E-ISSN: 2964-6219, P-ISSN: 2964-6200
Volume 4 Number 2 August 2026.
pp: 320-327. DOI: https://doi.org/10.58477/cj.v4i2.510.

Computer Journal

RESEARCH ARTICLE Open Access

Absent or Unretrievable: A Two-Layer Knowledge Management Framework for Retrieval-Critical Product Knowledge in a Grocery E-Commerce AI Shopping Assistant

Sinatrio Bimo Wahyudi [1*], **Dana Indra Sensuse** [2], **Sofian Lusa** [3], **Muhammad Hafizh Qurani** [4], **Yordan Yasin** [5], **Icha Mailinda** [6]

[1*,2,3,4,5,6] Department of Computer Science, Universitas Indonesia, Central Jakarta City, Special Capital Region of Jakarta, Indonesia.

*Correspondence email: sinatrio.bimo@ui.ac.id

Received: 25 July 2026.
Revised: 9 August 2026.
Accepted: 26 August 2026.
Published: 30 August 2026.

Full list of author information is available at the end of the article.

Abstract

AI shopping assistants increasingly employ agent-based retrieval, combining lexical search, structured filtering, and LLM-mediated selection. However, effective retrieval often requires knowledge beyond textual matching. This exploratory single-case study of an Indonesian grocery e-commerce assistant triangulates 279 observations, 52 failure traces, 74 practitioner-reported defects, a schema audit, and four practitioner interviews. The analysis identifies six retrieval-critical knowledge dimensions and reveals that externalization fails at two distinct layers. At the product data-model layer, essential fields (allergens, dietary constraints, age suitability) were absent and persisted despite architectural changes. At the retrieval-schema layer, existing knowledge failed to reach candidates: structured filters appeared in only 29% of calls, and correctly invoked filters often returned empty sets from non-empty pools. Addressing these layer-specific failures, the study proposes a structured knowledge framework based on the knowledge management process cycle, positioning GraphRAG as a future direction.

Keywords: AI Shopping Assistant; Agent-Based Retrieval; Knowledge Management; Knowledge Externalization; Product Knowledge Graph; Taxonomy.

Abstrak

Asisten belanja AI kian mengadopsi *agent-based retrieval* yang menggabungkan pencarian leksikal, penyaringan terstruktur, dan seleksi berbasis LLM. Namun, *retrieval* yang efektif memerlukan pengetahuan melampaui pencocokan teks. Studi kasus tunggal pada *e-commerce* bahan makanan di Indonesia ini melakukan triangulasi terhadap 279 observasi, 52 *failure traces*, 74 laporan defek, audit skema, dan wawancara praktisi. Hasilnya mengidentifikasi enam dimensi pengetahuan kritis dan menunjukkan kegagalan eksternalisasi pada dua lapisan. Pada lapisan model data produk, informasi alergen, batasan diet, dan kesesuaian usia tidak tersedia dalam skema. Pada lapisan skema *retrieval*, pengetahuan yang tersedia gagal dimanfaatkan: filter terstruktur hanya muncul pada 29% panggilan, dan filter yang tepat sering menghasilkan hasil kosong. Penelitian ini mengusulkan kerangka pengetahuan terstruktur berbasis siklus proses manajemen pengetahuan untuk mengatasi kegagalan tersebut, dengan menempatkan GraphRAG sebagai arah pengembangan masa depan.

Kata Kunci: Asisten Belanja AI; Agent-Based Retrieval; Manajemen Pengetahuan; Eksternalisasi Pengetahuan; Product Knowledge Graph; Taksonomi.

