Decomposing the Crowd for Targeted Role-Play: The PRISM Framework for Product Optimization

Anonymous ACL submission

Abstract

Product design success in consumer-facing markets depends on understanding user needs, yet traditional market research faces high costs and captures only explicit, surface-level requirements. Large Language Models (LLMs) offer scalable market simulation, but existing approaches lack statistical rigor and suffer from mode collapse. We introduce PRISM (Partitioned Role-play in Intelligent Simulated Markets), a framework that bridges unstructured LLM outputs and rigorous economic modeling by decomposing consumer profiles into orthogonal vectors aligned with microeconomic theory. PRISM establishes a theory-grounded diagnostic protocol evaluating dimensional independence and pricing rationality without large-scale ground truth data. We demonstrate that simulated feedback exhibits sufficient structural validity for iterative product optimization, showing clear improvements over baseline methods.

1 Introduction

In consumer-facing (2C) markets, product design success critically depends on accurately understanding user needs (Kohli and Jaworski, 1990; Langerak et al., 2004). However, traditional market research faces significant limitations: high execution costs create barriers to frequent iteration (Quan et al., 2023a, 2021), and methods primarily capture explicit, surface-level needs rather than underlying motivations (Price and Straker, 2017), providing only macro-level positioning guidance that fails to inform granular design decisions. As illustrated in **Figure 1 (left)**, this creates a fundamental disconnect between the dynamic, iterative nature of product design and the static, coarse-grained understanding of market needs.

The advent of Large Language Models (LLMs) has revolutionized computational social science, offering potential solutions to these challenges. As demonstrated by (Park et al., 2023) and (Horton, 2023), LLMs possess the latent capability to act as simulated economic agents, offering a scalable "in-silico" alternative to costly human subject experiments. Recent work has explored LLMs for sales dialogue agents (Jin et al., 2024; Chang and Chen, 2024; Cheng et al., 2025; Amarak et al., 2024) and market research augmentation (Wang et al., 2025a; Sarstedt et al., 2024), demonstrating their potential for understanding consumer behavior. However, bridging the gap between generative simulation and rigorous economic modeling requires adherence to foundational principles. According to Lancaster's characteristics theory (Lancaster, 1966) and McFadden's random utility models (McFadden, 1974), consumer choice is driven by the specific value assigned to orthogonal product attributes. Therefore, for a simulation to be valid, it must accurately model these structured preferences.

Currently, a critical gap exists in achieving this statistical rigor. While (Argyle et al., 2023) suggest LLMs can emulate diverse samples, standard implementations often suffer from mode collapse, regressing toward a generic "average" persona that fails to capture the distinct preference structures required by economic theory (Wang et al., 2025b). Compounding this is the challenge of validation: without accessible real-time ground truth data for hypothetical products, it is difficult to distinguish genuine economic reasoning from stochastic generation noise. This validation bottleneck prevents simulated feedback from serving as a reliable signal for decision-making.

To address these limitations, we introduce **PRISM** (Partitioned Role-play in Intelligent Simulated Markets). As shown in **Figure 1 (right)**, PRISM aligns with microeconomic theory by mathematically decomposing consumer profiles into orthogonal vectors, ensuring agents maintain distinct, non-overlapping preference structures. Furthermore, we establish a diagnostic protocol that eval-

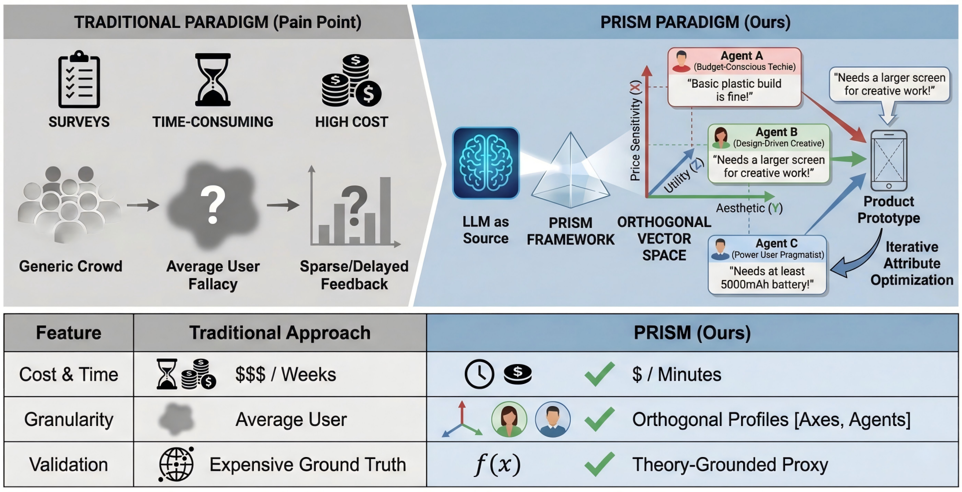

Figure 1: **Paradigm Shift in Market Simulation. Left:** Traditional market research relies on manual surveys that are costly and often result in an "average user fallacy." **Right:** Our proposed **PRISM** framework leverages LLMs to structure market heterogeneity.

uates internal consistency—such as dimensional independence—providing a proxy for fidelity even in the absence of external ground truth.

The main contributions of this work are summarized as follows:

- **PRISM Framework:** We propose a novel framework that structures market heterogeneity by decomposing consumer profiles into orthogonal vectors, bridging the gap between unstructured LLM outputs and rigorous economic modeling.
- **Diagnostic Evaluation:** We establish a theory-grounded proxy evaluation protocol to assess agent internal consistency—specifically dimensional independence and pricing rationality—enabling validation without expensive ground-truth data.
- **Application and Benchmark:** We demonstrate the utility of simulated feedback in a closed-loop product optimization task and release a comprehensive benchmark rooted in real-world e-commerce data to facilitate future research.

2 Related Work

2.1 Traditional Product Design Methodologies

Product design methodologies have evolved from technology-driven to user-centric approaches, with market research playing an increasingly central role (Li et al., 2014; Bangle et al., 2022). Quality Function Deployment (QFD) (Li et al., 2014), design thinking (Rodríguez-Salvador et al., 2016), and integrated customer-driven methodologies (Naik and Srinivasan, 2024; Choi et al., 2019) all emphasize early integration of customer insights. However, two persistent challenges remain unresolved: the high execution costs of traditional market research (Xu et al., 2020; Chadha et al., 2018) and the difficulty in capturing latent needs beyond explicit, surface-level requirements (Wang et al., 2022; Price and Straker, 2017). While big data and AI-driven methods have attempted to address these limitations (Quan et al., 2023b; Feng et al., 2020), they primarily enhance data collection efficiency rather than fundamentally solving the disconnect between static market understanding and dynamic design iteration.

2.2 Market Simulation and Prediction Methods

Machine learning approaches have explored market prediction through historical data analysis (Gaddam et al., 2024; Wei et al., 2024; Zhao et al., 2025), but these methods suffer from limited interpretability and discrete action spaces, constraining their application to recommendation and advertising rather than complex product design decisions (Tissera et al., 2024). Recent work has demonstrated LLMs' potential for market research augmentation (Wang et al., 2025a; Sarstedt et al., 2024) and sales dialogue agents (Jin et al., 2024; Chang and Chen, 2024; Cheng et al., 2025; Amarak et al., 2024), yet research on using LLMs to simulate market feedback for upstream product design optimization remains largely unexplored.

2.3 Theoretical Foundations for Market Simulation

Role-Playing Capabilities. Extensive research has demonstrated LLMs' ability to simulate diverse human groups, including dialogue, actions, and emotional states (Park et al., 2023; Argyle et al., 2023; Wang et al., 2025b; Marincioni et al., 2024; Shea et al., 2024). This capability reflects LLMs' deep understanding of human behavior acquired through large-scale text pretraining, revealing their potential for analyzing consumer psychology and predicting market responses.

Market Structure and Quantitative Analysis. Established theories provide the foundation for structured market decomposition and quantitative analysis. Reference price theory (Monroe, 1973) explains how consumers form internal price standards for purchase decisions, serving as the basis for modeling individual consumer group decision-making. Lancaster's characteristics theory (Lancaster, 1966) and hedonic price indexes (Court, 1939; Kotler, 2012) establish that consumer value perception is driven by orthogonal product attributes, enabling the structured decomposition of markets into finite consumer groups based on dimensional value preferences. McFadden's random utility models (McFadden, 1974) further provide the mathematical framework for quantitative analysis of consumer feedback in simulated markets. These theoretical foundations enable the decomposition of markets into structured consumer groups and the quantitative analysis of simulated market feedback, bridging the gap between unstructured LLM outputs and rigorous economic modeling.

3 The PRISM Framework

We propose **PRISM** (**P**artitioned **R**ole-play in **I**ntelligent **S**imulated **M**arkets), a closed-loop framework designed to decompose market heterogeneity into structured orthogonal vectors for targeted simulation and optimization. As illustrated in Figure 2, the framework consists of four main stages: initial product and market structuring, intelligent simulated market with consumer role-play, feedback analysis and evaluation, and product optimization through an iterative loop.

