iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75205

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

A Comparative Survey on Conversational LLM-Based Multi-Modal Product Recommendation Systems

Aditya M. Bhosale[1], Nihal K. Barkhade[2], Akshay V. Choudhari[3], Sandeep Warhade[4]
Department of Information Technology, Pune Institute of Computer Technology, Pune, India

Abstract: The exponential expansion of e-commerce platforms has increased accessibility but has also introduced challenges for users attempting to locate reliable, unbiased, and well- structured product information. Existing recommendation engines predominantly rely on static filtering algorithms that operate on historical behavioral data. These systems lack the ability to interpret user intent, adapt to contextual preferences, or justify their recommendations. As a result, users frequently invest significant time comparing products across platforms, manually validating reviews, and navigating inconsistent or biased information. This research proposes a Conversational LLM-Based Multi-Modal Product Research Assistant—an intelligent system that enables interactive, context-aware, and evidence-driven product exploration. The system integrates a domain-adapted Large Language Model (LLM) with multi-modal reasoning capabilities to analyze textual and visual product content. It leverages a Retrieval-Augmented Generation (RAG) pipeline to access real-time, verified data from e-commerce sources, ensuring factual consistency and transparency in the generated responses. The assistant also incorporates unbiased comparison logic and sustainability-aware evaluation aligned with the United Nations Sustainable Development Goal (SDG) 12: Responsible Consumption and Production. The outcome of this work is a scalable web-based assistant that supports natural conversational querying, displays relevant images and specifications, and enables users to make informed and responsible purchase decisions. The study demonstrates that integrating conversational AI with multi- modal retrieval can transform digital shopping into a transparent, personalized, and sustainable decision-making experience. Key- words: Conversational AI, Large Language Models (LLM), Multi- Modal Learn ing, Product Recommendation, Retrieval-Augmented Generation (RAG), Responsible Consumption, E-commerce.

Keywords: Conversational LLM, Multi-Modal Recommendation, RAG, Product Comparison, Sustainable AI

I. INTRODUCTION

E-commerce has evolved into a globally interconnected digital marketplace where millions of products, vendors, and pricing strategies coexist. While this ecosystem offers conve- nience and variety, users increasingly encounter information fragmentation, bias in reviews, and inconsistency in product specifications across platforms. The extensive availability of data paradoxically contributes to information overload, making it difficult for consumers to confidently determine which product best satisfies their needs. Traditional recommendation systems—such as collaborative filtering and content-based filtering—primarily rely on user purchase history, ratings, or simple textual similarity. These approaches suffer from inherent limitations: they lack contextual reasoning, do not understand user intent expressed through natural language, and cannot perform transparent cross-platform comparisons. Consequently, they provide suggestions that may appear accurate statistically but are not necessarily aligned with the user's true preferences or constraints. Recent advancements in Large Language Models (LLMs) present a significant opportunity to overcome these challenges. LLMs are capable of processing natural language queries, inferring user requirements, and synthesizing information from heterogeneous data sources. When combined with multi-modal learning, such models can reason over text as well as visual content—such as product images or specifications—enabling a more intuitive and human-like exploration of product options. The analyzed system, termed *Conversational LLM-Based Multi-Modal Product Research Assistant,* introduces an AI- driven product exploration workflow where users can interact using conversational language. The system retrieves verified information from multiple e-commerce platforms via APIs and utilizes a Retrieval-Augmented Generation (RAG) architecture to ensure that responses are accurate, transparent, and grounded in real-time data rather than hallucinated knowledge. Addition- ally, the system incorporates sustainability indicators—such as material recyclability and environmental impact—to support the United Nations Sustainable Development Goal (SDG) 12: Responsible Consumption and Production.

This work contributes toward bridging the gap between con- versational AI and e-commerce product research. By combining multi-modal inference, unbiased comparison, and sustainability- aware decision support, the analyzed system elevates digital shopping from transactional browsing to informed, responsible, and personalized decision-making.

II. LITERATURE SURVEY

A. *Web-scraping housing prices in real-time: The Covid-19 crisis in the UK*

1) Objective: The primary goal of this study is to demon- strate how web-scraped real-estate listing data can be leveraged as a high-frequency, near real-time indicator of macroeconomic activity, particularly during periods of volatile market behavior such as the Covid-19 pandemic. Traditional housing datasets suffer from long reporting lags, limiting their decision-making relevance. The re- search aims to show that web-scraped housing listings can serve as an early detection mechanism for emerging trends in supply, demand, and pricing.
2) Methodology: Daily listings were scraped from five leading UK real-estate platforms—*Zoopla*, *Rightmove*, *OnTheMarket*, *PropertyPal*, and *S1Homes*. Each listing included structured and semi-structured fields such as price, coordinates, property type, and textual description. Data cleaning included standardization of currency, nor- malization of area units, geolocation mapping, dedupli- cation, and removal of commercial/industrial listings. To approximate real transaction outcomes, scraped asking prices were linked with final closing prices recorded in the UK Land Registry. Matching was performed with a multi-step KNN algorithm that evaluated spatial distance, temporal proximity, and price similarity. The matched por- tion enabled computation of negotiation margins (listing price – transaction price) and allowed analysis of market sensitivity during lockdown fluctuations.
3) Contribution: The findings establish that real-time, web- scraped housing data can serve as a substitute or supple- ment to official house-price indexes. The study provides empirical proof that scraped data can reveal economic shocks weeks before governmental datasets. This has value for early policy intervention, market forecasting, and real- time monitoring of regional demand–supply elasticity.

