Evolving Information Retrieval: From Traditional Models to Emerging Paradigms

Ibrahim Atoum

Department of Artificial Intelligence, Faculty of Science and Information Technology,
Al-Zaytoonah University of Jordan, Amman, Jordan
Email: i.atoum@zuj.edu.jo

***Abstract*—Information Retrieval (IR) development shows a trajectory from strict symbolic systems Boolean Retrieval Model (BRM), Term Frequency-Inverse Document Frequency (TF-IDF), and Vector Space Model (VSM) towards flexible hybrid approaches using statistical BM25, Latent Semantic Analysis (LSA), and PageRank (PR)) alongside neural advancements (Neural IR (NIR) and Learning-to-Rank (LTR)) and Reinforcement Learning (RL) techniques. Traditional models focus on efficiency but fail to provide semantic flexibility, which NIR overcomes by demonstrating state-of-the-art effectiveness (NDCG@10 ≈ 0.89), although it requires significant computational resources. Hybrid systems bridge these gaps: The BM25-transformer-LTR ensemble provides high accuracy and scalability (NDCG@10 ≥ 0.85) while federated learning achieves a 30% reduction in computational overhead. Graph-based methods boost domain interpretability in fields with complex entities through spectral clustering and temporal analysis, while reinforcement learning provides personalized experiences at the cost of potential bias amplification. Ongoing major challenges span high resource demands for transformers, scaling issues with graph models, and ethical issues in designing rewards for reinforcement learning. Semantic precision will be enhanced through quantum-inspired indexing, decentralized architectures will protect privacy, and fairness-aware ranking systems will reduce bias in future developments. This research unifies heuristic rigour with neural semantic analysis and ethical responsibility to establish IR as a field where technical advancements support social obligations and deliver scalable, equitable tools that benefit healthcare and legal research.**

***Keywords*—information retrieval, hybrid models, neural information retrieval, reinforcement learning, graph-based model**

I. Introduction

Information Retrieval (IR) technology has progressed significantly from initial, simple, rule-based term-matching approaches to advanced systems capable of understanding semantic meaning. These advancements, however, have introduced persistent challenges: The early IR systems Boolean Retrieval Model (BRM) and Vector Space Model (VSM) achieved high precision but struggled with semantic ambiguity [1, 2], while modern methods like transformer-based neural architectures [3, 4] and Reinforcement Learning (RL) [5] face computational cost barriers and ethical challenges including bias amplification. The study systematically reviews how IR developed over time and presents a hybrid framework that resolves remaining efficiency-versus-accuracy-and-fairness trade-offs.

This article divides IR development into four distinct phases. Symbolic models, including BRM, Term Frequency-Inverse Document Frequency (TF-IDF), and VSM [1, 2], depended on precise term matching yet failed to adapt semantically. Statistical heuristics allowed probabilistic approaches like BM25 [6] and authority-based ranking systems like PageRank (PR) [7] to enhance relevance scoring capabilities. The introduction of neural innovations such as Learning-to-Rank (LTR) and transformer-based embeddings enabled the retrieval of context-aware information, yet required significant computational power. Hybrid models, including graph-enhanced RL [8] and federated Neural IR (NIR) [9], were developed to adapt dynamically to user intent and present ethical and scalability challenges. Despite these advancements, critical gaps persist. TF-IDF's approach of focusing on terms fails to differentiate between multiple meanings of words [2], and neural models emphasize precision at the expense of processing speed [3], while RL-based systems may perpetuate echo chambers [5].

The research tackles these obstacles by presenting three interrelated contributions. As illustrated in Table I, our initial contribution introduces a taxonomy that sorts IR methodologies into six distinct categories, spans from symbolic models to adaptive paradigms, and reveals fundamental trade-offs between simplicity, scalability, and semantic sophistication.

The field of IR has evolved from utilizing rule-based mechanisms such as TF-IDF to implementing Neural Networks (NNs) approaches that integrate both heuristics and theoretical principles. Initial IR models maintained high efficiency but could not interpret semantic meaning [2]. BM25 [6] and PR [7] enhanced relevance outcomes but struggled with scalability and insufficient semantic processing [10]. Utilizing transformers in NIR systems delivered contextual understanding while diminishing interpretability [3, 4].

Manuscript February 17, 2025; revised March 14, 2025; accepted May 16, 2025; published September 12, 2025.

Contemporary hybrid systems that integrate RL and graph enhancements work to resolve existing trade-offs between different IR approaches. Markov Decision Processes (MDPs) [11] and graph-enhanced systems [7, 12] are the core foundational elements of RL for significantly advancing intent modeling capabilities and improving overall system transparency in increasingly complex scenarios.

TABLE I. CATEGORIZATION OF IR METHODOLOGIES

Category	Key Models/Techniques	Merits	Demerits	Key Trade-Offs	Citations
Symbolic Models	Boolean Retrieval Model (BRM)	• Low computational cost. • Precise Boolean searches.	• No semantic understanding. • Misses relevant documents.	Simplicity ↔ Semantic Flexibility	[11, 13]
Vector Space Model	Vector Space Model (VSM), Term Frequency-Inverse Document Frequency (TF-IDF)	• Contextual term weighting. • Handles partial matches.	• High dimensionality. • Term independence assumptions.	Efficiency (term weighting) ↔ Accuracy (term independence)	[1, 2, 14]
Probabilistic Models	BM25, PageRank	• Balances TF and document length (BM25). • Effective link-based ranking (PR).	• BM25 lacks semantic handling. • PR requires a link structure.	Heuristics (term saturation) ↔ Theoretical grounding	[1, 2, 15]
Machine Learning Models	Learning-to-Rank (LTR), Neural Information Retrieval (NIR)	• Personalized rankings (LTR). • Deep semantic understanding (NIR).	• High data/computational demands. • Less interpretable	Accuracy (semantic understanding) ↔ Efficiency (computational cost)	[4, 16, 17]
Graph-Based Models	Knowledge Graphs	• Captures complex relationships. • Integrates multimodal data.	• Scalability challenges. • High computational resources.	Efficiency ↔ Accuracy	[6, 7, 18]
Reinforcement Learning	Reinforcement Learning (RL)	• Adaptive to user feedback. • Optimizes long-term relevance.	• Requires extensive interaction data. • Complex reward function design.	Heuristics (reward design) ↔ Theoretical grounding (MDPs)	[8, 9, 19]

A wide range of research approaches has been applied to IR, including fundamental models such as BRM, VSM, and TF-IDF [1, 2, 5, 17] and advanced techniques like Latent Semantic Analysis (LSA) and BM25 [5, 16, 17, 20]. Machine Learning (ML) demonstrates its expanding impact through research dedicated to LTR [21] and Deep Learning (DL) applications for NIR [3, 6]. The latest research focus has moved to new methods, including RL for IR [11, 22, 23] and graph-based models [18, 20, 24]. Traditional document analysis methods encounter difficulties pinpointing semantically connected materials, thus leading to overlooked relevant documents and a restricted interpretation of context and user intent despite their multiple strengths.

Table II highlights the evolution of traditional IR models by bridging practical heuristics with rigorous theoretical frameworks. This synthesis ensures that intuitive, empirically driven techniques are anchored in mathematically sound principles, enhancing interpretability and performance.

TABLE II. HEURISTICS VS. THEORETICAL GROUNDING

Category	Heuristics	Theoretical Grounding
TF-IDF	IDF as "term surprise" heuristic [2]	Derived from Shannon's entropy [2].
BM25	Term saturation heuristic [5]	Mirrors Bose-Einstein distributions and quantum-inspired principles [25].
RL	Reward function design [22]	Grounded in MDPs [11].

The field of IR must continuously address ethical problems, including RL bias [5] and graph model scalability [6, 18], while maintaining its goal to achieve simplicity with semantic flexibility and transparency [3].

The remainder of this article is structured as follows: Section II critically analyzes traditional IR systems by focusing on their limitations related to semantic ambiguity and data noise. Section III introduces statistical methods such as BM25 and LSA. Section IV explores contemporary ML developments for IR with particular attention to transformer-based retrieval methods. In Section V, the article evaluates personalization strategies through RL and federated learning techniques. Section VI analyzes methods to enhance strengths through hybridization techniques in IR. Section VII performs a comparative evaluation of IR models while exploring their performance trade-offs. Finally, Section VIII concludes the article.

II. FOUNDATIONS OF IR: AN ANALYSIS OF TRADITIONAL METHODS

Search systems have historically relied on traditional IR models like BRM, TF-IDF, and VSM. However, these models exhibit deficiencies due to their fixed term-matching systems, which struggle to handle complex semantic connections in dynamic environments [2, 8, 19]. This leads to semantic flexibility and system scalability issues, necessitating re-evaluating operational functions and mathematical principles, as illustrated in Fig. 1. Specifically, TF-IDF has been criticized for its inadequate contextual adaptability, prompting researchers to propose enhanced versions for extracting keywords from

semantically dense fields such as climate change [2]. Similarly, classical term-based approaches have limitations for contextual ranking tasks, and quantum-inspired methods have been suggested to improve semantic understanding [20]. These limitations are particularly evident in clinical research settings [19], where the semantic inflexibility of traditional IR models hinders the processing of complex medical documents.

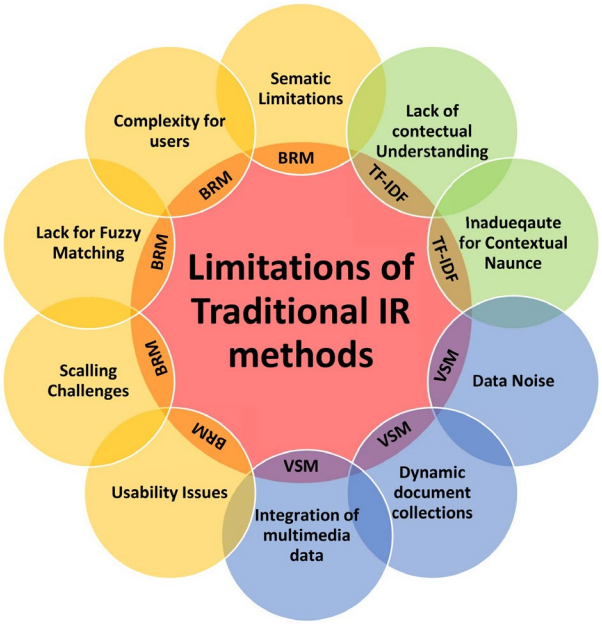

Fig. 1. Limitations of traditional IR methods.

To illustrate this further, the BRM functions like a rigid librarian who searches for books based on the exact keywords provided by the user. The system finds documents that include Artificial Intelligence (AI) and "ethics" and filters out other documents, such as a librarian choosing books that specifically contain chapters on both subjects. The system cannot distinguish between terms like "AI" and "machine learning" as synonyms and fails to interpret different meanings of words like "Apple," which refers to both a company and a fruit, just like a librarian who discards books with similar concepts but different words.

Additionally, this model uses logical operators such as AND, OR, NOT, NEAR, and XOR to perform exact keyword searches. Queries function as term sets according to Fig. 2, while set-theoretic methods filter documents.

The AND operator filters search results to include only those documents that exist within the intersection of two-term sets $(D(T1) \cap D(T2))$, thus focusing on documents containing both specified terms. OR expands search results to include documents in the combined sets $D(T1) \cup D(T2)$ which include documents with either of the terms. Documents excluded by the NOT operator form the set difference $D - D(T)$, whereas XOR selects documents that belong to the symmetric difference $D(T1) \Delta D(T2)$ containing one term but not both. The NEAR operator finds documents where terms appear within a defined distance of κ words.