3.1 Problem Formulation

We consider the design of a product defined by a set of distinct **attributes**, indexed by $j \in \{1, \dots, N\}$. These attributes represent the concrete specifications of the product (e.g., screen size, battery capacity, device weight).

However, consumer purchase decisions are typically driven by underlying dimensions of value rather than raw specifications alone. For instance, attributes like 'weight' and 'thickness' collectively contribute to the abstract dimension of 'Portability'. Therefore, we map the N attributes to a smaller set of M latent **dimensions**, indexed by $m \in \{1, \dots, M\}$. These dimensions represent the independent axes of user perception. Since dimensions serve as abstract generalizations of attributes, we naturally assume $M \leq N$.

We formally define a mapping function $d : \{1, \dots, N\} \to \{1, \dots, M\}$, where $d(j) = m$ indicates that the j-th attribute logically maps to the m-th dimension.

The market is composed of K distinct **consumer groups**, indexed by $i \in \{1, \dots, K\}$. Each group represents a unique combination of value preferences across the M dimensions. Our goal is to simulate the market response and find the optimal selling price p^* that maximizes the total profit $\pi(p)$:

$$\pi(p) = (p - c) \cdot \sum_{i=1}^{K} (N_i \cdot \mathbb{I}\{RP_i \geq p\}) \quad (1)$$

where c denotes the production cost, N_i is the population size of group i, $\mathbb{I}\{\cdot\}$ is the indicator function, and RP_i is the group's psychological reference price.

3.2 Decomposing the Crowd: Structural Market Modeling

Based on the formulation above, we perform a structural decomposition of the market using the M dimensions. Drawing on Conjoint Measurement theory, we discretize consumer preference on each dimension m into three distinct value stances: *negative* (-1), *neutral* (0), and *positive* (1).

Consequently, each consumer group i is uniquely identified by an orthogonal profile vector $\vec{x}_i = (x_{i,1}, \dots, x_{i,M}) \in \{-1, 0, 1\}^M$. The population size N_i is mathematically reconstructed from the statistical distribution of these dimensions:

$$N_i = N(\vec{x}_i) = N_{total} \cdot \prod_{m=1}^{M} \phi_m(x_{i,m}) \quad (2)$$

where N_{total} is the total market size, and $\phi_m(v)$ represents the proportion of the population holding value stance $v \in \{-1, 0, 1\}$ on dimension m.

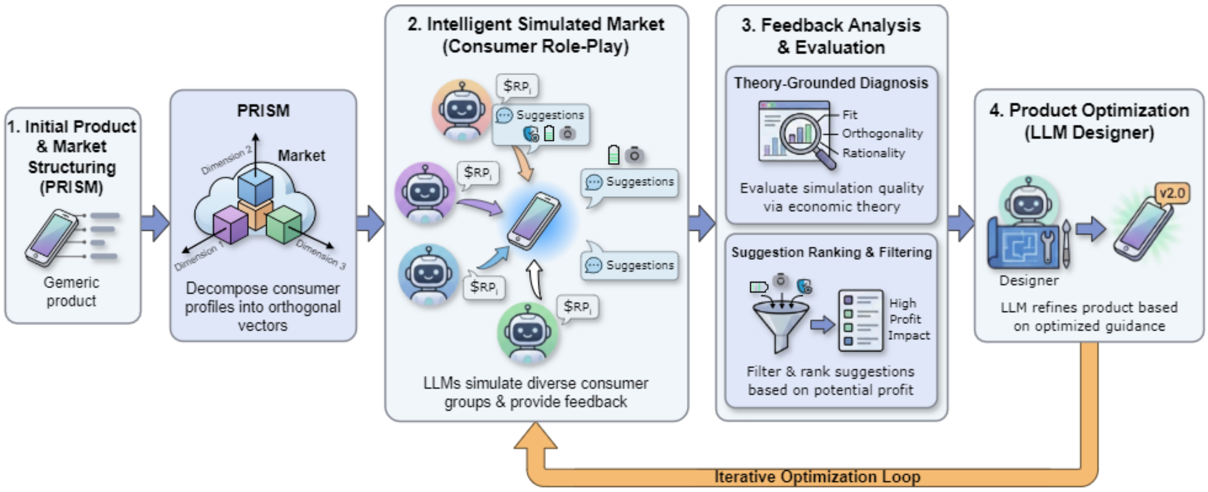

Figure 2: The PRISM framework: A closed-loop system for product design optimization through partitioned role-play in intelligent simulated markets. The framework consists of four main stages and the iterative optimization loop enables continuous improvement.

3.3 Targeted Role-Play and Simulation

With the market decomposed into structured vectors, we employ a Large Language Model (LLM), denoted as $\mathcal{L}$, to simulate consumer behavior through targeted role-play. To isolate the impact of preferences from varying financial capabilities, we normalize the consumption budget B for all groups.

The LLM is prompted with the specific profile vector $\vec{x}_i$ and budget B to predict the reference price:

$$RP_i = RP(\vec{x}_i) \sim p_{\mathcal{L}}(\cdot \mid \text{Prompt}(\vec{x}_i, B)) \quad (3)$$

The construction of prompts is detailed in Appendix B.

3.4 Theory-Grounded Diagnostic Evaluation

Ideally, the fidelity of a simulated market should be assessed by aligning it with large-scale empirical survey data (the gold standard). However, this approach is often hindered by prohibitively high data acquisition costs and privacy barriers. To address this, we propose a structure-based proxy evaluation framework to assess the quality of the simulated market indirectly but effectively.

Specifically, we decompose simulation fidelity into three critical capabilities of the LLM agent, forming the pillars of our evaluation:

Dimensional Orthogonality and Completeness: Whether the LLM can identify M consumer segmentation dimensions that are statistically independent (orthogonal) and collectively exhaustive, ensuring no significant market heterogeneity is overlooked.

Semantic Fidelity of Attribute Mapping: Whether the definition of the three-level preference features (negative, neutral, positive) on each dimension accurately maps to the latent psychological values of real consumers.

Rationality of Price Estimation: Whether the predicted reference price (RP) for a specific consumer group exhibits economic rationality and distributional alignment with ground-truth consumer behaviors.

To quantify these capabilities, we introduce a theoretical reference price model, $RP_{\text{the}}(\vec{x})$.

The Theoretical Model. Based on the linear additive assumption of Hedonic Pricing, we define the theoretical reference price for a group defined by vector $\vec{x}$ as:

$$RP_{\text{the}}(\vec{x}) = \beta_0 + \sum_{m=1}^{M} \left[\mathbb{I}(x_m = 1)\beta_m^+ - \mathbb{I}(x_m = -1)\beta_m^- \right] \quad (4)$$

Here, β_0 is the base reference price. We introduce asymmetric coefficients for the m-th dimension: β_m^+ represents the premium for positive perception, and β_m^- represents the penalty for negative perception.

Two-Stage Calibration. We estimate the parameter set $\Theta = \{\beta_0, \{\beta_m^+, \beta_m^-\}_{m=1}^M\}$ via a two-stage process:

1. **Initial Fitting**: We first obtain the initial parameters $\Theta^{(0)}$ by solving a constrained least

squares optimization problem to fit the model to the simulated data $\{(\vec{x}_i, RP_i)\}_{i=1}^{K}$:

$$\min_{\Theta^{(0)}} \sum_{i=1}^{K} \left(RP_{\text{the}}^{(0)}(\vec{x}_i) - RP_i \right)^2 \quad (5)$$

2. **Intercept Calibration**: We then calibrate the intercept β_0 such that the theoretical optimal price p^* aligns with the actual market price p_{actual} (see Algorithm 1 in Appendix A).

Diagnostic Metrics. Based on the calibrated parameters, we define four metrics to assess the three capabilities defined above.

1. Goodness-of-Fit (GF). Measures dimensional independence (1st Pillar) via the R^2 score:

$$GF = 1 - \frac{\sum_{i=1}^{K}(RP_i - RP_{\text{the}}^{(0)}(\vec{x}_i))^2}{\sum_{i=1}^{K}(RP_i - \bar{RP})^2} \quad (6)$$

2. Dimension Value (DV). Reflects the economic significance of dimensions (3rd Pillar):

$$DV = \frac{1}{2M} \sum_{m=1}^{M} \frac{|\beta_m^+| + |\beta_m^-|}{p_{\text{actual}}} \quad (7)$$

3. Definition Correctness (DC). Measures semantic fidelity (2nd Pillar) by checking coefficient consistency:

$$DC = \frac{1}{2M} \sum_{m=1}^{M} \left[\mathbb{I}(\beta_m^+ > 0) + \mathbb{I}(\beta_m^- > 0) \right] \quad (8)$$

4. Overall Bias (OB). Quantifies the rationality of base price estimation (3rd Pillar):

$$OB = \frac{|\beta_0 - \beta_0^{(0)}|}{p_{\text{actual}}} \quad (9)$$

While these proxy metrics do not directly measure the end-to-end gap between simulated and real markets, they offer a distinct advantage: **interpretability**. By rigorously evaluating the LLM's performance across the three pillars, this framework provides a granular diagnostic view of the simulation's reliability and robustness, effectively validating the constructed market structure.