B. *Web Scraping using Natural Language Processing: Exploit- ing Unstructured Text for Data Extraction and Analysis*

1) Objective: This study introduces a combined web- scraping and NLP framework to automatically collect unstructured text from digital platforms and convert it into structured, machine-interpretable representations. The objective is to streamline the transformation of large- volume textual data into actionable insights for businesses and researchers.
2) Methodology: The pipeline operates in two phases. In the extraction phase, static HTML content is parsed using BeautifulSoup and Scrapy, whereas Selenium is used to handle dynamic, JavaScript-rendered websites. Data collected includes customer reviews, blog text, and news articles. In the analysis phase, preprocessing techniques (tokenization, stop-word removal, stemming) prepare text for semantic analysis. Word embeddings are computed using Word2Vec (Skip-gram/CBOW), where the model maximizes contextual co-occurrence probability:

$$\text{TF-IDF} = \text{TF} \cdot \text{IDF}$$

NER confidence scoring is calculated as:

$$C = P(\text{NE}) - P(\text{not NE})$$

Sentiment classification uses logistic regression:

3) Contribution: The work provides a scalable blueprint for transforming open-web textual content into structured data assets. It emphasizes the need for ethical data practices and introduces a generalized approach applicable across domains including competitive intelligence, trend analysis, and review analytics.

C. *Increasing online shop revenues with web scraping: a case study for the wine sector*

1) Objective: The objective is to integrate web scraping into revenue optimization by continuously monitoring competitor prices and automating price adjustments to improve visibility and competitiveness while ensuring profit margins remain above a predefined threshold.
2) Methodology: An automated scraping system gathers product prices from Uvinum and Vivino. A PostgreSQL database stores internal product catalogs retrieved via authenticated API. The decision engine evaluates whether prices can be lowered while retaining profitability:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

$$\frac{(MinPriceSale - PurchasePrice)}{MinPriceSale} \times 100 > MinMargin$$

Products are classified into three priority actions—YES, CONSIDER, or NO. Execution pipelines are containerized using Docker and parallelized across multiple VMs to reduce scraping time.

3) Contribution: This study demonstrates how direct integra- tion of scraped data into pricing strategy can materially improve business KPIs such as conversion rate, profit margin, and inventory turnover. It also validates the feasibility of fully automated competitive pricing for SMEs.

D. Web Scraping based Product Comparison Model for E- Commerce Website

1) Objective: To build a transparent and vendor-agnostic comparison engine that eliminates manual searching by aggregating product details—including price, availability, and quality—from multiple online marketplaces.
2) Methodology: User input is transformed into platform- specific search URLs. BeautifulSoup handles static extraction, while Selenium manages dynamic content rendering. Extracted fields include product name, price, rating, image URL, and the direct product link. Mon- goDB stores the records to support scalability. The UI uses Node.js/Next.js and visualizes comparisons through tabular views and graphical pricing histograms.
3) Contribution: The approach reduces user decision fatigue and establishes a generalizable scraping and comparison framework. It encourages market price transparency and exposes price inflation strategies used by certain platforms.

E. Advanced Google Scholar Scraper: A Content-Based Fil- tering Approach for Literature Recommendation Using BERT

$$S(z) = \frac{1}{1+e^{-z}} \quad \text{where} \quad z = \beta_0 + \sum \beta_i X_i$$

1) ObjectObjective: To develop an automated academic literature discovery tool that returns semantically relevant research papers based on user keywords rather than simple text matching.
2) Methodology: Selenium automates Google Scholar navigation, followed by HTML parsing using BeautifulSoup. Text embeddings are generated using a pre-trained BERT model. Cosine similarity scores enable ranking of papers based on conceptual closeness rather than lexical similarity. Outputs—including similarity scores—are exported to CSV and delivered through a Tkinter GUI with progress tracking.
3) Contribution: The system reduces manual workload in literature review processes by improving retrieval precision. It bridges the gap between academic search engines and intelligent filtering tools based on semantic relevance.

F. Generating Contextual Variables From Web-Based Data for Health Research: Tutorial on Web Scraping, Text Mining, and Spatial Overlay Analysis

1) Objective: To introduce a systematic approach for gener- ating neighborhood-level health research variables using publicly available data when traditional health datasets are absent or incomplete.
2) Methodology: The WeTMS (Web scraping → Text mining → Spatial overlay) framework scrapes activity listings, cleans and tokenizes textual descriptions, classifies content into thematic categories, and applies geospatial overlay to map activities to Basic Health Areas (BHAs). Validation through manual sampling ensures reliability.
3) Contribution: The paper proves that domain knowledge and geospatial analytics can convert unstructured web data into meaningful public-health indicators for research, planning, and policymaking.

