DT166G, Presentation av ny Teknik - Group 15

Integrating Semantic with Keyword Product search in Online Marketplaces

MITTUNIVERSITETET

DSV Östersund

Examiner	Raja Khurram Shahzad	raja-khurram.shahzad@miun.se
Supervisor	Raja Khurram Shahzad	raja-khurram.shahzad@miun.se
Author	Albin Rönnkvist	alrn1700@student.miun.se
Author	Roberto Piran Amedi	roam2200@student.miun.se
Program	Programvaruteknik, 180hp	
Course	DT166G, Presentation av ny Teknik	
Field of study	Software Engineering	
Semester, year	HT, 2024	

DT166G, Presentation av ny Teknik
*Integrating Semantic with Keyword Product search in Online
Marketplaces*

Abstract

Product search in online marketplaces is complex, presenting several challenges in delivering relevant results, challenges that standalone keyword-based or semantic-based search models often fail to overcome. Keyword search performs well with exact term matching but struggles with context-dependent queries, while semantic search captures user intent yet may yield imprecise results when specific terms are important. This paper reviews recent literature on hybrid models, which integrate keyword and semantic models, to examine their architectures and evaluate their impact on product search result quality in online marketplaces. The findings reveal a common structural foundation among hybrid architectures and demonstrate significant improvements in recall, ranking quality, business performance, and other key metrics compared to standalone models. While these results confirm the value of hybrid models in improving the quality of product search results in online marketplaces, further research is required to understand the broader impact of hybrid models beyond this narrow yet highly relevant attribute.

DT166G, Presentation av ny Teknik
Integrating Semantic with Keyword Product search in Online
Marketplaces

1   Introduction

Online marketplaces have emerged as the dominant platforms for product searches worldwide [1, 2, 3, 4, 5]. With millions of products available and diverse customer needs, the search bar is critical in connecting users to the right products [4, 5, 6, 7]. As a result, creating a high-quality search experience has become essential for user satisfaction and business success [5]. Despite advancements in common approaches such as traditional keyword search and more recently adopted vector-based semantic search, accurately capturing user intent and delivering relevant search results remains a significant challenge [4, 5]. Hybrid search systems, which combine the strengths of both keyword and semantic search, have garnered increasing attention in recent years as a potential solution to address this challenge. However, to our knowledge, its application and impact on product search results within online marketplaces have not been thoroughly investigated. Thus, this study aims to explore the underlying architectures of hybrid product search systems and determine their impact on the quality of product search results in online marketplaces.

1.1   Background

A 2022 Harris Poll survey commissioned by Google Cloud revealed that retailers worldwide lose more than $2 trillion annually due to shoppers' inability to find what they are looking for. The same survey found that only 12% of U.S. shoppers consistently find exactly what they need, while just 11% find suitable alternatives using the search feature. Also, nearly 80% of consumers are likely to leave a site after an unsuccessful search, and 77% tend to avoid returning to websites with previous search difficulties [5]. In a more recent 2024 survey commissioned by Constructor, a U.S.-based company specializing in AI-powered search and product discovery, shoppers in the US and UK rated their e-commerce product search experience on a grading scale from A (excellent) to F (poor). The survey revealed that 42% of shoppers rated their experience a "C" or lower, highlighting widespread dissatisfaction with search results. Furthermore, 68% of shoppers expressed the need for improvements in retail search functions, particularly in the areas of understanding query intent to avoid frustrating reformulations with different terms [8].

The widespread dissatisfaction among shoppers underscores the importance and ongoing difficulties of delivering search functions that accurately capture user intent and deliver relevant results. One major challenge is the diverse range of user intentions, from specific product searches such as model numbers to more general queries like finding birthday gift ideas, which complicates precise matching [9]. The difficulty is further heightened by the fact that queries are often short, vague, or broad, and typically lack clear natural language structure, making it harder for search systems to interpret and return relevant results [4].
Online marketplaces in particular face two critical challenges that further complicate capturing user intent:
1. **Large and varied product catalogs**: As the number of products and categories grows, diversity increases, and further terminology, user intent, and context is introduced. This complexity makes it harder to fine-tune product search systems, leading to challenges in balancing precision (retrieving only the most relevant products) with recall (capturing a wide range of relevant products), which can result in less relevant search results. [3].
2. **Diverse user base:** Online marketplaces usually serve diverse user bases from various cultures, speaking different native languages and possessing expertise in different areas. These factors result in varied formulations, including differences in language structure and familiarity with industry-specific terms, often leading to mismatches between buyer queries and seller product descriptions [2, 10, 11, 12].

Two common approaches of implementing product search are the traditional keyword search and the more recently adopted semantic search [4], each with its strengths and weaknesses.
Keyword search models fundamentally rely on exact term matching, making them highly efficient when the

same terms are present in both the query and product description. However, this reliance on exact matching also causes keyword models to miss relevant results when different terms are used, especially in cases of misspellings or synonyms. Additionally, keyword models lack the capability of capturing the underlying intent behind queries, such as understanding semantics or context [2, 3, 9, 11, 13].

In contrast, semantic search models are more adept at interpreting the underlying meaning of a query. These models apply machine learning to convert textual representations of queries and products into embeddings—numerical vectors in a high-dimensional space that capture semantic meaning. These vector representations allow for calculating the semantic similarity between texts, rather than relying on exact terms [40]. While semantic models are strong at capturing general meaning, they sometimes miss important specific details like model numbers and brand names, resulting in less precise results [14, 15]. Furthermore, they have shown difficulties in handling other aspects such as negations and units of measurement [3, 16].

Given the known challenges in product search and the limitations of standalone keyword- and semantic-based search models, a hybrid search approach has gained attention as a possible solution. This approach combines the precision of exact term matching from keyword search and the understanding of broader intent from semantic search into one unified result.

Although hybrid models are designed to harness the strengths of both keyword and semantic approaches, evaluating their actual impact is essential to validate this advantage. In information retrieval (IR), including product search, result quality is assessed using both offline and online metrics.

Offline metrics are measured in a controlled setting before the system is launched for live users, ensuring that it provides value and meets quality standards. Common and useful offline metrics include Recall, Precision, Mean Reciprocal Rank (MRR), Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (NDCG). Recall is an order-unaware metric that looks only at the quantity of relevant items retrieved in relation to the total relevant items available in the dataset, ignoring their positions in the results. Similarly, Precision is also order-unaware, assessing result accuracy by examining the proportion of relevant items among those retrieved, without looking at the total number of relevant items in the dataset. In contrast, MRR, MAP, and NDCG are order-aware metrics, where the order of results affects the score. MRR focuses on the position of the first relevant result across queries, assessing how quickly users encounter a relevant item. MAP, on the other hand, considers multiple relevant items within each query, calculating precision each time a relevant item appears and then averaging these scores across all queries. It treats all relevant items equally, focusing on ranking them correctly overall, regardless of their grade of relevance. In contrast, NDCG prioritizes the ranking quality by giving higher scores to highly relevant items that appear near the top of the search results, emphasizing the importance of placing the most relevant items in prominent positions. This approach allows NDCG to account for varying levels of relevance, giving additional value to results that bring the most relevant items to the top [41].

While offline metrics establish a solid foundation, they are best reinforced by online metrics, which capture the real-world impact on live search systems as experienced by users [41]. These can include human evaluations and A/B testing. Human evaluations are often conducted by assessors who rate search results on a 3-point scale of good, fair, or poor. A/B tests often utilize a wide range of metrics to assess the value of search results, including business performance indicators like Gross Merchandise Value (GMV) and user engagement metrics, such as add-to-cart (ATC) rates.

With a range of metrics available to assess a hybrid approach, this study will review existing literature to further explore the architecture of hybrid product search systems and determine their impact on search result quality in online marketplaces.

1.2   Purpose

This study aims to review existing literature on hybrid product search systems in online marketplaces, to explore different architectures and determine their impact on search result quality.

1.3   Motivation

This research is driven by the well-known challenges in product search and the limitations of the widely used keyword- and semantic-based search models, as described in the background section of this paper. These models, while effective in specific contexts, often struggle to deliver accurate search results in complex environments [3]. This research is essential in determining whether a hybrid model can overcome the known limitations of the individual models and improve search result quality in complex online marketplaces, thereby increasing user satisfaction and retention, and ultimately driving sales [4, 5, 6].

1.4   Research Questions

- **Q1:** How can hybrid product search models, integrating semantic- and keyword-based models, be implemented in online marketplaces to leverage their complementary strengths?
- **Q2:** What is the impact of hybrid models on product search result quality compared to keyword and semantic models used independently?

2   Literature Review

2.1   Latest and Relevant articles on the selected topic

The selected studies are structured into subsections to organize the review. First, Section 2.1.1 explores hybrid architectures that integrate keyword models with semantic models. Then, Section 2.1.2 analyzes the impact of hybrid search systems on the quality of product search results, presenting relevant experiments and findings.

2.1.1 Hybrid Search Architecture

This section explores hybrid search architectures that combine keyword and semantic models. The following studies demonstrate how these hybrid models are applied and refined with the objective of improving product search results. The section begins by exploring foundational research that established the basis for future developments. It then moves to more recent studies conducted at online marketplaces, drawing inspiration from these foundational works. Lastly, it reviews the most current research, which expands on earlier findings and introduces new innovations in the field.

Foundational studies

The foundational studies that set the stage for future research in online marketplaces include the work by Nigam et al. [9] in 2019, who introduced a hybrid architecture at Amazon Marketplace that integrates a semantic model with a keyword model. For the semantic model, the authors employ a Siamese two-tower architecture, with one tower generating embeddings for queries and the other for products, each set at a dimension of 256. They employ an approximate k-Nearest Neighbor (k-NN) search method, a type of Approximate Nearest Neighbor (ANN) algorithm, using cosine similarity for matching and scoring. They do not provide specific details about the keyword model and also do not describe the hybrid re-ranking step in their architecture. Instead, they mention the general flow of a product search system using inverted index-based keyword matching, behavioral data matching and machine-learning ranking methods.
This foundational work set the stage for incorporating embedding-based semantic models into product search systems and is cited in a majority of subsequent studies within this literature review.

Motivated by the findings of Nigam et al. [9], which demonstrated the advantages of semantic models over keyword models in the retrieval stage of product search, Kuzi et al. [17] proposed a hybrid architecture in 2020 that combines keyword and semantic models for news article retrieval. The keyword model uses an inverted index with Okapi Best Matching 25 (BM25) for scoring. The semantic model employs a Siamese two-tower architecture, for query and product embeddings with 256 dimensions, generated by a BERT-based embedding model. They use an approximate k-NN search approach with the dot product for matching and scoring. The keyword and semantic model results are merged into a unified recall set, up to twice the size of the original document pool. Using Relevance Model 3 (RM3), this process selects the top-ranked documents from the combined results as the initial recall set for further processing.
Although the focus in this study is on news article retrieval, it provides insights applicable to product search used in subsequent research within this literature review.

Inspired by the hybrid semantic/keyword model presented by Kuzi et al. [17] the previous year, Gao et al. [15] proposed a complementary retrieval model called CLEAR, that uses a semantic model to complement a keyword model for general information retrieval. The keyword model utilizes an inverted index with BM25 for scoring. For the semantic model, they employ a Siamese two-tower architecture to generate embeddings using a BERT-based embedding model. The authors highlight that the semantic model incorporates residual training, a method where the semantic model is trained to address the gaps left by the keyword model specifically. These embeddings are indexed through a Maximum Inner Product Search (MIPS) index and

similarity scoring is performed using the dot product of query and document vectors. Documents retrieved from keyword and semantic models are combined into a unified recall set and are re-ranked using a weighted average of the individual scores.

While this study focuses on general information retrieval, its insights are also relevant to product search, influencing later research covered in this literature review.

In 2020, Huang et al. [18] presented a hybrid architecture for Facebook Search that integrates a semantic model with a keyword model. The keyword model utilizes an inverted index with Boolean matching for precise term-based retrieval. The semantic model employs a two-tower architecture using a cascade embedding model approach: first, a text embedding model filters candidates, and then a unified embedding model re-ranks these candidates based on social and location features. Facebook's retrieval engine, Unicorn, was extended to incorporate these semantic capabilities, using the Facebook AI Similarity Search (FAISS) library for efficient ANN search, with cosine similarity as the matching and scoring metric.

The authors developed a hybrid retrieval framework that scores documents using both keyword and semantic models, combining them into a unified system. The retrieved documents are then processed through a multi-stage ranking system to refine and produce the final ranked results. While this research initially focused on general Facebook search, it was later adapted for use in a product search context in subsequent studies discussed in this literature review.

One of the earlier hybrid models for product search in this literature review was presented by Zhang et al. [11] in 2020. The authors introduce DPSR, a personalized and semantic search model integrated into the keyword-based search system at JD Sports e-commerce platform. The keyword model uses an inverted index and incorporates query rewriting techniques. The semantic model adopts a two-tower architecture, generating embeddings for queries and products. A nearest neighbor search method is utilized, matching and scoring based on a dot product operation. The outputs from both the semantic and keyword models are merged and processed through the same ranking system and business logic to generate the final results. The ranking system employs an advanced Learning-To-Rank (LTR) approach, ensuring that the most relevant and personalized items are presented to users.

Another early product search hybrid model was introduced by Choi et al. [19] on Home Depot's e-commerce platform in 2020. The model consists of two sub-modules: a context encoder (semantic model) and a structured matching module (integration of both semantic and keyword models).

The context encoder uses a Siamese two-tower architecture based on DistilBERT to generate embeddings for the queries and product fields. The structured matching module (SMM) takes these embeddings and performs matching through element-wise operations (multiplication and subtraction) to measure semantic similarity between the query and product field embeddings. In addition to these element-wise operations, a binary matrix encodes exact matches between the terms in the query and the product fields. The combined outputs from the SMM are then processed through two fully connected layers with Rectified Linear Unit (ReLU) activation functions, refining the features into a final representation consisting of a relevance score for each product. The authors focus on candidate generation, as described above, and do not provide details on candidate re-ranking for producing a final sorted output.

Recent studies conducted at online marketplaces

Transitioning to later research in online marketplaces, inspired by the foundational studies in this review, Chang et al. [20] proposed a new semantic model in 2021 at Amazon Marketplace. This model replaced the one introduced by Nigam et al. [9], as part of the hybrid model, with focus on improving recall and reducing latency to more effectively manage large product catalogs.

For the semantic model, they apply a tree-based Extreme Multi-label Classification (XMC) method using the XR-Linear (PECOS) model, which enhances the match set by retrieving semantically relevant products.

DT166G, Presentation av ny Teknik
*Integrating Semantic with Keyword Product search in Online
Marketplaces*

While they focus on the XR-Linear (PECOS) model, they also reference the possibility of integrating it with the previous model, to improve the match set further. The authors explain that products retrieved from the keyword and semantic models are merged into a unified match set, which is then processed in a re-ranking phase. However, they do not specify the final ranking mechanism used in their system. They do not provide detailed information about the keyword model in their system either, although they mention the keyword model used in [15] as a typical approach, which involves using an inverted index with BM25 for scoring.