3.5 Closed-Loop Product Optimization

The final phase optimizes product design based on simulation feedback. The raw feedback consists of suggestions for the N attributes. Let $\text{Sug}_{i,j}$ denote a suggestion from group i regarding attribute j.

Feedback Filtering. We aggregate suggestions into *optimization directions*. An optimization direction $\mathcal{O}_{j,v}$ collects suggestions for a specific attribute j from groups holding value stance v on the corresponding dimension $d(j)$:

$$\mathcal{O}_{j,v} = \{ \text{Sug}_{i,j} \mid i \in \{1, \ldots, K\},\, x_{i,d(j)} = v \} \quad (10)$$

We retain the top-S suggestions for each direction.

Importance Ranking. To prioritize these directions, we assess the potential profit lift using a greedy strategy. We define the importance score I_i for a suggestion from group i based on two factors. First, the *Impact Probability*:

$$\mathcal{P}_i = \frac{p^{*2}}{(|RP_i - p^*| + \epsilon)^2} \quad (11)$$

Second, the *Magnitude*, defined by the group's total population proportion across all M dimensions:

$$\mathcal{M}_i = \frac{N_i}{N_{\text{total}}} = \prod_{m=1}^{M} \phi_m(x_{i,m}) \quad (12)$$

Consequently, the importance of a specific **optimization direction** $\mathcal{O}_{j,v}$ is the cumulative importance of all groups constituting that direction:

$$\begin{aligned} I_{j,v} &= \sum_{i:\, x_{i,d(j)}=v} (\mathcal{P}_i \cdot \mathcal{M}_i) \\ &= p^{*2} \sum_{i:\, x_{i,d(j)}=v} \frac{\prod_{m=1}^{M} \phi_m(x_{i,m})}{(|RP_i - p^*| + \epsilon)^2} \end{aligned} \quad (13)$$

We first rank optimization directions based on $I_{j,v}$, and then rank individual suggestions within each direction based on I_i.

4 Experiments

This section empirically evaluates the PRISM framework. We first validate the structural fidelity of the simulated markets using the diagnostic metrics defined in Section 3.4. Subsequently, we assess the practical efficacy of the closed-loop optimization by comparing profit outcomes against a zero-shot baseline. Finally, ablation studies verify the contribution of individual framework components.

4.1 Experimental Setup

Datasets and Configuration. Experiments are conducted on a real-world e-commerce dataset sourced from Flipkart (2015-2016). We filter the dataset to retain 140 consumer-facing (2C) products characterized by rich attribute descriptions. To anchor the simulation in economic reality, the consumption budget B is aligned with semi-annual average expenditure data derived from the 2025 Household Income and Expenditure Survey (National Bureau of Statistics), categorizing products into eight major sectors.

Implementation Details. We employ DeepSeek-R1-0528 as the core agent for market decomposition and simulation, utilizing DeepSeek-V3.1 for auxiliary classification tasks. Regarding the optimization hyperparameters defined in Section 3.5, we set the suggestion retention limit $S = 5$ and the smoothing constant $\epsilon = 0.01 \cdot p^*$. The profit maximization objective incorporates two simplifying assumptions to facilitate large-scale automated evaluation: (1) production cost c is normalized to zero, equating profit maximization with revenue maximization; and (2) optimization suggestions are constrained to maintain technical feasibility without degrading material quality.

Baselines. We benchmark PRISM against a **Zero-Shot** baseline. In this setting, the LLM optimizes product designs directly based on raw descriptions and general market knowledge, without access to the structured, granular feedback provided by the PRISM simulation.

4.2 Diagnostic Evaluation of Market Simulation

We first assess whether the LLM-generated markets possess sufficient structural integrity to serve as a proxy for optimization. Table 1 summarizes the diagnostic metrics across the test set.

Table 1: Diagnostic metrics for simulated markets (Avg. over 140 products). The high Definition Correctness (DC) relative to Goodness-of-Fit (GF) indicates that the simulation captures the *directionality* of preferences more accurately than precise numerical price points.

Metric	Average Value
Goodness-of-Fit (GF) (R^2)	0.4156
Dimension Value (DV)	0.2222
Definition Correctness (DC)	0.7851
Overall Bias (OB)	0.8976

Analysis of Fidelity. As indicated in Table 1, the simulation achieves a Definition Correctness (DC) of 0.7851, significantly exceeding random chance.

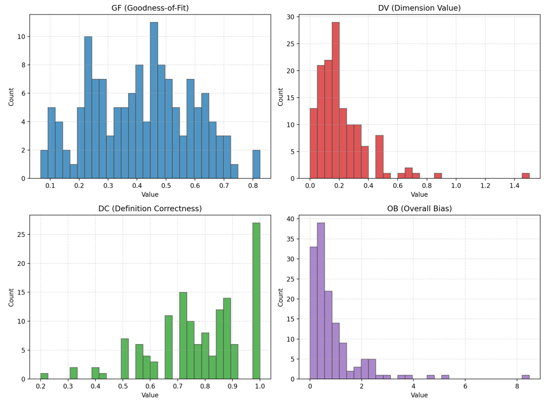

Figure 3: Distribution of diagnostic metrics (GF, DV, DC, OB) across all 140 products. Top-left: Goodness-of-Fit (GF) shows a multi-modal pattern with peaks around 0.2–0.25 and 0.45–0.5. Top-right: Dimension Value (DV) is heavily right-skewed, with most products concentrated in the 0.05–0.20 range. Bottom-left: Definition Correctness (DC) exhibits a strong peak at 1.0, with approximately 28 products achieving perfect definition correctness. Bottom-right: Overall Bias (OB) shows most products with low bias values (0–0.5).

Deeper analysis reveals that β^+ coefficients are positive in 81.09% of cases (489/603), while β^- coefficients are positive in 75.21% of cases (270/359). Overall, the combined positive rate for both β^+ and β^- is 78.90% (759/962), indicating strong semantic fidelity in capturing value directionality. This discrepancy between high semantic fidelity (DC) and moderate numerical fit (GF ≈ 0.42) suggests that while the agent may struggle with precise pricing variance, it successfully captures the fundamental *structure* of consumer preferences.

Furthermore, role-play consistency varies by value stance: positive value groups (β^+) exhibit higher robustness (81.09% positive rate) compared to negative value groups (β^-, 75.21% positive rate), indicating that the agent demonstrates greater stability when modeling positive value impacts than negative ones.

Figure 3 illustrates the distribution of diagnostic metrics across all 140 products. The GF distribution (top-left) shows a multi-modal pattern with peaks around 0.2–0.25 and 0.45–0.5, indicating varying levels of dimensional independence across products. The DV distribution (top-right) is heavily right-skewed, with most products concentrated in the 0.05–0.20 range, suggesting that while dimensions contribute to value perception, their economic significance varies substantially. The DC distribution (bottom-left) exhibits a strong peak at 1.0, with approximately 28 products achieving perfect definition correctness, while the OB distribution (bottom-right) shows most products with low bias values (0–0.5), indicating generally rational base price estimation.

4.3 Closed-Loop Optimization Performance

The ultimate validation of PRISM lies in its ability to guide product improvement. We compare the market performance of products optimized via the PRISM feedback loop against the Zero-Shot baseline.

Table 2: Comparative performance of closed-loop product optimization. PRISM demonstrates superior profit lift and success rates, validating the utility of structured, heterogeneous feedback over generic optimization.

Metric	PRISM	Zero-Shot
Improved Products Ratio (%)	**66.43**	52.14
Avg. Relative Improvement (%)	**21.79**	20.11

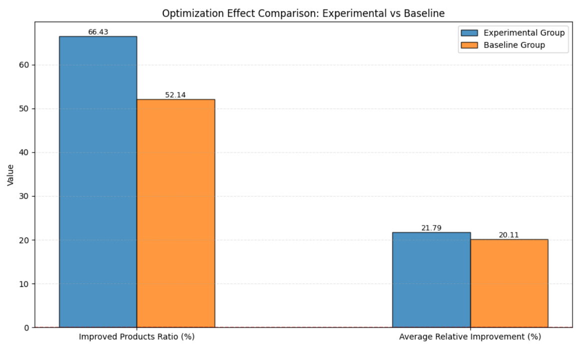

Figure 4: Optimization effect comparison: Experimental vs Baseline. The chart compares both the improved products ratio and average relative improvement between PRISM and Zero-Shot baseline.