G. Unlocking The Potential Of Web Data For Retailing Research

1) Objective: To popularize web data as an alternative research source by analyzing prior retailing studies and demonstrating practical acquisition methods.
2) Methodology: A systematic review of 28 research articles is conducted to categorize how web data were used in pricing studies, competitive analysis, consumer behavior modeling, and analytics innovation. The authors also design a mock digital store and provide scraping scripts and API examples for experimentation.
3) Contribution: The study lowers the technical barrier for researchers and provides a methodological roadmap that expands retail data analysis beyond proprietary datasets and limited geographical scopes.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

H. Multi-Level Cross-View Contrastive Learning for Knowledge-Aware Recommender System

1) Objective: The research aims to enhance the robustness of knowledge-aware recommender systems (KGRs) by resolving sparse supervision and embedding collapse. It introduces a multi-level, multi-view contrastive learning paradigm that ensures richer semantic and topological representation of users and items, enabling more stable generalization across domains.
2) Methodology: MCCLK constructs three complemen- tary graph representations of the same dataset: (i) a collaborative-view graph derived from LightGCN to embed user–item interactions, (ii) a semantic-view graph modeling entity and attribute semantics using a relation- aware GNN, and (iii) a structural-view graph capturing higher-order connectivity and global dependency paths using a path-aware encoder. Local contrastive learning aligns collaborative-item em- beddings with semantic-view embeddings, ensuring con- sistency of micro-level patterns (e.g., item similarity). Global contrastive learning aligns fused embeddings with structural-view representations, ensuring system-level understanding of long-range relationships. Optimization combines Bayesian Personalized Ranking (BPR) loss with view-based contrastive losses, improving representation discriminability and reducing embedding collapse.
3) Contribution: MCCLK is the first solution to unify self- supervised contrastive objectives across local and global graph perspectives. It significantly improves recommenda- tion precision under cold-start and sparse data conditions, outperforming KGAT, RippleNet, and KGIN on standard datasets including Book-Crossing, MovieLens-1M, and Last.FM.

I. LLM-as-a-Judge: Automated Evaluation of Search Query Parsing Using Large Language Models

1) Objective: To automate evaluation of structured search query parsers by replacing human evaluators with LLM- based reasoning agents that can evaluate semantic cor- rectness, attribute mapping accuracy, and contextual consistency of parsed queries.
2) Methodology: The framework introduces three evaluation modes: (1) Pointwise — evaluates each parsed output individually, (2) Pairwise — compares two candidate outputs and selects the better, (3) Pass/Fail — ensures correctness against rule-based validations. A novel Contextual Evaluation Prompt Routing (CEPR) mechanism dynamically selects evaluation prompts based on detected domain (e.g., automotive vs. real estate). The LLM is guided using structured evaluation rubrics that define correctness dimensions such as filter accuracy, loca- tion relevance, and semantic category matching. To reduce hallucinations, few-shot examples and rubric enforcement are incorporated.
3) Contribution**:** This is the first work to position LLMs as automated evaluators for structured NLP outputs. The system achieves 90% agreement with human experts, demonstrating that LLMs can reduce large-scale evaluation costs while maintaining semantic sensitivity.

J. CARE: Contextual Adaptation of Recommenders for LLM- Based Conversational Recommendation

1) Objective: The work aims to overcome the limitations of LLMs acting as standalone recommenders by introducing a hybrid system that fuses collaborative filtering predic- tions with conversational understanding and contextual reasoning.
2) Methodology: CARE introduces a two-step architecture: (i) a sequential recommender model predicts candidate items using learned embeddings from historical user in- teractions, (ii) an LLM performs context-aware reranking based on the ongoing conversation. Three contextual prompting strategies (Direct, Descriptive, Self-Reflective) and three re-ranking approaches (Expan- sion, Reranking, Selection-then-Reranking) are evaluated. The Selection-then-Reranking method improves relevance by combining user-item signals with conversation-derived semantic preferences.
3) Contribution: CARE demonstrates how structured rec- ommendation signals can be harmonized with natural language capabilities of LLMs, enabling accurate, person- alized, and explainable conversational recommendations while reducing LLM hallucination risks.

K. A Multimodal Recommender System Using Deep Learning Techniques Combining Review Texts and Images

1) Objective: To improve rating prediction accuracy in rec- ommender systems by modeling cross-modal relationships between visual product features and user-generated textual reviews.
2) Methodology: CAMRec consists of three major compo- nents: • a RoBERTa language encoder extracts contextual review semantics, • a VGG-16 backbone extracts visual features from images, and • a co-attention fusion layer jointly aligns image and text embeddings. The fused multi-modal representation is combined with user–item latent features inside an MLP network optimized with MSE loss and trained using Adam. Ablation studies validate that co-attention significantly improves learning of modality correlations.

3) Contribution: CAMRec bridges the gap between natural language understanding and visual perception in recom- mender systems. It also demonstrates that multimodal fusion is highly effective in scenarios where only limited textual or visual data are available.

L. *Web-Scale Semantic Product Search with Large Language Models*

1) Objective: To build a web-scale search mechanism that combines symbolic knowledge graph reasoning with the language understanding capabilities of LLMs, improving retrieval quality and interpretability.
2) Methodology: Product entities and relationships are extracted from text and used to build a structured knowl- edge graph (KG). KG embeddings are learned using TransE/RotatE, while textual descriptions are encoded using LLM embeddings (e.g., BERT, GPT). A contrastive alignment mechanism maps both representations into a unified semantic vector space. Retrieval scoring then con- siders both symbolic relationships and language semantics.
3) Contribution: The hybrid neuro-symbolic retrieval ap- proach overcomes cold-start limitations, improves contex- tual understanding, and produces explainable justifications— making product search more transparent and trustworthy.

