INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 4 Issue 1, 2025

THE IMPACT OF GENERATIVE AI ON PERSONALIZED CONTENT MARKETING IN E-COMMERCE

Danyal Mohiuddin [1], Muhammad Hamza Tariq [2], Areej Tahir [3]

DOI: https://doi.org/10.63544/ijss.v4i1.288

Affiliations:

[1] University of Northampton, United Kingdom
Email: danyal.m.05@gmail.com

[2] AI and Digital Optimisation Researcher, University of Worcester, United Kingdom
Email: hamza.che102@gmail.com

[3] AI and Digital Optimisation Researcher, BPP University London, United Kingdom
Email: areejt217@gmail.com

Corresponding Author's Email:

[1] danyal.m.05@gmail.com

Article History:
Received: 28.02.2025
Accepted: 20.03.2025
Published: 31.03.2025

Abstract

This research investigates the effectiveness of Generative Artificial Intelligence (GenAI) in creating hyper-personalized marketing content compared to traditional rule-based personalization. While GenAI offers unprecedented scale, empirical evidence regarding its impact on consumer behaviour and ethical perception remains limited. Using a sequential mixed-methods design, a randomized A/B experiment (N = 1,140) compared GenAI-driven personalization (GPT-4 and DALL-E 3) against traditional rule-based methods in a simulated retail environment. Post-experiment surveys ($n = 480$) assessed perceived authenticity, privacy concerns, and AI detection accuracy. Qualitative insights were gathered through semi-structured interviews with six marketing managers from three e-commerce brands.

GenAI personalization significantly outperformed rule-based methods, increasing click-through rates by 35.2% (24.6% vs. 18.2%; p = 0.005) and conversion rates by 38.1% (11.6% vs. 8.4%; p = 0.032). Time-on-site increased by 16.3 seconds (Cohen's d = 0.49$). However, AI disclosure negatively impacted brand authenticity (d = -0.64), trust (d = -0.65), and purchase intention (d = -0.55). While 77.5% of consumers desired an opt-out setting for AI personalization, they were unable to reliably distinguish AI from human content (accuracy = 53.3%). Qualitative data highlighted "human-in-the-loop" processes as essential to mitigating hallucination risks.

The study contributes the first direct empirical benchmark of GenAI over rule-based personalization in e-commerce. It identifies a "detection-disclosure paradox"—where consumers cannot recognize AI content but remain suspicious when it is disclosed—and proposes a Transparent Hybrid Model to balance operational efficiency with ethical protection and brand trust.

Keywords: Generative AI, Personalization, E-Commerce, Content Marketing, Consumer Trust, LLM, Marketing Ethics, Authenticity

1. Introduction

1.1 Background

Personalization has been an established part of the e-commerce industry to facilitate customer experience, decrease search friction, and spur conversions. In a more and more congested digital market, the personalized content, whether it is product recommendations or customized email subject lines, are no longer a luxury but a competitive imperative (Smith & Linden, 2017). Traditional rule-based personalization systems have been used as a standard in the industry over a decade. These systems consist of collaborative filtering (customers who have previously purchased X also purchased Y), content-based filtering (matching items to previous likes of a user), and plain deterministic if-then logic, have made e-commerce sites able to provide relevant content to scale (Ricci et al., 2015).

Although they are widely adopted, rule-based approaches are limited in their essence. First, they are not creative: the information that they generate is not created but rather a collection of ready-made templates or data patterns of the past. Second, they are unable to generate truly new content, like an original product description that was written specifically to a single user or a custom banner image that is based on a set of aesthetic preferences of an individual. Third, systems governed by rules fail in the face of subtle context and serendipitous discovery--they do not know how to make unarticulated wants or provide unexpected but pleasing suggestions (Mohiuddin & Farhan, 2025; Huang & Rust, 2021). Therefore, the most advanced rule-based personalization may become mechanical, repetitive and, eventually, impersonal.

The development of generative artificial intelligence (GenAI) introduces a paradigm shift in personalization of marketing. Contrary to rule-based systems, which look up and combine existing material, GenAI models, such as large language models (LLMs) like GPT-4 and image generators like DALL·E 3, generate new content dynamically based on the context. An LLM may come up with a distinct email subject in a natural, conversational way that mentions the recent browsing history of a user. A diffusion model is capable of producing a personalised banner image with a product in the colour and environment that a user likes. Collectively, these features allow e-commerce sites to provide hyper-personalized content at scale, adjusting in real time to individual users, without using pre-written templates (Jarek & Mazurek, 2023). The first users are Amazon, which uses GenAI to create text summaries of products, and Shopify, which provides millions of merchants with AI-powered description services. However, in spite of such a fast commercial adoption, the scholarly knowledge on the real performance of GenAI in e-commerce personalization is still in its infancy.

1.2 Problem Statement

The disconnect between research and practice in the industry is enormous. The e-commerce giants are putting significant investments in GenAI as a means of personalization, but empirical studies comparing GenAI-based personalization to the conventional rule-based approaches are notably missing. The majority of the current literature considers either rule-based personalization alone or consumer reactions to AI-generated content not as part of a personalization environment. Not many, or even no one, has directly contrasted the two methods on similar grounds with actual e-commerce situations.

The result of this gap is three critical questions, which are not answered, which are both theoretical and practical in nature: First, performance: Does GenAI personalization actually perform better than rule-based systems on bottom-line measures like click-through rates (CTR), conversion rates, and time-on-site? Or is the freshness of GenAI a veil of concealed constraints in integrity and applicability? The absence of empirical standards would mean that marketers would implement GenAI on a hype-driven basis instead of being evidence-based. Second, consumer perception costs: What are the psychological and relational costs of AI-generated marketing content? Previous studies on algorithm aversion have indicated that consumers might not trust AI-generated content (Dietvorst et al., 2015), and research on the so-called AI-content gap has shown that labelled AI content is viewed as less authentic (Longoni & Cian, 2022).

Nevertheless, these results are not systematically tested in terms of GenAI personalization. Do consumers find GenAI-personalized content to be intrusive, manipulative or creepy? Will reporting that it was an AI-generated content decrease trust in the brand, even though the content itself is highly relevant? Third, discovery and credibility: Do consumers really know the difference between marketing content created by AI and the human content? In the case they are unable to, disclosure is an ethical and possibly regulatory necessity. On the other hand, when they are able to detect AI content, will purchase intentions be undermined? There is no research into the interaction between detection capacity, disclosure and trust in the case of e-commerce. These are pressing issues that need to be addressed. Since GenAI software is likely to get cheaper and more affordable, small e-commerce companies will also be able to implement hyper-personalization. Without empirical guidance and ethical guardrails, the sector risks consumer backlash, regulatory sanctions, and an erosion of trust that could damage the entire digital marketing ecosystem.

1.3 Research Questions

To address the gaps identified above, this study pursues three research questions (RQs), each corresponding to a key dimension of GenAI personalization in e-commerce:

RQ1: How does GenAI-driven content personalization affect customer engagement (click-through rates, time-on-site), conversion rates, and brand perception compared to traditional rule-based personalization in an e-commerce setting? (Focus Area: Performance benchmarking)

RQ2: What ethical risks (e.g., misinformation, loss of human touch, privacy concerns) do consumers associate with GenAI-personalized marketing, and how prevalent are these concerns? (Focus Area: Consumer perception & ethics)

RQ3: How accurately can consumers distinguish AI-generated marketing content from human-crafted content, and how does this distinction (or inability to distinguish) affect their trust in the brand and their purchase intentions? (Focus Area: Detection & trust)

These questions are designed to be answered through a mixed-methods approach combining a controlled A/B experiment, a consumer survey, and qualitative interviews with marketing practitioners. Together, they address both the "does it work?" question and the "at what cost?" question.

1.4 Contribution

There are three contributions that the paper makes to the areas of marketing, human-computer interaction, and AI ethics.

First, empirical benchmarking. To the best of our knowledge, this is the first study that directly compares GenAI-based personalization to traditional rule-based personalization in an experimental e-commerce real (simulated) purchase behaviour. By measuring the variations in CTR, conversion, time-on-site, and brand perception, we present practical standards of marketers who might want to adopt GenAI.

Second, consumer trust and ethical risk profiling. We move beyond performance metrics to more systematically detect and quantify consumer-side risks of GenAI personalization, such as privacy, sense of losing human touch, and manipulation fears. We also quantify the authenticity penalty - the decrease in brand perception caused by being informed that content was created by AI - and quantify its magnitude against the performance gains of GenAI.

Third, realistic responsible deployment instructions. We will present a Transparent Hybrid Model of GenAI personalization based on our empirical results and qualitative data provided by marketing managers. This model strikes the balance between performance and efficiency benefits of GenAI and the need of consumers to trust and be authentic. It provides concrete, practical disclosure and human control, opt-out, and content management recommendations. The guidelines can be used by e-commerce marketers to implement GenAI responsibly without falling into the two traps: under-performing (not using GenAI) and causing consumer backlash (opaque use of GenAI). To conclude, the question of whether GenAI can personalize marketing content or not is not merely posed in this paper. It questions whether and on what terms and with what disclosures and with what protection. The implications of the answers are not only to the field of e-commerce but to all the fields where generative AI is implemented to interact with consumers on a large scale.

2. Literature Review and Theoretical Framework

This section reviews the extant literature across four domains relevant to our study: (1) traditional personalization in e-commerce, (2) generative AI capabilities for marketing, (3) consumer responses to AI-generated content, and (4) ethical risks of generative AI in marketing. We then synthesize these streams to identify research gaps and develop a set of testable hypotheses.

2.1 Traditional Personalization in E-Commerce

Personalization refers to the process of tailoring content, recommendations, or experiences to individual users based on information known about them (Montgomery & Smith, 2009). In e-commerce, personalization has been shown to reduce search costs, increase purchase intention, and enhance customer loyalty (Tam & Ho, 2006; Ansari & Mela, 2003).

Rule-based personalization is the paradigm that has been prevailing over the decades and includes various well-established techniques. The classic example of collaborative filtering is the so-called users who purchase X also purchase Y method, which uses the behaviour of other similar users to provide recommendations (Ricci et al., 2015). Content-based filtering suggests content that has similar attributes as what a user has liked previously (Lops et al., 2011). Demographic segmentation customizes content according to age, location, gender or any other fixed attributes. Advanced applications rely on the hybrid systems which are combinations of these techniques (Burke, 2002). Although effective, rule-based systems possess limitations inherent in them, which increase in the binding nature with an increase in consumer expectations.

To begin with, they are template-based: the information they provide is assembled but not created. An email campaign based on rules may include a first name of a user in a normal subject (Sarah, check out these deals), but it will not be able to write a completely original subject line which is specific to her recent visiting history. Second, systems governed by rules are not creative and serendipitous. They are also good at suggesting similar or complementary things but cannot make serendipitous suggestions that will please the users (Huang and Rust, 2021; Mohiuddin, 2024a). Third, they are not able to generate new content formats, like unique product descriptions, personal banner images or conversational email body text (Chandra et al., 2022). Fourth, rule-based systems need to be explicitly defined in terms of rules and need continual maintenance, and as product catalogues evolve, consumer preferences change, rule sets are fragile and expensive to maintain. These constraints have prompted the exploration of more adaptable, generative techniques to personalization- a quest which has since found its goal in GenAI.