Optimization Effectiveness. As shown in Table 2 and Figure 4, PRISM significantly outperforms the baseline, improving 66.43% of the products (93/140) compared to 52.14% (73/140) for the Zero-Shot baseline, representing a 14.29 percentage point improvement. The average relative improvement is 21.79% for PRISM versus 20.11% for the baseline, a difference of 1.68 percentage points. This underscores that aggregating feedback from diverse, structured personas yields more actionable and high-value signals than generic, unstructured optimization.

Complexity Analysis. We observe an inverted-U relationship between the number of adopted suggestions and performance gain: adopting 1–4 suggestions yields stable gains, whereas exceeding 5 leads to a performance drop (-4.37%) due to emerging feature conflicts. Additionally, products decomposed into 10–100 consumer groups exhibit the highest optimization potential, suggesting a "sweet spot" for market granularity where heterogeneity is captured without diluting the optimization signal.

4.4 Ablation Studies

To validate the design rationale of PRISM, we investigate the relationship between different numbers of dimensions and group counts with product optimization performance.

Figure 5 presents the relationship between the number of dimensions, group counts, and optimization performance. A key observation is that group count has minimal impact on the baseline performance across all dimension configurations. However, for the experimental group (PRISM), products with a larger number of dimensions (5) exhibit improved and more stable performance compared to those with fewer dimensions (3 or 4). Specifically, dimension 5 products consistently show higher positive improvement ratios across different group count ranges, with values ranging from 66.7% to 80.0%, while maintaining relatively stable performance. This suggests that a richer dimensional decomposition enables more effective optimization by capturing finer-grained market heterogeneity.

5 Discussion

The experimental results validate PRISM not merely as a simulation tool, but as a robust framework capable of imposing statistical rigor on unstructured LLM outputs. In this section, we interpret how the core components of our proposed framework contribute to the observed performance, linking back to the theoretical foundations established in our contributions.

5.1 Validity of the Theory-Grounded Proxy

A pivotal finding from our diagnostic evaluation is the alignment between the agents' behavior and standard economic theory, specifically regarding dimensional independence and pricing rationality. This suggests that while absolute numerical precision remains a challenge for LLMs, the framework successfully captures the directionality of consumer preferences. For closed-loop optimization, this structural validity is the governing factor. By

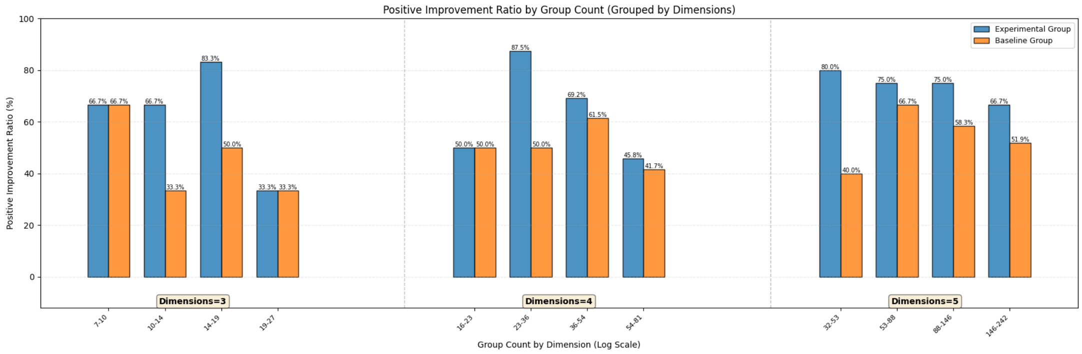

Figure 5: Positive improvement ratio by group count, grouped by dimensions. The chart displays the positive improvement ratio for products with 3, 4, and 5 dimensions across different group count ranges. Dimension 3 products are on the left, dimension 4 in the middle, and dimension 5 on the right. Each dimension is divided into 4 group count bins on a logarithmic scale.

verifying that agents exhibit internal consistency, we establish that the simulation can generate valid optimization gradients—distinguishing value creators from value detractors—even in the absence of large-scale ground truth data.

5.2 Efficacy of Orthogonal Decomposition

The significant performance gap between PRISM and the generic Zero-Shot baseline (14.29 percentage points in improved products ratio) vindicates our strategy of structuring market heterogeneity. By mathematically decomposing consumer profiles into orthogonal vectors, PRISM mitigates the mode collapse often observed in unstructured simulations, where agents default to an "average user" persona. This structural decomposition forces the system to acknowledge distinct, non-overlapping preferences. The ablation study (Figure 5) further validates this approach: products with richer dimensional decomposition (5 dimensions) exhibit more stable and improved optimization performance, while group count variations have minimal impact on baseline performance. This confirms that profit gains are achieved not by uniformly raising prices, but by sharpening the product's fit for specific high-value segments through structured market decomposition, essentially automating the logic of strategic market segmentation.

5.3 Future Work

To further enhance this framework, we envision three key expansions. First, integrating cost estimation models will allow the iterative simulation to optimize for net profit rather than revenue, providing a more holistic economic view. Second, introducing competitive agents will enable the simulation of dynamic market responses, such as pricing wars. Finally, enabling continuous learning from real-time feedback will allow the agent to refine its profiling accuracy over time, effectively closing the loop between simulation and reality.

6 Conclusion

In this paper, we introduced PRISM (Partitioned Role-play in Intelligent Simulated Markets), a framework designed to bridge the gap between unstructured LLM capabilities and rigorous economic modeling.

Our contributions are summarized as follows. First, we proposed a method to structure market heterogeneity by decomposing consumer profiles into orthogonal vectors, thereby introducing statistical stability into agent simulations. Second, we established a theory-grounded diagnostic evaluation framework. This approach allows for the rigorous assessment of an agent's internal consistency—specifically dimensional independence and pricing rationality—serving as a reliable proxy for fidelity without relying on expensive ground truth data. Third, we demonstrated the utility of this system in a validated closed-loop application, showing that simulated feedback provides sufficient structural validity to drive significant iterative product optimization.

PRISM represents a foundational step towards automated, data-driven market research. By providing a scalable environment for hypothesis testing and releasing a comprehensive benchmark rooted in real-world data, we aim to facilitate future research in economically aligned agent simulations.

Limitations

While PRISM establishes a novel paradigm for automated market simulation, our experimental design entails strategic trade-offs. First, to overcome the barriers of acquiring real-time transaction data, we relied on theory-grounded proxy metrics for validation; thus, while internally consistent, the optimized designs should be interpreted as high-confidence hypotheses pending empirical A/B testing. Second, we adopted simplified economic assumptions, such as static category-level budgets, to facilitate large-scale automated evaluation. These simplifications allow us to isolate the impact of consumer value perception—our primary focus—though future iterations would benefit from incorporating complex cost-revenue dynamics. Finally, our findings reflect the capabilities of current LLM architectures; as model reasoning evolves, the framework's parameters may require recalibration to maintain its high structural fidelity.

Acknowledgements

References

Ahmed Amarak, Mohamed Igamane, Tajjeeddine Rachidi, and Yousra Chtouki. 2024. Ai-based voice agent for automated sales calls. In *2024 International Conference on Decision Aid Sciences and Applications (DASA)*, pages 1–4. IEEE.

Lisa P Argyle, Ethan C Busby, Nancy Burgess, Joshua R Gubler, Christopher A Bail, and Thomas Wingfield. 2023. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351.

C. E. Bangle, M. Rosso, F. Montagna, and M. Cantamessa. 2022. Design methods review for smart product: Objectomy, a new approach. *Proceedings of the Design Society*, 2:2045–2054.

Charul Chadha, Kathryn A. Crowe, Christina L. Carmen, and Albert E. Patterson. 2018. Exploring an am-enabled combination-of-functions approach for modular product design. *Designs*, 2(4):37.

Wen-Yu Chang and Yun-Nung Chen. 2024. Injecting salesperson's dialogue strategies in large language models with chain-of-thought reasoning. *Preprint*, arXiv:2404.18564.

Sijia Cheng, Wen-Yu Chang, and Yun-Nung Chen. 2025. Exploring personality-aware interactions in salesperson dialogue agents. *Preprint*, arXiv:2504.18058.

Sung Chun Choi, Eun Suk Suh, and Chan Jung Park. 2019. Value chain and stakeholder-driven product platform design. *Systems Engineering*, 23(3):312–326.

Andrew Court. 1939. Hedonic price indexes with automotive examples. *dynamics of automobile demand.*

Yixiong Feng, Yuliang Zhao, Hao Zheng, Zhiwu Li, and Jianrong Tan. 2020. Data-driven product design toward intelligent manufacturing: A review. *International Journal of Advanced Robotic Systems*, 17(2):172988142091125.

Sandhya Rani Gaddam, Sarada Jayan, Pentakota Ravi, and Bilal Alatas. 2024. Data-driven sales optimization with regression and chaotic pattern search. *PeerJ Computer Science*, 10:e2144.