M. *Large Language Models for Information Retrieval: Chal- lenges and Chances*

1) Objective: To evaluate whether LLMs can independently perform structured reasoning required in Knowledge Graph Question Answering (KGQA) and whether they can replace symbolic reasoning systems.
2) Methodology: Benchmarking is performed on multiple datasets (WebQuestionsSP, LC-QuAD, MetaQA) across reasoning categories: simple lookups, constraint queries, composite queries, multi-hop logic, and semantic ambi- guity. LLM prompting modes include zero-shot learning, few-shot learning, chain-of-thought prompting, and graph- aware prompting.
3) Contribution: The study concludes that LLMs excel in natural language interpretation but require structured KG grounding to remain logically consistent and minimize hallucination.

N. *A Unified Framework for Multimodal Large Language Models in Recommendation and Retrieval*

1) Objective: To unify text, image, and structured knowl- edge representations into a single multimodal reasoning framework for retrieval and recommendation tasks.
2) Methodology: The framework uses multimodal encoders (GPT-4V, BLIP-2, Flamingo) to process user queries and product content. Cross-modal co-attention modules align different modalities in a shared representation space. Downstream retrieval tasks are evaluated using Recall@K, NDCG, and human-rated explanation quality.
3) Contribution: Enables explainable multimodal reasoning where LLMs justify recommendations using textual evi- dence or visual cues — a key step toward human-aligned AI.

O. *Semantic Product Search for Matching Structured Product Catalogs in E-Commerce*

1) Objective: To improve structured product search by encoding multi-field product data and aligning it with semantic meaning of user queries.
2) Methodology: Queries and product fields (title, brand, metadata, etc.) are independently encoded using a Siamese DistilBERT architecture. A structured matching module (SMM) aggregates field embeddings and improves candi- date retrieval accuracy.
3) Contribution: Significantly boosts recall quality during candidate generation by capturing semantics across multi- ple attributes rather than relying solely on product titles.

P. *Learning Variant Product Relationship and Variation At- tributes from E-Commerce Website Structures*

1) Objective: To automatically identify product variants (same product with differing features such as size or color) and extract variation attributes without human annotation.
2) Methodology: VARM combines DistilBERT-based encod- ing with RAG-based attribute extraction to detect variant relationships and differences between variant groups. This reduces manual effort in catalog curation.
3) Contribution: Automates variant detection at scale, im- proving catalog consistency and providing richer product metadata.

Q. *Effective Product Schema Matching and Duplicate Detec- tion with Large Language Models*

1) Objective: To eliminate redundant entries across product catalogs and align attributes between schemas via LLM- driven semantic inference.

2) Methodology: Sentence embeddings are generated for product attributes, and LLM-based classification deter- mines duplicate items. The framework achieves human- level schema alignment (F1 = 78.1%) and even surpasses humans in duplicate detection (F2 = 90.2%).
3) Contribution: Reduces manual schema matching time by 90% and ensures catalog consistency across vendors.

R. LaTeX-Numeric: Language-Agnostic Text Attribute Extraction for E-Commerce Numeric Attributes

1) Objective: To extract and normalize numeric values (e.g., "5 kg", "1080p", "13 MP") across multilingual product descriptions without requiring manual annotation.
2) Methodology: The system uses distant supervision and a multitask learning architecture to learn numeric aliases (kg vs. kilograms). It generalizes across five product categories and multiple languages.
3) Contribution: Enhances numeric attribute extraction accuracy by 20.2% without any manual labeling.

S. AE-smnsMLC: Multi-Label Classification with Semantic Matching and Negative Label Sampling

1) Objective: To simplify attribute extraction by reformu- lating the task from sequence labeling to multi-label classification, reducing annotation complexity.
2) Methodology: Dual encoders learn joint embeddings for attributes and product text. Negative sampling ensures fine-grained attribute discrimination.
3) Contribution: Reduces annotation overhead and improves attribute prediction across datasets.

T. ExtractGPT: Exploring the Potential of LLMs for Product Attribute Value Extraction

1) Objective: To evaluate how GPT-4 and LLaMA-3 perform attribute extraction under zero-shot and few-shot settings.
2) Methodology: Structured prompt templates are used to guide attribute extraction. GPT-4 attains an F1 score of 85% and outperforms PLM-based baselines by 5%.
3) Contribution: Shows strong generalization ability, making LLMs attractive for industrial-scale automation.

U. LLM-Ensemble: Optimal Large Language Model Ensemble Method for Product Attribute Extraction

1) Objective: To improve extraction accuracy by combining predictions from multiple LLMs using a theoretically optimal ensemble strategy.
2) Methodology: A Dawid–Skene-based optimization algo- rithm assigns model weights based on reliability across attribute categories. Evaluations show strong improve- ments in business KPIs including GMV and conversion rate.
3) Contribution: Demonstrates that combining multiple LLMs produces better results than any single LLM, making it suitable for enterprise-level deployments.