Building on the foundation established by Huang et al. [18] in Facebook Search, Liu et al. [21] introduced Que2Search in 2021, a semantic model integrated into the keyword-based product search system at Facebook Marketplace. The semantic model leverages a Siamese two-tower architecture, where query embeddings with 256 dimensions are generated using the XLM model and the product embeddings using the XLM-R model. The authors use Unicorn's existing semantic retrieval capabilities, developed in prior work by Huang et al. [18], which employ ANN search and cosine similarity for matching and scoring. The semantic model integrates its ANN search method with other retrieval methods such as the keyword retrieval method, combining them in a hybrid approach.
In ranking, the cosine similarity score is added as a ranking feature throughout all stages of Facebook Marketplace's two-stage ranking process. Initially, a lightweight Gradient Boosted Decision Tree (GBDT) ranker selects top candidates from the combined results of semantic search and other search methods. These candidates are further refined using a more sophisticated Deep Learning Recommendation Model (DLRM).

While previous research at Facebook [18, 21] used a semantic model to complement a keyword model for fuzzy matching, Li et al. [14] took a different approach, utilizing the keyword model to complement the semantic model for better relevance. The method was applied at Taobao Marketplace in 2021, to improve upon the existing semantic model in the hybrid product search system. The keyword model employs an inverted index. The new semantic model utilizes a two-tower architecture, where one tower generates embeddings for user-related information (combining query and user history) and the other for the product representation. The authors employ ANN search based on the inner product for matching and scoring. Additionally, they integrate a Boolean-based keyword matching module to filter the ANN search results, ensuring that items match key query terms such as brand and product type in the title. This was introduced to address the low controllability of search relevance in semantic models, a limitation they critique previous works [9, 11] for not discussing.
The products retrieved from the keyword, item-based collaborative filtering, and semantic models are merged into a unified candidate set without duplicates. This set then undergoes a multi-stage re-ranking process to produce the final result.

In 2022, Magnani et al. [3] presented a hybrid architecture for product search on Walmart Marketplace, which integrates a semantic model with a traditional keyword model. They highlight their approach of keeping the keyword and semantic models separate, drawing parallels to [9], while contrasting it with [15] which uses residual training to have the semantic model fill gaps left by the keyword model. The keyword model employs an inverted index with BM25 for scoring matches. For the semantic model, they utilize a Siamese two-tower architecture, where one tower generates embeddings for the query and the other for the products. After experimenting with various embedding models and dimensions, they ultimately chose DistilBERT and 256 dimensions. Retrieval is performed using ANN search with cosine similarity for matching and scoring. Although they acknowledge FAISS as a suitable tool, they opted for a commercially available managed ANN service in their implementation.
The authors explain that products retrieved from the keyword and semantic models are combined into a unified recall set, which is subsequently re-ranked using a Gradient Boosted Decision Tree (GBDT) model. This re-ranking process leverages various features, including query attributes, product information, and

relationships between the query and product, to produce the final ranked list of products presented to the user.

Most recent studies

Finally, the most recent studies are examined, building on earlier research and introducing new advancements in the field. In a 2024 study by Lin et al. [13], the authors build upon their earlier hybrid model at Walmart Marketplace [3], concentrating on improvements to address challenges associated with the semantic model, such as handling relevance, misspellings, and the accuracy of training data. They cite [9, 11, 14] as successful examples of embedding-based semantic models integrated into e-commerce search with positive outcomes. However, they also reference previous research [15, 21] to argue that downstream re-ranking systems may not always correctly rank results from semantic models, leading to irrelevant products being displayed. This supports their focus on enhancing relevance performance within the semantic model.

In the most recent study from October 2024, Kekuda et al. [26] presents a hybrid architecture for product search on Best Buy's e-commerce platform that integrates a semantic model with an existing keyword model. The keyword model uses an inverted index with TF-IDF scoring. For the semantic model, they implement a two-tower architecture to create embeddings for queries and products, using b3-small, a fine-tuned version of the RoBERTa model, for each tower. These embeddings are stored within Solr and retrieved using ANN search based on cosine similarity.

The authors mention that products retrieved from the keyword and semantic models are combined into a unified recall set. However, they do not provide detailed information on the specific re-ranking algorithm to sort these results.

2.1.2 Hybrid Search Impact on Product Search Results

This section analyzes the impact of hybrid models on the quality of product search results, based on a range of experiments and findings in reviewed studies. Thus, offering insights into how hybrid models affect recall, ranking accuracy, and other important factors in product search. As in the previous section, this section begins with foundational research, moves to recent studies in online marketplaces, and concludes with the latest research.

Foundational studies

One of the most essential foundational papers in this review, by Nigam et al. [9], performed online A/B tests on Amazon Marketplace across three product categories: toys and games, kitchen, and pets. The proposed semantic model was integrated as a component within the keyword-based production system and evaluated against a version without semantic capabilities. The authors state that the results showed statistically significant improvements in conversion rate, revenue, and other key performance indicators (KPIs), though specific figures were not provided. They further conclude that a combination of qualitative analysis of the improved search results and the observed increased business metrics validated that the proposed hybrid model effectively enhanced the user search experience.

Compared to the relatively limited information from online tests by Nigam et al. [9], Kuzi et al. [17] conducted an in-depth offline empirical analysis. The experiment was divided into four parts, focused on comparing the performance of their proposed hybrid model against other individual search models as baselines. The study utilized a Text REtrieval Conference (TREC) collection consisting of 441,676 news-wire documents from TREC Disks 1 & 2, standard benchmark datasets in information retrieval research. TREC topics 51-200 were used as queries, with the titles (short phrases/keywords) serving as the search input, aligning with the structure of product search. Performance was assessed using multiple metrics: Recall, MAP, #rel (total relevant documents retrieved), and Reliability of Improvement (RI). A two-tailed paired t-test was applied to

evaluate the statistical significance of performance differences, with a p-value threshold of 0.05.
The first experiment evaluated the recall performance of the hybrid model in comparison to the keyword and semantic models separately. The findings indicated that the keyword model achieved higher recall than the semantic model alone across all result list sizes. However, the hybrid model consistently improved recall across all list sizes, with an average increase of 5.55%. The authors suggest that although the semantic model alone is not as effective as the keyword model, it adds value by retrieving additional relevant documents when used as a component in the hybrid model.

The second experiment assessed the hybrid model's performance against the baseline keyword model across various list sizes (500, 1000, 1500, and 2000). The hybrid model consistently improved recall by around 2.8% and showed slight MAP gains, retrieving 3.0% to 3.2% more relevant documents than the keyword model alone. The RI measurement confirmed the stability of these recall improvements across different queries. An additional analysis tested if extending the keyword list length could achieve similar results. By merging a 2000-document semantic list with different keyword list lengths and clipping to the original size, the hybrid model consistently added around 700 unique relevant documents. This demonstrates that the semantic model identifies distinct items that the keyword model misses, even when extended.

The third experiment assessed the robustness of the hybrid model against the baseline keyword model using a result list of 1000 documents. The hybrid model improved recall for 50% of queries, with an average increase of 18%, while 40% showed no change, and 10% experienced a minor decrease of around 4%. Queries were grouped into four categories (Q1-Q4) based on their baseline performance. The largest improvement (14%) was observed in the poorest-performing group (Q1), while Q2 and Q3 saw smaller gains of 5.5% and 1.7%, respectively. The best-performing group (Q4) showed minimal improvement (0.5%), indicating limited added value when the baseline already performs well. Improvements in Q1, Q2, and Q3 were statistically significant. Further analysis indicated that longer queries, queries with higher IDF values (rare terms), and queries with greater term frequency variability benefited most from the hybrid model. The authors suggest a possible explanation being the ability of neural networks to learn semantics using multiple words.

The fourth experiment analyzed the differences between documents retrieved by the semantic and keyword models to understand their complementarity. Using a 1000-document result list, it showed that the semantic model often retrieved documents with minimal overlap compared to the keyword model, as indicated by low Jaccard index values. The semantic model focused on specific subtopics while the keyword model retrieved more general terms, demonstrating its ability to enhance specificity. t-Distributed Stochastic Neighbor Embedding (t-SNE) visualizations revealed that semantic documents formed distinct clusters, indicating they contribute unique information and increase diversity. The semantic model also retrieved longer documents, which the authors suggest may be because these are often penalized by the keyword model. A quantitative analysis of 50 queries confirmed the consistent low overlap and stable performance of the semantic model, supporting the hybrid model's effectiveness in expanding topic coverage and diversity.

The authors conclude that their in-depth empirical analysis demonstrates the effectiveness of the hybrid model, revealing the complementary strengths of the keyword and semantic models. They confirm that the semantic model successfully retrieves a substantial number of relevant documents that the keyword model does not capture, suggesting that semantic search is an effective enhancement and complement to traditional keyword search.

Similarly to the approach in Kuzi et al. [17], Gao et al. [15] conducted three offline experiments to evaluate their proposed hybrid model against keyword models (BM25, BM25+RM3), BERT-augmented keyword models (DeepCT, DeepCT+RM3), and a purely semantic model (BERT-Siamese). Two datasets were used: (1) the Microsoft MAchine Reading COmprehension (MS MARCO) Passage Ranking Dataset, containing 8.8 million passages and 0.5 million query-passage pairs, and (2) the TREC 2019 Deep Learning Track Queries dataset, which includes 43 queries with graded relevance levels manually judged by assessors. The evaluation focused on ranking accuracy using metrics such as MRR@10, MAP@1k, NDCG@10, and also

DT166G, Presentation av ny Teknik
Integrating Semantic with Keyword Product search in Online
Marketplaces

coverage using Recall@1k.

The first test aimed to evaluate the retrieval accuracy of the proposed hybrid model compared to the baselines. The results showed that the proposed hybrid model consistently outperformed all baseline models across all metrics and datasets, with statistically significant improvements. For instance, in terms of NDCG@10 on the TREC dataset, the proposed model (with score 0.699) outperformed BM25 by 38.1%, BM25+RM3 by 25.9%, DeepCT by 26.9%, DeepCT+RM3 by 16.3%, and BERT-Siamese by 17.7%. The authors suggest that the results demonstrate the limitations of keyword models due to their reliance on strict term matching, highlighting the proposed hybrid model's advantage in using embeddings for semantic-level, flexible matching. Additionally, the proposed hybrid model significantly outperforms the purely semantic model, indicating that semantic-only retrieval is insufficient for optimal result quality.

The second test aimed to assess the impact of the proposed hybrid model within a two-stage retrieval pipeline using a BERT re-ranker, compared to a BM25-based keyword model, evaluated through MRR@10 and NDCG@10 metrics across both datasets. The findings indicated that the proposed hybrid model, even without re-ranking, achieved performance levels nearly equivalent to the BM25 + BERT re-ranker setup, showcasing its robust initial retrieval capabilities. When integrated with the BERT re-ranker, the proposed hybrid model exhibited statistically significant improvements across all metrics and datasets over the BM25-based pipelines. The authors highlight that the proposed hybrid model's ability to produce strong initial rankings not only enhances recall but also improves the efficiency and scalability of the reranking process.

The third experiment served as a case study to examine how the proposed hybrid model addresses limitations of the BM25-based keyword model and identifies new challenges that arise when using semantic matching. The authors analyzed specific queries to illustrate the proposed hybrid model's ability to retrieve relevant documents that the keyword model missed, especially in cases of vocabulary mismatches. The proposed hybrid model's integration of semantic embeddings allowed it to resolve these issues, improving retrieval performance where the keyword model failed due to exact term mismatches, aligning with findings from Kuzi et al. [17]. However, the study also revealed challenges, where the proposed hybrid model occasionally retrieved semantically related but irrelevant documents—false positives that the keyword model did not retrieve. These false positives often shared thematic similarity with the query but did not meet the specific relevance criteria.

The authors conclude that traditional keyword retrieval models have difficulty grasping the semantic meaning of queries, while purely semantic models overlook crucial word-level details. Similar to Kuzi et al. [17], they find that the complementary strengths of their proposed hybrid model effectively bridge this gap by integrating a keyword model with a semantic model tailored to address the keyword model's shortcomings.

Similar to the detail-limited online experiments by Nigam et al. [9], Huang et al. [18] performed online A/B experiments for their proposed hybrid model on Facebook Search. The authors claim their results showed significant metrics gains compared to the existing production system, which lacked semantic capabilities, though specific details are not provided.

In contrast to earlier studies in this literature review that focused solely on online or offline experiments, Zhang et al. [11] conducted both offline and online evaluations of their semantic model, both independently and as part of a hybrid model.

The offline tests assessed the semantic model in isolation against a BM25-based keyword model. The evaluation used a dataset of 200,000 queries with relevance determined through human-labeled data, covering around 15 million products. The metrics included Top-K, which assesses the likelihood that a relevant item appears within the top K results out of a large random selection, and Area Under the Curve (AUC), which evaluates the model's ability to distinguish between relevant and non-relevant items, with a score of 1 indicating perfect distinction. The results showed that the semantic model, in its vanilla form, increased Top-1 by 16.9% and Top-10 by 3.2%, with an improvement in AUC score from 0.661 to 0.696. The authors suggest that these findings indicate the proposed pure semantic model achieves the highest retrieval

relevance.

The online tests involved both manual evaluation and A/B testing, where the proposed semantic model was integrated into the existing keyword-based production system, and compared to the baseline production system operating without semantic capabilities.

In the manual evaluation, human assessors rated the relevance of search results for a set of 500 long-tail queries using a 3-point scale: perfect, fair, and bad. The results showed that the proposed hybrid model reduced the percentage of "bad" cases from 17.86% to 13.70%, increased the percentage of "fair" cases from 26.04% to 33.28%, but showed a drop in "perfect" cases from 56.10% to 53.01%. Based on these results, the authors argue that the proposed hybrid model is especially effective in handling complex long-tail queries, which often require semantic matching.

A live A/B test was conducted on 10% of JD Sports site traffic over two weeks, using standard protocols. The results show that the proposed hybrid model improves the baseline for all core business search metrics, including user conversation rate (+10.03%), and gross merchandise value (+7.5%), as well as query rewrite rate (-9.99%), which the authors believe to be a good indicator of search result satisfaction.

The authors conclude that their proposed hybrid model produced significantly better results, especially for long-tail queries, which they claim was a challenge in their existing keyword search system. The proposed hybrid model was deployed into JD Sports production product search system.

In a different e-commerce context, Choi et al. [19] conducted offline experiments to evaluate the performance of their proposed hybrid model against keyword models (BM25, BM25F), a semantic model (Arc-II), and a hybrid model (Duet). Two primary datasets were used: (1) an internal click log dataset comprising 11.6 million entries from real-world user interactions on an e-commerce platform, and (2) a human-annotated dataset from Kaggle's Product Search Relevance (PSR) dataset. The test set consisted of 5,000 unique queries for each dataset. The evaluation focused on ranking accuracy using metrics such as NDCG, MAP, and MRR. Results showed that the proposed hybrid model outperformed all baselines across all metrics. For example, on the PSR dataset, the proposed hybrid model achieved an NDCG@5 score of 0.417, surpassing the Duet hybrid model's score of 0.408, the keyword models' score of 0.384, and the semantic model's score of 0.380. Statistical tests confirmed that these improvements were significant ($p < 0.05$). Additionally, the authors noted that the Duet hybrid model outperformed all other baselines aside from the proposed model, affirming the value of combining both semantic and keyword-based approaches.

