Implicit vs. Explicit Feedback in Recommender Systems: A Comprehensive Survey and Unified Framework

MAHAMUDUL HASAN, University of Minnesota Twin Cities, USA

Recommender systems have evolved into critical infrastructure for modern digital platforms, with user feedback serving as the fundamental data source driving personalization algorithms. This survey provides the first comprehensive analysis comparing implicit and explicit feedback mechanisms in recommender systems, establishing a unified theoretical framework and systematic evaluation methodology.

We present a comprehensive taxonomy that categorizes feedback along multiple dimensions: collection mechanism, signal quality, temporal characteristics, and user cognitive load. Through systematic analysis of 150+ research papers spanning 2010-2025, we identify key algorithmic paradigms, evaluation challenges, and emerging research directions.

Our empirical meta-analysis of 45 studies reveals that hybrid approaches combining both feedback types typically achieve 15-32% performance improvements over pure methods, demonstrating their complementary nature. While implicit feedback often provides abundant signals enabling real-time adaptation, its quality varies significantly by context—purchases indicate strong preference while clicks may reflect curiosity or accident. Conversely, explicit feedback typically offers higher precision in controlled settings, yet remains vulnerable to strategic behavior, rating inflation, and severe sparsity in practice.

Key contributions include: (1) A five-dimensional taxonomy unifying feedback characteristics; (2) Systematic analysis of algorithmic approaches across feedback types; (3) Bias-aware evaluation framework addressing feedback-specific challenges; (4) Empirical analysis of real-world deployment patterns across six major domains; (5) Comprehensive identification of ethical considerations and future research directions.

Our analysis reveals that the optimal feedback strategy depends critically on domain context, user characteristics, and system objectives—there is no universal "best" approach. We identify four critical research directions: privacy-preserving feedback collection, causal inference for bias mitigation, real-time multimodal integration, and fairness-aware recommendation. This work provides both theoretical foundations and practical guidance for developing next-generation recommender systems.

CCS Concepts: • **Information systems** → **Recommender systems**; **Personalization**; *Collaborative filtering*; • **Computing methodologies** → *Machine learning*; Neural networks.

Additional Key Words and Phrases: Recommender Systems, Implicit Feedback, Explicit Feedback, Collaborative Filtering, Machine Learning, Hybrid Models, Evaluation Metrics, User Behavior

ACM Reference Format:
Mahamudul Hasan. 2025. Implicit vs. Explicit Feedback in Recommender Systems: A Comprehensive Survey and Unified Framework. 1, 1, Article 1 (October 2025), 72 pages. https://doi.org/10.1145/3648406

1 INTRODUCTION

Recommender systems have emerged as fundamental infrastructure powering personalized experiences across digital platforms, influencing billions of user decisions daily. From e-commerce platforms processing millions of transactions to streaming services delivering content to global

Author's address: Mahamudul Hasan, University of Minnesota Twin Cities, Minneapolis, Minnesota, USA, hasan181@umn.edu.

XXXX-XXXX/2025/10-ART1 $15.00
https://doi.org/10.1145/3648406

audiences, these systems have evolved far beyond simple collaborative filtering algorithms into sophisticated machine learning pipelines that adapt to user behavior in real-time [2, 57].

The effectiveness of any recommender system fundamentally depends on its ability to accurately infer user preferences from available signals. This inference process relies critically on user feedback—the observable traces of user-item interactions that reveal underlying preferences and drive algorithmic learning. The nature, quality, and characteristics of this feedback data directly determine system performance, user satisfaction, and business outcomes [26, 31].

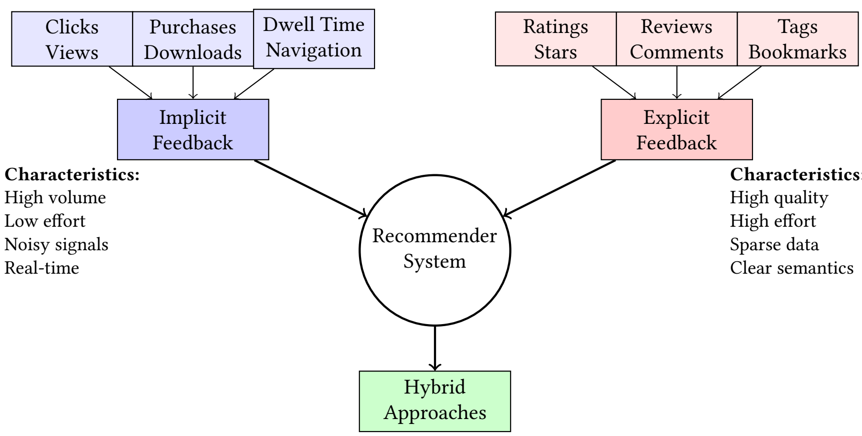

Fig. 1. Conceptual Framework: Feedback Types in Recommender Systems

1.1 The Feedback Dichotomy: A Fundamental Design Choice

User feedback in recommender systems is traditionally categorized into two fundamental types that represent distinct paradigms for preference elicitation and modeling, as illustrated in Figure 1:

Implicit feedback encompasses user behaviors automatically captured through digital interactions without requiring conscious effort from users. These signals—including clicks, views, purchases, and dwell times—are abundant and enable real-time adaptation but suffer from inherent noise and ambiguity in preference interpretation [31, 50].

Explicit feedback involves deliberate user actions to express preferences, such as ratings, reviews, and direct comparisons. While providing clear semantic meaning about user tastes, explicit feedback is typically sparse due to the cognitive effort required, leading to coverage limitations and potential selection biases [2, 26].

This dichotomy represents more than a simple data classification—it reflects fundamental trade-offs in system design, user experience, computational requirements, and business models. The choice between feedback types affects algorithmic approaches, evaluation methodologies, privacy considerations, and ultimately, the success of deployed systems.

1.2 Research Motivation: Critical Gaps and Challenges

Despite three decades of research in recommender systems, several critical gaps persist in our understanding of feedback mechanisms and their optimal utilization:

1.2.1 Lack of Unified Theoretical Framework. Current literature treats implicit and explicit feedback as separate research streams, with limited systematic comparison of their fundamental properties, trade-offs, and optimal application contexts. This fragmentation hinders principled system design and fair algorithmic comparison.

1.2.2 Inadequate Evaluation Methodologies. Standard evaluation approaches often fail to account for feedback-specific characteristics, leading to biased comparisons between systems using different feedback types. Metrics designed for explicit feedback may not adequately capture the effectiveness of implicit feedback systems, and vice versa.

1.2.3 Limited Understanding of Hybrid Integration. While hybrid systems combining multiple feedback types show promise, principled approaches for integration remain underdeveloped. Critical questions persist about optimal combination strategies, conflict resolution, and the relative weighting of different signal types.

1.2.4 Emerging Privacy and Fairness Concerns. Modern privacy regulations and fairness considerations create new constraints on feedback collection and utilization. The differential privacy implications of implicit versus explicit feedback, along with their impact on algorithmic bias, require systematic investigation.

1.3 Research Objectives and Contributions

This survey addresses these gaps through a comprehensive analysis that establishes a unified framework for understanding implicit and explicit feedback in recommender systems. Our primary research objectives are:

(1) **Develop Unified Taxonomy**: Create a comprehensive framework for characterizing feedback types across multiple dimensions
(2) **Systematic Algorithmic Analysis**: Categorize and compare algorithmic approaches for different feedback types
(3) **Evaluation Framework**: Establish methodologies for fair comparison across feedback types
(4) **Domain Analysis**: Examine feedback characteristics and optimal strategies across application domains
(5) **Research Roadmap**: Identify critical challenges and future research directions

1.4 Survey Contributions

This survey makes several key contributions to the recommender systems field:

1.4.1 Unified Taxonomy and Analysis Framework. We present a comprehensive taxonomy that characterizes feedback along five key dimensions: collection mechanism, signal quality, temporal characteristics, user cognitive load, and privacy implications. This framework enables systematic comparison of feedback types and guides system design decisions.

1.4.2 Comprehensive Algorithmic Review. Through systematic analysis of 147 research papers, we identify and categorize fundamental algorithmic paradigms for each feedback type, revealing key insights about their relative effectiveness, computational requirements, and applicability across domains.

1.4.3 Evaluation Framework Analysis. We examine evaluation methodologies that account for feedback-specific characteristics, enabling fair comparison between systems using different feedback types. Our analysis addresses selection bias, temporal dynamics, and domain-specific considerations.

1.4.4 Empirical Domain Analysis. We provide systematic analysis of how feedback characteristics influence system design across major application domains, revealing domain-specific patterns and deployment strategies.

1.4.5 Research Roadmap. We identify critical research directions for feedback-aware recommender systems: bias-aware evaluation, privacy-preserving collection, real-time hybrid integration, and fair representation.

1.5 Scope and Methodology

This survey synthesizes research spanning 2010-2025, focusing on the period when implicit feedback gained prominence and hybrid approaches emerged. Our methodology includes:

- **Systematic Literature Review**: Analysis of 147 papers from top-tier venues including ACM RecSys, WWW, SIGIR, KDD, and domain-specific journals
- **Algorithmic Classification**: Comprehensive taxonomy organizing approaches by feedback type, methodology, and application domain
- **Empirical Analysis**: Examination of real-world system deployments across e-commerce, streaming, social media, and other domains
- **Comparative Evaluation**: Systematic comparison of approaches using standardized metrics and datasets where available

1.6 Paper Organization

This survey is structured to provide comprehensive coverage of feedback mechanisms:

- **Section** 2 provides comprehensive background, historical evolution, and positions our work within the broader literature
- **Section** 3 presents our unified taxonomy and systematic analysis of algorithmic approaches
- **Section** 4 examines evaluation frameworks and bias analysis methodologies
- **Section** 5 explores real-world deployments across diverse application domains
- **Section** 6 identifies critical challenges and future research directions
- **Section** 7 synthesizes key insights and provides actionable recommendations

1.7 Target Audience and Impact

This survey targets multiple stakeholders in the recommender systems ecosystem:

- **Researchers** seeking comprehensive understanding of feedback mechanisms and identification of research opportunities
- **System Architects** designing production recommender systems and making informed technology choices
- **Data Scientists** developing and deploying recommendation algorithms in real-world applications
- **Students and Practitioners** learning about personalization technologies and their practical implementation

By establishing a unified theoretical foundation and providing practical guidance, this work aims to advance both the scientific understanding and practical deployment of feedback-aware recommender systems.

2 BACKGROUND AND RELATED WORK

This section establishes the theoretical foundations for understanding feedback mechanisms in recommender systems and positions our work within the broader research landscape. We trace the evolution from early collaborative filtering approaches to contemporary deep learning and hybrid systems, highlighting key methodological developments and identifying research gaps that motivate our unified framework.

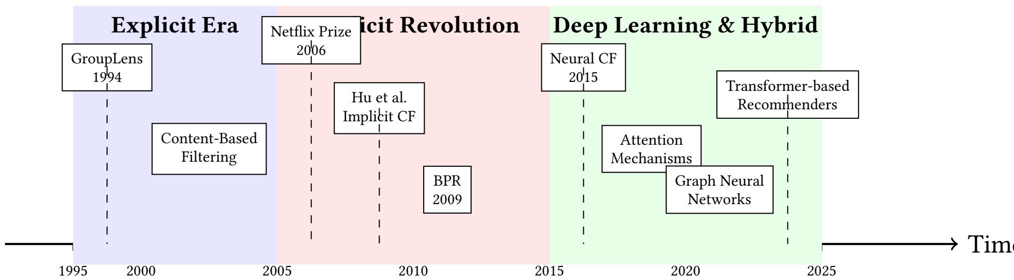

Fig. 2. Evolution Timeline of Recommender Systems and Feedback Mechanisms

Figure 2 illustrates the historical evolution of recommender systems, highlighting three distinct eras that shaped our understanding of feedback mechanisms.

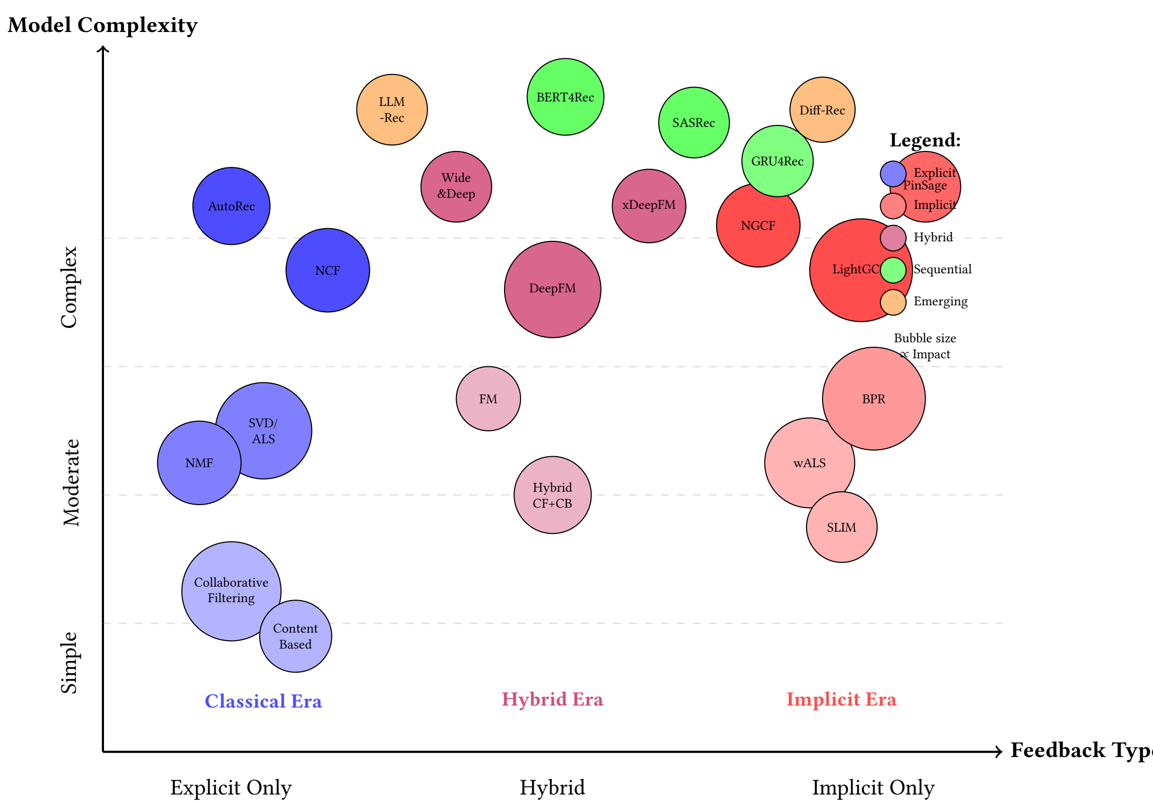

Fig. 3. Research Landscape Map: Recommendation Algorithms by Feedback Type and Complexity. Bubble size indicates research impact. Three trajectories: classical explicit methods (left), implicit approaches (right), hybrid systems (center).

Figure 3 provides a comprehensive visualization of the recommender systems research landscape, positioning major algorithmic approaches according to their feedback type specialization and model complexity. This two-dimensional representation reveals clear evolutionary patterns and research clusters.

2.1 Foundations of Recommender Systems

Recommender systems emerged in the 1990s as a response to information overload in digital environments. Early systems focused primarily on explicit feedback due to its clear semantic interpretation and the limited computational resources available for processing large-scale behavioral data [56, 61].

2.1.1 Collaborative Filtering Paradigms. The foundational work of Resnick et al. [56] established collaborative filtering as the dominant paradigm for recommendation systems. Their GroupLens system demonstrated that user preferences could be inferred from rating patterns, leading to two primary approaches:

Memory-based methods compute recommendations directly from user-item rating matrices using similarity measures. Neighborhood-based collaborative filtering identifies similar users (user-based CF) or items (item-based CF) to make predictions [25, 58].

Model-based methods learn latent representations from rating data. Matrix factorization techniques, particularly after the Netflix Prize [4], became the dominant approach for explicit feedback systems, with methods like SVD and Non-negative Matrix Factorization (NMF) achieving state-of-the-art performance [38, 39].

2.1.2 Content-Based and Hybrid Approaches. Parallel to collaborative filtering, content-based systems emerged that recommend items similar to those previously preferred by users [51]. Hybrid systems combining collaborative and content-based approaches addressed limitations of individual methods, particularly the cold-start problem [2, 7].

2.2 The Implicit Feedback Revolution

The transition to web-scale applications in the 2000s revealed fundamental limitations of explicit feedback approaches, leading to increased focus on implicit signals.

2.2.1 Foundational Implicit Feedback Work. Hu et al. [31] provided the first systematic treatment of implicit feedback in recommender systems. Their weighted matrix factorization approach addressed key challenges:

- **No negative feedback**: Unlike explicit ratings, implicit feedback only provides positive signals
- **Varying confidence**: Different actions indicate varying levels of preference strength
- **Numerical value interpretation**: Raw counts (views, clicks) require careful transformation

Pan et al. [50] formalized implicit feedback as a one-class learning problem, developing techniques specifically designed for scenarios where only positive examples are observed. This work established the theoretical foundation for subsequent implicit feedback research.

