SimRetail: A Persona-Informed Multi-Agent System for Autonomous Retail Assortment Planning

Demonstration Track

Sowmya Podila
Target Corporation
Minneapolis, United States
sowmya.podila@target.com

Andrew Yang
Target Corporation
Minneapolis, United States
andrew.yang@target.com

Bo Shen
Target Corporation
Minneapolis, United States
bo.shen@target.com

Matthew Nokleby
Target Corporation
Minneapolis, United States
matthew.nokleby@target.com

ABSTRACT

SimRetail is an LLM-based, persona-informed multi-agent system designed to optimize retail assortment selection. SimRetail comprises a *planning agent*—built on the React [7] framework to orchestrate interactions with subagents via structured reasoning, targeted tool use, and multi-step decision workflows—and a *scoring agent* that employs Nemotron synthetic personas [4] to score the outputs of the planning agent. The planning agent coordinates a set of specialist agents—including trend analysis, merchandising performance, and vendor item analysis—to generate item assortment selections. The scoring agent judges the resulting assortment against the feedback of the synthetic personas, allowing the user to choose the best assortment. We provide a user-friendly interface to SimRetail using Streamlit [6].

KEYWORDS

multi-agent systems; personas; assortment planning; LangGraph; retail decision support

ACM Reference Format:
Sowmya Podila, Andrew Yang, Bo Shen, and Matthew Nokleby. 2026. SimRetail: A Persona-Informed Multi-Agent System for Autonomous Retail Assortment Planning: Demonstration Track. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/http://doi.org/10.65109/SMUK4801

1 APPLICATION DOMAIN

Retail assortment planning requires selecting the right product mix for diverse customer segments while accounting for market trends, vendor constraints, and item availability. Currently, human merchandisers manually gather insights across multiple systems, resulting in slow and often inconsistent decisions.

SimRetail addresses the following problem: given a product category and target persona, determine (1) which products best adhere to current trends and maximize conversion, (2) which consumer segments exhibit strong purchase intent, (3) what factors drive their decisions, and (4) which products maintain overall category viability. Achieving this requires coordinated reasoning across trend intelligence, merchandising analytics, vendor evaluation, and supply chain constraints to optimize across these factors. However, even with historical data and demand forecasts it is difficult to predict which assortment items will resonate with consumers.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/http://doi.org/10.65109/SMUK4801

2 AUTONOMOUS MULTI-AGENT SYSTEM ARCHITECTURE

We address this problem in SimRetail with two agents: a *planning agent* and a *scoring* agent, which interact to provide an optimized retail assortment. The planning agent has a supervisor–specialist ReAct-style design, using "specialist" agents as tools. The scoring agent provides predicted demand metrics across multiple synthetic consumer personas, sourced from Nvidia's Nemotron personas dataset [4]. The interface presents the user with the optimized item selection using the generated demand metrics.

2.1 Planning Agent Design

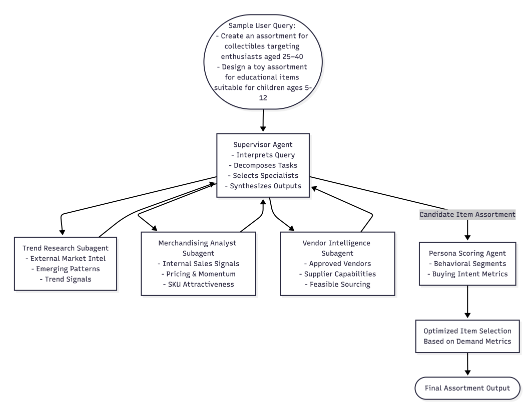

Figure 1: Planning and scoring agent interaction architecture.

As shown in Figure 1, the planning agent has a central supervisor agent that interprets user queries, decomposes tasks, selects relevant subagents autonomously, and synthesizes outputs. It interacts with its subagents over multiple rounds of coordination using the ReAct framework [7] as implemented in LangGraph [2]. Subagents are connected to domain-specific data sources and carry out specialized sub-tasks. Subagents include:

- **Trend Research Subagent**: Gathers market intelligence and emerging patterns using external sources (e.g., Exploding Topics) to identify relevant trends for the target category.
- **Merchandising Analyst Subagent**: Interprets internal sales and performance signals to assess SKU momentum, pricing, and product attractiveness.
- **Vendor Intelligence Subagent**: Reviews approved vendors, evaluates suppliers, and identifies sourcing options.

LangGraph enables persistent state, checkpointing, and conditional routing. ReAct-style reasoning allows agents to alternate between thought, tool invocation, and observation. Human-in-the-loop interrupt nodes support strategic approvals. All traces are monitored through Langfuse [5] for observability.