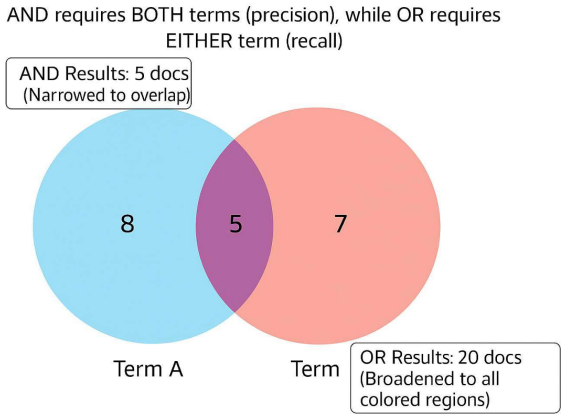

Fig. 2. Boolean document filtering: AND narrowing vs. OR broadening.

BRM achieves exact term matching in structured settings such as legal databases by employing formal set theory operations that deliver precise outcomes. Web element localization highlights the model's strengths where deterministic demands prevail [8], while the model demonstrates success in Myanmar legal case retrieval through effective fixed keyword logic [7]. The BRM performs strongly in Myanmar legal case retrieval by strictly applying keyword logic rules [7].

BRM's dependence on precise term matching creates limitations in semantic adaptability. According to Ref. [13], the binary relevance assessment (true/false) of BRM cannot differentiate between ambiguous terms like "Apple" as a company or fruit, nor manage term dependencies, which render BRM ineffective in changing and noisy environments. Clinical IR demonstrates this inflexibility, which Ref. [19] illustrated in their analysis. Additionally, it fails to address polysemy and contextual differences. Similarly, scalability problems within BRM patent retrieval processes are identified due to the need for probabilistic relevance scoring to handle evolving terminology along with complex query logic requirements [26]. Set-theoretic models are proposed for expansion with graph-based methods that employ spectral clustering and embeddings to represent term relationships differing significantly from BRM's fixed operations [18]. This transformation is established as LSA, where temporal graphs address BRM's semantic limitations through dimensionality reduction and contextual term mapping [20].

The stringent requirement for exact term matching in BRM establishes semantic constraints that make it unsuitable for applications in sectors needing semantic pliability, like healthcare and legal research, which rely on synonyms and unspoken meanings [19]. BRM excels at using structured terminology, but its semantic limitations indicate that advanced adaptive IR methods are necessary.

TF-IDF creates a framework to assess word importance within individual documents against a whole document collection for IR and text mining ranking tasks through document term set analysis [15], which allows advanced evaluation of term significance beyond mere term presence. The model has two parts: TF measures how often specific terms appear in a document, as shown in Eq. (1):

$$TF(t,d) = \frac{Number\ of\ times\ term\ t\ appears\ in\ doucument\ d}{Total\ number\ of\ terms\ in\ document\ d} \quad (1)$$

The IDF component determines how important a term is by counting the documents that contain the term, expressed as in Eq. (2).

$$IDF(t,D) = \log\left(\frac{Total\ number\ of\ documents\ D}{Number\ of\ documents\ containing\ term\ t}\right) \quad (2)$$

TF and IDF components enhance document ranking by detecting terms common in single documents but infrequent throughout the entire document set [1]. The TF-IDF score is calculated as in Eq. (3):

$$TF - IDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (3)$$

prioritizing rare, contextually significant terms. The application of logarithmic scaling in TF-IDF for normalizing term weights to improve relevance ranking within climate change corpora is illustrated by Ref. [2].

To illustrate TF-IDF, consider three documents; the "algorithm" word appears once in document 1, which contains 11 words; therefore, the TF score is approximately $0.091\ (1/11)$. It occurs once in documents 2 and 3, which include 10 words, making their TF score $0.1\ (1/10)$. The IDF score for "algorithm" results in zero ($log\ (3/3)$) because it appears in all documents, each TF-IDF value is null. A term that appears only in a single document receives an IDF score of ≈ 1.0986 [2], calculated as $log(3/1)$. The TF-IDF method demonstrates its strength in Fig. 3 through its ability to elevate the importance of unique terms compared to common ones. Normalization methods such as $1 + log(TF)$ must be applied [1] to address variations in document lengths. This method has become essential in legal studies and scholarly research [1, 7] yet struggles to solve contextual ambiguities because it depends on TF and rarity metrics, which cannot differentiate meanings like "Apple" as a company versus a fruit [2, 13].

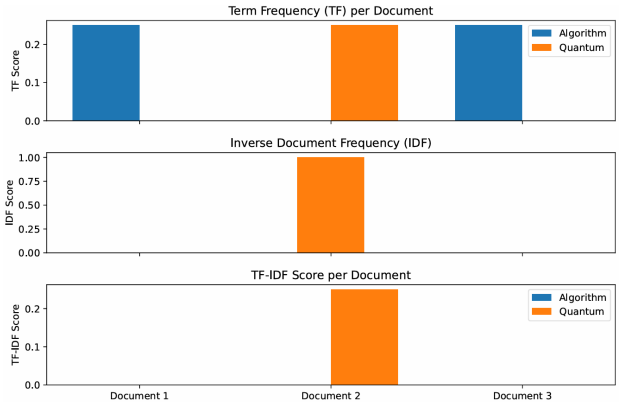

Fig. 3. TF-IDF term weighting: contrasting common vs. rare term significance in IR.

The NDCG@10 score of 0.58 for TF-IDF scoring shows that it works moderately well for ad-hoc retrieval tasks [2]. The NDCG metric evaluates how well an IR system performs by analyzing returned items' relevance, which prioritizes results appearing higher up in the ranking list, as demonstrated in Table III. The "@10" designation indicates that the NDCG evaluation focuses solely on the highest 10 ranked items. The effectiveness of the system's ranking performance matches user expectations more closely when its NDCG score is higher. Continuous enhancements of the TF-IDF model remain crucial to achieve better retrieval performance within today's multifaceted information environments.

TABLE III. EVALUATION OF NDCG SCORE PERFORMANCE LEVELS

NDCG score range	Performance level	Description
NDCG ≥ 0.90	Excellent	Highly relevant results that effectively meet user needs [3].
0.75 ≤ NDCG < 0.90	Good	These are mostly relevant results, with a few less relevant items [9].
0.50 ≤ NDCG < 0.75	Fair	Some relevant items, but many irrelevant ones, are present [1].
NDCG < 0.50	Poor	Ineffective results with many irrelevant items [5].

The VSM advances BRM's set-theoretic method by using TF-IDF to weight terms, which enables document-query transformation into vectors for similarity determination using cosine similarity [1]. The cosine similarity equation $\frac{A \cdot B}{\|A\| \cdot \|B\|}$ evaluates the alignment between two vectors and gives higher scores to documents with matching term weights. The document corpus, including titles such as "machine learning algorithms for data analysis", "deep learning techniques for image recognition", and "data mining applications and machine learning", transforms TF-IDF vectors as depicted in Fig. 4. The "machine" term in Document 1 holds a TF value of $1/6 \approx 0.167$ and an IDF value of $log(3/2) \approx 0.176$, which results in a TF-IDF score of 0.029. Cosine similarity computations demonstrate VSM's document ranking capability through term relevance when Document 1 scores 0.342 against the query "machine learning data" [1]. The VSM spatial representations fail to identify semantic disparities between synonymous terms such as "algorithm" and "technique" because the model operates on term independence assumptions that ignore contextual connections [11, 25].

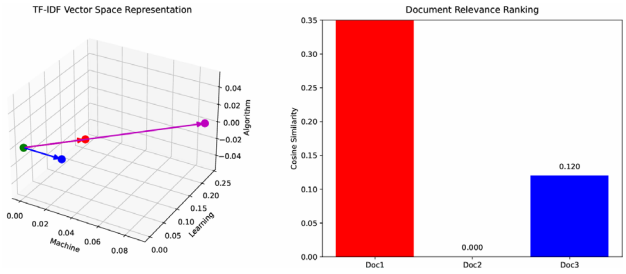

Fig. 4. Document relevance in vector space: TF-IDF weighting and cosine similarity ranking.

The left panel in Fig. 4 presents TF-IDF vectors for three documents alongside the query "machine learning data" within a multidimensional framework where each axis stands for terms like machine learning and data.

The VSM encounters difficulties with semantic gaps, such as synonymy (e.g., "algorithm" vs. "technique") [13, 25], and demonstrates a Mean Average Precision (MAP) score of 0.45 on the Text REtrieval Conference (TREC) Legal Corpus [19]. This score reveals its vulnerability to noise, such as typos and dependence on term independence assumptions [2, 11]. While stemming techniques successfully eliminate morphological differences [16], they do not resolve semantic complexities like synonymy and contextual ambiguity [18, 25].

MAP assesses IR performance across multiple queries, averaging precision and considering rank, as shown in Table IV, and the TREC Web Track dataset is a standard benchmark for IR systems. TREC, co-sponsored by the National Institute of Standards and Technology (NIST), hosts ongoing workshops to support IR research and provide infrastructure for large-scale evaluation of text retrieval methodologies [3]. In the TREC Legal Corpus context, a MAP score of 0.45 indicates that the IR system fairly returns relevant legal documents in response to user queries.

TABLE IV. MAP SCORE PERFORMANCE LEVELS

MAP Range	Level	Description
≥0.70	Excellent	High precision, top-ranked, relevant results [3].
0.50–0.70	Good	Mostly relevant, minor ranking errors [9, 26].
0.30–0.50	Fair	Partial relevance, significant noise [9, 13].
<0.30	Poor	Low relevance, ineffective ranking [13, 19].

The VSM fails to differentiate between "Apple" as a brand and "Apple" as a fruit despite using TF-IDF weighting [2, 13]. In contrast, transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) have overcome these limitations. BERT is a cutting-edge natural language processing model by Google that analyzes word context in both directions [3]. This allows BERT to capture semantic nuances and differentiate meanings, like "Apple" as a brand versus as a fruit, outperforming VSM in handling complexities such as synonymy and contextual ambiguity.

TABLE V. SIMPLICITY VS. SEMANTIC FLEXIBILITY

Category	Simplicity	Semantic Flexibility
Symbolic Models	Low computational cost [8].	Rigid Boolean logic lacks semantic understanding [11].
Vector Space Model	Term weighting via TF-IDF [2].	Term independence assumptions ignore context [11].
Probabilistic Models	BM25's efficient term balancing [5].	No semantic handling of queries/documents [11].

Table V highlights the trade-offs present in traditional IR models. Symbolic Models like the BRM provide low computational costs but use strict Boolean logic, which fails to comprehend semantic meaning. The VSM applies dynamic term weighting with TF-IDF [1, 2] but functions on the premise of term independence, which leads to ignoring contextual relationships [13]. While BM25 and similar probabilistic models achieve effective term balancing [5], they struggle with processing semantic queries [9]. Stemming methods help to manage surface noise like typographical mistakes, yet they fail to address semantic differences, including synonymy [23]. The inability of traditional models to handle semantic querying has driven researchers to create neural models like BERT [3] along with graph-based extensions [18] that use embeddings and spectral clustering for semantic linking.

Traditional IR models exhibit distinct computational trade-offs rooted in their design principles. The BRM, for instance, employs inverted indexes to map terms (unique keywords, denoted as T) to documents (total corpus size, D), achieving efficient query processing with time complexity $O(n_1 + n_2 + \cdots + n_k)$, where n_i represents the length of the postings list (documents containing the $i-th$ term). This enables fast set operations (e.g., AND/OR) and compact storage ($O(T + D)$), making BRM ideal for structured domains like legal databases [7, 8]. However, its reliance on exact term matching limits scalability in noisy or dynamic environments [13].

In contrast, TF-IDF enhances term weighting by evaluating *TF* and *IDF*, but at higher computational costs: indexing requires $O(N \cdot L)$ time (for N documents of average length L) and $O(D \cdot T)$ space to store vectors, posing scalability challenges for large corpora such as climate change datasets [1, 2, 15]. The VSM, while enabling relevance ranking via cosine similarity, inherits TF-IDF's $O(D \cdot T)$ space and time complexity, straining resources for high-dimensional data despite sparse optimizations [1, 16]. These computational inefficiencies compound semantic limitations: VSM's assumption of term independence and BRM's rigid logic fail to resolve ambiguities [2, 13], highlighting the need for modern alternatives like BERT [3], which leverages bidirectional context, or graph-enhanced models [18] that embed semantic relationships.