The ablation study demonstrated that the proposed hybrid model, which integrates the SMM component with DistilBERT, significantly outperforms DistilBERT alone, particularly for complex, multi-intent queries. The performance gains were statistically significant. Error analysis showed that the SMM component enhanced performance for query types such as units, materials, models, typos, and general queries. However, it was less effective in handling brand and color-related queries.

Recent studies conducted at online marketplaces

Shifting focus to experiments conducted in online marketplaces, Chang et al. [20] performed offline and online experiments of their proposed semantic model in isolation and also when integrated into a hybrid model. The offline tests evaluated the proposed semantic model's performance, focusing on recall, against two baselines: the semantic model developed by Nigam et al. [9] and a standard BM25-based keyword model, similar to the one used in Gao et al. [15]. The dataset comprised 1 billion query-product pairs derived from 240 million queries and 100 million products. A test set of 100,000 queries (yielding 182,000 query-product pairs) was sampled from the entire catalog of 100 million products. Recall metrics (Recall@10, Recall@50, Recall@100) were used to assess relevance. The results showed that the proposed semantic model consistently outperformed both baselines on all recall metrics, achieving a Recall@100 of 60.30%, compared to 36.19% for the semantic model and 21.72% for the keyword model.

The authors conducted an online A/B test at Amazon Marketplace, where the baseline employed a hybrid model combining a keyword model with candidates from the previous semantic model. This was compared to

DT166G, Presentation av ny Teknik
*Integrating Semantic with Keyword Product search in Online
Marketplaces*

the new hybrid model, with the proposed semantic model replacing the previous one. The results demonstrated that the new proposed hybrid model achieved major improvements in several KPIs compared to the baseline over an unspecified period.

The authors conclude that their proposed semantic model enhances product discovery in online retail search systems by delivering higher recall rates than standalone keyword-based models, standalone semantic models, and the existing hybrid model when replacing the semantic component.

In another online marketplace setting, Liu et al. [21] conducted online A/B tests at Facebook Marketplace, comparing their proposed hybrid model to the production baseline, which lacked semantic capabilities. The results demonstrated that the proposed hybrid model achieved over a 4% increase in product search engagements, with approximately 2.8% attributed to semantic retrieval and 1.24% from incorporating cosine similarity scoring as a ranking feature in the multi-stage ranking process.

The authors conclude that their proposed hybrid model significantly improved multiple metrics and was successfully deployed in production at Facebook Marketplace, supporting millions of product searches daily. Similar to previous conclusions in [15, 17], they also emphasize the importance of treating the semantic model as a component for semantic issues rather than a silver bullet for all retrieval challenges.

Similar to the hybrid model comparison experiments conducted in [20], Li et al. [14] performed online tests to evaluate the performance of their proposed hybrid model (using the proposed semantic model as the semantic component) against the existing production hybrid model (using a DNN-based semantic model) at Taobao Marketplace. The evaluation employed metrics including GMV, *Pgood* which measures the percentage of relevant products determined by an automated relevance model, and *Ph_good* evaluating relevance through human assessments. Results showed that the proposed hybrid model increased GMV by 0.77% and transaction count by 0.33%, indicating better business performance. Additionally, the proposed hybrid model improved *Pgood* by 1.0% and *Ph_good* by 0.35%, indicating more relevant search results and demonstrating the effectiveness of their approach using a keyword filter on the semantic results. The proposed hybrid model also raised the number of products entering pre-ranking and ranking stages by 22.53% and 36.76%, respectively, demonstrating more efficient resource use. The authors contextualize these figures by emphasizing that, considering the billions of daily transactions on Taobao Search, a 0.77% improvement equates to tens of millions in transaction value.

The authors conclude that the effectiveness of the proposed hybrid model has been verified and deployed in Taobao Product Search, serving hundreds of millions of users in real time.

Magnani et al. [3] conducted offline and online tests to evaluate the impact of their semantic model in isolation and also when integrated into a hybrid model at Walmart Marketplace.

The offline tests were conducted to evaluate the semantic model in isolation, comparing it against a keyword model baseline to assess recall differences. The recall results were generated using a test dataset of 140,000 queries from Walmart's logs, with relevant products identified based on user engagement and editorial feedback, drawn from a catalog of approximately 7 million products. The results demonstrated that the semantic model improved Recall@40 by up to 18.22% over the baseline, but also showed an -18.33% drop in Category Recall@40. These findings suggest that while the semantic model retrieves additional relevant products not captured by the keyword model, it does so with less category relevance.

The online tests were divided into two parts: manual evaluation and interleaving, comparing the proposed hybrid model to the existing production system, which lacked semantic capabilities and served as the baseline. The manual evaluation focused on tail queries, where human assessors rated the top-10 ranking results using a 3-point scale: not relevant, relevant with missing attributes, and perfect match. The results, based on the human ratings, demonstrated that the proposed hybrid model improved NDCG@5 by up to 2.02% and NDCG@10 by up to 2.88% compared to the baseline, with low p-values (< 0.05) indicating statistically significant improvements. This suggests the hybrid system enhances ranking quality, effectively

positioning the most relevant items at the top of the search results. The interleaving evaluation focused on user engagement, where users were presented with a mix of ranking results from both the hybrid model and the baseline. The metric used was ATC@40, measuring the number of add-to-carts within the top 40 ranking positions. The results demonstrated that the hybrid model achieved an ATC@40 lift of up to 0.54%, with low p-values indicating statistically significant improvements. This suggests that the hybrid model enhances the likelihood of users adding items to their carts, demonstrating its effectiveness in fulfilling user needs within a live production environment.

The authors conclude that the proposed hybrid model significantly enhanced the relevance of the product search engine and was deployed to production at Walmart Marketplace.

Most recent studies

In a more recent study by Lin et al. [13], aimed to replace the semantic model developed by Magnani et al. [3], the authors conducted online experiments to assess their new semantic model within Walmart's existing hybrid model. The experiments involved manual evaluations and A/B testing, comparing the proposed hybrid model to the current production hybrid model, which served as the baseline.

The manual evaluation focused on the top-10 ranked products where human assessors categorized the product relevance using a 3-point scale: exact match, substitute, and irrelevant. The results showed that the proposed hybrid model delivered a lift in NDCG@5 of up to 1.64% and NDCG@10 of up to 1.18% compared to the baseline, with low p-values (< 0.1), indicating statistically significant improvements. These results suggest that integrating the new semantic model enhances ranking quality.

The A/B testing focused on user engagement through live traffic tests at Walmart, utilizing the revenue lift metric, assessing the increase in revenue. The results demonstrated that the proposed hybrid model achieved up to a 0.43% revenue lift compared to the baseline, with statistically significant p-values (< 0.1). This indicates that the proposed hybrid model not only enhances the relevance of search results but also positively impacts user engagement and purchase behavior in a live production environment.

The authors summarize that the new semantic model demonstrated superiority over the existing one and was successfully integrated into the production hybrid product search system at Walmart.

In the most recent study within the product search context, Kekuda et al. [22] performed an online A/B test incorporating their proposed semantic model into the existing keyword-based search system, which lacked semantic matching. The test demonstrated a 3% increase in conversion rates, though further details were not provided. The hybrid model was deployed to production at Best Buy.

2.2   Strengths and Limitations of Existing Works

In this section, we explore the strengths and weaknesses in the literature, with particular emphasis on the hybrid architecture and experiments conducted. Additionally, we evaluate the relevance of the articles included in this narrative review to determine their influence on the overall findings.

TABLE I
STRENGTHS AND LIMITATIONS OF EXISTING WORKS

No.	Strengths	Limitations
[9]	*Experiments*: A/B tests were conducted in an e-commerce environment. *Relevance*: Research conducted on a large e-commerce platform, likely Amazon	*Experiments*: The A/B tests do not provide sufficient empirical data or detailed information needed for a comprehensive evaluation of the results.

	marketplace, as the author was affiliated with Amazon during the research, and the example result set included Amazon-specific features like Prime and a marketplace layout. Peer reviewed.	*Relevance*: Too focused on semantic capabilities in isolation and not on the hybrid approach as a whole.
[17]	*Architecture*: The use of open-source tools like Anserini and BERT-based models means that other researchers can easily reproduce the experiments. *Experiments*: The use of the TREC collection (a well-known and widely used benchmark in information retrieval research) ensures that the results are comparable to other research.	*Experiments*: Only based on offline tests, no online tests were conducted to gain insights in real-world impact. *Experiments & Relevance*: The dataset used in the experiments did not specifically match e-commerce, however it's similar in nature, making it appropriate for evaluating retrieval models. *Relevance*: Not peer reviewed.
[15]	*Experiments*: Robust offline tests using verified datasets, a variety of baselines and metrics, assessing both relevance and ranking quality. Specifically the datasets used were TREC and MS MARCO, which are reliable and respected benchmarks for evaluating IR systems.	*Experiments*: No online experiments conducted. *Relevance*: Not focused on product search or marketplaces. Not peer reviewed.
[18]	*Experiments*: In-depth analysis and experiments on optimizing embeddings. *Relevance*: Peer reviewed.	*Experiments*: The A/B tests lack clarity and do not offer enough details for proper evaluation. *Relevance*: It is not specific to marketplaces or e-commerce.
[11]	*Experiments*: Robust tests conducted on an e-commerce platform with a strong baseline, different metrics and both offline and online tests. *Relevance*: Research conducted on an e-commerce platform. Peer reviewed.	*Relevance*: Not specifically focused on marketplaces.

[19]	*Experiments*: The offline experiments were well-executed, utilizing diverse metrics and a relevant e-commerce dataset. It effectively demonstrated the overall performance of the model while highlighting specific areas of strength and weakness. *Relevance*: Research conducted in an e-commerce setting with focus on product search.	*Experiments*: No online experiments were conducted. *Relevance*: Not peer reviewed.
[20]	*Experiments*: Robust offline tests comparing the proposed model to other standalone methods. A/B tests were conducted on an e-commerce platform. *Relevance*: Research conducted on an e-commerce platform, likely Amazon marketplace, for the same reasons stated in [25]. Peer reviewed.	*Experiments*: The A/B tests do not provide sufficient empirical data or detailed information needed for a comprehensive evaluation of the results. Additionally, the offline tests only looked at Recall, which does not provide insights into ranking quality. *Relevance*: Too focused on semantic capabilities and not on the hybrid approach as a whole.
[21]	*Relevance*: Research conducted at Facebook Marketplace, a large online marketplace. Peer reviewed.	*Experiments*: The A/B tests lack clarity and do not offer enough details for proper evaluation.
[14]	*Experiments*: Experiments conducted on Taobao's marketplace platform with a highly relevant dataset and diverse metrics. Also gives insight into financial results. *Relevance*: Research conducted on Taobao, a large online marketplace. Peer reviewed.	*Experiments & Relevance*: The online experiments were conducted against an existing hybrid model, making it difficult to assess the impact of integrating semantic search.
[3]	*Architecture*: Detailed explanations of the mechanics of a hybrid system and its impacts. *Experiments*: Comprehensive experiments conducted at Walmart using varied and complementary metrics.	*Experiments*: The experiments focused solely on tail queries, leaving the system's performance across the entire range of search queries uncertain.

	Relevance: Research conducted at Walmart, an online marketplace. Peer reviewed.	
[13]	*Experiments*: Provides additional insights into the hybrid impacts, beyond [3], by incorporating revenue metrics in the online experiments. *Relevance*: Research conducted at Walmart, an online marketplace.	*Experiments*: The online experiments lack details on the scale of the tests, such as the duration of the A/B tests and the number of human assessors involved. *Relevance*: Not peer reviewed.
[22]	*Relevance*: Product search specific with implementation on a large e-commerce platform. *Relevance*: Very recent, October 2024.	*Experiments*: The A/B tests do not provide sufficient empirical data or detailed information needed for a comprehensive evaluation of the results. *Relevance*: Not specifically focused on marketplaces. Not peer reviewed.

2.3   Comparative analysis

Following the same structure as outlined earlier, this comparative analysis is divided into subsections. First, Section 2.3.1 provides a comparison of the hybrid models previously discussed. Then, Section 2.3.2 examines the impact of different hybrid models on product search results, based on experimental findings.

2.3.1 Hybrid Search Architecture

This section compares the architectures of the hybrid models discussed in the reviewed studies, focusing on their keyword and semantic components, as well as the hybridization and re-ranking processes.

Keyword model

Most hybrid architectures implement a traditional keyword search system as the core component of their retrieval mechanism. Nearly two-thirds of the studies [3, 11, 14, 15, 17, 18, 21, 22] explicitly mention employing an inverted index. Among these, a significant portion [3, 11, 15, 17, 18, 22] utilize BM25 as the scoring model for ranking matches.

Semantic model

In the hybrid architectures reviewed, the semantic models are based on embedding-retrieval models that capture the semantic meaning of queries and documents/products. A fundamental similarity across the studies is the use of the two-tower architecture, where one tower generates embeddings for the user queries and the other for the documents/products [3, 9, 11, 14, 15, 17, 18, 19, 21, 22]. A subset of these studies [3, 15, 17, 19, 21, 22, 25] specifically employ a Siamese two-tower architecture, where both towers share the same structure.

Various embedding models are used to generate these embedding representations. For instance, BERT-based models are used in studies [17, 15, 3, 22], XLM/XLM-R is utilized in [21], and XR-Linear in [20]. Standard embedding dimensions, such as 256, are found in studies [3, 9, 17], reflecting decisions to balance performance and computational efficiency. One study [18] takes a specialized approach by utilizing a

18

cascade embedding model, where an initial text embedding filters candidates, followed by a unified embedding model that refines them using additional features such as social and location data.

A common feature across the studies is the use of the ANN search algorithm, as seen in several studies [3, 9, 13, 14, 17, 18, 21, 22], which sacrifices a small degree of precision in exchange for significantly lower latency. The most common similarity matching and scoring technique is cosine similarity, which measures the angle between vectors to determine how closely they align, as seen in studies [3, 9, 18, 21, 22, 27]. Alternatively, some studies [11, 15, 17] use the dot product, which considers both the direction and magnitude of the vectors.

Hybridization and Re-Ranking
While many of the reviewed studies lack essential details about the hybridization and re-ranking processes, some common patterns emerge. The outputs from both models are often combined to create a unified candidate set [3, 9, 13, 14, 15, 17, 18, 19, 20, 22]. Following this merging, a re-ranking phase refines and ranks candidates based on relevance. Several studies incorporate machine learning models for re-ranking [3, 9, 11, 14, 21], notably GBDT [3, 14, 21], DLRM [21] and LTR models [11].