2.2 Persona-based Scoring Agent Design

The scoring agent employs synthetic personas to estimate the purchase intent of potential consumers and score the assortments generated by the planning agent. We use the Nemotron-Personas-USA dataset [4], which has 181,819 synthetic personas generated to match demographic profiles of the United States, each with detailed persona and behavioral profiles. Given these personas and an input assortment, the scoring agent carries out the following pipeline:

- **Archetype Classification**: All filtered personas are classified into 8 buyer archetypes (Enthusiast Collector, Practical Parent, Gift Buyer, Trend Follower, Budget Hunter, Quality Seeker, Casual Browser, Nostalgic Buyer) based on behavioral patterns, age, occupation, interests, and education level.
- **Individual LLM Scoring**: A LLM-based agent analyzes each sampled persona individually within their archetype context, generating buying intent scores (0–100%), purchase likelihood (yes/no), and behavioral drivers.
- **Viability Assessment**: Overall conversion rate, weighted average intent across archetypes, persona count, and archetype-level variance analysis generate comprehensive assortment recommendations with viability tiers (High, Marginal, Low) and specific product suggestions by buyer type.

Figure 2: Sample output of the persona-based scoring agent.

The scores are computed and displayed to the user as shown in Figure 2.

3 INTERACTIVE DEMONSTRATION FEATURES

Users are able to modify theme candidates as shown in Figure 3, adjust persona filters, explore archetype breakdowns, and observe the evolving execution graph, such as "Evaluate collectibles for kidults aged 25–40" for candidate item selections."

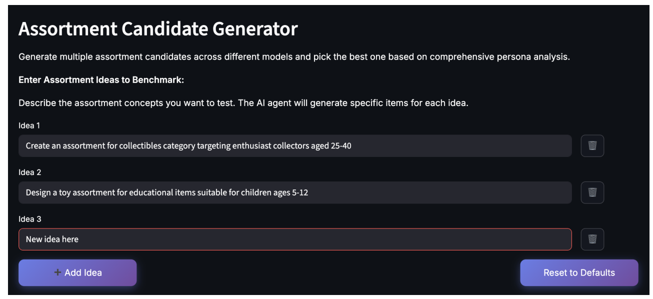

Figure 3: Sample user input themes.

The system dynamically executes the aforementioned LangGraph pipeline and displays the following user interactions:

- **Dashboard Visualizations**: Intent distributions by buyer archetype (Figure 2), persona breakdowns, individual persona analysis results, conversion rate ranges, and reasoning traces with variance indicators.
- **Archetype Analysis**: Detailed breakdown by 8 buyer types showing conversion rates, intent ranges, persona spans, and representative persona examples.
- **Proposed Assortment**: Specific product recommendations with pricing, rationale, and archetype appeal mapping.

4 RELATED WORK

Other works have incorporated LLMs and personas for retail-based applications. InteraSSort [1] endows an LLM with optimization tools (e.g. Cplex and Gurobi) to enable human-assisted iterations of assortment planning and refinement. PAARS [3] generates *retail-focused* personas from shopping histories to enable simulation and synthetic A/B tests.

5 BROADER IMPACT AND APPLICABILITY

While focused on retail, the system generalizes to any domain requiring persona-informed modeling, multi-source analytics, or interdisciplinary agent reasoning, including healthcare, finance, and education.

6 TECHNOLOGY STACK

The system utilizes: LangGraph for deterministic, stateful orchestration; OpenAI GPT-4o for agent execution and persona scoring; Streamlit and Plotly for visualization; and Python 3.12.

REFERENCES

[1] Saketh Reddy Karra and Theja Tulabandhula. 2023. InteraSSort: Interactive Assortment Planning Using Large Language Models. arXiv:2311.12241 [cs.AI] https://arxiv.org/abs/2311.12241
[2] LangChain Inc. 2024. LangGraph: Stateful Multi-Actor Applications. https://langchain-ai.github.io/langgraph/. Accessed 2026.
[3] Saab Mansour, Leonardo Perelli, Lorenzo Mainetti, George Davidson, and Stefano D'Amato. 2025. PAARS: Persona Aligned Agentic Retail Shoppers. arXiv:2503.24228 [cs.AI] https://arxiv.org/abs/2503.24228
[4] Yev Meyer and Dane Corneil. 2025. *Nemotron-Personas-USA: Synthetic Personas Aligned to Real-World Distributions.* https://huggingface.co/datasets/nvidia/Nemotron-Personas-USA
[5] Clemens Rawert, Marc Klingen, and Maximilian Deichmann. 2023. Langfuse — Open-Source LLM Engineering Platform. https://langfuse.com/ Software available from https://langfuse.com.
[6] Streamlit Inc. 2024. Streamlit: Build Data Apps. https://streamlit.io/. Accessed 2026.
[7] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2023. ReAct: Synergizing Reasoning and Acting in Language Models. In *International Conference on Learning Representations (ICLR).* https://openreview.net/forum?id=WE_vluYUL-X