Generative AI and Personalized Video Advertisements

Abstract

We study the effectiveness of personalized video advertisements created using Generative AI (GenAI). We run a mobile ad targeting field experiment on WhatsApp in partnership with a leading B2C e-commerce brand that sells eco-friendly sustainable products. We randomize users into receiving ads from one of three targeting conditions – 1) Generative AI-based personalized video ads, 2) personalized image ads, and 3) generic non-personalized video ads. The first group is our main treatment and the latter two serve as baselines. In the personalized treatment conditions, ad content is tailored to individual purchase histories, while in the generic treatment condition, a uniform brand message is delivered to all users. Our results show that Generative AI-based personalized video ads increase engagement by 6-9 percentage points over the baselines. We explore response heterogeneity across important user segments and find that GenAI-based ads increase engagement among potential "high-value" users and historically "non-engaged" users. Further, we use back-of-the-envelope calculations to highlight substantial cost savings and productivity benefits of GenAI-based personalized ad campaigns.

1 Introduction

Generative Artificial Intelligence (GenAI) is a promising new technology with the potential to disrupt multiple business verticals such as marketing, design, and innovation. Its application in marketing is particularly touted as transformational and is expected to influence core processes such as creative designs for ads, content creation for personalized targeting, customer support, and learning consumer feedback from user-generated content. According to Market.us, the market size of global GenAI in marketing was valued at USD 1.9 Billion in 2022. From 2023 to 2032, it is estimated to reach USD 22.1 Billion growing at a CAGR of 28.6%.

We empirically test the theoretical promise of GenAI for a core marketing outcome – advertisement effectiveness. We begin by designing a proof-of-concept that demonstrates how GenAI can be used to create personalized video advertisements, a task that was previously prohibited due to cost and effort. We then take this idea to the field and examine two pivotal questions. 1) Can GenAI-based personalized video ads enhance engagement metrics when pitted against traditional digital ad formats? 2) Do all users respond equally to these ads or is there heterogeneity across important user segments?

Our study is part of a multi-stakeholder research effort. We partner with three firms – 1) a leading Generative AI technology company to create personalized video ads with digital avatars (AI-Firm hereafter), 2) a leading Indian online retail brand selling eco-friendly sustainable products for which the ads were created, and 3) a delivery partner that provides marketing services via WhatsApp.

We run an ad targeting field experiment in partnership with the online retailer. We randomly assign the online retailer's existing customers to one of three experimental conditions – 1) GenAI-based personalized video ads, 2) personalized image ads, and 3) generic non-personalized video ads. The first group is our treatment and the latter two serve as baselines.

In the first experimental condition, i.e., the GenAI-based treatment, users receive personalized video ads created using AI-Firm's proprietary text-to-video generative AI engine. This process works in two steps. First, the technology company creates a high-fidelity digital avatar of the online retailer's brand ambassador based on pre-recorded template videos. Second, the text-to-video generative models personalize the content the brand ambassador communicates in the videos based on each user's historical data. This personalization follows the principles of 'behavior-based personalization' (Arora et al., 2008) such that messages are user-context dependent and are tailored at the ad-content level. For example, consider a user *Esha* who recently purchased eco-friendly dishwashing liquid from the retailer and now the retailer would like to upsell cleaner sheets to her. An excerpt of the personalized video message would have the following content, "Hi *Esha*, I hope you're enjoying chemical-free and eco-friendly living. Have you seen our new range of *Cleaner Sheets*? ...". The italicized words highlight the part of the message that is personalized based on the user's data. (Please refer to Figure A3 in the Online Appendix for the complete message with screenshots.) While scripting such text messages is straightforward, GenAI now allows us to create personalized content for videos with high fidelity and realism.

In contrast, users in the second condition, i.e., personalized image ads treatment receive similar messages to the GenAI group, but the content is delivered only using images and text. This format is the industry status quo and serves as an important baseline. The primary distinction between GenAI-based

personalized video ads and personalized image ads lies in the mode of personalized content delivery, which in the former case is video.

For both personalized treatment conditions, we create three sub-conditions to tailor message content based on the user's purchase history – (a) *upsell:* aimed at promoting new products to customers who have not yet made such purchases, (b) *revisit:* designed to encourage users to visit the retailer's website and explore new product selections if they haven't made recent purchases, and (c) *switch:* targeting signed-up prospects—users who have shared their contact information but haven't placed any orders yet—with the goal of persuading them to switch to the retailer's brand.

Lastly, the third condition involves all users receiving non-personalized generic video ads. In this group, every user receives identical video content conveying the brand's message about using their eco-friendly products. This condition allows us to control for a latent "video effect" and estimate the causal impact of *personalized* video content created using GenAI on engagement.

Results

We find that the engagement rate, measured using ad clicks, among users who received GenAI-based personalized video ads is 6-9 percentage points (p.p.) higher as compared to the users in the two baseline groups. These results are robust across demographics – gender and location, and across different model specifications.

Furthermore, our results reveal which customer segments benefit most from this novel ad format. We estimate heterogeneity in users' response to GenAI ads across key dimensions of consumer behavior – 1) RFM (Gupta et al., 2006), and 2) historical ad engagement (Rafieian, 2022). Our results show that GenAI-based personalized video ads perform better in engaging potentially "high-value" and previously "non-engaged" customers. The former includes customers who have been "moderately" involved in transacting with the retailer over the past year, i.e., they occupy the middle of the brand loyalty spectrum. The latter includes customers who have been sent targeted marketing messages in the past but have never engaged with them. These groups are potentially valuable customer segments and winning back such customers is crucial yet difficult (Ascarza, 2018). Our results show that GenAI-based personalized video ads can help make progress on this front.

It is important to contextualize the substantive findings of our study in light of the broader literature on advertising, personalization, and algorithmic aversion. Ex-ante the answer to our research question, "Are AI-generated personalized video ads more effective?" is theoretically ambiguous. On one hand, we know that personalization is important in marketing communication (Sahni et al., 2018; Zhang, 2011), with personalized advertisements having engagement rates that are almost twice as high as compared to generic ads (Tucker, 2014). On the other hand, increasing privacy concerns (Jin, 2018) and aversion to algorithms (Castelo et al., 2019; Dietvorst et al., 2015) make it unclear how consumers will react to AI-generated personalized video ads (Zhang, 2011). Personalization can backfire and instead of improving outcomes, these AI-generated ads could be perceived as invasive (Goldfarb and Tucker, 2011; Johnson et al., 2020). In light of these considerations, our results provide evidence that GenAI-based personalized video ads perform significantly better in engaging consumers. We provide a more nuanced discussion

regarding the implications of our results in Secion 6.

Contributions

Our work makes four direct contributions. First, our unique multi-stakeholder collaboration allows us to causally test the effectiveness of GenAI in a real-world field setting. Recent work, especially with Large Language Models (LLMs), has shown that Generative AI can boost productivity in writing tasks (Noy and Zhang, 2023), generate more financially viable solutions for business challenges (Boussioux et al., 2023), increase search engine advertising performance (Reisenbichler et al., 2023), and improve customer service outcomes (Brynjolfsson et al., 2023). We extend the scope of this literature by providing causal empirical evidence of GenAI improving core marketing outcomes.

Second, we present, to the best of our knowledge, the first empirical evidence of the impact of GenAI on video advertisements. Video content production and consumption have been growing at a staggering rate on social media platforms. Furthermore, industry studies have shown that users prefer videos over text and images for marketing communication. Current practice targets users with video ads that are personalized at the level of context, i.e., sending sports enthusiasts a Nike or Gatorade ad. GenAI allows for a marked shift in this practice by enabling the creation of video advertisements with personalized content. In our case, users receive personalized video communication from the retailer's brand ambassador with video content tailored for each user. This was previously unfeasible. We believe that our research is well-positioned in the broader context of this industry trend and has relevant implications for marketing managers.

Third, by documenting for whom GenAI-based personalized video ads work best, our granular results provide a roadmap for marketers aiming to deploy this new technology responsibly. We demonstrate that GenAI video ads reliably re-engage moderately active past customers and previously unmoved targets - two high-potential yet challenging segments. These novel findings delineate contexts where managers may gain the most from GenAI-based marketing efforts.

Finally, a key feature of our study is using GenAI for "guided personalization". In the current AI hype, it is tempting to leverage GenAI's full capabilities and generate ad creatives from scratch. However, without human oversight, this can lead to inconsistent, off-brand, or even inappropriate outputs. Recent high-profile failures of AI chatbots and LLMs illustrate the risks of an unconstrained approach. For example, reports have shown how Stable Diffusion, one of the most popular text-to-image engines, perpetuates harmful stereotypes (Nicoletti and Bass, 2023), how Google's Gemini distorts history in unacceptable ways (Robertson, 2024), and how LLMs hallucinate to conjure up inaccurate facts (Economist, 2024). In a marketing context, such AI misbehavior could seriously damage brand reputation and erode customer trust. By instead using GenAI in a more focussed and controlled fashion as we do, brands can harness the power of AI personalization while mitigating risk. Our experiment design, estimated treatment effects, and documented cost savings provide a valuable roadmap for marketers aiming to deploy this new technology at scale.

Related Work

Our work relates to several streams of literature. We build on the nascent literature that studies Generative AI applications in marketing. Recent papers include Burnap et al. (2023) who develop a novel method to improve car designs and Dew et al. (2022) who built a multi-modal variational autoencoder to generate new logo designs. Reisenbichler et al. (2022) show how Natural Language Generation methods can be used to improve SEO outcomes and Reisenbichler et al. (2023) further extend this scope by leveraging LLMs to design better search engine ads. Li et al. (2022) construct perceptual maps using LLMs. Brand et al. (2023) investigate whether LLMs can be used to understand consumer preferences and offer guidelines on how to use them as aids in marketing research. This body of work is itself situated within the larger ambit of the use of machine learning methods to improve marketing processes. This includes predicting product return rates using images (Dzyabura et al., 2023), designing hierarchical attention networks to generate more creative marketing content (Hong and Hoban, 2022), predicting emotions in music using convolutional neural networks (Fong et al., 2022), and identifying customer needs from user-generated content using deep neural networks (Timoshenko and Hauser, 2019). Our experiment extends this literature by providing real-world causal evidence on the effectiveness of Generative AI in improving advertising outcomes through targeted video ads.

