



Marketing Science Institute Working Paper Series 2025

Report No. 25-135

## Predicting Behaviors with Large Language Model (LLM)-Powered Digital Twins of Consumers

Bingqing Li, Qihong (Owen) Wei and Xin (Shane) Wang

“Predicting Behaviors with Large Language Model (LLM)-Powered Digital Twins of Consumers” © 2025

Bingqing Li, Qihong (Owen) Wei and Xin (Shane) Wang

MSI Working Papers are Distributed for the benefit of MSI corporate and academic members and the general public. Reports are not to be reproduced or published in any form or by any means, electronic or mechanical, without written permission.

**Predicting Behaviors with Large Language Model (LLM)-Powered  
Digital Twins of Consumers**

Bingqing Li

Qihong (Owen) Wei

Xin (Shane) Wang\*

Bingqing Li (bingqing@vt.edu) is a doctoral student at Pamplin College of Business, Virginia Tech. Qihong (Owen) Wei (owenwei1997@connect.hku.hk) is a doctoral student at Hong Kong University Business School. Xin (Shane) Wang (xswang@vt.edu) is Professor of Marketing at Pamplin College of Business, Virginia Tech. \*corresponding author.

Date of Submission: Oct 6, 2025

## ABSTRACT

Digital twins of consumers have emerged as a promising approach to simulate consumer thinking, feeling, and decision-making. Grounded in the psychological theory, which conceptualizes behavior as a function of both personal traits and contextual factors, this research proposes and validates a dual-component framework for constructing Large Language Model (LLM)-based consumer digital twins. Fine-tuning on consumer-specific data including user-generated content allows the model to internalize individual traits, preferences, cognitive and behavioral patterns, while retrieval-augmented generation (RAG) dynamically incorporates information specific to consumer context at inference. We demonstrate and validate the framework with Amazon e-commerce data, constructing 304 personified digital twins and evaluating their performance in predicting purchase decisions and review contents. The digital twins can predict future purchases with an average accuracy of 86% and to generate product reviews with strong semantic alignment to actual consumer-generated content (cosine similarity above 0.94). By aligning LLM adaptation techniques with the foundational psychological theory about behavior, our method enables psychologically grounded simulations of individual-level consumer behavior at a scale. This research contributes to the literature on generative AI, synthetic agents, and digital twins in consumer research, and at the same time, offers a new methodology for theory-driven modeling and privacy-compliant personalization in practice.

*Keywords:* large language model, artificial intelligence, digital twin, generative AI, user-generated content, consumer behavior

## INTRODUCTION

The concept of a digital twin, which originated in industrial design and proliferated with the advent of Industry 4.0, describes a dynamic, data-driven virtual replica that simulates the behaviors of its real-world counterpart, whether an object, system, or process, in response to environmental stimuli (Fuller et al. 2020). The applications of digital twins have since expanded to sectors such as manufacturing (Qi and Tao 2018), urban planning (Mohammadi and Taylor 2017), and healthcare (Liu et al. 2019). More recently, they have begun to take root in business and marketing, showing promise for accelerating product time-to-market, providing constant consumer insights, enhancing managerial decision-making, and even increasing revenue by 10% (Borden and Herlt 2022). Among emerging applications, digital twins of consumers have garnered growing interest due to their potential to simulate how individuals think, feel, and act across a variety of consumption contexts (Duerst, Colborne, and Velosa 2022).

Akin to a digital twin of an aircraft engine, which uses design parameters, sensor data, and historical performance records to mimic how the engine would respond to new external stimuli (e.g., temperature shifts, air pressures, fluid dynamics, etc.), a consumer digital twin draws from personal data (i.e., demographic, psychographic, behavioral, and transactional) to construct an imitating entity. It simulates consumer cognitions and behaviors, enabling researchers and practitioners to interact with it and predict how a real consumer might react to campaigns, products, advertisements, and service interactions prior to deployment (Henley 2024). Beyond prediction of outcomes, these simulations can also offer insight into the processes of how the consumer forms judgments and makes decisions. Such simulating capabilities hold profound implications for understanding consumer psychology and decision-making, predicting

future consumer behaviors, and improving experiences along the consumer journey by providing dynamic representations of individuals or segments (Castro et al. 2024). Efforts to model consumer behavior are not new. Researchers and practitioners have relied on statistical models and machine learning algorithms to predict outcomes such as purchase likelihood, churn probability, or click-through rates (e.g., Ascarza 2018; Fong, Fang, and Luo 2015; Netzer, Lattin, and Srinivasan 2008). While these methods are foundational and often effective, they only provide static representations of consumers, embodied by numbers, choices, or clusters, which limit deeper understanding of how consumers think, feel, and decide. In contrast, a consumer digital twin is designed to function as an interactive simulacrum, or a dynamic agent, that can engage in human-like exchanges, adapt to new information, and exhibit behavioral patterns consistent with those of its real-world counterpart.

Achieving this level of dynamic adaptability and generative interactivity is beyond the capabilities of conventional models. The rise of generative artificial intelligence (GenAI) and large language models (LLMs) has come in time and fueled excitement around the creation of digital twins that emulate human behaviors (Toubia et al. 2025). Unlike traditional predictive models, LLMs, a core implementation of GenAI, possess the ability to generate responses, make inferences, and interact using natural language. These models can be fine-tuned to reflect individual traits, adapted to specific tasks, and deployed as interactive agents, making them a promising computational foundation for constructing individual-level digital twins. These models serve not only as “silicon samples” for academic research in marketing and other social sciences, but also as practical tools for industry, where marketers are beginning to use GenAI to build digital replicas of individual consumers so that they can test and refine offerings before deploying them in the marketplace (Korst, Puntoni, and Toubia 2025). In this sense, digital twins

and GenAI heighten each other's value in marketing (Castro et al. 2024). Nonetheless, despite the promise and excitement surrounding digital twins, there remains a lack of robust methodologies to operationalize and implement digital twins that simulate individual consumers with high fidelity.

In response to this emerging opportunity and methodological gap, we propose a methodological framework for constructing LLM-based consumer digital twins. Our approach is grounded in classical psychological theories that conceptualize behavior as a function of both the traits of the individual person and the properties of the task environment (e.g., Lewin 1936; Simon 1990). Accordingly, we develop a two-component framework that reflects this dual influence: a digital twin model combining fine-tuning and retrieval augmented generation (RAG). Specifically, fine-tuning an LLM on an individual's personal data enables the model to internalize the individual's enduring preferences, communication style, cognitive tendencies, and behavioral patterns; this component captures the influence of individual traits on behavior. Simultaneously, RAG equips the model with the ability to search, retrieve, and incorporate relevant external information during inference, simulating how consumers seek and process contextual information in decision-making tasks; this component reflects the influence of task environment on behavior. As such, these mechanisms enable a digital twin to generate responses that are both personalized and context-aware, achieving a more psychologically grounded simulation of consumer behavior.

This study makes several contributions to research. First, it advances the proliferating literature on LLMs and GenAI in consumer research and marketing (e.g., De Freitas, Nave, and Puntoni 2025; Epp and Humphreys 2025; Huang and Rust 2025). While prior research has widely explored prompt engineering (Goli and Singh 2024; Li et al. 2024), retrieval

augmentation (Arora, Chakraborty, and Nishimura 2025), and fine-tuning (Ye, Yoganarasimhan, and Zheng 2025) as parallel methods for adapting general-purpose LLMs to marketing contexts, their integrated application remains underexamined. Our study highlights the complementary strengths of each approach: fine-tuning enables LLMs to model individual-level heterogeneity with consumer-specific data, while RAG enhances context relevance by grounding responses in dynamic, product-relevant information. Also, this dual-component approach contributes to consumer behavior modeling research by operationalizing foundational psychological theories through LLM-simulated behavior that is shaped by the interaction of personal traits and task environments (Lewin 1936; Simon 1990). Our framework proposes a computational means of capturing this interplay. Moreover, this study contributes to the nascent stream of work on digital twins in marketing (e.g., Levy 2023; Muschkiet, Paschmann, and Nissen 2022), and particularly, to the emerging vision of the realization of digital twins with GenAI (Korst et al. 2025; Sperling 2025). This work introduces a theory-driven framework for the development and application of GenAI-powered digital twins in modeling consumer behavior. As synthetic data is increasingly embraced as a valuable tool in consumer behavioral research, our work advances the promising yet underdeveloped research frontier of synthetic agents that simulate consumer responses to marketing stimuli.

From a managerial perspective, this consumer digital twin methodology, with LLMs' generative capability, domain adaptability, and fine-tuning potential, enables firms to gain consumer insights, forecast behavioral responses, and optimize marketing mix. This approach mitigates the high cost of real-world marketing actions by enabling pre-deployment testing of campaigns, products, and offers through virtual consumer agents. Consequently, firms can enhance personalization, improve consumer experience, and ultimately increase marketing

efficiency and return on investment. At the same time, our proposed approach offers a privacy-compliant solution for consumer analytics in an era of diminishing third-party data access and rising regulatory constraints (e.g., General Data Protection Regulation and California Consumer Privacy Act). As cookie-based tracking becomes increasingly restricted, conventional consumer analytics face growing challenges in acquiring necessary data (e.g., page views, clicks, session duration, and ad interactions). Consumer digital twins, built through first-party data in consumer relationship management (CRM) systems and publicly available user-generated content (UGC), allow firms to personalize, forecast, and refine marketing decisions without invasive data collection. The consumer digital twin framework provides a scalable and adaptive solution for understanding preferences and predicting behaviors at an individual level, informing various marketing decisions while fully complying with contemporary data regulation standards.

In the following sections, we first review the relevant literature on LLMs, synthetic data, digital twins, as well as the theoretical background of the dual influence of personal traits and contextual factors on consumer behavior. We then introduce our methodological framework for constructing consumer digital twins using LLM fine-tuning and RAG. Next, we demonstrate the technical implementation and empirical application of the consumer digital twin framework using e-commerce data, detailing our dataset construction, model setup, inference procedure, and evaluation metrics. Finally, we discuss the implications of our findings, outline limitations, and suggest future research directions.

## **RELEVANT LITERATURE AND THEORETICAL BACKGROUND**

Large Language Models in Marketing and Behavioral Research

Large Language Models (LLMs), built on advanced neural networks embedded in transformer architectures (Vaswani et al. 2017), have become central to the rapid expansion of generative artificial intelligence (GenAI). Trained on massive corpora and consisting of billions of parameters, these models can perform a variety of tasks, such as summarization, inference, and content generation (Brown et al. 2020; Radford et al. 2019). Accessible pre-trained models, both open-source (e.g., Llama and DeepSeek) and proprietary (e.g., GPT and Claude), have democratized LLMs to disciplines beyond computer science and engineering.

