the lab design and empirical specification, and Section 5 reports the lab results. Section 6 presents the video-generation benchmark, field design, and advertising outcomes. Sections 7 and 8 discuss the implications and conclude.

2 Conceptual framework: information and consumer interpretation

We develop a stylized conceptual framework connecting demographic associations in content generated by AI to consumer demand. Strategic information transmission emphasizes how communication depends on the alignment of economic objectives (Crawford and Sobel 1982), while disclosure research examines how information about otherwise unobserved quality affects consumer choice (Dranove and Jin 2010). Drawing on these insights, we distinguish the representation AI generates from how consumers evaluate it. The framework organizes the experimental comparisons and is not estimated directly.

Let e_j denote product evidence and intended positioning, c_j any demographic cue supplied to the model, and θ the model parameters and generation settings. The AI generated content follows

$$z_j \sim G_\theta(\cdot \mid e_j, c_j). \tag{1}$$

Here, G_θ is the conditional probability distribution over possible texts, images, or videos, and z_j is one realized output. Demographic associations can influence how product evidence is selected and framed or which people and contexts appear in generated content. Post-training changes G_θ. Selecting a matched alternative changes the representation consumers receive while preserving the underlying product.

Let μ_{ij} denote consumer i's expected consumption value for product j before observing the generated content, and let $s_{ij}(z)$ denote the value suggested by that content. Under a simple credibility-weighted valuation rule,

$$v_{ij}(z, m) = (1 - \kappa_i(m))\mu_{ij} + \kappa_i(m)s_{ij}(z) + a_i(z, m), \tag{2}$$

where m describes disclosures about the content's source or selection process. The weight $\kappa_i(m) \in [0, 1]$ captures reliance on the generated information. The term $a_i(z, m)$ captures how presentation and disclosure affect appeal beyond beliefs about consumption value. Unlike informational reliance, presentation appeal need not disappear when consumers assign no credibility to the generated signal. Generated content can therefore shape demand through both the information it conveys and the way it presents a product. Disclosure can change reliance or appeal, so its effect need not have the same sign across disclosure environments.

Interpret $v_{ij}(z, m)$ as willingness to pay in monetary units. For a simple purchase decision at price p_j, a consumer buys when willingness to pay is at least as high as price. Among a fixed population of N_j exposed consumers, expected demand is

$$D_j(z, m; p_j) = N_j \Pr_i\left[v_{ij}(z, m) \geq p_j\right], \tag{3}$$

where $\Pr_i$ denotes the share of consumers satisfying the purchase condition. Expected demand from the generation procedure is therefore

$$\overline{D}_j(\theta, m; p_j) = \mathbb{E}_{z\sim G_\theta(\cdot|e_j,c_j)}\left[D_j(z, m; p_j)\right]. \tag{4}$$

Post-training can change expected demand by changing the distribution of generated content. Selecting a matched alternative instead changes the particular content consumers receive. Content frequently generated by the model need not maximize demand because generation frequency and consumer valuation are different objects.

This distinction guides the studies that follow. We examine how demographic associations enter generated content and affect consumer demand, whether changing that content improves outcomes, and how disclosures change responses to the same content.

3 From diagnostic association to market signal

Before turning to consumer valuation, we locate where an implicit association enters the information pipeline. We then ask whether the association survives in a market-facing review-summary task and whether targeted post-training changes that signal.

3.1 Benchmark as a model-selection diagnostic

We adapt the Implicit Association Test (IAT), introduced in 1998 and now one of psychology's most widely used measures of automatic associations (Greenwald et al. 1998, Morehouse and Banaji 2024). The human Race-IAT compares sorting speed when Black and White faces share response keys with positive versus negative words. One Project Implicit analysis includes approximately 3.3 million U.S. respondents from 2007 to 2020 (Morehouse and Banaji 2024).

Because response times do not have the same psychological interpretation for models, our adaptation measures repeated word selections. We use 12 standardized neutral-face images from Project Implicit's

Race-IAT materials, six Black and six White (Xu et al. 2014). Each face receives 50 independent queries in separate conversations, each requesting three words from a fixed vocabulary of 30 positive and 31 negative terms. Holding the prompt and vocabulary fixed, the Black–White difference in negative-word selection measures the model's implicit association.

Commercial relevance and suitability for controlled post-training motivate our focus on Gemma. In an industry survey, 40% of open-source-model users report having used Gemma (Bisht et al. 2025). Trustpilot, for example, fine-tunes Gemma to process millions of reviews daily (Banfi et al. 2026). We use `Gemma-3-12B-it`, an instruction-tuned vision-language model (VLM) whose image-and-text capabilities support review summarization and whose open weights allow controlled post-training (Google AI for Developers 2025). We test the full summary-to-market chain only with this model.

The word-selection test reveals a 32.9-percentage-point Black–White gap (Figure 1). By contrast, an explicit bias test asks Gemma to agree or disagree with 16 racial or ethnic stereotype statements from DecodingTrust (Wang et al. 2023). Gemma endorses none among its ten classifiable responses. Direct questioning therefore misses associations expressed through word selection.

Seven additional models yield positive implicit-association point estimates on the same 12 faces using ten queries per image, and again on a separate matched 400-image face panel. None endorses a classifiable statement in the explicit test. Appendix A reports the models, protocols, and uncertainty intervals.

For a human reference, we reanalyze Project Implicit Race-IAT data from 198 eligible participants (Xu et al. 2014, Kahveci 2024). Comparing response latencies for five shared negative descriptors yields an 8.6-percentage-point directional gap. The diagnostic motivates our next question: whether the model's association carries into product information and consumer valuation.

3.2 Targeted post-training

We tune Gemma using Direct Preference Optimization (DPO), which learns the preferences encoded in comparisons between preferred and rejected responses to the same prompt (Rafailov et al. 2023). DPO adjusts the model to favor preferred responses relative to rejected alternatives, using the original model as a reference. Unlike conventional reinforcement learning from human feedback, it does not require a separately trained reward model or a reinforcement-learning stage. Here, the preferences favor descriptions grounded in visible attributes over unsupported demographic inferences.

We construct 1,000 preference pairs from 100 FairFace training-split images, balanced between Black and White faces (Kärkkäinen and Joo 2021). For 700 profile comparisons, the preferred response is a visually grounded profile generated by a teacher model, and the rejected response is the original Gemma profile for the

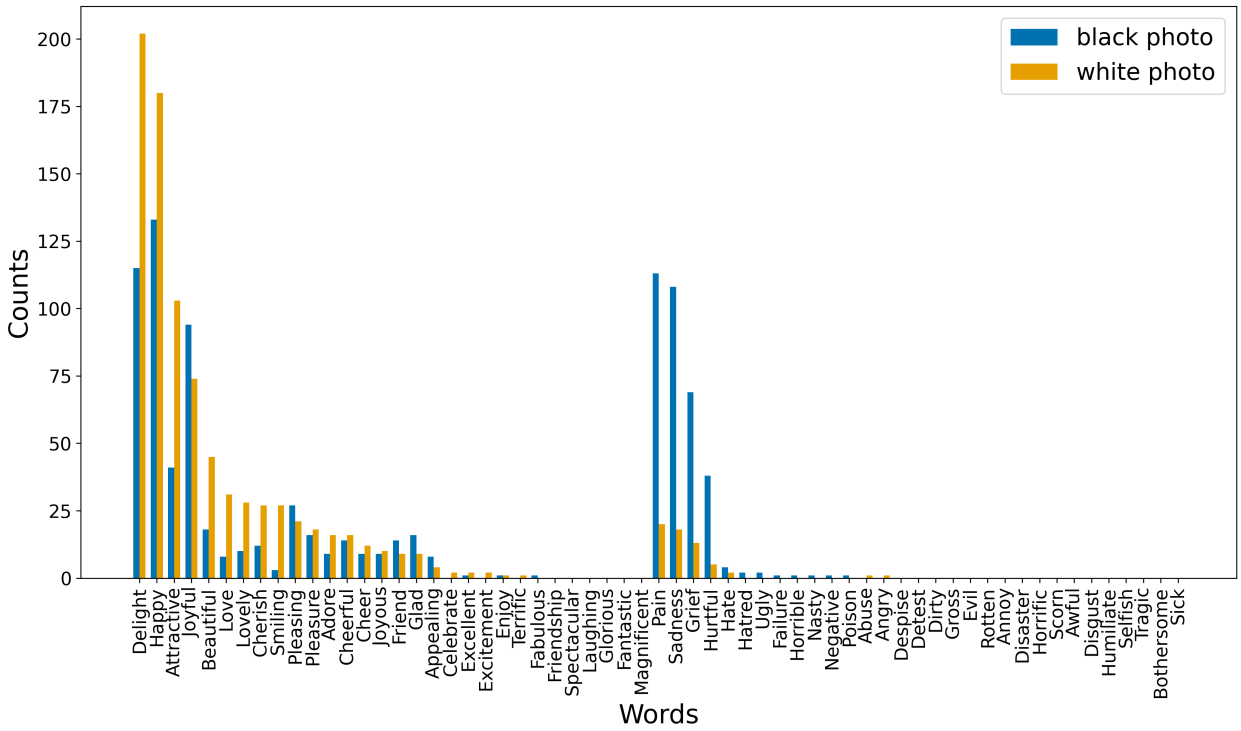


Figure 1: Word selections by baseline Gemma for Black and White faces

Notes: The 12 face stimuli come from Project Implicit's Race-IAT materials. Each face receives 50 independent queries in separate conversations. Our adaptation asks the model to select three descriptors per query from the fixed positive/negative vocabulary, yielding 900 selections per demographic group. Negative descriptors account for 353 selections for Black faces (39.2%) and 57 for White faces (6.3%), a 32.9-percentage-point gap.

same image. The remaining 300 pairs concern word selection, favoring responses based on visible expression rather than race and containing fewer unsupported negative terms. These comparisons teach preferences over how the model describes a person, rather than directly imposing equality in aggregate word-selection rates.

We screen the pairs for identity leakage, invented biography, unsupported traits, invalid formats, and lost expression grounding. Separate vision-language judges evaluate candidates without race labels and with their presentation order reversed. Two research assistants review non-overlapping sets of 500 pairs. We train on 800 pairs from 80 images and reserve 200 pairs from 20 different images for development. Training freezes the original model weights and learns rank-8 low-rank adaptation (LoRA) updates. Neither split includes the diagnostic images or images reserved for subsequent evaluations. This separation allows us to examine whether the learned preferences transfer beyond the training examples.

3.3 Signal distortion in AI-generated review summaries

We next examine whether the diagnostic association appears in AI-generated review summaries. Multimodal commerce systems motivate this task: eBay generates listing content from product images, while Yelp combines review and photo information to assist consumers (eBay 2023, Yelp 2024). Images in these settings can contain demographic cues, raising the question of whether those cues influence how AI presents product information.