John J Horton. 2023. Large language models as simulated economic agents: What can we learn from homo silicus? Working paper, National Bureau of Economic Research.

Chuhao Jin, Kening Ren, Lingzhen Kong, Xiting Wang, Ruihua Song, and Huan Chen. 2024. Persuading across diverse domains: a dataset and persuasion large language model. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1678–1706. Association for Computational Linguistics.

Ajay K. Kohli and Bernard J. Jaworski. 1990. Market orientation: The construct, research propositions, and managerial implications. *Journal of Marketing*, 54(2):1–18.

P Kotler. 2012. Marketing management/philip kotler. *Marketing*, 1009:4Ps.

Kelvin J Lancaster. 1966. A new approach to consumer theory. *Journal of political economy*, 74(2):132–157.

Fred Langerak, Erik Jan Hultink, and Henry SJ Robben. 2004. The impact of market orientation, product advantage, and launch proficiency on new product performance and organizational performance. *Journal of product innovation management*, 21(2):79–94.

Xiaolong Li, Wu Zhao, Yake Zheng, Rui Wang, and Chen Wang. 2014. Innovative product design based on comprehensive customer requirements of different cognitive levels. *The Scientific World Journal*, 2014:1–11.

Alessandro Marincioni, Myriana Miltiadous, Katerina Zacharia, Rick Heemskerk, Georgios Doukeris, Mike Preuss, and Giulio Barbero. 2024. The effect of llm-based npc emotional states on player emotions: An analysis of interactive game play. In *2024 IEEE Conference on Games (CoG)*, pages 1–6. IEEE.

Daniel McFadden. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, pages 105–142. Academic Press.

Kent B. Monroe. 1973. Buyers' subjective perceptions of price. *Journal of Marketing Research*, 10(1):70.

Gouri Naik and V. Srinivasan. 2024. Human- and design-centric source: comparison using requirements checklist. *Proceedings of the Design Society*, 4:65–74.

Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22. ACM.

R. Price and Karla Straker. 2017. The design movement: Two case studies from the edge of the discipline. *The Design Journal*, 20:S4565–S4574.

Huafeng Quan, Shaobo Li, Changchang Zeng, Hongjing Wei, and Jianjun Hu. 2021. Big data driven product design: A survey. *ArXiv*, abs/2109.11424.

Huafeng Quan, Shaobo Li, Changchang Zeng, Hongjing Wei, and Jianjun Hu. 2023a. Big data and ai-driven product design: A survey. *Applied Sciences*.

Huafeng Quan, Shaobo Li, Changchang Zeng, Hongjing Wei, and Jianjun Hu. 2023b. Big data and ai-driven product design: A survey. *Applied Sciences*, 13(16):9433.

Marisela Rodríguez-Salvador, Francisco Paredes, and Gaofeng Yi. 2016. Towards future customer experience: Trends and innovation in retail. *Foresight and STI Governance*, 10(3):18–28.

Marko Sarstedt, Susanne J. Adler, Lea Rau, and Bernd Schmitt. 2024. Using large language models to generate silicon samples in consumer and marketing research: Challenges, opportunities, and guidelines. *Psychology & Marketing*, 41(6):1254–1270.

Ryan Shea, Aymen Kallala, Xin Lucy Liu, Michael W. Morris, and Zhou Yu. 2024. Ace: A llm-based negotiation coaching system. *Preprint*, arXiv:2410.01555.

M. N. S. Tissera, P. P. G. D. Asanka, and R. A. C. P. Rajapakse. 2024. Enhancing customer segmentation using large language models (llms) and deterministic, independent-of-corpus embeddings (dice). In *2024 4th International Conference on Advanced Research in Computing (ICARC)*, pages 73–78.

Fanfan Wang, Qingjin Peng, Runhua Tan, Kang Wang, Shan Zhang, and Jianguang Sun. 2022. An attribute-based method for identifying implicit contradictions in product development. *Computer-Aided Design and Applications*, pages 439–455.

Mengxin Wang, Dennis J. Zhang, and Heng Zhang. 2025a. Large language models for market research: A data-augmentation approach. *Preprint*, arXiv:2412.19363.

Ye Wang, Tong Li, Meixuan Li, Ziyue Cheng, Ge Wang, Hanyue Kang, Yaling Deng, Hongjiang Xiao, and Yuan Zhang. 2025b. Rvbench: Role values benchmark for role-playing llms. *Computers in Human Behavior: Artificial Humans*, 5:100184.

Changshuai Wei, Benjamin Zelditch, Joyce Chen, Andre Assuncao Silva T Ribeiro, Jingyi Kenneth Tay, Borja Ocejo Elizondo, Sathiya Keerthi Selvaraj, Aman Gupta, and Licurgo Benemann De Almeida. 2024. Neural optimization with adaptive heuristics for intelligent marketing system. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, pages 5938–5949. ACM.

Ke Xu, Yingguang Li, Changqing Liu, Xu Liu, Xiaozhong Hao, James Gao, and Paul G. Maropoulos. 2020. Advanced data collection and analysis in data-driven manufacturing process. *Chinese Journal of Mechanical Engineering*, 33(1).

Liyang Zhao, Olurotimi Seton, Himadeep Reddy Reddivari, Suvendu Jena, Shadow Zhao, Rachit Kumar, and Changshuai Wei. 2025. Causal predictive optimization and generation for business ai. *Preprint*, arXiv:2505.09847.

A Intercept Calibration Algorithm and Proof

This appendix details the calibration process for the intercept β_0 of the theoretical model, as discussed in Section 3.4. The objective is to align the model's optimal market price with the observed real-world price p_{actual}.

A.1 Calibration Objective

We treat the theoretical reference price RP_{the} (defined in Eq. 4) and the total profit function $\pi(p)$ (defined in Eq. 1) as functions of the base intercept β_0. The optimal price under a specific intercept is given by:

$$p^*(\beta_0) = \arg\max_{p \geq 0} \ \pi(p; \beta_0) \tag{14}$$

The **Intercept Calibration Problem** is to find a value β_0^* such that $p^*(\beta_0^*) = p_{\text{actual}}$.

A.2 Calibration Algorithm

Since $p^*(\beta_0)$ is an implicit, step-wise function, we employ Algorithm 1. This method exploits the property that in a discrete demand model, the optimal price must align with the reference price of a specific consumer group.

Algorithm 1 Iterative Calibration for Intercept β_0

Require: Initial intercept $\beta_0^{(0)}$, observed price p_{actual}, dimension coefficients $\{\beta_m^+, \beta_m^-\}_{m=1}^M$, group profiles $\{\vec{x}_i\}_{i=1}^K$, group sizes $\{N_i\}_{i=1}^K$.
Ensure: Calibrated intercept β_0.

```
β_0 ← β_0^(0)
repeat
    {Update RPs based on Eq. 4}
    RP_i ← RP_the(x⃗_i; β_0) for all i ∈ {1, ..., K}.
    {Solve for optimal price based on Eq. 1}
    Enumerate candidate prices p ∈ {RP_i}_{i=1}^K.
    p* ← arg max_p π(p).
    {Identify purchasing set (groups willing to buy)}
    B ← {i ∈ {1, ..., K} | RP_i ≥ p*}.
    Calculate total quantity Q ← Σ_{i∈B} N_i.
    Adjustment Δ ← p_actual − p*.
    {Update Intercept}
    β_0 ← β_0 + Δ.
until p* = p_actual
return β_0
```

A.3 Proof of Convergence

Proposition 1. *The iterative process in Algorithm 1 monotonically adjusts β_0 and converges to a solution satisfying $p^* = p_{actual}$ in finite steps.*

Proof. **Notation.** Let subscripts (t) denote the state at iteration t. Let $RP_i^{(t)}$ be the reference price for group i. Let $p^{(t)}$ be the optimal price. We define the **purchasing set** at iteration t as $\mathcal{B}^{(t)} = \{i \mid RP_i^{(t)} \geq p^{(t)}\}$, representing the indices of all consumer groups willing to purchase at price $p^{(t)}$. Let $Q^{(t)} = \sum_{i \in \mathcal{B}^{(t)}} N_i$ be the total quantity demanded. The update rule is $\beta_0^{(t+1)} = \beta_0^{(t)} + \Delta^{(t)}$, where $\Delta^{(t)} = p_{\text{actual}} - p^{(t)}$. This implies a uniform shift: $RP_i^{(t+1)} = RP_i^{(t)} + \Delta^{(t)}$.

Monotonicity. We assume $\Delta^{(t)} > 0$ without loss of generality. Let $p^{(t)} = RP_a^{(t)}$ for some group a. Let the new optimal price be $p^{(t+1)} = RP_b^{(t+1)} = RP_b^{(t)} + \Delta^{(t)}$ for some group b.

From the optimality of $p^{(t)}$ at step t:

$$(RP_a^{(t)} - c) \cdot Q^{(t)} \geq (RP_b^{(t)} - c) \cdot Q_b \tag{15}$$

where Q_b is the quantity demanded at price $RP_b^{(t)}$.