III. RESEARCH GAPS

1) Existing studies on real-estate web scraping are con- strained by limited observation periods and geographic scope. Short data windows prevent modelling seasonal trends, and narrow regional coverage (mainly UK-centric) reduces generalizability. Additionally, the KNN-based matching strategy leaves a significant number of listings unmatched, limiting the completeness of analytics. Future work should increase multi-year coverage, extend scraping to international markets, and adopt ML-based probabilistic matching for higher linkage accuracy.[1]
2) Current scraping–NLP frameworks struggle with dynamic content, heterogeneous DOM structures, multilingual websites, and legal/ethical constraints. Large-scale noisy data leads to reduced accuracy in downstream analy- sis, especially during entity recognition and sentiment inference. More resilient pipelines leveraging transformer- based entity extraction, automatic selector generation, and ethical data governance are still underexplored.[2]
3) The revenue optimization case study is domain-specific, focusing only on wine marketplaces. Scraping speed is throttled due to anti-bot defenses, limiting scalability. Additionally, decision rules depend on user-defined mar- gin thresholds rather than predictive models. Future re- search should investigate proxy-based scraping, distributed crawlers, and learning-based dynamic pricing engines that adapt in real time.[3]
4) Product comparison scraping models rely heavily on manual selector creation for every new platform. They lack adaptability to layout changes and dynamic rendering. There is limited exploration of predictive analytics such as price forecasting or trend detection. Research opportu- nities exist in automated selector learning, multi-source fusion, and integrating recommender or

anomaly detection logic.[4]

5) Existing scholarly scraping and BERT-based literature rec- ommendation systems are constrained by limited database coverage, absence of multi-database aggregation, and shallow similarity modeling. They ignore citation context, knowledge graph relations, and time-evolving publication trends. Future studies may incorporate semantic caching, distributed scraping, and domain-adaptive model fine- tuning.[5]
6) In health research, current WeTMS workflows rely heavily on manual validation and assume uniform HTML struc- tures. Classification errors occur with ambiguous health activity descriptions, and spatial geocoding fails when metadata is incomplete. Machine learning-based automatic validation and triangulation across multiple data sources remain unexplored opportunities.[6]
7) In the retail domain, web scraping studies tend to be short-term and geographically biased (concentrated in Western markets). Limited data diversity, reliance on HTML scraping over structured APIs, and absence of longitudinal data pipelines restrict the generalizability of findings. Integrating web data with proprietary datasets and incorporating multimodal sources (images, reviews, social data) remains a key gap.[7]
8) Prior knowledge-aware recommenders struggle with sparse node interactions and single-view embeddings that fail to capture complementary semantics. Multimodal or multi- view learning for contrastive representation alignment is underexplored, leaving opportunities to incorporate visual, textual, and structural signals simultaneously.[8]
9) Existing query parsing evaluation depends on manual assessment and lacks semantic sensitivity. No standardized methodology exists for automatic, rubric-driven, domain- adaptive evaluation using LLMs. Systems do not quantify confidence or uncertainty in evaluation.[9]
10) Conversational recommendation approaches do not fully align LLM-generated reasoning with item-level embed- dings from domain recommenders. The interactions are shallow, and contextual refinement is not feedback-driven. There is a gap in developing hybrid systems that combine recommendation scores and dialogue reasoning in a closed- loop architecture.[10]
11) Prior multimodal recommender frameworks treat modal- ities independently or perform shallow concatenation instead of deep semantic alignment. There is limited exploration of co-attention or cross-modal dependency modeling, resulting in suboptimal personalization and higher bias toward dominant modalities.[11]
12) KG–LLM hybrid reasoning remains underdeveloped. Ear- lier works either rely solely on symbolic reasoning or LLM inference. Unified benchmarks and neuro-symbolic modeling remain open challenges, particularly for cold- start scenarios and interpretability.[12]
13) MLLM-based recommender systems lack scalable mul- timodal alignment mechanisms that support structured reasoning across text, images, and graph relational data. Retrieval justification and explainability (in natural lan- guage) remain limited.[13]
14) Structured semantic matching approaches treat all product attributes equally without explaining which fields con- tribute most to model decisions. They lack interpretability and generalization across domains with varying attribute completeness.[14]
15) Variant relationship extraction still depends on webpage structure cues and fails when structure varies across marketplaces. There is insufficient use of multimodal cues (images or metadata consistency) to identify variants at scale.[15]
16) Schema matching work has high computational overhead when applied to dynamic catalogs and lacks cross-lingual support. Decisions are generated without a transparent explanation of attribute alignment.[16]
17) Numeric extraction systems focus only on numeric values and unit normalization, ignoring contextual logic (e.g., differentiating "5 kg package weight" vs. "5 kg lifting capacity"). Better multimodal verification using images or packaging text remains unexplored.[17]
18) Multi-label attribute extraction formulations assume bal- anced datasets and fail to scale when attribute space is high-dimensional or inconsistent across categories. Handling noisy or overlapping labels is still a gap.[18]
19) LLM-based attribute extraction models are highly prompt- sensitive and may not generalize across unseen product types. Evaluation lacks robustness testing on multilingual and cross-market datasets.[19]
20) Ensemble-based LLM attribute extraction introduces model aggregation but lacks fairness analysis, interpretabil- ity in ensemble weighting, and deployment cost evaluation. Self-learning feedback loops are missing for continuous improvement.[20]