Most approaches keep the models distinct throughout the retrieval process, however, [14, 15] apply specialized techniques to complement the models before merging and re-ranking. For instance, [15] trains the semantic model to address gaps left by the keyword model, while [14] enhances the semantic model candidates with a keyword-based filter for better relevance.

2.3.2 Hybrid Search Impact on Product Search Results
This section compares the findings and methodologies of the experiments conducted in the reviewed studies, covering platforms and datasets, metrics, offline tests, and online evaluations, including human assessments and A/B tests. The comparison highlights the impact of hybrid models on product search results, as well as their strengths and limitations.

Platforms & datasets
The majority of the reviewed studies were conducted in online marketplaces [3, 13, 14, 20, 21, 25], while three studies were conducted on other types of e-commerce platforms [11, 19, 22]. These studies utilized comprehensive and relevant product datasets tailored to reflect real-world e-commerce environments.

The remaining three studies examined other platforms beyond traditional e-commerce and marketplaces. [18] investigated Facebook's search functionality, setting the foundation for its integration within Facebook Marketplace in later research [21]. Additionally, [17] analyzed news articles using the TREC collection, a well-established benchmark in information retrieval research. Another study [15] concentrated on general information retrieval, utilizing TREC and MS MARCO datasets, which are highly regarded benchmarks for assessing IR systems.

Offline experiments
The offline experiments conducted used a range of datasets and metrics, including Recall, MRR, MAP, NDCG, and AUC, among others. Statistical significance was assessed using p-values.

Several studies compared their proposed hybrid models against keyword- and semantic-based baselines [15, 17, 19, 20]. Other studies focused on evaluating their proposed semantic model in isolation against keyword models [3, 11]. Additionally, [19] further validated their hybrid model's effectiveness by comparing it against existing hybrid models.

For e-commerce and marketplace platforms, studies [20] and [19] compared their hybrid models against keyword- and semantic-based baselines. The hybrid models consistently improved key metrics like recall, MAP, and NDCG, indicating that the integration of semantic components enhanced relevance and ranking quality.

Studies [3] and [11] evaluated their proposed semantic models in isolation against keyword models. Both studies demonstrated improved recall metrics, showing that semantic models can retrieve additional relevant items that the keyword models fail to capture. However, [3] noted a limitation: while the semantic model retrieved more relevant items, it did so with less category-specific alignment.

[19] also tested their hybrid model against another existing hybrid model, showing improvements in NDCG, further validating the advantage of its approach over similar systems.

Studies examining other platforms revealed similar patterns. Studies [17] and [15] examined their proposed hybrid models against keyword- and semantic-based baselines, showing that hybrid models consistently improved recall, MRR and MAP scores across various datasets and query sizes. Thus, highlighting the effectiveness of hybrid models in improving relevance and ranking quality.

Online human evaluation experiments

The online human evaluation experiments utilized the NDCG metric, based on a 3-point scale assessed by human evaluators. Statistical significance was determined using p-values.

For e-commerce and marketplace platforms, [3] demonstrated that the hybrid model, compared to the existing production system lacking semantic capabilities, resulted in a significant improvement in NDCG, supported by statistically significant p-values. These findings indicate that integrating the semantic model improves ranking quality by surfacing more relevant items at the top of search results.

In another study, [11] evaluated their proposed hybrid model against the production system lacking semantic capabilities. The results showed a reduction in the percentage of "bad" cases, an increase in "fair" cases, and a slight decrease in "perfect" cases. The authors suggest that this outcome demonstrates the hybrid model's effectiveness, especially for handling complex, long-tail queries that require semantic matching. Additionally, [13] compared its proposed hybrid model against the existing production hybrid model and observed a lift in NDCG, also supported by low p-values.

Online human A/B experiments

The A/B tests employed metrics such as ATC, revenue lift, product search engagements, conversion rate, GMV, query rewrite rate, KPIs, *Pgood* (automated relevance assessment) and *Ph_good* (human relevance evaluation). Statistical significance was confirmed through p-values.

For e-commerce and marketplace platforms, studies [3, 9, 11, 21, 22] compared their hybrid models against the existing production systems lacking semantic components, showing consistently positive results. [3] demonstrated improvements in ATC with statistically significant p-values, suggesting the hybrid model increased the likelihood of users adding items to their carts. [21] reported a rise in product search engagements, indicating higher user interaction levels with the new system. [11] showed gains across several core business metrics, including user conversion rate, GMV, and query rewrite rate, which the authors highlighted as indicators of improved search satisfaction. [25] found statistically significant improvements in conversion rate, revenue, and other KPIs, supporting the value of integrating the semantic model with the existing keyword-based model. [22] recorded a 3% increase in conversion rates, further validating the hybrid model's business value.

Studies [13, 14, 20] evaluated updated proposed hybrid models by substituting the existing semantic models with their proposed semantic models, demonstrating significant improvements. [13] reported a revenue lift with low p-values, showing that the proposed hybrid model positively changes purchase behavior. [20]

observed substantial KPI enhancements, indicating the superiority of the proposed hybrid model. [14] recorded increases in GMV and transaction counts, alongside improvements in *Pgood* and *Ph_good* scores, confirming enhanced search relevance.

[18] evaluated its proposed hybrid model outside traditional e-commerce or marketplace contexts, testing it against the existing Facebook Search production system, which lacked semantic capabilities. The authors reported notable improvements in metrics, though additional details are needed for a more thorough evaluation.

Hybrid Model Strengths and Limitations Experiments
[17] revealed that the hybrid model is mainly effective for longer queries, queries with rare terms (high IDF values), and those with high term frequency variability. Additionally, [11] showed that the hybrid approach proved especially effective in handling complex, long-tail queries that require semantic matching.
This indicates that the hybrid model is mainly effective at managing complex search scenarios.

Furthermore, results in [17] showed that the semantic model often retrieved documents with minimal overlap compared to the keyword model. The semantic model also showed a tendency to retrieve longer documents, which the authors argue may be due to these documents being penalized by the keyword model. Additionally, [15] demonstrated that the hybrid model addressed vocabulary mismatches through the semantic component, retrieving relevant documents overlooked by the keyword model.
This shows the individual models' complementarity nature and supports the hybrid model's ability to improve diversity and recall.

However, the hybrid model also exhibited certain limitations. As [15] reported, the hybrid approach sometimes retrieved semantically related but irrelevant documents. While these documents shared thematic similarities with the query, they did not meet the specific relevance criteria, demonstrating a challenge in maintaining precision while expanding retrieval diversity. [11] also noted a slight decline in "perfect" cases for the hybrid model, suggesting it doesn't enhance results for queries that keyword-based models handle well, such as term-specific searches like model numbers. However, the overall improvement across the majority of cases strongly favored the hybrid model.

DT166G, Presentation av ny Teknik
*Integrating Semantic with Keyword Product search in Online
Marketplaces*

3   Identified Problem and Proposed Solution

This section presents a solution for improving product search result quality, applicable to online marketplaces. It starts by outlining the challenges within current product search systems and proposes a solution, highlighting the advantages of integrating semantic and keyword search models. The proposed solution is then compared with similar existing models. Lastly, the chapter concludes with an evaluation of the proposed solution's strengths and limitations.

3.1   Description

Previous research highlights several recurring challenges in product search systems and limitations of keyword and semantic models used independently. Experiments from the reviewed studies consistently demonstrated that integrating semantic capabilities into traditional keyword-based search systems addresses these limitations and improves product search results. The integration varies depending on the existing system, scale, available resources, and other factors. Large marketplaces such as Walmart, Taobao, Facebook, and Amazon employ complex hybrid architectures featuring specialized embedding models, multiple matching components, and advanced re-ranking mechanisms.

The literature review clearly demonstrated that hybrid models enhance product search results. However, their architectures can become complex, often demanding expertise and resources that may not be readily available to all marketplaces. To address this, we propose a lightweight, adaptable, and cost-effective hybrid model that integrates semantic and keyword models, providing a foundational framework for further customization on specific marketplace platforms. The proposed model is publicly available on GitHub [23].

The proposed model utilizes Elasticsearch, a widely used free and open-source search engine, as the foundation. The keyword component employs an inverted index and basic term matching with BM25 scoring, which has been widely adopted in previous studies. For the semantic component, a Siamese two-tower architecture is implemented to generate query and product embeddings with 768 dimensions. The choice of 768 dimensions enhances precision, though reducing the dimensions could be beneficial for larger-scale environments, as noted in prior research. OpenAI's *text-embedding-3-small* embedding model is used due to the simplicity of the embedding API, but the proposed system can be easily adapted to use other models, such as open-source BERT-based models commonly used in previous studies. Product embeddings are generated offline and stored in an Elasticsearch index in a dense vector field within an HNSW structure [24]. Query embeddings, on the other hand, are generated in real-time and used immediately for searching. Retrieval is performed using Elasticsearch's built-in approximate k-NN search and cosine similarity (recommended for OpenAI's embeddings [25]), a widely adopted method in previous studies.

To minimize latency, searches run in parallel across the keyword and semantic components. The individual result sets are then merged and re-ranked using a custom implementation of Reciprocal Rank Fusion (RRF) [26]. RRF was chosen for its effectiveness and simplicity, as it delivers strong results without the need for training or tuning a complex model, unlike most approaches used in existing studies. Additionally, it does not require relevance signals to be correlated, making it a flexible and efficient "plug-and-play" strategy [27, 28]. Although Elasticsearch offers a built-in RRF re-ranker, it is available only on a paid tier, making a custom implementation more accessible and cost-effective. The proposed solution is illustrated in Fig.1.
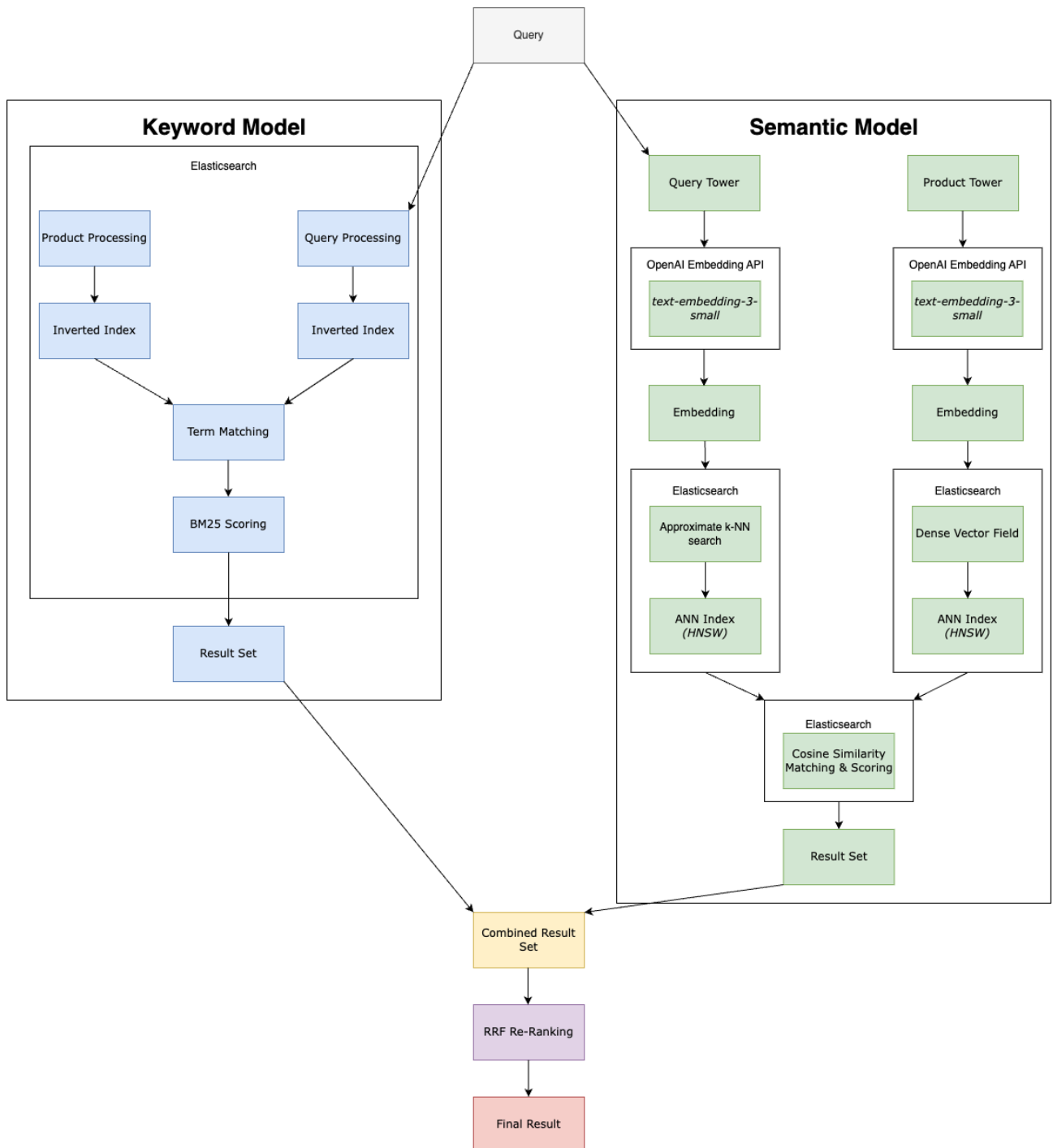
Fig. 1. Architecture of proposed solution.

3.2 Comparison with existing similar functionality methods /algorithms/ techniques

The proposed model builds upon previous studies discussed in this narrative review and is significantly influenced by the hybrid approach utilized in Elasticsearch [29], with custom implementations of several components to provide greater control and minimize costs. OpenSearch, a fork of Elasticsearch, also adopts a similar approach [30, 31].

3.3   Strengths and Limitations

The proposed model offers a strong foundation while also being flexible and allowing for further fine-tuning for different needs. Its strengths are its simplicity and cost-effectiveness, leveraging widely available technologies for seamless integration with existing systems, without requiring substantial resources or specialized training. Additionally, it's based on architectures in this literature review which has proved significant success in improving product search results, and it's also based on approaches recommended by popular open-source search engines.

While the proposed model offers a strong foundation, it may not be suitable for advanced use cases or environments that require highly specialized features or large-scale scalability. Additionally, the absence of empirical experiments on this specific model means its efficacy remains unverified in practical settings. It should be evaluated in offline tests, online human assessments and A/B testing, utilizing standard metrics like Recall, MRR, MAP, and NDCG.

4 Research Methodology

This section outlines the structured approach used to explore the core research areas related to hybrid product search systems in online marketplaces. It covers the search strategy, including the inclusion and exclusion criteria, as well as the development of keywords and search strings. The chosen research databases and tools are explained, along with the rationale for their selection. The collected literature is organized to provide a clearer overview of the field. Finally, the limitations and challenges encountered during the research process are discussed, followed by a summary of the methodology.

4.1 Procedures / Search Strategy

Begin by dividing the research topic into four core areas: the importance of online marketplaces and specifically the product search functionality, keyword search, semantic search and hybrid search. This approach ensures a structured focus on each key component of the study.