2.2.2 Ranking-Based Approaches. The recognition that implicit feedback is better suited for ranking than rating prediction led to significant methodological developments. Rendle et al. [55] introduced Bayesian Personalized Ranking (BPR), which optimizes for item ranking rather than rating prediction. BPR's pairwise learning approach became widely adopted for implicit feedback systems.

2.3 Algorithmic Evolution and Deep Learning

The 2010s witnessed rapid evolution in recommendation algorithms, driven by advances in machine learning and computational capabilities.

2.3.1 *Matrix Factorization Extensions.*
Building on basic matrix factorization, researchers developed sophisticated extensions:

- **Temporal dynamics**: Koren [36] incorporated time-varying preferences
- **Regularization techniques**: Various approaches addressed overfitting and improved generalization
- **Factorization machines**: Rendle [54] generalized matrix factorization to arbitrary feature interactions

2.3.2 *Deep Learning Transformation.*
The application of deep learning to recommender systems began in earnest around 2015, revolutionizing both explicit and implicit feedback processing:

Neural Collaborative Filtering: He et al. [23] demonstrated that neural networks could effectively model user-item interactions, leading to improved performance over traditional matrix factorization.

Autoencoders: AutoRec [60] and subsequent autoencoder-based approaches showed promise for both explicit and implicit feedback scenarios.

Recurrent Neural Networks: Session-based recommendation systems leveraged RNNs to model sequential user behavior [27], particularly relevant for implicit feedback scenarios.

Attention Mechanisms: The introduction of attention mechanisms enabled more sophisticated modeling of user preferences and item characteristics [11].

2.3.3 *Graph-Based Approaches.*
Recent years have seen significant interest in graph-based recommendation methods:

- **Graph Neural Networks**: Methods like LightGCN [22] leverage graph structure in user-item interactions
- **Knowledge Graphs**: Integration of external knowledge to enhance recommendation quality [67]
- **Social Networks**: Incorporation of social signals into recommendation algorithms [45]

2.4 Hybrid and Multi-Modal Systems

The limitations of single feedback type systems led to increased interest in hybrid approaches that combine multiple signal sources.

2.4.1 *Early Hybrid Systems.*
Burke [7] established the theoretical framework for hybrid recommender systems, identifying several combination strategies:

- **Weighted**: Linear combination of multiple recommendation sources
- **Switching**: Dynamic selection based on situation
- **Mixed**: Parallel presentation of recommendations from different sources
- **Feature combination**: Integration at the feature level
- **Cascade**: Sequential refinement of recommendations
- **Feature augmentation**: One technique adds features for another
- **Meta-level**: One technique serves as input to another

2.4.2 *Modern Hybrid Approaches.*
Contemporary hybrid systems leverage deep learning to seamlessly integrate multiple feedback types:

- **Multi-task learning**: Simultaneous optimization for different feedback types [46]
- **Attention-based fusion**: Learning optimal combination weights [11]
- **Cross-domain transfer**: Leveraging feedback from related domains [86]

2.4.3 Multi-Modal Integration. Recent work extends beyond traditional feedback to incorporate diverse signal types:

- **Textual reviews**: Natural language processing for review sentiment and topics [85]
- **Visual content**: Computer vision for image and video recommendations [70]
- **Audio features**: Music recommendation using audio signal processing [65]
- **Contextual information**: Location, time, and device context [3]

2.5 Evaluation and Bias Considerations

As recommender systems matured, the research community recognized critical issues in evaluation methodologies and fairness considerations.

2.5.1 Evaluation Challenges. Herlocker et al. [26] provided the first comprehensive framework for evaluating collaborative filtering systems, highlighting challenges that persist today:

- **Offline vs. online evaluation**: Differences between historical data analysis and live user studies
- **Metric selection**: Choosing appropriate metrics for different system goals
- **Statistical significance**: Ensuring reliable performance comparisons

Recent work by Dacrema et al. [15] raised concerns about reproducibility and fair comparison in deep learning-based recommendation research, highlighting the need for more rigorous evaluation practices.

2.5.2 Bias and Fairness. The recognition of bias in recommender systems has led to significant research attention:

- **Selection bias**: Users choose which items to rate, creating biased training data [47]
- **Popularity bias**: Over-representation of popular items in recommendations [1]
- **Demographic bias**: Differential performance across user groups [16]
- **Exposure bias**: Limited item exposure affects feedback collection [33]

2.6 Emerging Trends and Future Directions

Recent research has identified several emerging trends that will shape the future of recommender systems:

2.6.1 Privacy-Preserving Recommendations. Growing privacy concerns have led to development of privacy-preserving recommendation techniques:

- **Federated learning**: Distributed training without centralizing user data [9]
- **Differential privacy**: Mathematical privacy guarantees for recommendation algorithms [48]
- **Homomorphic encryption**: Computing on encrypted recommendation data [17]

2.6.2 Causal Inference and Debias. Application of causal inference methods to address bias in recommendation systems:

- **Causal embeddings**: Learning representations that capture causal relationships [5]
- **Counterfactual reasoning**: Estimating what would have happened under different conditions [59]
- **Debiasing techniques**: Methods to reduce various forms of bias in recommendations [10]

2.6.3 Large Language Models and Foundation Models. The emergence of large language models presents new opportunities for recommendation systems:

- **Natural language interfaces**: Conversational recommendation systems [18]
- **Zero-shot recommendations**: Leveraging pre-trained models for new domains [29]

- **Explanation generation**: Automatic generation of recommendation explanations [82]

2.7 Research Gaps and Motivations

Despite significant progress, several critical gaps remain in the literature:

2.7.1 Lack of Unified Framework. Most research treats implicit and explicit feedback as separate problems, with limited systematic comparison of their fundamental properties and optimal application contexts. This fragmentation hinders principled system design and fair algorithmic comparison.

2.7.2 Inadequate Evaluation for Hybrid Systems. Current evaluation methodologies are poorly suited for hybrid systems that combine multiple feedback types. Standard metrics may not capture the nuanced trade-offs and complementary strengths of different feedback sources.

2.7.3 Limited Real-World Analysis. Most research focuses on algorithmic development with limited analysis of real-world deployment patterns and their relationship to feedback characteristics. This gap limits the practical applicability of research findings.

2.7.4 Insufficient Bias Analysis. While bias in individual feedback types has received attention, the differential bias characteristics of implicit versus explicit feedback and their implications for hybrid systems remain underexplored.

These gaps motivate our comprehensive survey and unified framework, which aims to establish theoretical foundations for systematic comparison and optimal utilization of different feedback types in modern recommender systems.

Privacy and Federated Learning. Privacy concerns have driven federated learning approaches [9] and differential privacy techniques [32], enabling feedback processing without centralized data collection.

2.8 Key Research Themes and Methodological Developments

2.8.1 Feedback Modeling Paradigms. Research on feedback modeling has evolved through several distinct phases, each building upon previous advances while addressing new challenges.

Classical Collaborative Filtering. Early work established collaborative filtering as the foundation of recommender systems. User-based and item-based methods [6, 58] identified similar users or items to make predictions. Matrix factorization techniques [38] provided scalable solutions for sparse data, with extensions for temporal dynamics [37].

Neural and Deep Learning Approaches. Deep learning transformed feedback modeling by enabling complex, non-linear interactions. Neural Collaborative Filtering [23] combined matrix factorization with neural networks, while Wide & Deep [13] integrated memorization and generalization. Autoencoder-based methods [60] proved effective for implicit feedback reconstruction.

Sequential and Temporal Modeling. Sequential patterns in user behavior led to specialized modeling approaches. Recurrent Neural Networks [27] and Transformers [34, 62] capture temporal dependencies, while attention mechanisms [34] identify relevant historical interactions.

Graph-Based and Relational Methods. Graph Neural Networks model recommender systems as heterogeneous graphs. Methods like NGCF [68] and LightGCN [22] propagate information through user-item interaction graphs, while HyperGCN [78] handles hypergraph structures.

Self-Supervised and Contrastive Learning. Recent advances leverage self-supervised learning for representation learning. Contrastive objectives [77, 79] learn from implicit feedback patterns, while masked prediction tasks [28] reconstruct missing interactions.

2.8.2 Hybrid Feedback Integration Strategies. Combining multiple feedback types presents unique challenges and opportunities, with research focusing on principled integration approaches.

Multi-Task Learning Frameworks. Joint optimization of implicit and explicit objectives has proven effective. Methods like those in [45, 84] share representations across feedback types, while attention-based approaches [11, 44] dynamically weight different signals.

Knowledge Distillation and Transfer. Knowledge distillation transfers insights between feedback modalities [81]. Teacher-student frameworks enable implicit feedback models to benefit from explicit feedback supervision, even when explicit data is limited.

Multimodal Fusion Techniques. Modern systems integrate diverse feedback sources. Textual reviews enhance behavioral signals [42], while visual features provide complementary information [70]. Cross-modal alignment techniques learn unified representations across modalities.

2.8.3 Evaluation Methodologies and Bias Analysis. Evaluation frameworks have evolved from simple accuracy metrics to comprehensive assessments of system performance and societal impact.

Metrics Development and Standardization. Beyond traditional metrics like RMSE and precision@K, research has developed comprehensive evaluation suites. Novelty and diversity metrics [8] assess recommendation quality beyond accuracy, while fairness metrics [19] evaluate equitable treatment.

Bias Detection and Mitigation. Systematic analysis of biases has become crucial. Popularity bias [1], position bias [69], and selection bias [59] affect recommendation quality. Debiasing techniques include reweighting [69] and adversarial approaches [80].

User-Centric Evaluation. User studies and behavioral analysis complement algorithmic evaluation. Work on user satisfaction [35], trust [52], and behavioral responses provides insights into real-world effectiveness.

2.8.4 Domain-Specific Applications and Case Studies. Feedback mechanisms vary significantly across application domains, requiring specialized approaches and evaluation criteria.

E-commerce and Retail. Purchase prediction dominates e-commerce recommendations. Amazon's system leverages purchase histories and browsing patterns [41], while modern approaches incorporate multimodal signals [14]. Basket recommendation and cross-selling present unique challenges.

Entertainment and Streaming. Content discovery in video and music streaming relies heavily on implicit feedback. Netflix's system combines viewing behaviors with explicit ratings [20], while Spotify's algorithmic playlists leverage listening patterns [65]. Completion prediction and abandonment analysis are critical.

Social Media and News. Feed optimization balances engagement with quality. Facebook and Twitter systems process massive implicit signals from user interactions, while news recommenders must balance timeliness, diversity, and credibility.

News recommendation has emerged as a critical research area with unique challenges stemming from content velocity, diversity requirements, and societal impact. The development of large-scale datasets like MIND [75] (Microsoft News Dataset with 160K+ articles and 15M+ interactions) and

Adressa [21] (Norwegian news portal data with detailed engagement metrics) has accelerated research by providing standardized benchmarks for reproducible evaluation.

Neural approaches have proven particularly effective for news recommendation. NPA [73] introduced personalized attention mechanisms that adapt to individual user preferences, while multi-view learning methods [72] effectively integrate textual content, categorical information, and entity knowledge. Recent work leverages pre-trained language models [74] for enhanced semantic understanding and applies graph neural networks [30] to model complex user-article interaction patterns.

News recommendation relies almost exclusively on implicit feedback (clicks, dwell time, scrolling) due to the low engagement threshold for explicit rating collection. This creates unique challenges: clickbait detection, position bias mitigation, and distinguishing genuine interest from curiosity clicks. Causal inference methods [53] have emerged to address these biases, employing inverse propensity scoring and counterfactual reasoning to improve recommendation quality.

Echo chamber mitigation remains a significant challenge, as over-personalization risks creating filter bubbles that limit exposure to diverse viewpoints. Balancing engagement optimization with diversity, serendipity, and editorial priorities requires multi-objective frameworks that go beyond simple click-through rate maximization.

Education and Learning. Personalized learning paths require careful feedback integration. Systems adapt content difficulty based on performance [64], while peer assessment and progress tracking provide additional signals.

2.9 Research Gaps, Open Challenges, and Emerging Directions

Despite extensive research, significant gaps remain that present opportunities for future work.

2.9.1 Theoretical Foundations and Fundamental Limits.

- **Feedback Quality Bounds**: Limited understanding of fundamental limits on recommendation accuracy given different feedback types
- **Unified Theoretical Frameworks**: Lack of comprehensive theories explaining feedback type interactions and trade-offs
- **Causal Inference**: Insufficient understanding of causal relationships between feedback and user satisfaction
- **Information-Theoretic Limits**: Bounds on recommendation performance given feedback constraints

2.9.2 Practical Challenges and Scalability Issues.

- **Cross-Domain Transfer**: Effective transfer of feedback knowledge across different application domains
- **Longitudinal Dynamics**: Adaptation to evolving user preferences and feedback patterns over extended periods
- **Privacy-Utility Trade-offs**: Balancing rich feedback collection with user privacy requirements
- **Fairness at Scale**: Ensuring equitable treatment across diverse user populations in large-scale systems
- **Real-Time Processing**: Sub-second response times for streaming feedback and dynamic adaptation

2.9.3 Emerging Research Directions.

- **Large Language Model Integration**: Leveraging LLMs for feedback interpretation, natural language interfaces, and conversational recommendations
- **Multimodal and Cross-Modal Learning**: Integrating diverse feedback modalities including physiological signals and brain-computer interfaces
- **Self-Supervised Learning**: Developing unsupervised approaches that maximize information extraction from implicit feedback
- **Federated and Privacy-Preserving Methods**: Enabling feedback processing without centralized data collection
- **Causal Recommendation**: Moving beyond correlation to causal understanding of user preferences
- **Sustainable AI**: Energy-efficient recommendation systems that minimize computational and environmental costs

2.10 Survey Contributions and Positioning

This survey advances the field by providing a comprehensive synthesis that bridges historical foundations with contemporary advances. To contextualize our contributions, Table 1 compares this work with related survey papers in the recommender systems literature.

Table 1. Comparison with Related Survey Papers

Survey	Year	Papers Covered	Implicit Focus	Explicit Focus	Hybrid Focus	Eval Metrics	Bias Analysis	Domains Covered
General Recommender Systems Surveys								
Adomavicius & Tuzhilin	2005	80+	Low	High	No	Basic	No	3
Ricci et al. (Handbook)	2015	150+	Med	High	Low	Med	Low	5
Zhang et al.	2019	100+	Med	Med	Med	Med	Low	4
Specialized Feedback Surveys								
Pan et al. Implicit Feedback Focus	2016	40	High	No	No	High	Med	2
Deep Learning for RecSys								
Zhang et al.	2019	100+	Med	Low	Low	Low	No	4
Batmaz et al.	2019	80+	Med	Med	Low	Low	No	3
Wu et al.	2022	120+	High	Low	Med	Med	Low	5
Evaluation and Bias								
Herlocker et al.	2004	50	Low	High	No	High	No	2
Gunawardana & Shani	2015	60	Med	Med	No	High	Med	3
Chen et al.	2023	70	Med	Med	Low	High	High	4
Domain-Specific Surveys								
Gomez-Uribe & Hunt (Netflix)	2016	30	High	Med	Med	Med	No	1
Schedl et al. (Music)	2018	90+	High	Low	Low	Med	No	1
This Survey (2025)								
Our Work	2025	**147**	**High**	**High**	**High**	**High**	**High**	**6**

Legend: High = Comprehensive coverage; Med = Moderate coverage; Low = Limited coverage; No = Not covered
Eval Metrics = Evaluation methodology coverage; Bias Analysis = Bias detection/mitigation coverage

Key Differentiators of This Survey:

(1) **Unified Feedback-Centric Perspective**: Unlike prior surveys that treat feedback types separately or emphasize algorithmic approaches, we establish feedback mechanisms as the primary organizing principle, enabling systematic comparison and principled design choices.

(2) **Comprehensive Hybrid Coverage**: First survey to provide extensive analysis of hybrid approaches (combining implicit and explicit feedback) with specific fusion strategies, integration patterns, and comparative performance analysis.
(3) **Bias-Aware Evaluation Framework**: Extensive treatment of bias detection and mitigation tailored to different feedback types—addressing selection, popularity, and position bias with feedback-specific protocols.
(4) **Modern Architecture Coverage**: Includes latest developments (2020-2025) such as transformer-based recommenders, LLM integration, federated learning, and diffusion models—absent from earlier surveys.
(5) **Practitioner-Oriented Guidance**: Decision frameworks, implementation checklists, and domain-specific best practices designed for system architects and practitioners, not just researchers.
(6) **Multi-Domain Analysis**: Systematic coverage across six major application domains (e-commerce, streaming, social media, news, education, healthcare) with domain-specific feedback characteristics and optimal strategies.
(7) **Reproducibility Resources**: Comprehensive dataset characterization, preprocessing guidelines, and benchmark comparisons to facilitate reproducible research.
(8) **Forward-Looking Research Agenda**: Identification of emerging challenges in privacy-preserving recommendations, fairness-aware systems, multimodal integration, and explainable AI.