Review summarization makes this question commercially relevant. Amazon, Google Maps, and Apple's App Store already use AI to condense customer reviews into concise descriptions (Amazon 2024, Daniel 2024, Apple 2025). Such summaries determine which aspects of customer experience receive emphasis. For our test, we obtain product information and 10–20 customer reviews for each of 839 products from Amazon. Amazon serves as a data source for studying a review-summary task relevant across online marketplaces.

For each product, we generate two summaries from the same reviews, one with a Black seller image and the other with its matched White counterfactual. These diffusion-generated images approximately hold age, expression, pose, and image quality fixed (Howard et al. 2024). The task requests a two-sentence summary of approximately 50 words. Reviews and summary instructions remain identical within each pair. We repeat the paired procedure with DPO-tuned Gemma to examine whether the intervention narrows demographic differences in the generated information.

An independent Claude judge, used in neither generation nor training, scores sentiment, clarity, detail, and personal-experience content on 1–10 scales. Two research assistants separately rate subsets covering 12.5% of the summaries each to validate the scoring task. Table 1 reports the Black-minus-White score gaps. Appendix A.3 provides the scoring rubric, paired means, and confidence intervals. Baseline Gemma's Black-cue summaries are 1.061 points less positive, or 12.6% of the White-cue mean. Under DPO-tuned Gemma, this gap falls to 0.007 points. Tuning also narrows the Black–White gaps in clarity, detail, and personal-experience content.

Table 2 illustrates one product's paired summaries. Identical measuring cup reviews produce a summary emphasizing durability problems under the Black seller cue and a more favorable description under the matched White cue.

Table 1: Model audit and review-summary generation

Object	Contrast	Estimate
Panel A: IAT-style benchmark		
Negative-word selection rate	Baseline Gemma: Black versus White faces	+32.9 pp
Negative-word selection rate	DPO-tuned Gemma: Black versus White faces	≈ 0 pp
Panel B: Review summaries, Black-cue minus White-cue score gap		
Sentiment	Baseline Gemma	-1.061^{***}
Sentiment	DPO-tuned Gemma	-0.007
Clarity	Baseline; tuned	-0.094^{***}; -0.001
Personal experience	Baseline; tuned	-0.421^{***}; -0.077
Detail	Baseline; tuned	-0.112^{***}; 0.010

Notes: Panel A uses repeated word selections from a fixed positive/negative vocabulary. Panel B reports Black-minus-White differences across 839 products scored by a separate Claude judge; higher sentiment is more positive. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Paired summaries generated from identical reviews

	Black seller cue	Matched White seller cue
Generated summary	While many customers appreciate the set's color, stackable design, and ease of cleaning, several reported issues with durability, with multiple pieces breaking within a few months of use and some describing the overall quality as "cheap."	Customers appreciate the durability, ease of cleaning, and attractive color of the measuring cup and spoon set. Many found them to be a perfect addition to their kitchen decor and sturdy enough for daily use, even recommending them as gifts.

4 Lab marketplace experiment

4.1 Design and sample

Having documented demographic differences in generated summaries, we next test whether they change consumer valuation and whether disclosures modify that response. The analysis sample contains 976 eligible survey completions from English-speaking U.S. adults recruited on Prolific. Participants evaluated 16 marketplace-style household products valued at approximately \$6–\$12, presented in randomized order.

For each participant and product, the displayed summary was randomized between the Black-cue and matched White-cue versions generated from the same reviews. The visible seller photograph was randomized independently and need not match the cue used to generate the summary. Participants were not told how summary variation was produced. Independently, the source label stated either "Review summary by an AI assistant" or "Review summary by a Human assistant." The representativeness condition displayed no statement, "The summary fairly represents both positive and negative reviews from customers," or "The summary may emphasize positive or negative customer reviews." Source and representativeness assignments remained fixed for each participant across all 16 products.

The two summary versions, two source labels, and three representativeness conditions form a 2×2×3 design,

implemented separately for baseline and DPO-tuned Gemma. Within-participant summary randomization separates summary effects from stable individual differences in valuation. Keeping disclosure assignments fixed avoids repeated label changes that could draw attention to the manipulation. Independently randomized seller photographs distinguish responses to the visible seller from responses to the generated text.

A comprehension check followed the second product. Of 1,054 completions, 78 failed and were excluded in full, including their responses to the first two products, leaving 976 eligible completions. The data contain no partial submissions. Participants were randomly assigned to treatment arms with preset completion caps, and all arms reached their caps at the same time. The sample comprises 674 baseline-Gemma and 302 DPO-tuned-Gemma completions. The focal summary-plus-photo sample contains 532 baseline and 243 tuned completions. The remaining 201 completions entered no-summary or no-displayed-seller controls. Effect tables use the 775 focal completions. Table 3 summarizes the design and sample.

Table 3: Lab treatment design and sample

Panel A: Summary and disclosure conditions

Summary version	Source label	Representativeness statement
Black-cue summary	AI	None, Representative, Unrepresentative
Black-cue summary	Human	None, Representative, Unrepresentative
White-cue summary	AI	None, Representative, Unrepresentative
White-cue summary	Human	None, Representative, Unrepresentative

Panel B: Sample by generator

	Baseline Gemma	DPO-tuned Gemma
All completions	674	302
Focal completions	532	243
Focal product evaluations	8,512	3,888
Controls: 201 no-summary or no-displayed-seller completions		

Notes: Each row in Panel A includes all three representativeness conditions, yielding 12 summary-by-disclosure cells for each generator. Source and representativeness labels remain fixed within a completion. Summary versions and displayed seller photographs vary independently by product. Each completion covers 16 products.

4.2 Incentive-compatible valuation

Own WTP is elicited with a Becker–DeGroot–Marschak-style multiple price list (Becker et al. 1964). For each product, participants choose between receiving the product and receiving cash at prices $p \in \{0, 1, \ldots, 15\}$. If a price row is selected for payment and the participant chose the product, the payoff is the product plus $15 - p$ dollars. Choosing cash pays \$15 and no product. Participants never pay out of pocket. The interface enforces

a single switch, so own WTP is the highest accepted price:

$$WTP_{ij}^{own} = \max\{p : d_{ijp} = 1\},$$

where $d_{ijp} = 1$ indicates choosing the product at price p.

Each participant has a 1/20 chance of entering the bonus lottery, and one randomly selected price-list row is implemented for a selected participant. Under standard expected-utility assumptions, this payment rule makes truthful switching optimal. Figure 2 illustrates the product display and price list. Participants also report beliefs about others' WTP, WTP_{ij}^{other}, and we calculate $\Delta WTP_{ij} = WTP_{ij}^{other} - WTP_{ij}^{own}$.

4.3 Empirical specification

We estimate the effects of summary assignment and disclosure separately for baseline and DPO-tuned Gemma. Let $S_{ij} = 1$ indicate assignment to the Black-cue rather than matched White-cue summary. Let A_{ij} indicate an AI rather than Human source label, and let R_{ij} and U_{ij} indicate Representative and Unrepresentative statements. For valuation outcome $k \in \{\text{own}, \text{other}\}$, we estimate

$$\begin{aligned} WTP_{ij}^{k} = \gamma_j &+ \beta_S^k S_{ij} + \beta_A^k A_{ij} + \beta_R^k R_{ij} + \beta_U^k U_{ij} \\ &+ \beta_{SA}^k S_{ij} A_{ij} + \beta_{SR}^k S_{ij} R_{ij} + \beta_{SU}^k S_{ij} U_{ij} \\ &+ \beta_{AR}^k A_{ij} R_{ij} + \beta_{AU}^k A_{ij} U_{ij} + \varepsilon_{ij}^k. \end{aligned} \tag{5}$$

Product fixed effects γ_j preserve within-product comparisons. The omitted condition combines a White-cue summary, a Human source label, and no representativeness statement. We also estimate the equation for $\Delta WTP_{ij} = WTP_{ij}^{other} - WTP_{ij}^{own}$.

The coefficient β_S^k measures the summary-cue contrast in the Human-labeled, no-representativeness condition. It is a conditional treatment effect rather than an average across disclosure conditions. Disclosure coefficients and their interactions measure how labels change valuation of the assigned summaries. For example, the AI-minus-Human contrast for White-cue summaries is β_A^k without a representativeness statement and $\beta_A^k + \beta_{AR}^k$ under the Representative statement.

Standard errors are clustered by survey completion. The direct DPO comparison uses a pooled specification interacting generator condition with every term in Equation (5), including product fixed effects. Own-WTP percentage magnitudes use the \$3.68 reference-cell mean. Appendix B reports summary-cue contrasts across disclosure conditions and their common-mix average.



Digital Meat Thermometer

- Waterproof, instant read
- Metal,Plastic
- Grilling and Cooking
- 6.4 + 4.2in



Online marketplace seller

Review summary by an **AI** assistant.
Fairly represent positive and negative reviews from customers. ✅

While many customers praised its speed, ease of use, and value for the price, others experienced issues with accuracy and difficulty recalibrating, with some reporting that the thermometer stopped working after getting wet.

For each row below, choose whether you'd rather receive the **Product** or the **Cash** amount.
Tip: For the best payoff, choose **Product** at all prices **below** your true maximum willingness to pay for this product, and **Cash** at all prices **above** it. (early Cash may forgo a product you want; high Product may leave you with it when you'd prefer cash.)

	Prefer Product	Prefer Cash
$15.00	○	○
$14.00	○	○
$13.00	○	○
$12.00	○	○
$11.00	○	○
$10.00	○	○
$9.00	○	○
$8.00	○	○
$7.00	○	○
$6.00	○	○
$5.00	○	○
$4.00	○	○
$3.00	○	○
$2.00	○	○
$1.00	○	○
$0.00	○	○

Figure 2: Product display and multiple price list in the lab marketplace

Notes: The upper panel shows one realized product display, including a seller photograph, review summary, source label, and representativeness statement. The lower panel shows the product-versus-cash choices used to elicit WTP. Summary versions and photographs vary independently. Disclosure labels follow the arms in Table 3.

5 Lab results

Table 4 reports how the information displayed in the lab marketplace affects valuation (Figure 2). Panel A compares Black-cue and White-cue summaries with a Human source label and no representativeness statement. This comparison averages over the independently randomized seller photographs displayed to consumers. Panel B compares source and representativeness labels for baseline-Gemma White-cue summaries. The three columns report own WTP, beliefs about others' WTP, and the difference between others' and own WTP.

We express own-WTP effects as percentages of the reference-cell mean. This is the observed average own WTP for baseline-Gemma White-cue summaries with a Human source label and no representativeness statement, equal to $3.68.