In marketing and behavioral research, LLMs' capabilities in natural language processing, both generation and understanding, have attracted significant interest (Kshetri et al. 2024; Kumar et al. 2025). In natural language generation tasks, recent literature has demonstrated the ability of LLMs to produce effective marketing content that leads to higher engagement and conversion rates, including email communications (Angelopoulos, Lee, and Misra 2024), visual and audio content (Hartmann, Exner and Domdey 2025; Heitmann et al. 2025; Zhang, Zhou and Lee 2024), search engine optimization (Reisenbichler et al. 2022), and personalized advertisements (Kapoor and Kumar 2025). Even for artistic production and creative ideation, traditionally thought to be reserved for human mental faculty (Feuerriegel et al. 2024), LLMs are shown to be able to facilitate or even excel humans (e.g., Castelo et al. 2024; De Freitas et al. 2025; Tojin, Finkenstadt, and Folk 2023; Zhou and Lee 2024). In natural language understanding tasks, LLMs outperform both conventional lexicon-based and high-performing transfer learning models in sentiment classification accuracy and explainability (Krugmann and Hartmann 2024). They can also be used to identify which digital content headlines lead to higher engagement (Ye et al.

2025), predict the polarization of media content (Yoganarasimhan and Iakovetskaia 2024), and model consumer expectation and engagement with narrative media (Fong and Gui 2024).

These findings demonstrate the potential of LLMs to both understand and generate consumer-facing content across diverse platforms. Yet the utility of LLMs is not limited to consumer communication. Increasingly, LLMs are being deployed in research-facing roles, including the generation of synthetic data that simulates human participants. These applications open a novel approach for marketing and behavioral science to modeling and examining cognitive, emotional, and behavioral patterns of consumers. We next turn to the emerging domain of synthetic data and consumer digital twins.

### Synthetic Data and Consumer Digital Twins

With its generative capability, LLMs can simulate how humans would respond to questions or stimuli in different contexts and consequently create synthetic data that mimics the statistical properties of real-world human-generated datasets (Gui and Toubia 2023). For research examining human decision making and behavior, synthetic data is increasingly used to replicate or augment survey responses. For example, LLMs have been used to simulate human preferences and behaviors in economic experiments and social surveys (Argyle et al. 2023; Brand, Israeli, and Ngwe 2023; Horton 2023). In marketing, researchers explore how LLMs can assist in data generation in both quantitative and qualitative marketing research surveys (Arora et al. 2025; Qiu, Singh, and Srinivasan 2023). Recent research demonstrates that LLMs can conduct automated perceptual analysis in lieu of surveys with human participants (Li et al. 2024) and replicate human preferences in intertemporal choices (Goli and Singh 2024).

While synthetic data are capable of generating task-specific outputs, such as hypothetical responses to some survey questions, emerging work envisions a more versatile construct – synthetic agents, artificial entities designed to perform tasks or simulate behavior (Park et al. 2024). The former focuses on simulating *what* people do in a specific task, whereas the latter simulates *who* people are and *how* they behave across various situations. Consumer digital twins are a type of synthetic agent that is anchored in an individual’s personal data to replicate a specific consumer’s cognition, preferences, and decision-making patterns across time and contexts. The concept of consumer digital twins has been explored to improve consumer experience (Muschkiet et al. 2022) and reveal the drivers of brand affinity for individual consumers (Levy 2023). Recent efforts by Toubia et al. (2025) attempt to build digital twins with LLMs trained on individuals’ responses to an extensive survey of behavioral, psychological, and economic questions, creating a benchmark for training and evaluating LLM-based digital twins.

While extant studies demonstrate LLMs’ promising accuracy in replicating human perceptions and behaviors at aggregate level, current methods of generating synthetic data and building synthetic agents remain limited in their ability to capture the richness and variability of human behavior. For instance, Atari et al. (2024) show that LLMs ignore the psychological diversity of humans around the globe. Although adapting LLMs using techniques like prompt engineering and fine-tuning can partially mitigate such issues, synthetic data still tend to exhibit constrained variation, rendering LLM-based inferences unreliable (Bisbee et al. 2024). Indeed, findings suggest that LLMs show limited capability of predicting human behaviors at individual level (Qiu et al. 2023). Toubia et al. (2025) similarly observe that while digital twins are good at replication of average treatment effects, it is yet to know whether they can simulate non-normative behaviors and reflect the full diversity of domain-specific views. Moreover, existing

synthetic samples are rarely interactive, lacking the mechanism to seek information on demand, update beliefs and preferences, or engage in multi-turn reasoning. This restricts their utility for simulating dynamic consumer decision-making processes. In this research, we address these limitations by building behaviorally grounded digital twins that not only incorporate mechanisms for modeling individual heterogeneity but also dynamically integrate situational context.

## Dual Influence of Traits and Context on Consumer Behavior

Consumer behavior has long been understood as a product of the interaction between internal, stable personal dispositions and external, dynamic situational cues. This foundational view is formalized by Lewin's (1936) canonical equation  $B = f(P, E)$ , where B is behavior, P is person, and E is environment. This underpins many modern behavioral theories that emphasize person-situation interactions in shaping how individuals evaluate options, form preferences, and make choices. For example, the Cognitive-Affective Personality System theory explains the invariance of personality and the variability of behavior across situations by encoding situational features (Mischel and Shoda 1995). In consumer decision-making, this principle is reflected in the constructive choice framework (Bettman, Luce, and Payne 1998), which argues that preferences are not merely retrieved but actively constructed in response to contextual and task-specific constraints. Similarly, the adaptive decision-making model (Payne, Bettman, and Johnson 1993) suggests that consumers shift their strategies based on environmental conditions such as time pressure, complexity, or attribute framing. As summarized by Simon (1990), behavior is shaped by the interaction between the properties of the human information-processing system (e.g., traits, goals, and preferences) and the properties of task environments

(e.g., products, brands, and decision frames). To faithfully model consumer behaviors, both components are necessary.

However, few computational models for consumer behavior, from conventional statistical models to advanced LLM-based ones, have explicitly operationalized both individual traits and context as simultaneous drivers of behavior. To create more psychologically plausible simulations of consumer behavior, it is necessary to model both individual-level heterogeneity and contextual awareness. A digital twin framework grounded in person-environment theory offers a path forward. By simulating consumers as agentic entities whose decisions emerge from the dynamic interplay between enduring traits and evolving environments, such models better reflect how consumers actually think, feel, and act in the consumption contexts. In the ensuing subsection, we introduce the technical background on which LLM adaptation methods are suitable for operationalizing this dual-influence perspective.

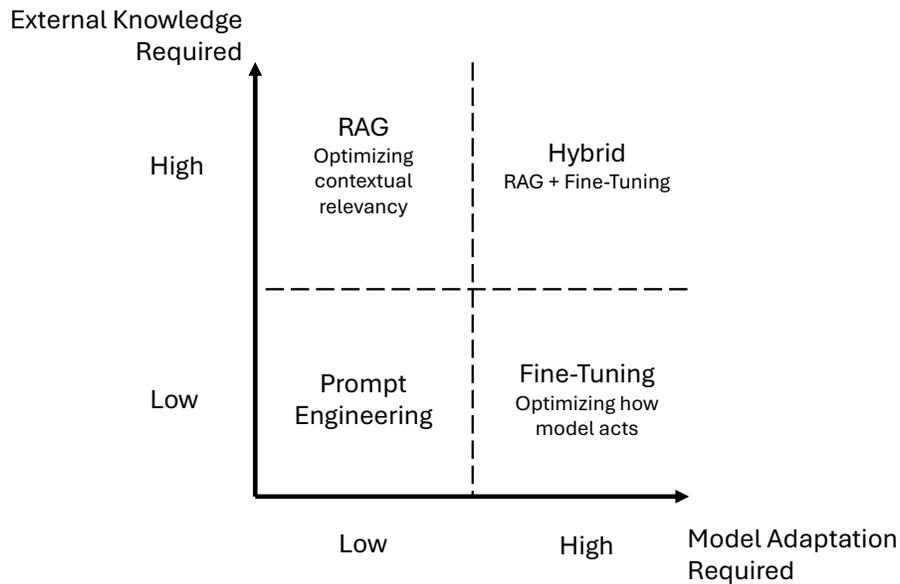
### Adaptation Techniques for Large Language Models

There are several methods to adapt general-purpose LLMs for specific roles, domains, and tasks. These adaptation strategies vary in their degree of model intervention and their reliance on external knowledge, leading to distinct optimization goals: some focus on improving how the model generates responses (e.g., linguistic style, reasoning pattern, and behavioral tendency), while others aim to enhance the relevance and factual grounding of outputs in specific decision contexts (Gao et al. 2023), as summarized in figure 1. The most widely used method is prompt engineering, through which carefully designed inputs guide the model to produce desired outputs without changing the model's underlying parameters (Reynolds and McDonnell 2021).

While easy to implement, prompt engineering offers limited modification and control over how the model reasons, leaving it insufficient for building digital twins that require a high level of personalization. Two more advanced methods, fine-tuning and retrieval-augmented generation (RAG), are particularly relevant for simulating consumer behavior, as they allow the model to reflect both personal traits and contextual cues.

**FIGURE 1**

**DIFFERENT LLM ADAPTATION TECHNIQUES**



Fine-tuning modifies the internal parameters of a pre-trained model by further training it on a small, specialized dataset (Howard and Ruder 2018). This process allows the model to internalize domain knowledge, new linguistic patterns, or user-specific preferences; hence, the model captures a more enduring modification on how it acts. In the context of digital twins and the broader domain of synthetic agents, fine-tuning enables the LLM to simulate stable traits that remain consistent across different decision contexts. Evidence shows that the reliability and

validity of synthetic personality traits are stronger in fine-tuned LLMs and that these traits can be adjusted along desired dimensions to mirror specific human personality profiles (Serapio-García et al. 2023).

In contrast, RAG augments a base LLM with access to an external knowledge retrieval system, allowing it to dynamically incorporate relevant, up-to-date information at inference time (Lewis et al. 2020). Rather than learning everything during training, RAG empowers the model to “look up” domain-specific or situationally relevant facts during generation. This is particularly important for simulating the influence of environmental stimuli on behavior, as consumers often make decisions in response to external content, such as product descriptions, social cues, reviews, and contextual constraints. Recent research highlights that RAG-generated synthetic data exhibits higher knowledge grounding and recovery rates compared to static generation (Arora et al. 2025).