More substantively, our research is related to the vast literature on ad personalization, targeting, and impact. This includes work on the effectiveness of search engine ads (Blake et al., 2015; Kalyanam et al., 2018), the efficacy of personalized online ads (Bleier and Eisenbeiss, 2015; Hoban and Bucklin, 2015), and the impact of re-targeted ads (Lambrecht and Tucker, 2013; Sahni et al., 2019). Of particular relevance is work that studies the impact of a new medium of advertisement. For example, Tucker (2015) studies the persuasiveness of viral video ads, Li and Lo (2015) investigates the impact of length and location of in-stream video ads on brand recognition, and Goodrich et al. (2015) examines the association between video ad characteristics and the ad's perceived intrusiveness. We add this literature by studying a new medium of communication – video ads with personalized content.

Our paper also relates to a growing empirical literature studying human-AI interaction and algorithmic aversion. Previous research has used data on exposures and consumer actions to study the effect of AI and algorithm exposure on consumer behavior, e.g., in the context of outbound sales calls from chat-bots on consumer purchases (Luo et al., 2019), AI as an employee feedback mechanism (Tong et al., 2021), and consumers' receptivity to medical AI (Longoni et al., 2019)). Our research explores the response to algorithm-generated ads in a field setting.

To summarize, our study adds to the broader marketing literature by – 1) illustrating a practical use-case of GenAI in advertising, 2) investigating the effectiveness of video ads with content personalized at the user level, and 3) exploring heterogeneity in consumer response to AI-generated ads.

2 Experiment Setting and Design

We test the effectiveness of GenAI-based ads in a multi-stakeholder research effort. We partner with three firms – 1) a leading Generative AI technology company to create hyper-personalized video ads, 2)

a leading Indian e-commerce brand selling eco-friendly products for which the ads were created, and 3) a delivery partner that provides marketing services via WhatsApp.

The Generative AI technology company is based in the US and India and provides GenAI solutions to enterprise customers. A big use case of their technology is creating digital avatars of brand ambassadors for personalized ad campaigns. These campaigns allow brands to create tailored messages for their customers depending on where they are in their customer journey. Since its launch, this firm has attracted many large and medium-sized media advertisers who use the company's technology to serve personalized video ads.

The e-commerce partner is a direct-to-consumer (B2C) brand that sells eco-friendly sustainable products for categories such as personal care, laundry, kitchen, and cleaning supplies. They are a medium-sized brand selling 24 unique products with an average price of about INR 200 ($\sim$ \$2.5). The brand regularly engages in marketing campaigns where it sends promotional coupons and brand awareness messages to its existing and prospective customers.

Our delivery partner helps B2C brands interact with their customers, share brand-related promotional materials, and send targeted marketing messages through apps such as WhatsApp and Facebook Messenger. Their expertise is in using chat-based apps for marketing communication. We focus on WhatsApp marketing in our experiment since that is one of the primary channels the e-commerce firm uses for brand outreach.

2.1 Design

We run a randomized ad experiment where the online retailer's existing customers are assigned to one of the three experimental conditions – 1) Generative AI-based personalized video ads (AI-video), 2) personalized image ads (P-Image), and 3) generic non-personalized video ads (N-Video). The first is our main treatment group and the latter two serve as important baselines/controls. Multiple aspects of our experimental study warrant further discussion. We provide details for each of them below. A schematic diagram providing an overview of the design process is shown in Figure 1.

Control conditions

Personalized image ads are the current industry status quo and the online retailer uses them frequently for targeting on WhatsApp. From a practitioner's perspective, any new advertisement format should perform better than this to be adopted into their marketing stack. Therefore, we use this as our first baseline. However, an issue with just using image ads as the control is that our main treatment actually delivers personalized video ads and not images. In the event that GenAI video ads outperform image ads, disentangling the impact of the content from the medium will be challenging. To address this, we introduce an additional baseline – generic non-personalized video ads. Users assigned to this group receive an identical generic video message. We compare the performance of GenAI video ads to both baselines.

One may argue that the right control/baseline should be human-created personalized video ads.

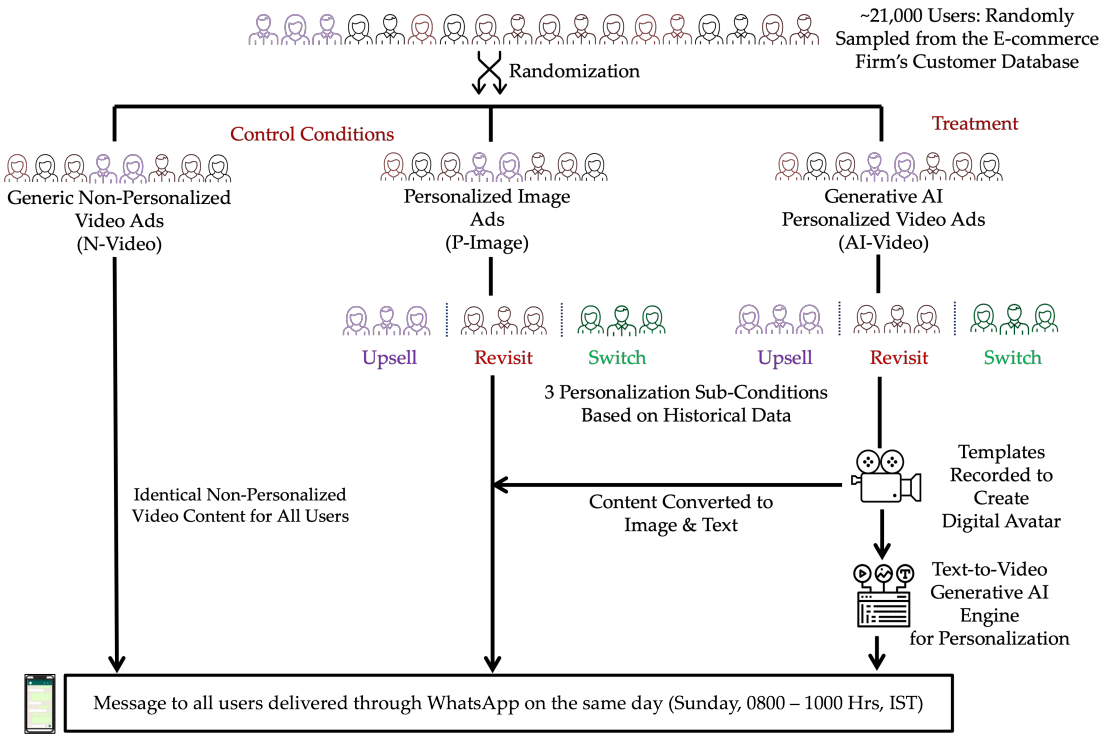


Figure 1: Overview of Experiment Design

However, the time and monetary costs prohibit creating and recording thousands of such ads. In fact, it is this limitation in the first place that serves as motivation for using GenAI to create personalized video content at scale.

Ad content & personalization

The ad content template was sketched by the marketing team of the online retailer. The objective of the campaign was to increase engagement among existing customers and the content was designed to reflect these goals.

Users in the treatment group receive personalized video messages from the brand ambassador created using the technology partner's text-to-video Generative AI engine. This process begins by first creating a high-fidelity digital avatar of the brand ambassador using pre-recorded templates. The templates are then fed into a text-to-video AI generator that personalizes the content the brand ambassador speaks. The specific personalized message depends on the user's historical data and context.

For our experiment, we create three personalized targeting sub-conditions in partnership with the marketing team of the retailer. The assignment rules to the sub-conditions are based on the user's stage in their customer journey. The three sub-conditions are – 1) *Upsell*, which targets consumers who have not yet purchased any of the products that the retailed intended to promote during the campaign (e.g., cleaning and laundry products), 2) *Revisit*, which encourages users who have not purchased anything from the retailer in the past month to visit the retailer's website and restock their supplies, and 3) *Switch*,

which targets signed-up prospects —users who have provided their contact information to the retailer but are yet to make any purchase—to switch to the retailer's brand. The personalized content in the video then includes the user's first name and the product that the retailer intended to promote to the particular user. No other part of the content was personalized.

Among the other two experiment arms, users in the personalized image group (P-Image) receive similar content as the GenAI group but with an image of the same brand ambassador accompanied by the text of the ad body. The ad content is adjusted for length and grammar that is appropriate for a text message while ensuring consistency in the meaning of the content. For the generic non-personalized ads groups (N-Video), a message was crafted to promote the brand's position on creating eco-friendly products. The message was then recorded live with the online retailer's actual brand ambassador.

All ads across the three treatment groups are created using the same brand ambassador. Figures A1a, A1b, and A2 in the Online Appendix show examples of three types of ads. Figure A3 provides screenshots and the script for one of the GenAI-based personalized video ads. Figure A4 provides screenshots and the script of the generic non-personalized video ad.

There are two points worth noting here. First, the assignment to treatment conditions is completely random and does not depend on the sub-conditions. That is, whether a user is assigned to the *switch*, *upsell*, or *revisit* depends on their historical purchases with the retailer, but the actual treatment that they receive (P-Image, N-Video, or AI-Video) is orthogonal to the sub-conditions.

Second, the content for the two personalized conditions is pre-scripted by the retailer's marketing team. GenAI here serves as an enabler of this personalization by creating high-fidelity video content. It *does not* decide what content is personalized. This feature, we believe, is a strength of our design as it permits controlled personalization, which is critical for any firm when testing out new technologies. Fine-grained control over what gets customized for users renders manual verification of each video unnecessary. An alternate approach would be to let GenAI decide the content and the creatives. However, currently, there is no automated system to ensure the fidelity of such content at scale, making the deployment of GenAI ads to real users very risky.

Delivery method

Ad messages are sent to users through WhatsApp by the delivery partner. WhatsApp is a popular communication app in India and has over 450 Million active users. We use WhatsApp for our study for two reasons. 1) The e-commerce brand in this study runs 2-3 WhatsApp ad campaigns each month, where each campaign targets anywhere between 20,000 – 50,000 users. This mitigates the concern for any channel-based novelty effects in our results since over 98% of the users in our study have received a WhatsApp marketing message from the retailer before. 2) While it is most commonly used for personal chats, it is becoming an increasingly popular tool among marketers for brand communication.

Outcome of interest

Our primary outcome of interest is ad engagement, which we measure using ad clicks. Data on ad delivery, receipt, and clicks are received by the delivery partner for each message they send. Ideally, we would like to measure ad success using down-the-funnel metrics such as purchases or revenue. However, the small magnitude of ads on purchases is well-documented (Lewis and Rao, 2015; Johnson, 2023). We don't have enough statistical power to detect effects on purchases as that would require participants in the range of hundreds of thousands if not millions. Nevertheless, ad engagement has been shown to be a leading indicator of customer purchases (Sahni et al., 2018; Sahni and Nair, 2020). We believe that our results on the effectiveness of GenAI ad videos on engagement are an important step forward in understanding the potential impact of this technology on core business outcomes.