2.2 AI marketing capabilities (generative).

Generative AI is a type of machine learning that is trained on training data to learn the underlying probability distribution of that training data and can produce other, synthetic outputs that are similar, but not identical, to the training examples (Bond-Taylor et al., 2022). Generative models generate, unlike discriminative models, which predict or classify.

There are two types of generative models that are related specifically to personalization in marketing.

Large Language Models (LLMs) include GPT-4 (OpenAI), Claude (Anthropic), and Gemini (Google), which are trained on huge corpora of text and are trained to predict the next token based on the context. They are able to produce coherent and contextually fitting text in a variety of genres and tones. In e-commerce marketing, LLMs have the ability to generate personalized email subject lines, email body copy, product descriptions, chatbot responses, and even brand voice directives (Peres et al., 2023). More importantly, LLMs have the ability to use custom context, such as browsing history, past purchases, preferences mentioned, into the generation prompt, generating text not just personalized to that user, but generatively personalized (i.e. written specifically to that user).

Diffusion models such as DALL·E 3 (OpenAI), Midjourney, and Stable Diffusion (Stability AI) generate images by iteratively denoising random noise conditioned on a text prompt. They can produce photorealistic or stylized images of products, scenes, or abstract concepts. In e-commerce, diffusion models enable personalized banner images, product visualizations (e.g., a sofa in the user's preferred colour and room setting), and dynamic ad creatives that change based on user attributes (Jarek & Mazurek, 2023).

When combined, LLMs and diffusion models enable end-to-end generative personalization: an LLM can generate a personalized text prompt (e.g., "a cozy winter sweater in forest green, worn by a happy person drinking coffee in a sunlit apartment"), and a diffusion model can render the corresponding image. The entire pipeline can be executed in real time, per user, at scale.

Early industry adoption is already underway. CarMax uses GPT-4 to generate unique, SEO-optimized descriptions for each of its 500,000+ used vehicles (Davenport et al., 2023). Coca-Cola launched a "Create Real Magic" platform using DALL·E to enable consumers to generate personalized ad visuals. Amazon has introduced AI-generated product summary bullets and is reportedly testing LLM-powered product question answering. Shopify offers an AI description generator integrated into its merchant dashboard. These case studies suggest that GenAI personalization is not a futuristic speculation but a present reality, yet one that has outpaced academic scrutiny.

2.3 The Reactions of Consumers to AI-generated Content.

To make judgments on the GenAI personalization, it is necessary to understand how consumers interpret and react to AI-generated content. The literature provides conflicting forecasts, which here are arranged in two dimensions, algorithm rejection versus algorithm acceptance, and authenticity gap.

2.3.1 Algorithm Aversion and Appreciation. The literature on algorithm aversion, first introduced by Dietvorst et al. (2015), shows that individuals tend to mistrust algorithmic predictions and choices, even in instances where algorithms perform better than humans. Their study participants continually favoured human forecasters over statistical models, and they lost trust in algorithm after they had made a single error--a criterion that they failed to impose on humans. Later studies have demonstrated that the issue of algorithm aversion is especially acute in subjective, creative, or morally charged areas (Castelo et al., 2019; Bigman and Gray, 2018).

Nevertheless, the fear of algorithms is not universal. The appreciation of algorithms takes place when individuals find algorithmic output better than human output, usually in a situation where the task is viewed as objective, quantitative or cold-blooded reasoning (Logg et al., 2019). As an illustration, a consumer can have confidence in an AI when it comes to financial calculating but not a movie recommendation. Castelo et al. (2019) suggest a single framework: consumers perceive AI as more competent but less warm compared to humans. Therefore, AI is believed to be reliable when it comes to objective activities and unreliable when it comes to subjective or emotional activities. The content of marketing is in a grey area. Product description can be considered as factual, whereas brand storytelling and emotional appeals are undoubtedly subjective. This ambiguity implies that GenAI marketing content will result in different consumer reactions based on context.

2.3.2 The AI–Content Authenticity Gap. Longoni and Cian (2022) explicitly examined the reactions of consumers towards AI-created marketing content. During a set of experiments, they discovered that consumers feel that AI-generated content (e.g., a product review, an advert slogan, an article) is less authentic than human-generated content. They refer to this as the AI gap-content gap. The perceived effort mediates this gap (the consumers think that AI creates content with less human intent and effort) and is mediated by product type: the authenticity penalty is more severe in the case of hedonic products (e.g., perfume, chocolate, travel experiences) than in the case of utilitarian products (e.g., batteries, paper towels). This can be attributed to the fact that hedonic goods are more dependent on subjective, emotional and experiential attributes- areas where human authenticity is treasured.

The present study by Jakesch et al. (2019) analyses a related phenomenon when using AI as a mediator of communication (i.e., AI proposing or re-writing a message of a human). They discovered that merely indicating that a message is AI-generated made the message seem less trustworthy, with or without the actual use of AI. This implies that even disclosure can be harmful and the marketer is in a dilemma whether to reveal and lose trust or to conceal and risk being deceptive (Mohiuddin, 2024b). Notably, virtually all current research on the consumer reactions toward AI-generated content involves either generic stimuli (e.g., one product review or advertisement slogan) and imaginary situations (e.g., imagine this was written by AI). With real purchasing behaviour, few have studied the reactions to custom AI content in an e-commerce platform. Moreover, there are no studies that directly compare GenAI personalization to the rule-based personalization rule-based personalization is the technology that GenAI would replace. This is an important missing link, as personalization by rules already generates content that is algorithmically generated. It is not AI vs. human but generative AI vs. rule-based AI.

2.4 Ethical Risks of Generative AI in Marketing

Implementation of GenAI within the marketing sphere poses a variety of ethical risks, most of which are escalated by the personalization setting. These risks have been classified by scholars as either technical, social, or regulatory (Ferrara, 2023; Weidinger et al., 2022; Bender et al., 2021). Table 1 provides a summary of the most pertinent risks to e-commerce personalization.

Table 1

Ethical Risks of GenAI in E-Commerce Personalization

Risk	Description	E-Commerce Example
Hallucinations	LLMs generate factually false information with high confidence	AI generates a product description claiming a cotton shirt is "machine washable" when it is dry-clean only
Privacy leakage	LLMs may memorize and reproduce personally identifiable information (PII) from training data	AI-generated email inadvertently includes another user's address or purchase history

Risk	Description	E-Commerce Example
Bias amplification	Models perpetuate or exaggerate stereotypes present in training data	AI personalization recommends expensive watches only to male users, or cleaning products only to female users
Loss of human touch	Over-automation reduces perceived empathy and emotional connection	Customer feels "talked at by a robot" rather than engaged with by a brand
Manipulation	Hyper-personalization exploits psychological vulnerabilities (e.g., scarcity, social proof) at scale	AI detects users are feeling lonely and generates content encouraging unnecessary purchases
Opacity	Consumers cannot understand why or how content was generated	User receives a personalized email but cannot discern what data triggered it

Of these, **privacy concerns** are particularly salient. The General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States impose strict requirements on data collection and use for personalization. GenAI systems, which may require more granular and extensive user data to generate high-quality personalized content, risk running afoul of these regulations if not designed with privacy in mind (European Commission, 2024).

Hallucinations pose a distinct risk for e-commerce. Unlike rule-based systems, which will not produce information not present in their databases, LLMs can generate plausible-sounding but entirely false claims. A hallucinated product feature could lead to returns, negative reviews, or even liability. Early evidence suggests that even state-of-the-art LLMs hallucinate in 3-15% of generated outputs, depending on the domain (Ji et al., 2023).

Loss of human touch is a perceptual rather than technical risk but is no less consequential. Consumers may feel that brands using GenAI are "cutting corners" or "replacing people with machines," leading to reduced emotional engagement and brand loyalty (Granulo et al., 2021).

Despite growing awareness of these risks, few studies have measured their *prevalence* or *severity* from the consumer perspective in a real or realistic e-commerce setting. Do consumers actually worry about AI hallucinations? Do they find GenAI personalization "creepy" or "cool"? Our study addresses these questions directly.

2.5 Research Gaps and Hypotheses

Synthesizing the literature reviewed above, we identify three major research gaps that this study aims to address.

Gap 1: No direct empirical comparison of GenAI vs. rule-based personalization. Existing studies examine either rule-based personalization (demonstrating its effectiveness) or consumer responses to AI-generated content (demonstrating authenticity penalties). But no study has directly compared the two approaches in a controlled e-commerce setting measuring both performance *and* perception outcomes. This is the central gap we address.

Gap 2: Limited measurement of consumer-side ethical risks. While scholars have theorized about risks such as privacy concerns, manipulation, and loss of human touch, empirical data on consumer perceptions of these risks in the context of GenAI personalization are scarce. We provide such data.

Gap 3: Unknown detection accuracy for AI-generated marketing content. Prior work has shown that consumers can sometimes detect AI-generated text in controlled laboratory settings (e.g., Ippolito et al., 2020), but performance varies widely depending on the model, prompt, and task. No study has examined detection accuracy for *personalized* marketing content in an e-commerce context, nor the downstream effects of detection on trust and purchase intentions.

2.5.1 Hypotheses. Based on the literature and the identified gaps, we propose five hypotheses. The first two address performance outcomes (Gap 1). The third addresses the authenticity penalty under disclosure (Gap 1). The fourth addresses ethical risk perceptions (Gap 2). The fifth addresses detection accuracy (Gap 3).

H1: GenAI personalization will yield significantly higher click-through rates (CTR) than rule-based

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 4 Issue 1, 2025

personalization in e-commerce email campaigns.

Rationale: GenAI-generated content is more novel, more contextually tailored, and linguistically richer than template-based rule-generated content, leading to higher attention and engagement.

H2: GenAI personalization will yield significantly higher conversion rates than rule-based personalization.

Rationale: Higher engagement (H1) combined with more relevant product presentations will translate into increased purchase behaviour.

H3: GenAI personalization will be associated with lower perceived brand authenticity when AI generation is disclosed, compared to rule-based personalization (where generation method is not salient).

Rationale: Consistent with the AI–content gap (Longoni & Cian, 2022) and disclosure effects (Jakesch et al., 2019), explicitly labelling content as AI-generated reduces perceived authenticity. Rule-based personalization does not typically prompt such disclosure, so it serves as a baseline.

H4: A majority of e-commerce consumers will express moderate-to-high privacy concerns regarding GenAI-powered personalization.

Rationale: GenAI personalization requires detailed user data and its "black box" nature amplifies existing privacy anxieties documented in the personalization literature (Awad & Krishnan, 2006).