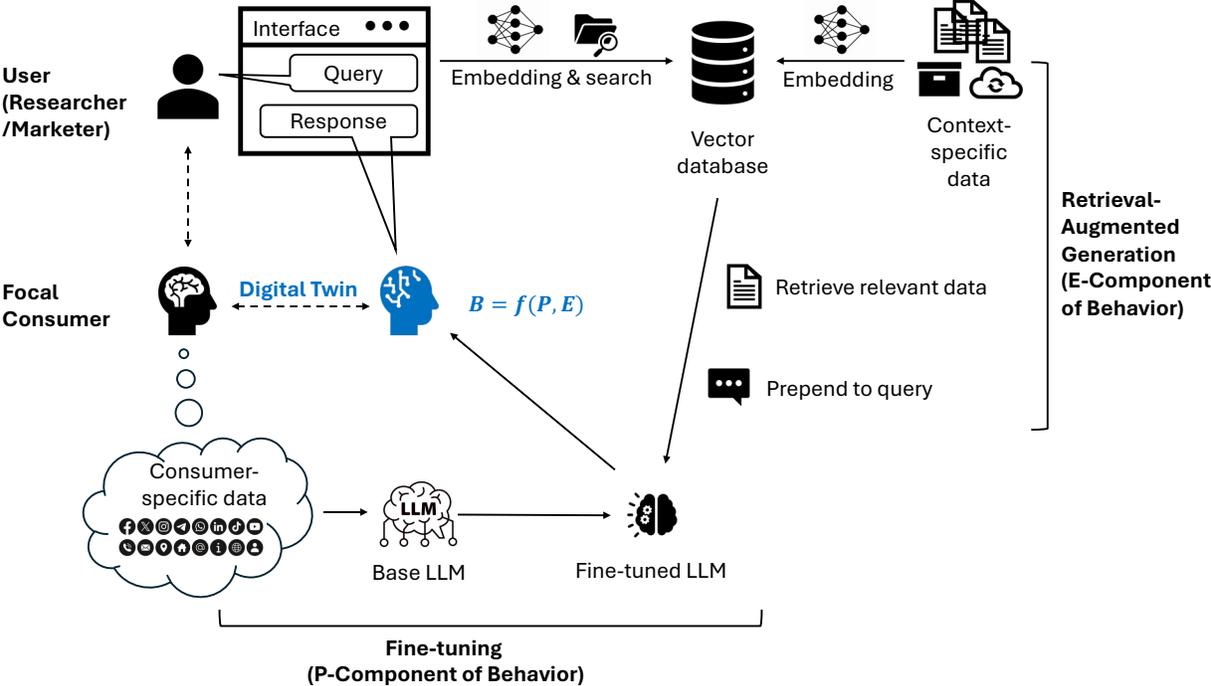
In this research, we adopt a hybrid approach to build digital twins of consumers: fine-tuning LLMs on historical personal data to internalize individual preferences and style, while using RAG to ground each decision in relevant product or situational information. This dual-component adaptation allows our simulated digital twins to reflect both enduring traits and dynamic contexts, aligning with psychological models of behavior as a function of person and environment.

## **METHODOLOGICAL FRAMEWORK**

We adapt large language models (LLMs) to construct consumer digital twins that can emulate individual patterns of thinking, feeling, and behaving in marketing-related contexts.

Figure 2 illustrates how digital twins can be realized through a combination of fine-tuning and RAG. Fine-tuning adapts a base LLM to consumer-specific data so that the LLM internalizes the individual’s traits, preferences, and behavioral tendencies; this is the P-component in Lewin’s (1936) behavioral model,  $B = f(P, E)$ . In parallel, RAG connects the LLM to context-specific data, such as product catalogs, brand reviews, or competitor data, enabling the model to simulate how a consumer searches for and incorporates relevant information during decision-making; this resembles the E-component of modeling consumer behavior. The following part details each methodological component of this framework.

**FIGURE 2**  
**METHODOLOGICAL FRAMEWORK**



Build Consumer Persona with Fine-Tuning

The first step is to “personify” a based LLM. The data for fine-tuning should provide insights into who the focal consumer is and how the consumer typically responds to marketing activities. Relevant data sources include sociodemographic attributes, purchase histories, and this consumer’s user-generated content (UGC). The more comprehensive and multidimensional the training data, the more behaviorally grounded the resulting digital twin becomes. While structured data from CRM systems and surveys provide demographic and transactional signals, unstructured UGC such as online reviews and social media posts reveal the consumer’s language, affect, beliefs, and decision-making logic in their own words and context.

A base LLM can be fine-tuned using several techniques. One widely used method is instruction fine-tuning, also known as supervised fine-tuning (SFT), in which the model is trained on instruction-response pairs to align its outputs with human intent and improve task specificity. Further alignment can be achieved through reinforcement learning from human feedback (RLHF), as demonstrated by Ouyang et al. (2022), where model outputs are iteratively refined based on preference judgments. To improve computational efficiency, Low-Rank Adaptation (LoRA), introduced by Hu et al. (2021), injects small trainable adapter layers into a frozen base model, allowing for parameter-efficient fine-tuning. More recently, QLoRA (Dettmers et al. 2023) has extended this approach by enabling fine-tuning on quantized models, significantly reducing resource demands and making training feasible on consumer-grade hardware. These methods form the foundation of fine-tuning pipelines, offering flexibility in balancing accuracy, efficiency, and scalability.

For constructing consumer digital twins, compact LLMs with approximately 7 to 10 billion parameters are sufficient, balancing the model capability, speed, and computation cost.

Representative models include Claude 3 Haiku, GPT-3.5 Turbo, Llama 2-7B, Mistral 7B, and Qwen-7B. Proprietary models usually support fine-tuning through API-based instruction fine-tuning. Open-source models, in contrast, provide greater flexibility, supporting advanced customization via LoRA, QLoRA, or full-parameter tuning. Although setting up the pipeline for open-source fine-tuning might be more complex than directly calling API, it enables local deployment and open-ended control. The prerequisite for fine-tuning is data formatting. Most frameworks follow the prompt-response or instruction-input-output format. As such, consumer-specific data must be pre-processed and converted to the appropriate format before model training.

Among a variety of consumer-specific data to personify an LLM, we consider UGC created by consumers especially critical. Consumers are often represented with artificial personas that are defined by inanimate tags and labels (e.g., female, 35 years old, pro-environmental, etc.); while useful, they lack psychological realism and fail to replicate the expressive diversity of actual consumers. To address these limitations and better capture individual-level heterogeneity, we turn to consumer UGC, such as product reviews and social media posts, that offer a rich and native source of psychological and behavioral data. UGC has long been used to infer consumer preferences, brand perceptions, choice patterns, and behavioral tendencies (e.g., Netzer, Lemaire, and Herzenstein 2019; Timoshenko and Hauser 2019; Tirunillai and Tellis 2014; Wang et al. 2022). Compared to structured data such as transactions and surveys, UGC captures consumers' thoughts, beliefs, experiences, and interactions in naturalistic settings, making it a more ecologically valid basis for modeling their decision-making processes and behavioral outcomes (Netzer et al. 2012). Thus, in this study, we use UGC to personify LLMs, enabling them to reflect individual-level heterogeneity as a step further toward consumer digital twins.

## Inject Domain Knowledge with Retrieval-Augmented Generation

While fine-tuning enables the LLM to internalize consumer-specific traits, preferences, and behavioral tendencies, effective consumer simulacra also requires the ability to acquire up-to-date, product-relevant knowledge. To embed the consumer digital twin within realistic decision-making contexts, we implement RAG to equip the LLM with dynamic information access via an external knowledge base, allowing the model to incorporate relevant product, brand, or category information to generate contextualized responses.

Integrating RAG into the digital twin framework starts with the construction of a knowledge base. It is tailored to the marketing context and may include structured and unstructured content such as product specifications, firm websites, promotional materials, consumer service FAQs, online reviews from other consumers, and other data sources reflecting brand, product, or category knowledge. These documents are embedded into a high-dimensional vector space, the same space where user queries are subsequently indexed for semantic search. This setup ensures that information retrieval is context-aware and continuously updatable, without retraining the underlying LLM.

The RAG architecture consists of two components: retrieval and generation. When a user interacts with the digital twin and sends a query, the retriever first encodes the input query and identifies the top- $k$  most semantically relevant information from the knowledge base (Gao et al. 2023). The retrieved content is prepended to the original user query and passed together to the generator, where the fine-tuned LLM produces a final response conditioned on both the consumer-specific model parameters and the retrieved contextual knowledge.

## Interact with the Digital Twin

A key advantage of LLM-powered digital twins lies in “interactivity”, setting them apart from conventional consumer profiling, propensity scoring, and predictive models. Rather than function as passive records of consumer attributes and past behaviors, digital twins act as agentic simulators. Marketers can directly interact with these virtual agents to forecast and test potential consumer responses. This interactivity transforms consumer analytics and modeling from retrospective, descriptive tools into generative, diagnostic instruments.

The interaction with a digital twin can take various formats, from structured prompts to open-ended dialogues. Marketers may input hypothetical scenarios such as new product introductions, pricing changes, and promotional messages, and observe how the digital twin responds. These responses can then be analyzed to infer likely behavioral outcomes (e.g., purchase intent, and click-through likelihood), as well as underlying cognitive or emotional reactions (e.g., sentiment, trust, and confusion). Moreover, the LLM-powered twin enables conversational probing, where marketers can iteratively refine messages and strategies. Such simulations allow firms to explore consumer journeys and perform pre-deployment testing of products and marketing actions without incurring real-world costs or risks.

Within the proposed methodological framework, the consumer digital twin is both personified through fine-tuning on individual UGC and contextualized via RAG-enabled access to external knowledge. Its outputs are therefore shaped by both the inherent characteristics of the simulated individual and the situational context of the marketing environment, enabling realistic, dynamic, and personalized simulation outputs.

## EMPIRICAL DEMONSTRATION

We showcase the proposed framework via an application of the consumer digital twin methodology in e-commerce context. We evaluate whether LLM-powered digital twins can effectively simulate individual consumers' purchase decisions and generate review content that reflects their preferences and styles. These two tasks operationalize the marketing goals of predicting behavioral outcomes and cognitive or affective reactions, respectively.

### Data Collection and Preprocessing

The data are collected from Amazon.com, the largest e-commerce platform globally. Our data source is the Amazon Reviews Dataset maintained by the McAuley Lab, including approximately 571.5 million reviews written between May 1996 and September 2023 (Hou et al. 2024). Each entry includes detailed information on both reviews (e.g., reviewer ID, review content, star rating, and helpfulness votes) and products (e.g., category, price, descriptions, specifications, and images).

The raw data is structured at the individual review level. To construct digital twins that reflect individual consumer's behavior and preferences, we first aggregate reviews at the reviewer level, resulting in a pool of 54.5 million unique consumer IDs. Review counts follow a right-skewed distribution, with a median of 136 per consumer ( $SD = 158$ ). To ensure sufficient data for fine-tuning robust consumer models, we retain only consumers with at least 400 reviews, yielding 1,021 qualified users. We further exclude users whose reviews are limited to books and

Kindle products. This exclusion criterion is based on two considerations: first, these users may systematically differ from general online shoppers and behave more like Goodreads (a book review platform) users; second, excluding them supports the goal of building category-agnostic digital twins. The final sample consists of 304 focal consumers, each contributing on average 552 reviews ( $SD = 138$ ), spanning an average of 3.71 product categories ( $SD = 1.79$ ). The mean review length is 38 words ( $SD = 52$ ).

In addition to review texts, we enrich each consumer's fine-tuning dataset with gender data. Specifically, we use consumer' usernames to infer gender. These names are then classified using the *facebook/bart-large-mnli* model hosted on Hugging Face, a natural language inference model capable of performing zero-shot classification (Lewis et al. 2019; Yin, Hay, and Roth 2019). Based on this pipeline, we infer gender for 304 users in the final sample: 150 are classified as female and 154 as male. Gender labels are incorporated into the fine-tuning input format to help guide the model to better emulate consumer traits.