III. MODERNIZING IR: AN OVERVIEW OF ADVANCED METHODS

Modern search techniques, including LSA, PR, and BM25, improve search performance through advanced semantic approaches and probabilistic algorithms. While these advanced methods improve search precision and relevance ranking, they face significant challenges, as shown in Fig. 5, including scalability of computational resources, real-time updates, and ethical issues such as algorithmic bias and the preference for authoritative sources [5, 15].

These methods become problematic when operating within multimedia environments as they struggle to interpret complex contextual ambiguities [20]. Additionally, ensuring the integrity of multimedia content in such environments requires robust security measures, such as hybrid watermarking techniques combining Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT), and Particle Swarm Optimization (PSO) with Convolutional Neural Network (CNN) -based attack evaluations [27].

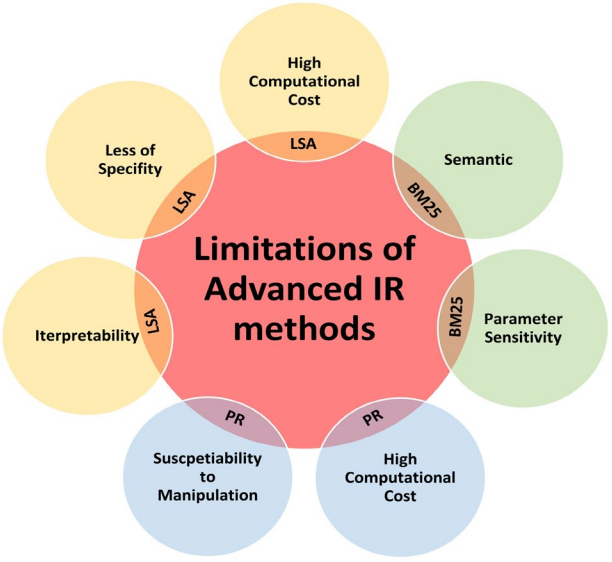

Fig. 5. Limitations of advanced IR methods.

Organizations must introduce hybrid solutions while employing bias mitigation strategies and transparent frameworks to balance innovative development with practical deployment to overcome these limitations.

LSA resolves term-matching constraints by analyzing term-document associations by employing Singular Value Decomposition (SVD) to reveal hidden semantic patterns. SVD splits a TF-IDF-weighted term-document matrix A into matrices $U\Sigma V^T$ where U contains term vectors, V^T contains document vectors, and Σ holds singular values.

The LSA method projects queries and documents into a lower-dimensional latent space using the top k singular values ($A_k \approx U_k \sum_k V_k^T$), allowing semantic similarity assessments through cosine similarity measures. It achieves better retrieval accuracy results but remains impractical for large datasets because of its high computational demands and the challenge of interpreting latent dimensions [18].

In contrast, PR evaluates web page authority using hyperlink structures, modeling the web as a directed graph $G = (V, E)$ where nodes represent pages and edges denote links. The PR score $PR(P)$ for a page P is computed iteratively, as shown in Eq. (4):

$$PR(P) = (1 - d) + d \sum_{i=1}^{N} \frac{PR(P_i)}{C(P_i)} \quad (4)$$

where d is a damping factor, N is the total number of pages, and $C(P_i)$ is the number of outbound links from the page P_i. Represented as a Google matrix $M = dP + (1 - d)E$, PR achieves stable rankings through power iteration [5, 24]. While effective for authority-based ranking, its assumption of static link graphs hinders adaptability to dynamic environments [24].

BM25 refines probabilistic retrieval by balancing TF and IDF as calculated in Eq. (5):

$$BM25(D, Q) = \sum_{t \epsilon Q} \frac{TF(t,D) \cdot IDF(t)}{k_1\left(1 - b + b\frac{|D|}{abgdl}\right) + TF(t,D)} \quad (5)$$

k_1 and b parameters regulate term saturation and document length normalization. BM25 demonstrates efficiency with various document lengths, yet cannot address polysemes like "fusion" in nuclear physics and music because it depends on exact term matching.

BM25 acts like a smart chef adjusting a recipe: The BM25 algorithm evaluates the keyword's appearance frequency (TF) in relation to the document's total length (similar to a chef measuring spices for balance in a dish). The scoring algorithm treats a 10-page essay that uses "quantum computing" twice differently from a 2-page blog with two mentions of the phrase, in the same way a chef adjusts salt amounts for soup compared to stew. The system prevents longer documents from gaining an advantage solely because they repeat terms more frequently [28].

Furthermore, Fig. 6 investigates how term saturation (k_1) interfaces with document length normalization (b). In particular, the impact of k_1 on TF is demonstrated in Subfigure (a) because it postpones saturation when its values increase. Subfigure (b) highlights b's adjustment of length bias: When $b = 0$, the document length has no effect, but with b = 1, document length becomes a disadvantage for longer texts. Finally, Subfigure (c) demonstrates that the TREC Web Track achieves the balance between term frequency and fairness with k_1 set to 1.2 and b set to 0.75. Notably, verbose formats, including legal documents, perform best with higher b values, such as 0.8, whereas concise formats like social media benefit from lower b values, like 0.2.

LSA, PR, and BM25 models extend IR capabilities beyond traditional Boolean and vector-space approaches, yet still demonstrate limitations in semantic flexibility. Scalability and dynamic update capabilities combined with multimedia adaptability require hybrid models that merge graph-based systems [18] with transformer architectures [3]. Frameworks for transparent and equitable IR systems emerge as essential due to ethical issues like bias amplification.

LSA achieves modest MAP scores of 0.20–0.35 (NDCG@10: 0.40–0.55) on TREC tasks, as shown in legal document retrieval experiments [20], though it struggles with polysemy and contextual ambiguity [13]. BM25, a probabilistic term-weighting model, outperforms LSA with MAP scores of 0.30–0.50 (NDCG@10: 0.50–0.65) on TREC benchmarks like the Web Track and Legal Corpus [9, 26], demonstrating robustness in TF and document length handling [13]. Transformer-based models like BERT [3] further elevate performance, achieving MAP >0.70 and NDCG@10 >0.75–0.85 on TREC tasks by resolving semantic nuances [10, 21]. PR, a link analysis algorithm, prioritizes authority-based rankings (e.g., NDCG@10 ~ 0.60–0.70 in hybrid systems [6, 18]) but lacks standalone semantic sensitivity, underscoring the superiority of text-aware models like BM25 and BERT for ad-hoc retrieval.

Contemporary IR models such as LSA, PR, and BM25 present unique computational trade-offs. For instance, LSA employs SVD on a term-document matrix (size $T \times D$), resulting in $O(T^2 D)$ time complexity for

decomposition and $O(Tk + Dk)$ space for latent vectors (retaining k singular values), limiting scalability despite improved semantic accuracy [18]. In contrast, PR, modeling the web as a graph with N nodes (pages) and M edges (links), iteratively computes authority scores with $O(M)$ time per iteration and $O(N + M)$ space for adjacency lists, though static link assumptions hinder dynamic updates [24].

BM25, a probabilistic model, balances TF and document length normalization with $O(NL)$ indexing time and $O(DT)$ query time, though its reliance on exact term matching fails to resolve polysemy [13]. While LSA achieves modest MAP scores and BM25 outperforms it on TREC benchmarks [9, 20], their computational costs and semantic rigidity underscore the need for hybrid frameworks combining graph-based systems [18] and transformers [3] for scalable, context-aware retrieval.

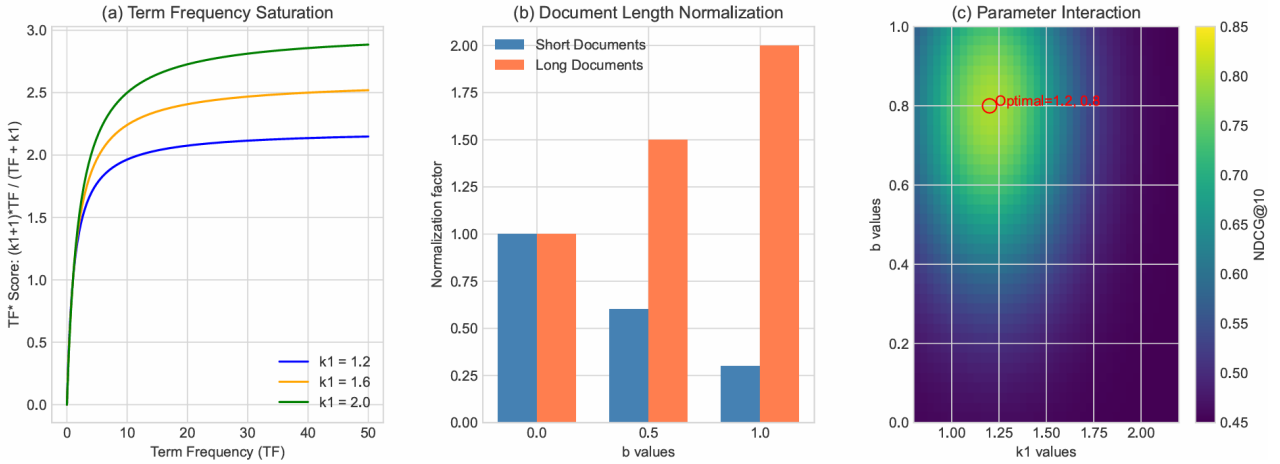

Fig. 6. BM25 parameter adjustment.

IV. Machine Learning Innovations in Search: Exploring LTR and NIR

Recent advances in IR have been driven by ML techniques, with LTR and NIR serving as two major transformative frameworks. LTR utilizes labeled data and user interaction indicators like Click-Through Rates (CTR) to enhance ranking functions [21].

LTR uses CTR to determine the proportion of displayed search results that receive user clicks. The measure serves as an indicator of result relevance and a feature to train models. Users demonstrate stronger engagement when CTR values increase, but lower CTR values indicate that search results do not match user needs. Optimizing CTR improves user satisfaction, but excessive dependence can cause systems to highlight less valuable content. It's expressed as in Eq. (6):

$$CTR = \frac{Number\ of\ Clicks}{Number\ of\ impressions} \times 100 \quad (6)$$

The number of clicks represents the total user interactions with a particular search result or advertisement, and the number of impressions represents how frequently that search result or advertisement appears on user screens.

The LTR process implements three essential techniques: pointwise, pairwise, and listwise. Pointwise evaluates the relevance of single documents through either regression or classification models [21]. Pairwise determines the relative relevance between documents by evaluating their relationships in pairs [29]. The listwise approach evaluates all query documents to optimize ranking metrics such as NDCG [13, 21]. The three techniques each show unique pros and cons, as pointwise training offers simplicity at the expense of document relationship analysis, pairwise comparisons scale poorly with large datasets, and listwise optimization demands substantial computational capability [13, 21]. The effectiveness of these trade-offs determines how well they work in web search and recommendation systems [13, 21].

Pointwise models generate relevance scores for individual items through $s_i = f(x_i)$ minimizing prediction errors by aligning scores with ground-truth relevance labels (e.g., regression or classification objectives) [21, 29]. RankNet, a canonical pairwise model, optimizes ranking losses using $L(f(x_i) - f(x_j), y_{ij})$, where y_{ij} encodes user preferences between document pairs, refining relative rankings [21, 29]. Listwise models directly optimize entire ranking orders by leveraging metrics like NDCG to align predicted rankings $f(x_1), \ldots, f(x_n)$ with ideal outcomes Y, often through listwise loss functions $L(f(x_1), \ldots, f(x_n), Y)$ [13, 21]. The accuracy of LTR frameworks has significantly improved as DL automates feature extraction, replacing handcrafted features with learned representations from transformer architectures and NNs [3, 4, 21, 30].