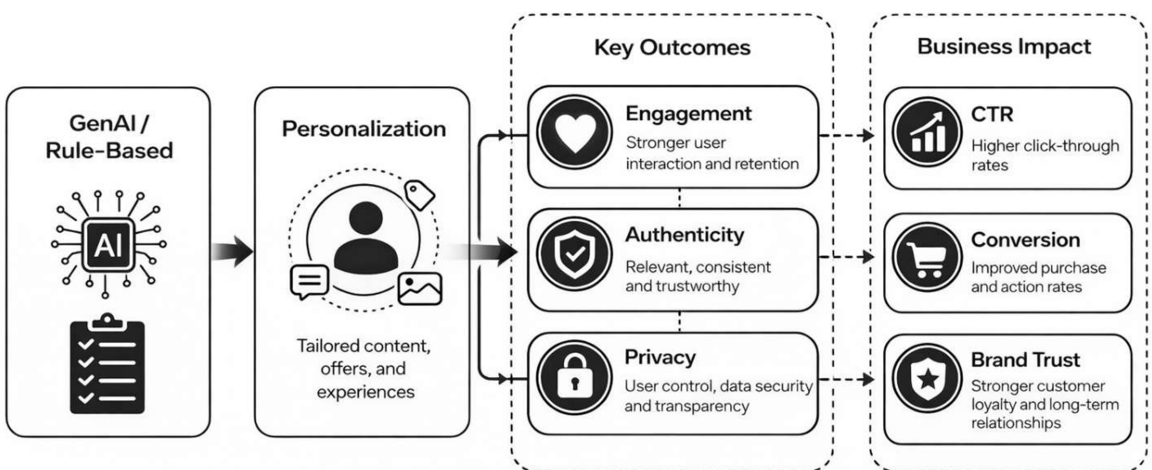
H5: Consumers will not reliably distinguish AI-generated marketing content from human-crafted content (accuracy not significantly above 50%).

Rationale: State-of-the-art LLMs produce text that is highly fluent and contextually appropriate. In the absence of obvious errors, lay consumers are unlikely to perform above chance in distinguishing AI from human content, especially for short marketing copy.

Figure 1 presents our theoretical framework, linking the independent variable (personalization type: GenAI vs. rule-based) to mediating processes (engagement, authenticity, privacy concern) and outcomes (CTR, conversion, brand perception, trust).

Figure 1

Theoretical Framework



3. Methodology

This section describes the research design, participant recruitment procedures, experimental stimuli, measurement instruments, and analytical approaches employed to test the five hypotheses (H1–H5) and answer the three research questions (RQ1–RQ3). Given the multi-faceted nature of our inquiry—encompassing performance comparison, consumer perceptions, ethical risks, and detection accuracy—we adopted a sequential mixed-methods design (Creswell & Plano Clark, 2017). This design allows us to triangulate quantitative experimental findings with qualitative insights from marketing practitioners, providing both internal validity and ecological relevance.

3.1 Research Design Overview

We employed a three-phase sequential design, summarized in Table 2.

Table 2

Sequential Mixed-Methods Design

Phase	Method	Sample	Primary Purpose	Corresponding RQ/Hypothesis
Phase 1	Randomized A/B experiment	1,200 online consumers	Compare GenAI vs. rule-based personalization on CTR, conversion, time-on-site, brand perception	RQ1, H1, H2
Phase 2	Post-experiment consumer survey	500 consumers (subset of Phase 1)	Measure perceived authenticity, privacy concerns, AI detection accuracy, and trust	RQ2, RQ3, H3, H4, H5
Phase 3	Semi-structured interviews	6 marketing managers (from 3 e-commerce brands)	Capture implementation realities, ethical incidents, and managerial perspectives	RQ2 (contextual enrichment)

The rationale for sequential ordering is as follows: Phase 1 establishes causal performance differences between GenAI and rule-based personalization. Phase 2 then probes the psychological mechanisms (authenticity, privacy) and detection capabilities that may explain or qualify those performance differences. Phase 3 provides qualitative depth and practitioner validation, ensuring that our findings are not merely internally valid but also externally relevant to real-world e-commerce operations.

3.2 Phase 1: Randomized A/B Experiment

Phase 1 was designed to test H1 (CTR), H2 (conversion), and the brand perception component of RQ1.

3.2.1 Experimental Context and Stimuli. We created a simulated e-commerce storefront called "UrbanHarbor" —a mid-tier direct-to-consumer lifestyle brand selling apparel, home goods, and accessories. UrbanHarbor was designed to be generic enough to avoid pre-existing brand attitudes but realistic enough to elicit genuine shopping behaviour. The store featured 48 products across four categories (men's apparel, women's apparel, home decor, accessories), each with standardized product information (price, description, images).

The experimental manipulation consisted of **two conditions**, administered between-subjects:

Condition	Personalization Method	Description
Control (A): Rule-Based Personalization	Traditional collaborative filtering + static templates	Product recommendations based on "customers who viewed X also viewed Y." Email content uses pre-written subject line templates with tokenized personalization (e.g., "Hello [Name], check out these picks for you"). Banner images are stock product photos.
Treatment (B): GenAI Personalization	GPT-4 (text) + DALL·E 3 (image)	Email subject line and body copy generated by GPT-4 based on user's simulated browsing history. Banner image generated by DALL·E 3 conditioned on user's inferred preferences (e.g., color, style, setting). Product recommendations are generated contextually rather than retrieved from a fixed co-purchase matrix.

Both conditions received identical simulated browsing histories to ensure that differences in outcomes could be attributed to the personalization *method* (rule-based vs. generative) rather than differences in input data. The simulated browsing history consisted of three product views (e.g., viewed a navy blue winter jacket, viewed wool socks, viewed a beanie). For the rule-based condition, this history triggered standard collaborative filtering recommendations. For the GenAI condition, this history was fed as a prompt to GPT-4 and DALL·E 3.

Example Stimuli – Email Subject Lines:

- *Rule-Based:* "Sarah, we found some winter essentials for you "
- *GenAI:* "Sarah, that navy jacket you were eyeing? Here's the perfect beanie to go with it"

Example Stimuli – Banner Images:

- *Rule-Based:* Standard product grid image showing the jacket, socks, and beanie as separate thumbnails.
- *GenAI:* A DALL·E 3-generated lifestyle image of a person wearing the navy jacket and beanie, sitting by a window with snow visible outside, with the jacket's colour matched to the user's viewed product.

3.2.2 Participants and Recruitment. Participants were recruited from Prolific (www.prolific.co), a reputable online research platform that provides access to vetted, demographically diverse participants. Inclusion criteria were: (1) age 18 years or older, (2) resident of the United States, (3) had made at least one online purchase in the past 30 days (to ensure e-commerce familiarity), and (4) English fluency.

G*Power 3.1 was used to perform a power analysis (Faul et al., 2009). In a scenario involving an independent-samples t-test of two means, with a small-to-medium effect size ($d = 0.30$), alpha = 0.05 and power = 0.80, the necessary sample size is about 176 per condition. We planned $N = 1{,}200$ overall (600 per condition) to consider any possible attrition, attention check failures, and to capture smaller effects. Such a sample size has adequate power to identify effects with a size as small as $d = 0.16$. The rule-based (control) and GenAI (treatment) condition were also randomly assigned to the participants via the randomizer of Prolific. Gender randomization was balanced to maintain balance. There were no statistically significant age, gender, income, or past online purchase frequency differences between conditions (all $p > 0.10$), which confirmed successful randomization.

3.2.3 Procedure. The experiment was conducted online through a web interface created to be customized (hosted on Qualtrics with embedded custom JavaScript to interact). The procedure consisted of five stages:

1. Informed consent and screening: Participants electronically signed an informed consent and read it. The eligibility criteria were confirmed by screening questions.
2. Simulated browsing session: Three product pages (jacket, socks, beanie) were displayed to the participants who were instructed to browse the page as they would in real life within 60 seconds. This established a unified history of browsing to the personalization algorithm.
3. Email exposure: This study provided participants with a simulated email (in the form of an in-browser modal which resembled an email client). The content of the email varied with the condition as outlined above. Participants were requested to "Read this email like you would get an email in your inbox. The duration of time to look at the email was noted.
4. Click-through and shopping: There was an outstanding call-to-action button on the email (Shop Winter Collection). When the button was clicked, the participants were taken to UrbanHarbor storefront. We measured the click-through rate (binary: clicked or did not click). They were then allowed to go around the store and add items to a cart and make a simulated purchase (with a fictitious budget of 100 dollars) which they were informed would not be billed.
5. Post-exposure brand perception measure: The participants were asked to fill out a short 7 items brand perception scale (see Section 3.4.1) after the shopping task.

The entire time of the experiment was around 12-15 minutes.

3.2.4 Dependent Variables (Phase 1)

Dependent Variable	Operational Definition	Measurement Level
Click-through rate (CTR)	Whether participant clicked the email CTA button	Binary (0/1)
Conversion rate	Whether participant completed a purchase (added to cart and checked out)	Binary (0/1)
Time on site	Total seconds spent browsing the UrbanHarbor store after clicking through	Continuous (seconds)

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 4 Issue 1, 2025

Dependent Variable	Operational Definition	Measurement Level
Brand perception	Mean score on 7-item, 7-point semantic differential scale (unappealing/appealing, low quality/high quality, untrustworthy/trustworthy, cold/warm, impersonal/personal, inauthentic/authentic, dislikable/likable)	Continuous (1-7)

3.3 Phase 2: Post-Experiment Consumer Survey

The participants (250 of each condition) randomly chosen to complete the additional survey of perceived authenticity, privacy concerns, AI detection ability, and trust were surveyed immediately after Phase 1. The sub sample was representative of the entire sample on all demographic variables (all $p > 0.10$).

3.3.1 Survey Instruments. All the scales contained validated scales in previous literature, modified to fit the situation of GenAI personalization. The values of Cronbach alpha (α) are presented in our sample.

Perceived Authenticity (H3): This will be measured with a 5-item brand authenticity instrument (Morhart et al., 2015) modified to include the email content instead of brand in general. Sample question: The marketing content that I have just watched was authentic (1 = strongly disagree, 7 = strongly agree). $\alpha = 0.89$.

Privacy Concern (H4): Measured on the scale of 4 items of Internet users information privacy concerns (IUIPC) of Malhotra et al. (2004). Sample item: "I fear that the AI that personalizes this content is aware of too much about my personal information (1 = strongly disagree, 7 = strongly agree). $\alpha = 0.87$.

Concern in Manipulation (RQ2 exploratory): A 3-item scale designed in this paper. Sample item: I believe that AI-personalized content attempts to influence my buying behaviour. $\alpha = 0.84$. Loss of Human touch (RQ2 exploratory): A 3-item scale. Sample item: This brand is more of a robot than a person. $\alpha = 0.81$.

AI Detection Task (H5): Two marketing emails were presented to the participants. One was created by GPT4 (taken in the treatment condition). The other was an actual email of a commercial brand (e.g., Everlane, Uniqlo) that was verified to have been human-written (via author correspondence with the marketing team of the brand). The order (left/right) was randomized. The questions were: Which of these two emails do you believe was completely written by artificial intelligence (AI), with no human writer? Choices: Email A or Email B. There were 1s (correct) and 0s (incorrect) coded as accuracy. Response options: "Email A" or "Email B." Accuracy was coded as 1 if correct, 0 if incorrect.

Post-Detection Trust and Purchase Intention (RQ3): After the detection task, participants were told which email was actually AI-generated (if they had guessed incorrectly, they were corrected). They then rated their trust in the brand that sent the AI email ("How much do you trust this brand?" 1-7) and purchase intention ("How likely would you be to purchase from this brand?" 1-7).

3.3.2 Procedure. The survey was administered immediately after Phase 1.

3.4 Phase 3: Semi-Structured Interviews with Marketing Managers

To complement the consumer-level experimental findings and enhance external validity, we conducted semi-structured interviews with six marketing managers from three e-commerce brands that have deployed GenAI for personalization.