To evaluate performance, we construct a testing set for each consumer composed of: (1) their 50 most recent reviews and associated products (i.e., hold-out review entries per consumer), and (2) 50 products randomly selected from the pool of products that do not appear in their purchase history. These 100 instances (50 purchased, 50 never-purchased products) serve as the ground truth for assessing the digital twin's predictive accuracy. With some overlapped purchases among the consumers, the entire testing dataset is constituted of 14,863 unique products. The average testing product has a numerical rating of 4.46 ( $SD = 0.36$ ), a total number of 505 reviews ( $SD = 1,978$ ), and a price of 23.46 USD ( $SD = 49.91$ ). During inference, product knowledge is injected via RAG to simulate real-world decision-making. Each product's contextual information (21 features in total), such as ratings, prices, categories, and aggregated

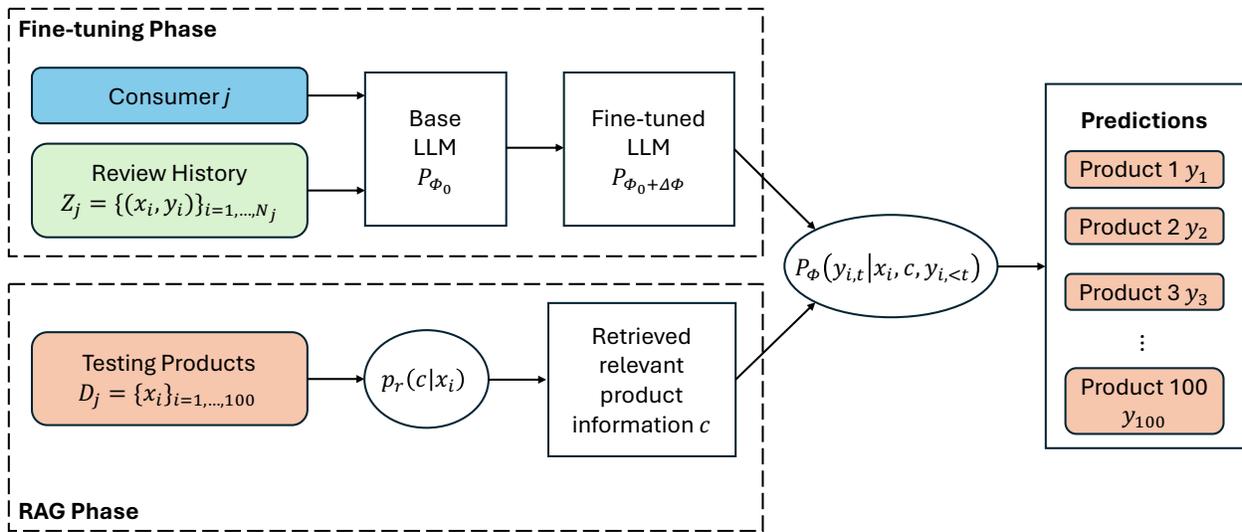
reviews, is embedded in the knowledge base and dynamically retrieved as part of the digital twin’s reasoning process.

### Model Setup and Inference Procedure

With the dataset organized at both consumer and product levels, we proceed to fine-tune an LLM on individual review history and evaluate the resulting digital twin’s ability to simulate purchase decisions and review behavior. Figure 3 outlines the implementation pipeline for constructing and testing consumer digital twins in the e-commerce context.

**FIGURE 3**

**IMPLEMENTATION OF E-COMMERCE CONSUMER DIGITAL TWIN**



The first step is to format the consumer review history according to the base model requirement. Given that consumer digital twins are use-case-specific rather than general-purpose agents, we opt for a lightweight model with adequate capability while being computationally

efficient. In this study, we adopt Qwen 2.5 7B, an open-source LLM comprising 7.61 billion parameters; and Qwen follows an instruction-input-output template. Accordingly, we develop a Python script to automatically convert each consumer’s historical reviews (excluding the 50 most recent reserved for validation) into fine-tuning records adhering to this structure.

Next, we set up the scheme for fine-tuning the LLM. We deploy the fine-tuning process using LoRA, a parameter-efficient tuning method. Let  $P_{\Phi_0}(y|x)$  parameterized by  $\Phi_0$  be a pre-trained LLM. When adapting this generic model to a digital twin simulating a particular consumer’s decision-making pattern and review-writing behavior, we fine-tune this model with the review history,  $Z_j = \{(x_i, y_i)\}_{i=1, \dots, N_j-50}$ , of the focal consumer  $j$ .  $x_i$  is a natural language query, inquiring about the consumer’s purchase and review regarding the product  $i$ ; and  $y_i$  is the corresponding review. Fine-tuning the LLM entails minimizing the cross-entropy loss on the training data (Hastie, Tibshirani, and Friedman 2009). In each round of training batch of fine-tuning, the initial pre-trained weights  $\Phi_0$  are updated by learning a low-rank perturbation  $\Delta\Phi(\Theta)$ , where  $\Theta$  contains the trainable LoRa parameters with  $\dim(\Theta) \ll \dim(\Phi_0)$ . The optimization objective maximizes the log-likelihood of generating the target review sequence token-by-token, conditional on the query:

$$\max_{\Theta} \sum_{(x_i, y_i) \in Z_j} \sum_{t \in \text{len}(y_i)} \log (P_{\Phi_0 + \Delta\Phi(\Theta)}(y_{i,t} | x_i, y_{i, < t})) \quad (1)$$

where  $t$  denotes each token in the target response sequence  $y_i$  in the training set  $Z_j$  and  $y_{i, < t}$  denotes the preceding tokens.

Specifically, we set the rank of the low-rank matrices to 8, a default balancing between expressiveness and computational efficiency. LoRA dropout rate is set to 0 to reduce the randomness of updates. For training configuration, we retain a default learning rate of 5e-5 and

experiment with different training epochs ranging from 3 to 10, a recommended range for small datasets. During inference, the temperature is set to 0.1, ensuring deterministic behavior by consistently selecting the most probable next token while preventing overly conservative predictions. This setting enhances output reproducibility, which is ideal for modeling stable digital consumer behavior.

Once the consumer-specific model is fine-tuned, we prepare the vectorized knowledge base for the testing products to support inference. Unlike fine-tuning, RAG does not require data to conform to a specific format. The product knowledge base includes both structured data (e.g., price, rating, category) and unstructured data (e.g., consumer reviews and product descriptions). All textual data are embedded into a high-dimensional vector space using BGE-M3, a multi-lingual embedding model via self-knowledge distillation (Chen et al. 2024).

We then build the pipeline to integrate the retrieved information as contextual knowledge. In making inference about the decision to purchase a certain product and subsequently writing a review for the purchased product, RAG first retrieves and summarizes relevant product information. This is the retrieval component, which assigns a relevance score  $p_r(c|x_k)$  to content chunk  $c$  based on its semantic similarity to the query  $x_k$  about the product  $k$  in the testing set  $T_j = \{x_k\}_{k=1,\dots,100}$  and returns top two truncated distributions. This retrieved content chunks are appended to the input prompt and then passed to the fine-tuned model, which produces output sequence based on the input prompt  $x_k$ , the retrieved content  $c$ , and previous tokens generated  $y_{k,<t}$ , denoted as  $P_\phi(y_{k,t}|x_k, c, y_{i,<t})$ . The inference objective for consumer  $j$  on a testing product  $k$  is defined as:

$$\hat{y}_{k,j} = \underset{y}{\operatorname{argmax}} \prod_{t \in \operatorname{len}(y_k)} \sum_{c \in \mathcal{Z}} p_r(c|x_k) P_\phi(y_{k,t}|x_k, c, y_{k,<t}) \quad (2)$$

where  $\hat{y}_{k,j}$  represents the most likely output sequence that contains a binary purchase decision (“yes” or “no”), and a review if the product is predicted to be purchased. The algorithm of building one LLM-powered digital twin is shown in appendix A.

Figure 4 shows an example prompt and the corresponding output from a fine-tuned consumer digital twin. The system message primes the digital twin by presenting a high-level summary of the consumer’s purchase history and introduces the prediction task. The user prompt

**FIGURE 4**

EXAMPLE OF A PAIR OF INPUT PROMPT AND OUTPUT RESPONSE



You are a customer who frequently purchases ['Clothing\_Shoes\_and\_Jewelry', 'Home\_and\_Kitchen' (INSERT CUSTOMER FREQUENTLY BOUGHT PRODUCT CATEGORY)]. Would you buy the product [INSERT PRODUCT INFORMATION]? First answer Yes or No. If Yes, please write a review for the product. If No, please just reply No.



Amazon Essentials Women's Long-Sleeve Lightweight Crewneck Sweater (Available in Plus Size), B082LNSCCJ



Yes.  
Very soft and comfortable sweater. Flattering fit and color is very pretty.



Verified Purchase.  
Review: “Color is very pretty and fits just right.”

specifies a new product and its identifier (ASIN), which are used to first fetch relevant product information from the RAG system. The consumer digital twin then makes a purchase decision and generates a review based on the consumer's purchase and review history, as well as product information.

## Main Results

We make inferences on 100 testing products for each of the 304 consumers according to the aforementioned protocol. We evaluate the predictive performance of consumer digital twins on two key tasks: first, a binary prediction on whether the consumer would purchase a product; second, an open-ended textual prediction on what review the consumer would write about the product.

*Predictive Accuracy of Purchase Decisions.* For binary predictions, we evaluate a total of 30,600 data points (i.e., 100 inferences for each of the 304 consumers) using standard classification metrics, including accuracy, precision, recall, and area under the ROC curve. Table 1 presents the evaluation results of the digital twin model in comparison to a base LLM and a base LLM augmented with product information. The digital twin model achieves an average accuracy of 0.859, precision of 0.960, and area under ROC curve of 0.859, representing a significant improvement over the base LLM, which yields an accuracy of 0.541. While connecting the base LLM with relevant product information improves its performance to an accuracy of 0.623, it is not comparable to the digital twin, which benefits from both contextual

awareness via product information and personalization through consumer-specific fine-tuning. The performance gain is particularly notable in recall, which increases from 0.085 in the base LLM to 0.745 in the digital twin, reflecting an improved ability to correctly identify positive cases (i.e., true purchases). The base model and the model with access to product information, when lacking the knowledge of the consumer’s inherent preferences, are too conservative to make positive predictions.

**TABLE 1**  
PERFORMANCE COMPARISON ACROSS MODELS

| Evaluation Metrics | Base LLM | Base LLM + Product Info | Digital Twin |
|--------------------|----------|-------------------------|--------------|
| Accuracy           | 0.541    | 0.623                   | 0.859        |
| Precision          | 0.749    | 0.757                   | 0.960        |
| Recall             | 0.085    | 0.288                   | 0.745        |
| F1-Score           | 0.141    | 0.361                   | 0.807        |
| AUC-ROC            | 0.541    | 0.623                   | 0.859        |

*Semantic Fidelity of Product Reviews.* For textual predictions, we assess the semantic similarity between the model-generated reviews and the actual consumer-authored reviews. This evaluation indicates whether the digital twin is able to not only predict what the consumer would buy but also replicate how they would articulate their experience. This analysis is restricted to the 10,520 true positive cases (i.e., the consumer did purchase, and the model predicted “yes”). We employ two approaches for semantic comparison.