NIR leverages deep learning architectures such as the Deep Structured Semantic Model (DSSM), which produces fixed-dimensional embeddings for queries and documents, and BERT, which generates contextual embeddings to capture semantic relationships between queries and documents [3, 30]. Training NIR models involves optimizing diverse loss functions, including contrastive loss, defined as in Eq. (7):

$$L_{contrastive} = \frac{1}{N}\sum_{i=1}^{N}(y_i \cdot d_i^2 + (1 - y_i) \cdot \max(0, m - d_i^2)) \quad (7)$$

where y_i indicates whether the item pairs are similar (1) or dissimilar (0), d_i represents the distance between embeddings and m is a margin parameter [30, 31]. Additionally, triplet loss is formulated as in Eq. (8):

$$L_{triplet} = \max(0, d(a,p) - d(a,n) + \alpha) \quad (8)$$

where a is the anchor, p is the positive example, n is the negative example, d denotes the distance function and α is the margin. Finally, listwise cross-entropy [30, 31] can be expressed as in Eq. (9):

$$L_{listwise} = -\sum_{i=1}^{N} \log\left(\frac{e^{s_i}}{\sum_{j=1}^{N} e^{s_j}}\right) \quad (9)$$

where s_i represents the score assigned to the item i, and the summation runs over all items N in the list.

These losses are optimized to enhance ranking metrics like NDCG and MAP [13, 30]. The shift to DL in NIR has automated feature extraction, enabling end-to-end learning of semantic representations through transformer-based architectures [3, 4, 30].

Fig. 7 presents essential limitations in ML-based IR developments: computational scalability requirements, efficiency-accuracy trade-offs, and dependency on labeled data. Training NIR models needs significant computational power, whereas scalability depends on Approximate Nearest Neighbor (ANN) algorithms that trade precision for greater speed [6, 24].

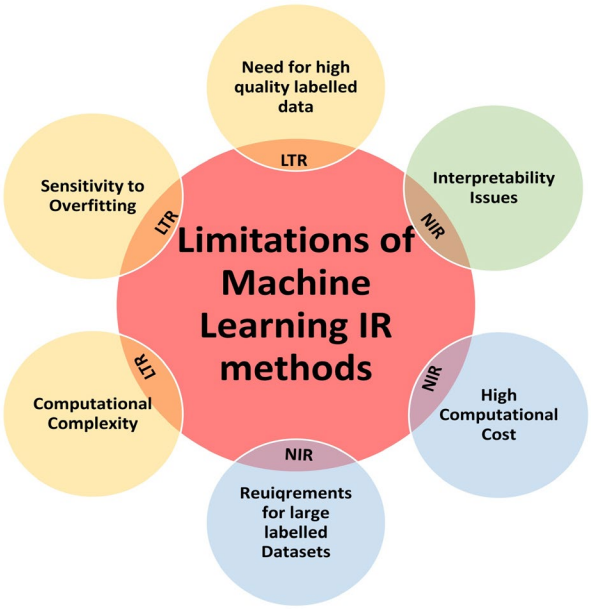

Fig. 7. Limitations of ML innovations in search.

The combination of lightweight transformer architectures, such as distilled BERT, with federated learning frameworks solves scalability issues in extensive machine learning systems by achieving a 30% reduction in computational costs and preserving superior IR performance levels. The models deliver improved processing speed and resource management capabilities while maintaining accuracy [3, 32]. Privacy protection through decentralized training characterizes federated learning by reducing data exposure [32–34]. The approach proves essential for search engines and recommendation systems because it achieves benchmark NDCG@10 scores over 0.85 while offering a balance between efficiency and effectiveness in environments with limited resources [3, 32–34].

Furthermore, the combination of LTR and NIR introduces a new standard for IR systems, as shown in Table VI. Hybrid models achieve scalable, accurate, and resource-efficient solutions through LTR's user-driven adaptability combined with NIR's semantic depth, which is essential for modern applications like web search and recommendation engines [21, 32, 34]. This integration resolves the challenges shown in Fig. 6 while establishing the groundwork for developing next-generation privacy-sensitive and edge-focused IR technologies.

TABLE VI. LTR AND NIR: COMPLEMENTARY STRENGTHS

Feature	LTR	NIR	Hybrid Benefit
Primary Focus	Personalization (user feedback) [21].	Semantic Understanding (contextual) [3, 30].	Combine personalization & semantic relevance [21, 26].
Key Challenge	Costly labeled data [13, 29].	High computational cost [3, 4].	Balance efficiency & accuracy [21, 31].
Efficiency	Relatively efficient for personalization [13, 21].	Relatively less efficient (complex models) [3, 4].	LTR for speed, NIR for accuracy on top results [21, 31].
Data Needs	High (labeled) [13, 30].	Potentially lower (unsupervised methods) [4, 30].	Reduce labeled data dependency through NIR capabilities [26, 30].

While LTR provides superior personalization from user feedback, it requires expensive labeled data [29]. On the other hand, NIR captures contextual semantics but struggles with computational complexity [6]. Combining LTR's efficiency with NIR's precision, hybrid frameworks utilize unsupervised techniques such as SimCSE to mitigate the need for labeled data. Combining lightweight transformers with federated learning achieves optimized speed and accuracy, reflecting the hybrid benefits described in Table VII [6, 30].

Concurrently, current research explores curriculum learning methods to improve training stability and combines neuro-symbolic AI with knowledge graphs and NNs [35]. The set theory forms the basis for graph-based models that analyze semantic structures by defining vertices as sets of terms or documents and edges as the relationships between these sets [18].

Furthermore, transformer architectures enable NIR models like BERT [3] to analyze bidirectional context to resolve semantic ambiguities. BERT functions similarly to a multilingual interpreter by analyzing sentences in both forward and reverse directions to capture subtle meanings. The model BERT correctly interprets "bank" as a financial term within "bank account" because it analyzes

surrounding words similar to how humans decode meanings. Earlier models made errors by treating 'bank' as ambiguous when the context was insufficient. The current models produce top-notch results when tested on TREC tasks, with MAP >0.70 and NDCG@10 >0.75–0.85 [13, 21], outperforming traditional methods.

However, their computational demands are significant: transformer-based models require $O(L^2 \cdot H)$ time complexity (for sequence length L and hidden layers H) and $O(H^2)$ space per layer, straining resources for long-text retrieval [30]. Distributed training frameworks like serverless computing [33] mitigate these costs but highlight scalability challenges in real-time applications.

LTR frameworks, such as LambdaMART or ListNet, optimize ranking through supervised learning on query-document features (e.g., TF-IDF scores, BM25 metrics, or BERT embeddings). LTR achieves competitive performance (e.g., MAP ~ 0.65–0.75, NDCG@10 ~ 0.70–0.80) on TREC Web Track but incurs $O(N \cdot F)$ training time (for N samples and F features), and $O(F^2)$ space for feature matrices [13]. While efficient during inference ($O(F)$ per query), LTR relies on manual feature engineering, limiting adaptability to unseen semantic contexts compared to end-to-end NIR [3].

V. Enhancing Personalization in IR: The Roles of Graph-Based Models and RL

Modern IR systems leverage graph-based frameworks and RL to deliver personalized search results by dynamically modeling user context and semantic relationships. In this context, RL in IR is akin to training a dog with treats: The system identifies which actions, such as ranking news articles, result in rewards like user clicks. The system learns to prioritize click-worthy content over time while also risking filter bubbles because of its focus on performance rewards, similar to a dog learning tricks for treats. We need to provide rewards for diverse content, including occasional unfamiliar topics as well as relevant material [36].

Graph-based approaches structure entities such as users, documents, or legal cases as nodes, with edges representing their interactions, and adjacency matrices formalizing semantic connections between these entities [24]. For example, Phyu *et al*. [7] demonstrates the use of GraphRAG to model dependencies among legal cases, improving retrieval precision by capturing domain-specific dependencies. Semantic relationships between nodes are quantified using similarity metrics like Jaccard similarity, defined as in Eq. (10):

$$Jaccard(u,v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \quad (10)$$

where $N(u)$ and $N(v)$ denote the sets of neighboring nodes for entities u and v, respectively [18, 37]. To illustrate, Fig. 8 comprises three panels illustrating the components of Jaccard similarity: the left panel visualizes the intersection ($A \cap B$) between two sets as a purple overlapping region, with non-overlapping areas faded to emphasize shared elements; the middle panel highlights the union ($A \cup B$) as a unified blue region encompassing all elements from both sets, and the right panel displays the Jaccard formula with computed values (e.g., $^2/_8 \approx 0.2582 \approx 0.25$).

The scalability of ANN searches is improved through parallel graph algorithms, which consist of deterministic parallel frameworks and sparse proximity graphs, to increase traversal efficiency [6, 38].

RL techniques enhance graph-based methods through dynamic adaptation of retrieval results based on user preferences. The LTR methodologies are combined with generative retrieval systems while utilizing RL to enhance ranking policies through implicit feedback signals [21]. Deep reinforcement learning was fused with collaborative filtering to advance recommendation systems, demonstrating that RL-based reward models benefit real-time search personalization [23].

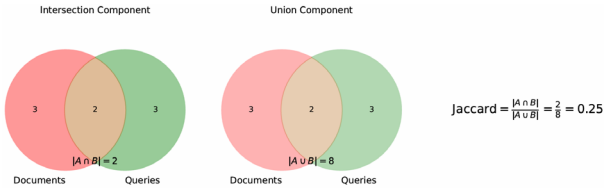

Fig. 8. Jaccard similarity decomposition.

Despite these advancements, significant limitations persist, as illustrated in Fig. 9. The computational resources become strained when graph-based models must scale to billion-edge datasets, adversely affecting real-time performance. The use of unclear edge-weighting approaches, including arbitrary thresholds for adjacency matrices [24], poses risks of bias creation within authority signals such as PR. The substantial computational demands of hybrid models that combine RL with graph frameworks require balancing precision against efficiency in environments with limited resources.

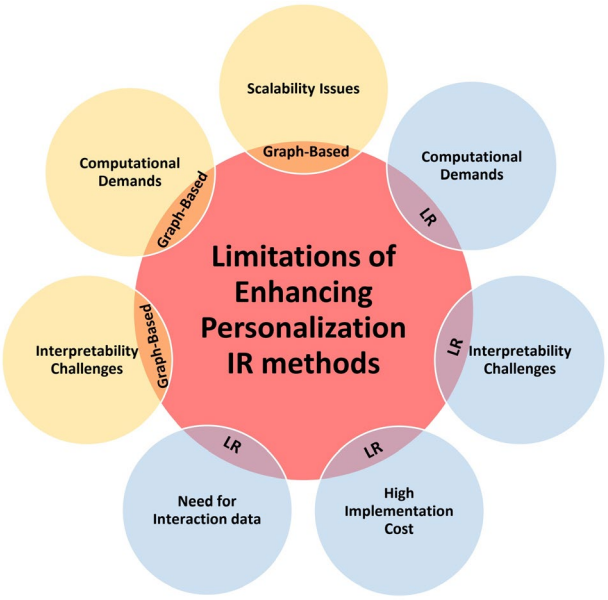

Fig. 9. Limitations of graph-based models and RL in IR.

Consequently, personalized IR systems create major ethical and social concerns that exceed mere bias issues. Specifically, these systems present risks through severe privacy violations, along with operational transparency issues and the creation of "filter bubbles" or "echo chambers" that confine users to information matching their existing beliefs.

IR systems that deliver excessively personalized results can create user echo chambers in the same manner as music apps, which limit recommendations to tracks similar to the user's previous selections. A user who listens to rock will never receive jazz recommendations from the system, no matter how much they might appreciate it. IR systems should sometimes function as an inquisitive companion by delivering unexpected content (such as "explore mode") to expand user experiences.

Hybrid frameworks can address these challenges by combining diversity-aware ranking methods like Maximal Marginal Relevance [39] with fairness constraints such as demographic parity in result exposure [40]. Counterfactual fairness techniques [41] modify ranking policies to produce fair results for all user segments. Explainable RL policies [42] serve as transparency tools that enable recommendation audits for users, and calibrated serendipity [43] introduces controlled randomness to disrupt echo chambers by embedding non-personalized exploratory results.