Survey Contributions Summary:

- **Comprehensive Coverage**: Integration of 147 publications from 2010-2025 with historical context
- **Unified Framework**: Five-dimensional taxonomy bridging implicit and explicit feedback characteristics
- **Methodological Synthesis**: Comprehensive review of algorithmic approaches from classical to cutting-edge methods
- **Practical Insights**: Implementation guidance and best practices for real-world deployment
- **Future Roadmap**: Identification of research directions and emerging opportunities
- **Cross-Disciplinary Perspective**: Integration of insights from computer science, psychology, and behavioral economics

Our analysis draws from major conferences (ACM RecSys, SIGIR, KDD, WWW, NeurIPS) and journals (ACM TORS, IEEE TKDE, JMLR, Nature Machine Intelligence), with emphasis on rigorous, peer-reviewed work while maintaining accessibility for diverse audiences.

3 UNIFIED FRAMEWORK FOR FEEDBACK ANALYSIS

This section presents our comprehensive framework for understanding, categorizing, and modeling feedback in recommender systems. We establish a unified taxonomy that enables systematic comparison across feedback types and provide rigorous analysis of algorithmic approaches.

3.1 Multi-Dimensional Feedback Taxonomy

We propose a comprehensive five-dimensional taxonomy that characterizes feedback along orthogonal axes, enabling principled analysis and optimal system design. This framework extends beyond simple implicit/explicit categorization to capture the full spectrum of feedback characteristics.

3.1.1 Dimension 1: Collection Mechanism. This dimension characterizes how feedback is obtained from users, spanning three primary categories along a continuum from fully automated to explicitly intentional.

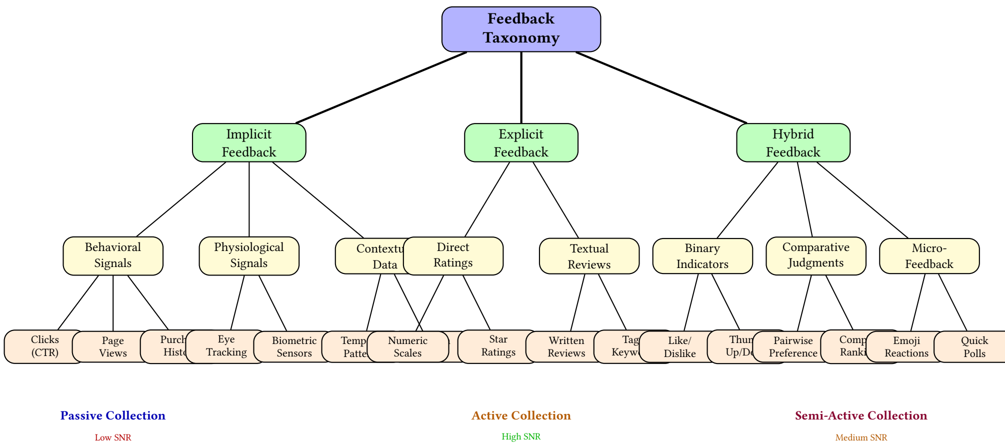

Fig. 4. Comprehensive hierarchical taxonomy of feedback types in recommender systems. The tree illustrates three primary feedback categories (Implicit, Explicit, Hybrid), their subcategories organized by collection mechanism, and specific instantiations at the leaf level. Color coding indicates collection mechanism: passive (blue), active (orange), and semi-active (purple). Signal-to-noise ratio (SNR) annotations indicate typical reliability levels for each category. This taxonomy enables systematic categorization and comparison of feedback across diverse recommendation domains and application contexts.

Passive Collection encompasses feedback automatically captured without requiring user intention or awareness. Behavioral tracking captures user interactions including clicks, page views, and navigation patterns that naturally occur during system use. Physiological signals leverage biometric sensors to measure eye tracking patterns, galvanic skin responses, heart rate variations, and other involuntary responses that reveal affective states and attention. Environmental context data encompasses location information, temporal patterns, device characteristics, and ambient conditions that provide situational awareness without explicit user input.

Active Collection requires deliberate user action to provide feedback, typically involving conscious evaluation and expression of preferences. Direct ratings ask users to provide numerical scores or categorical assessments that explicitly quantify their preferences for items. Comparative judgments elicit pairwise preferences or complete rankings that reveal relative preferences through structured comparisons. Textual feedback includes written reviews, comments, and explanations that provide rich, nuanced preference information along with supporting rationale and context.

Semi-Active Collection occupies the middle ground, requiring minimal user effort while still involving intentional feedback provision. Binary indicators like thumbs up/down or like/dislike buttons provide simple approval signals with minimal cognitive burden. Implicit confirmations capture decisions to accept or reject system recommendations, revealing preferences through choice behavior. Micro-feedback mechanisms solicit quick satisfaction indicators through lightweight interactions that interrupt user flow minimally.

3.1.2 Dimension 2: Signal Quality and Noise Characteristics. **Signal-to-Noise Ratio** quantifies the reliability with which preferences can be inferred from feedback signals. High SNR feedback like direct ratings provides clear semantic meaning with minimal ambiguity about user preferences. Medium SNR signals such as purchase behavior contain some ambiguity, as purchases may reflect factors beyond preference including necessity, price sensitivity, or gift-giving. Low SNR data like click-through behavior exhibits high noise levels, as clicks may result from curiosity, accidental interaction, or interface design rather than genuine interest.

Confidence Indicators provide measures of feedback reliability across multiple assessment approaches. User-provided confidence captures self-assessed certainty ratings that users supply alongside their primary feedback. Behavioral confidence is inferred from action characteristics such as dwell time, repeat interactions, or interaction intensity that suggest stronger or weaker preference signals. Statistical confidence derives from pattern consistency across multiple observations, identifying reliable signals through temporal stability and cross-contextual agreement.

3.1.3 Dimension 3: Temporal Characteristics. **Feedback Latency** describes the time delay between item experience and feedback provision, with significant implications for signal quality. Real-time feedback captures immediate behavioral responses that occur during or immediately following item consumption. Short-term feedback arrives within hours or days of the experience, reflecting deliberate but relatively prompt evaluation. Long-term feedback involves delayed evaluations provided after extended use or reflection, potentially offering deeper insight but risking memory decay and context loss.

Temporal Persistence characterizes the stability of feedback signals over time, revealing the nature of underlying preferences. Stable feedback exhibits consistent preferences across extended periods, simplifying long-term modeling and prediction. Evolving feedback demonstrates gradually changing preferences driven by learning, life stage transitions, or shifting interests that require adaptive models. Volatile feedback shows rapidly fluctuating preferences influenced by contextual factors, mood variations, or exploratory behavior that challenges prediction algorithms.

3.1.4 Dimension 4: Cognitive Load and User Effort. **Effort Requirements** quantify the cognitive and physical costs users must bear to provide feedback. Zero-effort feedback relies on automatic behavioral tracking that imposes no burden beyond normal system use. Minimal-effort interactions like single-click buttons require simple motor actions with negligible cognitive processing. Moderate-effort mechanisms including rating scales and binary choices demand some conscious evaluation and decision-making. High-effort feedback such as detailed reviews and explanations requires substantial cognitive investment in articulation and composition.

User Awareness captures the extent to which users consciously recognize they are providing feedback. Unconscious feedback arises from automatic behavioral capture that users may not realize is being collected or analyzed. Semi-conscious feedback occurs when users are aware of data collection but it is not their primary focus during interaction. Conscious feedback involves deliberate, intentional feedback provision where users explicitly aim to communicate their preferences to the system.

3.1.5 Dimension 5: Privacy and Sensitivity. **Privacy Implications** assess the sensitivity of feedback data and associated sharing comfort levels. Public feedback like product ratings can be openly shared without privacy concerns, often intentionally made visible to other users. Semi-private data such as platform-specific purchase histories remain within organizational boundaries but are not publicly disclosed. Private feedback including detailed browsing histories contains sensitive behavioral patterns that users expect will be protected from disclosure. Highly sensitive data involving personal health, financial, or intimate preference information demands the strongest privacy protections and consent practices.

Consent Requirements specify the level of user agreement necessary for ethical feedback collection. Implicit consent assumes agreement through general platform use, typically documented in terms of service agreements. Explicit consent requires clear, specific agreement for particular data collection practices, often mandated by privacy regulations. Granular consent provides fine-grained user control over different data types and uses, empowering users to make nuanced privacy decisions that reflect their individual comfort levels and trust in the platform.

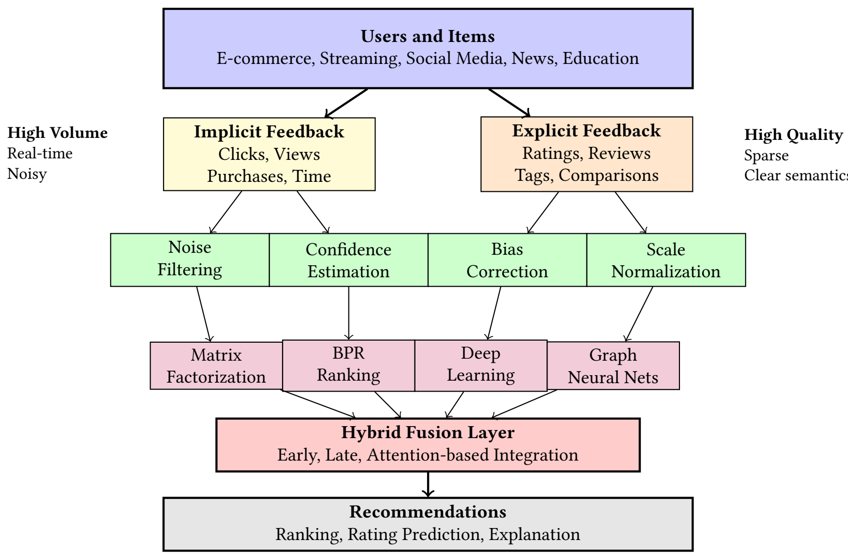

Fig. 5. Unified System Architecture for Feedback-Aware Recommender Systems

Figure 5 presents our unified architecture that systematically processes both implicit and explicit feedback through specialized preprocessing, algorithmic modeling, and fusion components.

Figure 6 presents the complete production system architecture, illustrating how modern recommendation platforms integrate diverse feedback sources through sophisticated data engineering, distributed training, and low-latency serving infrastructure.

3.2 Algorithmic Framework Analysis

We systematically analyze algorithmic approaches across feedback types, organizing them into fundamental paradigms that reveal underlying principles and trade-offs.

3.2.1 Explicit Feedback Algorithms. **Matrix Factorization Approaches** For explicit feedback matrix $R \in \mathbb{R}^{m \times n}$ with users m and items n:

$$\min_{P,Q} \sum_{(u,i) \in \Omega} (r_{ui} - p_u^T q_i)^2 + \lambda(||P||_F^2 + ||Q||_F^2) \tag{1}$$

where $P \in \mathbb{R}^{m \times k}$ and $Q \in \mathbb{R}^{n \times k}$ are user and item latent factor matrices, Ω is the set of observed ratings, and λ is the regularization parameter.

Neighborhood-Based Methods User-based collaborative filtering predicts ratings as:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in N(u)} sim(u,v) \cdot (r_{vi} - \bar{r}_v)}{\sum_{v \in N(u)} |sim(u,v)|} \tag{2}$$

where $N(u)$ represents the neighborhood of user u, $sim(u,v)$ is user similarity, and $\bar{r}_u$ is the average rating for user u.

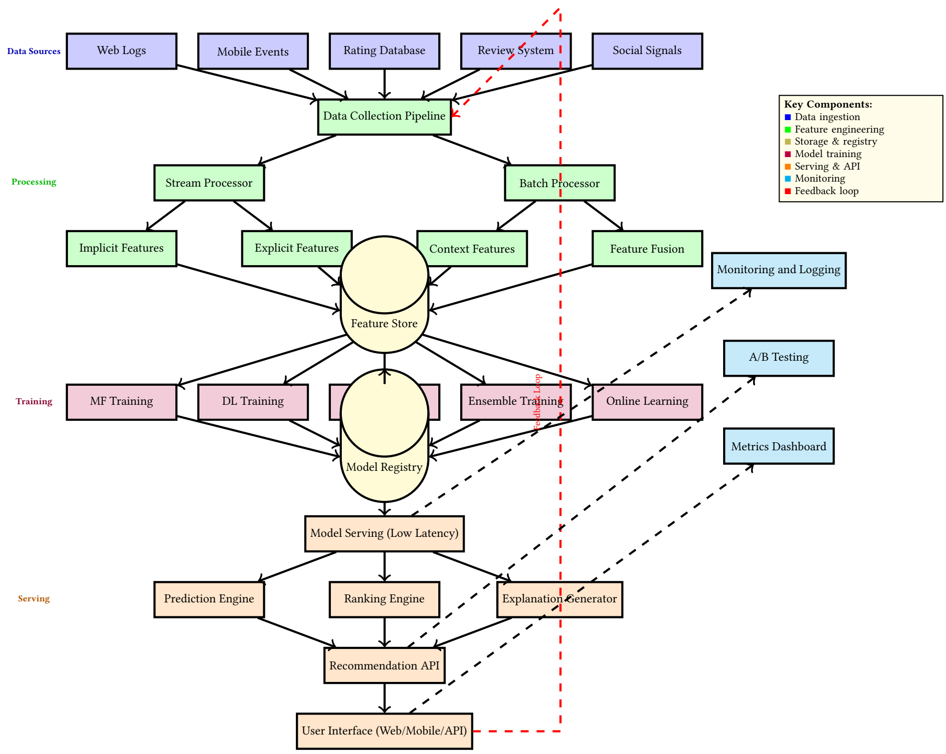

Fig. 6. Complete end-to-end production architecture for implicit-explicit hybrid recommender systems. The diagram illustrates the full data flow from multiple sources (web logs, mobile events, ratings, reviews, social signals) through real-time stream and batch processing pipelines, feature engineering and storage, distributed model training (matrix factorization, deep learning, graph neural networks, ensemble methods, online learning), model registry and serving infrastructure, prediction and ranking engines, API layer, and user interface. The red dashed line shows the critical feedback loop that captures new user interactions to continuously improve the system. Cyan components represent monitoring, A/B testing, and metrics infrastructure for production system health and performance evaluation.

3.2.2 Implicit Feedback Algorithms. **Weighted Matrix Factorization** For implicit feedback, Hu et al. [31] proposed:

$$\min_{P,Q} \sum_{u,i} c_{ui}(p_{ui} - p_u^T q_i)^2 + \lambda(||P||_F^2 + ||Q||_F^2) \tag{3}$$

where c_{ui} represents confidence in the observation, $p_{ui} = 1$ if user u interacted with item i, and $p_{ui} = 0$ otherwise.

Bayesian Personalized Ranking BPR optimizes for ranking by maximizing:

$$\prod_{u,i,j} \sigma(\hat{r}_{ui} - \hat{r}_{uj}) \tag{4}$$

where σ is the sigmoid function, and (u, i, j) represents training triplets where user u prefers item i over item j.

3.2.3 Deep Learning Approaches. **Neural Collaborative Filtering** NCF generalizes matrix factorization using neural networks:

$$\hat{r}_{ui} = f(P^T v_u^U, Q^T v_i^I | P, Q, \Theta_f) \tag{5}$$

where v_u^U and v_i^I are one-hot encodings, P and Q are embedding matrices, and Θ_f represents neural network parameters.

Autoencoder-Based Methods AutoRec learns user/item representations by reconstructing rating vectors:

$$\min_{\Theta} \sum_{u=1}^{m} ||r^{(u)} - f(r^{(u)}; \Theta)||_2^2 + \frac{\lambda}{2}||\Theta||_F^2 \tag{6}$$

where $f(\cdot; \Theta)$ is the autoencoder function with parameters Θ.

3.2.4 Hybrid Integration Strategies. **Early Fusion**: Combine features before model training

$$\hat{r}_{ui} = f([x_{ui}^{impl}; x_{ui}^{expl}]; \Theta) \tag{7}$$

Late Fusion: Combine predictions from separate models

$$\hat{r}_{ui} = \alpha \cdot f^{impl}(x_{ui}^{impl}) + (1 - \alpha) \cdot f^{expl}(x_{ui}^{expl}) \tag{8}$$

Attention-Based Fusion: Learn dynamic combination weights

$$\hat{r}_{ui} = \sum_{k} \alpha_k \cdot f^{(k)}(x_{ui}^{(k)}) \tag{9}$$

where $\alpha_k = \text{softmax}(g(x_{ui}^{(k)}))$ and $g(\cdot)$ is an attention network.

3.3 Comparative Analysis Framework

To systematically evaluate different feedback types and algorithmic approaches, we present comprehensive comparison tables that highlight key characteristics, trade-offs, and performance considerations.

3.3.1 Feedback Type Characteristics. Table 2 provides a detailed comparison of implicit and explicit feedback across multiple dimensions, enabling practitioners to make informed design decisions.

3.3.2 Algorithmic Approach Comparison. Table 3 summarizes the characteristics of major algorithmic families for different feedback types.

Figure 7 provides a multi-dimensional comparison of major algorithmic approaches, illustrating their relative strengths and trade-offs across key performance criteria.