1. Introduction

Retrieval-Augmented Generation (RAG) enables Large Language Models (LLMs) to generate responses by retrieving external knowledge from non-parametric memory (Lewis *et al.*, 2020). In e-commerce AI shopping assistants, this approach retrieves product-related information such as catalog data, descriptions, FAQs, and reviews. While vector-based retrieval is often assumed as the default implementation (Klesel & Wittmann, 2025), production-grade assistants in practice frequently combine lexical search, structured parameter filtering, and LLM-mediated candidate selection, with embedding components at varying stages of maturity. However, product retrieval in conversational e-commerce requires knowledge that transcends simple textual matching. Grocery retrieval is a particularly challenging setting where catalog text serves as a weak proxy for actual shopper needs (Xiao *et al.*, 2021). Work on large-scale platforms indicates that short grocery queries implicitly encode dietary preferences that similarity-based matching fails to recover (Magnani *et al.*, 2022). In the studied setting, user needs involve dietary restrictions, allergy concerns, age suitability, product availability, alternatives, and contextual use. These requirements correspond to the concept-level and attribute-level structures that systems like AliCoCo and AutoKnow were designed to represent (Luo *et al.*, 2020; Dong *et al.*, 2020). When such knowledge is absent, stored only as free text, or unavailable during retrieval, the system may return products that are textually related but functionally unsuitable. This challenge is further amplified in the Indonesian e-commerce landscape, where linguistic nuances, informal product naming, and specific dietary habits create a high barrier for standard retrieval systems. A first-order reading might attribute these failures to model capability or prompt formulation. However, the empirical corpus does not support this interpretation. Among observations evaluated in two successive rounds separated by prompt and configuration changes, 10 of 12 first-round failures persisted in the second round; in a curated subset enriched for failure cases, 39 of 41 persisted. This suggests that modifications to the model's instructions did not address the root cause, which remains relevant as conversational systems must handle natural, incomplete, and context-dependent requests. Prior work confirms that imperfect product schemas and knowledge representation remain significant challenges (Xiao *et al.*, 2021). Therefore, retrieval failures should be understood not merely as query ambiguity, but as fundamental product knowledge representation problems. From a Knowledge Management (KM) perspective, knowledge becomes useful to a system only after it has been externalized from tacit holders and moved through a sequence of processes. Externalization involves converting tacit knowledge into explicit, articulable forms (Nonaka & Takeuchi, 1995). Alavi and Leidner (2001) frame KM as four interdependent processes—creation, storage/retrieval, transfer, and application—distinguishing between knowledge as an object and knowledge as access to information. While AI can support these processes, it requires underlying knowledge to be captured in a usable form (Jarrahi *et al.*, 2023). In a grocery AI assistant, critical knowledge regarding allergen rules, substitution logic, and category boundaries is often held tacitly by category and operations teams or buried in free text. In KM terms, this knowledge has not been externalized into an object the retrieval stage can access, preventing its application during candidate selection.

This failure occurs at two separable representational layers with different remedies and organizational owners. While large platforms have converged on structured product knowledge—such as Alibaba's AliCoCo (Luo *et al.*, 2020) and Amazon's AutoKnow (Dong *et al.*, 2020)—there is a scarcity of research focusing on the diagnostic journey—identifying the specific layers where knowledge externalization breaks down in a production environment. Most studies assume a target state of structured knowledge, overlooking the intermediate failures that occur at the data-model and retrieval-schema layers. This study addresses this gap by examining which knowledge types are affected at each representational layer and the implications for organizational remedies, guided by two research questions:

RQ1: What types of product knowledge are critical for effective retrieval but insufficiently represented at either the product data-model layer or the retrieval-schema layer in a grocery e-commerce AI shopping assistant? RQ2: How can retrieval-critical product knowledge be externalized and structured into a knowledge framework so that it can act at the point of candidate formation in e-commerce AI shopping assistants? This study adopts an exploratory qualitative single-case study approach with taxonomy development. Retrieval-related failure cases are analyzed to identify recurring knowledge gaps interpreted through a KM lens. Rather than evaluating quantitative performance, this study develops a taxonomy and proposes a structured knowledge framework to organize product knowledge as a basis for improving retrieval effectiveness.

2. Method

This study adopts an exploratory qualitative single-case study approach combined with taxonomy development. It examines a grocery e-commerce AI shopping assistant that utilizes agent-based retrieval for conversational product discovery at a leading Indonesian e-grocery platform. The retrieval architecture in force during the observation period was characterized directly from execution traces and version-controlled source code, as presented in the subsequent architectural analysis. This section outlines the research design, data collection procedures, and the reflexive thematic analysis used to produce the taxonomy and the proposed knowledge framework.