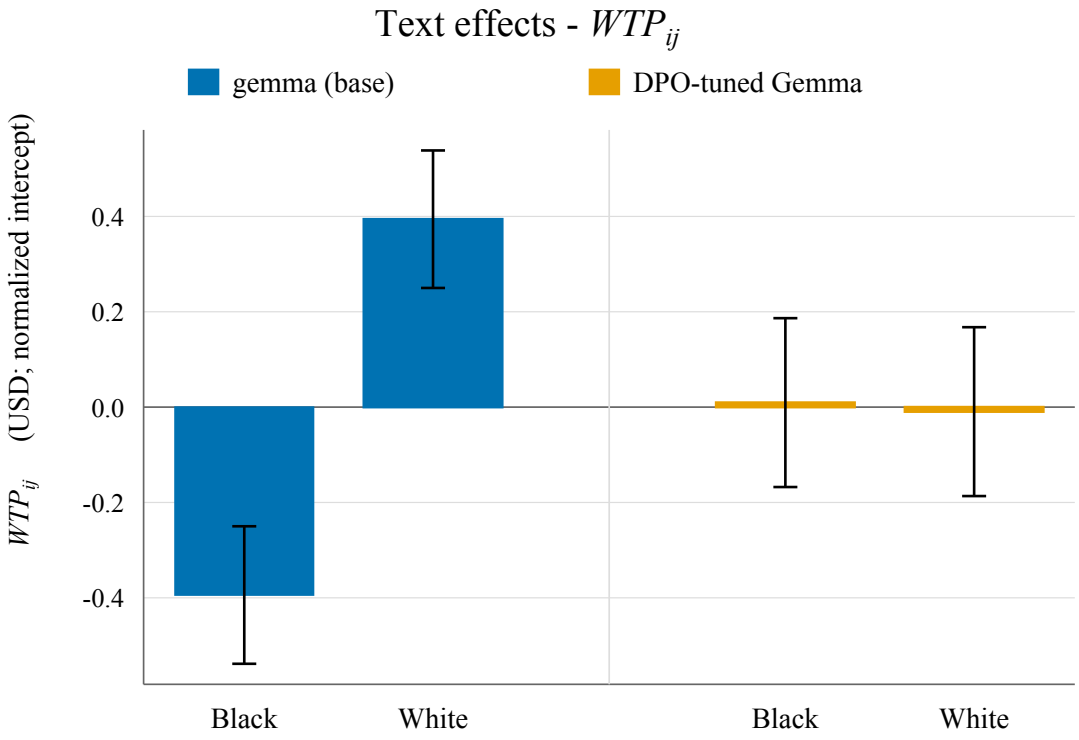


Figure 3: Own-WTP response to Black-cue and White-cue summaries by generator

Notes: Bars show fitted own WTP for Black-cue and White-cue summaries with a Human source label and no representativeness statement, centered at their within-generator midpoint. Whiskers extend one standard error of the corresponding Black-minus-White contrast above and below each bar.

Table 4: Main willingness-to-pay effects

	Own WTP	Others' WTP	Others − Own
Panel A: Black-cue versus White-cue summary contrast			
BlackTwinText, baseline Gemma	−0.788*** (0.144)	−0.924*** (0.143)	−0.136 (0.131)
BlackTwinText, DPO-tuned Gemma	0.019 (0.177)	−0.090 (0.168)	−0.109 (0.134)
Panel B: Disclosure effects, baseline Gemma			
AI-label effect under NoRepDisclosure	1.155*** (0.433)	0.345 (0.366)	−0.810** (0.331)
Representative effect under Human label	1.124*** (0.420)	0.282 (0.353)	−0.842** (0.327)
Unrepresentative effect under Human label	−0.602 (0.415)	−0.859** (0.388)	−0.257 (0.370)
AI–Human gap under Representative	−0.424 (0.339)	0.161 (0.301)	0.585** (0.258)
AI–Human gap under Unrepresentative	2.498*** (0.326)	1.603*** (0.324)	−0.895*** (0.306)
Product fixed effects	Yes	Yes	Yes
Clustered standard errors	Survey completion	Survey completion	Survey completion

Notes: Panel A reports *BlackTwinText* relative to *WhiteTwinText* in the Human-labeled, NoRepDisclosure condition. Panel B reports disclosure contrasts for baseline-Gemma White-cue summaries. Baseline results use 8,512 participant–product observations, and tuned results use 3,888. Estimates are dollars. Standard errors appear in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.1 Generated summaries change WTP

Panel A of Table 4 shows that, under baseline Gemma, the Black-cue summary lowers own WTP by \$0.788 and beliefs about others' WTP by \$0.924. The own-WTP reduction equals 21% of the \$3.68 reference-cell mean. Because products and reviews are held fixed and the displayed summary and seller photograph are independently randomized, this contrast isolates the valuation effect of the generated text.

Under the tuned generator, the corresponding own-WTP estimate is \$0.019 and statistically indistinguishable from zero. The direct comparison yields a significant \$0.807 increase in the Black-minus-White own-WTP contrast (95% CI: [\$0.360, \$1.255], $p = .000425$).[1] Together with the paired summary audit, these estimates show that DPO reduces demographic differences in summary sentiment and downstream valuation.

To summarize effects across disclosure conditions, we average the fitted summary-cue contrasts using the same observed disclosure distribution for both generators. The average contrast is −\$1.247 under baseline

[1]The pooled model interacts generator condition with every term in Equation (5), including product fixed effects, and uses survey-completion-clustered CR1 standard errors with t_{774} inference.

Gemma and −$0.056 after DPO, giving a change of $1.191 (95% CI: [$0.985, $1.397], $p < .001$). Appendix B reports the six disclosure-cell contrasts. Figure 3 visualizes the conditional comparison under a Human label and no representativeness statement.

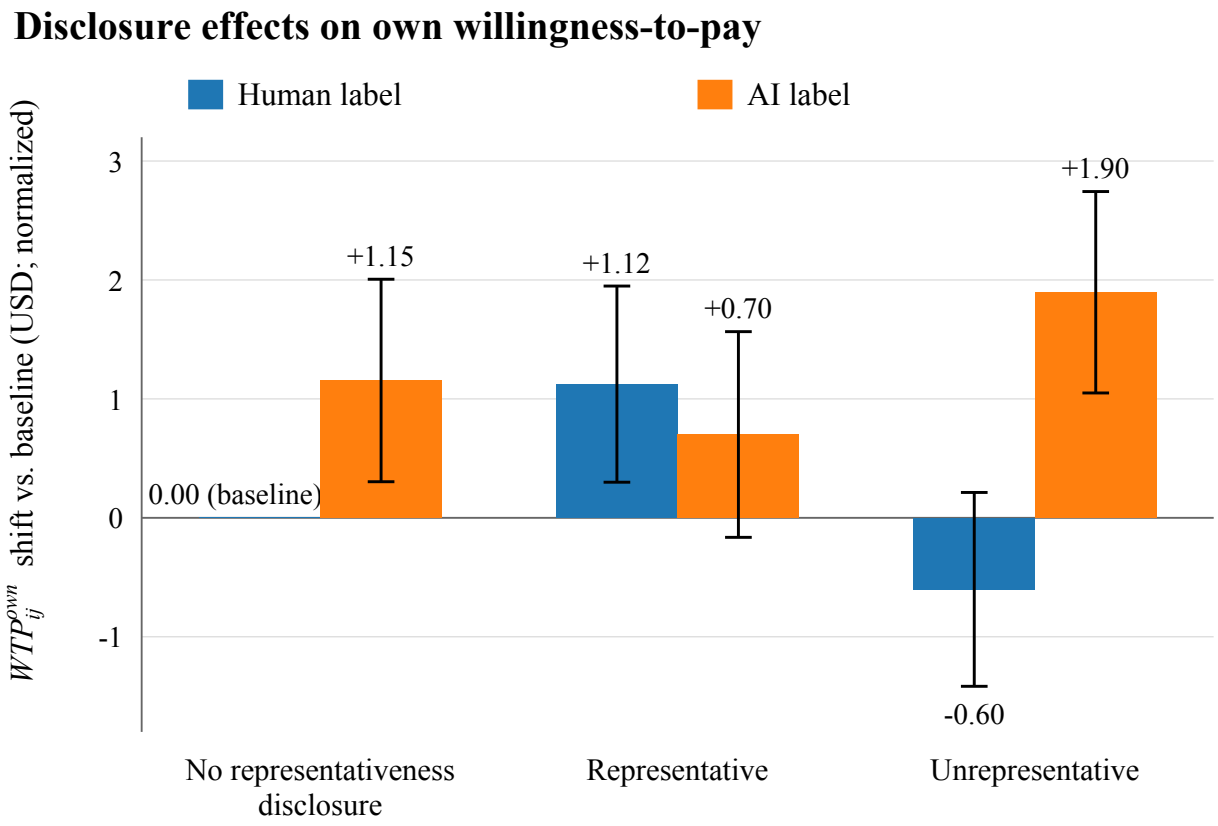


Figure 4: Source and representativeness disclosures jointly shape own WTP

Notes: Bars show own-WTP shifts for baseline-Gemma White-cue summaries relative to the Human-labeled, no-representativeness condition, normalized to zero. The AI-minus-Human differences are $1.15 without a representativeness statement, −$0.42 under Representative, and $2.50 under Unrepresentative. Estimates come from the own-WTP column of Table 4, include product fixed effects, and use 8,512 participant–product observations. Whiskers show 95% confidence intervals using standard errors clustered by survey completion. The reference condition is normalized to zero and has no whisker.

5.2 Source and representativeness are jointly interpreted

Panel B of Table 4 reports disclosure effects for baseline-Gemma White-cue summaries. With no representativeness statement, replacing the Human label with an AI label raises own WTP by $1.155, or 31% of the reference-cell mean. Among Human-labeled summaries, the Representative statement raises WTP by $1.124, or 31%, relative to no statement. The Unrepresentative estimate is −$0.602 and is not statistically

significant. The AI-minus-Human contrast is −$0.424 under the Representative statement and $2.498 under the Unrepresentative statement. The former is not statistically distinguishable from zero, whereas the latter is positive and statistically significant. Thus, consumer responses to the AI source label depend on how the summary's representativeness is described.

Post-survey responses are consistent with this interpretation. Perceived representativeness followed the intended ordering: Representative, no statement, then Unrepresentative. Respondents in the AI × Unrepresentative condition also found the summaries more helpful for determining WTP. Consumers thus appear to distinguish comprehensive coverage from decision usefulness: a selective summary may remain useful for valuation.

Prior work finds greater reliance on algorithmic than human advice (Logg et al. 2019), but also reduced willingness to use algorithms after observing errors (Dietvorst et al. 2015). Our disclosure contrasts show that the value of an AI label depends on how the summary is described as representing the underlying reviews. Source and representativeness statements jointly shape valuation of unchanged content. Figure 4 shows the level shifts and contrasts between AI and Human labels.

6 Field experiment: AI-generated video ads

6.1 Background and motivation

AI-generated content can shape demand from consumers' first impressions of a product. Advertising helps form these impressions, and social media has become a major channel through which firms reach potential customers. In 2025, social media captured 28% of UK digital advertising spending, and video accounted for 59% of spending within that channel (IAB UK 2026). Short-form video platforms such as TikTok integrate advertising into the feeds consumers browse for information and entertainment. WARC, an advertising research and market intelligence provider, reports that TikTok's U.S. advertising revenue is forecast to reach US$14.5 billion in 2026 (Huang 2026). Within these feeds, video ads use people, settings, and actions to communicate product attributes and positioning. These creative choices can shape consumers' first impressions and their willingness to explore a product, consistent with research linking advertising content to engagement and click-through (Lee et al. 2018).