For each of the four topic areas, identify relevant keywords that capture the main concepts. Use existing information and knowledge to gather the most appropriate terms. Document synonyms and related terms for each keyword to allow for a broader search.
Using a variety of related terms ensures that the search will capture a wide range of relevant documents, increasing the chances of finding results.

Combine the keywords and synonyms into search strings using Boolean operators to fine-tune the search results. Test different combinations of these keywords in selected databases.
Testing multiple combinations ensures that the search captures variations in terminology and increases the likelihood of relevant results.

Scan the initial search results quickly by reviewing titles, abstracts, and metadata to filter out irrelevant documents. Select documents that directly relate to the research question or address the core themes. Consider the relevance of dates to factor in new discoveries and look into the authors and organizations behind the resources to evaluate potential bias. Also look into the references and "cited by"-sections of the selected documents to potentially find more relevant documents.
This step helps in narrowing down the collection of documents, ensuring only the most relevant are taken into consideration.

Organize the selected documents by categorizing them according to their core ideas. Create a map to visualize which key areas are covered, to gain a clear understanding of the research landscape.
Mapping ensures that all areas of the research are adequately covered and helps identify any gaps in the literature.

After mapping, read and understand the relevant sections of the selected documents in greater detail to extract useful information for the study. Focus on sections that address the strengths, weaknesses, and use cases of each search approach (keyword, semantic, and hybrid).
This allows for a focused, efficient review, ensuring that only the most relevant sections are studied deeply.

Evaluate the extracted information using critical thinking to assess if the claims are true and if the contributions are valuable. Generate new ideas from the selected information to gather an oversight on what can be improved in the area. Such as if the author's claims can be supported better, if there are any gaps or unsolved problems, if there are alternative approaches and so on.

4.1.1 Inclusion & Exclusion Criteria

Specific criteria are applied to evaluate the relevance of the papers for review.

Inclusion criteria:
- The paper must be published in English to ensure accessibility and comprehension.
- Only papers published from 2019 onward are considered to ensure the inclusion of research utilizing modern techniques, except for foundational studies, which are included regardless of their publication date.
- The paper should focus on hybrid product search systems within online marketplaces. Research on other information retrieval areas may also be included if relevant to gather additional data.
  - Broad Area: Information Retrieval
  - Narrowed Area: E-commerce product search
  - Specific Focus within the Narrowed Area: Hybrid product search systems in online marketplaces
- The paper should propose a hybrid model for product search.
- The paper should include experiments evaluating the performance of the hybrid model, ideally comparing it with a standalone keyword-based model.
- The paper should be peer-reviewed, however very recent papers are exempt from this.

Exclusion criteria:
- Studies focused on virtual assistants, chatbots, or other conversational interfaces.
- Papers for which the full text is not accessible.

4.1.2 Keywords

Keywords are documented based on the above criteria to streamline the search process.

Keywords to include:
- **Online Marketplace** (e-marketplace, buy and sell platform)
- **E-commerce** (online shopping/retail/platform, digital economy)
- **Information Retrieval** (IR)
  - Search (retrieval, fetch, find, explore, discover, matching)
- **Product Search** (product retrieval, item search, search function)
- **Keyword Search** (lexical search, syntactical search, full-text search)
  - Keyword (token, term)
  - Term match, Boolean match, Exact match.
- **Semantic Search** (Vector search, vector-based search, Embedding search, Embedding-based retrieval, Neural search, AI search)
- **Hybrid Search** (Combined search, Multi-search, Merge-search, Unified search)

Keywords to exclude:
- **Conversational***:*
  - Q&A
  - Assistant
  - Chatbot

4.1.3 Search Strings

The keywords are combined to form advanced search strings using Boolean operators. The best search strings, which provided most relevant results, are the following:

- *(Online marketplace OR e-commerce) AND (Hybrid OR combined) AND (Keyword OR lexical) AND (Semantic OR vector OR embedding) AND (Search OR retrieval)*
- *Hybrid semantic product search*

See Appendix 1 for additional search strings and further details.

4.2   Tools/Sources/Databases/

A diverse range of research databases were evaluated to identify the most appropriate ones for this study. The selected research databases are detailed in Table II.

TABLE II
SELECTED RESEARCH DATABASES

Name	Type	Description
Google Scholar	Metadata	Google Scholar was selected due to its broad coverage and its ability to index a wide variety of sources, including articles, books, and conference papers. It has a user-friendly interface which makes it great for collecting a diverse range of relevant literature. Additionally, it has been recommended as a more appropriate tool for assessing computer science research compared to traditional citation databases like ISI Web of Science, which has been criticized for insufficient coverage of key CS publications [32].
Semantic Scholar	Metadata	Semantic Scholar was deemed suitable for this research due to its AI-powered capabilities which extract meaning from papers and surface it to the UI in different accessible ways, such as the dynamic Q&A tool [33]. It's especially useful in the beginning of the research to be able to scan resources efficiently.
IEEE Xplore	Full-text	IEEE Xplore was selected due to its large library, which contains nearly one-third of the world's current literature in computer science, systems engineering, and other related fields [34]. Given its extensive collection of peer-reviewed articles and conference papers in areas such as computer science, engineering, and technology, IEEE Xplore was selected as one of the primary full-text databases for this research.
ACM Digital Library	Full-text	The ACM Digital Library was included due to its wide coverage of computer science topics. Just like Google Scholar, it was also mentioned as a suitable alternative to ISI Web of Science for computer science research [32]. Another factor in the decision was ACM's Code of Ethics and Professional Conduct [35], along with its enforcement procedures [36], which ensures the information used in this research comes from a trustworthy platform.
Springer Link	Full-text	Springer Link was also included for its diverse range of scientific publications within computer science and similar fields. The database

DT166G, Presentation av ny Teknik
Integrating Semantic with Keyword Product search in Online
Marketplaces

		provides access to some full-text articles, books and conference proceedings.
ArXiv	Full-text	ArXiv was chosen as a primary database due to its vast collection of open-access preprints and research papers in computer science. It offers access to recent papers, providing timely insights into the latest advancements in the field.

Other search engines and databases, such as PubMed, were not included because their focus on life sciences was less relevant to e-commerce search technologies. ProQuest and Scopus were considered, but their overlap with the selected databases made them unnecessary for this study.

Some Swedish databases were also considered, such as DiVA and Libris, however an initial scan did not provide relevant results to this research.

DuckDuckGo and Google were selected to perform web searches. While Google is highly effective for finding a wide range of research sources, its ads and personalized results can obscure unbiased information and limit perspective [37, 38]. In contrast, DuckDuckGo avoids ads and tracking, promoting a more neutral search experience but with a smaller index and fewer advanced search options [39].

4.3   Organizing Literature

To provide a structured understanding of this body of work, we categorize the literature based on the type of platform where the hybrid models have been implemented and tested. This categorization results in three distinct groups: online marketplaces, e-commerce platforms, and other platforms. The relationships between the studies are highlighted through their references and citations, as detailed in the accompanying tables.

Research shown in Table III focuses on large-scale online marketplaces where the search experience is important for connecting users with products. These studies often reference each other, indicating a collaborative effort to refine hybrid models for marketplace environments.

TABLE III
RESEARCH CONDUCTED ON ONLINE MARKETPLACES

No.	Title	Authors	Year	Platform	References	Referenced By
[9]	Semantic Product Search	P. Nigam, Y. Song, V. Mohan, V. Lakshman, W. Ding, A. Shingavi, C. H. Teo, H. Gu, and B. Yin,	2019	Amazon Marketplace	-	[20, 3, 13, 14, 19, 22, 17]

[20]	Extreme Multi-label Learning for Semantic Matching in Product Search.	Wei-Cheng Chang, Daniel Jiang, Hsiang-Fu Yu, Choon Hui Teo, Jiong Zhang, Kai Zhong, Kedarnath Kolluri, Qie Hu, Nikhil Shandilya, Vyacheslav Ievgrafov, Japinder Singh, and Inderjit S. Dhillon.	2021	Amazon Marketplace	Compares its hybrid model to the hybrid model proposed in earlier research at Amazon [9]. The study also references [15] as an example of using a BM25-based keyword model which struggles with semantic-level matching. A similar keyword model was further used as a baseline in the experiments, demonstrating poor performance on a semantic product search dataset.	-
[21]	Que2Search: Fast and Accurate Query and Document Understanding for Search at Facebook.	Yiqun Liu, Kaushik Rangadurai, Yunzhong He, Siddarth Malreddy, Xunlong Gui, Xiaoyi Liu, and Fedor Borisyuk	2021	Facebook Marketplace	Builds upon a previous study conducted on Facebook's search engine [18], and extends it to Facebook Marketplace.	[3, 13, 22]
[14]	Embedding-based Product Retrieval in Taobao search	Sen Li, Fuyu Lv, Taiwei Jin, Guli Lin, Keping Yang, Xiaoyi Zeng, Xiao-Ming Wu, and Qianli Ma	2021	Taobao Marketplace	Makes several comparisons to previous research in Facebook Search [18], specifically noting the difference in semantic model use: Facebook complements the keyword model with it for fuzzy matching, while this study uses the keyword model to enhance the relevance of the semantic model. It critiques [9] and [11] for not discussing the low controllability of search relevance in semantic models, compared to keyword models.	[13, 22]

[3]	Semantic Retrieval at Walmart	A. Magnani, F. Liu, S. Chaidaroon, S. Yadav, P. R. Suram, A. Puthenputhussery, S. Chen, M. Xie, A. Kashi, T. Lee, and C. Liao	2022	Walmart Marketplace	Compares their approach of separating the models against the residual-based learning model used in [15]. Mentions similarities with the keyword and semantic integration used in [9]. Uses an approach mentioned in [21] for training the semantic model.	[13, 22]
[13]	Enhancing Relevance of Embedding-based Retrieval at Walmart	J. Lin, S. Yadav, F. Liu, N. Rossi, P. R. Suram, S. Chembolu	2024	Walmart Marketplace	Builds upon research in [3] to improve the hybrid model, where several authors from the previous study are also involved in this research. [9, 11, 14] are cited as examples of successfully integrating embedding-based semantic models into e-commerce search, with positive results. Uses [15, 21] to support the claim that downstream re-ranking systems may not always accurately rank results from semantic models, resulting in irrelevant products shown to the users. Furthermore, this is used to justify the focus on improving relevance performance in the semantic model. Utilizes an approach mentioned in [21] for training the semantic model.	-

Table IV includes research conducted on specialized e-commerce platforms where product catalogs and user queries present unique challenges. The studies often reference methodologies from online marketplace research but adapt them to fit the specific context of their e-commerce platforms.

TABLE IV
RESEARCH CONDUCTED ON E-COMMERCE PLATFORMS

No.	Title	Authors	Year	Platform	References	Referenced By

No.	Title	Authors	Year	Platform / Type	References	Referenced By
[11]	Towards personalized and semantic retrieval: An end-to-end solution for e-commerce search via embedding learning	H. Zhang, S. Wang, K. Zhang, Z. Tang, Y. Jiang, Y. Xiao, and W. Y. Yang	2020	JD Sports	-	[13, 14, 22]
[19]	Semantic Product Search for Matching Structured Product Catalogs in E-Commerce	J. I. Choi, S. Kallumadi, B. Mitra, E. Agichtein, and F. Javed	2020	Home Depot	Uses [9] as a reference in the introduction to explain common e-commerce search systems.	-
[22]	Embedding based retrieval for long tail search queries in ecommerce	Akshay Kekuda, Yuyang Zhang, and Arun Udayashankar	2024	Best Buy	The training dataset in the study was inspired by [21]. References [3, 9, 14] as examples of semantic models trained on clickstream data. However, this study avoided this approach because it overlooks low-engagement products, as it mainly captures user interactions, underrepresenting products with fewer clicks. This study acknowledges the usefulness of high-quality human-annotated data as described in [11], but avoids it due to the high cost.	-

Studies in Table V provide foundational theories and methods that are applicable across various platforms. They often serve as references for research in the more specialized categories (online marketplaces and e-commerce platforms).

TABLE V
RESEARCH CONDUCTED ON OTHER PLATFORMS

No.	Title	Authors	Year	Platform / Type	References	Referenced By

[17]	Leveraging Semantic and Lexical Matching to Improve the Recall of Document Retrieval Systems: A Hybrid Approach	S. Kuzi, M. Zhang, C. Li, M. Bendersky, and M. Najork	2020	News articles search	Cites [25] in the related work section to highlight that semantic models outperform keyword models in the retrieval stage of product search.	[15]
[15]	Complement Lexical Retrieval Model with Semantic Residual Embeddings	L. Gao, Z. Dai, T. Chen, Z. Fan, B. Van Durme, and J. Callan	2021	General information retrieval	References [17] as recent research that explores hybrid keyword/semantic models.	[3, 13, 20]
[18]	Embedding-based Retrieval in Facebook Search	Jui-Ting Huang, Ashish Sharma, Shuying Sun, Li Xia, David Zhang, Philip Pronin, Janani Padmanabhan, Giuseppe Ottaviano, and Linjun Yang	2020	Facebook Search	-	[3, 21, 14]