Spillovers from other campaigns

Finally, we pause all advertising campaigns by the retailer for three days beginning the day of the experiment. This allows us to cleanly estimate the impact of our treatment while minimizing spillovers from other channels that could potentially contaminate the causal findings.

Table 1: Balance check for key variables

	AI-Video (Treatment)	P-Image (Baseline)	N-Video (Baseline)	p-value
# of users	9,510	5,907	5,911	
Pre-treatment variables				
Gender (Female = 1)	0.61	0.60	0.60	0.153
	(0.49)	(0.49)	(0.49)	
Number of orders in the past year*	1.12	1.14	1.10	0.625
	(2.18)	(2.25)	(2.14)	
Total spend in the past year (INR)*	855.70	850.49	829.60	0.815
	(2336.43)	(2321.36)	(2269.87)	
Days since last purchase	237.38	238.60	233.84	0.485
	(126.13)	(128.03)	(126.98)	
Number of past marketing campaigns sent	2.28	2.27	2.27	0.480
March and April 2023	(0.54)	(0.53)	(0.54)	
Historical ad engagement rate	0.81	0.83	0.85	0.071
March and April 2023	(1.06)	(1.05)	(1.07)	

*1) This table shows the mean and standard deviation (in parentheses) for key pre-treatment variables across the three experimental groups. 2) The last column shows the p-value from ANOVA testing the difference in means across the three groups. 3) # of past marketing campaigns measures the number of marketing messages sent to the user through WhatsApp in the past two months before the experiment. 4) * indicates that the confidential numbers regarding orders and sales have been obfuscated by multiplying them with a common random number.*

2.2 Data

We randomize 21,328 customers of the e-commerce into the three experiment arms described above - 1) AI-Video (GenAI-based treatment), 2) P-Image (baseline), and 3) N-Video (baseline). We observe the historical transactions of these customers over the past year with the online retailer, which we use for designing 'behavior-based personalization' sub-conditions. We create three personalized targeting

sub-conditions – 1) *Upsell*, 2) *Revisit*, and 3) *Switch*. All users in the generic non-personalized condition receive an identical video message.

The experiment was run on a Sunday in summer 2023, and messages were sent between 8:00 AM - 10:00 AM IST. We collected post-treatment one week after the experiment. The outcome data are at a user level in which we observe the time the message was sent, the time it was delivered, and the time the user clicked on it (conditional on the user clicking).

Table 1 provides the means and standard deviations of key pre-treatment variables across the experiment groups. We do balance checks across these variables to verify that the randomization was done correctly. The last column shows the p-values from ANOVA testing the difference in means across the experiment groups. We do not detect any systematic difference in the variables across the experimental groups.

3 Average Treatment Effect

We estimate treatment effects using the following regression:

$$y_i = \alpha + \beta \times T_i + \epsilon_i \tag{1}$$

where y_i is a binary variable that takes the value 1 if the user i clicks on the ad and 0 otherwise. β is our coefficient of interest which measures the impact of the treatment T_i, where T_i is also a binary variable that takes the value 1 if user i is assigned to the GenAI ad treatment group and 0 otherwise.

Our main results are shown in Table 2. Each column shows estimates from a linear probability model in which we regress our outcome 'Ad Click' on the treatment indicator as described in Equation 1. For ease of exposition, we split the comparison with the two baseline variables. In Column 1, the baseline (control) is the industry status quo – personalized image ads (P-Image), and in Column 2 it is generic non-personalized videos (N-Video). We find that the experimental group which receives Generative AI-based personalized video ads has a significantly higher ad click rate. In Columns 3 and 4, we add controls for demographics – gender and geographic state along with a measure of historical pre-treatment total purchases made by the consumer from the online retailer. Again, our results are robust to these controls and we find that ad clicks in the treatment group are 6-9 percentage points (p.p.) higher than those of the two baselines.

Section C of the Online Appendix shows further robustness for our results including running a logistic regression on the binary dependent variable (ad click) and a Fisher's exact test. We also test whether there is any differential impact based on gender and state of residence. Figures E10 and E11 in the Online Appendix show that the results are stable and robust across these demographic variables. Furthermore, we explore potential heterogeneity in the time users take to click (conditional on clicking) after receiving the message on WhatsApp. More than 90% of clicks happen within one hour of receiving the message and there is no substantial difference among the three experimental conditions. Section D in the Online Appendix shows the results.

Table 2: Treatment effect of personalized ad videos created using Generative-AI on engagement

Dependent Variable:	Ad Click			
Baseline/Control:	P-Image	N-Video	P-Image	N-Video
Model:	(1)	(2)	(3)	(4)
Variables				
(Intercept)	0.0946***	0.1230***	0.1447***	0.1535***
	(0.0046)	(0.0048)	(0.0171)	(0.0181)
Treatment	0.0935***	0.0651***	0.0938***	0.0653***
	(0.0059)	(0.0061)	(0.0059)	(0.0061)
Gender (F=1)			-0.0205***	-0.0223***
			(0.0060)	(0.0062)
Total spend (last 1 year)			0.0037***	0.0050***
			(0.0009)	(0.0010)
State			✓	✓
Fit statistics				
Observations	15,417	15,421	15,417	15,421
R^2	0.01600	0.00734	0.02078	0.01223
Log-Likelihood	-5,971.1	-6,471.7	-5,933.6	-6,433.6

*Each column shows the estimates from a linear probability model estimated using OLS. The baseline/control in columns 1 and 3 is the personalized image group. The baseline/control in columns 2 and 4 is the generic non-personalized video group. IID Standard Errors are in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

4 Consumer Response Heterogeneity

Our main results show an effective proof-of-concept design for leveraging GenAI to create personalized video advertisements, a task that was unachievable at scale previously. While our aggregate results are encouraging, the individual-level impact may vary with the user's characteristics. In this section, we consider two potential user-level moderators that may affect the response to personalized ads: 1) RFM – Recency, Frequency, and Monetary (Gupta et al., 2006), and 2) engagement with past ad campaigns (Rafieian, 2022).

We preface this discussion by noting that our analysis here is largely exploratory. Our data does not allow direct testing of the underlying mechanisms. Hence, while do connect to extant literature to motivate our analysis, we refrain from making claims about the specific behavioral mechanisms at play. Nevertheless, we believe that the heterogeneity results described below are insightful for researchers to probe further and actionable for managers to implement.

4.1 RFM

RFM is a segmentation scheme that classifies users based on their spending habits with the firm. They provide interpretable user segments for firms to test different marketing strategies for different sub-groups of users. Firms can then optimize resource allocation over their entire customer base with tailored strategies over sub-groups.

For each user in the experiment, we calculate the RFM measures using historical pre-treatment data from the past year. We discretize these variables by splitting them into three bins – low, medium, and

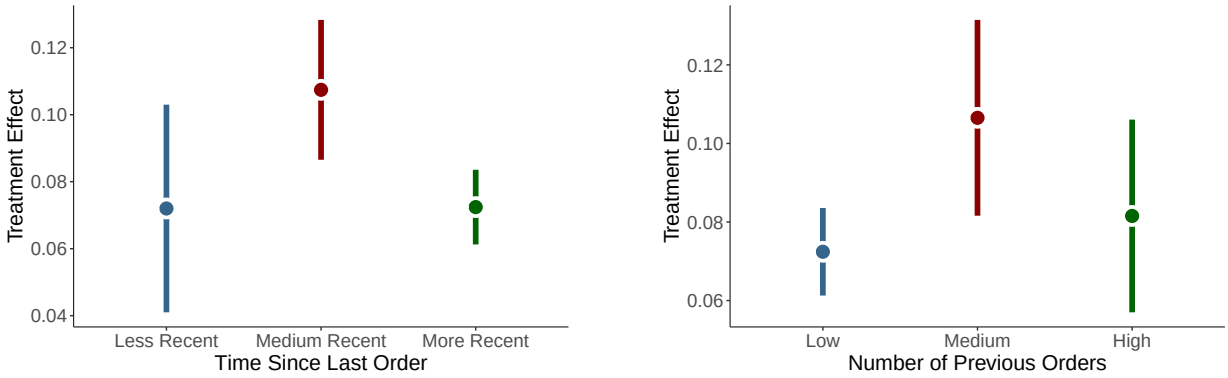
high to make the results more interpretable. We then estimate heterogeneous treatment effects across the three bins. To make exposition easier, we combine both the baseline groups (P-Image and N-Video) to create a single control group. The treatment group remains the same, AI-Video. Figure 2 shows the results for the recency and frequency measures. The results for the measure of monetary expenditure are similar to the results for purchase frequency and are shown in the Appendix.

We find that the effect of our treatment on ad clicks is strong and positive across all measures and bins. Interestingly though, we find that GenAI-based personalized ad videos engage users with medium levels of RFM measures at a higher rate. Specifically, users who have had moderate purchases and sales with the retailer in the past, i.e., neither too high nor too low, have substantially higher ad engagement rates. Similarly, users who have *not* purchased from the online retailer in the past six months (Medium Recent in the left panel), have a significantly higher engagement rate with GenAI-based ads. Typically, we find that engagement rates are highest for the most loyal customers; hence, the incremental impact of marketing interventions tends to be subdued. However, if GenAI can help create engaging personalized videos, then it would provide a valuable resource for retailers to engage these consumers. These results are in sync with prior literature on targeting which suggests that it may be more beneficial to target people with the highest sensitivity to targeting (Ascarza, 2018).

4.2 Historical Ad Engagement

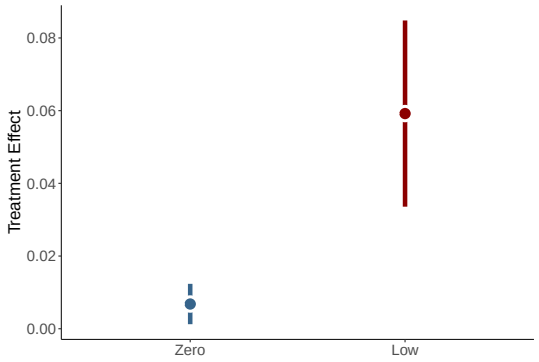
The e-commerce firm regularly engages in targeted campaigns, sending brand promotions and targeted messages to its users through WhatsApp. We use historical campaigns from March and April 2023 to segment the user base in our campaign by past ad engagement levels. Specifically, conditional on a user being sent a marketing message on WhatsApp, we segment users into three groups based on past ad engagement levels – 1) Zero engagement, 2) Low engagement, and 3) High engagement, where engagement is measured using ad clicks. Out of the 21,328 users in our experiment, all but 251 users were sent at least one WhatsApp marketing message in the two months prior to the experiment.