First, we use a classic vector-based method as a baseline. Each review is embedded using a pre-trained Word2Vec model (trained on the Google News corpus), and semantic similarity is computed as the cosine similarity between the word vectors of the actual and simulated reviews. This approach assesses the lexical-level semantic alignment. The cosine similarity scores across

the matched pairs of reviews have a mean similarity of 0.943 with a standard deviation of 0.052, indicating that the digital-twin-generated reviews preserve core semantic content and verbal habits of the real consumers.

Second, we employ contextual embedding model, which further captures lexical ordering, deeper semantic meaning, and contextual dependencies in language. Specifically, we use Sentence-BERT within the BGE-M3 embedding space to quantify the semantic alignment between generated and actual consumer reviews. The model yields an average cosine similarity of 0.650 with a standard deviation of 0.125, indicating that the digital twins generate review content that reflects the original tone, sentiment, and evaluative detail of the consumer. Compared to co-occurrence-based models such as Word2Vec, contextual embeddings provide a more stringent assessment, as they evaluate meaning in relation to the full linguistic structure rather than mere word co-occurrence. Together, these results provide converging evidence that consumer digital twins not only learn what a person would purchase but also how they would describe it, thereby capturing both behavioral and expressive dimensions of consumer behavior.

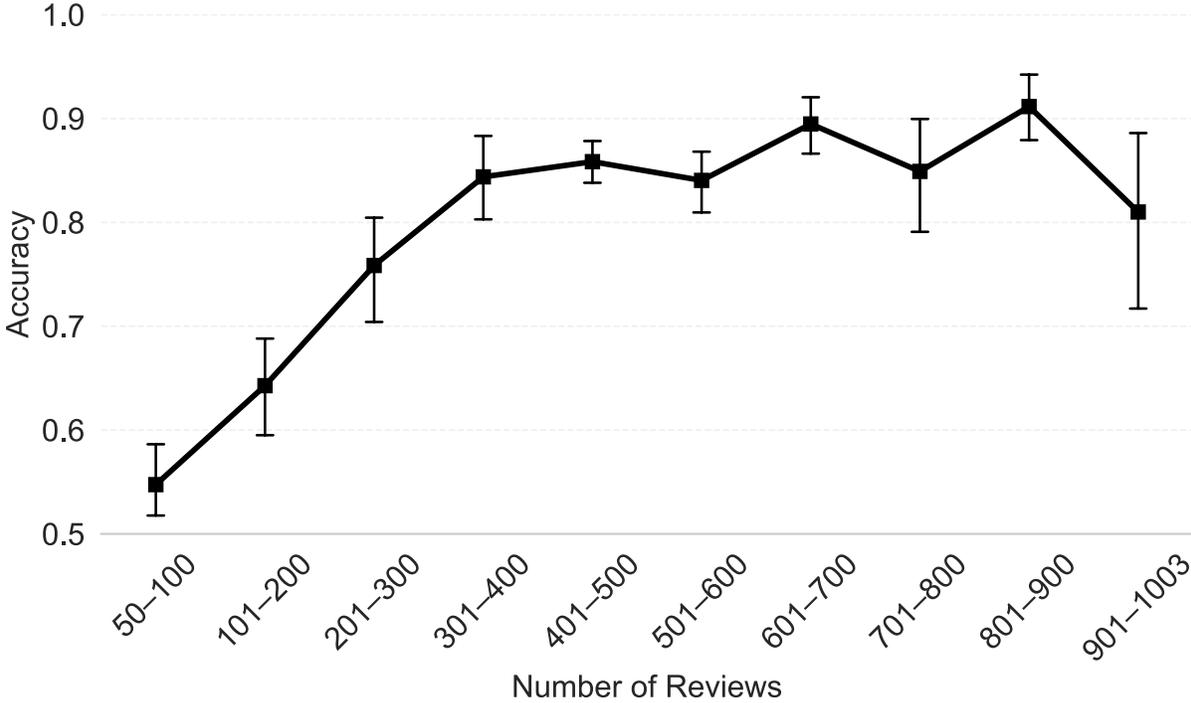
#### Heterogeneity in Model Performance

*Effect of Review Volume on Digital Twin Accuracy.* To examine how the volume of consumer-specific data influences the predictive performance of digital twins, we conduct a sensitivity analysis on the number of reviews used for fine-tuning with model accuracy. Among the 304 consumers included in our main analysis (each with more than 400 reviews), the correlation between review volume and accuracy is small and statistically non-significant ( $r = 0.046$ ,  $p\text{-value} = 0.426$ ), suggesting limited variation above the baseline threshold.

To explore model behavior below this threshold, we conduct a focused analysis on low-data scenarios. Because the population of consumers with fewer than 400 reviews is large, fine-tuning digital twins for all of them would be computationally prohibitive. Instead, we randomly sample 50 consumers in each of the following bins: 50-100, 101-200, 201-300, and 301-400 reviews. We do not go further below 50 reviews for the fine-tuning data as recommended (OpenAI n.d.). We apply the same modeling pipeline and evaluate performance on 100 held-out product inferences per consumer. As shown in figure 5, accuracy improves substantially with the increasing number of reviews, rising from 0.547 for consumers with around 50-100 reviews to 0.834 for those with 301-400 reviews. This result suggests that richer behavioral history enhances the model’s ability to internalize individual preferences and simulate decision-making

**FIGURE 5**

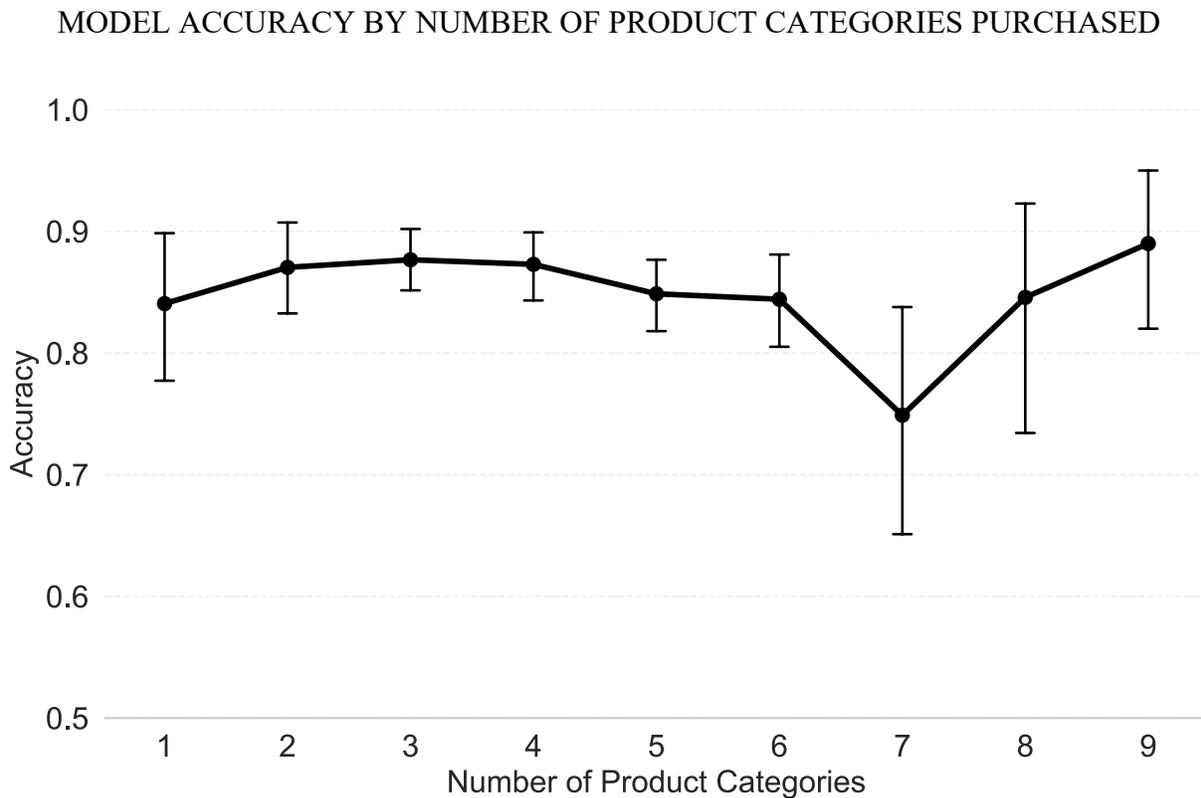
**MODEL ACCURACY BY NUMBER OF REVIEWS**



patterns. Yet, while performance improves steadily as the number of reviews increases, it reaches a plateau around 300 reviews. The accuracy improvement with increasing review volume saturates between 400 and 900 reviews, beyond which additional reviews induce noticeably diminishing returns.

*Effect of Product Category Breadth on Model Accuracy.* We also examine whether the diversity of a consumer’s product purchases affects the predictive accuracy of their digital twin. Specifically, we calculate the number of unique product categories reviewed by each consumer and correlate it with model accuracy. As shown in figure 6, accuracy is relatively stable across consumers with different category breadths. While there is a slight downward trend as the number of product categories increases, the Pearson correlation between category count and

**FIGURE 6**



accuracy is weak and statistically non-significant ( $r = -0.069$ ,  $p = 0.229$ ). These results suggest that the digital twin model performs robustly across consumers with varying content diversity. The stability of predictive performance, even for consumers with limited or highly specialized review patterns, indicates the generalizability of the digital twin architecture across diverse consumer types.

### Tradeoff Between Computational Cost and Performance Gain

While the digital twin framework demonstrates strong predictive accuracy, in practice, it is important to consider the computational resources required to achieve these performance gains. To examine this tradeoff, we evaluate model performance across different levels of training duration, measured in the number of fine-tuning epochs.

Table 2 reports average model accuracy, precision, recall, F1-score, and AUC-ROC, along with the associated computational time and cost, for the digital twins fine-tuned over 3, 5, and 10 training epochs. Accuracy rises from 0.683 at 3 epochs to 0.703 at 5 epochs, and substantially to 0.859 at 10 epochs. Recall shows similar improvement, increasing from 0.419 to 0.745, indicating enhanced sensitivity to identifying likely purchases.

These improvements, nonetheless, come with increased computational cost. In our setup using a cloud-based NVIDIA GeForce RTX 4090D (with an hourly rate of 0.27 US dollars), the average training time per consumer twin increases from approximately 8 minutes at 3 epochs to 28 minutes at 10 epochs. Correspondingly, cost per digital twin increases from \$0.036 to \$0.128. GPU memory utilization and energy consumption scale proportionally with training duration.