2.1 Research Design

The unit of analysis consists of retrieval-related failure cases within the studied assistant. A single-case qualitative design was selected because the specific types of product knowledge critical for retrieval are novel and not yet conceptualized in prior literature. Case-based research is particularly suited for abductive theory building, allowing for iteration between empirical observation and conceptual framing (Eisenhardt, 1989; Eisenhardt & Graebner, 2007). Taxonomy development is integrated as the primary analytical outcome. The study follows the empirical-to-conceptual procedure established by Nickerson *et al.* (2013), where characteristics and categories are derived from observed objects—in this case, failure cases—and refined iteratively against ending conditions. This ensures the resulting six dimensions are methodologically defensible and grounded in operational reality.

2.2 Data Collection and Selection

The primary data consist of curated test cases from an evaluation master sheet, recording user queries, expected retrieval behaviors, actual results, observed failures, and evaluation notes. The master sheet contains 279 test-case observations across 254 scenario-based codes. The unit of observation is a single query–expectation–outcome triple. Applying exclusion criteria removed 25 observations related to greeting behavior, tone, or out-of-scope refusals. Of the remaining 254, 60 were recorded as failures (forming the primary corpus) and 194 were recorded as passes (retained as a contrast set). Execution traces were available for 52 of the 60 failures. This purposive selection isolates cases that reveal product knowledge gaps rather than unrelated defects. Seven supporting evidence types were used to triangulate the primary corpus, as summarized in Table 1.

Table 1. Data Sources and Roles

Data Source	Role	Related RQ
Evaluation master sheet	Primary corpus for retrieval-related failure cases	RQ1
Execution traces (52 failures)	Tool invocations, search paths, candidate counts	RQ1
Source-code and schema audit	Field availability, indexing, filter usage	RQ1, RQ2
Version-control history	Dating of architectural components to builds	RQ1
Practitioner defect log	Independently generated failure evidence	RQ1
Category coverage test	Best-seller retrieval coverage across categories	RQ1
Semi-structured interviews	Practitioner interpretation and knowledge ownership	RQ1, RQ2

To strengthen the qualitative findings, semi-structured interviews were conducted with four key practitioners: a Lead Product Manager, a Senior Backend Engineer, a Search Engineer, and a Category Operations Specialist. This ensured that both technical constraints and business logic were represented in the analysis.

2.3 Data Analysis

Failure cases were analyzed using Braun and Clarke's (2006) six-phase reflexive thematic analysis. This method was chosen for its flexibility in identifying latent patterns of meaning without a pre-existing coding frame. The analysis was operationalized into five concrete stages (Table 2).

Table 2. Coding and Framework Development Procedure

Stage	Focus	Output
Initial Coding	Identify observed retrieval failures from query-outcome mismatches.	Initial failure code
Knowledge Gap Inference	Determine gap mode from trace fields (parameters, pool size) using fixed decision rules.	Product knowledge gap

Category Grouping	Group similar gaps across cases into candidate categories.	Preliminary taxonomy category	
Evidence Checking	Refine categories against traces, schema evidence, and stakeholder notes.	Refined taxonomy category	
Framework Construction	Organize final categories into the KM process cycle (storage, retrieval, application).	Structured knowledge framework	

Stage 2 was strengthened for the 52 traced failures by applying a fixed decision rule: determining whether structured filters were sent, whether candidate pools were empty, or whether filters yielded no results from non-empty pools. To ensure methodological rigor, the study employed data triangulation by cross-referencing failure cases with execution traces and practitioner insights. Furthermore, member checking was performed by presenting the final taxonomy to the interviewed practitioners to verify that the dimensions accurately reflected the system's structural gaps. Preliminary evidence suggests these failures are not attributable to prompt formulation alone. In a curated subset of 50 observations, 39 of 41 first-round failures persisted despite prompt and configuration changes between evaluation rounds, underscoring the structural nature of the identified knowledge representation issues.