Among those who were targeted in the past, $\sim$57% never engaged with any of the previous campaigns. This presents a unique opportunity for us to study if these new GenAI-based personalized ad videos can improve engagement among this set of presumably "non-engaged" users. Figure 2c shows the treatment effect for users with zero and low ad engagement levels in the past. We find that for both these groups, the treatment effects are positive and significant compared to the users in control. Users who have never engaged in the past have a small but significant positive treatment response of $\sim$0.5 p.p. and users who have had limited engagement in the past have a sizable $\sim$6 p.p. increase in engagement with GenAI-based personalized ads.



(a) Recency: Time since last order

(b) Frequency: # of previous orders



(c) Engagement with previous ads

Figure 2: Heterogeneity in treatment effects by purchase recency and frequency and previous ad engagement level

5 Cost-Benefit Analysis

Our key result shows the incremental effects of targeting users with GenAI-based personalized video ads over two standard baselines – 1) personalized image ads, and 2) generic non-personalized video ads. A natural extension is comparing GenAI-personalized ads with human-personalized ads. However, creating such ads at scale is prohibitively expensive, a limitation that serves as motivation for using GenAI in the first place. We present a simplified cost-benefit analysis using *back-of-the-envelope* calculations to provide intuition for the potential productivity gains from GenAI personalization.

Consider a scenario where a firm wants to run a campaign targeting 10,000 consumers with personalized video ads featuring a celebrity endorser. We base our calculations on the Indian market but present the figures in USD for ease of comparison. The following exposition is based on our conversations with the AI-Firm, industry experts, and independent market research.[1]

[1]These cost calculations are from an advertiser's perspective. We assume that an advertiser can hire GenAI as a service and it does not have to invest in creating new GenAI models from scratch.

First, the traditional approach – hiring a celebrity typically costs ∼120,000 for a full 8-hour day, during which the celebrity can shoot ∼1,000 personalized 30-second videos[2] if they work non-stop. To make 10,000 personalized video ads, it would cost $1.2M, equating to $120 per video.[3]

Now, consider GenAI – hiring the celebrity for a few hours to record template videos plus the cost of creating a high-fidelity digital avatar totals $120,000. This is a one-time fixed cost, which we assume to be the same as the 1-day cost of working with the celebrity for simplicity. Once the digital avatar is created, the benefits of GenAI set in. Creating 10,000 personalized videos from templates costs <$10,000, making the total campaign cost ∼$130,000, which is substantially lower than human-created videos. GenAI can produce personalized video ads at nearly $1/10^{th}$ the cost ($13 vs $120). Furthermore, the benefits of GenAI compound at scale. For instance, creating 100,000 personalized video ads with GenAI would cost $220,000, compared to $12M for human-created videos.

Beyond cost savings, GenAI also offers significant productivity gains through faster iteration, greater creative flexibility, and streamlined workflows. The entire process from conceptualization to ad delivery is much faster with GenAI since the bulk of the work is automated once the templates are created. Institutionalizing GenAI into the firm's marketing stack can allow marketing managers to scale personalized campaigns more efficiently.

However, there are also some caveats to consider. Our cost-benefit estimates rely on simplifying assumptions about production processes, talent costs, and data infrastructure, and may not capture the full complexity and hidden costs of implementation at scale. Additionally, the quality and impact of GenAI videos relative to human-created ads are uncertain. Authenticity concerns Nunes et al. (2021), privacy preferences (Jin, 2018), and algorithmic aversion (Dietvorst et al., 2015; Castelo et al., 2019) could limit engagement with AI-created content. More research is needed on GenAI's strengths, limitations, and boundary conditions in different marketing contexts.

Despite these challenges, the gains on the cost and productivity fronts are clear. GenAI makes personalized video advertising accessible and scalable in a way that was previously unimaginable. To summarize our results, we find that GenAI-based ads have engagement rates that are 6-9 p.p. higher as compared to the baselines. In addition to higher engagement rates, GenAI provides substantial cost and productivity benefits over human-personalized video ads, which are prohibitive at this scale.

6 Implications and Discussion

Our paper has important implications for marketing research and practice by demonstrating the potential of GenAI in creating engaging, personalized video ads at scale, a task previously prohibitively expensive. These gains highlight GenAI's impact on core marketing processes and its ability to deliver more relevant, tailored ad experiences to consumers. As the technology matures, one can imagine a world where users browsing Instagram can see ads with personalized content directly in their feeds. However, given that there is more hype than science surrounding the potential impacts of GenAI, our

[2]This figure is based on our conversation with the AI-Firm, Geyser (2024), and WDC (2024).

[3]These per-video numbers are comparable to personalized celebrity message services such as Tring and Cameo.

encouraging results merit a more nuanced discussion. We deliberate on three important themes to position our findings in the context of broader business and policy discussions on the implications of GenAI.

1. **Novelty effects.** Our results, showing a 6-9 p.p. increase in engagement with GenAI-based ads, are from a single-exposure experiment where most users likely encountered GenAI-personalized ads for the first time. This makes our findings susceptible to novelty effects (Kohavi et al., 2020). Eventually, as the prevalence of these GenAI ads increases, the initial allure of novelty can wane, and diminishing returns due to consumer fatigue may set in. This, however, is neither a feature nor a bug of our experimental design, but just an artifact of human behavior. Our paper provides a foundation for testing GenAI in ad creation and targeting. As firms gain insights into customer preferences, they should optimize their ad load and mix to maximize long-term benefits, balancing novel GenAI ads with more traditional formats to maintain engagement.

2. **Privacy concerns.** Concerns about privacy, data ownership, and regulation dominate debates on online advertising. Our study sidestepped these direct constraints by utilizing the retailer's first-party data customers to create targeted advertisements. This, however, does not preclude that there may still be perceived concerns about privacy among the customers, a topic that we think is important to study more deeply in itself. Nevertheless, at the very least, access to first-party data allows the initial exploration of GenAI in creating engaging personalized ads. As the market for AI-generated content matures and access to tools democratizes, firms should actively engage in field experiments, consumer surveys, and controlled lab studies to better understand consumers' perceptions of this content. Firms must balance hyper-personalization with transparency around data usage.

3. **Scalability and generalizability.** On one hand, GenAI's ability to create customizable video content at scale led the retailer to explore this space for personalized ads. On the other hand, as the volume of AI-generated content increases and inundates our online feeds, ensuring consistent quality and brand alignment can become challenging. We approach this impending trade-off cautiously by leveraging Generative AI in close partnership with the retailer's marketing team. The marketing team designed the creative templates and the AI is then used to personalize them for thousands of customers in a guided fashion. Human-AI collaboration certified quality consistency and granular first-party data ensured relevance for targeted users. However, both these features limit the range and diversity of content, potentially leaving out broader audiences for targeting. As brands seek to scale this approach for other types of ad campaigns, such as new customer acquisition or competitor targeting, it will be imperative to build safeguards that certify content quality and fidelity.

7 Conclusions

Advertising continues to evolve, and the introduction of Generative AI (GenAI) in crafting personalized video advertisements demonstrates a promising advancement. In this paper, we study the effectiveness

of Generative AI in the context of video advertisements for targeting. We partner with a leading Generative AI firm based in the US and India to create personalized video advertisements using Generative AI for an online retailer that sells eco-friendly sustainable products in India. Specifically, we use GenAI to first create a high-fidelity digital avatar of the retailer's brand ambassador. We then use the AI-Firm's proprietary text-to-video generative AI engine to personalize the content the ambassador says for each customer.

To test these ads, we run a mobile targeting field experiment on WhatsApp where we randomize the online retailer's customers into receiving marketing messages from one of three conditions – 1) GenAI-based personalized video ads, 2) personalized image ads, and 3) generic video ads. The first group is our treatment. The latter two are the baseline groups that allow us to control for the industry status quo in mobile targeting (personalized images) and the medium of the message (videos) respectively.

We find that the engagement rate, as measured by ad clicks, in the treatment group is 6-9 p.p. is higher than that of the two control groups. The effect is robust across demographics including gender and the state of residence. Further, we estimate heterogeneous treatment effects across two important dimensions of consumer behavior – 1) RFM, and 2) historical ad engagement. We find encouraging results for both dimensions. For RFM, we find that users who have had "moderate" purchases and spend with the retailer have larger treatment effects as compared to users with low or high levels of past transactions. We believe that this provides a new opportunity for firms to engage with these potentially "high-value" customers. Furthermore, we find that GenAI-based personalized video ads are able to engage historically "non-engaged" users at a much higher rate as compared to other targeting strategies.

To the best of our knowledge, we provide the first set of real-world causal results on the effectiveness of Generative AI as an important technology for video-based marketing. The findings provide timely, actionable insights for marketers as video and AI shape modern advertising. We think that two direct avenues flow from this paper – 1) studying the performance of GenAI-based ads on "down-the-funnel" metrics such as purchases and revenue, and 2) expanding the scope of GenAI in the ad creation process. More broadly, an important companion to this paper, and related papers studying the impact of this technology, would be a study on the hidden, and potentially negative, effects of Generative AI. This includes concerns for privacy and data ownership from the consumer's side and concerns of skill-biased technological change on the labor side. Our results offer an initial blueprint for deploying these capabilities responsibly. But given GenAI's nascency, ongoing research must continue balancing potential benefits against costs. We believe that the findings of this paper provide insightful action items for practitioners. Finally, we hope that the ideas presented here provide fertile ground for further exploration by researchers.

Disclosure Statement

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

References

Arora, Neeraj, Xavier Dreze, Anindya Ghose, James D. Hess, Raghuram Iyengar, Bing Jing, Yogesh Joshi, V. Kumar, Nicholas Lurie, Scott Neslin, S. Sajeesh, Meng Su, Niladri Syam, Jacquelyn Thomas, and Z. John Zhang, "Putting one-to-one marketing to work: Personalization, customization, and choice," *Marketing Letters*, Dec 2008, *19* (3), 305–321.

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

Blake, Thomas, Chris Nosko, and Steven Tadelis, "Consumer heterogeneity and paid search effectiveness: A large-scale field experiment," *Econometrica*, 2015, *83* (1), 155–174.

Bleier, Alexander and Maik Eisenbeiss, "Personalized online advertising effectiveness: The interplay of what, when, and where," *Marketing Science*, 2015, *34* (5), 669–688.

Boussioux, Leonard, Jacqueline N. Lane, Miaomiao Zhang, Vladimir Jacimovic, and Karim R. Lakhani, "The Crowdless Future? How Generative AI Is Shaping the Future of Human Crowdsourcing," *SSRN*, Aug 2023.