IV. FUTURE WORK

1) Future work in real-estate web scraping should focus on extending data collection beyond short-term pandemic periods to enable long-term modelling of seasonal pricing patterns and economic cycles. Incorporating regional diversity (e.g., Scotland, Wales, Northern Ireland, Euro- pean and Asian markets) would enhance generalizability. Improving the listing-to-transaction

matching algorithm using deep learning–based similarity matching or geo- spatial clustering could significantly increase match ratios. Integrating demand-side indicators (Google Trends, mortgage loan approvals, consumer sentiment data) could enable predictive "now-casting" models that estimate housing-market shifts in advance.[1]

2) To improve scraping–NLP pipelines, future research should explore automated DOM adaptation mechanisms using ML-based selector generation, enabling scrapers to self-adjust when website structures change. Improved entity extraction can be achieved using transformer-based multilingual NER and domain-adaptive sentiment models. Additionally, incorporating legal-aware scraping modules that continuously monitor robots.txt, GDPR flags, and rate-limit policies would establish ethical, compliant large- scale data collection.[2]
3) Future enhancements to pricing intelligence systems should include integration of distributed scraping architectures (proxy rotation, stealth browser automation, serverless crawling) to circumvent scraping throttling mechanisms. Forecasting models such as reinforcement learning, ARIMA, or demand elasticity estimation can be used to dynamically optimize product pricing based on sales performance, seasonal behavior, and competitor fluctuations. Integration into real-time dashboards with alerts would improve managerial decision-making.[3]
4) Product comparison systems can be extended with auto- mated marketplace discovery using crawling-based search to detect new e-commerce sources without manual config- uration. A recommendation module can be integrated to suggest best-value or sustainable alternatives. Furthermore, real-time price trend prediction could be introduced using time-series models to notify users when a product is expected to drop in price.[4]
5) Literature recommendation systems can evolve from static scraping to intelligent multi-database aggregation integrat- ing platforms such as IEEE Xplore, Scopus, PubMed, and Semantic Scholar. Incorporating citation network analysis and knowledge graph reasoning could improve semantic ranking beyond pure embedding similarity. Distributed processing and caching mechanisms would allow support for heavy query loads and continuous monitoring of new publications.[5]
6) For contextual variable generation in public health, machine-learning-driven validation and annotation could replace manual verification, increasing scalability. In- tegration of satellite imagery and social media data could capture location-based contextual variables more accurately. Automating geocoding with multi-source en- richment (OpenStreetMap, Google APIs) would increase spatial accuracy and real-time update potential.[6]
7) For retail analytics, future research should explore lon- gitudinal data aggregation pipelines that monitor price volatility over weeks or months. Combining scraped web data with transactional customer data can improve demand prediction models. LLM-assisted scraping agents could automate extraction logic and adapt dynamically to different website structures.[7]
8) MCCLK can be extended to incorporate temporal knowl- edge graphs, where embeddings evolve as new data arrives. Interpretable contrastive learning modules could be devel- oped to highlight why certain items were recommended. A multi-modal version of MCCLK could combine image or review text embeddings with KG entities, improving understanding of item characteristics.[8]
9) LLM-as-a-Judge could evolve into a hybrid human–LLM evaluation model, where uncertain cases are escalated to human reviewers. Introducing calibrated confidence scoring and bias/fairness auditing would enhance its reliability and transparency in production environments.[9]
10) CARE can be extended to support end-to-end differen- tiable co-training, where recommender models and LLM reasoning continuously improve each other. Conversational feedback loops could be incorporated, enabling the system to learn user preferences dynamically and update latent embeddings in real-time.[10]
11) For multimodal recommenders like CAMRec, future work may incorporate transformer-based visual encoders (e.g., ViT, CLIP) and sentiment-aware review fusion. Supporting multi-modal user feedback (speech, images, interaction logs) could improve personalization and reduce cold-start effects.[11]
12) Hybrid KG–LLM systems should be extended toward neuro-symbolic co-training frameworks where LLMs generate graph expansions while graphs constrain rea- soning. Explainability modules should convert reasoning pathways into natural-language explanations to improve user trust.[12]
13) MLLM frameworks should incorporate efficient multi- modal fine-tuning (LoRA, adapter layers) for domain adap- tation and introduce fairness-aware ranking to prevent bias toward visually appealing products. Real-time justification generation would increase user confidence.[13]
14) Structured semantic matching models should incorporate interpretability layers that visualize which product fields strongly influence similarity decisions. Cross-domain experiments are needed to generalize across multiple product categories with varying metadata quality.[14]