3. Results and Discussion

3.1 Results

The findings were derived from qualitative coding of retrieval-related failure cases in a single e-commerce AI shopping assistant. From 279 test-case observations, 60 retrieval-relevant failures form the primary corpus, and 194 passes were retained as a contrast set; execution traces support 52 of the 60 failures. The analysis moved from observed failure to knowledge gap, taxonomy category, and structured knowledge framework. Verified Retrieval Architecture: Because the interpretation of every failure depends on the mechanism that produced it, the retrieval architecture was verified directly. Version-control history identifies the build that produced the observations. In it, an LLM agent invoked a product-search tool querying an Elasticsearch-backed inventory service over two search paths: keyword and category. Across 49 recorded debug blocks, only these two paths appear; no embedding or vector path was active in the product-retrieval traces. Consequently, the phrase "textually related but functionally unsuitable" describes the mechanism literally: failures occurred because semantic similarity was not the operative mechanism, establishing the baseline for the subsequent analysis. Retrieval Failure Patterns and Coding Results: The findings were derived by applying a five-stage procedure to the 60 primary failure cases, operationalizing Braun and Clarke's reflexive thematic analysis. Coding showed that failures were recurring patterns: the system retrieved textually related products yet could not apply product knowledge needed for eligibility, substitution, attribute-based ranking, category enforcement, user-language mapping, or contextual use. For example, a request for peanut-allergy-safe jam resulted in zero products because the system lacked a fallback to relax search terms while preserving hard constraints. Across traced cases, structured knowledge was applied in only 12 of 41 parsed search invocations (29%). Table 3 summarizes the abstraction for all six dimensions.

Table 3. Abstraction Ladder from Failure Codes to Taxonomy Dimensions

Representative Case	First-order code	Inferred knowledge gap	Dimension	Gap layer
Peanut-allergy jam returns zero results despite correct exclusion.	Constraint correctly applied to query rewritten beyond coverage.	No fallback preserving hard constraints under recall collapse.	T1. Constraint and Eligibility	L1 + Execution
OOS herb returns no alternative despite existing links.	OOS item not resolved to relevant alternative.	Substitution not modeled as product-to-product relations.	T2. Substitution and Alternative	L2
Best-selling coffee returns no results with filter applied.	Filter correctly invoked; attribute not populated for candidates.	Attributes are unevenly queryable or not sortable.	T3. Attribute and Specification	L2

Cheapest groceries return party popper and cleaning kit.	Non-grocery items cross category boundary.	Category boundaries and concept mapping not enforced.	T4. Category and Concept Mapping	L2
Tennis ball matched to rice-flour snack "bola".	Informal phrase matched by substring rather than concept.	No alias-mapping layer between language and catalog.	T5. User-Language-to-Product-Concept	L1
Breakfast request under budget not resolved to product set.	Situational query not decomposed into product set.	Use-case and basket context not represented in schema.	T6. Use-Case and Contextual	L1

Taxonomy of Retrieval-Critical Product Knowledge: Following the inductive logic of Gioia *et al*. (2013), dimensions were built upward from the data and validated against practitioner interviews and schema audits. Six dimensions resulted:
T1. Constraint and Eligibility: Decides admissibility under health, age, or safety constraints.
T2. Substitution and Alternative: Provides valid alternatives when products are unavailable.
T3. Attribute and Specification: Enables filtering and ranking by specifications (e.g., price, popularity).
T4. Category and Concept Mapping: Enforces category boundaries to prevent textual override.
T5. User-Language-to-Product-Concept: Maps slang and functional phrases to catalog concepts.
T6. Use-Case and Contextual (Emergent): Decomposes situational use cases and maintains context.
The taxonomy structure is visualized in the thematic code map (Figure 1) and the knowledge hierarchy chart (Figure 2). Dimension assignment for a 35-case subset showed T3, T1, and T4 as strongly grounded, while T2, T5, and T6 are supported by converging evidence from practitioner defect logs and interviews.

Figure 1. Thematic code map of product knowledge dimensions.

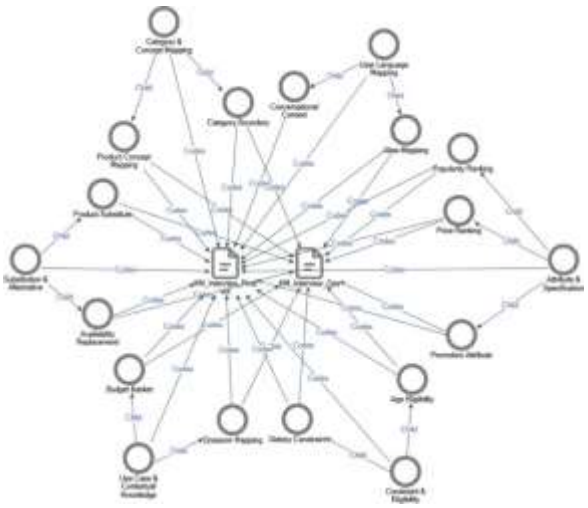

Figure 2. Knowledge hierarchy chart showing internal structure of T1-T6.