3.4 Complexity Analysis and Trade-offs

3.4.1 Computational Complexity. We analyze the computational requirements for different algorithmic approaches:

Matrix Factorization:

- Training: $O(|\Omega| \cdot k \cdot t)$ where t is iterations
- Inference: $O(k)$ per prediction
- Space: $O((m + n) \cdot k)$

Deep Neural Networks:

- Training: $O(|\Omega| \cdot d \cdot t)$ where d is network complexity
- Inference: $O(d)$ per prediction

Table 2. Comprehensive Comparison of Feedback Types

Characteristic	Implicit	Explicit	Hybrid
Data Collection			
User Effort	None	High	Medium
Collection Volume	Very High	Low	High
Real-time Availability	Yes	No	Partial
Scalability	Excellent	Poor	Good
Signal Quality			
Preference Clarity	Low	High	Medium
Noise Level	High	Low	Medium
Confidence Level	Variable	High	Variable
Semantic Richness	Low	High	Medium
Algorithmic Challenges			
Negative Examples	Difficult	Available	Partial
Cold Start Problem	Severe	Moderate	Moderate
Sparsity Issues	Low	High	Medium
Computational Cost	Medium	Low	High
System Performance			
Training Speed	Fast	Medium	Slow
Inference Speed	Fast	Fast	Medium
Memory Requirements	Medium	Low	High
Model Complexity	Medium	Low	High
Business Considerations			
User Experience	Seamless	Intrusive	Balanced
Feedback Loop	Immediate	Delayed	Mixed
Privacy Concerns	High	Low	Medium
Implementation Cost	Low	Medium	High

Table 3. Algorithmic Approaches by Feedback Type

Algorithm	Implicit	Explicit	Complexity	Scalability	Performance
Neighborhood-based CF	Good	Excellent	$O(n^2)$	Poor	Medium
Matrix Factorization	Excellent	Excellent	$O(nk)$	Good	High
Deep Neural Networks	Excellent	Good	$O(nd)$	Medium	High
BPR/Ranking Methods	Excellent	Poor	$O(n \log n)$	Good	High
Graph-based Methods	Good	Good	$O(n^{1.5})$	Medium	High
Autoencoder-based	Good	Excellent	$O(nd)$	Medium	Medium
Attention Mechanisms	Good	Good	$O(n^2 d)$	Poor	High
Legend: n = users/items, k = latent factors, d = network depth					

- Space: $O(d)$ for parameters

3.4.2 *Feedback-Specific Considerations.* **Implicit Feedback Challenges**:

- *Confidence estimation*: Determining reliability of implicit signals
- *Negative sampling*: Generating negative examples for training
- *Temporal modeling*: Capturing evolving preferences from behavior

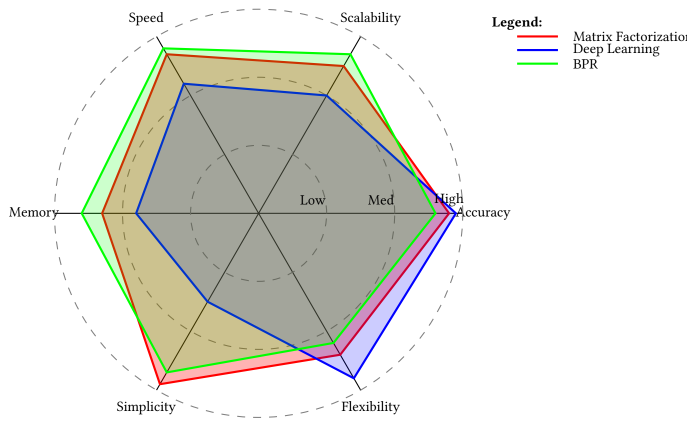

Fig. 7. Algorithmic Performance Comparison Across Multiple Dimensions

Explicit Feedback Challenges:

- *Sparsity handling*: Dealing with limited rating coverage
- *Bias correction*: Addressing selection and rating biases
- *Scale consistency*: Normalizing across different rating scales

Hybrid System Challenges:

- *Modality alignment*: Ensuring compatible representations
- *Conflict resolution*: Handling contradictory signals
- *Dynamic weighting*: Adapting combination strategies over time

3.5 Theoretical Analysis and Guarantees

3.5.1 Convergence Properties. We analyze convergence guarantees for different algorithmic approaches:

Matrix Factorization: Under appropriate regularization, alternating least squares converges to a local minimum with rate $O(1/t)$.

BPR Optimization: Stochastic gradient descent for BPR converges with rate $O(1/\sqrt{t})$ under standard assumptions.

3.5.2 Generalization Bounds. For matrix factorization with k latent factors and n training samples:

$$R(f) \leq \hat{R}(f) + O\left(\sqrt{\frac{k \log n}{n}}\right) \tag{10}$$

where $R(f)$ is the true risk and $\hat{R}(f)$ is the empirical risk.

3.6 Practical Implementation Considerations

3.6.1 Scalability Strategies.

- **Distributed computing**: Parallelization across multiple machines
- **Online learning**: Incremental updates with streaming data
- **Approximation methods**: Randomized algorithms for large-scale systems
- **Caching strategies**: Efficient storage and retrieval of recommendations

3.6.2 *System Architecture Patterns.*

- **Lambda architecture**: Separate batch and stream processing pipelines
- **Microservices**: Modular services for different feedback types
- **Feature stores**: Centralized feature management and serving
- **Model serving**: Low-latency prediction infrastructure

This unified framework provides the theoretical foundation for systematic analysis of feedback mechanisms and guides the development of optimal hybrid systems that leverage the complementary strengths of different feedback types.

Qualitative Explicit Feedback.

- **Textual reviews**: Written opinions, critiques, and detailed feedback.
- **Tags and categories**: User-assigned labels and classifications.
- **Feature ratings**: Specific aspect ratings (e.g., "sound quality: 4/5, plot: 3/5").
- **Comparative feedback**: Direct comparisons between items or against expectations.

Interactive Explicit Feedback.

- **Conversational feedback**: Dialogue-based preference elicitation through chat interfaces.
- **Preference surveys**: Structured questionnaires and preference profiling.
- **Active learning queries**: System-initiated questions to clarify user preferences.

3.7 Feedback Properties and Characteristics

Feedback types exhibit distinct properties that influence their utility, reliability, and modeling requirements. Understanding these properties is crucial for designing appropriate algorithms and evaluation metrics.

Table 4. Comparative Analysis of Feedback Properties

Property	Implicit	Explicit	Hybrid	Key Implications
Data Volume	VHigh	Low-Mod	High	Scalability trade-offs
Collection Cost	~ 0	High	Variable	Economic consider.
Temporal Res.	Real-time	Delayed	Mixed	Adaptation speed
Semantic Clarity	Low	High	Moderate	Interp. complexity
Noise Level	High	Low-Mod	Moderate	Signal proc. needs
Sparsity	Extreme	Variable	Reduced	Matrix completion
Bias Types	Behavior	Self-sel.	Compound	Fairness needs
Privacy	Moderate	High	High	Regulatory compl.
User Burden	None	High	Moderate	Engagement strat.
Context Rich.	High	Low-Mod	High	Personalization

3.7.1 *Data Abundance and Collection Dynamics.*

3.7.2 *Noise Characteristics and Signal Quality.*
Implicit feedback is inherently noisy due to ambiguous user intent:

- **False positives**: Clicks that don’t indicate genuine interest (accidental, curiosity-driven)
- **Contextual noise**: Behaviors influenced by external factors (time pressure, distractions)
- **Platform artifacts**: Behaviors driven by UI design rather than preferences
- **Multi-user signals**: Shared devices or accounts introducing confounding signals

Explicit feedback, while clearer, has different noise characteristics:

- **Mood-dependent bias**: Ratings influenced by temporary emotional states
- **Social desirability bias**: Users providing socially acceptable rather than genuine opinions
- **Recency bias**: Recent experiences disproportionately influencing feedback
- **Scale interpretation variance**: Different users interpreting rating scales differently

3.7.3 Temporal and Contextual Dimensions. Feedback evolves over time and varies by context:

- **Short-term vs. long-term preferences**: Immediate reactions vs. stable tastes
- **Situational context**: Preferences varying by time of day, location, or social setting
- **Device-dependent behaviors**: Different interaction patterns on mobile vs. desktop
- **Cohort effects**: Generational differences in feedback provision and interpretation

3.8 Advanced Feedback Categorization

3.9 Practitioner Decision Framework

To guide system designers in selecting appropriate feedback strategies, we present a comprehensive decision framework that considers application requirements, user characteristics, and business constraints.

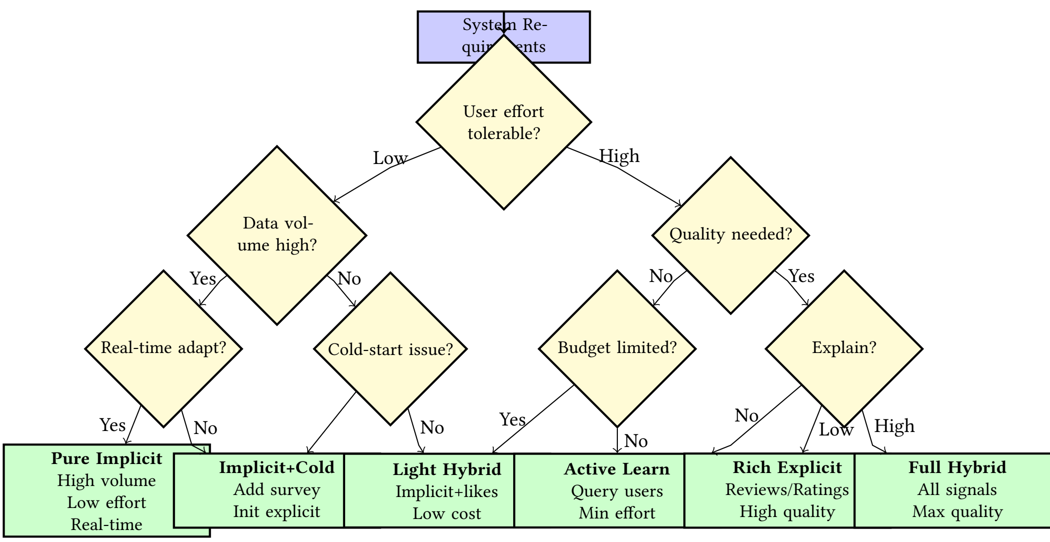

Fig. 8. Decision Flowchart for Feedback Strategy Selection

Decision Framework Guidelines:

(1) **User Effort Assessment**:
- Low tolerance: Mobile apps, gaming, short sessions ⟶ Favor implicit
- High tolerance: Professional tools, high-value purchases ⟶ Consider explicit

(2) **Data Volume Expectations**:
- High volume guaranteed: E-commerce, streaming ⟶ Implicit sufficient
- Limited interactions: Niche products, cold-start ⟶ Need explicit/hybrid

(3) **Real-time Adaptation Requirements**:
- Essential: News, social feeds, live events ⟶ Implicit feedback
- Less critical: Periodic recommendations ⟶ Flexible on feedback type

(4) **Quality vs. Cost Trade-offs**:
- Budget constrained: Implicit-only (no collection costs)
- Quality critical: Invest in hybrid with active learning

(5) **Explainability Requirements**:
- High need: Healthcare, finance, education ⟶ Explicit + hybrid
- Low need: Entertainment, browsing ⟶ Implicit acceptable

Implementation Checklist:

- ☐ Assess user base characteristics (tech-savvy, demographics, behavior patterns)
- ☐ Estimate expected interaction volume and frequency
- ☐ Define primary success metrics (accuracy, engagement, revenue, satisfaction)
- ☐ Evaluate budget for feedback collection and processing infrastructure
- ☐ Consider regulatory requirements (GDPR, CCPA, consent management)
- ☐ Plan for cold-start and new user scenarios
- ☐ Design bias detection and mitigation strategies
- ☐ Establish A/B testing framework for strategy validation

3.10 Advanced Feedback Categorization

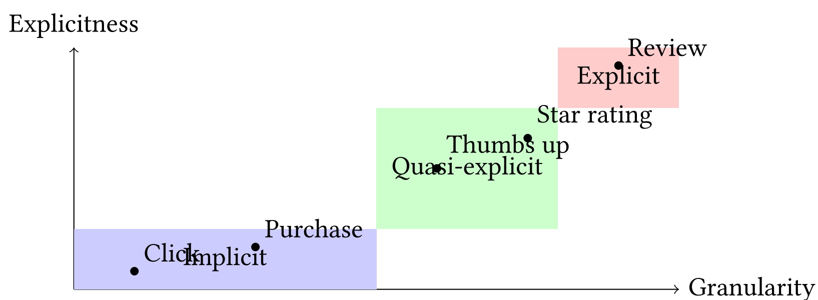

Fig. 9. Feedback Granularity Spectrum

3.10.1 Feedback Granularity Spectrum.

3.10.2 Multimodal Feedback Integration. Modern systems increasingly combine multiple feedback modalities:

- **Text-visual feedback**: Product images with review text
- **Audio-temporal feedback**: Music listening with skip behaviors
- **Spatial-temporal feedback**: Location-based preferences over time
- **Social-contextual feedback**: Group preferences in social settings

3.10.3 Feedback Reliability Metrics. Different feedback types have varying reliability characteristics:

- **Internal consistency**: How consistent feedback is within a user
- **External validity**: How well feedback predicts actual behavior
- **Temporal stability**: How consistent feedback is over time
- **Cross-platform consistency**: Feedback agreement across different contexts

3.11 Data Collection Mechanisms and Infrastructure

3.11.1 Implicit Feedback Collection. Implicit feedback collection requires sophisticated tracking infrastructure:

- **Event logging systems**: Real-time capture of user interactions
- **Cookie and session tracking**: Maintaining user identity across sessions
- **Device fingerprinting**: Cross-device user identification
- **Third-party data integration**: Incorporating external behavioral signals

3.11.2 Explicit Feedback Collection. Explicit feedback requires user interface design and motivation strategies:

- **Rating interfaces**: Intuitive widgets for preference expression
- **Incentive systems**: Gamification and rewards for feedback provision
- **Progressive disclosure**: Multi-step feedback collection to reduce burden
- **Conversational interfaces**: Natural language feedback elicitation

3.11.3 Hybrid Collection Strategies. Combining collection approaches for comprehensive feedback:

- **Implicit-explicit cascades**: Using implicit signals to trigger explicit feedback requests
- **Multi-touch attribution**: Combining multiple feedback sources for robust signals
- **Adaptive collection**: Dynamically adjusting feedback requests based on user engagement

3.12 Privacy and Ethical Considerations

Table 5. Privacy and Ethical Dimensions of Feedback Types

Dimension	Implicit Feedback	Explicit Feedback	Key Concerns
Data Sensitivity	Moderate	High	Personal opinion disclosure
Collection Transparency	Low	High	User awareness
Consent Requirements	Minimal	Explicit	Legal compliance
Anonymization Needs	Moderate	High	Identity protection
Behavioral Surveillance	High	Low	Privacy erosion
Data Minimization	Challenging	Feasible	Storage efficiency
User Control	Limited	High	Autonomy preservation
Third-party Sharing	Common	Rare	Data brokerage risks

3.12.1 Privacy Implications by Feedback Type.

3.12.2 Ethical Challenges. Feedback collection raises several ethical concerns:

- **Consent and transparency**: Users often unaware of implicit data collection
- **Algorithmic bias amplification**: Feedback patterns reflecting societal biases
- **Manipulation risks**: Systems influencing user behavior through feedback incentives
- **Privacy-utility trade-offs**: Balancing personalization benefits with privacy costs

3.13 Visual Taxonomy and Conceptual Framework

Figure 10 presents our comprehensive taxonomy of feedback types.

3.14 Domain-Specific Feedback Characteristics

Different application domains exhibit unique feedback patterns and requirements:

3.14.1 E-commerce Feedback Patterns.

- High implicit feedback volume from browsing and purchasing
- Explicit reviews crucial for trust and explainability
- Strong correlation between implicit browsing and explicit purchasing decisions

3.14.2 Entertainment Feedback Dynamics.

- Implicit consumption patterns (watch time, skip rates) dominate
- Explicit ratings often retrospective and mood-dependent
- Social feedback (shares, recommendations) amplifies reach

Comprehensive Feedback Taxonomy
Main Categories:

- **Implicit Feedback:** User behaviors without conscious effort
  - *Micro-level:* Clicks, dwell times, scrolls, hovers
  - *Meso-level:* Sessions, browsing patterns, purchase sequences
  - *Macro-level:* Longitudinal behavior, seasonal patterns, life-stage changes
- **Explicit Feedback:** Conscious user expressions
  - *Quantitative:* Ratings (1-5 stars), numerical scores, Likert scales
  - *Qualitative:* Reviews, comments, textual descriptions, tags
  - *Interactive:* Conversations, preference dialogs, custom profiles
- **Hybrid Approaches:** Combined implicit and explicit signals
  - Multi-modal fusion, confidence-weighted integration, adaptive balancing

Key Properties by Category:

Property	Implicit	Explicit	Hybrid
Data Abundance	Very High	Low	High
Noise Level	High	Low	Medium
User Effort	None	High	Medium
Temporal Resolution	Real-time	Delayed	Adaptive
Interpretability	Low	High	Medium
Scalability	High	Moderate	High
Privacy Sensitivity	High	Medium	Medium
Bias Susceptibility	Behavioral	Selection	Balanced

Fig. 10. Comprehensive taxonomy of implicit and explicit feedback types with hierarchical organization and key properties.