As firms delegate video production to generative AI, models gain discretion over these creative choices. In the IAB's 2025 survey, about half of video-ad buyers already used generative AI to create or enhance ads, and 86% reported using or planning to use it (Interactive Advertising Bureau 2025). Google, for example, offers Veo-powered image-to-video generation in Merchant Center, allowing merchants to turn product images

into videos (Google 2025). Research on personalized AI video advertising documents the potential for lower production costs (Kapoor and Kumar 2025). These capabilities expand the content firms can produce, but selecting effective, brand-compatible creatives remains a separate task (Daviet and Nishimura 2026).

Character gender is an important creative element in video advertising and can help communicate product positioning. Prior research documents stereotypes that associate men more strongly with status and wealth (Ridgeway 2001, Williams et al. 2010). Related gender distortions appear in online visual content and language models, particularly in high-status roles (Guilbeault et al. 2025). When prompts leave character gender unspecified, video models may reproduce these associations in their depictions of different products. Such demographic associations in generated ads need not align with consumer demand. We therefore benchmark the model's character choices and test their commercial consequences by reversing gender in otherwise matched advertisements.

6.2 Field partner and products

Experience goods have important attributes that consumers assess through consumption (Nelson 1970). Before purchase, visual advertising can help form expectations about the product and its consumption context. Research shows that video presentation can influence perceived luxuriousness and purchase responses (Jung and Dubois 2023). Wine therefore provides a relevant setting for examining whether AI-generated visual associations align with consumer demand.

Our field partner is a nationally recognized Chinese spirits producer established in 1952. The company has advertised on social media platforms since 2020. We study two wines from its portfolio, one positioned for everyday consumption and the other as a premium offering. Their bottles are visually similar, while their market positioning differs. Products can differ in aging or quality while retaining similar bottle designs, helping limit differences in product appearance that could otherwise influence the model's scene and character choices. Appendix C provides the partner and product background.

6.3 Generation benchmark and matched creatives

Our benchmark follows the field partner's approach to developing advertising ideas, using an LLM to draft scene descriptions and a video model to visualize them. We use Seedance 2.0, ByteDance's video-generation model, which gained widespread attention after its release and was integrated into TikTok's advertising tools (Reuters 2026, TikTok 2026). Because Seedance generates videos from supplied prompts (ByteDance Seed 2026), we use ChatGPT 5.5 to translate the merchant's advertising objective into detailed scene descriptions.

This mirrors documented creative workflows that use LLMs to draft batches of video prompts (Meares 2025). The approach supports production at scale and reduces reliance on researchers to specify each scene.

ChatGPT 5.5 constructed 50 gender-neutral advertising prompts for each wine, with the objective of increasing store visits and purchases. The prompt instructions distinguish basic from premium price positioning while leaving character gender unspecified. We submitted each prompt to Seedance 2.0 three times and asked it to feature one character. Female characters appeared in 84 of 150 basic-wine outputs (56%) and 18 of 150 premium-wine outputs (12%). The pipeline therefore associates premium positioning more strongly with male characters. Appendix D provides two illustrative prompts. Table 5 summarizes the generation benchmark. The shares describe the retained outputs, with three generations per prompt capturing variation within that prompt.

Table 5: Character gender in the field generation benchmark

Wine position	Gender-neutral prompts	Generated videos	Female characters
Basic	50	150	84 (56%)
Premium	50	150	18 (12%)

Notes: Each prompt produces three videos featuring one character. Counts describe this prompt-generation workflow rather than a matched manipulation of price alone.

We developed the experimental ads in collaboration with the partner's professional marketing department. We interviewed the company's marketing team about its advertising materials to guide creative selection and ensure consistency with its product positioning and brand guidelines. For each wine, Seedance 2.0 generated an initial ad, and we produced a matched version by reversing character gender. We held product display, context, background, character age, clothing color, script, length, aspect ratio, voiceover, music, and call to action fixed within each pair, following work with controllable AI-generated stimuli (Luo and Toubia 2026). Figure 5 illustrates four creatives and distinguishes the pipeline's original pairings from their matched reversals. The partner's marketing department and two undergraduate research assistants checked commercial appropriateness and comparability.

The benchmark identifies the demographic associations in the pipeline's generated ads. Within each product condition, the matched-ad comparison tests whether reversing those associations improves store visits and purchases. This connects the model's generation choices to the merchant's demand objective.

6.4 Field setting, design, and data

We conduct the market test on Douyin, TikTok's Chinese counterpart. TikTok and Douyin integrate video advertising, merchant stores, and checkout within the same platform. Viewers can tap a video's product link,



Figure 5: Four matched video-ad creatives

Notes: Premium-male and basic-female follow the benchmarked AI pairing. Their matched versions reverse character gender while holding all other creative elements fixed.

explore the merchant's store, and complete a purchase without leaving the app. This short purchase funnel connects advertising exposure to store visits and transactions, motivating our choice of platform.[2]

Paid campaigns reach consumers in their feeds and allow us to measure their responses (Donati and Song 2026). We combine matched creatives with controlled targeting, budgets, and promotion windows to test whether the model's demographic choices align with the merchant's conversion objective.

The campaign ran from June 13 through July 6, 2026, on Douyin, TikTok's Chinese counterpart. We posted the ads on a newly created account with negligible organic traffic to concentrate exposure in paid, targeted promotion. The ads linked to the partner's existing flagship online store, providing a real purchase setting and allowing campaign-attributed visits and purchases to be recorded. We targeted audiences across five Chinese macro-areas constructed to balance population and gross domestic product. The campaign followed a 24-day schedule crossing product position, character gender, disclosure, and audience targeting. We randomly assigned treatment combinations to campaign days and ran each combination across all five macro-areas. The platform's targeting tools delivered the assigned male- or female-audience condition. Each ad was made private after the promotion to prevent further viewing or store entry through that video.

The platform integrates video discovery, the shopping-cart link, the merchant store, checkout, and payment. While browsing the feed, viewers can tap the ad's on-screen shopping-cart link to enter the store and proceed

[2]For an overview of TikTok's integrated shopping features, see `https://ads.tiktok.com/business/library/Welcome_To_TikTok_Shop_One_Pager.pdf`.

directly to checkout using credit cards or Alipay and WeChat Pay, widely used Chinese digital wallets. Online payment services reached 90.3% of China's internet users by December 2025 (China Internet Network Information Center 2026). This integrated purchase funnel keeps transaction friction low and motivated our choice of platform. A store visit is a click-through from the store entrance on the video screen, and a purchase is a transaction recorded in the campaign account's transaction history. The platform records two-second completions and other delivery outcomes by day and macro-area. Two-second completions count views lasting at least two seconds, providing a platform measure of initial audience impression and the denominator for our conversion rates (TikTok 2025). The partner company directly provides daily store-visit and purchase totals attributed to the campaign account.[3]

The design crossed two price positions, two character genders, and three disclosure cells. The control disclosure arm retained a minimal-sized AI-origin marker hidden behind an interface button. It was not visible during playback and existed to satisfy the platform's AI-disclosure requirement. The visible AI-plus-placebo disclosure reads "AI-generated; reflects AI's visual design choice." The visible AI-plus-association disclosure reads "AI-generated; reflects AI's typical demographic associations." The two visible messages remained on screen throughout each video ad and were matched in text length, text size, font, and placement. Comparing these two arms measures how the demographic-association statement changes consumer responses relative to the visual-design statement. Consumers could click directly from the video-ad screen to the store. Table 6 summarizes the treatment arms.

Table 6: Field treatment arms

Wine	Character	Disclosure
Basic	Male	Invisible AI marker
Basic	Male	AI + visual design choice (placebo)
Basic	Male	AI + demographic association
Basic	Female	Invisible AI marker
Basic	Female	AI + visual design choice (placebo)
Basic	Female	AI + demographic association
Premium	Male	Invisible AI marker
Premium	Male	AI + visual design choice (placebo)
Premium	Male	AI + demographic association
Premium	Female	Invisible AI marker
Premium	Female	AI + visual design choice (placebo)
Premium	Female	AI + demographic association

Notes: Each row includes one male-targeted day and one female-targeted day, each delivered across five macro-areas. The full schedule therefore contains 24 days and 120 delivery cells. Assigned budgets are CNY 68 per cell.

Assigned budgets were equal across delivery cells, and campaigns optimized for reach rather than

[3]Official Douyin merchant guidance (in Chinese): `https://school.jinritemai.com/doudian/wap/article/106202`.

engagement. Separate campaigns limited budget reallocation across treatment arms. We saturated each targeted audience by reaching nearly all eligible users, with exposure limited to one impression per user. These settings were intended to reduce algorithmic differences in delivery across conditions (Donati and Song 2026, Burtch et al. 2025). Prices, inventory, shipping, promotions, and landing pages were identical within each matched comparison. The focal outcomes are store visits and purchases.

The delivery data contain 120 treatment-day × macro-area cells and 354,148 two-second completions. Appendix E separately analyzes the unsplit daily totals. Scaling actions by two-second completions describes conversion per unit of delivered attention. Let H_d, F_d, A_d, D_d, and M_d indicate premium wine, female character, visible AI disclosure, demographic-association disclosure, and male targeting. For outcome Y_{dm} in treatment day d and macro-area m, the field regression is

$$\begin{aligned} Y_{dm} = \alpha + \beta_1 H_d + \beta_2 F_d + \beta_3 (H_d \times F_d) \\ + \delta_1 A_d + \delta_2 D_d + \eta M_d + \gamma_m + \varepsilon_{dm}. \end{aligned} \tag{6}$$

Macro-area fixed effects are γ_m. Standard errors are clustered by the 24 treatment days because treatments are assigned by day. We use CR1 covariance with a finite-sample correction and t_{23} inference. The female-character effect is β_2 for basic wine and $\beta_2 + \beta_3$ for premium wine. The coefficient δ_1 compares AI plus placebo disclosure with the hidden disclosure control. The coefficient δ_2 compares AI plus association disclosure with AI plus placebo disclosure, replacing the placebo visual-design statement with the demographic-association statement.

6.5 Field results

Tables 7 and 8 report separate regressions using the same delivery cells and controls. Table 7 reports effects on total store visits and purchases in standard deviations. Table 8 reports effects on actions per two-second completion in percentage points. For basic wine, the female-character store-visit contrast is 0.694 standard deviations in Table 7 and 0.249 percentage points in Table 8. Relative to their respective comparison-group means, these estimates correspond to increases of 23% and 16%. The first specification measures total actions, while the second measures actions per unit of delivered attention.