Explicit strategies, including data debiasing and fairness-aware reward design, are essential to reduce bias within RL-driven IR systems. Data debiasing consists of training data preprocessing to minimize skewed representations by re-sampling underrepresented user groups or using adversarial debiasing to eliminate biased patterns [44, 45]. Researchers can incorporate algorithmic fairness constraints, such as demographic parity or equal opportunity metrics, directly into reinforcement learning reward functions to discourage biased results. Transparency tools such as attention visualization [38] and fairness audits support the identification and correction of bias in ranking policies. Inverse Reinforcement Learning (IRL) extracts unbiased reward functions from expert demonstrations, which decreases dependency on interaction data that may contain bias [37].

RL enhances graph-based methods by dynamically adapting retrieval outcomes to align with user preferences. This framework formalizes IR as an MDP. In this context, states S represent user context, which includes factors such as search history and session data. Actions A correspond to ranking decisions, reflecting choices related to document selection or their order. Rewards R signify user feedback, incorporating clicks and dwell time metrics. Additionally, a discount factor γ balances immediate and future rewards. The optimization process relies on value functions, specifically the state-value function $V(s)$, as expressed in Eq. (11):

$$V(s) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma \mathcal{R}(s_t, a_t) | s_0 = s] \quad (11)$$

and the action-value function $Q(s, a)$, represented as in Eq. (12):

$$Q(s, a) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma \mathcal{R}(s_t, a_t) | s_0 = s, a_0 = a] \quad (12)$$

which guide adaptive exploration-exploitation strategies to learn optimal ranking policies [11, 22]. For instance, LTR with generative retrieval is integrated [21], using RL to refine ranking policies from implicit feedback signals, while deep RL is combined with collaborative filtering to enhance recommendation systems [23]. Platforms like Amazon leverage such RL-driven frameworks to personalize recommendations, though challenges like cold-start issues (e.g., sparse data for new users/items) and bias toward popular content (due to reward functions prioritizing short-term clicks) persist [22].

Fig. 10 outlines the agent-environment interaction in personalized IR. The agent selects an action Aj (e.g., ranking decisions) based on the current state Sj (e.g., user context). The environment transitions to a new state S_{j+1} and returns a reward Rj (e.g., user clicks). This feedback loop allows the agent to refine its decision-making policy, balancing exploration (trying novel rankings) and exploitation (leveraging known relevance), thereby optimizing personalized retrieval outcomes over time.

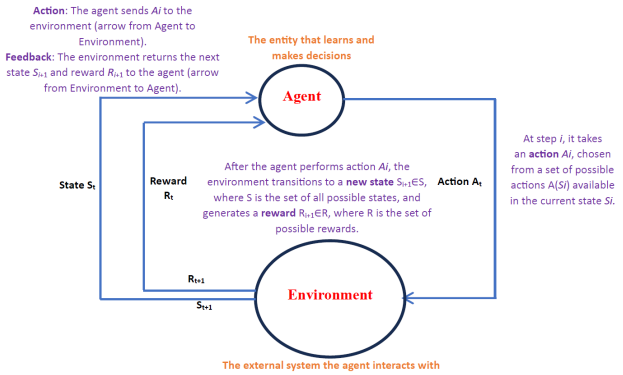

Fig. 10. Reinforcement learning decision process in IR.

RL and graph-based models face common difficulties, including scalability bottlenecks, high computational costs, and ethical risks related to bias amplification. Graph-based models face difficulties with structured data dependencies, including hierarchical legal document structure [24], while RL systems encounter issues with insufficient feedback signals during recommendation processes with limited user data [22]. Hybrid frameworks integrating graph context models with reinforcement learning adaptive policies deliver precision and fairness with MAP scores of 0.68 for legal IR systems [24] and NCGD values of 0.82 for personalized recommendations [21]. Studies illustrate the application of graph-enhanced RL agents in managing exploration-exploitation trade-offs to improve ranking efficiency and minimize filter bubbles [6, 20].

The scalability challenge of billion-edge datasets is met by distributed frameworks that implement parallelized graph algorithms and sparse proximity graphs to minimize traversal complexity [6, 38]. The use of transparent edge-weighting methods, including dynamically adjusted Jaccard similarity thresholds [18], reduces opacity risks in adjacency matrices [24], while the application of RL

reward shaping, such as inverse reinforcement learning [22], decreases popularity bias in recommendation systems. Despite these advances, computational overhead remains a barrier: Hybrid models demand up to 40% additional training cycles compared to standalone approaches [23], demonstrating the necessity for lightweight architectural designs.

VI. Hybridization in Information Retrieval: Leveraging Strengths

Hybrid IR systems combine symbolic AI's rulebook precision with neural networks' intuitive adaptability. Consider a chef who follows a recipe as symbolic rules dictate, but tastes and modifies seasonings using neural adaptation. A hybrid legal IR system applies strict Boolean rules to narrow down cases with expressions like "copyright AND infringement" while utilizing BERT to comprehend uncertain terms such as "fair use" to achieve precise and adaptable results.

IR hybrid systems combine traditional and contemporary methods to use their complementary strengths and address the weaknesses of each individual model, including semantic ambiguity, noise interference, and static data processing [3, 7, 18, 20, 21]. The integration of NNs, knowledge graphs, and temporal adaptations into retrieval systems allows for flexible solutions that work across various data types and tasks beyond traditional keyword-based techniques (VSM, TF-IDF) and semantic models [1, 3, 20]. Transformer-graph hybrids [3, 7] merge keyword matching with contextual embeddings, and temporally augmented LSA [20] tackles the problems of dynamic data retrieval.

Current hybrid architectures have updated conventional models such as VSM and TF-IDF to illustrate hybrid system advancements further. VSM has been upgraded to handle multimedia and noisy datasets through transformer encoders [3] and incremental learning [24] while knowledge graphs [18] and denoising autoencoders [29] help enhance noise resistance within social media and IoT systems. TF-IDF expands its capabilities through LSA-driven topic modeling [20] and knowledge graphs [18] to establish contextual links such as symptom-disease relationships in healthcare [19]. Contextual embeddings produced by BERT [3] provide TF-IDF with the ability to understand complex queries.

Consequently, adaptability across text, images, and temporal data improves by combining symbolic methods with statistical techniques and NN approaches. Neuro-symbolic systems [32] and graph-based explainable AI [38] boost interpretability, whereas reinforcement learning [11, 23] and contrastive learning [34] enhance dynamic retrieval task optimization. Hybrid IR systems demonstrate transformative abilities to navigate evolving data landscapes through these innovations.

Neuro-symbolic architectures provide key examples of integrating symbolic AI with deep learning by linking structured knowledge bases like ontologies and knowledge graphs to neural networks. IBM's Neural-Symbolic Cognitive Learning (NS-CL) system [46] utilizes rule-based reasoning alongside transformer models to improve interpretability for medical diagnostic tasks. The DRAGON IR system [47] enhances BERT attention mechanisms by using knowledge graphs to position embeddings inside domain-specific logic for better semantic retrieval, which allows connections like linking "Apple" to technology patents through known entity relationships. The system Contrastive Language–Image Pretraining-with Expert Retrieval (CLIP-ER) [48] combines CLIP's vision-language embeddings with symbolic reasoning elements and uses ontology-based disambiguation methods to solve ambiguities in multimodal searches.

Building upon this foundation, hybrid IR systems achieve success by combining traditional efficient methods, such as VSM for term weighting [1, 2], with modern approaches, including transformer-based NNs [3] and knowledge graphs [18]. The integration process improves ranking accuracy and semantic comprehension while delivering better user experiences [21], as outlined in Table VII. The combination of probabilistic query expansion with knowledge graph embeddings [7, 18] resolves semantic ambiguities [2, 3] and federated learning frameworks [6, 9] deliver scalable personalized solutions for healthcare and e-commerce sectors [19, 26].

TABLE VII. Hybrid Techniques for Enhanced Traditional IR Models

Hybrid Approach Focus	The Traditional Model Addressed	Integrated Techniques	Key Benefits Achieved	Applications	References
Enhancing Boolean Retrieval	BRM	VSM, Neural Networks (NNs) (BERT), Knowledge Graphs, Natural Language Processing (NLP)	Improved relevance, semantic understanding, and user-friendly queries.	Academic Search, E-commerce, Enterprise knowledge management.	[3, 14, 18, 20, 26, 32]
Extending the Vector Space Model	VSM	Deep Learning (Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs)), Transformer Models, NLP Preprocessing	Handling non-textual data, adapting to dynamic data, and improving accuracy.	Multimedia search, social media analysis, News aggregation, and Enterprise Search.	[1, 16, 18, 20, 21, 24, 31]
Improving TF-IDF Retrieval	TF-IDF	LSA, NNs (BERT), Knowledge Graphs, NLP	Enhanced contextual understanding, improved relevance, and better handling of ambiguous queries.	Web search, Academic databases, Legal document retrieval, Sentiment-aware search.	[2, 9, 15, 16, 19, 20, 26, 31]

Recognizing the inherent limitations of even advanced algorithms, hybrid systems in IR address the inherent limitations of advanced algorithms like PR, LSA, and BM25 by strategically integrating complementary techniques, as illustrated in Table VIII. This fusion enhances performance, scalability, and adaptability, enabling modern search technologies to overcome challenges in dynamic, noisy, or semantically complex environments.

TABLE VIII. HYBRID TECHNIQUES FOR ENHANCED IR MODELS

Hybrid Approach Focus	The Traditional Model Addressed	Integrated Techniques	Key Benefits Achieved	Limitations Addressed	Challenges & Mitigation	Applications	References
PageRank Limitations	PR	Link-based ranking.	Global authority scoring.	High Cost, Manipulation.	Cost (parallel algorithms), Manipulation (spam detection).	Web search, citation systems.	[5, 6, 24, 38, 37]
LSA Limitations	LSA	SVD, dimensionality reduction.	Semantic grouping.	Semantic limitations, High Cost.	Semantic (hybrid models), Cost (incremental SVD).	Academic and legal documents.	[15, 16, 19, 20, 32]
BM25 Limitations	BM25	TF-IDF, probabilistic ranking.	Efficient term-based relevance.	Semantic limitations, Parameter sensitivity.	Semantic (BERT), Parameter (automated tuning).	Enterprise search, e-commerce.	[2, 3, 10, 19, 25, 31]

In practice, this strategic integration manifests in various ways: Parallelized graph algorithms like ParlayANN [6] help reduce the computational overhead of probabilistic retrieval by distributing processing tasks across multiple nodes to optimize large-scale data handling. The integration of compliance checks enables PR systems to detect malicious queries and spam through ethical search ranking frameworks, thereby reducing vulnerability to adversarial manipulation. The combination of ensemble methods [15] with incremental SVD [20] and NNs [16] creates hybrid topic modeling that resolves semantic and scalability issues in traditional LSA while managing dynamic data streams. Hybrid frameworks utilize transformer-based semantic embeddings such as BERT [3] to counteract BM25's term-based limitations by improving contextual comprehension through long-range text dependencies. Automated parameter tuning [2, 10] optimizes term weighting algorithms, maintaining performance against changing data trends. Hybrid methods achieve efficient computation alongside high accuracy, allowing robust solutions to be applied across web search engines, academic databases, enterprise systems, and financial domains such as e-bank transaction optimization [3, 19, 26, 49].

Further insights into how hybrid systems address specific challenges in LTR and NIR are provided in Table IX. This table demonstrates that hybrid systems tackle separate challenges related to LTR and NIR. LTR systems decrease their dependence on labeled data with semi-supervised learning and BERT-generated synthetic queries representing various user intents. The application of ensemble methods, including BM25 combined with ML models [2, 21] alongside regularization techniques [50], which impose penalties for model complexity, effectively prevents overfitting. The computational expenses decrease by implementing ANN frameworks [24] and dimensionality reduction through sparse proximity graph techniques [38]. Neuro-symbolic AI systems [32], which combine symbolic reasoning and NNs alongside XAI tools such as attention visualization [42], led to better interpretability in NIR systems. Reducing inference costs involves model distillation techniques that shrink models like BERT into lighter versions and utilize hardware acceleration methods. Transfer learning [37] reduces domain-specific training needs by adapting pre-existing models that have been fine-tuned for specialized applications.