The map illustrates the empirical grounding of the taxonomy. Each child code (representing specific product knowledge) is linked to the practitioner interview sessions by "Codes" relations and to its respective parent dimension (T1–T6) by "Child" relations. ini membuktikan bahwa setiap kategori bukan sekadar label teoretis, melainkan berakar pada data wawancara praktisi. This hierarchy visualizes the internal depth and relative significance of each dimension. The block area for each dimension reflects its coding volume within the corpus, showing that dimensions like Attribute (T3) and Constraint (T1) have more complex sub-structures compared to emergent dimensions like Use-Case (T6).

3.2 Discussion

Applying the six dimensions to the contrast set of successful retrievals confirms they are not an artifact of failure; both outcomes depend on how knowledge is represented. The lexical path succeeds only when catalog text aligns with user needs but fails consistently for constraint and substitution requirements that are not expressible through the existing parameter surface. In knowledge management terms,

externalization has failed for these dimensions. Proposed Knowledge Framework for Retrieval Improvement: To answer RQ2, the taxonomy was translated into a structured knowledge framework (Figure 3) Organized along the KM process cycle of Alavi and Leidner (2001), the framework positions retrieval improvement as a knowledge management problem rather than a prompt-engineering issue:

1) Knowledge Creation (Stages 1-2): Tacit knowledge from practitioners is externalized into explicit forms.
2) Storage and Retrieval (Stage 3): Encodes semantics into a Product Knowledge Graph using typed relations (e.g., *SUITABLE_FOR*, *CAN_SUBSTITUTE*).
3) Transfer and Application (Stages 4-5): A hybrid retrieval core activates parallel paths—lexical recall and structured graph retrieval—converging at candidate ranking.
4) Feedback (Stage 6): Closes the cycle by returning usage evidence to creation.

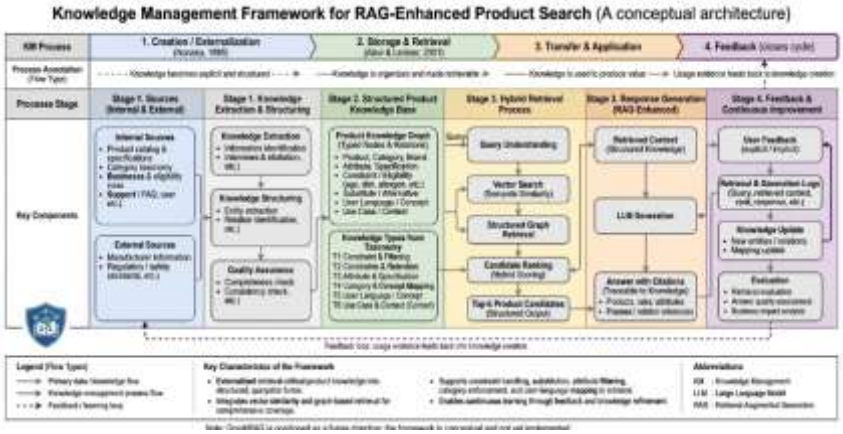

Figure 3. Proposed structured knowledge management framework for RAG-enhanced product search.

The framework augments rather than replaces the existing path. Where similarity fails (e.g., eligibility gating), the structured layer supplies the missing capability during candidate formation. This contrastive analysis explains why a structured representation, traversed by GraphRAG in combination with vector retrieval, is the appropriate response to the identified failures.

4. Conclusion

This study investigated the underlying causes of retrieval failures in e-commerce AI shopping assistants—specifically why systems retrieve textually related yet functionally unsuitable products. Answering RQ1, the analysis produced a taxonomy of six retrieval-critical product knowledge dimensions. The central finding is that retrieval improvement in conversational commerce is primarily a knowledge management (KM) problem. Most failures occurred at the candidate formation stage because the required knowledge was either absent from the data model (L1 failure) or present but un-retrievable due to schema limitations (L2 failure). Answering RQ2, the study proposed a structured knowledge framework (Figure 3) that externalizes product knowledge into rules, relations, and mappings. The framework introduces a hybrid approach where structured knowledge supplies the capabilities that lexical similarity alone cannot express. While the study is bounded by its exploratory single-case scope, it identifies that different failure layers require different organizational remedies. Future work should focus on implementing the hybrid GraphRAG framework and evaluating the governance needed to keep externalized knowledge accurate over time.