The categories and relationships between the literature are illustrated in Fig. 2, offering an overview of how the papers are interconnected. The graph represents the relationships using directed edges, where arrows point from the referencing paper to the referenced paper. The nodes are numbered based on their number in the reference list of this study, with platform types differentiated by color.
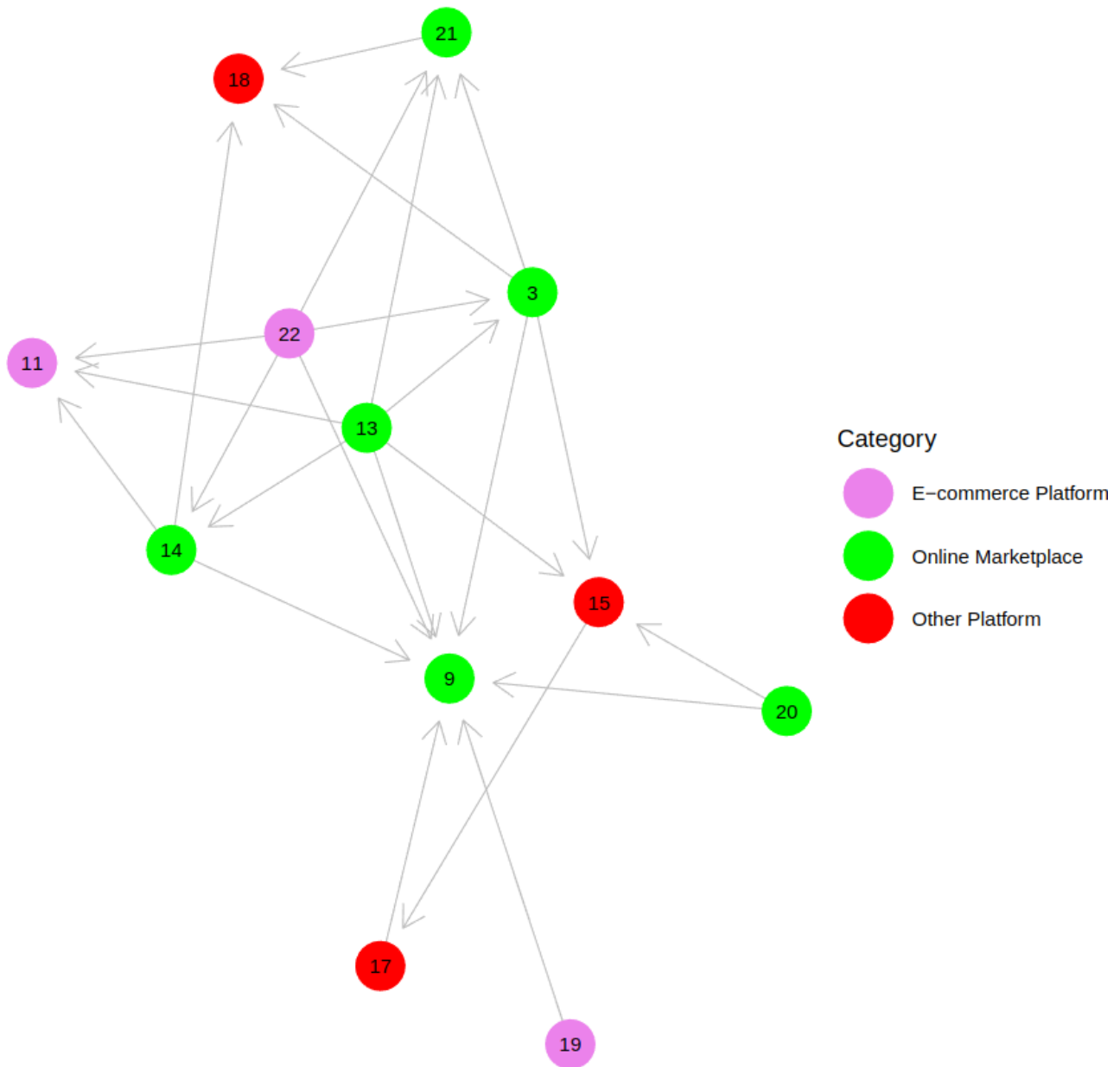
Fig. 2. Literature relationship graph categorized by platform type

4.4 Limitations/Emerged Obstacles

The field uses a broad range of terminology, leading to potential missed matches, requiring the use of multiple keywords during the search process. This made it difficult to find papers within the "Specific Focus within the Narrowed Area" inclusion criteria, as described earlier.

Most papers placed excessive emphasis on the semantic model, which limited the ability to gather comprehensive information about the overall hybrid architecture and made it challenging to fairly assess the experiments.

4.5 Summary of Procedures

The research has been systematically divided into four core areas: online marketplaces, keyword search, semantic search, and hybrid search. Keywords and synonyms were identified to broaden the search scope.

Boolean operators were used to construct search strings, which were tested across various databases. Initial results were scanned quickly through titles, abstracts, and introductions to select relevant documents. The initial set of relevant documents were then further skimmed to narrow the results down further. The final set of relevant documents were thoroughly reviewed to gain a comprehensive understanding of the content. Additionally, the documents were categorized by platform type and publication year, and the relationships between them were mapped to provide a clearer overview of advancements in the field.

5   Results and Discussion

To address the research questions of this literature review, twelve studies were analyzed to explore the architecture of hybrid models, and their impact on product search result quality. As shown in Fig. 3, six of these studies focused on online marketplaces, and three were centered on other types of e-commerce platforms, both focusing on product search. Additionally, three studies investigated platforms outside of e-commerce and online marketplaces, but remained relevant as foundational research applied in product search-focused studies within the review.
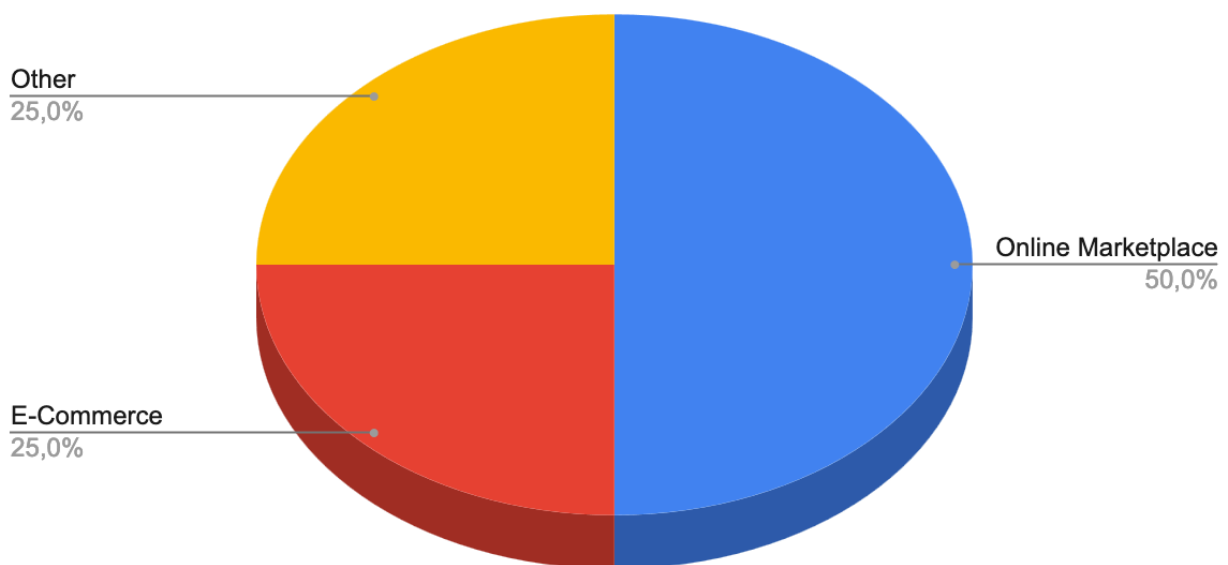
Fig. 3: Distribution of reviewed studies.

The reviewed studies and previous research highlight the complexity of product search in online marketplaces that requires systems capable of managing both precision and context, as neither keyword nor semantic models alone can consistently meet users' diverse intents. This is demonstrated by several offline and online experiments in the reviewed studies. Given these complexities, the need for hybrid models becomes apparent.

While different hybrid models were proposed in the reviewed studies, a consistent pattern emerged among them, as shown in Fig. 4. Most hybrid architectures analyzed rely on a keyword model as the core retrieval mechanism, utilizing an inverted index and BM25 scoring. Semantic models are often integrated using a two-tower architecture with a BERT-based embedding model. ANN search is employed for faster retrieval, with cosine similarity serving as the main matching and scoring method. After retrieving results from both the keyword and semantic models, merging and re-ranking mechanisms are applied to refine the final output, often applying machine learning models. However, most studies primarily focused on the matching phase, providing minimal information on the re-ranking phase, making it difficult to provide further details.
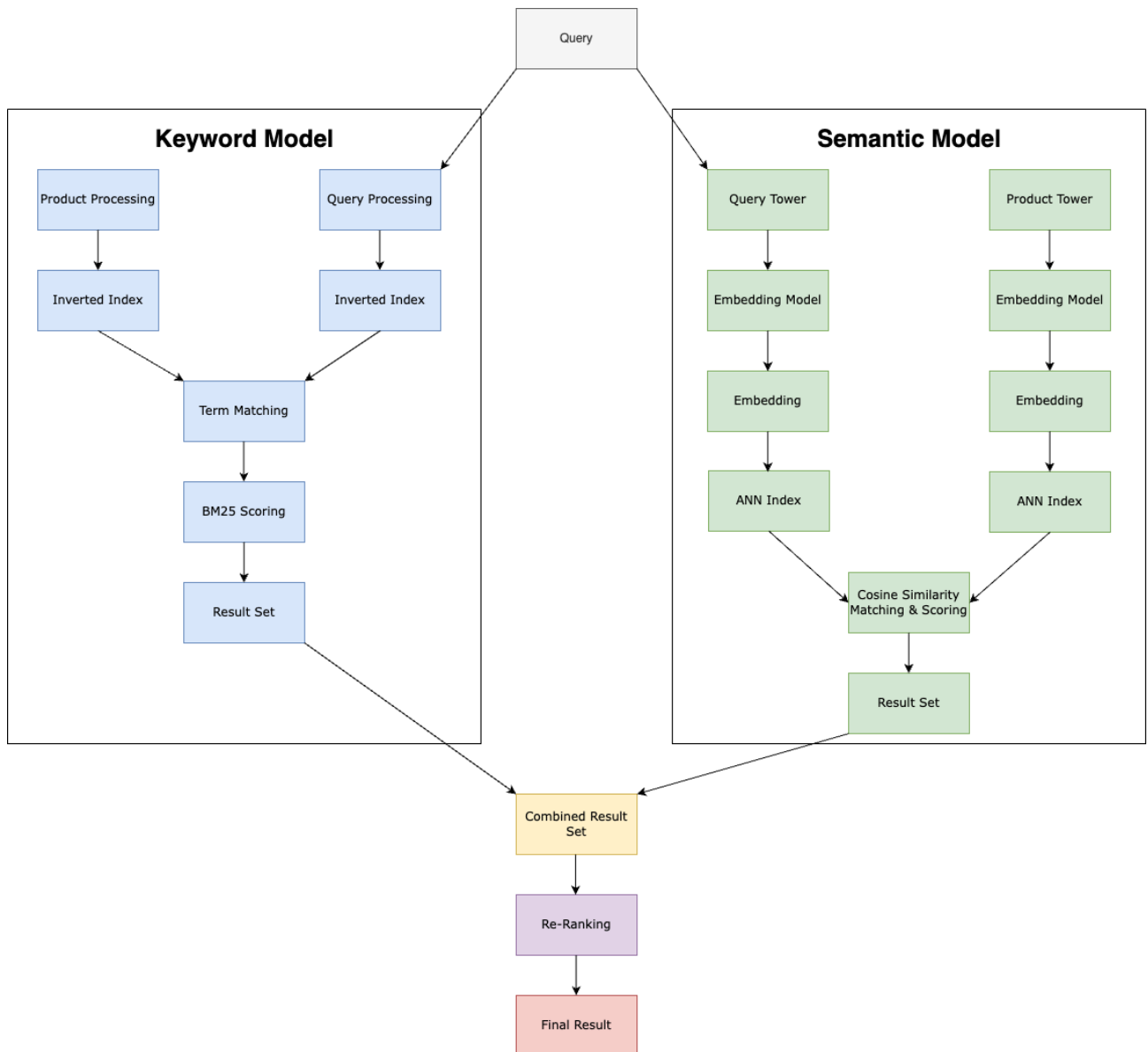
Fig. 4: Common hybrid model architecture.

To evaluate the performance of the hybrid models, the reviewed studies conducted several offline and online experiments, comparing them to keyword and semantic models in isolation. The authors presented similar results, showing that their proposed hybrid models significantly improved product search results across various metrics, with the primary metrics illustrated in Fig. 5. These hybrid models particularly demonstrated improvements in handling complex and long-tail queries, and successfully addressed limitations of traditional keyword-based models. As shown in Fig. 6, the majority of proposed hybrid models were deployed into production environments, supporting millions of users and searches daily. This underscores the practical, real-world applicability of integrating semantic models with keyword-based product search systems in online marketplaces.
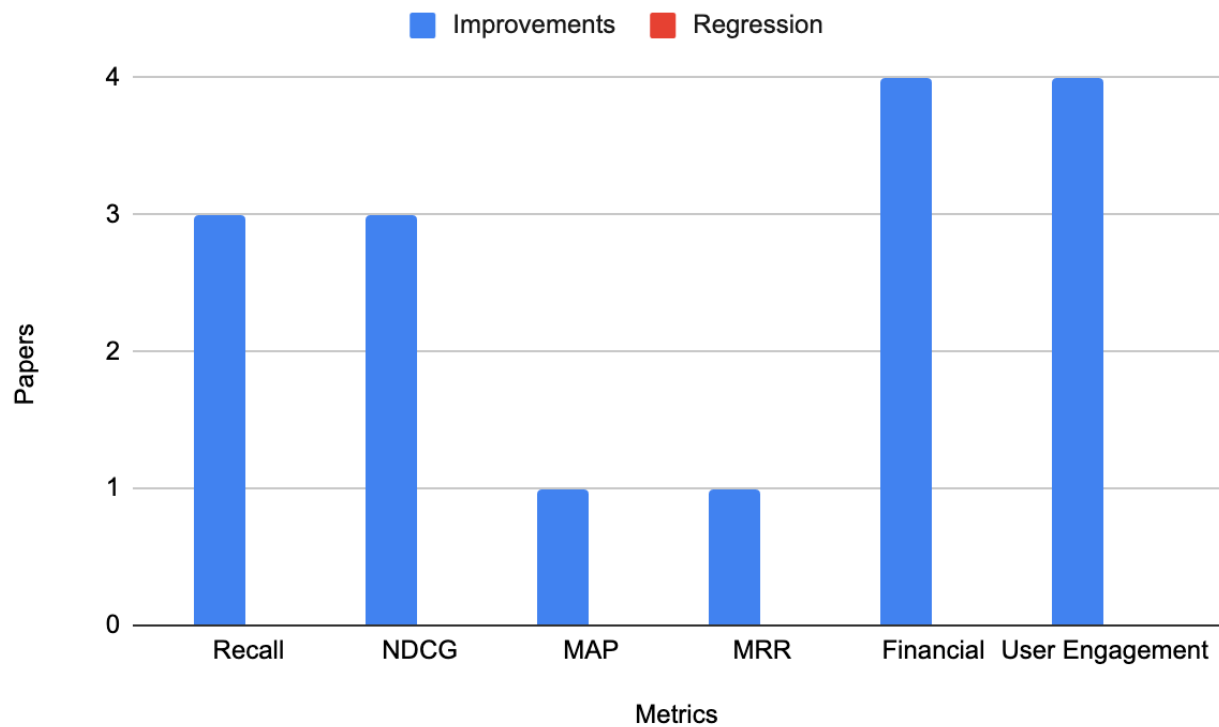
Fig. 5: Reported improvements and regressions in core metrics, focusing exclusively on experiments conducted within online marketplaces and e-commerce platforms.
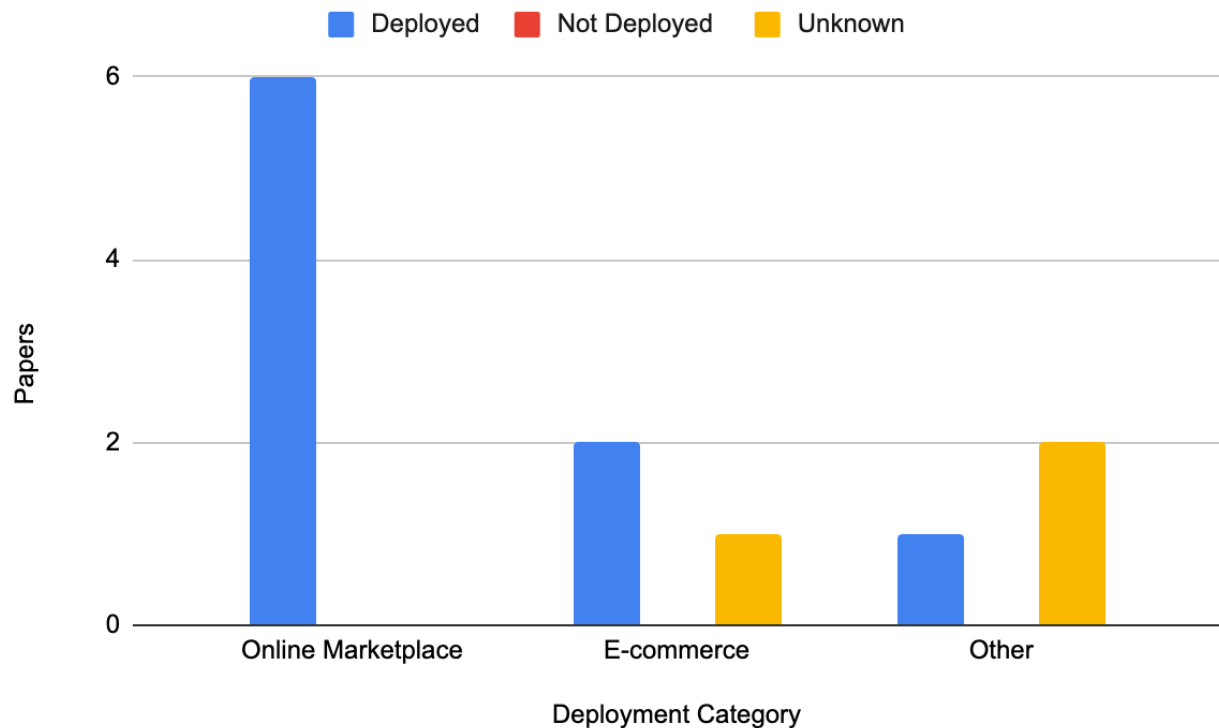
Fig. 6: Deployment of Hybrid Search Models in Real-World Environments.