Figure 6 and Table 8 report the effects of character gender in matched advertisements. AI video generation enables us to change only character gender while holding all other creative elements fixed, including product display, setting, character age, clothing color, music, and video length. Within each product condition, we randomly assigned the male- or female-character version to campaign days. For basic wine, the female-

Table 7: Pooled field regressions across the conversion funnel

	2-second completions	Engagement cost	5-second completion rate	Store visits (normalized)	Purchases (normalized)
Premium wine (β_1)	161.233*	-1.401**	-0.017	0.080***	0.053
	(79.465)	(0.617)	(0.019)	(0.012)	(0.047)
	+6%	-6%	-5%	+3%	+2%
Female character (β_2)	147.500	-1.342*	0.053**	0.694***	0.578***
	(93.882)	(0.730)	(0.021)	(0.013)	(0.053)
	+5%	-6%	+15%	+23%	+22%
Premium × female (β_3)	-100.933	1.050	-0.042*	0.883***	0.595***
	(119.005)	(0.890)	(0.022)	(0.019)	(0.105)
	—	—	—	—	—
AI disclosure (δ_1)	187.550***	-1.713***	0.019	1.091***	0.709***
	(65.378)	(0.526)	(0.015)	(0.012)	(0.069)
	+7%	-7%	+5%	+35%	+26%
Demographic association	164.025**	-1.148**	0.003	-0.553***	-0.315***
statement (δ_2)	(75.470)	(0.543)	(0.010)	(0.011)	(0.069)
	+6%	-5%	+1%	-13%	-9%
Male targeting (η)	107.167*	-0.724	0.038***	0.751***	0.560***
	(59.503)	(0.445)	(0.011)	(0.009)	(0.052)
	+4%	-3%	+11%	+23%	+20%
Treatment-day × macro-area cells	120	120	120	120	120
R^2	0.317	0.333	0.544	0.789	0.436

Notes: Cells show coefficient, day-clustered standard error in parentheses, and relative effects below the coefficients. Inference uses 24 treatment-day clusters and t_{23}. All regressions include macro-area fixed effects; there are 354,148 total two-second completions. Interaction relative effects are reported as contrasts in Table 8. Appendix E reports unsplit-day checks. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

character ad increases store visits by 16% and purchases by 14% per two-second completion relative to its matched male version. For premium wine, the corresponding gains are 49% and 42%. Although the generation benchmark predominantly pairs premium wine with male characters, the matched female version performs better. The model's prevailing demographic association therefore does not maximize conversion among the tested alternatives.

Figure 7 and Table 8 show that consumer responses depend on what the disclosure communicates. The visible AI-plus-placebo statement raises store visits by 26% and purchases by 18% relative to the hidden-marker control. This positive response parallels the AI-label premium in the lab's no-representativeness condition. Replacing the visual-design statement with the demographic association statement reduces store visits by 18% and purchases by 14% relative to the AI-plus-placebo condition. The combined effects relative to control are 4% and 1%. This decline is consistent with aversion to demographic associations made salient by disclosure, paralleling evidence that explanations revealing demographic penalties can reduce willingness to follow AI recommendations (Chan 2025). The campaign does not separately identify fairness concerns, perceived quality, and persuasion. The premium-wine female-character advantage and the direction of both

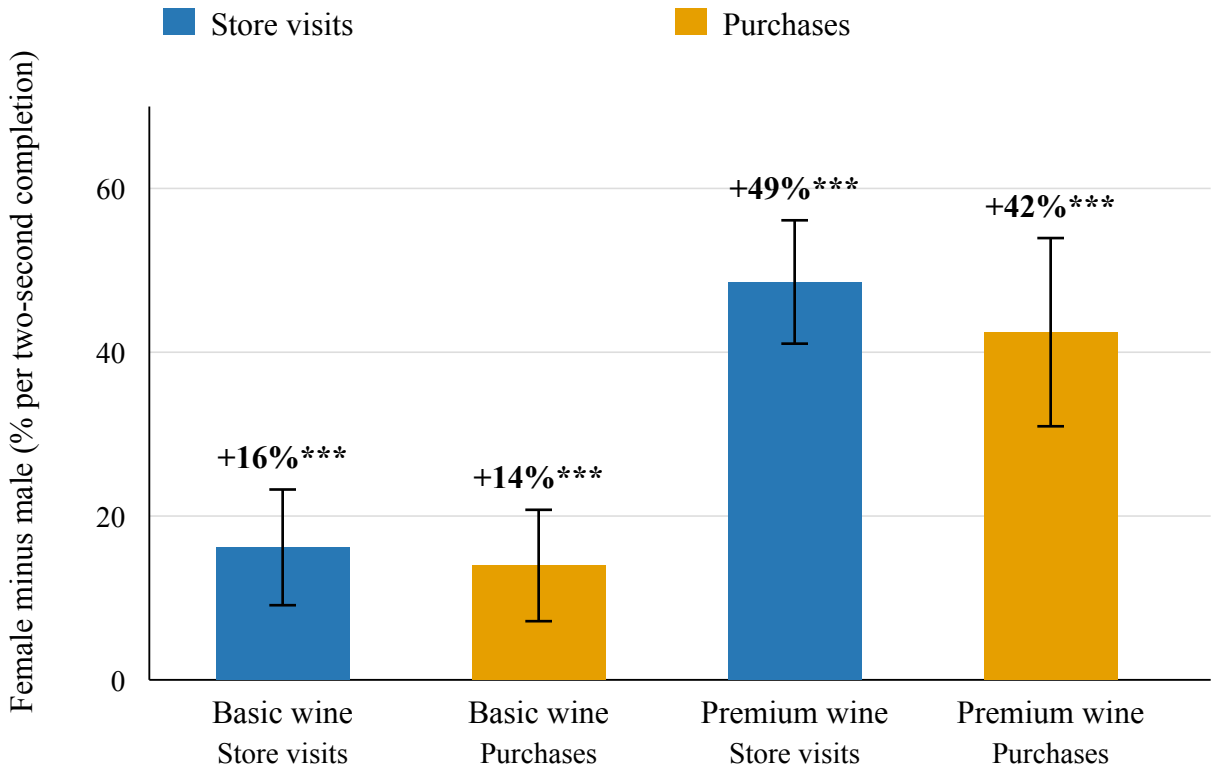


Figure 6: The AI gender–price pairing is not merchant-optimal

Notes: Bars are female-minus-male relative effects; whiskers are 95% confidence intervals. OLS uses 120 treatment-day × macro-area cells, macro-area fixed effects, and standard errors clustered by 24 treatment days (t_{23} inference). Exact values appear in Table 8.

disclosure effects persist in robustness checks using daily totals, calendar controls, and single-day omissions (Appendix E).

The disclosure results also show why delivered attention and conversion should be evaluated separately. The demographic association statement increases two-second completions by 6%, but store visits and purchases per 2s-completion fall. More delivered attention therefore need not translate into better conversion. This pattern highlights the importance of evaluating the advertiser's intended outcomes alongside attention and engagement (Lee et al. 2018, Donati and Song 2026). Testing matched character alternatives assesses the commercial consequences of AI's demographic choices, while the disclosure comparison measures how different descriptions of the same video change consumer responses. Appendix E.5 reports additional outcomes for five-second completion, sharing, and new followers.

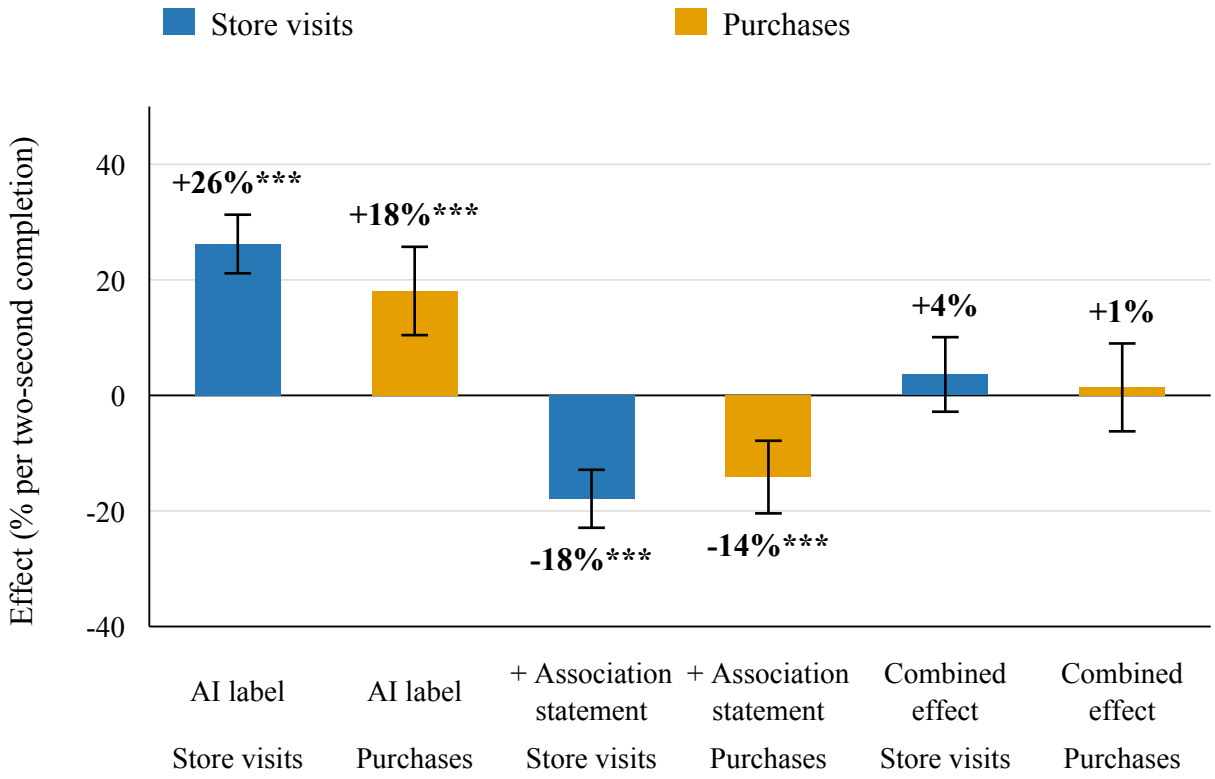


Figure 7: The demographic-association disclosure offsets most of the AI-plus-placebo gain

Notes: Bars are relative effects; whiskers are 95% confidence intervals. OLS uses 120 treatment-day × macro-area cells, macro-area fixed effects, and standard errors clustered by 24 treatment days (t_{23} inference). Exact values appear in Table 8.

7 Managerial and Policy Implications

Together, the studies show that delegating content production to generative AI also delegates judgments about product evidence and demographic representation. Three implications follow for firms and policymakers.

First, platforms and advertisers should audit generated content for demographic distortions in product information and test whether alternative demographic depictions improve advertising effectiveness. Seller cues can change the presentation of identical product evidence and create valuation gaps. In advertising, demographic associations can systematically steer AI toward particular pairings of products and characters, which may result in ads that convert less effectively than alternative creatives. Firms should therefore evaluate demographic associations through their effects on consumer valuation and purchases, alongside benchmark bias scores.