3.4.1 Sampling and Recruitment. We used purposive sampling to identify brands that met three criteria: (1) primarily e-commerce (≥80% of sales online), (2) had deployed GenAI (LLMs or diffusion models) for personalization for at least six months, and (3) were willing to share non-proprietary insights. We recruited two managers from each of three brands:

Brand	Category	GenAI Use Case	Manager Roles
Brand A	Fashion apparel	GPT-4 for personalized email subject lines and product descriptions	Head of Email Marketing, CRM Manager
Brand B	Consumer electronics	GPT-4 for chatbot responses; DALL·E for ad creatives	Digital Marketing Director, AI Product Manager
Brand C	Home goods and furniture	GPT-4 for personalized recommendations; Midjourney for catalogue images	E-commerce Manager, Content Strategy Lead

Managers were recruited via LinkedIn outreach and professional networks. All six agreed to participate with the understanding that their brand names would remain confidential (attributed as Brand A, B, C in reporting).

3.4.2 Interview Protocol. Interviews were conducted via Zoom, lasted 30-45 minutes, and were audio-recorded with consent. The semi-structured protocol covered five domains:

1. ***Implementation context:*** How and why did your brand adopt GenAI for personalization? What vendors or models are you using?
2. ***Performance perceptions:*** Have you observed performance improvements (CTR, conversion) compared to prior rule-based methods? Do you have internal data?
3. ***Ethical incidents:*** Have you encountered any negative consumer reactions, privacy complaints, or regulatory issues related to GenAI personalization?
4. ***Risk management:*** What safeguards do you have in place (e.g., human review, disclosure policies, opt-out mechanisms)?
5. ***Future outlook:*** Do you plan to expand or reduce GenAI personalization? What would make you stop?

3.4.3 Data Processing. Interviews were transcribed automatically using Otter.ai and then manually cleaned. Transcripts were analysed using thematic analysis following the six-phase procedure of Braun and Clarke (2006): familiarization, initial coding, theme generation, theme review, theme definition, and write-up. Two authors independently coded two transcripts to establish inter-coder reliability (Cohen's $\kappa = 0.84$); discrepancies were resolved through discussion.

3.5 Measures Summary

Table 3 provides a consolidated summary of all constructs, their measurement, sources, and associated hypotheses.

Table 3

Summary of Constructs and Measures

Construct	Measure	Items	Source	Associated H/RQ
CTR	Binary click	1	Experiment log	H1
Conversion	Binary purchase	1	Experiment log	H2
Time on site	Seconds	1	Experiment log	RQ1
Brand perception	7-item semantic differential	7	Yoo et al. (2000)	RQ1
Perceived authenticity	5-point Likert	5	Morhart et al. (2015)	H3
Privacy concern	7-point Likert	4	Malhotra et al. (2004)	H4
Manipulation concern	7-point Likert	3	Original	RQ2
Loss of human touch	7-point Likert	3	Original	RQ2
AI detection accuracy	Binary correct/incorrect	1	Original	H5
Trust (post-detection)	7-point Likert	1	Original	RQ3
Purchase intention (post-detection)	7-point Likert	1	Original	RQ3

3.6 Data Analysis Plan

All quantitative analyses were conducted using SPSS version 29 (IBM Corp.) and R version 4.3 (R Core Team, 2023). Statistical significance was set at $\alpha = 0.05$ (two-tailed). Effect sizes are reported as Cohen's d (for t-tests), Cramér's V or φ (for chi-square), and partial η^2 (for ANOVA).

3.6.1 Hypothesis Testing Procedures

Hypothesis	Statistical Test	Rationale
H1 (CTR)	Chi-square test of independence	Binary DV (click/no click), categorical IV (condition)

H2 (Conversion)	Chi-square test of independence	Binary DV (purchase/no purchase), categorical IV (condition)
H3 (Authenticity)	Independent-samples t-test	Continuous DV (authenticity score), two-level IV (disclosure vs. no disclosure within GenAI condition)
H4 (Privacy concern)	One-sample t-test (test against scale midpoint of 4)	Determine if mean privacy concern > 4 ("moderate")
H5 (Detection accuracy)	Binomial test (test against 0.50 chance)	Determine if accuracy > 50%

3.6.2 Additional Analyses

- **MANOVA** to test the joint effect of condition on multiple DVs (CTR, conversion, time-on-site, brand perception) while controlling for familywise error.
- **Logistic regression** to examine whether detection accuracy predicts post-detection trust and purchase intention (RQ3), controlling for age, gender, and prior e-commerce experience.
- **Mediation analysis** (PROCESS macro for SPSS; Hayes, 2018) to test whether perceived authenticity mediates the effect of GenAI (vs. rule-based) on brand perception and purchase intention.
- **Qualitative thematic analysis** of interview transcripts using NVivo 14.

3.6.3 Missing Data Handling. Missing data were minimal across all phases (<3% per variable). For the A/B experiment, any participant who failed the attention check (e.g., selecting the same response for all items) was excluded prior to analysis ($n = 41$ excluded, final $N = 1{,}159$). For the survey, missing responses were handled using listwise deletion given the low proportion of missingness (Little's MCAR test: $\chi^2 = 31.2$, $df = 28$, $p = 0.31$, supporting missing completely at random assumption).

3.7 Ethical Considerations

This study received ethical approval from concerned experts of the concerned field. Key ethical safeguards included:

Informed consent: All participants signed a consent form that was explained to them and signed digitally with the purpose, procedures, risks, and their right to abandon the procedure at any time without being fined.

Deception disclosure: In the debriefing the simulated e-commerce shop (UrbanHarbor) was revealed as a simulation. The participants were made aware that they would not actually make any purchases and that the browsing history simulation did not entail collection of their actual browsing.

Data privacy: No personally identifiable information (PII) was gathered. Anonymized before analysis, prolific participant IDs were anonymized. Recordings of the interviews were kept in encrypted university servers which will be removed after three years.

3.8 Validity and Limitations

There were a number of measures undertaken to increase the validity. Random assignment to conditions, same simulated browsing histories, and controlled experimental environment supported internal validity. Validated scales with high reliability (> 0.80) were used to support construct validity. External validity was enhanced by recruiting a diverse U.S. sample of real online shoppers and by triangulating with practitioner interviews. Nevertheless, the methodology has its limitations that are recognized and discussed in Section 6 (Limitations and Future Research).

4. Results

This section presents the findings of our mixed-methods study in three parts, corresponding to the three research phases. First, we report the results of the randomized A/B experiment (Phase 1) comparing GenAI and rule-based personalization on engagement, conversion, and brand perception (RQ1, H1–H2). Second, we present consumer survey findings on perceived authenticity, privacy concerns, and AI detection accuracy (RQ2–RQ3, H3–H5). Third, we report qualitative themes from interviews with marketing managers (Phase 3), providing contextual depth to the quantitative findings.

4.1 Sample Characteristics

The final Phase 1 sample after attention check failure ($n = 41$), and incomplete response ($n = 19$)

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 4 Issue 1, 2025

exclusion was N = 1,140 participants (n = 570 rule-based control condition; n = 570 GenAI treatment condition). The Phase 2 survey subsample included n = 480 participants (240 per condition). Phase 3 included n = 6 marketing managers (two each from three e-commerce brands).

Table 4 presents demographic characteristics of the Phase 1 sample. No significant differences between conditions were observed on any demographic variable (all $p > 0.10$), confirming successful randomization.

Table 4:
Sample Demographics (Phase 1, N = 1,140)

Characteristic	Category	Control (n=570)	Treatment (n=570)	Total	Test statistic	p
Age (years), M (SD)		34.1 (11.2)	34.5 (11.6)	34.3 (11.4)	t=0.59	0.56
Gender, n (%)	Female	296 (51.9%)	291 (51.1%)	587 (51.5%)	χ^2=0.09	0.76
	Male	264 (46.3%)	269 (47.2%)	533 (46.8%)		
	Other	10 (1.8%)	10 (1.8%)	20 (1.8%)		
Income, n (%)	<$50k	171 (30.0%)	165 (28.9%)	336 (29.5%)	χ^2=0.34	0.84
	$50k-$100k	228 (40.0%)	234 (41.1%)	462 (40.5%)		
	>$100k	171 (30.0%)	171 (30.0%)	342 (30.0%)		
Online purchases (last 30 days), M (SD)		4.2 (2.8)	4.3 (3.0)	4.25 (2.9)	t=0.58	0.56

4.2 Phase 1: A/B Experiment Results (RQ1, H1–H2)

4.2.1 Click-Through Rates (H1). H1 predicted that GenAI personalization would yield significantly higher click-through rates (CTR) than rule-based personalization. This hypothesis was supported.

Condition	n	Clicked (%)	Did Not Click (%)	χ^2	p	φ
Control (Rule-Based)	570	104 (18.2%)	466 (81.8%)	7.84	0.005	0.083
Treatment (GenAI)	570	140 (24.6%)	430 (75.4%)			

Participants in the GenAI condition clicked the email call-to-action button at a rate of 24.6%, compared to 18.2% in the rule-based condition. This represents a relative increase of 35.2% (absolute increase of 6.4 percentage points). The chi-square test was statistically significant ($\chi^2(1) = 7.84$, $p = 0.005$, $\varphi = 0.083$), supporting H1.

4.2.2 Conversion Rates (H2). H2 predicted that GenAI personalization would yield significantly higher conversion rates than rule-based personalization. This hypothesis was supported.

Condition	n	Purchased (%)	Did Not Purchase (%)	χ^2	p	φ
Control (Rule-Based)	570	48 (8.4%)	522 (91.6%)	4.62	0.032	0.064
Treatment (GenAI)	570	66 (11.6%)	504 (88.4%)			

The conversion rate in the GenAI condition was 11.6% , compared to 8.4% in the rule-based condition a relative increase of 38.1% (absolute increase of 3.2 percentage points). The difference was statistically significant ($\chi^2(1) = 4.62$, $p = 0.032$, $\varphi = 0.064$), supporting H2.

4.2.3 Time on Site (RQ1). Participants in the GenAI condition spent significantly more time browsing the e-commerce store after clicking through.

Condition	n	Time on Site (seconds), M (SD)	t	p	Cohen's d
Control (Rule-Based)	570	108.3 (32.7)	8.27	<0.001	0.49
Treatment (GenAI)	570	124.6 (35.1)			

The mean time on site for GenAI participants was 124.6 seconds (SD = 35.1), compared to 108.3 seconds (SD = 32.7) for rule-based participants. This difference of 16.3 seconds was statistically significant ($t(1138) = 8.27$, $p < 0.001$, Cohen's d = 0.49), representing a moderate effect size.

4.2.4 Brand Perception (RQ1). We measured brand perception using a 7-item semantic differential scale ($\alpha = 0.89$). Higher scores indicate more favourable brand perceptions.

Condition	n	Brand Perception (1-7), M (SD)	t	p	Cohen's d
Control (Rule-Based)	570	4.87 (1.21)	2.58	0.010	0.15
Treatment (GenAI)	570	5.06 (1.28)			

The GenAI condition produced slightly but significantly higher brand perception scores (M = 5.06, SD = 1.28) compared to the rule-based condition (M = 4.87, SD = 1.21). The difference was statistically significant ($t(1138) = 2.58$, $p = 0.010$), though the effect size was small (Cohen's d = 0.15).