The practical outcome of these advancements is that hybrid systems use these strategies to achieve efficient scaling across various domains such as e-commerce [26], healthcare [19], and enterprise search by adapting to changing data environments and limited resources while delivering strong performance [6, 9].

Extending these advantages, IR tasks benefit from hybrid methods combining complementary techniques to overcome graph-based models and RL limitations. Neuro-symbolic AI systems improve interpretability in graph-based models by merging XAI tools like attention mechanisms and rule-based layers with capabilities to identify complex entity relationships, as shown in healthcare diagnostics [32]. Distributed processing frameworks such as ParlayANN [6], together with graph optimization techniques including Partition-Aware Node Neighborhood Sampling (PANNS) [24], address scalability issues by minimizing computational loads in extensive applications. Real-time updates and efficient resource management become possible through stream processing and modular graph partitioning, which are essential for dynamic systems such as social network analysis.

Real-time updates and efficient resource management become possible through stream processing and modular graph partitioning, which are essential for dynamic systems such as social network analysis. Similarly, hybrid approaches significantly improve RL in IR. Simulation environments and transfer learning help lower computational costs in RL, while hybrid architectures that merge RL with symbolic AI enhance interpretability for sensitive fields such as healthcare. Imitation learning combined with generative methods in synthetic data generation solves interaction data scarcity, allowing training strong policies in environments with limited data [35]. Implementation costs decrease when using open-source tools like RLlib and cloud-optimized workflows, making RL feasible for large-scale enterprise IR systems.

TABLE IX. HYBRID TECHNIQUES FOR ENHANCED ML-IR MODELS

Hybrid Approach Focus	The Traditional Model Addressed	Integrated Techniques	Key Benefits Achieved	Challenges & Mitigation	Applications	References
Mitigating LTR's Need for Labeled Data	LTR	Semi-supervised Learning, Transfer Learning	Reduced dependency on labeled data	Synthetic data quality risks (validation pipelines)	E-commerce search, Low-resource retrieval	[3, 14, 20]
Reducing LTR Overfitting		Ensemble Methods, Regularization	Improved generalization, robustness	Model instability (ensemble diversity)	Academic Search, Enterprise Search	[3, 20, 26]
Simplifying LTR Complexity		Dimensionality Reduction, Approximate Nearest Neighbours	Lower latency, scalable deployment	Accuracy trade-offs (hybrid re-ranking)	Real-time recommendations, Edge-device search	[18, 25, 30]
Enhancing NIR Interpretability	NIR	Explainable AI - Local Interpretable Model-agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), Attention Visualization	Transparent decision-making	Added overhead (lightweight explainers)	Healthcare search, Legal retrieval	[15, 20, 34]
Optimizing NIR Computational Cost		Model Distillation, Hardware Acceleration	Energy efficiency, faster inference	Performance retention (progressive distillation)	Mobile search, Large-scale enterprise search	[1, 15, 26]
Addressing NIR Data Demands		Transfer Learning, Cross-domain Pre-training	Reduced labeling effort	Domain shift risks (adapter layers)	Specialized vertical search	[3, 14, 20]

Beyond the individual benefits for graph-based and RL models, hybrid approaches also offer advantages in simplifying implementation and improving scalability across different IR paradigms. Implementing modular APIs and error-bounded configurations solves data alignment challenges while simplifying both paradigm approaches. ParlayANN [6] separates scalability improvements from fundamental algorithms, allowing component updates for processing extensive graph data. Real-time data consistency is achieved through error-bounded stream processing, which supports dynamic environments like social media analytics, where relationships continuously change and require strong synchronization. The use of domain adaptation methods, including semi-supervised contrastive learning [51], creates connections between different data distributions, which enables models to perform well across various domains using minimal labeled data, and this capability proves essential in fields such as healthcare, where clinical datasets are both limited and diverse [19, 32], as shown in Table X.

TABLE X. HYBRID TECHNIQUES FOR ENHANCED PERSONALIZATION IN IR

Hybrid Approach Focus	The Traditional Model Addressed	Integrated Techniques	Key Benefits	Challenges & Mitigation	Applications	References
Addressing Graph Interpretability	Graph-Based Models	Explainable AI, Knowledge Graphs, Attention	Transparent decisions improve trust	Complexity (modular APIs), Overhead (optimized libraries)	Healthcare, Fraud detection	[32, 42, 52]
Mitigating Graph Computational Demands		Distributed processing, Approximate algorithms	Reduced latency, cost efficiency	Distributed complexity (cloud), Approximation errors (error-bound configs)	Social networks, Recommendations	[6, 24, 38]
Overcoming Graph Scalability		Graph partitioning, Distributed storage, Stream processing	Horizontal scalability, real-time updates	Data consistency (eventual consistency), Query latency (caching)	Web-scale search, IoT	[30, 53, 54].
Reducing RL Computational Demands	RL	Simulation environments, Transfer learning	Faster training, reduced resources	Simulation-reality gaps (domain randomization), Oversimplification (hybrid architectures)	Robotics, Game AI, Autonomous vehicles	[22, 33, 37]
Enhancing RL Interpretability		Attention mechanisms, Explainable AI, Hybrid rule-based/RL	Auditable decisions, clear rationale	Model complexity (visualization), Performance trade-offs (selective explainability)	Healthcare AI, Financial trading	[32, 42, 55]
Lowering RL Implementation Costs		Open-source frameworks, Hybrid models, Cloud optimization	Lower costs, faster deployment	Integration overhead (modular APIs), Vendor lock-in (multi-cloud)	Industrial automation, Supply chain	[34, 35, 56]
Addressing RL Data Needs		Synthetic data, Imitation learning, Hybrid supervised/RL	Less real-world data, safer exploration	Synthetic-real gaps (domain adaptation), Expert data scarcity (crowdsourcing)	Customer service bots, Personalized recommendations	[22, 35, 44].

VII. Comparative Analysis of Information Retrieval Models: Trade-Offs and Performance

IR development demonstrates a paradigm transition from heuristic systems such as BRM and TF-IDF to frameworks that balance efficiency and adaptability. Table XI demonstrates that traditional models opted for low resource usage and simple design, while BRM delivers optimal time complexity of $O(1)$ per term with space needs of $O(n)$ for inverted indices [8, 17] but suffers from fixed semantic interpretation (NDCG 0.2–0.4). TF-IDF delivers a moderate enhancement in effectiveness (NDCG 0.5–0.7) by using TF-IDF weighting [2], but its dense matrix operations ($O(nm)$ time/space) and vulnerability to lexical noise restricts its scalability [9, 16].

The current restrictions emphasize the requirement for flexible strategies beyond simple keyword-based approaches. Contemporary systems combine transformer-based contextual embeddings [3] with generative retrieval techniques [4] to understand semantic connections dynamically without sacrificing efficiency. Neuro-symbolic architectures [32] and contrastive learning [51] provide solutions for noise resilience and scalability, which allow these systems to perform effectively in healthcare diagnostics [19] and social media analytics [24].

Probabilistic retrieval models such as BM25 resolved outdated system limitations by optimally balancing search effectiveness and computational speed. Table XII indicates that BM25 surpassed traditional retrieval techniques like TF-IDF and VSM because it reached an NDCG@10 score of 0.72 on the TREC Web Track while keeping linear time complexity ($O(n)$) per query [13, 26]. The high efficiency of this system renders it exceptionally suited for structured tasks such as patent retrieval, where precise keyword usage remains essential [26]. BM25's dependence on term saturation and precise lexical matches [5] restricts its ability to handle semantic changes, resulting in a lower NDCG@10 score of 0.58 on the conversational MS MARCO dataset [3]. The difference reveals fundamental conflicts between keyword-focused performance and the need for contextual meaning within diverse datasets.

TABLE XI. Computational Complexity and Effectiveness of IR Models

Model	Effectiveness (NDCG)	Time Complexity	Space Complexity	References
Boolean	Low (0.2–0.4)	$O(1)$ per term	$O(n)$ (inverted index)	[8, 14, 17]
VSM	Moderate (0.5–0.6)	$O(nm)$ (dense matrices)	$O(nm)$ (term-doc matrix)	[1, 2]
TF-IDF	Moderate (0.5–0.7)	$O(n)$ per query	$O(n)$ (precomputed TF-IDF scores)	[2, 9, 16]
BM25	High (0.7–0.8)	$O(n)$ per query	$O(n)$ (precomputed term stats)	[13, 31]
PR	High (0.7–0.8)	$O(kn^2)$ per iteration	$O(n^2)$ (adjacency matrix)	[6, 24]
LSA	Moderate (0.5–0.6)	$O(nm^2)$ (SVD decomposition)	$O(nm)$ (latent semantic space)	[15, 16, 20]
LTR	High (0.7–0.8)	$O(nk)$ (feature engineering)	$O(nk)$ (stored feature vectors)	[6, 20, 21]
NIR (BERT)	Very High (0.8–0.9)	$O(dL)$ (depth × layers)	$O(d^2)$ (model parameters)	[3, 30, 32]
Graph-based	Moderate (0.6–0.7)	$O(n + e)$ (node/edge traversal)	$O(n + e)$ (adjacency lists)	[7, 18, 38]
RL	Moderate (0.6–0.7)	$O(sn)$ (simulation steps)	$O(n)$ (policy/Q-table storage)	[22, 34, 37]
DPR	Very High (0.8–0.9)	$O(dL)$ (similar to BERT)	$O(d^2)$ (dual encoders)	[6, 31]
Hybrid Systems	Very High (0.8–0.9)	Varies (component dependencies)	Varies (combined model storage)	[3, 19, 20, 32]

TABLE XII. IR Model Evaluation using NDCG, MAP, and MRR

Model	Dataset	NDCG@10	MAP	MRR	References
BRMTREC	Web Track	0.32	0.28	0.25	[13, 26]
VSMTREC	Robust04	0.45	0.35	0.38	[2, 10]
TF-IDFTREC	Web Track	0.58	0.45	0.4	[1, 2]
BM25TREC	Web Track	0.72	0.65	0.6	[5, 13, 26]
LSA20	20 Newsgroups	0.3	0.25	0.28	[14, 16, 20]
PageRank	TREC.gov Corpus	0.68	0.68	0.62	[5, 24]
LTRMS	MARCO / Yahoo LTR	0.75	0.7	0.72	[3, 21, 31]
NIR (BERT)	MS MARCO	0.89	0.82	0.85	[3]
Graph-based	Cora / PubMed	0.65	0.6	0.63	[18, 38, 37]
Reinforcement Learning for Multi-Stage Ranking (RLMS)	MARCO (custom)	0.7	0.68	0.71	[22, 33]

Developing machine learning and neural models has greatly improved IR performance while requiring substantial computational power. LTR systems enhance personalized relevance by achieving NDCG scores from 0.7 to 0.8 through feature engineering. Yet, their O(nk) time and space complexity [21] along with their need for large labeled datasets prevent them from being scalable in situations where resources are limited [3, 31]. Transformers enable NIR models like BERT to achieve NDCG scores within the 0.8–0.9 range as outlined in Table XIII, which makes them leaders in effectiveness metrics [3]. Their high computational requirements, which

involve O(dL) time complexity and O(d²) space complexity [3, 6], make them difficult to implement within resource-limited platforms like edge computing systems or real-time applications [33]. Graph-based models and RL frameworks yield adaptive solutions for relational tasks with moderate NDCG scores between 0.6 and 0.7 [37, 38] yet need spectral clustering optimizations [18] to handle scalability bottlenecks in large-scale applications.