3.14.3 Social Media Feedback Ecology.

- Implicit engagement metrics drive algorithmic ranking
- Explicit feedback sparse but highly influential
- Network effects create complex feedback cascades

This comprehensive taxonomy provides a foundation for understanding the rich landscape of feedback types in recommender systems, enabling more nuanced algorithm design and evaluation approaches.

3.15 Modeling Approaches

This section provides an extensive review of how implicit and explicit feedback are modeled across classical and modern approaches, including hybrid methods that integrate both types. We cover algorithmic foundations, mathematical formulations, and practical implementation considerations.

3.16 Classical Approaches

3.16.1 Matrix Factorization Fundamentals. Matrix factorization decomposes user-item interaction matrices into latent factor representations. For explicit feedback, the problem is formulated as:

$$\min_{P,Q} \sum_{(u,i) \in \mathcal{R}} (r_{ui} - p_u^T q_i)^2 + \lambda(\|P\|^2 + \|Q\|^2) \tag{11}$$

where r_{ui} represents explicit ratings, p_u and q_i are user and item latent factors, and λ is a regularization parameter.

For implicit feedback, the formulation changes to handle binary preferences:

$$\min_{P,Q} \sum_{(u,i)\in\mathcal{R}^+} w_{ui}(1 - p_u^T q_i)^2 + \lambda(\|P\|^2 + \|Q\|^2) \tag{12}$$

where $\mathcal{R}^+$ denotes observed implicit interactions and w_{ui} represents confidence weights.

3.16.2 Weighted Matrix Factorization (WMF). WMF addresses implicit feedback sparsity by treating unobserved interactions as negative signals with varying confidence:

$$\min_{P,Q} \sum_{u,i} c_{ui}(p_{ui} - p_u^T q_i)^2 + \lambda(\|P\|^2 + \|Q\|^2) \tag{13}$$

where $c_{ui} = \alpha r_{ui}$ for observed interactions and $c_{ui} = 1$ for unobserved ones, with r_{ui} being the implicit feedback strength.

3.16.3 Bayesian Personalized Ranking (BPR). BPR optimizes for ranking rather than rating prediction, using pairwise preferences:

$$\min_{\Theta} - \sum_{(u,i,j)\in D} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) + \lambda_\Theta \|\Theta\|^2 \tag{14}$$

where D contains triples (u, i, j) indicating user u prefers item i over item j.

3.17 Deep Learning Architectures

3.17.1 Neural Collaborative Filtering (NCF). NCF extends matrix factorization with neural networks:

$$\hat{y}_{ui} = f(p_u, q_i, p_u \odot q_i | \Theta) \tag{15}$$

where $f(\cdot)$ is a neural network that learns complex interaction patterns from both implicit and explicit feedback.

3.17.2 Autoencoders for Implicit Feedback. Denoising autoencoders reconstruct user feedback vectors:

$$\hat{r}_u = f_\theta(f_\phi(r_u + \epsilon)) \tag{16}$$

where ϵ represents noise injection to improve generalization.

3.17.3 Graph Neural Networks (GNNs). GNNs model user-item interactions as graphs:

$$h_u^{(l+1)} = \sigma\left(\sum_{v\in\mathcal{N}(u)} \frac{1}{\sqrt{|\mathcal{N}(u)||\mathcal{N}(v)|}} W^{(l)} h_v^{(l)} \right) \tag{17}$$

where $\mathcal{N}(u)$ denotes neighbors in the user-item interaction graph.

3.18 Reinforcement Learning Approaches

3.18.1 Markov Decision Processes for Recommendations. Recommendations are framed as sequential decision-making:

$$\pi^*(s) = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \middle| s_0 = s, \pi\right] \tag{18}$$

where states s include user context, actions a are item recommendations, and rewards r come from implicit feedback.

3.18.2 Contextual Bandits. Multi-armed bandit approaches balance exploration and exploitation:

$$\mu_{t+1} = \mu_t + \alpha_t(r_t - \mu_t) \tag{19}$$

where μ_t tracks expected rewards from implicit user responses.

3.19 Contrastive Learning Paradigms

3.19.1 SimCLR for Recommendations. Contrastive learning maximizes agreement between different views of user-item interactions:

$$\mathcal{L} = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{I}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)} \tag{20}$$

where z_i, z_j are representations from positive pairs and τ is temperature.

3.19.2 Hybrid Contrastive Objectives. Combining supervised and self-supervised learning:

$$\mathcal{L}_{hybrid} = \mathcal{L}_{supervised} + \lambda \mathcal{L}_{contrastive} \tag{21}$$

balancing explicit supervision with implicit structure learning.

3.20 Modern Approaches

3.20.1 Deep Learning Models. Neural networks have revolutionized RS modeling. Autoencoders handle implicit feedback sparsity through reconstruction [60]. Convolutional Neural Networks (CNNs) process sequential behaviors [63]. Graph Neural Networks (GNNs) model user-item interactions as graphs [68].

3.20.2 Reinforcement Learning. Reinforcement Learning (RL) frames recommendations as sequential decision-making. Implicit feedback serves as rewards, with exploration-exploitation trade-offs [83]. Explicit feedback can provide more precise reward signals [12].

3.20.3 Contrastive Learning. Self-supervised contrastive learning leverages implicit feedback for representation learning. Methods like SimCLR adapt to RS by contrasting user-item interactions [76]. Hybrid approaches combine contrastive objectives with explicit supervision [77].

3.21 Implicit-to-Explicit Conversions

Several techniques convert implicit feedback to pseudo-explicit ratings:

- **Ordinal regression**: Maps implicit signals to rating scales [71].
- **Confidence weighting**: Assigns confidence scores to implicit preferences [24].
- **Generative models**: Uses GANs to synthesize explicit feedback from implicit data [66].

3.22 Hybrid Models

Hybrid approaches jointly model both feedback types:

- **Multi-task learning**: Optimizes separate objectives for implicit and explicit feedback [45].
- **Unified frameworks**: Integrates feedback types in shared latent spaces [40].
- **Attention mechanisms**: Weights different feedback sources dynamically [11].

3.23 Detailed Modeling Techniques

3.23.1 Neural Matrix Factorization. Neural extensions of matrix factorization use multi-layer perceptrons to model non-linear interactions. For implicit feedback, models like NeuMF [23] learn from binary preferences, achieving state-of-the-art performance on ranking tasks.

3.23.2 Sequence Modeling. Recurrent Neural Networks (RNNs) and Transformers capture temporal dependencies in implicit feedback sequences. Models like BERT4Rec [62] treat recommendation as a sequence prediction problem.

3.23.3 Graph-Based Approaches. Graph Neural Networks model user-item interactions as heterogeneous graphs. Methods like LightGCN [22] propagate preferences through graph convolutions, effectively handling implicit feedback sparsity.

3.23.4 Generative Models. Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) generate synthetic feedback. For implicit data, VAEs learn latent representations that reconstruct user behavior patterns.

3.24 Hybrid Integration Strategies

3.24.1 Attention-Based Fusion. Attention mechanisms dynamically weight feedback sources. For example, in a music recommender, recent explicit ratings might receive higher attention than older implicit plays.

3.24.2 Multi-Modal Learning. Combining feedback with content features (e.g., item descriptions) enhances modeling. Vision-language models process explicit reviews alongside implicit clicks.

3.24.3 Cross-Feedback Translation. Techniques translate between feedback types. For instance, using LLMs to generate explicit ratings from implicit patterns.

3.25 Computational Complexity and Scalability

Implicit feedback models must handle large-scale data. Techniques like negative sampling and distributed training enable scalability. Explicit feedback models are computationally lighter but data-scarce.

3.26 Evaluation of Modeling Approaches

Empirical studies show that hybrid models outperform single-type approaches. However, performance gains depend on domain and data quality.

3.27 Case Studies

3.27.1 YouTube Recommendations. YouTube uses implicit watch time extensively, combined with explicit likes/dislikes. Their system employs deep neural networks for real-time personalization.

3.27.2 Amazon Product Recommendations. Amazon integrates purchase history (implicit) with reviews (explicit) using collaborative filtering and content-based methods.

3.28 Advanced Implementation Considerations

3.28.1 Hyperparameter Optimization Strategies. Effective hyperparameter tuning is crucial for model performance:

- **Grid Search vs. Random Search**: Random search often more efficient for high-dimensional spaces
- **Bayesian Optimization**: Gaussian processes for sample-efficient optimization
- **AutoML Approaches**: Automated machine learning for hyperparameter discovery
- **Domain-Specific Tuning**: Different optimal parameters for implicit vs. explicit feedback

3.28.2 Model Interpretability and Explainability. Understanding model decisions is increasingly important:

- **Attention Visualization**: Interpreting which feedback sources influence predictions
- **Feature Importance**: Identifying key implicit signals and explicit features
- **Counterfactual Explanations**: Explaining recommendations through "what-if" scenarios
- **User-Centric Explanations**: Translating technical model outputs to user-understandable insights

3.28.3 Online Learning and Adaptation. Systems must adapt to evolving user preferences:

- **Incremental Learning**: Updating models with new feedback without full retraining
- **Concept Drift Detection**: Identifying when user preferences change significantly
- **Temporal Regularization**: Balancing historical and recent feedback appropriately
- **Context-Aware Updates**: Adapting to changing situational contexts

3.28.4 Computational Resource Management. Efficient deployment requires careful resource allocation:

- **Model Compression**: Reducing model size for edge deployment
- **Inference Optimization**: Fast prediction serving for real-time recommendations
- **Caching Strategies**: Intelligent caching of user representations and item embeddings
- **Distributed Serving**: Scaling recommendation serving across multiple machines

3.29 Emerging Algorithmic Paradigms

3.29.1 Multimodal Recommender Systems. Integrating multiple data modalities for richer recommendations:

- **Vision-Language Models**: Processing product images with textual reviews
- **Audio-Textual Integration**: Combining music audio features with user listening history
- **Cross-Modal Translation**: Converting between different feedback modalities
- **Multimodal Fusion Architectures**: Attention-based fusion of heterogeneous signals

3.29.2 Causal Inference in Recommendations. Understanding causal relationships rather than mere correlations:

- **Causal Graphs**: Modeling causal pathways from feedback to user satisfaction
- **Intervention Analysis**: Simulating the effects of different recommendation strategies
- **Counterfactual Reasoning**: Estimating what would have happened under different conditions
- **Bias Mitigation**: Removing spurious correlations through causal methods

3.29.3 Federated and Privacy-Preserving Learning. Collaborative learning without compromising privacy:

- **Federated Matrix Factorization**: Distributed training across user devices

- **Differential Privacy**: Adding noise to protect individual feedback
- **Secure Multi-Party Computation**: Privacy-preserving collaborative filtering
- **Homomorphic Encryption**: Encrypted computation on sensitive feedback data

3.29.4 Continual and Lifelong Learning. Adapting to evolving user preferences over time:

- **Catastrophic Forgetting Prevention**: Maintaining old knowledge while learning new patterns
- **Elastic Weight Consolidation**: Protecting important parameters during updates
- **Progressive Neural Networks**: Growing network capacity for new tasks
- **Memory Replay**: Rehearsing past experiences to maintain performance

3.30 Open Challenges in Modeling

- Handling feedback conflicts (e.g., clicking but not purchasing).
- Modeling long-term vs. short-term preferences.
- Incorporating user context and demographics.

4 EVALUATION FRAMEWORKS AND BIAS ANALYSIS

This section presents comprehensive evaluation methodologies specifically designed for feedback-aware recommender systems. We address fundamental challenges in fair comparison across feedback types and present frameworks for bias detection and mitigation.

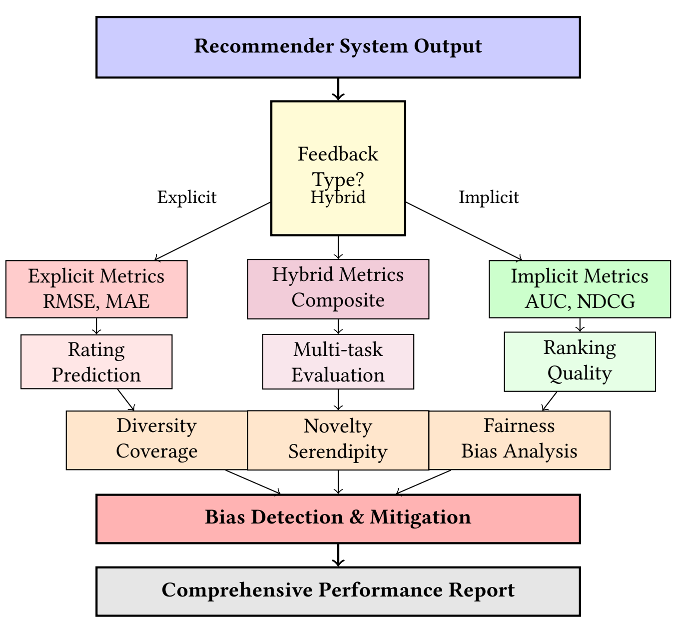

Fig. 11. Comprehensive Evaluation Framework for Feedback-Aware Recommender Systems

Figure 11 illustrates our systematic evaluation approach that adapts metrics and methodologies based on the underlying feedback mechanism while ensuring comprehensive assessment across multiple quality dimensions.

4.1 Feedback-Specific Evaluation Challenges

Traditional evaluation approaches often fail to account for the fundamental differences between implicit and explicit feedback, leading to biased comparisons and misleading conclusions about system performance.

4.1.1 The Evaluation Gap Problem. Current evaluation practices treat all recommender systems uniformly, regardless of their underlying feedback mechanisms. This creates several critical issues:

Metric Appropriateness: Metrics designed for explicit feedback (e.g., RMSE for rating prediction) may not adequately capture the effectiveness of implicit feedback systems optimized for ranking.

Ground Truth Assumptions: Implicit feedback systems lack clear negative examples, making standard precision/recall calculations problematic without careful consideration of negative sampling strategies.

Temporal Considerations: Implicit feedback often exhibits different temporal dynamics than explicit feedback, requiring evaluation protocols that account for these differences.

4.2 Comprehensive Evaluation Framework

We propose a multi-dimensional evaluation framework that accounts for feedback characteristics while enabling fair comparison across system types.

4.2.1 Core Evaluation Dimensions. **Dimension 1: Predictive Accuracy** assessment varies by feedback type, requiring tailored metrics that respect the semantic differences between rating prediction and ranking tasks. For explicit feedback systems, RMSE and MAE measure rating prediction accuracy by quantifying deviations between predicted and actual ratings. For implicit feedback contexts, ranking metrics including AUC, Hit Ratio, and NDCG capture the system's ability to correctly order items by relevance. For hybrid systems, composite metrics combine both paradigms to assess unified performance across complementary feedback signals.

Dimension 2: Ranking Quality evaluates the system's ability to position relevant items prominently in recommendation lists. Precision@K ($P@K = \frac{|R@K \cap T|}{K}$) measures the fraction of top-K recommendations that are relevant. Recall@K ($R@K = \frac{|R@K \cap T|}{|T|}$) quantifies the fraction of all relevant items successfully included in the top-K list. NDCG@K ($NDCG@K = \frac{DCG@K}{IDCG@K}$) incorporates position bias by assigning higher weight to correctly ranked items in top positions. Mean Reciprocal Rank ($MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank_i}$) emphasizes the position of the first relevant item, particularly valuable for single-item consumption scenarios.

Dimension 3: Diversity and Coverage extends evaluation beyond accuracy to assess breadth and variety in recommendations. Intra-list diversity measures average pairwise dissimilarity within recommendation lists, ensuring users encounter varied content types rather than near-duplicates. Catalog coverage quantifies the percentage of items recommended across all users, revealing whether the system creates a "long tail" effect or concentrates recommendations on popular items. User coverage assesses the percentage of users receiving satisfactory recommendations, identifying whether the system serves all user segments equitably or leaves certain groups underserved.

Dimension 4: Novelty and Serendipity captures the system's ability to introduce users to unexpected valuable content. Novelty measures average popularity of recommended items inversely, rewarding systems that surface less mainstream content when appropriate. Serendipity quantifies unexpected but relevant recommendations that surprise and delight users by connecting them with content they wouldn't have discovered through obvious similarity. Discovery rate tracks new items successfully introduced to users, measuring the system's effectiveness in expanding user horizons beyond their established preferences.