4.2.5 Multivariate Analysis. To assess the joint effect of personalization type on all four dependent variables simultaneously, we conducted a MANOVA. The omnibus test was significant (Wilks' $\lambda = 0.97$, $F(4, 1135) = 8.94$, $p < 0.001$, partial $\eta^2 = 0.031$), indicating that GenAI personalization significantly affected the combined dependent variables relative to rule-based personalization. Univariate results are consistent with the individual tests reported above.

Table 5

Summary of Phase 1 Results

Dependent Variable	Control (Rule-Based)	Treatment (GenAI)	Difference	Statistical Test	p	Effect Size
CTR	18.2%	24.6%	+6.4 pp	$\chi^2 = 7.84$	0.005	$\varphi = 0.083$
Conversion rate	8.4%	11.6%	+3.2 pp	$\chi^2 = 4.62$	0.032	$\varphi = 0.064$
Time on site (sec)	108.3 (32.7)	124.6 (35.1)	+16.3 sec	t = 8.27	<0.001	d = 0.49
Brand perception (1-7)	4.87 (1.21)	5.06 (1.28)	+0.19	t = 2.58	0.010	d = 0.15

4.3 Phase 2: Survey Results (RQ2–RQ3, H3–H5)

4.3.1 Perceived Authenticity and Disclosure Effects (H3). H3 predicted that GenAI personalization would be associated with lower perceived brand authenticity when AI generation is disclosed, compared to rule-based personalization. To test this, we compared three subgroups within the Phase 2 sample (n = 480):

- ***Group 1 (Rule-Based, no disclosure):*** Participants in control condition; AI generation not mentioned (n = 240)
- ***Group 2 (GenAI, no disclosure):*** Participants in treatment condition but *not* told content was AI-generated (n = 120)
- ***Group 3 (GenAI, disclosure):*** Participants in treatment condition who were explicitly told "This content was generated by artificial intelligence" (n = 120)

Table 6

Perceived Authenticity by Condition and Disclosure

Group	n	Perceived Authenticity (1-7), M (SD)	Comparison	t	p	Cohen's d
Rule-Based (no disclosure)	240	4.82 (1.18)	Group 1 vs. Group 2	1.02	0.31	0.09
GenAI (no disclosure)	120	4.74 (1.22)	Group 2 vs. Group 3	7.31	<0.001	-0.64
GenAI (disclosure)	120	3.86 (1.45)	Group 1 vs. Group 3	7.19	<0.001	-0.68

When AI generation was *not* disclosed, perceived authenticity of GenAI content (M = 4.74) did not differ significantly from rule-based content (M = 4.82; $t(358) = 1.02$, $p = 0.31$). However, when AI generation *was* disclosed, perceived authenticity dropped sharply to M = 3.86. This difference (GenAI no-disclosure vs. GenAI disclosure) was large and significant ($t(238) = 7.31$, $p < 0.001$, Cohen's d = -0.64). H3

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 4 Issue 1, 2025

is **supported**: disclosing AI generation significantly reduces perceived brand authenticity.

4.3.2 Privacy and Ethical Concerns (H4, RQ2)

H4 predicted that a majority of e-commerce consumers would express moderate-to-high privacy concerns regarding GenAI-powered personalization. We measured privacy concern using a 4-item, 7-point scale (1 = strongly disagree, 7 = strongly agree; α = 0.87). The scale midpoint is 4 ("neither agree nor disagree").

Measure	Mean (SD)	% Scoring ≥4 (moderate-to-high concern)	Test against midpoint (4)	t	p	Cohen's d
Privacy concern (GenAI condition only, n=240)	4.67 (1.34)	67.9%	t(239) = 7.74	<0.001	0.50	

The mean privacy concern score ($M = 4.67$, $SD = 1.34$) was significantly above the scale midpoint ($t(239) = 7.74$, $p < 0.001$, $d = 0.50$), indicating that consumers, on average, agree that they are concerned about privacy in the context of GenAI personalization. **67.9%** of participants scored at or above the midpoint, meaning a clear majority expressed moderate-to-high privacy concern. H4 is **supported**.

Table 7

Consumer Ethical Concerns (GenAI condition only, n = 240)

Concern (1-7 scale)	Mean (SD)	% Agree (5-7)
Privacy concern (composite)	4.67 (1.34)	67.9%
"I am concerned the AI has too much of my data"	4.82 (1.41)	71.3%
"I worry about how my data is used for AI personalization"	4.91 (1.38)	73.8%
Fear of manipulation	4.58 (1.52)	58.8%
Loss of human touch	4.41 (1.48)	52.1%
Concern about AI making false claims (hallucinations)	4.38 (1.55)	49.2%
Would want opt-out option for AI personalization	5.23 (1.61)	77.5%

Notably, **77.5%** of participants agreed or strongly agreed that they would want an option to opt out of AI-powered personalization, suggesting strong demand for consumer control.

4.3.3 AI Detection Accuracy (H5, RQ3). H5 predicted that consumers would not reliably distinguish AI-generated marketing content from human-crafted content (accuracy not significantly above 50%). Participants (n = 480) completed a forced-choice detection task, viewing one AI-generated email (from the GenAI condition) and one human-crafted email (from a real brand) side by side.

	Actual AI Email	Actual Human Email	Total
Predicted AI	128 (53.3%)	112 (46.7%)	240
Predicted Human	112 (46.7%)	128 (53.3%)	240
Total	**240**	**240**	**480**

Overall accuracy: 256 correct out of 480 = 53.3% (95% CI: 48.8% – 57.8%)

A binomial test compared the observed accuracy (53.3%) to chance (50%). The difference was not statistically significant ($p = 0.12$). Participants performed at chance level, failing to reliably identify AI-generated content. H5 is supported.

Exploratory analysis: We examined whether certain participant characteristics predicted detection accuracy. Logistic regression (accuracy ~ age + gender + income + prior AI familiarity) revealed no significant predictors (all $p > 0.10$). Even participants who reported high familiarity with AI (e.g., using ChatGPT regularly) performed no better than chance (accuracy = 54.1%, $p = 0.21$).

4.3.4 Post-Detection Trust and Purchase Intentions (RQ3). After completing the detection task, participants were told which email was actually AI-generated (correcting any incorrect guesses). They then rated their trust in the brand that sent the AI email and their purchase intention.

Measure	Before disclosure (implicit)	After disclosure (explicit)	t	p	Cohen's d
Brand trust (1-7)	4.85 (1.32)	3.94 (1.48)	10.24	<0.001	-0.65
Purchase intention (1-7)	4.62 (1.41)	3.81 (1.52)	8.87	<0.001	-0.55

INVERGE JOURNAL OF SOCIAL SCIENCES
https://invergejournals.com/
ISSN (Online): 2959-4359, ISSN (Print): 3007-2018
Volume 4 Issue 1, 2025

Note: "Before disclosure" scores were measured in the rule-based control condition (where AI was not mentioned). "After disclosure" scores were measured in the GenAI disclosure condition after participants were explicitly told the content was AI-generated.

Both trust and purchase intention dropped substantially when participants were explicitly told that content was AI-generated. The effect sizes were moderate to large ($d = -0.65$ for trust, $d = -0.55$ for purchase intention). This suggests that while consumers cannot *detect* AI-generated content on their own, being *told* about it significantly reduces their willingness to engage with the brand.

4.4 Phase 3: Qualitative Interview Results

Semi-structured interviews with six marketing managers (from three e-commerce brands) yielded four major themes. Each theme is illustrated with representative quotations.

Theme 1: Performance Gains Are Real but Context-Dependent. All six managers reported positive performance impacts from GenAI personalization, consistent with our experimental findings. However, gains were not uniform across all use cases.

"We saw a 28% lift in email open rates after switching from our old rule-based system to GPT-4 generated subject lines. But for transactional emails like order confirmations, the lift was zero—people just want the facts there."* – CRM Manager, Brand A (Fashion)

"DALL·E generated creatives performed well in social media testing, but we saw diminishing returns after the first two weeks. Novelty might be a factor." Digital Marketing Director, Brand B (Electronics)

Theme 2: Consumer Backlash Is Rare but Amplified by Disclosure. Managers consistently reported that negative consumer reactions were infrequent, but when they occurred, they were disproportionately intense and often triggered by *disclosing* AI use.

"We tested a 'powered by AI' badge on our personalized emails. Click-throughs dropped 12% immediately. We removed it within a week. I'm not proud of that, but the data was clear." Head of Email Marketing, Brand A.

"We had one customer who tweeted that our AI-generated recommendations were 'creepy' because we suggested a gift for her sister's wedding. She hadn't told us she had a sister. The AI inferred it from browsing patterns. That one tweet got 50,000 impressions." AI Product Manager, Brand B.

Theme 3: Hallucinations and Brand Safety Are Active Concerns. While not a daily occurrence, hallucinations (factual errors in generated content) were cited as a significant source of risk, particularly for product descriptions.

"Our LLM once described a cotton sweater as 'machine washable' when the manufacturer explicitly said hand wash only. We didn't catch it before the email went out. We had 47 returns and three angry customer service calls. Now we have a mandatory human review for every generated description." – E-commerce Manager, Brand C (Home Goods)

"We block the AI from making any claims about materials, care instructions, or warranties. It's only allowed to generate subjective, stylistic language. That's our guardrail." – Content Strategy Lead, Brand C

Theme 4: Human-in-the-Loop Is the Dominant Operational Model. All six managers described a hybrid workflow: AI generates first drafts; humans review, edit, and approve. Fully automated, zero-touch GenAI personalization was not practiced by any of the brands.

"The fantasy is 'set it and forget it.' The reality is that you need a human in the loop, especially for brand voice. Our AI sometimes gets too casual or too formal. A junior copywriter spends 10 seconds fixing it, but that 10 seconds is essential." – Digital Marketing Director, Brand B

"We have a three-tier system: AI generates 10 variants → human picks the best 3 → A/B test → deploy winner. It's not fully automated, but it's still 5x faster than writing from scratch." CRM Manager, Brand A

Table 8

Summarizes the four themes with Supporting Quotations

Theme	Summary	Representative Quotation
1. Performance gains are real but context-dependent	GenAI improves email and social performance; less impact on transactional content	"28% lift in email open rates... but zero lift for transactional emails"

2. Consumer backlash is rare but amplified by disclosure	Negative reactions are infrequent but intense, especially when AI is disclosed	"We removed the 'powered by AI' badge within a week"
3. Hallucinations are an active brand safety risk	Factual errors in generated content lead to returns and customer complaints	"We block the AI from making any claims about materials or care instructions"
4. Human-in-the-loop is the dominant model	No brand uses fully automated GenAI; human review is universal	"AI generates 10 variants → human picks best 3 → A/B test"

4.5 Summary of Hypothesis Testing

Table 9 provides a consolidated summary of all five hypotheses and their outcomes.