15) Variant extraction systems should leverage visual similarity modeling to detect product variants when text metadata is insufficient. A contrastive learning backbone can improve robustness across brands and marketplaces.[15]
16) Schema-matching frameworks should support multilingual alignment, low-resource schema learning, and continual learning when catalogs change. Integrating explainability dashboards can allow auditors to review attribute-mapping decisions.[16]
17) Numeric attribute extraction should evolve toward multi- modal verification, combining text extraction with com- puter vision (e.g., detecting numeric labels on product packaging). Context-aware disambiguation should distin- guish between similar numeric attributes (e.g., product volume vs. shipping weight).[17]
18) Future work for multi-label attribute extraction should incorporate contrastive pretraining, hierarchical label organization, and dynamic label space expansion to accommodate evolving attribute taxonomies.[18]
19) For ExtractGPT-style attribute extraction, improvements could be made using hybrid retrieval–generation pipelines, prompt optimization through reinforcement learning, and cross-lingual generalization for global e-commerce plat- forms.[19]
20) For LLM ensembles, future work could incorporate reinforcement learning with human feedback (RLHF) and cost-aware ensemble optimization to choose the cheapest model that meets accuracy targets. Continuous improvement through active learning loops remains an unexplored opportunity.[20]

V. COMPARATIVE ANALYSIS AND DISCUSSION

A comprehensive comparative evaluation of the reviewed literature highlights a clear evolution of methodologies across three major research dimensions: (i) intelligent data acquisition via web scraping and NLP, (ii) deep learning and Large Language Model (LLM)-based recommendation architectures, and (iii) multimodal product reasoning and conversational interaction.

The first group of studies [1]-[7] predominantly focuses on data acquisition and preprocessing. These works demonstrate how modern scraping frameworks and automated extraction pipelines support large-scale data collection from heterogeneous sources. Approaches vary from rule-driven scraping to dynamic browser automation using Selenium, and semi-structured extraction using NLP-based parsing. Although valuable for enabling access to real-time product or market information, these methods often depend heavily on platform-specific DOM structures and lack adaptability to layout changes or anti-bot mechanisms. Additionally, most studies are domain-constrained (e.g., wine pricing, housing prices, academic papers), limiting general applicability across e-commerce ecosystems.

The second cluster of research [8]-[13] transitions toward deep learning–driven representation learning and knowledge- aware recommendation. Models such as BERT-based encoders, co-attention multimodal networks (CAMRec), and MCCLK apply self-supervision, graph learning, or cross-view contrastive learning to improve personalization accuracy. These approaches outperform traditional collaborative filtering by incorporating semantic relationships, contextual signals, and user intent. However, technical challenges remain: high computational complexity, lack of interpretability, reliance on dense interaction histories, and limited responsiveness to real-time conversational inputs. They improve recommendation accuracy but still struggle to justify their predictions in a human-understandable manner.

The third and most recent trend [14]-[20] introduces mul- timodal LLM-driven frameworks capable of reasoning over heterogeneous signals including product images, knowledge graphs, and conversation states. These approaches leverage advanced architectures such as GPT-4V, Flamingo, and RAG- enhanced prompting to perform semantic alignment between text, images, and metadata. Unlike earlier systems, LLM-based models introduce transparency by generating natural-language justifications for the recommendations. However, limitations still persist—LLM inference latency, hallucination risks, depen- dence on external retrieval for factual grounding, and challenges in optimizing responses under real-time interactive settings.

Overall, the comparative synthesis reveals a gradual paradigm shift:

Rule-based Extraction → Deep Representation Learning → LLM-driven Multimodal Reasoning

Web scraping methods excel at obtaining extensive product information but lack contextual understanding. Deep-learning models understand preference dynamics, yet they are computationally expensive and non-interactive. Multimodal LLM- based systems support interactive, transparent, and personalized recommendations, but require careful optimization to balance accuracy, latency, cost, and sustainability.

These observations emphasize the need for a unified system one capable of:
1) retrieving factual product data in real time,
2) reasoning across multimodal information (text, image, and metadata),
3) enabling interactive decision support through conversa- tional LLM interfaces, and
4) providing transparency, explainability, and bias-aware response generation.

This motivates the reviewed Conversational LLM-based Multi-Modal Product Research Assistant, which bridges the identified gaps by integrating Retrieval-Augmented Generation (RAG) with multimodal embeddings. This system positions itself as a next-generation assistant that transforms static recommendation pipelines into adaptive, human-centric, and sustainable decision-support systems.

VI. CONCLUSION

This survey presented an extensive comparative analysis of research spanning web data acquisition, deep learning–driven recommender systems, and emerging Large Language Model (LLM)-based multimodal reasoning frameworks. The findings reveal an evident evolution in the design of intelligent product exploration and recommendation systems. Traditional rule- based scraping and filtering methods provided access to product information but lacked contextual awareness, adaptability to platform changes, and interpretability. Deep-learning ap- proaches, particularly those leveraging representation learning, knowledge graphs, and multimodal fusion, enabled stronger personalization and semantic modeling but often suffered from computational overhead, sparse supervision, and limited transparency in decision logic.

More recent advancements employing LLMs and multimodal architectures demonstrate significant progress toward conversa- tional and explainable recommendation ecosystems. Techniques such as Retrieval-Augmented Generation (RAG), cross-modal embedding alignment, and LLM-based re-ranking enable systems to reason jointly over structured and unstructured sources, integrating textual reviews, visual content, and product metadata. However, these systems still face challenges involving hallucination control, inference latency, domain adaptation, and the need for efficient bias-aware optimization to ensure trustworthiness and fairness.

The comparative insights from this survey highlight a growing need for a unified architecture that combines:

1) scalable and ethical data acquisition,
2) multimodal representation learning and cross-domain alignment,
3) contextual and interactive dialogue-based reasoning through LLMs, and
4) explainable and unbiased recommendation outputs that align with sustainable decision-making.