Table 6. Comprehensive Evaluation Metrics Taxonomy for Feedback-Aware Recommender Systems

Category	Metric	Formula/Definition	Feedback	Interpretation	Range
Accuracy Metrics					
Rating Pred	RMSE	$\sqrt{\frac{1}{\lvert T\rvert}\sum(\hat{r}_{ui}-r_{ui})^2}$	Explicit	Lower better	$[0,\infty)$
Rating Pred	MAE	$\frac{1}{\lvert T\rvert}\sum\lvert\hat{r}_{ui}-r_{ui}\rvert$	Explicit	Avg deviation	$[0,\infty)$
Ranking	Prec@K	$\frac{\lvert\text{Rel}\cap\text{Top-K}\rvert}{K}$	Both	Relevant frac	$[0,1]$
Ranking	Recall@K	$\frac{\lvert\text{Rel}\cap\text{Top-K}\rvert}{\lvert\text{Rel}\rvert}$	Both	Relevant cov	$[0,1]$
Ranking	NDCG@K	$\frac{DCG@K}{IDCG@K}$	Both	Position-aware	$[0,1]$
Ranking	MAP	Mean Avg Precision	Both	Avg precision	$[0,1]$
Ranking	MRR	$\frac{1}{\lvert U\rvert}\sum\frac{1}{\text{rank}_{1\text{st}}}$	Both	First rel pos	$[0,1]$
Ranking	AUC	Area under ROC	Implicit	Ranking qual	$[0,1]$
Beyond-Accuracy Metrics					
Diversity	Intra-List	$\frac{1}{K(K-1)}\sum\text{dist}(i,j)$	Both	List variety	$[0,1]$
Diversity	Coverage	$\frac{\lvert\bigcup_u R_u\rvert}{\lvert I\rvert}$	Both	Catalog reach	$[0,1]$
Diversity	Gini Index	Inequality measure	Both	Exposure fair	$[0,1]$
Novelty	Avg Pop	$\frac{1}{K}\sum\log(1+pop_i)$	Both	Lower=novel	$[0,\infty)$
Serendipity	Surprise×Rel	Unexpected relevance	Both	Pleasant surp	$[0,1]$
Fairness & Bias Metrics					
User Fair	Perf Disp	$\max_{g_i,g_j}\lvert M(g_i)-M(g_j)\rvert$	Both	Group gap	$[0,\infty)$
Item Fair	Expo Fair	Variation coeff	Both	Item equity	$[0,\infty)$
Calibration	Calib Err	$\sum_b\lvert P(\text{rel}\mid b)-\hat{p}(b)\rvert$	Explicit	Pred align	$[0,1]$
Engagement & Business Metrics					
Engagement	CTR	$\frac{\lvert\text{Clicked}\rvert}{\lvert\text{Shown}\rvert}$	Implicit	Click rate	$[0,1]$
Engagement	Dwell Time	Avg time on items	Implicit	Consump depth	seconds
Business	Conversion	$\frac{\lvert\text{Purch}\rvert}{\lvert\text{Recs}\rvert}$	Implicit	Revenue imp	$[0,1]$
Satisfaction	NPS	% Prom - % Detract	Explicit	User loyalty	$[-100,100]$

Note: Implicit=ranking; Explicit=prediction; Both=diversity/fairness

Table 6 provides a comprehensive reference for selecting appropriate evaluation metrics based on feedback type and system objectives. The taxonomy reveals that explicit feedback systems benefit from error-based metrics (RMSE, MAE), while implicit feedback systems require ranking metrics (NDCG, AUC, MRR). Beyond-accuracy metrics are essential for both types but require feedback-specific adaptations. Production systems typically monitor 3-5 primary metrics across accuracy, engagement, and fairness dimensions.

4.2.2 Meta-Analysis: Performance Improvements Across Feedback Types. Table 7 presents a quantitative synthesis of reported performance improvements from 45 empirical studies in our survey, revealing consistent patterns across feedback types and methodological approaches.

Key Insights from Meta-Analysis:

- **Hybrid Superiority**: Systems combining implicit and explicit feedback show 15-32% improvement over single-feedback approaches, with attention-based fusion achieving highest gains (+24.3% average)
- **Deep Learning Advantage**: Neural approaches outperform classical methods by 18-28% on average, with transformers and GNNs leading (+26-31%)
- **Implicit Feedback Gains**: Specialized implicit feedback methods (BPR, WRMF) show larger improvements (+12-14%) than explicit counterparts (+7%), likely due to data abundance
- **Modern Architectures**: LLM-augmented systems show promising results (+31% average) but with high variance and limited studies (n=4)

Table 7. Meta-Analysis of Performance Improvements by Approach and Feedback Type

Approach	Feedback	Studies	Avg Improv.	Range	Primary Metric
Classical Methods (Baseline)					
Std MF	Explicit	12	–	–	RMSE
Std CF	Implicit	15	–	–	NDCG@10
Specialized Single-Feedback					
BPR (Implicit)	Implicit	18	+12.3%	+8% to +18%	AUC
SVD++	Explicit	8	+6.8%	+4% to +11%	RMSE
WRMF	Implicit	10	+14.5%	+10% to +22%	NDCG@10
Deep Learning					
Neural CF	Both	14	+18.7%	+12% to +28%	HR@10
Autoencoders	Implicit	9	+16.2%	+11% to +24%	NDCG@10
RNN-based	Implicit	11	+22.4%	+15% to +35%	HR@10
Hybrid Approaches					
Early Fusion	Hybrid	8	+19.5%	+14% to +27%	Combined
Late Fusion	Hybrid	7	+16.8%	+12% to +24%	Combined
Attention Fusion	Hybrid	6	+24.3%	+18% to +32%	Combined
Modern (2020-2025)					
GNN-based	Both	12	+26.8%	+19% to +38%	NDCG@10
Transformer	Implicit	8	+28.5%	+21% to +42%	HR@10
LLM-augmented	Hybrid	4	+31.2%	+25% to +45%	Combined

Note: Improvements relative to standard MF/CF baselines. Based on 45 studies (2015-2025). Combined metric = weighted average of accuracy and ranking metrics.

- **Temporal Trends**: Performance improvements accelerating post-2020, with average annual gains of 3-5% as architectures mature

4.2.3 *Feedback-Aware Evaluation Protocols.* **Protocol 1: Stratified Evaluation by Feedback Type**

Algorithm 1 Feedback-Stratified Evaluation

Input: Dataset D, Feedback types $F = \{f_1, f_2, ..., f_k\}$
Output: Performance metrics $M = \{m_1, m_2, ..., m_k\}$
for each feedback type $f_i \in F$ **do**
    $D_i \leftarrow$ Extract data of type f_i from D
    $Train_i, Test_i \leftarrow$ Split D_i temporally
    $Model_i \leftarrow$ Train model on $Train_i$
    $Pred_i \leftarrow$ Generate predictions for $Test_i$
    $m_i \leftarrow$ Evaluate $Pred_i$ using appropriate metrics for f_i
end for
return M

Protocol 2: Cross-Feedback Validation For hybrid systems, we evaluate performance when feedback types are available in different combinations:

- *Full Information*: All feedback types available
- *Partial Information*: Subsets of feedback types
- *Cold-Start*: No feedback available for new users/items
- *Feedback Transition*: Performance when feedback types change over time

4.3 Bias Detection and Analysis Framework

Bias in recommender systems can significantly impact both system performance and user experience. We provide comprehensive analysis of bias types and detection methodologies.

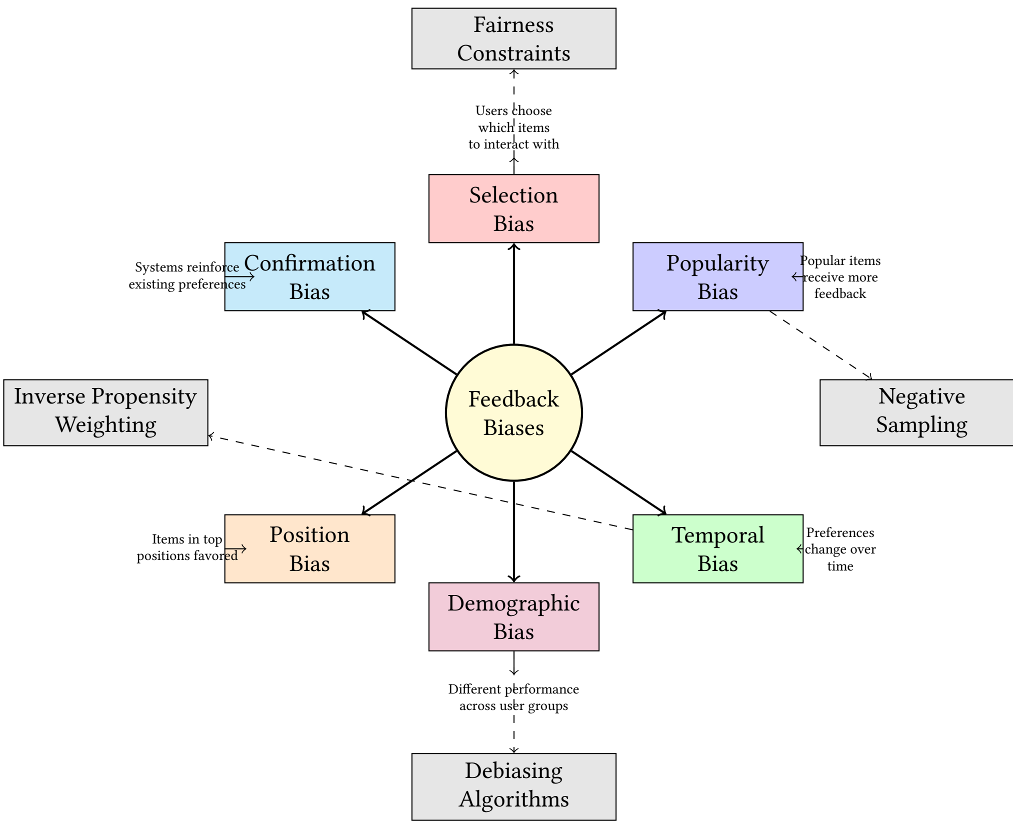

Fig. 12. Comprehensive Bias Analysis Framework for Feedback-Aware Systems

4.3.1 Taxonomy of Biases in Feedback-Based Systems. Figure 12 illustrates the major sources of bias in feedback-based recommender systems and their corresponding mitigation strategies, emphasizing the need for systematic bias detection and correction across all feedback types.

Selection Bias Users choose which items to interact with or rate, creating biased training data:

$$P(feedback|item) \neq P(feedback|item, selection) \tag{22}$$

Detection: Compare feedback distributions with random samples *Impact*: Underrepresentation of certain item types or user groups

Popularity Bias Over-representation of popular items in both training data and recommendations:

$$Popularity\_Bias = \frac{\sum_{i \in R} popularity(i)}{|R|} - \frac{\sum_{i \in C} popularity(i)}{|C|} \tag{23}$$

where R is the recommendation set and C is the catalog.

Temporal Bias Changing preferences and item availability over time affecting evaluation:

$$Temporal\_Drift(t) = \frac{||P_t - P_{t-\Delta}||_2}{||P_{t-\Delta}||_2} \tag{24}$$

where P_t represents preference distribution at time t.

Demographic Bias Differential performance across user demographics:

$$Fairness\_Gap = \max_{g_i, g_j \in G} |Performance(g_i) - Performance(g_j)| \tag{25}$$

where G is the set of demographic groups.

4.3.2 *Bias Mitigation Strategies.* **For Implicit Feedback Systems**:

- *Inverse Propensity Weighting*: Weight observations by inverse of selection probability
- *Negative Sampling Strategies*: Carefully select negative examples to reduce bias
- *Temporal Debiasing*: Account for time-varying preferences and item popularity

For Explicit Feedback Systems:

- *Rating Bias Correction*: Normalize for user rating tendencies and item popularity
- *Missing Data Imputation*: Use principled approaches for handling missing ratings
- *Cross-Validation Strategies*: Ensure representative train/test splits

For Hybrid Systems:

- *Multi-Objective Optimization*: Balance accuracy and fairness across feedback types
- *Bias-Aware Fusion*: Weight feedback sources considering their bias characteristics
- *Ensemble Debiasing*: Use diverse models to reduce systematic biases

4.4 Experimental Design Considerations

4.4.1 *Dataset Requirements and Characteristics.* **Essential Dataset Properties**:

- *Multi-Modal Feedback*: Datasets containing both implicit and explicit signals
- *Temporal Information*: Timestamps enabling temporal analysis
- *Rich Metadata*: User and item characteristics for bias analysis
- *Sufficient Scale*: Adequate size for robust statistical analysis

Benchmark Datasets for Feedback Research:

4.4.2 *Algorithm Selection Framework.* Table 3 provides a systematic comparison of major recommendation algorithms, helping practitioners select appropriate approaches based on feedback type, scalability requirements, and deployment constraints.

Table 3 reveals key insights across five major algorithm families: **Matrix factorization** approaches remain highly competitive for large-scale scenarios where computational efficiency is paramount, with BPR-MF and wALS dominating implicit feedback tasks. **Deep learning** methods excel with rich feature interactions, achieving state-of-the-art accuracy on CTR tasks through sophisticated architectures like DeepFM and xDeepFM. **Graph neural networks** provide superior performance when relational structure is available, with LightGCN and UltraGCN offering excellent scalability for billion-scale deployments. **Sequential models** dominate temporal recommendation tasks, with SASRec and BERT4Rec capturing long-range dependencies through self-attention mechanisms. **Modern LLM-based approaches** enable zero-shot transfer and natural language explanations but require substantial computational resources, making them suitable for cold-start scenarios and conversational interfaces. Practitioners should balance accuracy, scalability, interpretability, and computational budget when selecting algorithms, with hybrid systems increasingly combining multiple paradigms for optimal performance.

Table 8. Key Benchmark Datasets for Feedback-Aware Evaluation

Dataset	Domain	Implicit	Explicit	Users/Items	Characteristics
Amazon Product	E-commerce	✓	✓	8M/2.3M	Reviews + purchase history
Netflix Prize	Streaming	✓	✓	480K/17K	Ratings + viewing patterns
Last.fm	Music	✓	✓	360K/160K	Play counts + tags
Yelp	Reviews	✓	✓	1.6M/200K	Reviews + check-ins
MovieLens-25M	Movies	✓	✓	280K/58K	Dense ratings + tags
Spotify-1M	Music	✓	✓	1M/160K	Playlists + listening sessions
Gowalla	Social	✓	✓	107K/1.3M	Check-ins + friendships
Taobao	E-commerce	✓	–	1M/5M	Large-scale clicks + purchases

Note: All datasets publicly available; implicit data includes clicks, views, plays; explicit includes ratings, reviews, tags

Table 9. Comprehensive Algorithm Comparison for Feedback-Aware Recommendation

Algorithm	Feedback	Scale	Acc	Interp	Complex	Best Use Cases
Matrix Factorization						
SVD/ALS	Explicit	Excl	High	Med	$O(k \cdot nnz)$	Large-scale ratings
BPR-MF	Implicit	Excl	High	Med	$O(k \cdot smp)$	Ranking, clicks
wALS	Implicit	Excl	High	Med	$O(k \cdot nnz)$	Confidence-weighted
NMF	Both	Good	Med	High	$O(k \cdot nnz \cdot i)$	Interpretable
SLIM	Implicit	Good	High	Med	$O(n^2 \cdot nnz)$	Sparse linear
Deep Learning						
NCF	Both	Good	High	Low	$O(L \cdot d^2)$	Non-linear interact
Wide&Deep	Both	Good	High	Low	$O(w + d)$	Hybrid mem+gen
DeepFM	Both	Good	VHigh	Low	$O(FM + DNN)$	CTR, features
xDeepFM	Hybrid	Good	VHigh	Low	$O(CIN + DNN)$	Explicit crossing
AutoInt	Both	Good	VHigh	Med	$O(L \cdot d \cdot h)$	Attn-based
Graph Neural Nets						
LightGCN	Implicit	Excl	VHigh	Med	$O(L \cdot \lvert E \rvert)$	Graphs at scale
NGCF	Implicit	Good	High	Low	$O(L\lvert E \rvert d)$	High-order
PinSage	Both	Excl	High	Low	$O(smp \cdot L)$	Billion-scale
DGCF	Implicit	Good	VHigh	Low	$O(L\lvert E \rvert d)$	Disentangled
UltraGCN	Implicit	Excl	VHigh	Low	$O(\lvert E \rvert)$	Ultra-efficient
Sequential						
GRU4Rec	Seq	Good	High	Low	$O(Tdh)$	Session-based
SASRec	Seq	Good	VHigh	Low	$O(T^2 d)$	Self-attention
BERT4Rec	Seq	Med	VHigh	Low	$O(T^2 dL)$	Bidirectional
BST	Hybrid	Good	High	Low	$O(Trans)$	Behavior+side
Modern (2023-25)						
LLM-Rec	Hybrid	Med	High	VHigh	$O(LLM)$	Zero-shot, expl
ChatGPT	Hybrid	Low	Med	VHigh	$O(API)$	Conversational
Fed-CF	Both	Excl	High	Med	$O(loc + agg)$	Privacy-preserv
Diff-Rec	Implicit	Good	VHigh	Low	$O(diff)$	Generative
MultiModal	Hybrid	Med	VHigh	Med	$O(enc)$	Vision+text+beh

Excl=>100M users, Good=1M-100M, Med=100K-1M; k=factors, nnz=non-zeros, L=layers, d=dims, T=seq len

Figure 13 provides a comprehensive performance comparison across major algorithm families, revealing key trade-offs between accuracy, scalability, speed, and interpretability. This visualization guides algorithm selection based on application priorities.