5.1 Goal Fulfillment

The literature review has successfully fulfilled its purpose by exploring existing research on hybrid product search systems to determine its impact on product search results. The review compared various approaches of integrating semantic models with keyword-based models into hybrid models. The hybrid models were compared with other standalone models, in a wide variety of offline and online tests, using a range of relevant metrics. The results from these tests were compared and analyzed to form a conclusion on the impact of hybrid search systems on product search result quality in online marketplaces.

5.2 Analysis of Methodology

The three studies conducted outside of online marketplaces and e-commerce were included to provide additional data on hybrid models and to support further product search research in this review, even though they themselves did not focus specifically on product search. Although the review consistently distinguishes between these studies, this differentiation may introduce some confusion for the reader and shift attention away from product search. However, gathering more data was considered necessary to gain a better understanding of hybrid models and was also thought to provide insights applicable to product search scenarios.

Some hybrid architectures included additional matching components, such as behavioral data, which could divert focus from the complementary relationship between keyword and semantic models. Despite this, these studies still demonstrate the impact of integrating semantic models into keyword-dominant search systems. Most of the selected studies focused primarily on the semantic model within the hybrid architecture, providing limited details about the keyword model and re-ranking processes, which could have offered valuable insights. Additionally, our limited knowledge in the area of semantic retrieval and machine learning negatively impacts the discussion of hybrid architectures. Nevertheless, this section was included to provide an overview and facilitate a better understanding of the experiments evaluating hybrid model performance, which was the primary goal of the review.

The experiments differed across the reviewed studies, with a range of metrics employed, varying levels of detail provided, and overall inconsistent quality. This variability made it difficult to pinpoint and categorize common factors in which the hybrid model excels or underperforms.

The obstacles discussed in Section 4.4 could have been handled more efficiently from the start. Documenting synonyms and related terms helped in handling the wide variety in terminology, but using a tool designed for this purpose would have made the process more efficient.
Earlier exploration of the references and "cited by" sections of relevant papers might have simplified the search for papers within the narrowed area, yet even with this approach used later on, identifying enough relevant studies remained difficult. This led to the necessity of employing a regressive search strategy to gather additional data from a broader area.

6   Conclusions and Future Recommendations

This literature review has offered a thorough overview of the impact of hybrid models on product search results in online marketplaces. Experimental findings consistently demonstrated that hybrid models significantly improved product search result quality, compared to standalone keyword and semantic models, as measured across various standard metrics.

While the review offered in-depth insights on its primary focus—product search result quality—it deliberately omitted several other important aspects of product search systems. To fully evaluate these systems, future research should consider how hybrid models impact complexity, maintenance, resource usage, costs, latency and other factors. For example, a largely unexplored challenge is handling pagination when merging results from two separate candidate sets, a method frequently employed by the hybrid models in this review. Though the primary focus of the review was on search result quality, it also examined hybrid architectures, pointing out both shared foundations and distinct features. However, it lacked sufficient detail on the crucial re-ranking stage and deliberately omitted comparisons that could help identify the best-performing model. Future research should place greater emphasis on hybrid architectures, especially the re-ranking stage, and include comparisons within the context of online marketplaces.

This assignment has broadened our understanding of product search challenges and expanded our knowledge in IR and the metrics used to evaluate IR models. It has deepened our insights into keyword- and semantic-based models, as well as the implementation of hybrid architectures. Additionally, we have gained a clearer view of where hybrid models can effectively address product search challenges and observed their impact in real-world applications. Furthermore, we have learned to define an engaging topic and conduct a literature review on relevant research. We gained insight into the structure of research papers and developed the ability to understand and critically assess the literature. Finally, we improved our ability to craft a well-structured report with a clear and cohesive flow.

References

[1] Department of Commerce - United States of America. "2024 eCommerce Size & Sales Forecast", 2024. [Online]
Available at: https://www.trade.gov/ecommerce-sales-size-forecast
[Accessed: 13th of September 2024]

[2] Department of Commerce - United States of America. "Understanding eCommerce Marketplaces", date unknown. [Online]
Available at: https://www.trade.gov/understanding-ecommerce-marketplaces
[Accessed: 13th of September 2024]

[3] A. Magnani, F. Liu, S. Chaidaroon, S. Yadav, P. R. Suram, A. Puthenputhussery, S. Chen, M. Xie, A. Kashi, T. Lee, and C. Liao, "Semantic Retrieval at Walmart," in Proc. 28th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD '22), New York, NY, USA, 2022, pp. 3495–3503. doi: 10.1145/3534678.3539164.

[4] N. Choudhary, N. Rao, S. Katariya, K. Subbian, and C. K. Reddy, "ANTHEM: Attentive Hyperbolic Entity Model for Product Search," in Proc. 15th ACM Int. Conf. Web Search and Data Mining (WSDM '22), New York, NY, USA, 2022, pp. 161–171. doi: 10.1145/3488560.3498456.

[5] Google Cloud, "New research: Search abandonment continues to vex retailers worldwide", 2023. [Online]
Available at: https://cloud.google.com/blog/topics/retail/new-research-on-search-abandonment-in-retail
[Accessed: 12th September 2024]

[6] C. Luo, X. Tang, H. Lu, Y. Xie, H. Liu, Z. Dai, L. Cui, A. Joshi, S. Nag, Y. Li, Z. Li, R. Goutam, J. Tang, H. Zhang, and Q. He, "Exploring Query Understanding for Amazon Product Search," arXiv, Aug. 2024.
 doi: 10.48550/arXiv.2408.02215.

[7] A. Ahuja, N. Rao, S. Katariya, K. Subbian, and C. K. Reddy, "Language-Agnostic Representation Learning for Product Search on E-Commerce Platforms," in Proc. 13th Int. Conf. Web Search and Data Mining (WSDM '20), New York, NY, USA, 2020, pp. 7-15. https://doi.org/10.1145/3336191.3371852.

[8] Constructor. "State of Ecommerce Search and Product Discovery 2024", 2024. [Online]
Available at:
https://info.constructor.io/hubfs/2024-state-of-ecommerce.pdf?utm_campaign=Content_FY2024Q3_Survey_Report&utm_medium=email&_hsenc=p2ANqtz-8hsHbAuCP-kpo5SYY3DQgeU11r2EyfKtVSGsSqsewqmR8hALitkSZ4UoE8t8n_IeUOAedFFJUgb6yXHGF31MwsUU-aZw&_hsmi=324814027&utm_content=324814027&utm_source=hs_automation&hsCtaTracking=47a42651-9863-4f21-8e4a-b7f973d8557f%7Ced6d2e4b-6383-4cfa-a0b9-7f9741cfb395
[Accessed: 22nd of october 2024]

[9] P. Nigam, Y. Song, V. Mohan, V. Lakshman, W. Ding, A. Shingavi, C. H. Teo, H. Gu, and B. Yin, "Semantic Product Search," in Proc. 25th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD '19), Anchorage, AK, USA, Aug. 2019, pp. 2876-2885. doi: 10.1145/3292500.3330759.

[10] T. Wimmer and M. Scholz, "Online Product Descriptions – Boost for your Sales?," in Proc. 14th Int. Conf. Wirtschaftsinformatik, Siegen, Germany, Feb. 2019, pp. 498-512.

[11] H. Zhang, S. Wang, K. Zhang, Z. Tang, Y. Jiang, Y. Xiao, and W. Y. Yang, "Towards personalized and semantic retrieval: An end-to-end solution for e-commerce search via embedding learning," in Proc. 43rd Int. ACM SIGIR Conf. Research and Development in Information Retrieval, Xi'an, China, Jul. 2020, pp. 2407-2416. doi: 10.1145/3397271.3401446.

[12] M. Jbene, S. Tigani, R. Saadane, and A. Chehri, "Deep Neural Network and Boosting Based Hybrid Quality Ranking for e-Commerce Product Search," Big Data Cogn. Comput., vol. 5, no. 3, p. 35, 2021.
https://doi.org/10.3390/bdcc5030035.

[13] J. Lin, S. Yadav, F. Liu, N. Rossi, P. R. Suram, S. Chembolu, and C. Liao, "Enhancing Relevance of Embedding-based Retrieval at Walmart," arXiv, 2024. doi: 10.48550/arXiv.2408.04884.

DT166G, Presentation av ny Teknik
*Integrating Semantic with Keyword Product search in Online
Marketplaces*

[14] S. Li, F. Lv, T. Jin, G. Lin, K. Yang, X. Zeng, X.-M. Wu, and Q. Ma, "Embedding-based Product Retrieval in Taobao Search," in Proc. 27th ACM SIGKDD Conf. Knowledge Discovery & Data Mining (KDD '21), New York, NY, USA, 2021, pp. 3181-3189. https://doi.org/10.1145/3447548.3467101.

[15] L. Gao, Z. Dai, T. Chen, Z. Fan, B. Van Durme, and J. Callan, "Complement Lexical Retrieval Model with Semantic Residual Embeddings," in Advances in Information Retrieval, D. Hiemstra, M. F. Moens, J. Mothe, R. Perego, M. Potthast, and F. Sebastiani, Eds., vol. 12656 of Lecture Notes in Computer Science, Cham, Switzerland: Springer, 2021, pp. 144-157. doi: 10.1007/978-3-030-72113-8_10.

[16] Elastic Search. "Search relevance tuning: Balancing keyword and semantic search", 2024. [Online] Available at: https://www.elastic.co/search-labs/blog/search-relevance-tuning-in-semantic-search [Accessed: 13th of September 2024]

[17] S. Kuzi, M. Zhang, C. Li, M. Bendersky, and M. Najork, "Leveraging Semantic and Lexical Matching to Improve the Recall of Document Retrieval Systems: A Hybrid Approach," arXiv, 2020. doi: 10.48550/arXiv.2010.01195.

[18] J.-T. Huang, A. Sharma, S. Sun, L. Xia, D. Zhang, P. Pronin, J. Padmanabhan, G. Ottaviano, and L. Yang, "Embedding-based Retrieval in Facebook Search," in Proc. 26th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining (KDD '20), New York, NY, USA, 2020, pp. 2553-2561. https://doi.org/10.1145/3394486.3403305.

[19] J. I. Choi, S. Kallumadi, B. Mitra, E. Agichtein, and F. Javed, "Semantic Product Search for Matching Structured Product Catalogs in E-Commerce," arXiv, 2020. doi: 10.48550/arXiv.2008.08180.

[20] W.-C. Chang, D. Jiang, H.-F. Yu, C. H. Teo, J. Zhang, K. Zhong, K. Kolluri, Q. Hu, N. Shandilya, V. Ievgrafov, J. Singh, and I. S. Dhillon, "Extreme Multi-label Learning for Semantic Matching in Product Search," in Proc. 27th ACM SIGKDD Conf. Knowledge Discovery & Data Mining (KDD '21), New York, NY, USA, 2021, pp. 2643-2651. https://doi.org/10.1145/3447548.3467092.

[21] Y. Liu, K. Rangadurai, Y. He, S. Malreddy, X. Gui, X. Liu, and F. Borisyuk, "Que2Search: Fast and Accurate Query and Document Understanding for Search at Facebook," in Proc. 27th ACM SIGKDD Conf. Knowledge Discovery & Data Mining (KDD '21), New York, NY, USA, 2021, pp. 3376-3384. https://doi.org/10.1145/3447548.3467127.

[22] A. Kekuda, Y. Zhang, and A. Udayashankar, "Embedding based retrieval for long tail search queries in ecommerce," in Proc. 18th ACM Conf. on Recommender Systems (RecSys '24), New York, NY, USA, 2024, pp. 771-774. https://doi.org/10.1145/3640457.3688039.

[23] Albin Rönnkvist / Roberto Piran Amedi, "Course_DT166G_PresentingNewTechnology". GitHub repository, [Online]. Available: https://github.com/albinronnkvist/Course_DT166G_PresentingNewTechnology. [Accessed: 11th of October 2024]

[24] Y. A. Malkov and D. A. Yashunin, "Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 4, pp. 824-836, 1 April 2020, doi: 10.1109/TPAMI.2018.2889473.
keywords: {Routing;Complexity theory;Search problems;Data models;Approximation algorithms;Biological system modeling;Brain modeling;Graph and tree search strategies;artificial intelligence;information search and retrieval;information storage and retrieval;information technology and systems;search process;graphs and networks;data structures;nearest neighbor search;big data;approximate search;similarity search},

[25] OpenAI Platform. "Vector Embeddings FAQ", n.d. [Online] Available at: https://platform.openai.com/docs/guides/embeddings/faq [Accessed: 11th of October 2024]

[26] G. V. Cormack, C. L. A. Clarke, and S. Büttcher, "Reciprocal Rank Fusion outperforms Condorcet and individual Rank Learning Methods," in Proc. 32nd ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR '09), Boston, MA, USA, 2009, pp. 3495-3503. doi: 10.1145/3534678.3539164.

