# Large Language Models & Consumer Segmentation

www.bobhurling.com

## Research Papers

Li, Y., Liu, Y., & Yu, M. (2025). Consumer segmentation with large language models. *Journal of Retailing and Consumer Services, 82*, 104078. https://doi.org/10.1016/j.jretconser.2024.104078

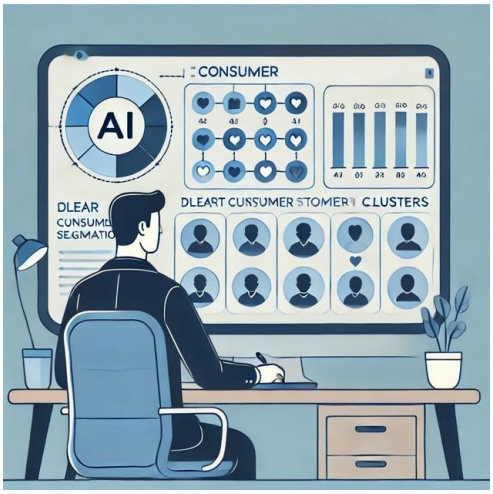
## Overview

Traditional methods of **consumer segmentation** often fail to capture the complexity of consumer preferences, particularly in text based survey responses. This study explores how **Large Language Models (LLMs)** can improve consumer segmentation by embedding and analysing textual data from surveys. The authors claim that **LLMs enhance clustering accuracy**, leading to more refined consumer profiles. Additionally, the study introduces **AI-driven persona chatbots** that simulate consumer preferences with high accuracy, providing an innovative tool for marketing strategy development.

## Key Findings & Insights

| Clusters (k) | Silhouette Coefficient (LLM-Based) | Silhouette Coefficient (Traditional) |
|---|---|---|
| 2 | 0.158 | 0.077 |
| 3 | 0.112 | 0.052 |
| 4 | 0.096 | 0.041 |
| 5 | 0.094 | 0.045 |

LLMs outperform traditional clustering models: using LLM results in more accurate consumer segmentation than conventional methods like One-Hot Encoding. See silhouette coefficients in table, which measure clustering quality; higher values indicate better segmentation




Text-based data holds untapped potential: instead of treating responses as independent variables, LLMs analyse semantic connections between survey answers, improving segmentation quality.

Chatbot personas effectively simulate consumers: AI-driven consumer profiles achieved a 10% error rate, suggesting that LLMs might be able to replicate real consumer behaviours with high accuracy.

LLMs reduce redundancy and bias in survey analysis: traditional surveys often contain redundant or overlapping questions. LLMs identify and resolve semantic redundancies, leading to cleaner, more reliable consumer insights.

Marketing implications: AI-generated consumer personas can be used to test marketing strategies, enabling businesses to simulate and predict how different segments will respond to campaigns.

## Methodology

Dataset: 500 consumer responses from a Chinese liquor company survey. Embedding Process: LLMs converted survey responses into vector representations, enabling semantic clustering. Clustering Approach: K-means clustering identified three primary consumer segments: Connoisseur Buyers – prioritise quality, cultural value. Business Entertainers – emphasise price and brand recognition. Social Drinkers – influenced by endorsements and social trends. Persona Chatbots: AI-generated consumer personas were tested for accuracy by comparing their preferences to real survey data.

## Conclusion

This study suggests that LLMs revolutionize consumer segmentation, offering a more nuanced, accurate, and scalable approach to understanding consumer behaviour. By leveraging AI-powered segmentation, businesses may be able to: Improve market targeting with more precise consumer personas. Reduce survey bias by allowing AI to identify meaningful patterns. Use AI-driven consumer simulations to refine marketing strategies before implementation. LLMs offer a new frontier in marketing research, moving beyond traditional segmentation methods to unlock deeper consumer insights and predictive analytics.