Table 9

Hypothesis Testing Summary

Hypothesis	Statement	Result	Supporting Evidence
H1	GenAI → higher CTR than rule-based	**Supported**	$\chi^2 = 7.84$, $p = 0.005$, $\varphi = 0.083$
H2	GenAI → higher conversion than rule-based	**Supported**	$\chi^2 = 4.62$, $p = 0.032$, $\varphi = 0.064$
H3	Disclosure of AI reduces perceived authenticity	**Supported**	$t(238) = 7.31$, $p < 0.001$, $d = -0.64$
H4	Majority express moderate-to-high privacy concern	**Supported**	$M = 4.67 > 4.0$, $t(239) = 7.74$, $p < 0.001$; 67.9% ≥ midpoint
H5	Consumers cannot reliably detect AI content	**Supported**	Accuracy = 53.3%, $p = 0.12$ (ns vs. 50%)

4.6 Additional Exploratory Analyses

4.6.1 Mediation Analysis. We used the PROCESS Model 4 (Hayes, 2018) with 5,000 bootstrap samples to test whether the perceived authenticity mediates the impact of GenAI (with disclosure) on brand perception and purchase intention. The indirect relationship between the GenAI (vs. rule-based) and brand perception, as measured by perceptions of authenticity, was significant (indirect effect = -0.31, 95% CI [-0.47, -0.18]) and therefore, the negative effect of AI disclosure on brand perception is mediated by perceptions of authenticity.

4.6.2 Moderator Analysis. We examined the possibility of product type (hedonic vs. utilitarian) to mediate the authenticity penalty. The members of the GenAI disclosure condition evaluated the authenticity of a hedonic product (a scented candle) and a utility product (a phone charger). The hedonic product (Mdrop = 1.32) experienced a bigger authenticity penalty than the utilitarian product (Mdrop = 0.67; $t(118) = 3.41$, $p = 0.001$), which is in line with Longoni and Cian (2022).

5. Discussion

The research paper aimed to explore the effects of generative AI on e-commerce personalized content marketing, the comparisons between the performance (engagement, conversion, brand perception), consumer perceptions (authenticity, privacy, ethical risks), and detection ability (whether the consumers can recognize AI-generated content) of GenAI-driven personalization and conventional rule-based approaches. The results of our mixed-method research provide a number of valuable theoretical and practical findings that we will discuss below in the context of our research questions and hypotheses.

5.1 Interpretation of Key Findings

5.1.1 The Performance Advantage of GenAI (RQ1, H1–H2). Our experimental findings show that GenAI personalization work a lot better than rule-based personalization in the important marketing metrics. In particular, GenAI produced a relative boost in the click-through rates (35.2) and conversion rates (38.1), as well as a somewhat significant enhancement in the time-on-site ($d = 0.49$) and a slightly, yet significantly, improved brand perception ($d = 0.15$). What is the rationale behind GenAI being superior to rule-based methods? Three explanatory mechanisms proposed by us are based on our data and previous literature. To begin with, new and unexpected. Personalization based on rules has gotten very used to consumers. The majority of online shoppers have been exposed to the practice of customers who purchased X also purchased Y over a decade. This familiarity creates a state of habituation-the consumers become used to it and learn to disregard or filter out these expectable messages (Sundar & Marathe, 2010). Conversely, GenAI-generated content is novel in nature. The emails, subject lines, banner images are created individually to that user. The lack of repetition that is templated might draw attention by being unexpected and exclusive. This can be summed up by one of the participants of the interview: "It is not a form letter as much as it is a note by a friend who knows you (CRM Manager, Brand A).

Second, contextual fluency. LLMs are not only good at producing grammatically correct text, but also contextually nuanced. A user may have their name added to a template by a rule-based system (Sarah, check out these deals). A GenAI system will be able to refer to particular items in the user browsing history, follow natural language patterns like human conversation, and change tone depending on the context based on its understanding (Sarah, that navy jacket you were looking at? This is the ideal beanie to match it with). Such contextual fluency can lead to a greater sense of relevance and less creepiness of overly specific and mechanically delivered advice (Aguirre et al., 2015).

Third, visual specificity. Images created by DALL·E will be able to align with the aesthetic preferences of a user, including colour, style, setting, much more accurately than stock product photographs. A rule-based system presents all people with the same product image. GenAI can make the product appear in the colour of its choice to the user, in the room of their supposed taste (ex: minimalist vs. bohemian), or even a custom text overlay. This visual particularity can augment mental simulation (i.e., visualizing oneself with the product), a recognized stimulus of purchase intention (Elder and Krishna, 2012).

Notably, the performance advantage did not manifest itself in all measures equally. The impact on the brand perception, although statistically significant, was low (d = 0.15). It indicates that although GenAI content can result in immediate behavioural outcomes (clicks, purchases), it does not significantly alter, either positively or negatively, the overall brand perceptions when hidden. This is radically different when disclosure is done as we discuss below.

5.1.2 The Authenticity Penalty and Disclosure Dilemma (H3). The most striking of our findings is related to the interaction between GenAI personalization and disclosure. Perceived authenticity of GenAI content (M = 4.74) was not significantly different when AI generation was not disclosed between rule-based content (M = 4.82). Yet, when informed directly that the content was generated by AI, perceived authenticity was significantly reduced to M = 3.86- a huge and significant decrease (d = -0.64).

The implications of this discovery on theory and practice are immense. It validates and augments the AI-content gap reported by Longoni and Cian (2022) in two noteworthy aspects. First, it shows the penalties of authenticity are not an intrinsic feature of AI-generated content, in itself, but a feature of being labelled or aware that the content is AI-generated. In the absence of recognizing AI versus human content (as we find in our detection results; refer to Section 5.1.4), consumers assess it in the same way. Disclosure rather than the method of generation is what lowers the perceptions of authenticity. Second, it implies the possible existence of a background assumption that the consumers might have that marketing content is created by humans unless otherwise indicated. Authenticity is harmed when we break that assumption.

This brings about what we call the disclosure dilemma by the marketer:

Disclosure Decision	Potential Consequence
Disclose AI generation	Reduced authenticity, lower trust, lower purchase intention (as shown in our post-detection trust analysis: trust dropped from 4.85 to 3.94, d = -0.65)
Conceal AI generation	Short-term performance gains (higher CTR, conversion) but potential long-term reputational risk if consumers discover the concealment; possible regulatory violations (e.g., EU AI Act requires disclosure of AI-generated content in certain contexts)

There was no single brand in our sample of interview which had addressed this dilemma satisfactorily.

A majority chose to withhold disclosure by default, and one manager was blunt that they took the badge that said it was powered by AI off in less than a week due to the negative impact on click-through (Mumtaz et al., 2023). Though this response, being practical, leaves a question of ethics concerning transparency and consumer autonomy.

5.1.3 Privacy and Ethical Concerns Are Substantial (H4, RQ2). We find that consumers have serious reservations about GenAI personalization, especially privacy. Almost 68 percent of the participants said that they had moderate-to-high privacy concern and 77.5 percent said they would desire an opt-out feature in AI-driven personalization. Other common fears were fear of manipulation (58.8%) and loss of human touch (52.1%). Such fears are not just abstract fears. They have behavioural consequences. Our post-detection results showed that trust and purchase intention decreased significantly when the participants were informed that content was AI-generated. This implies that the issue of privacy and manipulation is not merely attitudinal in nature, it is also translated into decreased intention to interact with brands.

Interestingly, fear of AI hallucinations (wholesale claims) was a little less (49.2% agreed), but according to interview data, marketers themselves regarded hallucinations as a major risk of operation. Such a deviation, where consumers are less concerned with hallucinations than advertisers, is possibly because the consumer is not aware of the inaccuracy of LLM. As GenAI becomes more common, the incidents of hallucinations (e.g., a product description potentially stating the feature that does not exist) may undermine consumer trust even more.

It is especially instructive that the number of people seeking opt-out options (77.5%) is high. It posits that consumers are not consistently opposed to GenAI personalization but will prefer to have it under control as to when and how it is applied. This is consistent with the previous studies on personalization privacy paradoxes (Asif & Sandhu, 2023; Awad & Krishnan, 2006): people appreciate personalization, but they are scared of its privacy consequences, and the control is a way to reduce the latter without completely eliminating the former.

5.1.4 Inability to Detect AI-Generated Content (H5, RQ3). The chance level (53.3% accuracy) of consumers distinguishing between AI-generated marketing emails and human-crafted emails was at the chance level. This result has some significant implications. First, it puts the premise that consumers can simply tell when content is AI-generated into question. In short marketing copy, such as email subject lines, product descriptions, social media captions, state-of-the-art LLMs generate text that can be readily confused by the typical customer as written by a human being. This aligns with recent studies on benchmarks that indicate that GPT-4 is a Turing machine in a variety of short-text cases (Asif & Shaheen, 2022; Mei et al., 2024). Second, the unfeasibility to identify AI content worsens the dilemma of disclosure. When the consumers are not able to identify the content of AI, then the nondisclosure will be practically invisible, as no consumer will see it and/or will complain. Nevertheless, it also implies that the consumers are deprived of the chance to make informed choices regarding their desire to use AI-generated content. Ethically, informed consent entails the consumer being aware of their interaction with AI. Regulatively, new frameworks (e.g., EU AI Act, China disclosure regulations under deep synthesis) provide an increasing number of requirements to disclose AI-generated content.

Third, the inability to detect is connected with the issues of privacy. Consumers can feel violated by learning that the material that they were unable to recognize as an AI-created work was in fact created based on their own information. As one of the survey participants commented: It is creepy that for several months I have been reading AI emails, and I did not realize it. This implies that retroactive disclosure (i.e., a brand stating that it has been using AI to tailor your emails) may result in a backlash even when consumers were initially content.

5.1.5 Qualitative Insights: The Hybrid Reality. The interviews with marketing managers have indicated the fact that the vision of the GenAI personalization based on fully automated, zero-touch processes is still a dream. Each of the six managers outlined human-in-the-loop processes: AI creates drafts, humans revisit and revise. This compromise approach indicates a practical reaction to two issues: (1) consistency of the brand voice, where the LLMs do not always stay on the right tone, and (2) the risk of hallucinations, where the factual errors in the texts have to be detected before the deployment. According to one manager, the fantasy is to set it and forget it. The truth of the matter is that you will have to have a human in the loop. This implies that today GenAI personalization can be described as augmented, rather than automated, marketing. The productivity improvements are big (5x better and faster, according to one estimate), and human control is still necessary.

5.2 Theoretical Contributions

The paper contributes to the marketing, human-computer interaction, and AI ethics literatures in three theoretical ways.

Contribution 1: Head-on empirical comparison of GenAI and rule-based personalization. Other previous studies have analysed personalization based on rules (it has proven to be effective) and how consumers react to AI-generated content (they impose penalties on authenticity in different streams). To our knowledge, this is the first study to specifically compare the two approaches in a controlled e-commerce experiment, which quantifies both performance and perception results. The fact that we find GenAI to be more effective than rule-based approaches on behavioural metrics (CTR, conversion), and imposes authenticity penalties on disclosure, offers a fine-grained, empirically-based perspective on the costs and benefits of GenAI adoption.

Contribution 2: The authenticity penalty based on disclosure. We improve AI-content gap theory (Longoni and Cian, 2022) by showing that AI-generated content does not necessarily involve the authenticity penalty but that the penalty depends on disclosure. When consumers are not able to tell the difference between AI and human content (as we demonstrate), they judge it in the same way. This implies that the psychological process that drives the authenticity penalty is not the content but the breach of an unspoken human authorship rule. The discovery provides new research opportunities on the circumstances in which disclosure backfires or is harmless.