Brand, James, Ayelet Israeli, and Donald Ngwe, "Using GPT for market research," Mar 2023.

Brynjolfsson, Erik, Danielle Li, and Lindsey R. Raymond, "Generative AI at Work," *NBER Working Paper*, April 2023.

Burnap, Alex, John R Hauser, and Artem Timoshenko, "Product aesthetic design: A machine learning augmentation," *Marketing Science*, 2023.

Castelo, Noah, Maarten W Bos, and Donald R Lehmann, "Task-dependent algorithm aversion," *Journal of Marketing Research*, 2019, *56* (5), 809–825.

Dew, Ryan, Asim Ansari, and Olivier Toubia, "Letting Logos Speak: Leveraging Multiview Representation Learning for Data-Driven Branding and Logo Design," *Marketing Science*, 2022, *41* (2), 401–425.

Dietvorst, Berkeley J, Joseph P Simmons, and Cade Massey, "Algorithm aversion: people erroneously avoid algorithms after seeing them err.," *Journal of Experimental Psychology: General*, 2015, *144* (1), 114.

Dzyabura, Daria, Siham El Kihal, John R. Hauser, and Marat Ibragimov, "Leveraging the Power of Images in Managing Product Return Rates," *Marketing Science*, 2023.

Economist, Feb 2024.

Fong, Hortense, Vineet Kumar, and K. Sudhir, "A Theory-Based Interpretable Deep Learning Architecture for Music Emotion," *SSRN*, 2022.

Geyser, Werner, "Influencer rates: How much do influencers really cost in 2024?," Jan 2024.

Goldfarb, Avi and Catherine E Tucker, "Privacy regulation and online advertising," *Management Science*, 2011, *57* (1), 57–71.

Goodrich, Kendall, Shu Z. Schiller, and Dennis Galletta, "Consumer Reactions to Intrusiveness Of Online-Video Advertisements," *Journal of Advertising Research*, 2015, *55* (1), 37–50.

Gupta, Sunil, Dominique Hanssens, Bruce Hardie, Wiliam Kahn, V. Kumar, Nathaniel Lin, Nalini Ravishanker, and S. Sriram, "Modeling Customer Lifetime Value," *Journal of Service Research*, 2006, *9* (2), 139–155.

Hoban, Paul R. and Randolph E. Bucklin, "Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment," *Journal of Marketing Research*, 2015, *52* (3), 375–393.

Hong, Jiyeon and Paul R. Hoban, "Writing More Compelling Creative Appeals: A Deep Learning-Based Approach," *Marketing Science*, 2022, *41* (5), 941–965.

Jin, Ginger Zhe, "Artificial intelligence and consumer privacy," in "The economics of artificial intelligence: An agenda," University of Chicago Press, 2018, pp. 439–462.

Johnson, Garrett A., "Inferno: A guide to field experiments in online display advertising," *Journal of Economics & Management Strategy*, 2023, *32* (3), 469–490.

Johnson, Garrett A, Scott K Shriver, and Shaoyin Du, "Consumer privacy choice in online advertising: Who opts out and at what cost to industry?," *Marketing Science*, 2020, *39* (1), 33–51.

Kalyanam, Kirthi, John McAteer, Jonathan Marek, James Hodges, and Lifeng Lin, "Cross channel effects of search engine advertising on brick & mortar retail sales: Meta analysis of large scale field experiments on Google. com," *Quantitative Marketing and Economics*, 2018, *16*, 1–42.

Kohavi, Ron, Diane Tang, and Ya Xu, *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*, Cambridge University Press, 2020.

Lambrecht, Anja and Catherine Tucker, "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, 2013, *50* (5), 561–576.

Lewis, Randall A. and Justin M. Rao, " The Unfavorable Economics of Measuring the Returns to Advertising *," *The Quarterly Journal of Economics*, 07 2015, *130* (4), 1941–1973.

Li, Hao and Hui-Yi Lo, "Do You Recognize Its Brand? The Effectiveness of Online In-Stream Video Advertisements," *Journal of Advertising*, 2015, *44* (3), 208–218.

Li, Peiyao, Noah Castelo, Zsolt Katona, and Miklos Sarvary, "Language Models for Automated Market Research: A New Way to Generate Perceptual Maps," *Available at SSRN 4241291*, 2022.

Longoni, Chiara, Andrea Bonezzi, and Carey K Morewedge, "Resistance to medical artificial intelligence," *Journal of Consumer Research*, 2019, *46* (4), 629–650.

Luo, Xueming, Siliang Tong, Zheng Fang, and Zhe Qu, "Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases," *Marketing Science*, 2019, *38* (6), 937–947.

Nicoletti, Leonardo and Dina Bass, "Humans are Biased. Generative AI is Even Worse.," Jun 2023.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

Nunes, Joseph C., Andrea Ordanini, and Gaia Giambastiani, "The Concept of Authenticity: What It Means to Consumers," *Journal of Marketing*, 2021, *85* (4), 1–20.

Rafieian, Omid, "Optimizing user engagement through adaptive ad sequencing," *Marketing Science*, 2022.

Reisenbichler, Martin, Thomas Reutterer, and David A. Schweidel, "Applying Large Language Models to Sponsored Search Advertising," *MSI Working Paper Series*, 2023.

_ , _ , _ , and Daniel Dan, "Frontiers: Supporting Content Marketing with Natural Language Generation," *Marketing Science*, 2022, *41* (3), 441–452.

Robertson, Adi, "Google apologizes for "missing the mark" after Gemini generated racially diverse Nazis," Feb 2024.

Sahni, Navdeep S. and Harikesh S. Nair, "Sponsorship Disclosure and Consumer Deception: Experimental Evidence from Native Advertising in Mobile Search," *Marketing Science*, 2020, *39* (1), 5–32.

Sahni, Navdeep S, S Christian Wheeler, and Pradeep Chintagunta, "Personalization in email marketing: The role of noninformative advertising content," *Marketing Science*, 2018, *37* (2), 236–258.

Sahni, Navdeep S., Sridhar Narayanan, and Kirthi Kalyanam, "An Experimental Investigation of the Effects of Retargeted Advertising: The Role of Frequency and Timing," *Journal of Marketing Research*, 2019, *56* (3), 401–418.

Timoshenko, Artem and John R Hauser, "Identifying customer needs from user-generated content," *Marketing Science*, 2019, *38* (1), 1–20.

Tong, Siliang, Nan Jia, Xueming Luo, and Zheng Fang, "The Janus face of artificial intelligence feedback: Deployment versus disclosure effects on employee performance," *Strategic Management Journal*, 2021, *42* (9), 1600–1631.

Tucker, Catherine E., "Social Networks, Personalized Advertising, and Privacy Controls," *Journal of Marketing Research*, 2014, *51* (5), 546–562.

_ , "The Reach and Persuasiveness of Viral Video Ads," *Marketing Science*, 2015, *34* (2), 281–296.

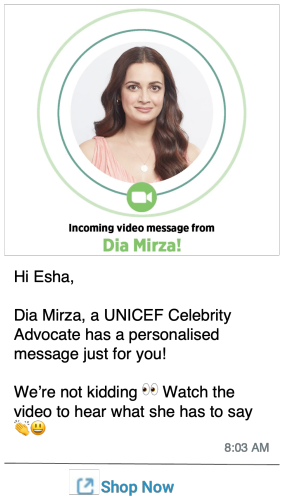
WDC, "Influencer marketing costs in India 2024: A comprehensive guide," Feb 2024.

Zhang, Juanjuan, "The perils of behavior-based personalization," *Marketing Science*, 2011, *30* (1), 170–186.

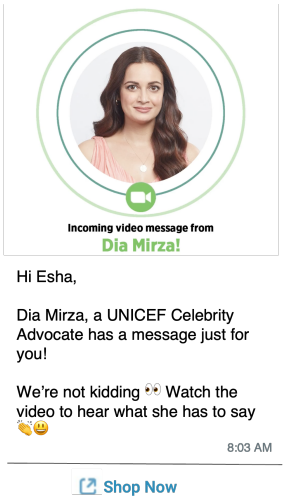
Online Appendix:
Generative AI and Personalized Video Advertisements

A Ad Samples

Consider a representative consumer of the online retailer, Esha. Figures A1a, A1b, and A2 show the ad copies Esha would receive if she were in the GenAI personalized video treatment (AI-Video), the generic non-personalized treatment (N-Video), or the personalized image treatment (P-Image) respectively. The first is our main treatment and the remaining two are the baseline treatment arms. Across all conditions, the ads feature the same brand ambassador of the online retailer.



(a) Sample Personalized Video Ad (Treatment)



(b) Sample Generic Video Ad (Baseline)

Figure A1: Sample Ads from Treatment and Generic Video Baseline

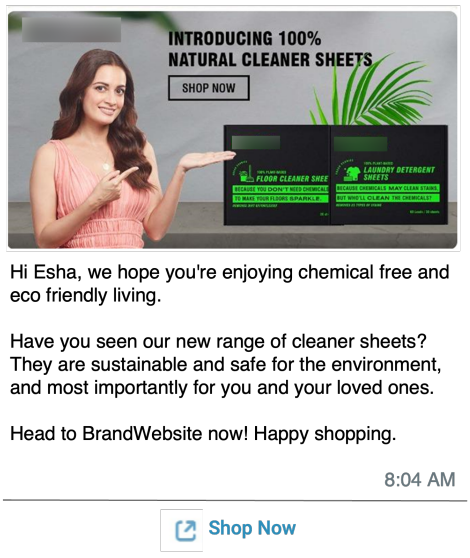


Figure A2: Sample Personalized Image Ad (Baseline)

Figure A3 shows the screenshots from one of the GenAI-based personalized video messages that upsell the customer with a new product, *Cleaner Sheets* that she has not purchased before. The GenAI engine personalizes two parts of the video content – 1) the *name* of the customer and 2) the *product* that the brand ambassador is promoting. The technology firm's proprietary text-to-video GenAI engine ensures high fidelity creation of such personalized videos.