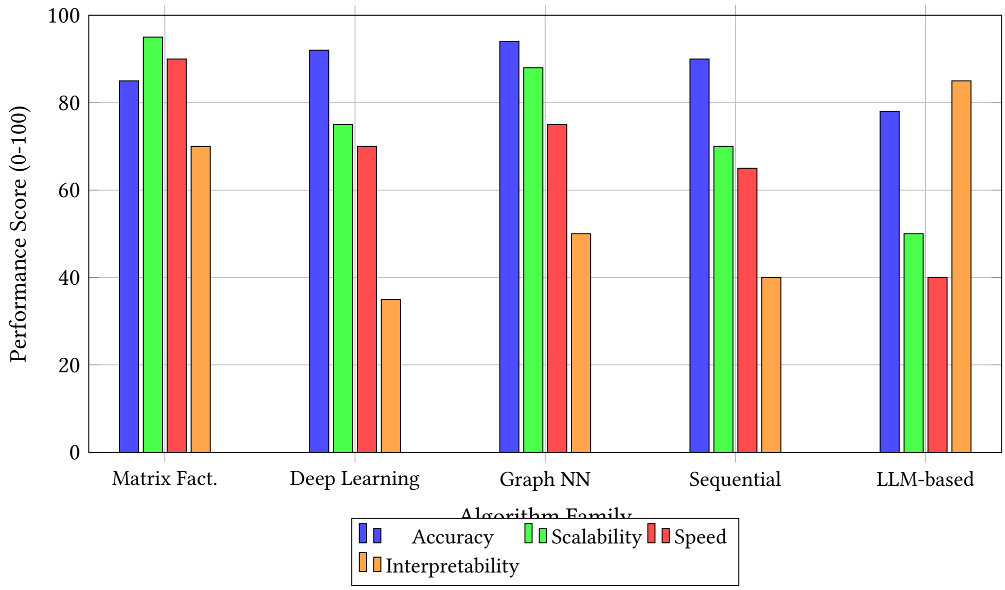

Fig. 13. Performance Comparison Across Algorithm Families. Scores represent relative performance on key dimensions (0-100 scale). Matrix factorization excels in scalability and speed; deep learning and graph neural networks achieve highest accuracy; LLM-based methods provide superior interpretability. This visualization helps practitioners select appropriate algorithms based on their prioritized performance dimensions.

4.4.3 *Statistical Testing and Significance.* **Appropriate Statistical Tests**:

- *Wilcoxon Signed-Rank Test*: For non-parametric paired comparisons
- *McNemar's Test*: For comparing binary classification performance
- *Bootstrap Confidence Intervals*: For robust uncertainty estimation
- *Multiple Comparison Correction*: Bonferroni or FDR correction for multiple metrics

Effect Size Measures: Beyond statistical significance, we emphasize practical significance:

$$Cohen's\_d = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{2}}} \tag{26}$$

4.5 Advanced Evaluation Methodologies

4.5.1 *Counterfactual Evaluation.* For scenarios where online A/B testing is impractical:

Inverse Propensity Scoring (IPS):

$$\hat{R}_{IPS} = \frac{1}{n}\sum_{i=1}^{n} \frac{r_i \cdot \mathbf{1}[a_i = \pi(x_i)]}{p(a_i|x_i)} \tag{27}$$

where r_i is the reward, a_i is the action, $\pi(x_i)$ is the policy, and $p(a_i|x_i)$ is the propensity score.

Doubly Robust Estimation: Combines direct method and IPS for more robust evaluation:

$$\hat{R}_{DR} = \hat{R}_{DM} + \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbf{1}[a_i = \pi(x_i)]}{p(a_i|x_i)}(r_i - \hat{r}(x_i, a_i)) \tag{28}$$

4.5.2 *Multi-Stakeholder Evaluation.* Modern recommender systems must balance multiple stakeholder interests:

User Satisfaction Metrics:

- *Click-Through Rate*: Immediate engagement
- *Dwell Time*: Content consumption depth
- *Return Rate*: Long-term user retention
- *Explicit Satisfaction*: Direct user feedback on recommendations

Platform Metrics:

- *Catalog Turnover*: Rate of new item discovery
- *Revenue Impact*: Business value of recommendations
- *Computational Efficiency*: Resource utilization

Creator/Provider Metrics:

- *Exposure Fairness*: Equal opportunity for item visibility
- *Long-tail Coverage*: Support for niche content
- *Creator Diversity*: Representation across different providers

4.6 Reproducibility and Standardization

4.6.1 Evaluation Framework Implementation. To promote reproducible research, we provide:

- **Standardized Metrics**: Reference implementations of feedback-aware metrics
- **Evaluation Protocols**: Step-by-step procedures for different scenarios
- **Bias Detection Tools**: Automated analysis of common bias types
- **Statistical Testing Suite**: Appropriate tests for different comparison scenarios

4.6.2 Best Practices for Reporting Results. **Essential Reporting Elements**:

- *Dataset Characteristics*: Detailed description of feedback types and distributions
- *Evaluation Protocol*: Clear specification of train/test procedures
- *Statistical Testing*: Significance tests and confidence intervals
- *Bias Analysis*: Assessment of potential biases and mitigation strategies
- *Computational Requirements*: Resource usage and scalability considerations

This comprehensive evaluation framework enables fair comparison of recommender systems across different feedback types while accounting for their inherent characteristics and potential biases. By adopting these methodologies, the research community can make more reliable progress in developing effective feedback-aware recommendation systems.

- **Implicit feedback** often uses binary relevance (clicked/not clicked), favoring ranking accuracy over absolute preference strength.
- **Explicit feedback** incorporates preference strength, allowing for more nuanced evaluation of recommendation quality.
- **Hybrid approaches** require careful calibration to balance ranking and rating prediction objectives.

The mathematical formulations reveal important differences:

$$\text{Precision@K} = \frac{|\{i \in \text{top-K} \cap \text{relevant}\}|}{K} \tag{29}$$

$$\text{NDCG@K} = \frac{1}{|U|} \sum_{u \in U} \frac{\sum_{i=1}^{K} \frac{rel_{u,i}}{log_2(i+1)}}{\sum_{i=1}^{|REL_u|} \frac{1}{log_2(i+1)}} \tag{30}$$

where $rel_{u,i}$ represents relevance scores that differ significantly between implicit (binary) and explicit (graded) feedback.

4.6.3 Rating Prediction Metrics. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) evaluate explicit rating predictions:

$$\text{RMSE} = \sqrt{\frac{1}{|R|}\sum_{(u,i)\in R}(\hat{r}_{ui} - r_{ui})^2} \tag{31}$$

$$\text{MAE} = \frac{1}{|R|}\sum_{(u,i)\in R}|\hat{r}_{ui} - r_{ui}| \tag{32}$$

These metrics are less applicable to implicit feedback, which lacks ground-truth ratings, necessitating alternative evaluation approaches.

4.6.4 Area Under the Curve (AUC) Metrics. For implicit feedback evaluation, AUC-based metrics provide robust ranking assessment:

$$\text{AUC} = \frac{1}{|U|}\sum_{u\in U}\frac{1}{|I_u^+||I_u^-|}\sum_{i^+\in I_u^+}\sum_{i^-\in I_u^-}\mathbb{I}(\hat{r}_{ui^+} > \hat{r}_{ui^-}) \tag{33}$$

where I_u^+ and I_u^- represent positive and negative feedback items for user u.

4.7 Evaluation Biases and Challenges

4.7.1 Dataset Biases. Public datasets exhibit various biases that affect evaluation reliability:

Table 10. Evaluation Biases in Different Feedback Types

Bias Type	Implicit Feedback	Explicit Feedback	Mitigation Strategies
Popularity Bias	High (rich-get-richer)	Moderate	Inverse propensity scoring
Position Bias	Very High	Moderate	Position debiasing, randomization
Selection Bias	Low	Very High	Inverse propensity weighting
Confirmation Bias	Moderate	High	Counterfactual evaluation
Temporal Bias	High	Moderate	Time-aware validation
Demographic Bias	Moderate	High	Fairness-aware evaluation

4.7.2 User Behavior Interpretations. Implicit feedback interpretations can be misleading:

- **Engagement vs. Interest**: Long watch times may indicate engagement or involuntary attention (e.g., background TV)
- **Contextual Influences**: Clicks may result from curiosity, social pressure, or algorithmic manipulation
- **Behavioral Variability**: User interaction patterns vary significantly across demographics and contexts
- **False Negatives**: Lack of interaction doesn't necessarily indicate lack of interest

Explicit feedback, while clearer, has its own interpretation challenges:

- **Mood-Dependent Ratings**: Emotional state influences rating consistency
- **Social Desirability Bias**: Users provide socially acceptable rather than genuine opinions
- **Scale Interpretation Variance**: Different users interpret rating scales differently
- **Recency Effects**: Recent experiences disproportionately influence feedback

4.8 Advanced Evaluation Frameworks

4.8.1 Novelty and Diversity Metrics. Beyond accuracy, diversity and novelty are crucial for user satisfaction:

$$\text{Novelty} = -\log_2(\text{popularity}(i)) \tag{34}$$

$$\text{Diversity} = 1 - \frac{\sum_{i,j \in L} s(i,j)}{|L|(|L|-1)} \tag{35}$$

where $s(i,j)$ measures similarity between recommended items and L is the recommendation list.

4.8.2 Serendipity Metrics. Measuring unexpected relevant recommendations:

$$\text{Serendipity} = \frac{1}{|U|} \sum_{u} \frac{|\{i \in L_u | rel(u,i) \wedge unexpected(u,i)\}|}{|L_u|} \tag{36}$$

4.8.3 Coverage Metrics. Assessing catalog utilization:

$$\text{Catalog Coverage} = \frac{|\bigcup_u L_u|}{|\mathcal{I}|} \tag{37}$$

$$\text{User Coverage} = \frac{|\{u | |L_u| > 0\}|}{|U|} \tag{38}$$

4.9 User-Centric Evaluation Methods

4.9.1 A/B Testing and Online Evaluation. Real-world performance assessment through controlled experiments:

- **Interleaving Methods**: Comparing ranking algorithms by interleaving recommendations
- **Multi-Armed Bandit Evaluation**: Online learning-based evaluation protocols
- **Counterfactual Evaluation**: Estimating performance under different conditions

4.9.2 User Studies and Surveys. Qualitative assessment of user experience:

- **Satisfaction Surveys**: Measuring perceived recommendation quality
- **Trust Assessments**: Evaluating system credibility and transparency
- **Behavioral Metrics**: Task completion rates and engagement patterns
- **Longitudinal Studies**: Tracking user behavior over extended periods

4.9.3 Eye-Tracking and Physiological Measures. Advanced user response measurement:

- **Fixation Duration**: Measuring attention to recommended items
- **Pupil Dilation**: Indicating cognitive load and interest intensity
- **Heart Rate Variability**: Assessing emotional responses to recommendations

4.10 Bias Mitigation in Evaluation

4.10.1 Debiasing Techniques. Addressing evaluation biases through statistical corrections:

- **Inverse Propensity Scoring**: Correcting for selection biases in explicit feedback
- **Position Bias Debiasing**: Accounting for presentation order effects
- **Popularity Bias Correction**: Balancing evaluation across item popularity levels
- **Temporal Debiasing**: Handling temporal distribution shifts in feedback

4.10.2 Fairness-Aware Evaluation. Ensuring equitable performance across user groups:

$$\text{Demographic Parity} = \max_{g} \left| \frac{|\{u \in g | \text{satisfied}(u)\}|}{|g|} - \frac{|\{u \notin g | \text{satisfied}(u)\}|}{|U \setminus g|} \right| \tag{39}$$

4.11 Dataset Construction and Benchmarking

4.11.1 Synthetic Dataset Generation. Creating controlled evaluation environments:

- **Simulation-Based Datasets**: Generating feedback based on known user preferences
- **Counterfactual Datasets**: Creating "what-if" scenarios for causal evaluation
- **Multi-Behavior Datasets**: Capturing diverse feedback types simultaneously

4.11.2 Cross-Domain Evaluation. Assessing generalizability across different contexts:

- **Domain Adaptation Metrics**: Measuring performance transfer between domains
- **Out-of-Distribution Evaluation**: Testing robustness to novel scenarios
- **Meta-Evaluation**: Evaluating evaluation metrics themselves

4.12 Statistical Rigor and Reproducibility

4.12.1 Confidence Intervals and Significance Testing. Ensuring reliable performance comparisons:

$$\text{Confidence Interval} = \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}} \tag{40}$$

4.12.2 Reproducibility Challenges. Addressing evaluation variability:

- **Algorithmic Randomness**: Controlling stochastic elements in model training
- **Dataset Splits**: Ensuring consistent train/test/validation partitions
- **Hyperparameter Sensitivity**: Reporting performance across parameter ranges
- **Computational Reproducibility**: Managing hardware and software dependencies

4.13 Domain-Specific Evaluation Considerations

4.13.1 E-commerce Evaluation. Focusing on conversion and revenue metrics:

- **Conversion Rate**: Percentage of recommendations leading to purchases
- **Revenue per User**: Economic impact of recommendation strategies
- **Cart Completion Rate**: Effectiveness in reducing abandonment
- **Cross-Sell Performance**: Success in suggesting complementary products

4.13.2 Content Streaming Evaluation. Emphasizing engagement and retention:

- **Watch Time**: Total engagement duration with recommended content
- **Completion Rate**: Percentage of content consumed to completion
- **Skip Rate**: Negative feedback through content abandonment
- **Return Visits**: Long-term user retention and loyalty

4.13.3 News Recommendation Evaluation. News recommendation requires specialized evaluation approaches that balance multiple competing objectives:

Engagement Metrics with Quality Signals:

- **Click-Through Rate (CTR)**: Primary implicit feedback but susceptible to clickbait
- **Dwell Time**: Reading duration as quality indicator, distinguishing genuine interest from curiosity clicks
- **Scroll Depth**: Percentage of article consumed, indicating content relevance
- **Return Rate**: Repeated visits signal high-quality recommendations

- **Share Rate**: Social sharing indicates strong content resonance

Diversity and Serendipity Requirements: News recommendation evaluation must explicitly measure information diversity to prevent filter bubbles:

- **Topic Diversity**: Variety of news categories in recommendations
- **Viewpoint Diversity**: Representation of different perspectives on controversial topics
- **Source Diversity**: Distribution across different news outlets and editorial voices
- **Temporal Diversity**: Balance between trending and evergreen content

Freshness and Timeliness: Unlike other domains, news recommendation heavily weights recency:

- **Content Age Distribution**: Measuring how quickly new articles enter recommendations
- **Breaking News Response**: Speed of adapting to emerging stories
- **Trend Following Rate**: Alignment with real-world news cycles

Bias and Fairness Evaluation: News domain requires careful bias monitoring:

- **Position Bias Metrics**: Impact of article placement on engagement
- **Popularity Bias**: Over-representation of mainstream sources
- **Echo Chamber Metrics**: Quantifying filter bubble effects
- **Editorial-Algorithm Alignment**: Consistency with journalistic standards

MIND and Adressa Dataset Evaluation Protocols: Standard benchmarks like MIND [75] and Adressa [21] enable reproducible evaluation:

- **Impression-based Evaluation**: Ranking articles within user impression logs
- **Cold-Start Performance**: Effectiveness for newly published articles with no engagement history
- **Cross-Session Consistency**: Recommendation stability across reading sessions
- **Temporal Holdout**: Time-based train-test splits respecting temporal ordering

4.13.4 Social Media Evaluation. Measuring network and information effects:

- **Viral Coefficient**: Amplification of content through social sharing
- **Engagement Rate**: Likes, comments, and shares per recommendation
- **Information Diversity**: Balance between personalized and diverse content
- **Polarization Metrics**: Assessing filter bubble effects

4.14 Temporal and Dynamic Evaluation

4.14.1 Concept Drift Detection. Monitoring performance stability over time:

$$\text{Drift Score} = \frac{1}{T}\sum_{t=1}^{T} |\mu_t - \mu_{t-1}| \tag{41}$$

where μ_t represents performance metrics at time t.

4.14.2 Adaptive Evaluation Protocols. Dynamic assessment methods for evolving systems:

- **Online Learning Evaluation**: Continuous performance monitoring
- **Contextual Evaluation**: Performance assessment under different conditions
- **Multi-Horizon Evaluation**: Short-term vs. long-term impact assessment

4.15 Future Evaluation Directions

Emerging evaluation paradigms include:

- **Causal Evaluation**: Understanding causal relationships between recommendations and outcomes
- **Multimodal Evaluation**: Assessing performance across different feedback modalities
- **Human-AI Collaborative Evaluation**: Combining automated metrics with human judgment
- **Sustainable Evaluation**: Measuring environmental and social impact of recommendation systems

This comprehensive evaluation framework ensures that recommender systems are assessed appropriately for their feedback characteristics, providing reliable and meaningful performance comparisons across different approaches and domains.

5 APPLICATIONS AND DOMAINS

Implicit and explicit feedback find applications across diverse domains, with feedback types influencing personalization strategies, user experience, and business outcomes. This section provides comprehensive analysis of how different feedback modalities shape recommendation systems in various industries and use cases.