[27] Elastic Search Labs. "Improving information retrieval in the Elastic Stack: Hybrid retrieval" 2023. [Online] Available at: https://www.elastic.co/search-labs/blog/improving-information-retrieval-elastic-stack-hybrid [Accessed: 11th of October 2024]

[28]   Krystian Safjan's Blog. "Rank Fusion Algorithms - From Simple to Advanced". 2023. [Online]
Available at: https://safjan.com/Rank-fusion-algorithms-from-simple-to-advanced/
[Accessed: 11th of October 2024]

[29]   Elastic Search Labs. "Hybrid Search: Combined Full-Text and kNN Results". n.d. [Online]
Available at: https://www.elastic.co/search-labs/tutorials/search-tutorial/vector-search/hybrid-search
[Accessed: 11th of October 2024]

[30]   OpenSearch. "Hybrid Search". n.d. [Online]
Available at: https://opensearch.org/docs/latest/search-plugins/hybrid-search/
[Accessed: 11th of October 2024]

[31]   OpenSearch. "Neural Search Tutorial". n.d. [Online]
Available at: https://opensearch.org/docs/latest/search-plugins/neural-search-tutorial/
[Accessed: 11th of October 2024]

[32]   B. Meyer, C. Choppy, J. Staunstrup, and J. Van Leeuwen, "Research Evaluation for Computer Science," Commun.
ACM, vol. 52, no. 4, pp. 31–34, Apr. 2009, doi: 10.1145/1498765.1498780.

[33]   Semantic Scholar FAQ: "What is the advantage of using Semantic Scholar instead of other academic search engines?",
date unknown.
Available at: https://www.semanticscholar.org/faq#advantages
[Accessed: 13th of September 2024]

[34]   IEEE, "IEEE Xplore digital Library: 2021 Subscription Options.", 2021. [PDF]
Available At:
https://innovate.ieee.org/wp-content/uploads/2021/01/2021-Subscription-Options-Catalog-Product-Listing.pdf
[Accessed: 12th September 2024]

[35]   ACM Digital Library, " ACM Code of Ethics and Professional Conduct", 2018. [Online]
Available at: https://www.acm.org/code-of-ethics
[Accessed: 12th September 2024]

[36]   ACM Digital Library, " Association for Computing Machinery Code of Ethics and Professional Conduct Enforcement
Policy/Procedure", 2018. [Online]
Available at: https://www.acm.org/code-of-ethics/enforcement-procedures
[Accessed: 12th September 2024]

[37]   Google Search. "How ads work on Google search", date unknown. [Online]
Available at: https://www.google.com/search/howsearchworks/our-approach/ads-on-search/
[Accessed: 15th September 2024]

[38]   Google Search. "Personalization & Google Search results", date unknown. [Online]
Available at: https://support.google.com/websearch/answer/12410098?hl=en
[Accessed: 15th September 2024]

[39]   DuckDuckGo Help Pages. "DuckDuckGo Web tracking Protections", 2024.
Available at: https://duckduckgo.com/duckduckgo-help-pages/privacy/web-tracking-protections/
[Accessed: 15th September 2024]

[40]   O. Espejel, "Getting Started With Embeddings," Hugging Face, Jun. 23, 2022. [Online]. Available:
https://huggingface.co/blog/getting-started-with-embeddings.
[Accessed: Nov. 1, 2024]

[41]   Laura Carnevali, "Evaluation Measures in Information Retrieval," Pinecone, Jun. 30, 2023. [Online]. Available:
https://www.pinecone.io/learn/offline-evaluation/
[Accessed Nov. 2, 2024]

DT166G, Presentation av ny Teknik
*Integrating Semantic with Keyword Product search in Online
Marketplaces*

Appendix 1: Search schedule for information seeking

9-13-2024	ACM Digital Library	(E-commerce OR online shopping) AND (Hybrid Search OR combined search) AND (Keyword search OR traditional search) AND (Semantic Search OR vector search)	63864	Too focused on E-commerce and too little on product search
9-13-2024	ACM Digital Library	(E-commerce OR Online shopping) OR (Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	327177	Too broad and focused on personalization
9-13-2024	ACM Digital Library	(Online marketplace OR e-commerce) AND (Hybrid OR combined) AND (Keyword OR lexical) AND (Semantic OR vector OR embedding) AND (Search OR retrieval)	77192	Some similar results, though not much within our scope.
9-13-2024	ACM Digital Library	(Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	68713	Too broad
9-13-2024	ACM Digital Library	Product Search AND Semantic Search AND NOT Personalized AND NOT QR AND NOT Q&A AND NOT Assistants	10498	Narrowed it down, but too little focus on search technology
9-13-2024	ACM Digital Library	hybrid semantic product search	13442	Quite narrow, resulted in some of the most relevant results
9-13-2024	Google Scholar	(E-commerce OR online shopping) AND (Hybrid Search OR combined search) AND (Keyword search OR traditional search) AND (Semantic Search OR vector search)	27200	Some relevant results.
9-13-2024	Google Scholar	*(E-commerce OR Online shopping) OR (Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search*	21900	It produced some relevant results, but many were from before 2019.

DT166G, Presentation av ny Teknik
*Integrating Semantic with Keyword Product search in Online
Marketplaces*

		OR lexical search) AND *(Semantic Search OR vector search OR* *embedding search)*		
9-13-2024	Google Scholar	(Online marketplace OR e-commerce) AND (Hybrid OR combined) AND (Keyword OR lexical) AND (Semantic OR vector OR embedding) AND (Search OR retrieval)	36000	A number of relevant results emerged from this search string.
9-13-2024	Google Scholar	Product Search AND Semantic Search AND NOT Personalized AND NOT QR AND NOT Q&A AND NOT Assistants	5150	No relevant results. Mostly focused on personalization.
9-13-2024	Google Scholar	(Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	23500	It produced some relevant results, but many were from before 2019.
9-13-2024	Google Scholar	hybrid semantic product search	482000	Some relevant results. Some published before 2019.
9-13-2024	Semantic Scholar	(E-commerce OR online shopping) AND (Hybrid Search OR combined search) AND (Keyword search OR traditional search) AND (Semantic Search OR vector search)	280	Some Relevant results.
9-13-2024	Semantic Scholar	*(E-commerce OR Online shopping) OR* *(Product search OR Item search OR Information Retrieval) AND* *(Hybrid Search OR combined search or union search) AND* *(Keyword Search OR traditional search OR lexical search) AND* *(Semantic Search OR vector search OR embedding search)*	250	Some Relevant results.
9-13-2024	Semantic Scholar	(Online marketplace OR e-commerce) AND (Hybrid OR combined) AND (Keyword OR lexical) AND (Semantic OR vector OR embedding) AND (Search OR retrieval)	255	A number of relevant results emerged from this search string.
9-13-2024	Semantic Scholar	Product Search AND Semantic Search AND NOT Personalized AND NOT QR AND NOT Q&A AND NOT Assistants	124000	Some relevant results. Majority not.

9-13-2024	Semantic Scholar	(Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	255	Some relevant results.
9-13-2024	Semantic Scholar	hybrid semantic product search	4690	Some relevant results.
9-13-2024	Springer Link	(E-commerce OR online shopping) AND (Hybrid Search OR combined search) AND (Keyword search OR traditional search) AND (Semantic Search OR vector search)	5916	Results often referred to search methods for documents.
9-13-2024	Springer Link	*(E-commerce OR Online shopping) OR (Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)*	5917	Most results are not in the scope of our research.
9-13-2024	Springer Link	(Online marketplace OR e-commerce) AND (Hybrid OR combined) AND (Keyword OR lexical) AND (Semantic OR vector OR embedding) AND (Search OR retrieval)	3186	Some relevant results, though limited to subscribers.
9-13-2024	Springer Link	(Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	5375	Most results are not in the scope of our research.
9-13-2024	Springer Link	Product Search AND Semantic Search AND NOT Personalized AND NOT QR AND NOT Q&A AND NOT Assistants	33302	Majority of results regard other semantic or hybrid use cases than search.
9-13-2024	Springer Link	hybrid semantic product search	12190	Results regarded other hybrid systems than search.

9-15-2024	IEEE Xplore	*(E-commerce OR Online shopping) OR (Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)*	21987	Results were too broad for our research.
9-15-2024	IEEE Xplore	(Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	22194	Too broad for our research.
9-15-2024	IEEE Xplore	(Online marketplace OR e-commerce) AND (Hybrid OR combined) AND (Keyword OR lexical) AND (Semantic OR vector OR embedding) AND (Search OR retrieval)		Most of the research did not focus on hybrid search.
9-15-2024	IEEE Xplore	Product Search AND Semantic Search AND NOT Personalized AND NOT QR AND NOT Q&A AND NOT Assistants	750	Some relevant results.
9-15-2024	IEEE Xplore	(Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	103	Results were too broad for our research.
9-15-2024	IEEE Xplore	hybrid semantic product search	32	Too broad results for our research. Most regarding semantic in other areas.
9-15-2024	ArXiv	(E-commerce OR online shopping) AND (Hybrid Search OR combined search) AND (Keyword search OR traditional search) AND (Semantic Search OR vector search)	N/A	Search strings not compatible with ArXivs search system.
9-15-2024	ArXiv	(E-commerce OR Online shopping) OR (Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search	N/A	Search strings not compatible with ArXivs search system.

		OR lexical search) AND (Semantic Search OR vector search OR embedding search)		
9-15-2024	ArXiv	Product Search AND Semantic Search AND NOT Personalized AND NOT QR AND NOT Q&A AND NOT Assistants	N/A	Search strings not compatible with ArXivs search system.
9-15-2024	ArXiv	(Online marketplace OR e-commerce) AND (Hybrid OR combined) AND (Keyword OR lexical) AND (Semantic OR vector OR embedding) AND (Search OR retrieval)	N/A	Search strings not compatible with ArXivs search system.
9-15-2024	ArXiv	(Product search OR Item search OR Information Retrieval) AND (Hybrid Search OR combined search or union search) AND (Keyword Search OR traditional search OR lexical search) AND (Semantic Search OR vector search OR embedding search)	N/A	Search strings not compatible with ArXivs search system.
9-15-2024	ArXiv	hybrid semantic product search	2	No relevant results.

DT166G, Presentation av ny Teknik
Integrating Semantic with Keyword Product search in Online Marketplaces

Time Plan

ASSIGNMENT	OWNER	START	END	DURATION	% COMPLETED
Assignment 1					
Select research area	Albin, Piran	24-09-04	24-09-04	0	100%
Identify search terms	Albin, Piran	24-09-05	24-09-06	1	100%
Choose full-text databases	Albin, Piran	24-09-06	24-09-06	0	100%
Find relevant resources, explain why	Albin, Piran	24-09-07	24-09-08	1	100%
Indicate the best search string(s)	Albin, Piran	24-09-09	24-09-09	0	100%
Fill in search template	Albin, Piran	24-09-10	24-09-10	0	100%
Write Introduction	Albin, Piran	24-09-11	24-09-11	0	100%
Write Background	Albin, Piran	24-09-12	24-09-13	1	100%
Risk analysis	Albin, Piran	24-09-14	24-09-14	0	100%
Submit Assignment 1	Albin, Piran	24-09-15	24-09-15	0	100%
Assignment 2					
Update Risk Analysis	Albin, Piran	24-09-16	24-09-16	0	100%
Write Purpose/Research Objectives (Section 1.2)	Albin, Piran	24-09-17	24-09-19	2	100%
Write Motivation (Section 1.3)	Albin, Piran	24-09-20	24-09-22	2	100%
Write Research Problem (Section 1.4)	Albin, Piran	24-09-23	24-09-25	2	100%
Write Possible Research Questions (Section 1.4)	Albin, Piran	24-09-26	24-09-29	3	100%
Submit Assignment 2	Albin, Piran	24-09-29	24-09-29	0	100%
Assignment 3					
Write Literature Review (Sections 2 and 3)	Albin, Piran	24-09-30	24-10-11	11	100%
Submit Mid Assignment	Albin, Piran	24-10-11	24-10-11	0	100%
Write Results, Discussion & Conclusion (Sections 5 & 6)	Albin, Piran	24-10-12	24-10-23	11	100%
Write remaining parts	Albin, Piran	24-10-24	24-10-26	2	100%
Submit Peer Review	Albin, Piran	24-10-26	24-10-27	1	100%
Submit Report For Opposition	Albin, Piran	24-10-27	24-10-27	0	100%
Assignment 4					
Prepare Oral Presentation	Albin, Piran	24-10-28	24-10-31	0	100%
Give Oral Presentation	Albin, Piran	24-10-31	24-10-31	0	100%
Submit Opposition	Albin, Piran	24-10-31	24-10-31	0	100%
Submit Final Report & Presentation Slides	Albin, Piran	24-11-03	24-11-03	0	100%

Risk Analysis

STRENGTHS	WEAKNESSES
● **Experience with Information Retrieval:** We have experience with programming and search engine technologies, such as Elasticsearch, enabling the testing and simulation of hybrid search solutions. ● **Hybrid Search Experience in Online Marketplaces:** Our background in implementing hybrid search within an online marketplace provides valuable insights for in-depth review and discussion.	● **Limited theoretical knowledge:** We have limited theoretical knowledge within AI and semantic search. Diving further into those areas will be necessary and time consuming.
OPPORTUNITIES	THREATS
● **Growing interest in AI:** With expanding interest in AI applications for product search, there is a strong demand for insights in this area. Writing this report allows us to contribute valuable perspectives to a trending field ● **Existing research and tools:** A substantial body of research already exists on this topic, offering a strong foundation for our report.	● **Lack of detail**: Proprietary software limitations and other factors can prevent the inclusion of essential details in the reviewed research. ● **Rapid Technological Change**: The fields of AI, semantic search, and hybrid search are rapidly advancing, which poses a risk that certain findings may become outdated.

How to handle threats and weaknesses
- **Limited Theoretical Knowledge**: With limited background knowledge in AI and semantic search, it will be necessary to allocate additional time for study.
    - **Prioritize key concepts:** Focus the research on core ideas that are essential for understanding semantic and hybrid search. This targeted approach can help manage time more effectively.
    - **Start with simpler resources:** Begin with accessible materials such as blog posts or introductory videos, which can provide a foundation before moving on to more advanced research. This method could promote gradual learning which reduces the initial complexity of the subject knowledge before researching in more advanced material.
- **Lack of detail:**
    - **Focus on Open-Source Alternatives**: Whenever possible, focus on research that utilizes open-source software or publicly accessible tools, as these often come with extensive documentation and offer the potential for reproducibility.
    - **Review complementary resources:** Additional research from complementary sources can be included to fill any informational gaps.
- **Rapid Technological Change**:
    - **Focus on Foundational Principles**: Emphasize foundational concepts and techniques in AI and search that are less likely to become obsolete. This ensures that the report remains relevant over time, even as specific technologies evolve.
    - **Review Recent Research**: Review the latest research and advancements to ensure that the findings reflect current developments.