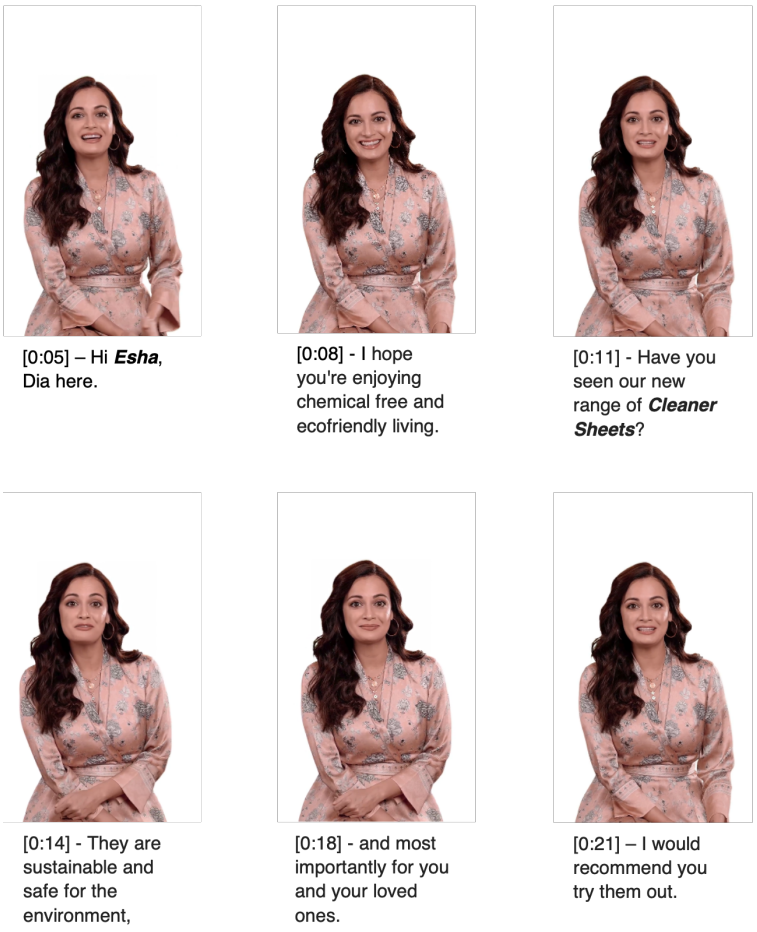


Figure A3: Screenshots from GenAI-based personalized video with the 'Upsell' message

Figure A4 shows the screenshots from the generic non-personalized video ad. Note that no part of the content here is personalized. All users in this group (N-Video) receive the same message recorded live by the brand ambassador.

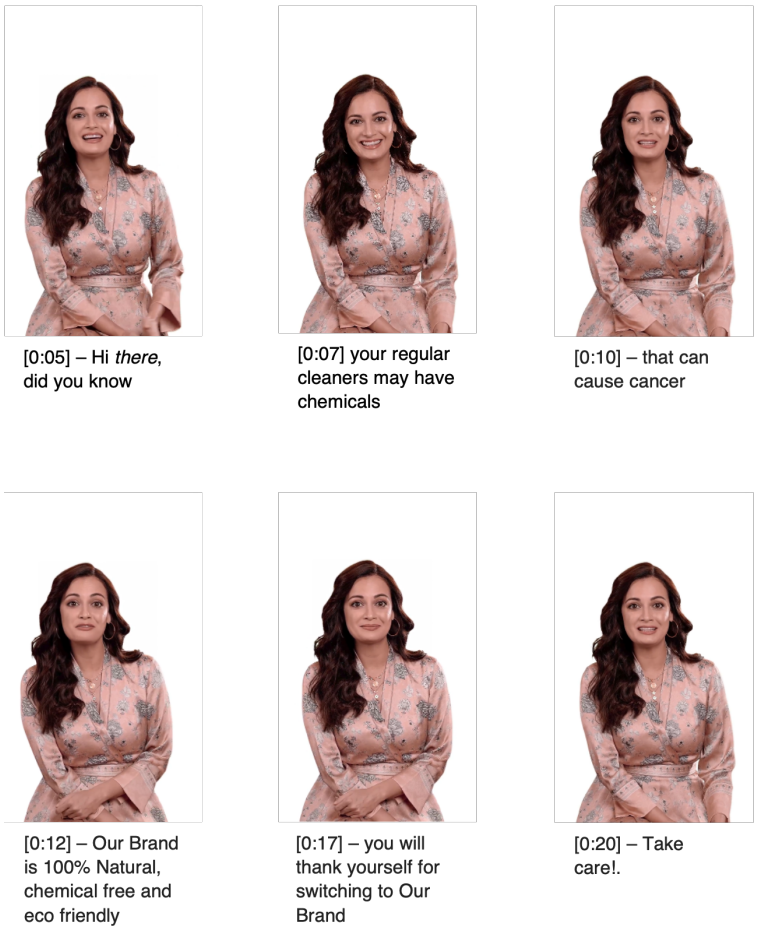


Figure A4: Screenshots from the generic non-personalized video ad

B Geographic Distribution of Participants

To ensure the generalizability of our experiment, we target a random sample of the online retailer's customers from all over India. Figure B5 shows the percentage of customers from different states. Maharashtra, Karnataka, Tamil Nadu, Telangana, and Delhi are the top 5 states in terms of participants in the experiment.

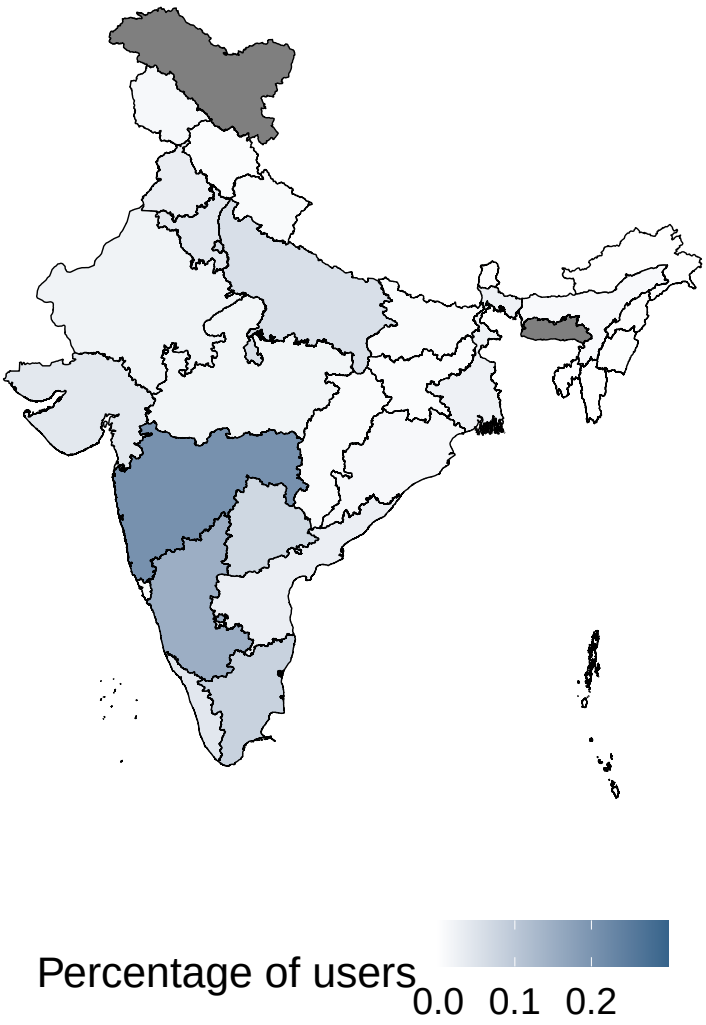


Figure B5: Distribution of participants by state

Figure B6 shows the percentage of customers from the top 20 cities. These are among the biggest cities in India in terms of population.

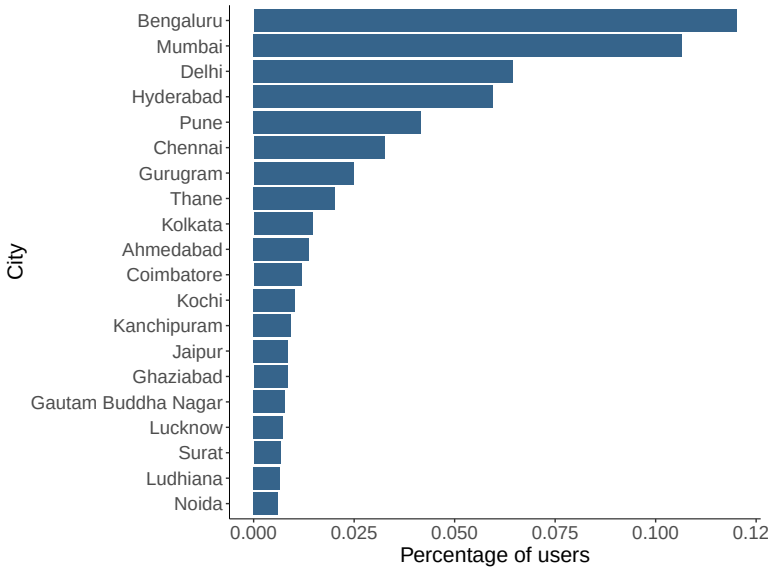


Figure B6: Distribution of experiment participants by city

C Robustness of ATE to model specification

We test the robustness of our main results (shown in Table 2) to alternative specifications. Table C1 estimates the treatment effects using logistic regression. The dependent variable is a binary indicator for ad click. The key independent variable is the binary treatment indicator of whether the user was targeted using a GenAI-based personalized video ad or not. Column 1 shows the treatment effect compared to the personalized image baseline and column 2 shows the treatment effect compared to the generic non-personalized video baseline. Columns 3 and 4 include demographic and behavioral controls for further robustness checks. Across all specifications, we find that the treatment effect is consistent.

Table C1: Robustness to model specification: treatment effect of Generative AI-based personalized video ads on engagement using Logistic regression

Dependent Variable:	Ad Click			
Baseline	P-Image	N-Video	P-Image	N-Video
Model:	(1)	(2)	(3)	(4)
Variables				
(Intercept)	-2.258***	-1.964***	-1.904***	-1.758***
	(0.0445)	(0.0396)	(0.1262)	(0.1277)
Treatment	0.7960***	0.5021***	0.8020***	0.5059***
	(0.0516)	(0.0475)	(0.0518)	(0.0477)
Gender (F=1)			-0.1614***	-0.1641***
			(0.0467)	(0.0452)
Total spend (last 1 year)			0.0285***	0.0358***
			(0.0072)	(0.0069)
State			✓	✓
Fit statistics				
Observations	15,417	15,421	15,417	15,421
Log-Likelihood	-6,447.6	-6,801.8	-6,408.4	-6,762.9

*Each column shows the estimates from a logistic regression of ad click on a binary treatment indicator (plus controls). The baseline/control in columns 1 and 3 is the personalized image group. The baseline/control in columns 2 and 4 is the generic non-personalized video group. IID standard errors in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table C2 replicates the results above but combines all the treatments into a single categorical variable that takes on three values: 1) P-Image (Intercept), 2) N-Video, and 3) AI-Video. We then run the following regression:

$$y_i = \alpha + \beta_1 \times V_i + \beta_2 \times (P_i \times V_i) + \epsilon_i \tag{2}$$

where α, the intercept is the baseline engagement rate in the P-Image condition, β_1 captures the effect of receiving a video message, V_i (N-Video in our case), and β_2 captures the effect of a personalized P_i video message V_i given by $P_i \times V_i$ (AI-Video in the experiment).