From the optimality of $p^{(t+1)}$ at step $t+1$:

$$(RP_b^{(t+1)} - c) \cdot Q_b \geq (RP_a^{(t+1)} - c) \cdot Q^{(t)} \tag{16}$$

Substituting $RP^{(t+1)} = RP^{(t)} + \Delta^{(t)}$ into Eq. (16) and subtracting Eq. (15) yields:

$$\Delta^{(t)} \cdot Q_b \geq \Delta^{(t)} \cdot Q^{(t)} \tag{17}$$

Since $\Delta^{(t)} > 0$, we have $Q_b \geq Q^{(t)}$. In standard demand theory, a non-decreasing quantity implies a non-increasing price point, thus $RP_b^{(t)} \leq RP_a^{(t)} = p^{(t)}$. Examining the new adjustment:

$$\Delta^{(t+1)} = p_{\text{actual}} - p^{(t+1)} = p^{(t)} - RP_b^{(t)} \geq 0 \tag{18}$$

The sign of Δ preserves monotonicity.

Convergence. Since $\{\beta_0^{(t)}\}$ is monotonic and $p^{(t)}$ is selected from a finite discrete set, the algorithm must reach the target in finite steps. □

B Prompt Templates

Note: In the actual experiments, all prompts, product documents, and product category classifications were in Chinese. The following are English translations for reference.

B.1 Phase 1: Market Segmentation

This phase decomposes the market into orthogonal dimensions and defines consumer groups.

Table 3: Prompt template used in Phase 1 (Market Segmentation).

Phase 1: Market Segmentation

System: You are a market analysis expert specializing in product market segmentation analysis based on Hedonic Price Index theory. You need to deeply understand the product's value attributes and segment user groups accordingly.
User:
`<task>`
Based on Hedonic Price Index theory, perform market segmentation analysis for this product.
`</task>`
`<analysis_steps>`
1. Extract N independent orthogonal neutral objective attributes from the product document. Attributes should be neutral objective properties without value judgments, and most attributes should have two-sided nature. Also consider attributes inherent to the product category itself.
2. Define all potential user groups who have demand for this product based on the product's value proposition.
3. Find M user dimensions corresponding to one or more attributes, where M ≤ N. Prioritize attributes that most significantly affect user value judgments. Ensure all user dimensions are strictly orthogonal and independent. All attributes must be assigned to a user dimension.
4. For each user dimension, provide three-part division criteria based on the value perception direction of positive attributes.
`</analysis_steps>`
`<principles>`
`<value_attribute_principles>`
According to Hedonic Price Index theory, a user's value perception of a product is the sum of their value perceptions of each feature attribute, meaning product value can be decomposed into attribute values. Based on this theory, multiple attributes can be defined for each product, with the following constraints:

Continued on next page

Table 3 – continued from previous page

Phase 1: Market Segmentation
1. Attribute definitions must extract neutral objective properties from product documents (including hard design, core features, value proposition, etc., which may contain promotional descriptions), without value judgments, ensuring most attributes have two-sided nature (positive for some users, negative for others). 2. Attribute descriptions should only objectively describe the factual manifestation of the attribute on the product, without discussing usage experience or purpose, and without value evaluation. 3. Do not only focus on properties introduced in product documents; also consider attributes inherent to the product category itself (e.g., height-increasing attribute of high heels, cushioning attribute of sports shoes). 4. User perceptions of different attributes should be as independent as possible, i.e., there should be no situation where users perceiving attribute A as more valuable necessarily perceive attribute B as more or less valuable. 5. Should not miss any attribute that independently affects the overall product value. 6. Should not include attributes that almost do not affect users' value perception of the product. </value_attribute_principles> <user_group_principles> All users in the market who have demand for the product are called the potential user group (as long as a user might purchase the product, they should be considered part of the user population). An accurate criterion for determining whether any user belongs to the potential user group must be provided. The criterion should satisfy the following constraints: 1. Based solely on the written criterion, it should be possible to accurately determine for each user in the overall market whether they belong to the potential user group. 2. The potential user group criterion should be based on users' own hard objective conditions. The criterion should not be too complex, a rough division is sufficient. The population range defined by the criterion may be slightly larger than the actual potential group with real purchase demand, but must not miss any user who might purchase. 3. You need to think about various groups of people in the world, carefully analyze and deeply consider whether they have demand for this product. 4. Finally, check that it does not conflict with any group divided by user dimensions. </user_group_principles> <user_dimension_principles> For the overall potential user group with demand for the product, the overall can be divided according to individual situations of a certain feature or attribute, and the feature or attribute used is called a dimension. If the overall is divided into k parts on each dimension, then setting M dimensions can divide the overall into k^M parts, called user groups. To comprehensively and efficiently analyze all user groups with different perceptions of a product in the market, user dimensions should correspond to product attributes, i.e., user differences on each dimension affect and only affect users' perception of corresponding attributes. Since dimensions must be orthogonal and independent, two attributes with negatively correlated perceptions must correspond to the same dimension and must be specially marked. For each dimension, positive attributes and negative attributes need to be distinguished: - **Positive attributes**: Negative value groups perceive reduced product value due to this attribute, positive value groups perceive increased value, neutral value groups have no perception of this attribute. - **Negative attributes**: Negative value groups perceive positive value from this attribute, positive value groups perceive negative value, neutral value groups have no perception of this attribute (i.e., opposite to positive attributes' value perception direction). Based on N attributes, corresponding M user dimensions should be found, satisfying the following constraints: 1. $M \leq N$, i.e., the number of user dimensions does not exceed the number of attributes. 2. Each user dimension must have at least one positive attribute corresponding to it; negative attributes can be empty (i.e., there can be no negative attributes). Each attribute must correspond to exactly one user dimension. 3. Each user dimension affects and only affects its corresponding attributes (including positive and negative attributes), with almost no effect on other attributes. 4. User dimensions are strictly independent and orthogonal, i.e., there should be no situation where when a user has value a on dimension A, the probability of having value b on dimension B greatly increases or decreases. 5. The naming of each user dimension should be from the user's perspective, not from the product's perspective. </user_dimension_principles> <dimension_division_principles> On each user dimension, divide the overall potential user group into three parts, corresponding to perceptions of the current product's performance on the dimension's corresponding **positive attributes** as: - Negative value: Users resist the product's performance on this attribute; the product's performance on this attribute significantly reduces users' purchase intention and psychological price. - Neutral value: Users do not need this attribute; the product's performance on this attribute has almost no effect on users' purchase intention and psychological price.

Continued on next page

Table 3 – continued from previous page

Phase 1: Market Segmentation
- Positive value: Users identify with this attribute; the product's performance on this attribute significantly increases users' purchase intention and psychological price. Note: For the dimension's **negative attributes**, the value perception direction is opposite to positive attributes, i.e.: - Negative value groups perceive positive value from negative attributes. - Neutral value groups still perceive neutral value from negative attributes. - Positive value groups perceive negative value from negative attributes. Provide division criteria for the three parts. These three division criteria should satisfy the following constraints: 1. Based solely on the written criterion for a single part, it should be possible to accurately determine for each user in the overall whether they belong to this part. 2. After classifying each part according to the first constraint, **the union of the three parts should equal the overall without omission**, and the three parts should **have no intersection**. 3. Groups divided by each part must not conflict with the potential user group's overall criterion. 4. You need to put yourself in the shoes of real users, think about what value and burden this product can bring to such users, and because of the product's performance on each attribute related to the current dimension, what kind of users would like this product more, and what kind of users would dislike it more. 5. You need to ensure that all users in each part will 100% form value perceptions consistent with the part's expectations, i.e., if actual research is conducted on users in the three parts, all users' changes in purchase intention (decrease, unchanged, increase) due to the product's performance on all related attributes should match expectations. This means the criterion must strictly determine the part's user value perception, with no exceptions, not just that there might be people with expected perceptions. 6. Pay special attention that negative value groups must **significantly reduce purchase intention and psychological price** due to the product's performance on all positive-related attributes, and the reduction should be significantly greater than neutral value groups. **If such a population does not exist, honestly set it as an empty set**. 7. Each criterion can only write objective factors about users themselves such as lifestyle, personality traits, physical state, etc. Do not directly state how users view the product's performance on related attributes, but the objective factors you write must be detailed and comprehensive enough that these objective factors necessarily lead to value perceptions of the product's performance on all related attributes consistent with expectations. 8. When almost no group in the overall potential user group holds a certain perception of a set of related attributes, you can set that part as an empty set, writing the corresponding criterion as an empty string "" without forcing a creation of an unreasonable criterion with many exceptions, which especially may occur in the negative value part. Each user dimension **can have at most one part as an empty set**. 9. In writing each part's criterion, you must directly describe the people in that part, e.g., "people who meet... conditions" or "users who satisfy... conditions". It is forbidden to directly indicate in the criterion which category (negative value, neutral value, positive value) the part belongs to, including directly using these three names, or using words like "neutral" or "low intention" that describe the part's classification. 10. For each part's criterion, you need to provide a rationale, explaining why all users defined by this criterion **necessarily** perceive each attribute in the related attribute set as negative value, neutral value, or positive value, i.e., why they resist, do not need, or identify with an attribute. The rationale must completely cover all positive-related and negative-related attributes. The rationale should be based on the correctness of consumer division itself, be concise and clear, with obvious reasonableness, and contain no logical fallacies. The conclusion must be a proof of necessity; if the conclusion can only reach "may produce... perception", it indicates exceptions in the criterion, requiring narrowing and clarifying the criterion scope to ensure exclusion of exceptions. If a criterion is an empty set, the corresponding rationale should also be an empty string. You need to be alert and avoid the following pitfalls: 1. Do not directly assume that an attribute's impact on user value perception is positive just because the current product has made some optimizations on that attribute; start from actual user usage experience. For example, even if a pair of sandals has been treated for durability and stability, for people who exercise frequently, comfort may still correspond to negative value. 2. Do not define people without demand for an attribute as negative value groups; negative value must correspond to the attribute's existence causing user resistance, unwillingness to purchase, or willingness to purchase only at a lower price. People without corresponding demand generally correspond to neutral value rather than negative value. 3. Do not consider factors other than usage experience that affect purchase intention; for example, an attribute's existence causing the product to be expensive should not be a reason for user resistance. In such cases, if users have no demand for the attribute, regardless of how much the attribute increases price, such users should be classified as neutral value rather than negative value. 4. For attributes like convenience and ease of use, do not mistakenly think that people with corresponding professional skills will resist such attributes. Such attributes generally mean neutral value for professionals rather than negative value. 5. When product documents say "can...", it does not mean this is the only usage method. For example, when clothing products say "machine washable", it obviously does not mean they cannot be hand-washed; preferring hand-washing cannot be a reason to resist this attribute.