**TABLE 2**

COMPARISON ACROSS DIGITAL TWINS OF DIFFERENT TRAINING EPOCHS

| Model Performance             | 3 Epochs | 5 Epochs | 10 Epochs |
|-------------------------------|----------|----------|-----------|
| Accuracy                      | 0.683    | 0.703    | 0.859     |
| Precision                     | 0.730    | 0.776    | 0.960     |
| Recall                        | 0.419    | 0.424    | 0.745     |
| F1                            | 0.457    | 0.476    | 0.807     |
| AUC-ROC                       | 0.683    | 0.702    | 0.859     |
| Computational Resources       |          |          |           |
| Time Per Consumer (mins/secs) | 7'56"    | 14'33"   | 28'28"    |
| Cost Per Consumer (USD)       | 0.036    | 0.066    | 0.128     |

## GENERAL DISCUSSION

### Implications for Research

This research advances the current conceptual and methodological development of consumer digital twins, as well as the broader literature on consumer behavior modelling. It merges insights from behavioral research, psychological theory, and machine learning (especially LLMs) to build synthetic agent that truthfully simulate consumer behavior.

First, our study highlights the critical role of psychological theory in guiding the development of LLMs for behavioral modeling. As machine learning methods become increasingly prevalent in marketing and consumer research, there is a growing need to design and evaluate model architectures not only based on predictive performance but also through the lens of psychological alignment and behavioral plausibility. Using psychological theories to construct and validate machine learning models becomes increasingly essential (Feng, Li and

Zhang 2025). By grounding model design in foundational theories, such as the person-environment interaction framework, we demonstrate how psychological insight can inform the construction of LLM-based systems that better simulate how consumers think, feel, and decide, encouraging future research to construct models that align with consumer psychology.

Second, our work extends the growing literature on LLMs and GenAI consumer research (e.g., De Freitas et al. 2025; Epp and Humphreys 2025). While prior studies have explored prompt engineering, retrieval augmentation, and fine-tuning as isolated, comparative adaptation strategies (e.g., Arora et al. 2025; Goli and Singh 2024; Ye et al. 2025), we demonstrate their complementary strengths through an integrated application. Fine-tuning allows digital twins to internalize individual-level heterogeneity, while retrieval-augmented generation enables dynamic context adaptation. Our approach addresses prior limitations around homogeneity and lack of realism in synthetic consumer responses by enabling digital twins to exhibit diverse psychological and behavioral patterns that reflect real-world variability.

Moreover, our work joins and expands the dialogue on synthetic data and digital twins (e.g., Argyle et al. 2023; Gui and Toubia 2023) by shifting from single-instance simulations to continuous, agent-based modeling. Whereas prior research primarily focuses on generating synthetic responses to surveys or stimuli, digital twins act as persistent, interactive simulacra of individuals that can be probed in longitudinal and contextualized experiments. We demonstrate the feasibility of conducting synthetic consumer research at scale. By simulating both behavioral outcomes and expressive judgments (e.g., product reviews), digital twins offer a viable alternative to costly experiments and surveys. This allows researchers to test behavioral hypotheses, conduct pretests, or simulate choice architectures using AI-generated agents. In doing so, our work contributes to a broader movement toward generative consumer research and

demonstrates how AI can augment, not replace, the theory-driven modeling of consumer decision-making.

Finally, this research also contributes to the body of work leveraging user-generated content (UGC) to understand consumer behavior. Previous consumer research has long treated UGC as a valuable window into consumers' psychology and behavior. Researchers have used linguistic cues in reviews, social media posts, and other forms of digital expression to infer consumer sentiment, personality traits, identity signaling, and decision-making processes. Our work builds on this foundation by demonstrating that UGC is not only a useful diagnostic lens but also a generative substrate for modeling individual consumer behavior. By fine-tuning LLMs on the review history of individual users, we show that UGC can be repurposed to construct personalized digital twins capable of predicting future choices and expressions. This approach advances the utility of UGC in consumer research: rather than merely interpreting what consumers have said or done, we simulate how they are likely to act in new contexts. In doing so, our study shifts the role of UGC from static observational data to dynamic modeling input, enabling richer, more individualized representations of consumer cognition, preference formation, and behavioral tendencies.

### Implications for Practice

The development and validation of a functioning, scalable prototype for consumer digital twins carry substantial implications for marketing practice. This research moves beyond conceptual speculation and offers a working framework for firms to simulate individual consumer behavior with high fidelity. By fine-tuning LLMs on consumer-specific data and

equipping them with contextual knowledge through RAG, organizations can now construct personalized, dynamic agents that reflect how real consumers think, feel, and act. These digital twins are not just theoretical constructs, they can be operationalized at scale using off-the-shelf models and consumer data already available to most firms, enabling widespread adoption without significant infrastructural costs.

First, digital twins can significantly enhance the development and execution of marketing decisions. Because they simulate the behavior of real consumers, they can serve as synthetic testbeds for various forms of consumer research. For example, during new product development, marketers can expose digital twins to product descriptions, branding concepts, or price points to assess likely reactions and refine offerings before investing in full-scale launches. Similarly, email campaigns, ad copy, or promotional content generated via GenAI can be iteratively tested on these digital twins to evaluate emotional tone, relevance, or likelihood of purchase. This fusion of synthetic consumers and synthetic content enables a new form of marketing automation, where entire campaign pipelines can be optimized in silico before reaching real consumers. The feedback loop between generative content and simulated consumer response enables consumer targeting and personalization that traditional A/B testing or focus groups cannot match. Operationally, our findings on the trade-off between computational cost and performance offer guidelines for deployment. For instance, marketers with limited resources may opt for lower-epoch fine-tuning or focus on high-value consumers with rich behavioral history. Moreover, the robustness of the digital twin performance across varying category breadths suggests that these models are applicable to both niche and broad-market consumers.

Also, consumer digital twins help address a critical organizational challenge: fragmented consumer data. In many companies, consumer data exists in silos (Fogarty 2022; Vinogradov

2025), dispersed across departments (e.g., sales, marketing, and customer service), platforms (e.g., web analytics, marketing automation system, and CRM), and channels. Digital twins offer a way to consolidate these scattered signals into a coherent, behaviorally grounded representation of each consumer. Rather than relying on a CRM record here, a web analytics log there, and a survey score elsewhere, marketers can use the digital twin as a single source of behavioral truth. This allows for the creation of a 360-degree view of each customer, not just in terms of demographics or purchase history, but in terms of simulated behavior and future potential. It represents a shift from descriptive to generative consumer profiles, unlocking powerful applications in segmentation, targeting, and lifetime value modeling.

Importantly, consumer digital twins offer a privacy-preserving alternative to traditional targeting and personalization approaches, many of which rely heavily on cookie-based tracking and third-party data. As regulatory frameworks such as the GDPR, CCPA, and the phasing out of third-party cookies by major browsers and mobile apps reshape the digital advertising landscape, firms face growing constraints in how they collect, store, and use consumer data (Ahuja et al. 2022; Wieringa et al. 2021). Our framework aligns with these shifts by enabling personalization without intrusive data collection. Digital twins can be built exclusively from consumers' own first-party data and publicly available UGC. Rather than fighting the tide of increasing privacy restrictions, firms can use digital twins to future-proof their personalization strategies, delivering personalized consumer experiences while preserving consumer privacy and meeting legal obligations.

## Limitations and Future Research Directions

Despite its contributions, this study has several limitations that open avenues for future research. First, our digital twins are trained exclusively on text-based UGC, specifically consumer reviews. While text reviews are rich in cognitive and affective information, they may not capture the full behavioral spectrum of real-world consumers. Many aspects of consumer decision-making, such as browsing paths, visual attention, cart abandonment, and non-verbal expressions, remain outside the scope of text. Future research could enhance behavioral realism by incorporating multimodal data sources, such as clickstream logs, visual content (e.g., product photos, video testimonials), or audio interactions (e.g., voice-based customer service exchanges). With the advancement in multimodal large language models (MLLMs), deep learning algorithms that can understand and generate various forms of content ranging across text, images, video, audio, and more (Yin et al. 2023; Zhang et al. 2024), research can explore how these modalities could improve the dimensions of behavioral modeling and enable more accurate simulations of mental and behavioral processes.

Second, while behavioral and transactional data of consumers included provide insights into consumer preferences and characteristics, the demographic and psychographic data available for model training are limited. Our method of inferring gender from usernames using a zero-shot classifier is a pragmatic but imperfect solution that may introduce misclassification and reflect underlying biases in model training. Future work should try to acquire sources for socio-demographic and psychographic data, such as CRM databases, metadata, or self-reporting surveys. A broader demographic spectrum (e.g., age, region, education, income) would allow researchers to investigate how these attributes influence the fidelity and fairness of digital twins. Additionally, psychometric enrichment using standardized personality assessments, motivations,

or cognitive style inventories could enable richer representations of consumer mindsets, especially for modeling non-normative behavior.

Third, our empirical validation uses Amazon e-commerce data, which is accessible and parsimonious for the purpose of proof of concept. The generalizability of our framework to other consumer contexts, such as customer services, subscription-based platforms, or B2B decision-making, remains to be tested. Each of these domains introduces unique challenges, including longer decision cycles, group-based choices, and intangible evaluation criteria. Future research should expand the digital twin architecture to such contexts. Moreover, while our models accurately predict purchase decisions and replicate actual reviews, they do not account for why consumers chose not to purchase an item. Although digital twins can be queried to probe hypothetical reasoning paths, the absence of ground truth data limits validation. Future work could investigate more structured inference protocols or combine twin-generated responses with passive tracking or survey follow-ups to validate whether digital twin is equally accurate in judgment-forming and decision-making processes.

Technically, our modeling choices were guided by parsimony and replicability. We fine-tuned open-source compact LLMs and implemented a basic RAG setup to illustrate a broadly applicable and computationally feasible approach. However, this means we did not benchmark against proprietary or more powerful models (e.g., GPT-4, Claude, Gemini) that might offer higher performance in certain tasks. Nor did we experiment with more advanced RAG implementations, such as memory-augmented retrievers, graph-based indexing, or multi-hop reasoning (Gao et al. 2023). While our results show that even relatively simple configurations can yield high accuracy, future research could compare architectures to determine when best-in-class performance is needed.

## CONCLUSION

This study presents a methodological framework for constructing digital twins of consumers using large language models adapted with fine-tuning and RAG. By combining consumer-specific data for personification and product-relevant information for contextualization, these digital twins can simulate consumer reactions to various marketing stimuli at the individual level. Empirical results on e-commerce data show strong predictive performance in both purchase decision and product reviews. Additionally, we evaluate the tradeoff between model performance and computational cost, offering practical guidance for scalable implementation. Beyond predictive accuracy, the approach introduces an interactive approach to consumer modeling, opening new possibilities for pre-deployment testing and personalization. While further work is needed to assess generalizability across different marketing contexts and integrate more diverse sources of consumer data, our findings contribute to the growing interest in generative AI for marketing and highlight the potential of LLM-based digital twins as a tool for consumer understanding and marketing decision-making.

## APPENDIX A

### ALGORITHM FOR BUILDING AND TESTING A DIGITAL TWIN MODEL

The algorithm of building and testing one LLM-powered digital twin is illustrated below.