Table C2: Treatment effect of personalized ad videos created using Generative-AI on engagement

Dependent Variable:	Ad Click			
Model:	(1) OLS	(2) OLS	(3) Logit	(4) Logit
Variables				
(Intercept)	0.0946***	0.1329***	-2.258***	-1.972***
	(0.0045)	(0.0146)	(0.0445)	(0.1143)
N-Video	0.0284***	0.0283***	0.2939***	0.2948***
	(0.0064)	(0.0064)	(0.0595)	(0.0597)
AI-Video	0.0935***	0.0937***	0.7960***	0.8010***
	(0.0058)	(0.0058)	(0.0516)	(0.0518)
Gender (F=1)		-0.0204***		-0.1669***
		(0.0050)		(0.0405)
Hist. total sales		0.0038***		0.0304***
		(0.0008)		(0.0062)
State		✓		✓
Fit statistics				
Observations	21,328	21,328	21,328	21,328
R^2	0.01350	0.01739		
Log-Likelihood	-7,804.8	-7,762.7	-8,651.4	-8,608.4

*The outcome variable is binary, 'ad-click' and the treatment variable is categorical which takes three values – 1) P-Image (Intercept), 2) N-Video, and 3) AI-Video. Columns 2 and 4 include controls for demographics and historical sales over the past year with the online retailer. IID Standard Errors are in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

We further estimate the robustness of our results using Fisherian Randomization. We then test whether the treatment had any impact by randomly permuting the treatment indicator and re-estimating the treatment effect for each permutation. We do this separately for both control conditions. Figure C7 shows the histogram of treatment effects estimated across the permutations along with the observed treatment effect estimated using the realized treatment assignment. These permutations take the N-Video arm as the control and AI-Video as treatment. We find that the observed estimate is significantly far from the distribution of permuted estimates. Figure C8 repeats the analysis with P-Image as the control. We see similar results in this case as well.

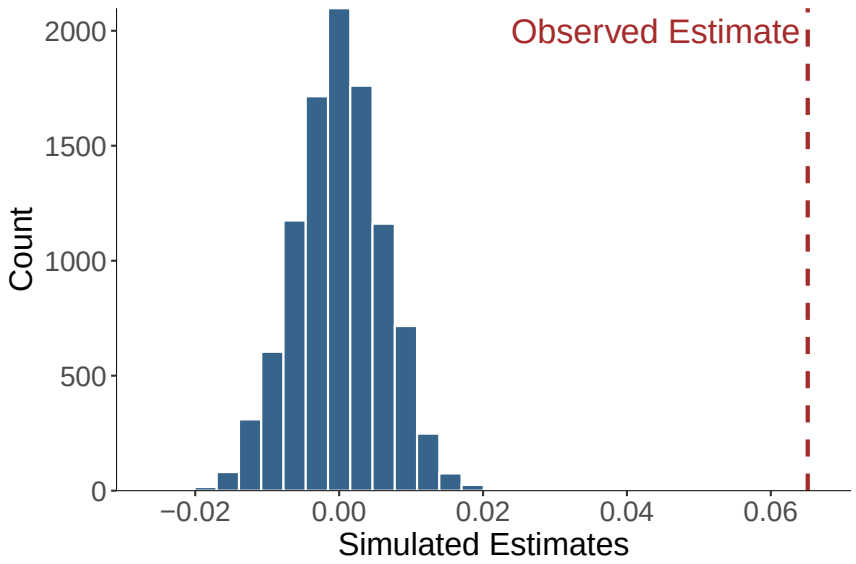


Figure C7: Fisher's Randomization Test with N-Video as Control

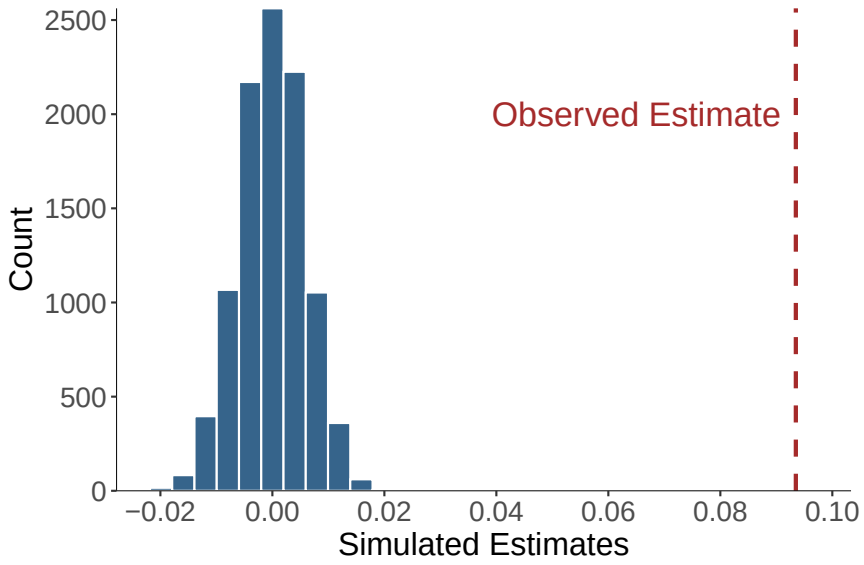
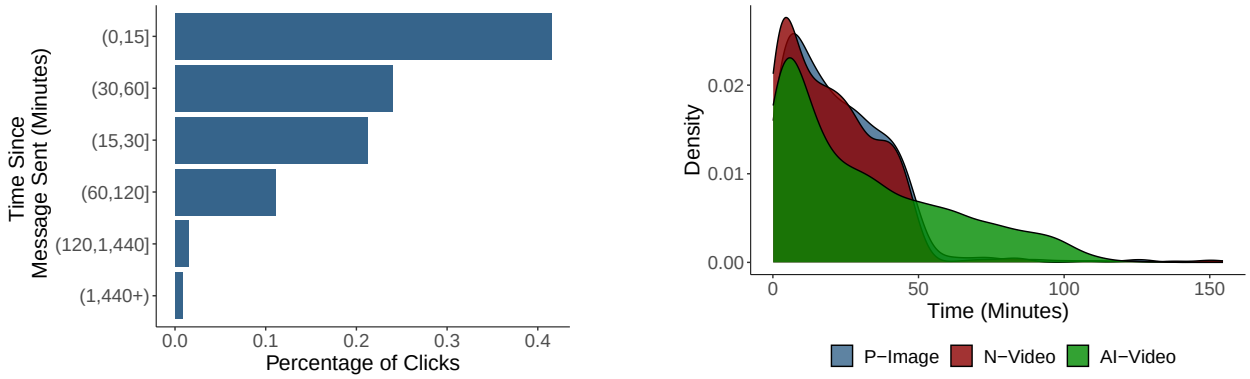


Figure C8: Fisher's Randomization Test with P-Image as Control

D Time to click

We explore heterogeneity in the time users take to click on the ad after receiving the message on WhatsApp, conditional on the user clicking. There are two points worth highlighting. First, over 95% of the clicks across all arms happen within two hours of receiving a message. Figure D9a shows the percentage of clicks received in different time windows after receiving the message. This is an important artifact of WhatsApp-based marketing.

Second, we don't find a significant difference in the distribution of time to click. Figure D9b plots the density in time to click (conditional on clicking) for the three experimental conditions. Although the main treatment group shows a slightly fatter tail, it fizzles out at the 2-hour mark. It is important to note that these results cannot be interpreted causally since they condition on a post-treatment variable. Nevertheless, we find these results helpful in better understanding WhatsApp-based marketing strategies where the literature is limited.



(a) Proportion of ad clicks within a given time frame

(b) Time to click after receiving the message

Figure D9: Timing of ad clicks after receiving the message on WhatsApp

In summary, we find that GenAI-based personalized video ads have a substantially higher engagement rate as compared to non-personalized video ads and personalized image ads. The effect is most pronounced for ads up-selling new products to consumers who recently purchased from the retailer. Further, there is no substantial difference in the time it takes to click on the ad (conditional on clicking) after receiving the message.

E Robustness of Treatment Effect Across Demographics

We test for any differential impact of the treatment across consumer demographics – gender and state of residence. We estimate heterogeneous treatment effects across genders by interacting the gender dummy (which takes the value 1 when the gender is female) with the treatment indicator. The results are shown in Figure E10. We find no substantial difference in the impact across male and female users.

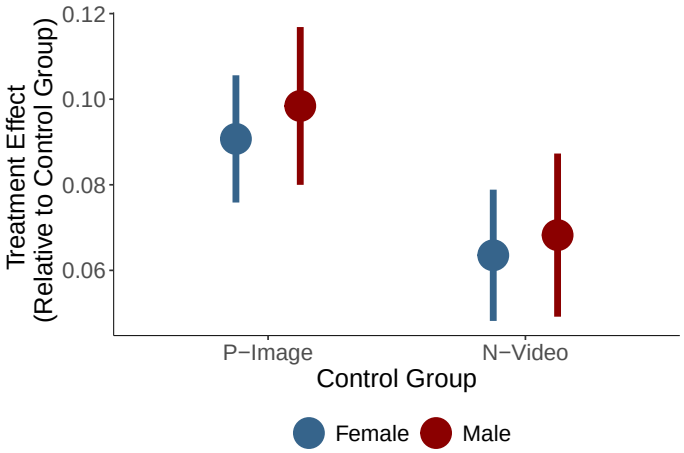


Figure E10: Heterogeneity by gender

State of residence is a high-dimensional categorical variable, with 33 unique values. We test for heterogeneity using a random effects regression that allows for partial pooling of information across states. We estimate the treatment effects separately with the two baselines and we allow them to vary by state. We estimate the model using Restricted Maximum Likelihood. Treatment effects across states are shown in Figure E11. While there is some heterogeneity across states, they are largely consistent and vary around the global mean as shown by the dashed lines in the plot.