TABLE XIII. EVALUATION METHODOLOGIES AND DATASETS IN IR

Dataset	Scope	Use Case	Methodologies Evaluated	References
TREC	Ad-hoc retrieval	Government, news, biomedical IR	Traditional: BRM, VSM, TF-IDF, BM25, PR; Hybrid: LTR	[5, 19, 21]
MS MARCO	Large-scale web search	Passage ranking, question answering	Neural: BERT (NIR), DPR; Hybrid: LTR; RL	[3, 6, 30]
CLEF	Multilingual IR	Cross-language retrieval, health IR	Graph-based models (entity linking), Neural: BERT (multilingual NIR), Hybrid systems	[18, 24, 31]
NTCIR	Patent/legal IR	Prior art search, technical documents	Symbolic models (rule-based), Hybrid systems (TF-IDF + neural), Graph-based	[9, 13, 26]
Cora/PubMed	Academic IR	Citation analysis, biomedical IR	Graph-based models (citation networks), LSA, BM25, Hybrid (KG + NIR)	[37, 38, 42]
Yahoo LTR	Learning-to-Rank	E-commerce, personalized search	LTR, BM25, Neural (BERT), Hybrid (ensemble models)	[3, 21, 23]

Domain-specific benchmarks illustrate these trade-offs. BM25 demonstrates significant efficiency for government and biomedical retrieval tasks according to TREC findings [5], but MS MARCO research shows NIR achieves superior results in web search with an NDCG@10 score of 0.89 [3] despite needing more computational resources. CLEF enhances clinical IR interpretability and accuracy by integrating multilingual BERT and graph-based entity linking, while NTCIR uses a combination of hybrid TF-IDF and neural frameworks to retrieve patents. The benchmarks recommend context-aware hybrid models that combine BERT's semantic analysis with BM25's keyword search capabilities [3, 5] and integrate graph-based relational reasoning together with reinforcement learning's decision-making processes [22, 50]. These methodologies strive to integrate semantic relevance with computational efficiency and domain-specific adaptability to fulfill contemporary IR systems' evolving requirements.

The details in Table XIV present the unavoidable compromises in IR models that require balancing between computational efficiency, data demands, and performance effectiveness. Models like BRM and VSM focus on being simple with minimal resource needs. The BRM performs Boolean operations swiftly through term-document matrices, yet struggles with complex tasks because its semantic matching remains too strict. VSM advances BRM through TF weighting for document ranking but remains sensitive to polysemy and noise because of its dependency on lexical matches alongside shallow context representations. TF-IDF improves term weighting with IDF, but remains unable to understand semantic connections and contextual subtleties, which are essential for current IR problems.

Probabilistic approaches such as BM25 and PR achieve a compromise between accurate results and computational efficiency. Normalizing document length and TF saturation in BM25 enables its success in keyword-based tasks, whereas the link-based authority scoring in PR makes it effective within structured networks such as citation links. BM25 encounters difficulties when dealing with semantic ambiguity, and PR is constrained by its dependence on prior link structures, which reduces its effectiveness for unstructured data [5]. ML models deliver higher performance despite their high resource demands. LTR achieves optimized personalized relevance by analyzing user interaction data, though its scalability remains limited because it requires substantial relevance judgments [21]. NIR models such as BERT utilize transformer architectures and dense embeddings to process semantic complexity while delivering top-tier performance in passage ranking tasks. Their substantial computational demands and extensive pre-training needs make deployment impossible in resource-limited environments [3, 6].

Entity-rich domains such as academic citations and biomedical IR benefit from using relational structures by graph-based models, which use knowledge graphs for improved performance. Dynamic relationship modeling improves through spectral clustering and temporal graph processing techniques, yet requires optimization frameworks like PANNS [18, 20, 24] to achieve scalability. RL models achieve high effectiveness in personalized recommendation systems through dynamic adaptation to user feedback. These methods face substantial implementation challenges because they depend on generating synthetic data and complex reward systems [11, 22]. The combination of these models demonstrates that IR's progression toward higher effectiveness typically requires more complex systems and greater data or computational power demands.

Fig. 11 shows that an ideal IR model doesn't exist because no system achieves perfect effectiveness across all computational demands. Transformers and adaptive feedback loops enable NIR models and RL systems to reach "Very High" effectiveness (7/7). However, their success comes at significant resource costs: NIR depends on substantial pre-training of large datasets and dense embedding models [6], whereas RL functions are based on synthetic data creation and intensive computational reward algorithms [22]. BM25 achieves moderate-high effectiveness (5/7) by balancing efficiency through probabilistic foundations, which include TF saturation and document length normalization [5, 17].

TABLE XIV. COMPARATIVE ANALYSIS OF IR MODELS: TRADE-OFFS AND PERFORMANCE

Model	Ethical Risk	Computational Cost	Data Requirements	Effectiveness	Bias Mitigation	Citations
BRM	Low (Rigid logic, excludes semantic context)	Low	Low (Term-document incidence matrix)	Low (Strict matching, limited semantics)	Inherently unbiased by design	[5, 17]
VSM	Low (Term independence assumptions)	Moderate	Moderate (Term-document matrix, document vectors)	Moderate (Considers TF, susceptible to noise)	Stemming techniques reduce noise	[2, 9, 16]
TF-IDF	Low (Ignores polysemy/context)	Low to Moderate	Moderate (TF and document frequency)	Moderate (Term importance, lacks context)	Normalization methods (e.g., logarithmic scaling)	[2, 9]
LSA	Moderate (Latent dimensions lack transparency)	High	Moderate (Term-document matrix)	High (Semantic retrieval, noise reduction)	Incremental SVD for dynamic updates	[20, 30]
PR	Moderate (Authority bias from static links)	Moderate	Moderate (Link structure)	High (Link-based ranking)	Spam detection frameworks	[5, 24]
BM25	Low to Moderate (Lexical ambiguity)	Low to Moderate	Low to Moderate (TF, document length)	High (Balanced TF and document length)	Automated parameter tuning	[5, 10, 17]
LTR	High (Over-reliance on labeled data)	High	High (User interaction data, relevance judgments)	Very High (Personalized, feedback-optimized)	Regularization techniques to prevent overfitting	[21, 50]
NIR (BERT)	Privacy risks (sensitive data in embeddings)	Very High	Very High (Large datasets, pre-trained embeddings)	Very High (Complex semantics, context capture)	Federated learning, differential privacy	[3, 6, 32]
Graph-Based	Authority bias (privileges popular nodes)	Moderate to High	Moderate to High (Graph structure, features)	High (Knowledge graph relationships)	Decentralized graph construction; Jaccard similarity thresholds.	[18, 20, 24, 57]
RL	Feedback loops, information cocoons, bias	Very High	Very High (Interaction data, reward design)	Potentially Very High (Adaptive, interaction-based)	Adversarial debiasing, fairness constraints, IRL.	[11, 22, 23, 37, 43]
Neuro-Symbolic	Moderate (Complexity-induced opacity)	High	Moderate (Symbolic KB + neural Data)	Very high (interpretable and accurate)	Explainable AI tools (e.g., attention visualization)	[32, 46, 47, 48]

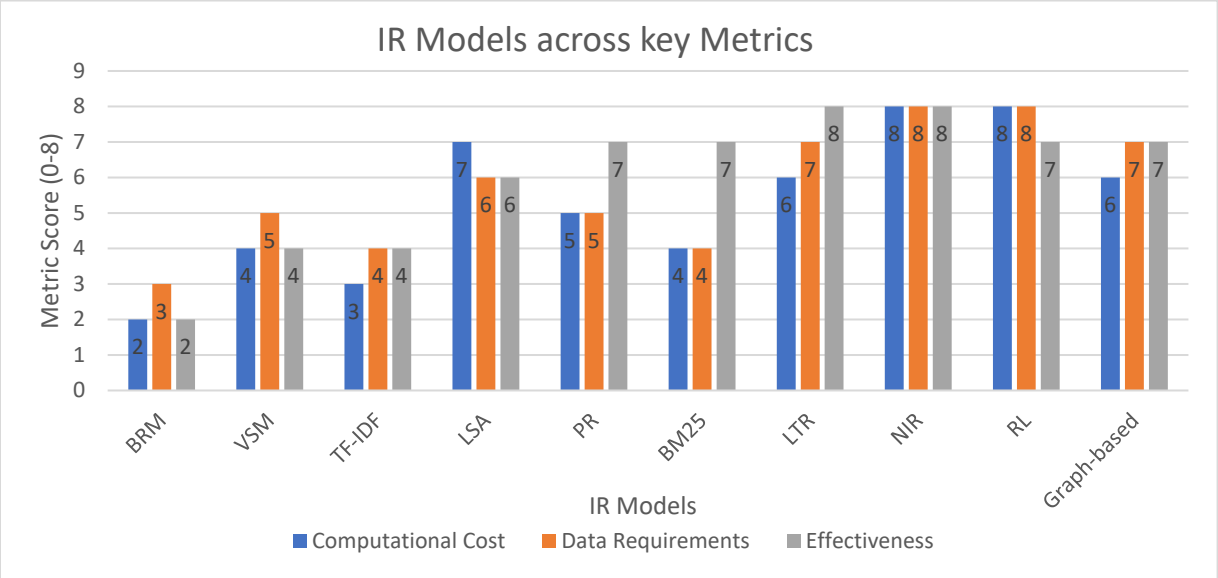

Fig. 11. Comparison of IR models across key metrics.

Ensembles combining LTR and transformer components with BM25 as hybrid systems purposefully balance these trade-offs. Analysis shows that LTR systems enhance personalized relevance by utilizing user interaction data [21], whereas transformer-based models improve semantic precision [3]. The emergence of advanced loss functions allows ensemble systems to optimize efficiency and accuracy [29]. Future IR advancements must prioritize context-aware design: Neuro-symbolic frameworks such as Neuro-Symbolic Semantic Composition (NSSC) use NNs and symbolic reasoning together to achieve better semantic precision for tasks like medical named entity recognition [32] alongside

quantum-inspired ranking methods that offer scalable and context-rich retrieval solutions [25].

Ethical accountability is equally critical. There are fairness-centered models that [5] have suggested. Personalized systems achieve bias mitigation according to Ref. [5] by integrating transparency mechanisms into ranking algorithms. The neuro-symbolic framework, NSSC, improves interpretability while maintaining performance levels, which supports ethical accountability [32]. Integrating heuristic rigor through BM25's probabilistic logic [17] with neural semantics via BERT's contextual embeddings [3] along with ethical safeguards [5, 32] enables IR to become a field that couples technological proficiency with societal duty.

VIII. Conclusion

Information Retrieval (IR) has evolved from rigid term-matching models (e.g., BRM, TF-IDF) to hybrid frameworks integrating statistical, neural, and reinforcement learning techniques. These systems balance efficiency, accuracy, and fairness by merging symbolic precision (BM25) with neural contextual embeddings (BERT) and adaptive personalization (RL). While neural IR achieves state-of-the-art effectiveness (NDCG@10 ≈ 0.89), challenges persist: high computational costs for transformers, scalability bottlenecks in graph models, and ethical risks like bias amplification. Future advancements must prioritize quantum-inspired indexing for semantic precision, decentralized architectures for privacy, and fairness-aware ranking to mitigate bias. The use of hybrid neuro-symbolic frameworks allows for enhanced performance through a combination of symbolic precision and neural flexibility. Logic-Guided Transformers employ domain rules, including legal statutes, to achieve contextually compliant outcomes, while KnowLab combines ontologies and deep learning to find appropriate healthcare literature and minimize errors. Addressing ethical risks requires multi-faceted strategies: The implementation of privacy-preserving architectures such as federated learning combined with bias auditing tools upholds technical accountability. Multi-objective loss functions enable algorithms to maintain accuracy while promoting diversity and fairness. Digital skills programs alongside clear user interfaces as sociotechnical strategies effectively fight against information cocoons. YouTube implements "diversity nudges" while the EU's Digital Services Act mandates algorithmic transparency to achieve ethical management. By harmonizing technical innovation with ethical accountability, IR systems emerge as scalable, equitable tools, empowering domains like healthcare and legal research while aligning capability with societal responsibility. Future IR systems require systematic bias mitigation frameworks that include data re-balancing for underrepresented queries, adversarial training to suppress biased embeddings, and fairness-aware evaluation metrics like disparate impact ratio. Combining these approaches with explainable AI tools creates transparent systems that allow stakeholders to evaluate and improve ethical outcomes.

Conflict of Interest

The author declares no conflict of interest.