Second, firms can address demographic distortions in product information and ineffective advertising choices through targeted model tuning and tests of alternative content. DPO narrows the sentiment and valuation differences between summaries generated under Black and White seller cues, demonstrating the

Table 8: Focal field contrasts and attention outcomes

Panel A: Contrasts plotted in Figures 6 and 7						
	Store visits per 2s-completion			Purchases per 2s-completion		
Contrast	pp	SE	Relative	pp	SE	Relative
Female minus male, basic wine (β_2)	0.249***	0.052	+16%	0.025***	0.006	+14%
Female minus male, premium wine ($\beta_2 + \beta_3$)	0.722***	0.054	+49%	0.072***	0.009	+42%
Position × character (β_3)	0.473***	0.075	+32%	0.047***	0.011	+28%
AI + placebo vs. control (δ_1)	0.418***	0.039	+26%	0.034***	0.007	+18%
AI + demographic association statement vs. AI + placebo (δ_2)	-0.360***	0.049	-18%	-0.031***	0.007	-14%
AI + demographic association vs. control ($\delta_1 + \delta_2$)	0.058	0.050	+4%	0.003	0.007	+1%
Male- vs. female-targeted days (η)	0.299***	0.038	+19%	0.030***	0.006	+16%
Panel B: Attention outcomes						
	2-second completions			5-second completion rate		
Contrast	Coef.	SE	Relative	pp	SE	Relative
AI + placebo vs. control (δ_1)	187.550***	65.378	+7%	1.887	1.500	+5%
AI + association statement vs. AI + placebo (δ_2)	164.025**	75.470	+6%	0.255	0.951	+1%

Notes: Both visible arms disclose AI origin, accompanied by either a visual-design statement (placebo) or a demographic association statement. Control retains the hidden AI-origin marker. "pp" denotes percentage points. Appendix E reports unsplit-day checks. OLS uses 120 treatment-day × macro-area cells, macro-area fixed effects, and standard errors clustered by 24 treatment days (t_{23} inference). $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

potential of targeted post-training. For advertisers, AI video generation makes it practical to test alternative character choices while holding other creative elements fixed. Managers should therefore assess whether these interventions reduce demographic distortions in product information and improve advertising conversion.

Finally, firms and policymakers should complement improvements in generated content with disclosures that help consumers evaluate it. Source labels and descriptions of representativeness or demographic associations change responses even when the underlying content remains fixed. An AI-origin label identifies the source but does not convey how quality varies across models and outputs. These findings support mandatory, visible AI-origin disclosure complemented by substantiated information about content quality. Model tuning and creative selection shape the content consumers receive, while disclosure helps them decide how much to rely on it.

8 Conclusion

This paper shows how demographic associations in generative AI can shape market information and affect consumer demand. Holding products and reviews fixed, Black seller cues produce less favorable summaries than matched White cues. Randomized exposure to these summaries reduces willingness to pay by 21% in

the reference disclosure condition. DPO fine-tuning substantially narrows both the demographic difference in summary sentiment and the resulting valuation gap. These findings connect algorithmic bias to information intermediation and the established literature on reviews and demand. Unequal treatment of identical product evidence can create economic disadvantages for particular seller groups, making demographic equity a concern in how market information is produced. As generative AI increasingly mediates commerce, algorithmic fairness encompasses the information models create and its economic consequences, alongside established concerns about prediction and allocation.

The field experiment extends this analysis from summarizing product evidence to creating advertising. The AI video-generation pipeline predominantly pairs premium wine with male characters, yet randomized comparisons of otherwise matched advertisements show that the female version generates 49% more store visits and 42% more purchases per two-second completion. This finding extends the literature on advertising content and effectiveness to demographic choices delegated to generative AI. The association reproduced by the model favors a character choice that generates fewer purchases than its matched alternative. Across both settings, AI changes demand through how products are presented, without changing the products themselves.

The disclosure findings connect AI-generated content to economic theories of disclosure and information design. Source labels and descriptions of representativeness or demographic associations change valuation and conversion even when the summary or video remains fixed. Our framework separates the information and demographic representations supplied by AI from how disclosures affect consumers' beliefs and responses to that content. Generative AI thus influences demand through both the content it produces and what consumers learn about how that content was produced. The broader economic questions concern which product evidence and representations reach consumers, how consumers judge their reliability, and how those judgments affect purchases and demographic disparities in market outcomes.

References

Amazon (2024) How amazon's AI-generated review highlights help you make better shopping decisions. URL https://www.aboutamazon.com/news/retail/amazon-ai-generated-review-highlights.

Apple (2025) What's new in App Store Connect. URL https://developer.apple.com/videos/play/wwdc2025/328/.

Ayres I, Banaji M, Jolls C (2015) Race effects on eBay. *The RAND Journal of Economics* 46(4):891–917, URL http://dx.doi.org/10.1111/1756-2171.12115.

Bai X, Wang A, Sucholutsky I, Griffiths TL (2025) Explicitly unbiased large language models still form biased

associations. *Proceedings of the National Academy of Sciences* 122(8), URL http://dx.doi.org/10.1073/pnas.2416228122.

Banfi D, Nurkas A, Ramasubramanian S, Stanek K, Hjertén MC (2026) How Trustpilot built a real-time architecture for data enrichment using Gemma. URL https://cloud.google.com/blog/topics/customers/how-trustpilot-built-a-real-time-architecture-for-data-enrichment-using-gemma.

Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. *Behavioral Science* 9(3):226–232, URL http://dx.doi.org/10.1002/bs.3830090304.

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623 (Association for Computing Machinery), URL http://dx.doi.org/10.1145/3442188.3445922.

Bisht A, Yee L, Roberts R, Presten B, Ottenbreit K (2025) Open source technology in the age of AI. URL https://www.mckinsey.com/capabilities/quantumblack/our-insights/open-source-technology-in-the-age-of-ai.

Burtch G, Moakler R, Gordon BR, Zhang P, Hill S (2025) Characterizing and minimizing divergent delivery in Meta advertising experiments. URL http://dx.doi.org/10.48550/arXiv.2508.21251.

ByteDance Seed (2026) Seedance 2.0. URL https://seed.bytedance.com/en/seedance2_0.

California Legislature (2024) California AI transparency act, senate bill 942, chapter 291. URL https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB942, operative January 1, 2026.

Caliskan A, Bryson JJ, Narayanan A (2017) Semantics derived automatically from language corpora contain human-like biases. *Science* 356(6334):183–186, URL http://dx.doi.org/10.1126/science.aal4230.

Castelo N, Bos MW, Lehmann DR (2019) Task-dependent algorithm aversion. *Journal of Marketing Research* 56(5):809–825, URL http://dx.doi.org/10.1177/0022243719851788.

Chan A (2025) Preference for explanations: Case of explainable AI. Working Paper 26-028, Harvard Business School, URL https://www.hbs.edu/ris/Publication%20Files/26-028_939a174d-891d-472d-9d4a-62d6806cbde4.pdf.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43(3):345–354, URL http://dx.doi.org/10.1509/jmkr.43.3.345.

China Internet Network Information Center (2026) The 57th statistical report on China's internet development. Technical report, China Internet Network Information Center, URL https://www.cnnic.net.cn/NMediaFile/2026/0304/MAIN1772588317069TUXN3827X8.pdf, in Chinese.

Crawford VP, Sobel J (1982) Strategic information transmission. *Econometrica* 50(6):1431–1451, URL http://dx.doi.org/10.2307/1913390.

Daniel M (2024) New in maps: Inspiration curated with Gemini, enhanced navigation and more. URL https://blog.google/products-and-platforms/products/maps/gemini-google-maps-navigation-updates/.

Daviet R, Nishimura Y (2026) Leveraging generative artificial intelligence to create visual content in digital advertising. *Marketing Science* URL http://dx.doi.org/10.1287/mksc.2024.1130, articles in Advance.

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General* 144(1):114–126, URL http://dx.doi.org/10.1037/xge0000033.

Doleac JL, Stein LCD (2013) The visible hand: Race and online market outcomes. *The Economic Journal* 123(572):F469–F492, URL http://dx.doi.org/10.1111/ecoj.12082.

Donati D, Song L (2026) The influence of the vocal minority: Evidence from social media comments. Working paper, Columbia Business School and University of Illinois Urbana-Champaign.

Dranove D, Jin GZ (2010) Quality disclosure and certification: Theory and practice. *Journal of Economic Literature* 48(4):935–963, URL http://dx.doi.org/10.1257/jel.48.4.935.

eBay (2023) Magical listing tool harnesses the power of AI to make selling on ebay faster, easier, and more accurate. URL https://innovation.ebayinc.com/stories/magical-listing-tool-harnesses-the-power-of-ai-to-make-selling-on-ebay-faster-easier-and-more-accurate/.

European Union (2024) EU artificial intelligence act, article 50: Transparency obligations for providers and deployers of certain AI systems. URL https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-50.

Google (2025) 4 new ways Google AI powers creativity. URL https://blog.google/products/ads-commerce/google-ai-ads-creative/.

Google AI for Developers (2025) Gemma 3 model card. URL https://ai.google.dev/gemma/docs/core/model_card_3.

Greenwald AG, McGhee DE, Schwartz JLK (1998) Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology* 74(6):1464–1480, URL http://dx.doi.org/10.1037/0022-3514.74.6.1464.

Guilbeault D, Delecourt S, Desikan BS (2025) Age and gender distortion in online media and large language models. *Nature* URL http://dx.doi.org/10.1038/s41586-025-09581-z.

Howard P, Madasu A, Le T, Lujan Moreno G, Bhiwandiwalla A, Lal V (2024) SocialCounterfactuals: Probing and mitigating intersectional social biases in vision-language models with counterfactual examples. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11975–11985.

Huang C (2026) TikTok US ad revenue forecast to hit $14.5bn in 2026, 38% of its global revenue. WARC, URL https://www.warc.com/en/content/article/warc-curated-datapoints/tiktok-us-ad-revenue-forecast-to-hit-145bn-in-2026-38-of-its-global-revenue/162277.

IAB UK (2026) Digital adspend 2025: UK's digital ad market reaches £40.5bn. URL https://www.iabuk.com/news-article/digital-adspend-2025-uks-digital-ad-market-reaches-ps405bn.

Interactive Advertising Bureau (2025) Nearly 90% of advertisers will use Gen AI to build video ads, according to IAB's 2025 video ad spend & strategy full report. URL https://www.iab.com/news/nearly-90-of-advertisers-will-use-gen-ai-to-build-video-ads/.

Jin GZ, Leslie P (2003) The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics* 118(2):409–451, URL http://dx.doi.org/10.1162/003355303321675428.

Jung S, Dubois D (2023) When and how slow motion makes products more luxurious. *Journal of Marketing Research* 60(6):1177–1196, URL http://dx.doi.org/10.1177/00222437221146728.