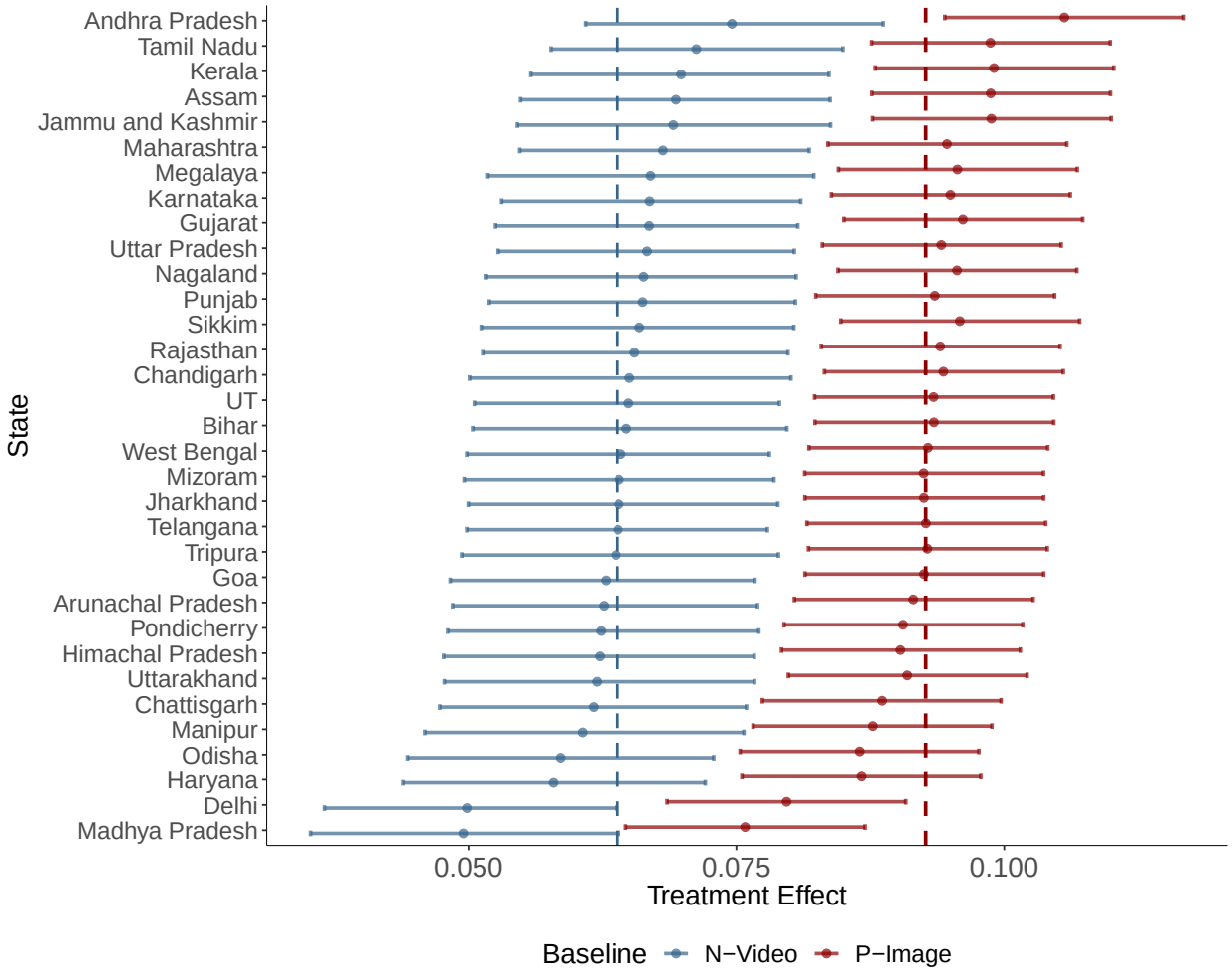


Figure E11: Treatment effects across all Indian states

F Performance Across Sub-Conditions

We split the aggregate results across the three sub-conditions described in Figure 1, i.e., 1) *Upsell*, 2) *Revisit*, and 3) *Switch*. It is worth noting that while the experiment treatment is randomly assigned and orthogonal to each sub-condition, assignment to a specific sub-condition is *not* random. The whole premise of personalization is based on sending users targeted messages based on their historical involvement with the retailer. Hence, we cannot directly compare the treatment effects across sub-conditions. For example, we cannot causally claim that AI-Video treatment is more effective in the case of *revisit* messages as compared to *switch* messages. This effect could be, in part, driven by the different mix of users assigned to each sub-condition. However, we can compare AI-Video treatment with N-Video and P-Image, within each sub-condition since conditional on being in a sub-condition, users are randomly assigned to one of the three treatment groups.

Figure F12 shows the treatment effects across the different sub-conditions. We find that within each sub-condition, GenAI-based videos significantly outperform personalized image ads and generic non-personalized video ads.

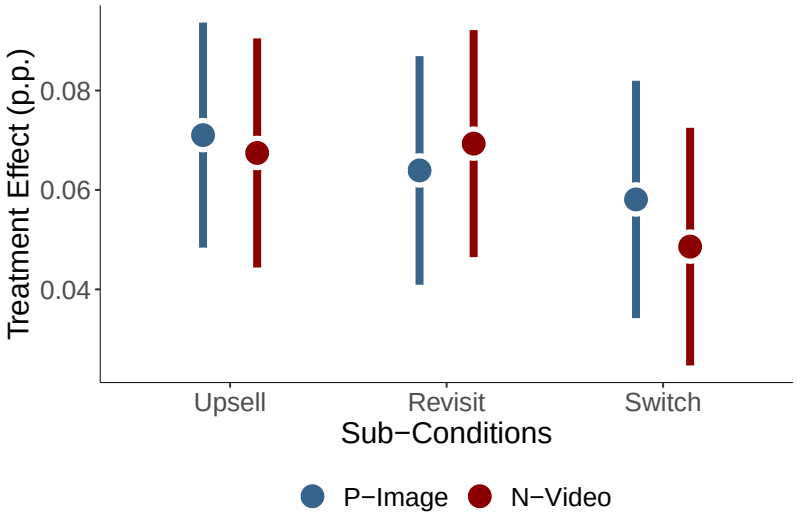


Figure F12: Treatment effect across different sub-conditions

G HTE Based on RFM Measures

We estimate heterogeneous treatment effects across user segments based on pre-treatment Recency, Frequency, and Monetary (RFM) measures. We use one year of historical data to calculate user-level RFM measures and then segment users into three bins based on each of these measures – 1) Low, 2) Medium, and 3) High. For each measure, we estimate heterogeneity in consumer response using OLS. To make exposition easier, we combine both the baseline groups into one big control group. The treatment group is the same, GenAI-based personalized video treatment

The results are shown in Table G3. We estimate the treatment effect separately for each user bin and measure. Columns 1-3 present the treatment effects for users classified into Low, Medium, and High bins based on Recency (R). Columns 4-6 repeat this analysis for Frequency (F), and Columns 7-9 show results for Monetary (M). For all three measures, we find that users in the middle of the distribution, that is users in the "medium" bin have significantly higher engagement with GenAI-based ads as compared to users in the other two bins.

Table G3: Heterogeneity in Treatment Effects Across User Segments Based on RFM Measures

Dependent Variable:					**Ad Click**				
Measure:		**Recency**			**Frequency**			**Monetary**	
Bin:	Low	Medium	High	Low	Medium	High	Low	Medium	High
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Variables									
(Intercept)	0.1611***	0.0944***	0.1039***	0.1039***	0.1178***	0.1210***	0.1040***	0.1163***	0.1223***
	(0.0105)	(0.0072)	(0.0038)	(0.0038)	(0.0085)	(0.0084)	(0.0038)	(0.0085)	(0.0085)
Treatment	0.0720***	0.1074***	0.0724***	0.0724***	0.1065***	0.0815***	0.0724***	0.1098***	0.0785***
	(0.0158)	(0.0106)	(0.0057)	(0.0057)	(0.0127)	(0.0125)	(0.0057)	(0.0127)	(0.0126)
Fit statistics									
Observations	2,504	4,262	14,562	14,562	3,383	3,383	14,590	3,369	3,369
R^2	0.00821	0.02333	0.01101	0.01101	0.02033	0.01239	0.01101	0.02160	0.01149
Log-Likelihood	-1,214.1	-1,522.4	-4,999.2	-4,999.2	-1,412.3	-1,365.9	-5,012.5	-1,402.6	-1,362.3

*This table shows heterogeneous treatment effects for users based on RFM measures. We segment users into 3 bins for each RFM measure and then estimate the treatment effect separately for each bin. Columns 1-3 show the treatment effect for users classified into Low, Medium, and High bins based on the recency of purchase with the online retailer. All columns show results from estimating a linear probability model using OLS. Both baseline groups are combined into a single control group. The treatment group is GenAI-based personalized video ads. IID Standard Errors are in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

H Treatment Effect by Historical Ad Engagement Levels

The e-commerce firm regularly engages in targeted campaigns, sending brand promotions and targeted messages to its users through WhatsApp. We use historical campaigns from March and April 2023 to segment the user base in our campaign by past ad engagement levels. Specifically, conditional on a user being sent a marketing message on WhatsApp, we segment users into three groups based on past ad engagement levels – 1) Zero engagement, 2) Low engagement, and 3) High engagement, where engagement is measured using ad clicks. The segmentation is shown in Table H4.

Table H4: User Segmentation Based on Historical Ad Engagement

Historical Ad Engagement	# Users	Historical Ad Clicks (Mean)
Unmatched	251	-
High	6,366	2.31
Low	2,726	1.00
Zero	11,985	0.00

The "Unmatched" group includes users who were not sent any marketing message from the firm in March and April 2023.

We then estimate the treatment effect by comparing our main treatment, GenAI-based personalized video ads, to a joint control group that combines both baselines. The results are shown in Table H5.

Table H5: Treatment effect split by engagement levels with past campaigns

Dependent Variable:	Ad Click		
Past Ad Engagement	Zero	Low	High
Model:	(1)	(2)	(3)
Variables			
(Intercept)	0.0214***	0.1072***	0.2689***
	(0.0019)	(0.0087)	(0.0078)
Treatment	0.0068**	0.0592***	0.2427***
	(0.0028)	(0.0131)	(0.0118)
Fit statistics			
Observations	11,985	2,726	6,366
R^2	0.00048	0.00748	0.06179
Log-Likelihood	5,384.8	-914.96	-4,209.7

*This table shows heterogeneous treatment effects for users based on engagement with past campaigns. We segment users into 3 bins based on their historical ad engagement rate. Column 1 shows the treatment effect for users who have never engaged with any past targeted campaign, conditional on being sent at least one targeted ad on WhatsApp. Columns 2 and 3 show treatment effects for users with low and high past engagement levels. All columns show results from estimating a linear probability model using OLS. Both baseline groups are combined into a single control group. The treatment group is GenAI-based personalized video ads. IID Standard Errors are in parentheses. Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*