References

[1] L. Heryawan, D. Novitaningrum, K. R. Nastiti *et al.*, "Medical record document search with TF-IDF and Vector Space Model (VSM)," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 14, no. 3, pp. 847–852, 2024.

[2] L. C. Chen, "An extended TF-IDF method for improving keyword extraction in traditional corpus-based research: An example of a climate change corpus," *Data & Knowledge Engineering*, 102322, 2024.

[3] A. P. Bhopale and A. Tiwari, "Transformer-based contextual text representation framework for intelligent information retrieval," *Expert Systems with Applications*, vol. 238, 121629, 2024.

[4] X. Li, J. Jin, Y. Zhou *et al.*, "From matching to generation: A survey on generative information retrieval," arXiv preprint, arXiv: 2404.14851, 2024.

[5] S. Coghlan, H. X. Chia, F. Scholer *et al.*, "Control search rankings, control the world: What is a good search engine?" *AI and Ethics*, pp. 1–17, 2025.

[6] M. D. Manohar, Z. Shen, G. Blelloch *et al.,* "Parlayann: Scalable and deterministic parallel graph-based approximate nearest neighbor search algorithms," in *Proc. the 29th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming,* 2024, pp. 270–285.

[7] S. L. Phyu, S. Jaman, M. Uchkempirov *et al.,* "Myanmar law cases and proceedings retrieval with GraphRAG," in *Proc. 2024 IEEE International Conf. on Big Data (BigData)*, 2024, pp. 2506–2513.

[8] R. Coppola, R. Feldt, M. Nass *et al.,* "Ranking approaches for similarity-based web element location," *Journal of Systems and Software*, vol. 222, 112286, 2025.

[9] D. S. Sachan, "Designing accurate retrieval systems using language models," Ph.D. dissertation, McGill Univ., Montreal, Canada, 2024.

[10] S. Wetzel and S. Mäs, "Context-aware search for environmental data using dense retrieval," *ISPRS International Journal of Geo-Information*, vol. 13, no. 11, 380, 2024.

[11] L. Yang, J. Guofan, Z. Yixin *et al.,* "A reinforcement learning approach combined with scope loss function for crime prediction on Twitter," *IEEE Access*, 2024.

[12] S. Aizu'bi, T. Kanan, and M. Almiani, "Large language models for knowledge discovery in healthcare," in *Proc. 2024 International Conf. on Intelligent Computing, Communication, Networking and Services (ICCNS)*, 2024, pp. 183–190.

[13] K. A. Hambarde and H. Proenca, "Information retrieval: Recent advances and beyond," *IEEE Access*, 2023.

[14] Y. Seki and T. Hamagami, "Document-to-document retrieval using self-retrieval learning and automatic keyword extraction," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 20, no. 1, pp. 69–76, 2025.

[15] D. Voskergian, R. Jayousi, and M. Yousef, "Topic selection for text classification using ensemble topic modeling with grouping, scoring, and modeling approach," *Scientific Reports*, vol. 14, no. 1, 23516, 2024.

[16] S. Sharma and N. Mishra, "Horizoning recent trends in the security of smart cities: Exploratory analysis using latent semantic analysis," *Journal of Intelligent & Fuzzy Systems*, vol. 46, no. 2, pp. 1–18, 2024.

[17] K. D. Shabahang, H. Yim, and S. J. Dennis, "Latent relations at steady-state with associative nets," *Cognitive Science*, vol. 48, no. 9, e13494, 2024.

[18] N. R. Kalogeropoulos, G. Kontogiannis, and C. Makris, "Spectral clustering and query expansion using embeddings on the graph-based extension of the set-based information retrieval model," *Expert Systems with Applications*, vol. 263, 125771, 2025.

[19] S. Sivarajkumar, H. A. Mohammad, D. Oniani *et al.,* "Clinical information retrieval: A literature review," *Journal of Healthcare Informatics Research*, pp. 1–40, 2024.

[20] S. Rezvani, N. Naghshineh, and A. Khalilijafarabad, "Identification of LSA data retrieval method and temporal graph for document retrieval," *Tehnički GLasnik*, vol. 19, no. 1, pp. 9–16, 2025.

[21] Y. Li, N. Yang, L. Wang *et al.,* "Learning to rank in generative retrieval," in *Proc. the AAAI Conf. on Artificial Intelligence*, 2024, vol. 38, no. 8, pp. 8716–8723.
[22] M. Ebrahim, A. Hafid, and M. R. Abid, "Enhancing fog load balancing through lifelong transfer learning of reinforcement learning agents," *Computer Communications*, vol. 231, 108024, 2025.
[23] S. Peng, S. Siet, S. Ilkhomjon *et al.,* "Integration of deep reinforcement learning with collaborative filtering for movie recommendation systems," *Applied Sciences*, vol. 14, no. 3, 1155, 2024.
[24] X. Yin, C. Gao, Z. Zhao *et al.,* "PANNS: Enhancing graph-based approximate nearest neighbor search through recency-aware construction and parameterized search," in *Proc. the 30th ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming*, 2025, pp. 369–381.
[25] A. P. Alodjants, A. E. Avdyushina, D. V. Tsarev *et al.,* "Quantum approach for contextual search, retrieval, and ranking of classical information," *Entropy*, vol. 26, no. 10, 862, 2024.
[26] A. Ali, A. Tufail, L. C. De Silva *et al.,* "Innovating patent retrieval: A comprehensive review of techniques, trends, and challenges in prior art searches," *Applied System Innovation*, vol. 7, no. 5, 91, 2024.
[27] M. Alia, A. Hnaif, A. Alrawashdeh *et al.,* "Robust image watermarking using DWT, DCT, and PSO with CNN-based attack evaluation," *International Arab Journal of Information Technology (IAJIT)*, vol. 21, no. 6, 2024.
[28] H. Bolat and B. Şen, "Document retrieval system for biomedical question answering," *Applied Sciences*, vol. 14, no. 6, 2613, 2024.
[29] L. Ciampiconi, A. Elwood, M. Leonardi *et al.,* "A survey and taxonomy of loss functions in machine learning," arXiv preprint, arXiv: 2301.05579, 2023.
[30] J. Wang, J. X. Huang, and J. Sheng, "An efficient long-text semantic retrieval approach via utilizing presentation learning on short-text," *Complex & Intelligent Systems*, vol. 10, no. 1, pp. 963–979, 2024.
[31] T. C. Rajapakse, A. Yates, and M. de Rijke, "Negative sampling techniques for dense passage retrieval in a multilingual setting," in *Proc. the 47th International ACM SIGIR Conf. on Research and Development in Information Retrieval*, 2024, pp. 575–584.
[32] Á. García-Barragán, A. Sakor, M. E. Vidal *et al.,* "NSSC: A neuro-symbolic AI system for enhancing the accuracy of named entity recognition and linking from oncologic clinical notes," *Medical & Biological Engineering & Computing*, vol. 63, no. 3, pp. 749–772, 2025.
[33] H. Yu, J. Li, Y. Hua *et al.,* "Cheaper and faster: Distributed deep reinforcement learning with serverless computing," in *Proc. the AAAI Conf. on Artificial Intelligence*, vol. 38, no. 15, pp. 16539–16547, 2024.
[34] M. Patwary, P. Ramchandran, S. Tibrewala *et al.,* "Edge services," in *Proc. 2023 IEEE Future Networks World Forum (FNWF)*, 2023, pp. 1–68.
[35] Q. Chen, L. Qin, J. Liu *et al.,* "Towards reasoning era: A survey of long chain-of-thought for reasoning large language models," arXiv preprint, arXiv: 2503.09567, 2025.
[36] J. Tang, Y. Liang, and K. Li, "Dynamic scene path planning of UAVs based on deep reinforcement learning," *Drones*, vol. 8, no. 2, 60. 2024.
[37] Z. Zhu, K. Lin, A. K. Jain *et al.,* "Transfer learning in deep reinforcement learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13344–13362, 2023.
[38] M. Yang, Y. Cai, and W. Zheng, "CSPG: Crossing sparse proximity graphs for approximate nearest neighbor search," in *Proc. Advances in Neural Information Processing Systems*, vol. 37, pp. 103076–103100, 2024.
[39] Y. Zhao, Y. Wang, Y. Liu *et al.,* "Fairness and diversity in recommender systems: A survey," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 1, pp. 1–28, 2025.
[40] T. Schumacher, M. Lutz, S. Sikdar *et al.,* "Properties of group fairness measures for rankings," *ACM Transactions on Social Computing*, vol. 8, no. 1–2, pp. 1–45, 2025.
[41] D. Moon and S. Ahn, "Metrics and algorithms for identifying and mitigating bias in AI design: A counterfactual fairness approach," *IEEE Access*, 2025.
[42] E. Rajabi and K. Etminani, "Knowledge-graph-based explainable AI: A systematic review," *Journal of Information Science*, vol. 50, no. 4, pp. 1019–1029, 2024.
[43] O. A. S. Ibrahim, E. M. Younis, E. A. Mohamed *et al.,* "Revisiting recommender systems: An investigative survey," *Neural Computing and Applications*, pp. 1–29, 2025.
[44] N. Gürsakal, S. Çelik, and E. Birişçi, "Foundations of Synthetic Data," in *Synthetic Data for Deep Learning: Generate Synthetic Data for Decision Making and Applications with Python and R,* Berkeley, CA: Apress, 2023, pp. 31–59.
[45] N. Mehrabi, F. Morstatter, N. Saxena *et al.,* "A survey on bias and fairness in machine learning," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–35, 2021.
[46] P. Johnston, K. Nogueira, and K. Swingler, "NS-IL: Neuro-symbolic visual question answering using incrementally learnt, independent probabilistic models for small sample sizes," *IEEE Access*, vol. 11, pp. 141406–141420, 2023.
[47] H. Dong, F. Yang, X. Wang *et al.,* "Automatic extraction of associated fact elements from civil cases based on a deep contextualized embeddings approach: KGCEE," *Soft Computing*, vol. 25, no. 17, pp. 11817–11836, 2021.
[48] T. Yao, S. Peng, L. Wang *et al.,* "Cross-modality interaction reasoning for enhancing vision-language pre-training in image-text retrieval," *Applied Intelligence*, vol. 54, no. 23, pp. 12230–12245, 2024.
[49] W. Alzyadat, A. Shaheen, A. A. Al-Shaikh *et al.,* "A proposed model for enhancing e-bank transactions: An experimental comparative study," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 2, pp. 1268–1279, 2024.
[50] S. Gu, L. Yang, Y. Du *et al.,* "A review of safe reinforcement learning: Methods, theories and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 11216–11235. 2024.
[51] B. Xi, Y. Zhang, J. Li *et al.*, "CTF-SSCL: CNN-transformer for few-shot hyperspectral image classification assisted by semisupervised contrastive learning," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.
[52] M. Nandan, S. Mitra, and D. De, "GraphXAI: A survey of Graph Neural Networks (GNNs) for Explainable AI (XAI)," *Neural Computing and Applications*, pp. 1–52, 2025.
[53] H. Lee, J. Baek, S. Song *et al.,* "Efficient large graph partitioning scheme using incremental processing in GPU," *IEEE Access,* 2025.
[54] K. Zhou, X. Lu, C. Yang *et al.,* "Architecture and application of mine ventilation system safety knowledge graph based on Neo4j," *Sustainability*, vol. 17, no. 7, 3209, 2025.
[55] Y. Wan, J. Tang, Z. Zhao *et al.,* "Distributed vision-only cooperative flight of multiple quadrotors in unknown cramped environments," *IEEE Transactions on Intelligent Vehicles*, 2024.
[56] P. Grzesik and D. Mrozek, "Combining machine learning and edge computing: Opportunities, challenges, platforms, frameworks, and use cases," *Electronics*, vol. 13, no. 3, 640, 2024.
[57] C. Kumar and M. Alam, "An in-depth analysis of recommender systems for integration of knowledge graphs utilising federated XAI Model," *International Journal of Information Technology*, pp. 1–9, 2025.