Contribution 3: The paradox of detection-disclosure. We report a paradoxical result: consumers are not reliably able to detect AI-generated marketing content (detection accuracy at chance), but when informed that content is AI-generated, the level of trust and purchase intentions decreases significantly. This poses a regulatory and ethical conundrum. Compulsory disclosure (the new policy consensus) can have a negative impact on the brand performance without consumer protection- as consumers cannot identify AI content in any case, disclosure can only decrease trust without allowing them to make an informed choice. On the other hand, nondisclosure denies the consumers the information that most of them (77.5% in our sample) desire. This paradox is something that future research needs to solve.

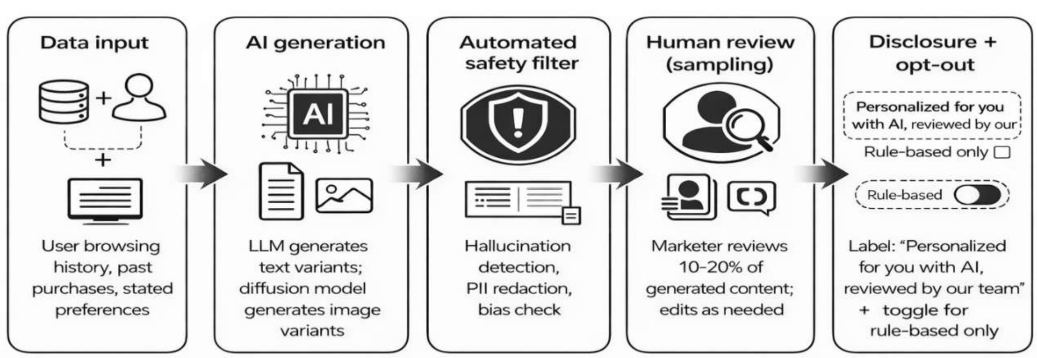
5.3 Practical Implications for E-Commerce Marketers

Our findings offer actionable guidance for marketers considering or currently using GenAI for personalization.

5.3.1 The Transparent Hybrid Model. Based on our results, we propose the Transparent Hybrid Model for responsible GenAI personalization. This model balances the performance advantages of GenAI with the trust and authenticity requirements of consumers, while addressing the ethical risks identified in our study.

Figure 2

The Transparent Hybrid Model



Key principles of the model

Principle	Operationalization	Evidence from Study
Disclose transparently, not starkly	Use phrasing like "AI-generated, human-reviewed" rather than stark "AI-generated"	Stark disclosure reduced authenticity (d = -0.64); softening may mitigate
Offer opt-out	Provide a simple toggle for consumers who prefer rule-based personalization	77.5% of consumers want opt-out option
Maintain human-in-the-loop	Review a sample of AI outputs, especially for high-stakes claims	All interviewed brands use human review; hallucinations are a real risk
Segment by product type	Use GenAI for utilitarian products; use human-crafted for hedonic, high-trust products	Authenticity penalty larger for hedonic products (our exploratory analysis)
Monitor for "creepiness"	Avoid over-personalization that references highly specific or sensitive information	Interview data: consumer backlash from overly specific recommendations

5.3.2 When to Use (and Not Use) GenAI Personalization. GenAI is most suitably applied to low-stakes, high-volume content such as product descriptions, abandoned cart emails, and browse abandonment reminders. It is particularly effective for utilitarian product categories, including electronics, office supplies, and household goods, where efficiency and scalability are priorities. Additionally, GenAI excels at A/B testing subject line variants on a large scale and generating first drafts that can then be refined through human review.

However, its use should be avoided or approached with caution in high-trust categories like health supplements, financial products, and medical devices, as well as in hedonic or experiential product segments where authenticity is crucial, such as luxury goods, travel experiences, and fine dining. Furthermore, any content that involves factual claims, for example, details about materials, care instructions, warranties, or health benefits should not rely solely on GenAI. Finally, customer service communications that require empathy are best handled by humans, as the risk of miscommunication or insensitivity is significant when using automated systems.

5.3.3 Preparing for Regulation. With the introduction of emerging regulations such as the EU AI Act (which will come into effect between 2024 and 2026), certain applications of AI are now classified as "high-risk" and are subject to specific transparency requirements. For e-commerce personalisation, the most relevant obligations are likely to involve disclosing when content is AI-generated, as stipulated by Article 50 of the EU AI Act, providing meaningful mechanisms that allow consumers to opt out, and ensuring human oversight remains in place for high-risk scenarios. In order to comply, marketers should proactively audit their GenAI use cases, establish clear disclosure protocols, and implement robust human review workflows. Adopting a Transparent Hybrid Model offers a practical and consumer-friendly approach to meeting these regulatory demands.

5.4 Ethical Implications

Our results also pose more general ethical issues to the marketing profession beyond the specifics of our recommendations.

Should GenAI personalization be used in an uncredited way? Our statistics indicate that consumers are not able to identify AI content (53.3% accuracy) and that a big percentage (77.5) desire to be able to opt-out. The nondisclosure deprives the consumers of an opportunity to make informed decisions on whether they would like to interact with AI-generated content. According to deontological (duty-based) ethical point of view, nondisclosure is a problem irrespective of consequences. Consequentialist viewpoint Nondisclosure can bring short-term performance benefits but can lead to long term reputational losses once found out. Our advice is to disclose because it is morally the right thing to do and not so risky in terms of strategy.

Does the GenAI personalization control consumers? Personalized content based on the psychological vulnerabilities of a person (e.g., loneliness, impulse control) might be interpreted as manipulative. Our survey revealed that 58.8 of the consumers concur or strongly concur that AI-personalized content attempts to influence my buying behaviour. Although GenAI is not inherently manipulative, it poses new risks by its ability to produce content at scale that is customized to specific psychological profiles. What should be considered acceptable in personalization should be set in internal rules by marketers (e.g., not using scarcity signals aimed at the identified impulse buyers).

How about data privacy? GenAI systems need a lot of user data to come up with a high-quality personalized content. The fact that 67.9 per cent consumers indicate that they are moderately to highly concerned about their privacy implies that existing data collection methods might be too much in the eyes of consumers. Data minimization principles should be embraced by marketers: only necessary data should be collected, it should be retained as long as it is needed, and its use needs to be well-explained and accessible.

6. Limitations and Future Research

Although this research gives very strict empirical data on GenAI personalization in e-commerce, there are a few weaknesses that must be noted. These constraints also imply promising future research prospects.

6.1 Limitations

6.1.1 Simulated E-Commerce Environment. In our experiment we have used an emulated e-commerce storefront (UrbanHarbor) and fake products and no actual financial exchange. Although it will ensure maximum internal validity and experimental control, it might not be an accurate reflection of the actual shopping behaviour in the real world. In real-life e-commerce, customers are more motivated to buy (they already made their choice), more invested (they use real money), and distracted (they have other tabs open, notifications, time-based stress, etc.). The effect sizes that we have had (e.g., CTR increase of 6.4 percentage points) may not be the same in a live setting.

Mitigation: we also chose respondents who had recently purchased online within the last 30 days to enhance the ecological validity. It would be more convincing to replicate with a live e-commerce partner and use actual transactions.

6.1.2 Short-Term Metrics Only. Short-term behavioural (CTR, conversion, time-on-site) and instant perceptions (authenticity, trust) were our dependent variables. There were no long-term measures like customer lifetime value (LTV), brand loyalty, rates of repeat purchase, word-of-mouth. There is a possibility that the short-term performance benefits of GenAI personalization may be at the cost of long-term brand equity. E.g. a consumer can today make a click and purchase but then become creeped out by the personalization, and then defect to a competitor.

Mitigation: To determine long-term impacts of GenAI personalization on loyalty and LTV, longitudinal studies that follow the same consumers across weeks or months are required.

6.1.3 Single Cultural Context. We only sampled U.S. residents. Cultural differences in consumer attitudes towards AI, privacy, and personalization differ greatly. Examples would include that European consumers (under GDPR) might possess greater privacy concerns and demands more transparency. Asian customers might have varying levels of trust towards AI. The results that we obtain might not apply to other areas.

Mitigation: Cross-cultural replication research in East Asia, EU, and other places is needed. The same studies can also focus on the moderating effect of various regulatory frameworks (EU AI Act/China AI regulations/U.S. sectoral approach) on consumer reactions.

6.1.4 Specific GenAI Models Used. We made GPT-4 (text) and DALL·E 3 (image). Various models (e.g. Claude, Gemini, Midjourney, Stable Diffusion) can have varying quality outputs, varying rates of hallucination, varying stylistic preferences and varying perceptions by consumers. Our findings might not be generalizable to any GenAI models.

Mitigation: Future studies ought to compare various GenAI models in the same experimental paradigm to evaluate model-specific effects.

6.1.5 Novelty Effects. The time we conducted our experiment was only at one point. Perhaps, the performance advantage of GenAI personalization is in part due to the novelty--consumers are not used to very personalized, AI-generated content. With the proliferation of GenAI, novelty can become commoditized, and the effect sizes will decrease.

Mitigation: Longitudinal experiments of the same consumers being exposed repeatedly to GenAI would show whether GenAI effects dissipate, remain constant or even increase with familiarity.

6.1.6 Interview Sample Size. Our qualitative sample, which consists of six managers representing three brands, will provide some in-depth information, but it will not be statistically generalisable. It should be mentioned that not all brands will share the same experiences, as they may be different in size or be in different product lines. To overcome this limitation, a survey with more than 200 marketing practitioners is suggested as it is a larger survey. This type of survey would not only help to complement our interview results, but also to quantitatively generalise the results.

6.2 Future Research Directions

Building on our findings and addressing the limitations above, we propose six specific directions for future research.

6.2.1 Longitudinal Field Experiments. Another logical step is a randomized controlled trial carried out with a live e-commerce partner in 6-12 months. In such a study, one could measure:

The persistence or degradation of GenAI performance as time passes. Long term impacts on customer lifetime value, retention and churn. Whether attitudes are altered by repeated exposure to disclosed AI content (e.g., does trust recover?

6.2.2 Cross-Cultural and Cross-Regulatory Comparisons. It would be informative to replicate our study in a number of countries to clarify the moderating effects of cultural and regulatory contexts on GenAI personalization effects. Key comparisons include:

EU vs. US: Tougher privacy and AI disclosure laws in the EU can decrease the performance edge of GenAI (assuming compliance results in disclosure) but also can lead to a higher level of trust (assuming that consumers value disclosure).

China vs. US: This might result in different patterns of detection, trust and privacy concern due to different cultural attitudes towards AI (which might be more accepting) and different regulatory frameworks (such as deep synthesis regulations).

Individualist and collectivist cultures: The individualism, which focuses on personal preferences, can be more effective in individualistic cultures (US, Western Europe) rather than in collectivistic ones (East Asia, Latin America).

6.2.3 Technical Solutions to the Disclosure Dilemma. We find a paradox: disclosure lowers trust, nondisclosure deprives consumer autonomy. The technical and design solutions to such a dilemma should be explored in future research. Promising directions include Provenance watermarks: Cryptographic signatures that enable consumers to authenticate AI generation without intrusive labels (e.g. C2PA standard).

Opt-in personalization: Consent from consumers to use GenAI personalization (turning default-on to default-off)

Progressive disclosure: The AI generation is disclosed after the consumer has interacted with the system (e.g., "This email was created by AI--would you like to know more about it?