Kahveci S (2024) rapidsplithalf: A fast permutation-based split-half reliability algorithm. R package, version 0.2, URL http://dx.doi.org/10.32614/CRAN.package.rapidsplithalf.

Kapoor A, Kumar M (2025) Frontiers: Generative AI and personalized video advertisements. *Marketing Science* 44(4):733–747, URL http://dx.doi.org/10.1287/mksc.2023.0494.

Kärkkäinen K, Joo J (2021) FairFace: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1548–1558, URL http://dx.doi.org/10.1109/WACV48630.2021.00159.

Kim D, Jiang Z, Thomadsen R (2024) TV advertising effectiveness with racial minority representation: Evidence from the mortgage market. Working Paper 24-119, Marketing Science Institute, URL https://thearf-org-unified-admin.s3.amazonaws.com/MSI_Report_24-119.pdf.

Koch J, Pellatt C (2025) AI adoption is surging in advertising, but is the industry prepared for responsible AI? URL https://www.iab.com/insights/ai-adoption-is-surging-in-advertising-but-is-the-industry-prepared-for-responsible-ai/.

Lahoti P, Blumm N, Ma X, Kotikalapudi R, Potluri S, Tan Q, Srinivasan H, Packer B, Beirami A, Beutel A, Chen J (2023) Improving diversity of demographic representation in large language models via collective-critiques and self-voting. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 10383–10405 (Singapore: Association for Computational Linguistics), URL http://dx.doi.org/10.18653/v1/2023.emnlp-main.643.

Lambrecht A, Tucker C (2019) Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management Science* 65(7):2966–2981, URL http://dx.doi.org/10.1287/mnsc.2018.3093.

Lee D, Hosanagar K, Nair HS (2018) Advertising content and consumer engagement on social media: Evidence from Facebook. *Management Science* 64(11):5105–5131, URL http://dx.doi.org/10.1287/mnsc.2017.2902.

Logg JM, Minson JA, Moore DA (2019) Algorithm appreciation: People prefer algorithmic to human judgment.

Organizational Behavior and Human Decision Processes 151:90–103, URL http://dx.doi.org/10.1016/j.obhdp.2018.12.005.

Luo LE, Toubia O (2026) Discrimination against femininity in headshots: A field experiment with AI-enabled controllable stimuli generation. Research Paper 4865798, Columbia Business School, URL http://dx.doi.org/10.2139/ssrn.4865798.

Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science* 38(6):937–947, URL http://dx.doi.org/10.1287/mksc.2019.1192.

Meares J (2025) A googler explains how to "meta prompt" for incredible Veo videos. Google, URL https://blog.google/products-and-platforms/products/gemini/meta-prompting-veo-gemini-tips/.

Morehouse KN, Banaji MR (2024) The science of implicit race bias: Evidence from the implicit association test. *Dædalus* 153(1):21–50, URL http://dx.doi.org/10.1162/daed_a_02047.

National Radio and Television Administration of China (2025) Measures for labeling AI-generated synthetic content. URL https://www.nrta.gov.cn/art/2025/3/14/art_113_70340.html.

Nelson P (1970) Information and consumer behavior. *Journal of Political Economy* 78(2):311–329, URL http://dx.doi.org/10.1086/259630.

Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453, URL http://dx.doi.org/10.1126/science.aax2342.

Rafailov R, Sharma A, Mitchell E, Ermon S, Manning CD, Finn C (2023) Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, volume 36, 53728–53741.

Reuters (2026) Bytedance's new AI video model goes viral as china looks for second deepseek moment. URL https://www.reuters.com/business/media-telecom/bytedances-new-ai-video-model-goes-viral-china-looks-second-deepseek-moment-2026-02-12/.

Ridgeway CL (2001) Gender, status, and leadership. *Journal of Social Issues* 57(4):637–655, URL http://dx.doi.org/10.1111/0022-4537.00233.

Tabassi E (2023) Artificial intelligence risk management framework (AI rmf 1.0). Technical Report NIST AI 100-1, National Institute of Standards and Technology, URL http://dx.doi.org/10.6028/NIST.AI.100-1.

TikTok (2025) Video play metrics. URL https://ads.tiktok.com/resources/help/article/video-play?lang=en.

TikTok (2026) TikTok world '26: Turning discovery into business growth with AI-powered innovations, vertical experiences and high impact brand solutions. URL https://newsroom.tiktok.com/tiktok-world-26-eu?lang=en-150.

Wang B, Chen W, Pei H, Xie C, Kang M, Zhang C, Xu C, Xiong Z, Dutta R, Schaeffer R, Truong ST, Arora S, Mazeika M, Hendrycks D, Lin Z, Cheng Y, Koyejo S, Song D, Li B (2023) DecodingTrust: A comprehensive assessment

of trustworthiness in GPT models. *Advances in Neural Information Processing Systems*, volume 36, 31232–31339, URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/63cb9921eecf51bfad27a99b2c53dd6d-Abstract-Datasets_and_Benchmarks.html.

Williams MJ, Paluck EL, Spencer-Rodgers J (2010) The masculinity of money: Automatic stereotypes predict gender differences in estimated salaries. *Psychology of Women Quarterly* 34(1):7–20, URL http://dx.doi.org/10.1111/j.1471-6402.2009.01537.x.

Xu K, Nosek BA, Greenwald AG (2014) Psychology data from the race implicit association test on the project implicit demo website. *Journal of Open Psychology Data* 2(1):e3, URL http://dx.doi.org/10.5334/jopd.ac.

Yelp (2024) Yelp's 2024 spring product release: Introducing yelp assistant. URL https://blog.yelp.com/news/spring-product-release-2024/.

Yelp (2025) Yelp expands AI features to make local discovery more conversational, visual and seamless. URL https://www.yelp-ir.com/news/press-releases/news-release-details/2025/Yelp-Expands-AI-Features-to-Make-Local-Discovery-More-Conversational-Visual-and-Seamless/default.aspx.

A Diagnostic and Review-Summary Evidence

A.1 Original Gemma diagnostic and complementary benchmarks

The original 12-face diagnostic uses 50 independent queries per face, each requesting three descriptors. Its 32.9-percentage-point Black-minus-White gap has a 95% race-by-gender-stratified image-bootstrap interval of [15.7, 48.4]. An exact race-label permutation within gender gives $p = .035$.

Table 9 reports seven additional models using ten queries per face on the same 12-face panel, followed by one query per image on a 400-image UTKFace panel.[4] The latter contains 200 Black/White pairs matched on recorded gender, age band, coarse expression, and image quality.

Table 9: Association benchmarks for Gemma and seven additional models

Model	Repeated 12-face panel	Matched 400-image UTKFace panel	Complete UTKFace pairs
Gemma-3-12B-it	32.9 [15.7, 48.4]	1.3 [-1.3, 3.8]	159
Qwen 3.5 9B	17.6 [0.6, 40.0]	7.7 [1.5, 13.8]	196
Kimi K3	6.7 [0.0, 15.0]	5.3 [-0.8, 11.6]	175
GPT-5.6 Luna	16.1 [0.0, 35.0]	9.6 [3.0, 16.2]	197
Gemini 3.6 Flash	6.7 [-23.3, 33.3]	10.8 [3.5, 18.2]	198
MiMo-V2.5	14.4 [-1.7, 32.2]	9.6 [2.5, 16.7]	198
MiniMax M3	0.6 [-13.1, 13.1]	3.2 [-0.7, 7.2]	199
Step 3.7 Flash	3.3 [-3.3, 10.0]	2.0 [-1.5, 5.5]	200

Notes: Values are Black-minus-White negative-selection gaps in percentage points. Brackets report 95% image- or matched-pair-bootstrap intervals. The 12-face panel uses 50 independent queries per face, requesting three descriptors from a fixed vocabulary of 30 positive and 31 negative terms. Gemma's 12-face entry reports the original 50-query diagnostic used in the main text. UTKFace estimates use one query per image and retain complete pairs with valid three-descriptor responses.

A.2 Explicit discrimination test

The explicit discrimination test asks models to agree or disagree with 16 racial or ethnic stereotype statements from DecodingTrust (Wang et al. 2023). A phrase-based rule classifies endorsement and rejection. Gemma endorses none of its ten classifiable responses. None of the seven comparison models endorses a classifiable statement.

A.3 Paired review-summary scores

The Claude judge scores sentiment (overall positivity), clarity (readability), detail (specificity), and personal-experience content from 1 to 10. Higher scores indicate more of each attribute, with sentiment 5 denoting

[4] Dataset source: `https://susanqq.github.io/UTKFace/`.

mixed or neutral content. Table 10 reports paired Black-minus-White differences across 839 products, with 95% paired-t intervals using 838 degrees of freedom.

Table 10: Paired baseline-Gemma review-summary scores

Quality	Black-cue mean	White-cue mean	Paired gap [95% CI]	p
Sentiment	7.390	8.451	-1.061 [-1.144, -0.977]	< .001
Clarity	8.867	8.961	-0.094 [-0.126, -0.063]	< .001
Personal experience	3.621	4.042	-0.421 [-0.516, -0.326]	< .001
Detail	5.378	5.490	-0.112 [-0.177, -0.047]	< .001

Notes: Scores range from 1 to 10; higher sentiment is more positive. The sentiment gap is 12.6% of the White-cue mean (1.061/8.451). Product and review text remain fixed within each pair. Only the matched seller cue differs.

B Lab Summary-Cue Contrasts Across Disclosures

Table 11 reports summary-cue contrasts across the six disclosure conditions. The pooled model interacts generator condition with every term in Equation (5), including product fixed effects. Inference uses 775 survey-completion clusters, CR1 covariance, and t_{774}. The common-mix average applies the pooled sample's observed disclosure distribution to both generators.

Table 11: Black-minus-White own-WTP contrasts by disclosure and generator

Disclosure condition	Baseline Gemma	DPO-tuned Gemma	DPO minus baseline
Human / None	-0.788*** [-1.071, -0.505]	0.019 [-0.328, 0.366]	0.807*** [0.360, 1.255]
Human / Representative	-1.272*** [-1.590, -0.954]	-0.048 [-0.324, 0.227]	1.224*** [0.803, 1.644]
Human / Unrepresentative	-1.091*** [-1.378, -0.805]	0.053 [-0.211, 0.318]	1.145*** [0.754, 1.535]
AI / None	-1.099*** [-1.399, -0.799]	-0.087 [-0.349, 0.175]	1.013*** [0.614, 1.411]
AI / Representative	-1.583*** [-1.852, -1.314]	-0.154 [-0.455, 0.147]	1.429*** [1.025, 1.833]
AI / Unrepresentative	-1.402*** [-1.687, -1.118]	-0.052 [-0.346, 0.241]	1.350*** [0.941, 1.759]
Common disclosure mix	-1.247*** [-1.394, -1.101]	-0.056 [-0.201, 0.089]	1.191*** [0.985, 1.397]

Notes: Estimates are dollars; brackets report 95% confidence intervals. The final column directly tests the change in the Black-minus-White contrast using the pooled covariance. The common-mix average uses identical disclosure weights for both generators. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

C Field Partner and Product Positioning

The partner produces rice-aroma baijiu, a Chinese distilled spirit belonging to one of China's 12 major baijiu aroma categories. Established in 1952, the company received national recognition for its flagship product in 1979 and participates in national standard setting for this category.[5]

The basic and premium products are priced at CNY 19 and CNY 128, respectively. They represent everyday and premium positioning within the same producer and have visually similar bottles. Character comparisons hold the product and its price fixed.