Continued on next page

Table 3 – continued from previous page

Phase 1: Market Segmentation
6. If two attributes cannot perfectly determine users' value perception of them simultaneously using the division on the same user dimension, i.e., cannot be completely correlated, they should be considered uncorrelated, divided into two dimensions respectively, or one of them should be abandoned. </dimension_division_principles> </principles> <product_info> [Product document including hard_design, core_features, value_proposition] </product_info> <output_format> Output must be in JSON format with the following structure: {...} </output_format

B.2 Phase 2: Proportion Estimation

This phase estimates the proportion of each value stance (negative, neutral, positive) within each dimension.

Table 4: Prompt template used in Phase 2 (Proportion Estimation).

Phase 2: Proportion Estimation
System: You are a market research expert specializing in estimating population proportions based on user characteristics. You need to deeply analyze the feature distribution across user dimensions and provide accurate population proportion estimates. **User:** <task> Based on the provided potential user group and the three-part division criteria for each user dimension, estimate the proportions of negative value, neutral value, and positive value groups within each dimension in the overall potential user group. </task> <principles> <proportion_estimation_principles> 1. Deeply analyze the universality of dimension features. Consider the real feature distribution of various groups in society, avoiding stereotypes or biases. 2. Carefully analyze the user characteristics described in each dimension's division criteria, and consider the actual population size and relative proportions that meet these criteria. 3. Consider correlations between different features. Some features may co-occur more frequently, while others may be mutually exclusive. 4. Refer to statistical knowledge and life experience, considering that feature distributions often follow long-tail, normal, or power-law patterns. 5. For each dimension, deeply consider the real proportional relationships among negative value, neutral value, and positive value groups, considering feature scarcity and universality. 6. Proportion numbers should be based on logical reasoning and reasonable assumptions, not random guesses. </proportion_estimation_principles> <constraints> 1. The sum of three proportions for each dimension must equal 1.000. 2. Each proportion should be a pure decimal, rounded to three decimal places. 3. Each proportion represents the proportion of the current part in the current dimension within the overall potential user group. 4. Proportions should be as close as possible to real-world feature distributions. </constraints> </principles> <user_group> [Potential user group definition] </user_group> <dimensions>

Continued on next page

Table 4 – continued from previous page

Phase 2: Proportion Estimation

`[Dimension information with three-part division criteria]`

</dimensions>

`<output_format>`

Output must be an array format, directly outputting an array of M objects, each containing:

- dimension_name: dimension name;

- proportions: array of three floats representing negative value, neutral value, and positive value proportions.

</output_format

B.3 Phase 3: Consumer Feedback

This phase simulates consumer role-play to obtain reference prices and optimization suggestions.

Table 5: Prompt template used in Phase 3 (Consumer Feedback).

Phase 3: Consumer Feedback

System: You are now the specified consumer yourself. You will always use first-person "I" to express yourself, not from an observer's perspective or third-person analysis. You have rigorous self-reflection capabilities and can derive credible profiles, value assessments, psychological prices, and optimization suggestions based on given materials, real situations, and logical reasoning.

User:

<task>

You need to construct "my" user profile based on consumer_group_definition and complete subsequent analysis. consumer_group_definition contains:

1. potential_user_group: defines the potential user group that "I" must belong to.

2. Each Dimension: the content is "conditions for belonging to this part", representing the conditions that "I" must satisfy on this dimension.

You must ensure that "my" profile completely satisfies all conditions in potential_user_group and all Dimensions, then complete the following steps, all using first-person "I":

1) Construct a unique and representative [*user profile*], confirming I satisfy all group criteria.

2) Output a [*user profile rationale*], itemizing full proof that the current user profile satisfies the definition requirements of potential_user_group and the specific criteria of each Dimension.

3) For each attribute in product_attributes, analyze what effects this attribute's performance in the current product will have in my actual use.

4) Score each attribute (-2 to 2).

5) Provide targeted optimization suggestions for each attribute.

6) Give a [*psychological price*] (the maximum price at which I am willing to purchase).

</task>

<constraints>

`[Understanding and Mapping of consumer_group_definition]`

1. consumer_group_definition contains two parts:

a) potential_user_group: defines the overall potential user group that "I" must belong to.

b) Each Dimension: the content of each Dimension is "conditions for belonging to this part", i.e., conditions that "I" must satisfy.

2. You must parse all conditions in consumer_group_definition item by item, mapping each definition point to specific, verifiable facts about "me", achieving one-to-one correspondence with no omissions.

`[User Profile Evaluation Standards]`

1. You should cover key information affecting purchase demand and usage experience: identity background, life structure and rhythm, common environments and tools, abilities and mental characteristics, social relationships and roles, information sources and trust preferences, typical scenarios and limitations related to the product, frequency and duration of related behaviors, disposable resources and key constraints, etc.

2. Each piece of profile information you state must be unique and specific, prohibiting range/selective/probabilistic expressions (no expressions like "between... and...", "is... or...", "possibly/probably/one of/between.../depending on the situation", etc.).

3. Profile information should be self-consistent without conflicts, and ensure "I" indisputably satisfy all criteria in consumer_group_definition, while being reasonable and credible in real life.

Continued on next page

Table 5 – continued from previous page

Phase 3: Consumer Feedback

[User Profile Rationale Evaluation Standards]

1. In the user profile rationale, you must itemize and fully prove that the current user profile satisfies the following requirements:

a) Prove that "I" satisfy all conditions defined in potential_user_group.

b) For each Dimension, prove that "I" satisfy the "conditions for belonging to this part" in that Dimension.

2. The rationale must be clear and organized, with obvious reasonableness, which must be based on the user profile itself actually meeting these conditions.

[Attribute Analysis Evaluation Standards]

1. For each attribute in product_attributes, based on my real usage scenarios, analyze what effects (including positive and negative effects) the attribute's performance on the current product will have on me.

2. You must clearly explain through what mechanism each attribute's effects are produced, in what scenarios they occur, and how they relate to my profile elements (identity background, life scenarios, ability characteristics, etc.).

3. Based on the effect analysis, give accurate value scores (-2 to 2) for each attribute, where -2 represents strong negative value (the attribute's existence severely reduces my psychological price for the current product), -1 represents negative value (the attribute's existence slightly reduces my psychological price), 0 represents no effect, 1 represents positive value (the attribute's existence slightly increases my psychological price), 2 represents strong positive value (the attribute's existence significantly increases my psychological price). Scores must be logically consistent with effect analysis.

4. You should provide targeted optimization suggestions for each attribute, explaining how to improve the attribute to increase my psychological price. Suggestions should be specific and actionable. **Note: Do not suggest reducing product cost or lowering selling price.**

5. If you think an attribute already needs no optimization, you can omit the suggestion, writing it as an empty string "".

[Psychological Price Evaluation Standards]

1. You can only make psychological price judgments based on all attributes' value scores, budget constraints and other constraints given in consumption_budget, prohibiting reference to any external selling prices, costs, discounts, competitors, or market conditions.