---

**Algorithm:** An LLM-Powered Digital Twin to Make Purchase Decisions and Write Product Reviews

---

**Input:** Training set  $\mathcal{Z} = \{(x_i, y_i)\}_{i=1}^{N-50}$ , testing set  $\mathcal{T} = \{(x_k, y_k)\}_{k=1}^{100}$ , pretrained model  $P_{\Phi_0}$ , LoRA adapter configuration, product corpus  $\mathcal{P}$ , learning rate  $\eta$ , epochs  $E$ , batch size  $B$

**Output:** Digital twin model  $P_{\Phi}$  and prediction  $\hat{y}$

```
1 Fine-Tuning with LoRA:
2 for  $e = 1$  to  $E$  do
3   for  $b = 1$  to  $\lfloor \frac{N}{B} \rfloor$  do
4     Sample mini-batch  $\mathcal{B}_b = \{(x_i, y_i)\}_{i=1}^B$ ;
5     for each  $(x_i, y_i) \in \mathcal{B}_b$  do
6       Predict  $\hat{y}_i = P_{\Phi}(y_i|x_i, y_{i,<t})$ ;
7       Compute batch loss  $\mathcal{L}_b = \frac{1}{B} \sum_{i=1}^B \mathcal{L}(y_i, \hat{y}_i)$ ;
8       Update LoRA parameters  $\Delta\Phi(\theta)$  using gradient descent:
9        $\Delta\Phi(\theta) \leftarrow \Delta\Phi(\theta) - \eta \nabla_{\Delta\Phi(\theta)} \mathcal{L}_b$ ;
10    Record epoch-level loss:  $\mathcal{L}^{(e)} = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(y_i, \hat{y}_i)$ ;
11 Inference with Retrieval Augmented Generation:
12 for each  $(x_k, y_k) \in \mathcal{T}$  do
13   Use product name + ASIN to retrieve  $c = \text{RAG}(\text{product}, \mathcal{P})$ ;
14   Construct prompt  $x = \text{Concat}(x_k, c)$ ;
15   Predict  $\hat{y}_k = P_{\Phi}(y_k|x_k, c, y_{k,<t})$ ;
16 return  $P_{\Phi}$  and  $\hat{y}_k$ 
```

---

## REFERENCES

- Ahuja, Kabir, Thomas Bauer, Caroline Meder, and Oliver Gediehn (2022), “As the Cookie Crumbles, Three Strategies for Advertisers to Thrive,” *Mckinsey.com*, <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/as-the-cookie-crumbles-three-strategies-for-advertisers-to-thrive>.
- Angelopoulos, Panagiotis, Kevin Lee, and Sanjog Misra (2024), “Causal Alignment: Augmenting Language Models with A/B Tests,” *SSRN*, <http://dx.doi.org/10.2139/ssrn.4781850>.
- Argyle, Lisa P., Ethan C. Busby, Nancy Fulda, Joshua R. Gubler, Christopher Rytting, and David Wingate (2023), “Out of One, Many: Using Language Models to Simulate Human Samples,” *Political analysis: an annual publication of the Methodology Section of the American Political Science Association*, 31 (3), 337–51, <http://dx.doi.org/10.1017/pan.2023.2>.
- Arora, Neeraj, Ishita Chakraborty, and Yohei Nishimura (2025), “AI-Human Hybrids for Marketing Research: Leveraging LLMs as Collaborators,” *Journal of Marketing*, forthcoming, <http://dx.doi.org/10.1177/00222429241276529>.
- Ascarza, Eva (2018), “Retention Futility: Targeting High-Risk Customers Might Be Ineffective,” *Journal of Marketing Research, Journal of marketing research*, 55 (1), 80–98, <http://dx.doi.org/10.1509/jmr.16.0163>.
- Atari, Mohammad, Mona J. Xue, Peter S. Park, Damián E. Blasi, and Joseph Henrich. (2024), “Which Humans?,” *PsyArXiv*, preprint, June 20, <https://doi.org/10.31234/osf.io/5b26t>.
- Bettman, James R., Mary Frances Luce, and John W. Payne (1998), “Constructive Consumer

Choice Processes,” *Journal of Consumer Research*, 25 (3), 187–217,  
<http://dx.doi.org/10.1086/209535>.

Bisbee, James, Joshua D. Clinton, Cassy Dorff, Brenton Kenkel, and Jennifer M. Larson (2024),  
“Synthetic Replacements for Human Survey Data? The Perils of Large Language Models,”  
*Political analysis: an annual publication of the Methodology Section of the American  
Political Science Association*, 32 (4), 1–16, <https://doi.org/10.1017/pan.2024.5>.

Borden, Kimberly and Anna Herlt (2022), “Digital Twins: What Could They Do for Your  
Business?,” *McKinsey*, [https://www.mckinsey.com/capabilities/operations/our-  
insights/digital-twins-what-could-they-do-for-your-business](https://www.mckinsey.com/capabilities/operations/our-insights/digital-twins-what-could-they-do-for-your-business).

Brand, James, Ayelet Israeli, and Donald Ngwe (2023), “Using LLMs for Market Research,”  
Working Paper No. 23-062, Harvard Business School, Boston, MA 02163,  
<http://dx.doi.org/10.2139/ssrn.4395751>.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla  
Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini  
Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya  
Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen,  
Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner,  
Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei (2020), “Language  
Models Are Few-Shot Learners,” *arXiv [cs.CL]*,  
<https://doi.org/10.48550/arXiv.2005.14165>.

Büschken, Joachim and Greg M. Allenby (2016), “Sentence-Based Text Analysis for Customer  
Reviews,” *Marketing Science*, 35 (6), 953–75, <http://dx.doi.org/10.1287/mksc.2016.0993>.

Castelo, Noah, Zsolt Katona, Peiyao Li, and Miklos Sarvary (2024), “How AI Outperforms

Humans at Creative Idea Generation,” *SSRN*, <http://dx.doi.org/10.2139/ssrn.4751779>.

Castro, Antonio, Guilherme Cruz, Victoria Bough, and Sanchit Tiwari (2024), “Enhancing the Customer Journey with Gen AI–Powered Digital Twins,” *Mckinsey*, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/enhancing-the-customer-journey-with-gen-ai-powered-digital-twins>.

Chen, Jianlv, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu (2024), “BGE M3-Embedding: Multi-Lingual, Multi-Functionality, Multi-Granularity Text Embeddings Through Self-Knowledge Distillation,” *arXiv* [cs.CL], <https://doi.org/10.48550/arXiv.2402.03216>.

De Freitas, Julian, Gideon Nave, and Stefano Puntoni (2025), “Ideation with Generative AI—in Consumer Research and Beyond,” *Journal of Consumer Research*, 52 (1), 18–31, <http://dx.doi.org/10.1093/jcr/ucaf012>.

Dettmers, Tim, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer (2023), “QLoRA: Efficient Finetuning of Quantized LLMs,” *arXiv* [cs.CL], <https://doi.org/10.48550/arXiv.2305.14314>.

Duerst, Michelle, Christine Colborne, and Alfonso Velosa (2022), “A Digital Twin of a Customer Predicts the Best Consumer Experience,” *Gartner Business Quarterly*.

Epp, Amber M. and Ashlee Humphreys (2025), “Collaborating with Generative AI in Consumer Culture Research,” *Journal of Consumer Research*, 52 (1), 32–48, <http://dx.doi.org/10.1093/jcr/ucaf014>.

Feng, Xiaohang (flora), Charis X. Li, and Shunyuan Zhang (2025), “Visual Uniqueness in Peer-to-Peer Marketplaces: Machine Learning Model Development, Validation, and Application,” *Journal of Consumer Research*, <http://dx.doi.org/10.1093/jcr/ucaf021>.

- Feuerriegel, Stefan, Jochen Hartmann, Christian Janiesch, and Patrick Zschech (2024), “Generative AI,” *Business & Information Systems Engineering*, 66 (1), 111–26, <http://dx.doi.org/10.1007/s12599-023-00834-7>.
- Fogarty, Neal (2022), “What Are Data Silos (and How Do They Impact Marketers)?,” *Adverity.com*, <https://www.adverity.com/blog/what-are-data-silos-and-how-do-they-impact-marketers>.
- Fong, Hortense and George Gui (2024), “Modeling Story Expectations to Understand Engagement: A Generative Framework Using LLMs,” *SSRN*, Working Paper No. 5053346, Columbia Business School, New York, NY 10027, <http://dx.doi.org/10.2139/ssrn.5053346>.
- Fong, Nathan M., Zheng Fang, and Xueming Luo (2015), “Geo-Conquering: Competitive Locational Targeting of Mobile Promotions,” *Journal of Marketing Research*, 52 (5), 726–35, <http://dx.doi.org/10.1509/jmr.14.0229>.
- Fuller, Aidan, Zhong Fan, Charles Day, and Chris Barlow (2020), “Digital Twin: Enabling Technologies, Challenges and Open Research,” *IEEE Access*, 8, 108952–71, <http://dx.doi.org/10.1109/access.2020.2998358>.
- Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang (2023), “Retrieval-Augmented Generation for Large Language Models: A Survey,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.2312.10997>.
- Goli, Ali and Amandeep Singh (2024), “Frontiers: Can Large Language Models Capture Human Preferences?,” *Marketing Science*, 43 (4), 709–22, <http://dx.doi.org/10.1287/mksc.2023.0306>.
- Gui, George and Olivier Toubia (2023), “The Challenge of Using LLMs to Simulate Human Behavior: A Causal Inference Perspective,” *SSRN*, Working Paper No. 4650172, Columbia

Business School, New York, NY 10027, <http://dx.doi.org/10.2139/ssrn.4650172>.

Hartmann, Jochen, Yannick Exner, and Samuel Domdey (2025), “The Power of Generative

Marketing: Can Generative AI Create Superhuman Visual Marketing Content?,”

*International Journal of Research in Marketing*, 42 (1), 13–31,

<http://dx.doi.org/10.1016/j.ijresmar.2024.09.002>.

Hastie, Trevor, Robert Tibshirani, and J. H. Friedman (2009), *The Elements of Statistical*

*Learning: Data Mining, Inference, and Prediction*, 2nd ed., New York, NY: Springer,

<http://dx.doi.org/10.1007/b94608>.

Heitmann, Mark, Tijmen P. J. Jansen, Martin Reisenbichler, and David A. Schweidel (2025),

“Picture Perfect: Engaging Customers with Visual Generative AI,” *Journal of Marketing*,

forthcoming, <http://dx.doi.org/10.1177/00222429251356993>.

Henley, Melissa (2024), “Digital Twins: Can They Transform Customer Experience?,”

*Cmswire.com*, <https://www.cmswire.com/customer-experience/digital-twins-can-they-transform-customer-experience/>.

Horton, J. J. (2023), “Large Language Models as Simulated Economic Agents: What Can We

Learn from Homo Silicus?,” *SSRN*, Working Paper No. w31122, National Bureau of

Economic Research, Cambridge, MA 02138, <https://doi.org/10.2139/ssrn.4413859>.

Hou, Yupeng, Jiacheng Li, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley (2024),

“Bridging Language and Items for Retrieval and Recommendation,” *arXiv [cs.CL]*,

<https://doi.org/10.48550/arXiv.2403.03952>.

Howard, Jeremy and Sebastian Ruder (2018), “Universal Language Model Fine-Tuning for Text

Classification,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.1801.06146>.

Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu

- Wang, and Weizhu Chen (2021), “LoRA: Low-Rank Adaptation of Large Language Models,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.2106.09685>.
- Huang, Ming-Hui and Roland T. Rust (2025), “The GenAI Future of Consumer Research,” *Journal of Consumer Research*, 52 (1), 4–17, <http://dx.doi.org/10.1093/jcr/ucaf013>.
- Kapoor, Anuj and Madhav Kumar (2025), “Frontiers: Generative AI and Personalized Video Advertisements,” *Marketing Science*, 44 (4), 733–47, <http://dx.doi.org/10.1287/mksc.2023.0494>.
- Korst, Jeremy, Stefano Puntoni, and Olivier Toubia (2025), “How Gen AI Is Transforming Market Research,” *Harvard Business Review*, <https://hbr.org/2025/05/how-gen-ai-is-transforming-market-research>.
- Krugmann, Jan Ole and Jochen Hartmann (2024), “Sentiment Analysis in the Age of Generative AI,” *Customer Needs and Solutions*, 11 (1), <http://dx.doi.org/10.1007/s40547-024-00143-4>.
- Kshetri, Nir, Yogesh K. Dwivedi, Thomas H. Davenport, and Niki Panteli (2024), “Generative Artificial Intelligence in Marketing: Applications, Opportunities, Challenges, and Research Agenda,” *International Journal of Information Management*, 75 (102716), <http://dx.doi.org/10.1016/j.ijinfomgt.2023.102716>.
- Kumar, V., Philip Kotler, Shaphali Gupta, and Bharath Rajan (2025), “Generative AI in Marketing: Promises, Perils, and Public Policy Implications,” *Journal of Public Policy & Marketing*, 44 (3), 309–31, <http://dx.doi.org/10.1177/07439156241286499>.
- Le Mens, Gaël and Aina Gallego (2025), “Positioning Political Texts with Large Language Models by Asking and Averaging,” *Political Analysis*, 1–9, <https://doi.org/10.1017/pan.2024.29>.
- Levy, Samuel (2023), “Digital Twins: A Generative Approach for Counterfactual Customer

Analytics,” Carnegie Mellon University, Pittsburgh, PA 15213.

Lewin, Kurt (1936), *Principles of Topological Psychology*, New York, NY: McGraw-Hill.

Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer

Levy, Ves Stoyanov, and Luke Zettlemoyer (2019), “BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.1910.13461>.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-Tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela (2020), “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” *arXiv [cs.CL]*, <https://arxiv.org/abs/2005.11401>.

Li, Peiyao, Noah Castelo, Zsolt Katona, and Miklos Sarvary (2024), “Frontiers: Determining the Validity of Large Language Models for Automated Perceptual Analysis,” *Marketing Science*, 43 (2), 254–66, <http://dx.doi.org/10.1287/mksc.2023.0454>.

Liu, Ying, Lin Zhang, Yuan Yang, Longfei Zhou, Lei Ren, Fei Wang, Rong Liu, Zhibo Pang, and M. Jamal Deen (2019), “A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin,” *IEEE Access*, 7, 49088–101, <https://doi.org/10.1109/access.2019.2909828>.

Mckinsey (2022), “Digital Twins: What Could They Do for Your Business?,” *Mckinsey*, <https://www.mckinsey.com/capabilities/operations/our-insights/digital-twins-what-could-they-do-for-your-business>.

Mischel, Walter and Yuichi Shoda (1995), “A Cognitive-Affective System Theory of Personality: Reconceptualizing Situations, Dispositions, Dynamics, and Invariance in Personality Structure,” *Psychological Review*, 102 (2), 246–68,

<http://dx.doi.org/10.1037/0033-295x.102.2.246>.

Mohammadi, Neda and John E. Taylor (2017), “Smart City Digital Twins,” in *2017 IEEE*

*Symposium Series on Computational Intelligence (SSCI)*, IEEE, 1–5,

<https://doi.org/10.1109/ssci.2017.8285439>.

Muschkiet, M., J. W. Paschmann, and A. Nissen (2022), “Towards Human Digital Twins for

Improving Customer Experience,” in *International Conference on Information Systems*

*(ICIS)*, Copenhagen, Denmark.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), “Mine Your Own

Business: Market-Structure Surveillance through Text Mining,” *Marketing Science*, 31 (3),

521–43, <http://dx.doi.org/10.1287/mksc.1120.0713>.

Netzer, Oded, James M. Lattin, and V. Srinivasan (2008), “A Hidden Markov Model of

Customer Relationship Dynamics,” *Marketing Science*, 27 (2), 185–204,

<http://dx.doi.org/10.1287/mksc.1070.0294>.

Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019), “When Words Sweat: Identifying

Signals for Loan Default in the Text of Loan Applications,” *Journal of Marketing*

*Research*, 56 (6), 960–80, <http://dx.doi.org/10.1177/0022243719852959>.

OpenAI “Supervised Fine-Tuning,” *OpenAI Platform*,

<https://platform.openai.com/docs/guides/supervised-fine-tuning>.

Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin,

Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob

Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul

Christiano, Jan Leike, and Ryan Lowe (2022), “Training Language Models to Follow

Instructions with Human Feedback,” *arXiv [cs.CL]*,

<https://doi.org/10.48550/arXiv.2203.02155>.

Park, Joon Sung, Carolyn Q. Zou, Aaron Shaw, Benjamin Mako Hill, Carrie Cai, Meredith Ringel Morris, Robb Willer, Percy Liang, and Michael S. Bernstein (2024), “Generative Agent Simulations of 1,000 People,” *arXiv [cs.AI]*,

<https://doi.org/10.48550/arXiv.2411.10109>.

Payne, John W., James R. Bettman, and Eric J. Johnson (1993), *The Adaptive Decision Maker*, Cambridge, England: Cambridge University Press.

Qi, Qinglin and Fei Tao (2018), “Digital Twin and Big Data towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison,” *IEEE Access*, 6, 3585–93,

<http://dx.doi.org/10.1109/access.2018.2793265>.

Qiu, Liying, Param Vir Singh, and Kannan Srinivasan (2023), “Consumer Risk Preferences Elicitation from Large Language Models,” *SSRN*, <http://dx.doi.org/10.2139/ssrn.4526072>.

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever (2019), “Language Models Are Unsupervised Multitask Learners,” *OpenAI Blog*, 1 (8).

Reisenbichler, Martin, Thomas Reutterer, David A. Schweidel, and Daniel Dan (2022),

“Frontiers: Supporting Content Marketing with Natural Language Generation,” *Marketing Science*, 41 (3), 441–52, <http://dx.doi.org/10.1287/mksc.2022.1354>.

Reynolds, Laria and Kyle McDonell (2021), “Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm,” *arXiv [cs.CL]*, <http://arxiv.org/abs/2102.07350>.

Ringel, Daniel (2023), “Creating Synthetic Experts with Generative Artificial Intelligence,” *SSRN*, <https://doi.org/10.2139/ssrn.4542949>.

Serapio-García, Greg, Mustafa Safdari, Clément Crepy, Luning Sun, Stephen Fitz, Peter

Romero, Marwa Abdulhai, Aleksandra Faust, and Maja Matarić (2023), “Personality Traits

- in Large Language Models,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.2307.00184>.
- Simon, H. A. (1990), “Invariants of Human Behavior,” *Annual Review of Psychology*, 41 (1), 1–19, <http://dx.doi.org/10.1146/annurev.ps.41.020190.000245>.
- Sperling, Jonathan (2025), “AI-Generated Digital Twins: Shaping the Future of Business,” *Columbia AI Summit*, <https://business.columbia.edu/insights/digital-future/ai-agent-digital-twin>.
- Timoshenko, Artem and John R. Hauser (2019), “Identifying Customer Needs from User-Generated Content,” *Marketing Science*, 38 (1), 1–20, <http://dx.doi.org/10.1287/mksc.2018.1123>.
- Tirunillai, Seshadri and Gerard J. Tellis (2014), “Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation,” *Journal of Marketing Research*, 51 (4), 463–79, <http://dx.doi.org/10.1509/jmr.12.0106>.
- Tojin, Daniel J., Josh Finkenstadt, and Lokesh Folk (2023), “How Generative AI Can Augment Human Creativity,” *Harvard Business Review*, 101 (4), 56–64.
- Toubia, Olivier, George Z. Gui, Tianyi Peng, Daniel J. Merlau, Ang Li, and Haozhe Chen (2025), “Twin-2K-500: A Dataset for Building Digital Twins of over 2,000 People Based on Their Answers to over 500 Questions,” *arXiv [cs.CY]*, <https://doi.org/10.48550/arXiv.2505.17479>.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin (2017), “Attention Is All You Need,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.1706.03762>.
- Vinogradov, Roman (2025), “Data Silos: Why They’re a Problem & How to Fix 2025,” *Improvado.io*, <https://improvado.io/blog/data-silos>.

- Wang, Xin (Shane), Jiaxiu He, David J. Curry, and Jun Hyun (Joseph) Ryoo (2022), “Attribute Embedding: Learning Hierarchical Representations of Product Attributes from Consumer Reviews,” *Journal of Marketing*, 86 (6), 155–75, <http://dx.doi.org/10.1177/00222429211047822>.
- Wieringa, Jaap, P. K. Kannan, Xiao Ma, Thomas Reutterer, Hans Risselada, and Bernd Skiera (2021), “Data Analytics in a Privacy-Concerned World,” *Journal of Business Research*, 122, 915–25.
- Ye, Zikun, Hema Yoganarasimhan, and Yufeng Zheng (2025), “LOLA: LLM-Assisted Online Learning Algorithm for Content Experiments,” *Marketing Science*, 0 (0), <https://pubsonline.informs.org/doi/abs/10.1287/mksc.2024.0990>.
- Yin, Shukang, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen (2023), “A Survey on Multimodal Large Language Models,” *arXiv [cs.CV]*, <https://doi.org/10.48550/arXiv.2306.13549>.
- Yin, Wenpeng, Jamaal Hay, and Dan Roth (2019), “Benchmarking Zero-Shot Text Classification: Datasets, Evaluation and Entailment Approach,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.1909.00161>.
- Yoganarasimhan, Hema and Irina Iakovetskaia (2024), “From Feeds to Inboxes: A Comparative Study of Polarization in Facebook and Email News Sharing,” *Management Science*, 70 (9), 6461–72, <http://dx.doi.org/10.1287/mnsc.2023.04134>.
- Zhang, Duzhen, Yahan Yu, Jiahua Dong, Chenxing Li, Dan Su, Chenhui Chu, and Dong Yu (2024), “MM-LLMs: Recent Advances in MultiModal Large Language Models,” *arXiv [cs.CL]*, <https://doi.org/10.48550/arXiv.2401.13601>.
- Zhang, Xiaoke, Mi Zhou, and Gene Moo Lee (2024), “Generative AI and Creator Economy:

Investigating the Effects of AI-Generated Voice on Online Video Creation,” *SSRN*,  
<http://dx.doi.org/10.2139/ssrn.4676705>.

Zhou, Eric and Dokyun Lee (2024), “Generative Artificial Intelligence, Human Creativity, and Art,” *PNAS Nexus*, 3 (3), gae052, <http://dx.doi.org/10.1093/pnasnexus/pgae052>.