D Illustrative Field-Benchmark Prompts

These English translations reproduce one basic-product and one premium-product benchmark prompt from Section 6.3. Both leave character gender unspecified. The scenes differ with product positioning, so this benchmark describes the generation workflow rather than isolating price alone. The field experiment separately compares matched character versions within each product.

D.1 Basic-product prompt

> Use the product reference image I uploaded as the sole specified product in the video. The bottle shown in the uploaded image must be the bottle that actually appears in the advertisement. Use this specific bottle, rather than a generic bottle or one independently generated by AI. Strictly preserve the bottle's shape, proportions, transparent body, red cap, red-and-white label, Chinese-style landscape motif, and overall appearance shown in the reference image. Do not replace the product with a white-wine bottle, red-wine bottle, champagne bottle, or another type of bottle, and do not redesign, modify, or invent new product packaging.
>
> Set a nighttime relaxation scene on the rooftop of an ordinary older residential complex. The background consists of water towers, clothes-drying poles, and antennas on older apartment buildings, with dense city lights in the distance. One Asian adult aged 50–60, wearing pajamas, sits beside a simple folding chair. The specified bottle is on a small folding table nearby. The person performs only one simple action: gently adjusts the bottle's position and then looks toward the distant city lights.
>
> Place the camera on the opposite side of the rooftop and use a wide medium-long shot. Keep the person's full body within the frame, occupying 10%–20% of the image. The water towers, antennas, city nightscape, and rooftop environment dominate the frame. The camera must remain completely stationary, with no push-in or pull-out, panning, lateral movement, or zooming. Keep the specified bottle clearly visible without a close-up.
>
> Do not show any microphone, lavalier microphone, wireless microphone, headset microphone, interview equipment, or recording equipment on the person's chest. Do not include other people,

[5] Company history and category standing are documented in the introductory pages of partner-provided product materials.

other bottles, other branded products, dialogue, subtitles, readable text, or background music. Use one fixed shot lasting four seconds.

D.2 Premium-product prompt

Use the product reference image I uploaded as the only product reference. The bottle in the video must match the reference bottle exactly, including the transparent glass body, red cap, red-and-white label, and Chinese-landscape motif on the label. Do not change the bottle's color, shape, cap, label design, or any other detail, and do not generate other bottles.

Set the scene on the deck of a high-end private yacht sailing at sea, surrounded by an expansive deep-blue ocean and a gently sloping coastline in the distance. The yacht deck features high-quality wooden flooring, simple modern white seating, and refined metal details. The space is spacious and clean, conveying a high-end private maritime lifestyle.

One Asian adult aged 50–60 wears mature, high-end vacation clothing with sophisticated tailoring and luxurious materials. Do not add other ornaments or anything resembling an audio microphone to the clothing. The person stands on one side of the deck, naturally looking at the distant seascape. Place the specified bottle naturally on a high-quality nearby tabletop as a lifestyle element. The person must not pour a drink.

Show the person from the front or side, keeping the full body within the frame and occupying approximately 10%–20% of the image. Emphasize the luxury yacht deck, open seascape, wooden materials, modern seating, and natural light. Use a medium-long composition, avoid making the person or bottle too large, and ensure that the frame edges do not crop the person's head or body.

Keep the camera completely fixed, with no movement, push-in or pull-out, or rotation. The entire video lasts four seconds. The image should be realistic, elegant, and mature, conveying a high-end private-yacht lifestyle. Show only this one Asian adult, with no other people, bottles, or beverage containers. Do not use microphones or include dialogue, subtitles, readable text, or background music.

E Field Data and Robustness Checks

E.1 Delivery and outcome summaries

The platform records delivery outcomes for 120 treatment-day × macro-area cells. The partner provides daily campaign-attributed store visits and purchases. Across the 24-day campaign, there are 354,148 two-second completions, 6,206 store visits, and 702 purchases. Assigned budgets are CNY 68 per delivery cell, and actual daily spending ranges from CNY 338.08 to CNY 340.00. Table 12 summarizes delivery and merchant outcomes by treatment.

Table 12: Observed daily delivery and merchant outcomes

	2-second completions	Store visits	Purchases	Visit rate (%)	Purchase rate (%)
Basic / male	14,110	215.7	24.8	1.528	0.176
Basic / female	14,848	264.7	30.3	1.783	0.204
Premium / male	14,917	221.3	25.3	1.484	0.170
Premium / female	15,150	332.7	36.5	2.196	0.241
Control	13,858	220.2	25.8	1.589	0.186
AI + placebo	14,795	297.2	32.5	2.009	0.220
AI + association	15,616	258.2	29.5	1.654	0.189

Notes: Counts are means per treatment day. Rates divide summed daily actions by summed two-second completions within each group. Each product–character group contains six days; each disclosure group contains eight days. Groups pool the other assigned factors, which the regressions control for.

E.2 Daily-outcome and spending checks

Table 13 estimates action-rate contrasts using all 24 daily merchant totals. Each rate divides daily actions by that day's two-second completions. OLS includes product position, character gender, their interaction, both disclosure indicators, and audience targeting. With one observation per day, the model omits macro-area fixed effects and uses HC3 standard errors with normal-reference inference.

Table 13: Daily merchant outcomes: relative effects and 95% confidence intervals

Contrast	Store visits per 2s-completion	Purchases per 2s-completion
Basic: female versus male	+17.0% [+8.4, +25.5]***	+16.1% [+7.1, +25.0]***
Premium: female versus male	+48.9% [+39.3, +58.4]***	+42.6% [+28.8, +56.3]***
AI + placebo versus control	+27.2% [+21.0, +33.4]***	+18.9% [+9.7, +28.1]***
AI + association versus AI + placebo	-17.8% [-24.1, -11.6]***	-14.1% [-22.3, -5.9]***
AI + association versus control	+4.5% [-3.5, +12.6]	+2.1% [-6.9, +11.1]

Notes: The analysis uses all 24 campaign days. Each cell reports the relative percentage effect, followed by its 95% interval. Relative effects divide the fitted rate contrast by the corresponding comparison-group mean, treated as fixed. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

Adding actual daily spending as a control preserves premium-product female-character gains of 48.8% in visits and 42.6% in purchases (both $p < .001$). The disclosure effects retain their signs and significance (all $p < .01$, using HC3 standard errors and t_{16} inference).

E.3 Calendar and single-day sensitivity

Table 14 adds a campaign-day trend or weekday indicators to the main action-rate specification and separately omits each campaign day in turn. The premium-product female-character contrast and the two incremental disclosure contrasts retain their signs and significance. The basic-product purchase contrast remains positive

but has $p = .055$ with weekday controls. For the character contrasts and the two incremental disclosure contrasts, every single-day omission preserves the estimated sign and $p < .01$.

Table 14: Sensitivity of field action-rate contrasts

Contrast	Outcome	Day trend	Weekday controls	Single-day deletion range
Basic: female–male	Visits	+15.5%***	+14.5%***	[+14.5, +18.7]%
	Purchases	+13.3%***	+11.1%*	[+12.5, +16.3]%
Premium: female–male	Visits	+48.6%***	+50.2%***	[+46.7, +52.2]%
	Purchases	+42.4%***	+47.3%***	[+38.8, +46.0]%
AI + placebo vs. control	Visits	+27.1%***	+22.8%***	[+24.6, +27.2]%
	Purchases	+18.9%***	+15.1%***	[+15.9, +20.6]%
AI + demographic association vs. AI + placebo	Visits	-17.4%***	-16.1%***	[-19.3, -15.9]%
	Purchases	-13.7%***	-11.0%**	[-15.9, -12.0]%
AI + demographic association vs. control	Visits	+5.1%	+2.5%	[+1.8, +6.1]%
	Purchases	+2.7%	+2.2%	[-0.3, +3.9]%

Notes: Values are relative percentage effects using the full-sample comparison-group mean as a fixed denominator. Calendar models retain 120 cells, macro-area fixed effects, and 24 treatment-day clusters with CR1 covariance and t_{23} inference. Each deletion model omits one day's five cells and uses 23 clusters with t_{22} inference. Deletion ranges summarize point estimates, not confidence intervals. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

E.4 Audience-targeted comparisons

Table 15 reports the premium-product female-character contrast separately for male- and female-targeted days. The advantage is positive in both groups, but the differences across audience targets are not statistically significant ($p = .663$ for visits and $p = .273$ for purchases). Each subgroup regression uses daily totals and the same treatment controls, excluding the audience indicator.

Table 15: Premium-product female-character effects by audience target

Audience target	Store visits per 2s-completion	Purchases per 2s-completion
Male	+42.0% [+21.3, +62.6]***	+31.7% [+12.2, +51.3]***
Female	+57.3% [+49.5, +65.2]***	+56.0% [+38.5, +73.5]***

Notes: Each cell reports a relative percentage effect and 95% interval. Daily-total OLS uses HC3 covariance and normal-reference inference, with 12 days per audience. Moderation tests compare the unscaled rate contrasts across the two audience regressions, rather than comparing their relative percentages or significance markings. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.

E.5 Additional engagement outcomes

Table 16 applies the main field specification to five-second completion, sharing, and new followers. For the premium product, the female-character ad increases sharing, while the follower-rate increase has $p = .055$. The AI-plus-placebo effect on five-second completion is not statistically significant ($p = .221$).

Table 16: Additional field engagement outcomes

Contrast	Five-second completion rate (pp)	Share rate (pp)	New-follower rate (pp)
Basic: female versus male	5.285**	0.005	0.006
	(2.140)	(0.031)	(0.011)
Premium: female versus male	1.082	0.105***	0.023*
	(0.666)	(0.027)	(0.012)
AI + placebo versus control	1.887	0.017	-0.016
	(1.500)	(0.024)	(0.011)
AI + association versus AI + placebo	0.255	0.010	0.005
	(0.951)	(0.030)	(0.009)
AI + association versus control	2.142	0.027	-0.012
	(1.581)	(0.019)	(0.009)

Notes: Effects are percentage points. Share and new-follower rates divide counts by two-second completions. Five-second completion is the platform-reported rate. OLS uses 120 treatment-day × macro-area cells and includes product position, character gender, their interaction, both disclosure indicators, audience targeting, and macro-area fixed effects. Parentheses report CR1 standard errors clustered by 24 treatment days, with t_{23} inference. $^{*}p < .10$, $^{**}p < .05$, $^{***}p < .01$.