2. If the product belongs to household consumption expenditure categories, you must consider the annual budget for such expenditure and all necessary consumer goods and services included in such expenditure, and conservatively evaluate the reasonable proportion of the current product in that budget.

3. Psychological price must be expressed as a RMB amount number (without any units or symbols), as a pure floating-point number with two decimal places. This value should be derivable and explainable from all attributes' value scores and constraints, and be verifiable.

4. All your expressions (including list items) must use first-person "I".

</constraints>

<consumption_budget>

[Consumption budget information or indication that product is not in consumption expenditure category]

</consumption_budget>

<product_info>

<modified_name>[*Product name*]</modified_name>

<hard_design>[*Hard design*]</hard_design>

<core_features>[*Core features*]</core_features>

<value_proposition>[*Value proposition*]</value_proposition>

</product_info>

<product_attributes>

[List of product attributes with descriptions]

</product_attributes>

[Consumer group definition with potential_user_group and all Dimensions]

`<output_format>`

Output must be JSON with fields: user_profile, user_profile_rationale, attribute_analysis (array), psychological_price.

</output_format

B.4 Phase 4: Product Optimization

This phase optimizes product design based on aggregated market feedback.

Table 6: Prompt template used in Phase 4 (Product Optimization).

Phase 4: Product Optimization

System: You are a product manager responsible for optimizing products based on market feedback results, aiming to maximize market return (number of purchasers × product selling price).

User:

<task>

You are a product manager responsible for optimizing products based on market feedback results, aiming to maximize market return.

Please complete the optimization task according to the following steps:

1. Carefully read the complete current product design document.
2. Understand the product's user dimension divisions and attribute system.
3. Analyze the attribute optimization priority table, understanding the meaning and importance of each optimization direction.
4. Reference optimization suggestions to think about how to improve product design.
5. Ensure technical feasibility: optimized product design must be producible and non-contradictory.
6. Generate optimized product design document.
7. Generate suggestion analysis for each optimization direction (whether adopted and reasons).
8. Actively argue for the feasibility of current optimization.

</task>

<current_product_document>

<hard_design>[*Hard design*]</hard_design>

<core_features>[*Core features*]</core_features>

<value_proposition>[*Value proposition*]</value_proposition>

</current_product_document>

<user_dimensions>

[User dimension information with division criteria and proportions]

</user_dimensions>

<product_attributes>

[Product attributes with corresponding dimensions]

</product_attributes>

<market_research_explanation>

Based on several product attributes, we identified a series of corresponding user dimensions. On each user dimension, we divided users into three parts: those who perceive related attributes as resistance (negative value), indifference (no value), and preference (positive value). We then selected representative consumers from various user dimension value combinations for research, obtaining their valuations of the current product, satisfaction scores for each attribute, and optimization suggestions for each attribute.

</market_research_explanation>

[Optimization priority table in XML format]

<optimization_table_explanation>

This table lists optimization directions for each product attribute, sorted by importance in descending order. Each optimization direction (item) contains the following fields:

1. **attr_name_and_direction**: Description of the optimization direction, formatted as "Attribute name - Optimization direction". There are three types of optimization directions:

- "Reduce negative impact on negative value groups": Targeting groups with negative value on the corresponding user dimension for this attribute.

- "Enhance appeal to neutral value groups": Targeting groups with neutral value on the corresponding user dimension for this attribute.

- "Increase satisfaction for positive value groups": Targeting groups with positive value on the corresponding user dimension for this attribute.

2. **current_performance**: The product's performance on this attribute before optimization, i.e., the specific description of this attribute.

3. **user_criteria**: Definition of the optimization target group; all groups providing suggestions satisfy this definition. This is the criterion for the corresponding user dimension of this attribute at this value (negative value or positive value).

4. **consumer_proportion**: The proportion of consumer groups corresponding to the current optimization direction, i.e., the sum of proportions of all groups satisfying user_criteria.

Continued on next page

Table 6 – continued from previous page

Phase 4: Product Optimization

5. **importance_score**: Importance indicator for the current optimization direction, calculated as:
$I = p^2 \times \sum(\text{group proportion}/(|\text{group valuation} - \text{optimal price}| + \epsilon)^2)$
where p is the optimal price and $\epsilon = p/100$. Larger values indicate higher importance; this table is sorted by this indicator in descending order.

6. **reference_consumer_list**: Several representative consumer objects from the optimization direction group, obtaining their experience feedback. Contains at most 5 reference consumers, sorted by group importance indicator in descending order. Each reference consumer (reference_consumer) contains:

- **profile**: The consumer's user profile.
- **proportion**: The proportion of people similar to this consumer (i.e., the group's proportion).
- **effect**: The consumer's usage effect experience.
- **score**: The consumer's score for usage experience, values -2, -1, 0, 1, 2, corresponding to strong resistance, resistance, indifference, preference, strong preference respectively.
- **suggestion**: The consumer's suggestion for optimizing this attribute.
- **valuation**: The consumer's valuation of the entire product, indicating the price below which the consumer is willing to purchase.

`</optimization_table_explanation>`

`<optimization_rules>`

1. The sole purpose of optimization is to increase each user group's valuation of the product.
2. Think independently while appropriately referencing optimization suggestions.
3. Technical feasibility requirements: ensure optimized design is producible and non-contradictory.
4. Document formality requirements: optimized documents must be formal versions without any editing traces.

`</optimization_rules>`

`<output_format>`

Output must include: hard_design, core_features, value_proposition, feasibility_analysis, suggestion_analysis.

`</output_format`

B.5 Baseline: Zero-Shot Product Optimization

The baseline method optimizes products directly based on product documents and general market knowledge, without access to structured, granular feedback from PRISM simulation.

Table 7: Prompt template used in the Zero-Shot baseline.

Baseline: Zero-Shot Product Optimization

System: You are a senior product manager responsible for product optimization and improvement, specializing in deeply identifying user pain points and designing targeted solutions. You need to perform minor optimizations based on product documents, ensuring that optimized designs are technically feasible, producible, and cost-effective. Most importantly, all optimizations must directly address user pain points and solve actual problems users encounter when using the product.

User:

`<task>`

You are a senior product manager responsible for product optimization and improvement, aiming to enhance product market competitiveness and user value.

Please complete the optimization task according to the following steps:

1. **Deeply analyze user pain points**: Carefully read the complete current product design document, identify core pain points users may encounter when using this product from user usage scenarios, including but not limited to:
   – User troubles caused by missing or incomplete functions;
   – Design flaws affecting user experience;
   – Material or process issues affecting product durability or comfort;
   – Inconvenience in use or maintenance difficulties;
   – Insufficient safety or reliability;
   – Unreasonable cost-effectiveness.

Evaluation criteria: Only consider optimization when pain points reach medium or higher severity and can be solved through mature, feasible technical solutions. If pain points are minor or already largely resolved, maintain the status quo.

2. **Minor optimization design**: Design minor, targeted optimization solutions for identified user pain points, ensuring:

Continued on next page

Table 7 – continued from previous page

Baseline: Zero-Shot Product Optimization
– Each optimization point directly corresponds to a clear user pain point; – Optimization solutions can effectively solve or alleviate the pain point; – Optimized improvements are visible and perceivable to users; – Prioritize solving key pain points affecting core user experience. 3. **Ensure technical feasibility**: Optimized product design must be producible and non-contradictory. 4. **Ensure cost reasonableness**: Optimization should not lead to significant cost increases; prioritize cost-effective improvement solutions. If cost increases exceed 5%, evaluate whether optimization is worthwhile. 5. **Generate optimized product design document**. 6. **Generate before-and-after comparison**: Clearly show the before-and-after comparison of each optimization point. 7. **Actively argue for the feasibility of current optimization**. </task> <current_product_document> <hard_design>[*Hard design document*]</hard_design> <core_features>[*Core features document*]</core_features> <value_proposition>[*Value proposition document*]</value_proposition> </current_product_document> <optimization_rules> 1. **Pain point orientation principle**: All optimizations must directly address user pain points. Each optimization point should correspond to a clear user problem. Only optimize when pain points reach medium or higher severity. 2. **Minor optimization principle**: If optimization is performed, make minor improvements, avoiding major changes. 3. **Visibility requirement**: Optimized improvements should be visible and perceivable to users. 4. **Technical feasibility requirement**: Ensure optimized product design is producible and non-contradictory. 5. **Cost reasonableness requirement**: Optimization should not lead to significant cost increases. Prioritize cost-effective solutions. 6. **Targeted improvement**: Optimizations should be targeted improvements, each change should have a clear reason and corresponding user pain point. 7. **Maintain product core positioning**: Optimization should not change the product's core positioning and main functional direction. 8. **Document formality requirement**: Optimized documents must be formal versions without any editing traces. </optimization_rules> <output_format> Output must include: hard_design, core_features, value_proposition, feasibility_analysis, optimization_comparison, optimization_summary. </output_format