In response to these identified gaps, the reviewed Conversational LLM-Based Multi-Modal Product Research Assistant advances the field by integrating real-time retrieval, multimodal analysis, and natural-language interaction. The system aims to transform product discovery from static list-based recommendations into a transparent, interpretable, and user-driven exploration process. Furthermore, its explicit alignment with UN SDG 12 — Responsible Consumption and Production — positions it as a socially responsible framework that enables consumers to make informed and sustainable purchasing decisions.

Going forward, future enhancements will include optimizing inference efficiency, reducing carbon footprint of LLM deployment, improving fairness and bias detection in attribute extraction, and incorporating domain-adaptive multimodal fine- tuning. This work builds the foundation for next-generation recommendation systems that are not only intelligent and context- aware, but also human-centric, trustworthy, and sustainable.

REFERENCES

[1] J.-C. Bricongne, B. Meunier, and S. Pouget, "Web- scraping housing prices in real-time: The Covid-19 crisis in the UK," Journal of Housing Economics, 2023
[2] V. Pichiyan, S. Muthulingam, G. Sathar, et al., "Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis," Procedia Computer Science, vol. 230, pp. 193–202, 2023.
[3] O. Jorge, A. Pons, J. Rius, C. Vintró, J. Mateo, and J. Vilaplana, "Increasing online shop revenues with web scraping: A case study for the wine sector," British Food Journal, vol. 122, no. 11, pp. 3383–3401, 2020.
[4] M. D'Souza, S. Desai, D. Agrawl, F. Joshi, et al., "Web scraping based product comparison model for e- commerce website," Journal of Emerging Technologies and Innovative Research, vol. 11, no. 4, 2024.
[5] P. Bandi, "Advanced Google Scholar Scraper: A content- based filtering approach for literature recommendation using BERT," M.Sc. research project, National College of Ireland, 2023
[6] P. Galvez-Hernandez, A. Gonzalez-Viana, L. Gonzalez- de Paz, et al., "Generating contextual variables from web-based data for health research," JMIR Public Health Surveillance, vol. 10, 2024.
[7] J. Y. Guyt, H. Datta, and J. Boegershausen, "Unlocking the potential of web data for retailing research," Journal of Retailing, vol. 100, pp. 130–147, 2024. for LLM-Based Conversational Recommendation," arXiv preprint arXiv:2508.13889, 2025.
[8] J.-K. Kim et al., "A Knowledge Graph and Large Language Model-Based Hybrid Recommender System," in Advances in Intelligent Systems and Computing, Springer, 2023.
[9] Y. Wang, Z. Chu, X. Ouyang, et al., "Enhancing Recom- mender Systems with Large Language Model Reasoning Graphs," arXiv preprint arXiv:2308.10835, 2023.
[10] X. Ren, W. Wei, L. Xia, et al., "Representation Learning with Large Language Models for Recommendation," arXiv preprint arXiv:2310.15950, 2023.
[11] A. Nigam, P. Jain, and A. Gupta, "Semantic Product Search for Matching Structured Product Catalogs in E- Commerce," arXiv preprint arXiv:2008.08180, 2020.

[12] Y. Zhang, X. Liu, and C. Yu, "Learning Variant Product Relationship and Variation Attributes from E-Commerce Website Structures," in Proc. ACM Web Conference (WWW), 2022.

[13] H. Chen, M. Zhang, and D. Song, "Effective Product Schema Matching and Duplicate Detection with Large Language Models," arXiv preprint arXiv:2104.09576, 2021

[14] X. Wang, L. Zhao, and R. Li, "AE-smnsMLC: Multi- Label Classification with Semantic Matching and Nega- tive Label Sampling for Product Attribute Value Extrac- tion," arXiv preprint arXiv:2310.07137, 2023.

[15] A. Brinkmann, R. Shraga, and C. Bizer, "ExtractGPT: Exploring the Potential of Large Language Models for Product Attribute Value Extraction," arXiv preprint arXiv:2310.12537, 2024.

[16] C. Fang, X. Li, Z. Fan, et al., "LLM-Ensemble: Optimal Large Language Model Ensemble Method for Product Attribute Value Extraction," in Proc. 47th Int. ACM SIGIR Conference, 2024.

[17] Z. He, M. Huang, and Q. Lin, "LaTeX-Numeric: Language-Agnostic Text Attribute Extraction for E-Commerce Numeric Attributes," arXiv preprint arXiv:2403.00863, 2024.

[18] M. S. Baysan, S. Uysal, ˙I. ˙Is¸lek, Ç . Ç ıg˘ Karaman, and T. Gu¨ngo¨r, "LLM-as-a-Judge: Automated evaluation of search query parsing using large language models," Frontiers in Big Data, vol. 8, 2025.

[19] E. Jeong, X. Li, A. (Eunyoung) Kwon, S. Park, Q. Li, and J. Kim, "A Multimodal Recommender System Using Deep Learning Techniques Combining Review Texts and Images," Applied Sciences, vol. 14, no. 20, 2024.

[20] C. Li, Y. Deng, H. Hu, S.-K. Ng, M.-Y. Kan, and H.Li, "CARE: Contextual Adaptation of Recommenders