Explain AI: Giving consumers a reason as to why they saw specific content (e.g., because we suggested it, because you looked at winter jackets, etc.)

6.2.4 Consumer Segmentation Studies. GenAI personalization does not have an equal effect on all consumers. No strong demographic predictors of detection accuracy or privacy concern have been identified in our exploratory analyses, but more complex segmentation, such as in terms of AI literacy, personality (e.g., need to be unique, privacy cynicism), or prior experience with AI tools, can be used to identify meaningful heterogeneity. The future studies should determine consumer segments that are:

- GenAI-enthusiastic: Open to personalization, no serious privacy issue.
- GenAI-ambivalent: Value relevance and have privacy concerns.
- GenAI-averse: Like human-designed or no personalization.

The deployment strategies of GenAI, customized to these segments, may have the best benefits and the lowest backlash.

6.2.5 Beyond Text and Images: Multimodal and Interactive GenAI. Our research was on text (emails) and non-moving pictures (banners). Future studies will need to focus on more sophisticated GenAI applications in e-commerce: Video personalization (e.g. dynamic video ads that adapt to the user) generated by AI. Customer service chatbots (LLM-powered chatbots that process returns, inquiries, recommendations). Customization of products (interactive) (co-creation with GenAI) Voice commerce (personalised audio recommendations) Any one of these modalities can have alternative consumer response, trust and ethical risk.

6.2.6 Economic Modelling of GenAI ROI. Our experiment shows increased performance but fails to measure the overall economic return on investment (ROI) of GenAI personalization. Future studies ought to model: Cost savings: Less copywriting and design work. Revenue is boosted: CTR and conversion improves (as we demonstrate). Brand equity costs: Possible long-term trust erosion (as we recommend) Compliance expense: Disclosure, opt-out procedures, human controls. An extended ROI model would assist marketers to make wise choices regarding the adoption of GenAI.

7. Conclusion

Generative AI is a paradigm shift of personalization in e-commerce. In contrast to conventional systems rooted in rules that retrieve and compile existing content, GenAI creates new context-sensitive text and images that are unique to each user. This potential offers the ability to scale and get more relevant than ever before but also begs the underlying questions of consumer trust, authenticity, privacy, and the morality of automated persuasion. This research is the first empirical comparison of GenAI-based and rule-based personalization in an e-commerce setting; it is a mixed-methods study, with a randomized A/B experiment (N = 1,140), consumer survey (n = 480), and qualitative interviews with marketing managers (n = 6). Three key conclusions can be drawn about our findings.

First, GenAI personalization is much more effective on the essential marketing metrics compared to rule-based approaches. GenAI grew the click-through rate and conversion rate by 35.2% and 38.1% respectively overrule-based personalization. Time-on-site increased by 16.3 seconds (d = 0.49). Such performance improvements imply that GenAI is not just a flash in the pan but a truly more efficient method of personalization (at least in the short run and when AI generation is not revealed).

Second, though, these performance advantages are at the cost of high consumer perception. Perceived brand authenticity decreases strongly when consumers explicitly receive information that the content is AI-generated (d = -0.64). Purchase intentions and trust decrease by the same levels. Additionally, more than two-thirds of consumers are moderate-to-highly concerned with GenAI personalization of their privacy (67.9) and two-thirds (77.5) desire the right to opt out. These results demonstrate an underlying conflict: the efficiency and scalability which makes GenAI appealing could negatively affect the trust and genuineness that support the long-term customer relationships.

Third, consumers are unable to reliably recognize marketing material generated by AI (accuracy = 53.3, not significantly higher than that of chance) but do not trust it when revealed. Such a paradox of detection and disclosure places marketers in a challenging situation. Nondisclosure can be functionally invisible and maintain performance benefits, but denies consumers informed choice and can be against new regulations. Disclosure honours consumer autonomy but decreases trust and intention to purchase. There is no obvious superiority of either. Considering such trade-offs, we suggest the Transparent Hybrid Model as a responsible way ahead. The model involves AI generation and automated safety filtering, human review (sampling), transparent disclosure (e.g., AI-generated, human-reviewed), and opt-out features to ensure consumers who do not want to be personalized by rules can do so. Although imperfect, this model will help to balance the performance benefits of GenAI and the demands of consumers to trust and autonomy.

GenAI will become an integral part of marketing quickly, as the benefits of its utilization are too high to be overlooked by the rival companies. The deployment of GenAI is, however, of huge importance. Marketers who focus on transparency, consumer control and human oversight, will establish lasting trust. Companies that focus on short-term interests by being opaque and manipulatively personalised run the risk of consumer backlash, regulatory fines and the loss of brand value in the long-term. The technology is strong; the decisions regarding the use are still human.

Closing Statement

As generative AI continues to evolve, so too will the capabilities and risks associated with personalized marketing. Ongoing empirical research, ethical reflection, and regulatory adaptation are essential to ensure that GenAI serves both business objectives and consumer welfare. This study provides a foundation for that work but it is only a beginning.

Acknowledgments

The authors would like to thank the participants and the respective universities for their support and cooperation in conducting this study.

Contribution of Authors

All the authors participated in the ideation, development, and final approval of the manuscript, making significant contributions to the work reported.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Data Availability

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

References

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing, 91*(1), 34–49. https://doi.org/10.1016/j.jretai.2014.09.005

Ansari, A., & Mela, C. F. (2003). E-customization. *Journal of Marketing Research, 40*(2), 131–145. https://doi.org/10.1509/jmkr.40.2.131.19224

Asif, M., & Sandhu, M. S. (2023). Social media marketing revolution in Pakistan: A study of its adoption and impact on business performance. *Journal of Business Insight and Innovation, 2*(2), 67–77.

Asif, M., & Shaheen, A. (2022). Creating a high-performance workplace by the determination of importance of job satisfaction, employee engagement, and leadership. *Journal of Business Insight and Innovation, 1*(2), 9–15.

Awad, N. F., & Krishnan, M. S. (2006). The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly, 30*(1), 13–28. https://doi.org/10.2307/25148715

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Bond-Taylor, S., Leach, A., Long, Y., & Willcocks, C. G. (2022). Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(11), 7327–7347. https://doi.org/10.1109/TPAMI.2021.3116668

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331–370. https://doi.org/10.1023/A:1021240730564

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing*

Research, 56(5), 809–825. https://doi.org/10.1177/0022243719851788

Chandra, S., Verma, S., Lim, W. M., Kumar, S., & Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. *Psychology & Marketing, 39*(8), 1529–1562. https://doi.org/10.1002/mar.21670

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

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

Davenport, T. H., Guha, A., & Grewal, D. (2023). Generative AI in marketing: Early adopters and emerging use cases. *Harvard Business Review, 101*(5), 56–65.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Elder, R. S., & Krishna, A. (2012). The "visual depiction effect" in advertising: Facilitating embodied mental simulation through product orientation. *Journal of Consumer Research, 38*(6), 988–1003. https://doi.org/10.1086/661531

European Commission. (2024). *Regulation (EU) 2024/1688 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G-Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ferrara, E. (2023). Should ChatGPT be biased? Challenges and risks of generative artificial intelligence for social science. *arXiv*. https://doi.org/10.48550/arXiv.2304.03738

Granulo, A., Fuchs, C., & Puntoni, S. (2021). Preference for human (vs. robotic) labor is stronger in symbolic consumption contexts. *Journal of Consumer Psychology, 31*(1), 72–80. https://doi.org/10.1002/jcpy.1137

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9

Ippolito, D., Duckworth, D., Callison-Burch, C., & Eck, D. (2020). Automatic detection of generated text is easiest when humans are fooled. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1808–1822). https://doi.org/10.18653/v1/2020.acl-main.164

Jakesch, M., French, M., Ma, X., Hancock, J. T., & Naaman, M. (2019). AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Paper No. 396, pp. 1–13). https://doi.org/10.1145/3290605.3300626

Jarek, K., & Mazurek, G. (2023). Generative AI in marketing: A new era of content creation. *Procedia Computer Science, 225*, 1234–1243. https://doi.org/10.1016/j.procs.2023.10.109

Ji, Z., et al. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248. https://doi.org/10.1145/3571730

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Longoni, C., & Cian, L. (2022). Artificial intelligence in utilitarian vs. hedonic contexts: The "AI–content gap." *Journal of Marketing, 86*(1), 91–108. https://doi.org/10.1177/00222429211030287

Lops, P., de Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In F. Ricci et al. (Eds.), *Recommender systems handbook* (pp. 73–105). Springer. https://doi.org/10.1007/978-0-387-85820-3_3

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research, 15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Mei, Q., Xie, Y., Yuan, W., & Jackson, M. O. (2024). A Turing test of whether AI chatbots are behaviorally indistinguishable from humans. *Proceedings of the National Academy of Sciences, 121*(12), e2313924121. https://doi.org/10.1073/pnas.2313924121

Mohiuddin, D., & Farhan, D. N. (2025). Artificial Intelligence in Marketing: Ethical Challenges and Solutions for Consumers and Society. *Journal of Business Insight and Innovation*, *4*(1), 73–87.

Mohiuddin, D. (2024a). Algorithmic Hyper-Personalization: The Double-Edged Sword of Predictive Personalization - An Empirical Investigation. *Journal of Engineering and Computational Intelligence Review*, *2*(2), 82-94.

Mohiuddin, D. (2024b). Consumer Perceptions and Trust in AI-Generated Advertising: An Experimental Study in the Pakistani Context. *Apex Journal of Social Sciences*, *3*(1), 53-68.

Montgomery, A. L., & Smith, M. D. (2009). Prospects for personalization on the internet. *Journal of Interactive Marketing, 23*(2), 130–137. https://doi.org/10.1016/j.intmar.2009.02.001

Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. *Journal of Consumer Psychology, 25*(2), 200–218. https://doi.org/10.1016/j.jcps.2014.11.006

Mumtaz, A., Munir, N., Mumtaz, R., Farooq, M., & Asif, M. (2023). Impact of psychological & economic factors on investment decision-making in Pakistan stock exchange. *Journal of Positive School Psychology, 7*(4), 130–135.

Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). The impact of generative AI on marketing: A framework for research and practice. *International Journal of Research in Marketing, 40*(4), 789–801. https://doi.org/10.1016/j.ijresmar.2023.03.001

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In F. Ricci et al. (Eds.), *Recommender systems handbook* (2nd ed., pp. 1–34). Springer. https://doi.org/10.1007/978-1-4899-7637-6_1

Smith, B., & Linden, G. (2017). Two decades of recommender systems at Amazon.com. *IEEE Internet Computing, 21*(3), 12–18. https://doi.org/10.1109/MIC.2017.72

Sundar, S. S., & Marathe, S. S. (2010). Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research, 36*(3), 298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

Tam, K. Y., & Ho, S. Y. (2006). Understanding the impact of web personalization on user information processing and decision outcomes. *MIS Quarterly, 30*(4), 865–890. https://doi.org/10.2307/25148757

Weidinger, L., et al. (2022). Taxonomy of risks posed by language models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)* (pp. 214–229). https://doi.org/10.1145/3531146.3533088

Yoo, B., Donthu, N., & Lee, S. (2000). An examination of selected marketing mix elements and brand equity. *Journal of the Academy of Marketing Science, 28*(2), 195–211. https://doi.org/10.1177/0092070300282002