References

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly, 25*(1), 107–136. https://doi.org/10.2307/3250961

Balakrishnan, J., & Dwivedi, Y. K. (2024). Conversational commerce: Entering the next stage of AI-powered digital assistants. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04049-5

Barnett, S., *et al*. (2024). Seven failure points when engineering a retrieval augmented generation system. In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering – Software Engineering for AI (CAIN)* (pp. 194–199). https://doi.org/10.1145/3644815.3644945

Benita, K. (2024). Implementation of RAG in chatbot systems for enhanced real-time customer support in e-commerce. In *Proceedings of the International Conference on Advanced Computing and Robotics Systems*. IEEE. https://doi.org/10.1109/ICACRS62842.2024.10841586

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chen, Y. (2025). Application of retrieval-augmented generation for interactive industrial knowledge management via a large language model. *Computer Standards & Interfaces, 94*, 103995. https://doi.org/10.1016/j.csi.2025.103995

Desai, S., Yao, H., Porwal, U., & Lee, K. (2026). INSPIRE: Intent-aware neural sponsored product retrieval for e-commerce. In *Proceedings of the ACM SIGIR Workshop on eCommerce (ECOM'26)*. arXiv:2606.23889

Dong, X. L., *et al*. (2020). AutoKnow: Self-driving knowledge collection for products of thousands of types. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2724–2734). https://doi.org/10.1145/3394486.3403323

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review, 14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal, 50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods, 16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Hogan, A., *et al*. (2021). Knowledge graphs. *ACM Computing Surveys, 54*(4), 1–37. https://doi.org/10.1145/3447772

Jarrahi, M. H., *et al*. (2023). Artificial intelligence and knowledge management: A partnership between human and AI. *Business Horizons, 66*(1), 87–99. https://doi.org/10.1016/j.bushor.2022.03.002

Klesel, M., & Wittmann, H. F. (2025). Retrieval-augmented generation (RAG). *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-025-00945-3

Lewis, P., *et al*. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems, 33* (pp. 9459–9474).

Luo, X., *et al*. (2020). AliCoCo: Alibaba e-commerce cognitive concept net. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data* (pp. 313–327). https://doi.org/10.1145/3318464.3386132

Magnani, A., *et al*. (2022). Semantic retrieval at Walmart. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3495–3503). https://doi.org/10.1145/3534678.3539164

Myung, J., Park, J., & Han, J. (2025). HyST: LLM-powered hybrid retrieval over semi-structured tabular data. In *Proceedings of the 2nd EARL Workshop on Evaluating and Applying Recommender Systems with LLMs*. arXiv:2508.18048

Ngai, E. W. T., *et al*. (2021). An intelligent knowledge-based chatbot for customer service. *Electronic Commerce Research and Applications, 50*, 101098. https://doi.org/10.1016/j.elerap.2021.101098

Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems, 22*(3), 336–359. https://doi.org/10.1057/ejis.2012.26

Nonaka, I., & Takeuchi, H. (1995). *The knowledge-creating company: How Japanese companies create the dynamics of innovation*. Oxford University Press.

Nurmi, P., *et al*. (2008). Product retrieval for grocery stores. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 781–782). https://doi.org/10.1145/1390334.1390491

Xiao, L., *et al*. (2021). End-to-end conversational search for online shopping with utterance transfer. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.281

Zhu, Y., Vedula, N., & Malmasi, S. (2025). Hint-augmented re-ranking: Efficient product search using LLM-based query decomposition. *arXiv*. https://arxiv.org/abs/2511.13994

How Cites

Wahyudi, S. B., Sensuse, D. I., Lusa, S., Qurani, M. H., Yasin, Y., & Mailinda, I. (2026). Absent or Unretrievable: A Two-Layer Knowledge Management Framework for Retrieval-Critical Product Knowledge in a Grocery E-Commerce AI Shopping Assistant. *Computer Journal*, 320-327. https://journal.ypmma.org/cj/article/view/510**.**

Publisher's Note