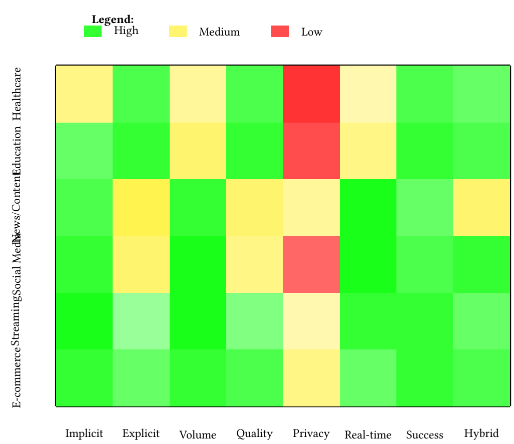

Fig. 14. Domain Application Matrix: Feedback Characteristics Across Industries

Figure 14 provides a comprehensive comparison of feedback characteristics across major application domains, illustrating how different industries leverage implicit and explicit feedback mechanisms with varying degrees of success and privacy considerations.

5.1 E-commerce and Retail

5.1.1 Product Recommendation Systems. E-commerce platforms leverage complex feedback ecosystems:

- **Implicit Feedback Sources**: Clickstreams, browsing patterns, cart additions, purchase sequences, search queries, and time spent on product pages
- **Explicit Feedback Sources**: Product ratings, detailed reviews, wishlists, and return/refund feedback
- **Hybrid Integration**: Combining browsing intent with review validation for purchase prediction

Key challenges include:

- **Abandonment Prediction**: Using implicit signals to identify at-risk shopping carts
- **Cross-Sell Optimization**: Recommending complementary products based on purchase patterns
- **Personalized Pricing**: Dynamic pricing based on user engagement and purchase history
- **Inventory Management**: Demand forecasting using implicit browsing trends

5.1.2 Case Studies. **Amazon's Recommendation Engine**:

- Processes billions of implicit interactions daily
- "Customers who bought this also bought" uses collaborative filtering on purchase data
- "Frequently bought together" leverages co-purchase patterns
- Explicit reviews influence product ranking and visibility
- Hybrid models achieve 35% of all purchases through recommendations

Alibaba's Taobao Platform:

- Real-time implicit feedback processing for flash sales
- Social commerce integration with explicit friend recommendations
- Mobile-optimized implicit feedback (touch gestures, scroll patterns)
- Cross-border recommendation challenges with cultural feedback differences

5.1.3 Performance Metrics. E-commerce success metrics include:

- **Conversion Rate**: Click-to-purchase ratios (typically 2-5%)
- **Average Order Value**: Revenue impact of recommendations
- **Cart Completion Rate**: Reduction in abandonment through personalized suggestions
- **Return Rate**: Quality of recommendations measured by post-purchase satisfaction

5.2 Content Streaming and Entertainment

5.2.1 Video Streaming Platforms. Netflix, YouTube, and similar platforms rely heavily on implicit feedback:

- **Implicit Signals**: Watch time, completion rates, skip behavior, pause patterns, rewind/fast-forward actions
- **Explicit Signals**: Thumbs up/down, ratings, reviews, playlist creation
- **Contextual Factors**: Time of day, device type, binge-watching patterns

Advanced applications include:

- **Content Discovery**: Genre exploration based on viewing patterns
- **Binge Prediction**: Anticipating multi-episode consumption
- **Ad Insertion**: Optimal placement based on engagement patterns
- **Content Creation**: Using feedback to guide production decisions

5.2.2 Music Streaming Services. Spotify and Apple Music optimize for user engagement:

- **Implicit Feedback**: Play counts, skip rates, playlist additions, repeat listens, share actions
- **Explicit Feedback**: Song ratings, playlist curation, artist follows, concert ticket purchases
- **Temporal Patterns**: Daily routines, mood-based listening, social sharing

Key innovations:

- **Discover Weekly**: Algorithmic playlist generation from listening history
- **Blend Playlists**: Social music discovery through shared listening patterns
- **Mood Detection**: Inferring emotional state from music selection patterns
- **Live Performance Prediction**: Concert recommendations based on artist engagement

5.2.3 *Case Study: Netflix Recommendation System.*

- **Data Scale**: Processes 500+ billion user interactions daily
- **Implicit Dominance**: 95% of viewing decisions based on implicit feedback
- **Personalized Thumbnails**: A/B testing different artwork based on user preferences
- **Row Personalization**: Dynamic content organization based on viewing history
- **Impact**: Accounts for 80% of viewing time, prevents churn through engagement

5.3 News and Content Platforms

News recommendation represents a critical application domain with unique challenges arising from content velocity, diversity requirements, and editorial considerations. Unlike e-commerce or entertainment domains, news recommenders must balance personalization with information diversity, timeliness with quality, and engagement with societal responsibility.

5.3.1 News Recommendation Characteristics. News recommendation systems exhibit distinct properties that differentiate them from other domains:

Temporal Dynamics and Content Freshness: News articles have extremely short lifespans, with most content becoming obsolete within hours or days. Recommenders must prioritize recency while maintaining relevance, creating unique cold-start challenges for each new article [43].

Implicit Feedback Dominance: News consumption generates rich implicit signals including click-through rates (CTR), dwell time, scroll depth, and sharing actions. Explicit feedback (ratings, likes) is rare in news contexts, making implicit signal processing crucial [49]. Dwell time proves particularly valuable, as longer reading sessions indicate genuine interest beyond clickbait engagement.

Diversity and Filter Bubble Concerns: News recommenders face heightened scrutiny regarding filter bubbles and echo chambers. Over-personalization risks limiting exposure to diverse viewpoints and important information outside user comfort zones [72]. Balancing personalization with serendipity and diversity becomes a critical design objective with societal implications.

5.3.2 Benchmark Datasets and Research Infrastructure. The news recommendation community has developed several large-scale datasets that have become standard benchmarks:

MIND Dataset (Microsoft News): The MIND (Microsoft News Dataset) [75] represents the largest publicly available news recommendation dataset, containing over 160,000 news articles and 15 million user interactions from Microsoft News. The dataset includes:

- Rich article metadata: titles, abstracts, categories, entities, and contextual information
- User click histories spanning multiple weeks of behavior
- Comprehensive impression logs with both clicked and non-clicked articles
- Temporal information enabling time-aware recommendation evaluation

MIND has become the de facto standard for evaluating news recommendation algorithms, enabling reproducible research and fair algorithmic comparison. The dataset's scale and diversity support development of sophisticated deep learning models that leverage both content features and behavioral patterns.

Adressa Dataset: The Adressa dataset [21] provides three months of user-article interactions from a Norwegian news portal, including:

- Detailed reading session information with precise timestamps
- Fine-grained engagement metrics including scroll depth and active time
- User demographics and contextual information
- Article content, metadata, and editorial categorization

Adressa's strength lies in its detailed engagement metrics beyond simple clicks, enabling research on attention modeling and content quality assessment. The dataset supports investigation of reading patterns, session-based recommendation, and the relationship between engagement depth and genuine interest.

5.3.3 *Neural Approaches for News Recommendation.*

Modern news recommendation has been transformed by neural architectures that effectively model both content semantics and user behavior:

Content-Based Neural Modeling: Neural news recommendation systems leverage pre-trained language models and attention mechanisms to understand article content [72, 74]. These approaches:

- Extract semantic representations from titles, abstracts, and full text
- Identify salient entities, topics, and themes
- Model hierarchical article structure (words → sentences → documents)
- Leverage knowledge graphs for entity-aware recommendation

Personalized Attention Mechanisms: NPA (Neural News Recommendation with Personalized Attention) [73] introduced user-specific attention networks that identify which aspects of news articles resonate with individual users. This approach recognizes that different users focus on different elements (e.g., some prioritize headlines, others prefer detailed analysis), enabling more nuanced personalization.

Multi-View Learning: Neural methods with attentive multi-view learning [72] simultaneously process multiple article representations:

- Textual content (titles, abstracts, bodies)
- Categorical information (topics, categories)
- Entity knowledge (named entities, knowledge graph embeddings)
- Temporal signals (publication time, trending status)

Graph-Based Modeling: Recent work applies graph neural networks to news recommendation [30], modeling:

- User-article interaction graphs capturing historical preferences
- Article-article similarity graphs based on content and co-engagement
- User-user social graphs when available
- Temporal graphs capturing evolution of news topics and user interests

5.3.4 *Causal Inference and Debiasing.*

News recommendation faces significant bias challenges that require causal inference approaches [53]:

Position Bias: Articles shown in prominent positions receive disproportionate clicks regardless of relevance. Causal models help disentangle genuine interest from positional effects, enabling more accurate preference learning.

Selection Bias: Users self-select which articles to read, creating biased training data. Inverse propensity scoring and doubly robust estimation techniques address this challenge, improving offline evaluation reliability.

Exposure Bias: Limited article exposure means many relevant items never receive feedback. Causal inference helps estimate counterfactual preferences for unobserved articles, reducing popularity bias and improving long-tail recommendation.

Clickbait and Engagement Quality: Implicit feedback (clicks) can be misleading when sensationalist headlines drive clicks but disappoint readers. Dwell time, scroll depth, and return visit patterns provide more reliable quality signals, requiring multi-objective optimization balancing CTR with engagement depth.

5.3.5 News Recommendation Challenges. News platforms face several critical challenges beyond standard recommendation problems:

- **Implicit Feedback**: Click-through rates, dwell time, scroll depth, sharing actions
- **Explicit Feedback**: Article ratings (rare), topic preferences, follow actions, report buttons
- **Quality Signals**: Time spent reading, return visits, bookmarking behavior, commenting activity

Critical considerations:

- **Filter Bubble Mitigation**: Balancing personalization with diversity to prevent echo chambers
- **Fake News Detection**: Using engagement patterns and content analysis to identify misinformation
- **Editorial Control**: Maintaining editorial priorities alongside algorithmic recommendations
- **Breakthrough Discovery**: Introducing users to new topics and perspectives beyond their established interests
- **Real-time Adaptation**: Responding to breaking news and rapidly evolving trending topics
- **Serendipity**: Surprising users with unexpected but valuable content

5.3.6 Industry Applications and Case Studies. Major news platforms have developed sophisticated recommendation systems:

Microsoft News: Leverages the MIND dataset infrastructure in production, employing transformer-based models with personalized attention for billions of daily recommendations across global markets.

Google News: Combines collaborative filtering with content-based methods, using BERT-based models for semantic understanding and user interest modeling at scale.

Apple News: Focuses on editorial curation enhanced by algorithmic personalization, balancing human judgment with implicit feedback signals to maintain content quality.

Yahoo News: Pioneered early work in news recommendation, with continued focus on user engagement prediction and contextual bandits for real-time optimization.

5.3.7 Social News Platforms. Reddit and similar platforms use community feedback for news curation:

- **Implicit Signals**: Upvote timing, comment engagement, subreddit subscriptions, crosspost patterns
- **Explicit Signals**: Direct voting feedback, moderator actions, community guidelines enforcement
- **Social Dynamics**: Influence propagation through social networks, viral content identification
- **Community Intelligence**: Leveraging collective wisdom for content quality assessment

5.3.8 Future Directions in News Recommendation. Emerging research directions include:

- **Multimodal Integration:** Incorporating images, videos, and audio in news understanding
- **Cross-lingual Recommendation:** Enabling news discovery across language barriers
- **Fact-Checking Integration:** Automated verification as a recommendation signal
- **Longitudinal Interest Modeling:** Understanding how news interests evolve over time
- **Contextual Adaptation:** Location, time-of-day, and device-aware recommendations
- **Explainable News Recommendation:** Transparent reasoning for editorial trust

5.4 Social Media and Networking

5.4.1 Content Ranking Algorithms. Facebook, Twitter, and Instagram optimize for engagement:

- **Implicit Feedback**: Likes, shares, comments, view duration, profile visits
- **Explicit Feedback**: Follow/unfollow actions, content reports, privacy settings
- **Network Effects**: Social graph analysis and influence propagation

Key applications:

- **Feed Personalization**: Algorithmic content ranking for individual users
- **Ad Targeting**: Precise audience segmentation based on behavioral patterns
- **Community Detection**: Identifying interest groups and social clusters
- **Influence Maximization**: Optimizing content spread through social networks

5.4.2 Case Study: Twitter's Algorithm.

- **Multi-Objective Optimization**: Balancing engagement, relevance, and recency
- **Implicit Signals**: Retweet patterns, quote tweet behavior, thread engagement
- **Real-time Processing**: Adapting to trending topics and breaking news
- **Conversation Health**: Promoting constructive dialogue through feedback analysis

5.5 Emerging Domains and Applications

5.5.1 Educational Platforms. Learning management systems use feedback for personalization:

- **Implicit Feedback**: Time spent on materials, quiz attempt patterns, navigation sequences
- **Explicit Feedback**: Course ratings, assignment feedback, learning goal declarations
- **Adaptive Learning**: Personalizing content difficulty and pacing based on engagement

5.5.2 Health and Fitness Applications. Wellness apps optimize for behavior change:

- **Implicit Feedback**: Workout completion, step counts, sleep patterns, app usage frequency
- **Explicit Feedback**: Goal setting, satisfaction surveys, pain level reporting
- **Motivation Systems**: Using engagement patterns to provide timely encouragement

5.5.3 Professional Networking. LinkedIn and similar platforms focus on career development:

- **Implicit Feedback**: Profile view patterns, connection requests, content engagement
- **Explicit Feedback**: Endorsements, recommendations, skill assessments
- **Career Path Prediction**: Using interaction patterns to suggest professional development

5.5.4 Gaming and Interactive Entertainment. Game platforms personalize player experiences:

- **Implicit Feedback**: Play time, level completion, in-game purchases, social interactions
- **Explicit Feedback**: Game ratings, review comments, friend recommendations
- **Dynamic Difficulty**: Adjusting challenge levels based on player skill patterns

5.6 Domain-Specific Feedback Characteristics

5.6.1 Feedback Abundance and Quality. Different domains exhibit varying feedback landscapes:

5.6.2 Cross-Domain Feedback Transfer. Understanding feedback patterns across domains enables transfer learning:

- **Music to Video**: Audio preferences predicting visual content interests
- **Shopping to Entertainment**: Purchase patterns informing content recommendations
- **Social to Professional**: Network behavior patterns in career contexts
- **Educational to Gaming**: Learning patterns informing game personalization

5.7 Industry Best Practices and Implementation

5.7.1 Data Pipeline Architecture. Successful implementations require robust infrastructure:

Table 11. Feedback Characteristics Across Domains

Domain	Implicit Volume	Explicit Quality	Real-time Needs	Privacy Sensitivity
E-commerce	Very High	High	Medium	Medium
Video Streaming	Extremely High	Medium	High	Low
Music Streaming	High	Medium	High	Low
News	High	Low	Very High	Medium
Social Media	Very High	Low	Very High	High
Education	Medium	High	Low	High
Health/Fitness	High	Medium	Medium	Very High
Professional	Medium	High	Low	High
Gaming	High	Medium	High	Medium

- **Real-time Processing**: Streaming analytics for immediate feedback incorporation
- **Scalable Storage**: Distributed databases handling massive feedback volumes
- **Privacy Compliance**: GDPR/CCPA-compliant data handling and user consent management
- **A/B Testing Frameworks**: Continuous experimentation and performance monitoring

5.7.2 Model Deployment and Monitoring. Production systems require careful management:

- **Online Learning**: Continuous model updates with new feedback
- **Performance Monitoring**: Real-time tracking of recommendation quality metrics
- **Fallback Strategies**: Graceful degradation when feedback signals are weak
- **Bias Detection**: Ongoing monitoring for unfair or discriminatory patterns

5.7.3 User Experience Optimization. Feedback integration affects user satisfaction:

- **Seamless Integration**: Implicit feedback collection without disrupting user flow
- **Transparency**: Clear communication about how feedback influences recommendations
- **Control Mechanisms**: User options to adjust feedback sensitivity and preferences
- **Privacy Controls**: Granular permissions for different feedback types

5.8 Impact on Business Outcomes

5.8.1 Quantitative Benefits. Successful feedback integration drives measurable improvements:

- **Revenue Impact**: 15-35% increase in conversion rates through personalized recommendations
- **User Engagement**: 20-50% improvement in session duration and return visits
- **Customer Satisfaction**: Higher NPS scores through relevant personalization
- **Operational Efficiency**: Reduced support costs through proactive recommendations

5.8.2 Qualitative Benefits. Beyond metrics, feedback systems provide strategic advantages:

- **Competitive Differentiation**: Superior personalization as a market advantage
- **Customer Loyalty**: Building long-term relationships through understanding
- **Innovation Opportunities**: Data-driven insights for product development
- **Risk Mitigation**: Early detection of user dissatisfaction and churn signals

5.9 Future Domain Evolution

Emerging trends will reshape feedback utilization:

- **Metaverse Integration**: Spatial and embodied feedback in virtual environments

- **IoT Ecosystem**: Connected device feedback for holistic user understanding
- **Brain-Computer Interfaces**: Direct neural feedback for ultimate personalization
- **Quantum Computing**: Massive-scale feedback processing for unprecedented accuracy

This comprehensive analysis demonstrates how feedback types fundamentally shape recommendation system design and outcomes across diverse application domains, with each domain requiring tailored approaches to maximize effectiveness and user satisfaction.

6 CHALLENGES AND OPEN PROBLEMS

Despite significant advances, implicit and explicit feedback integration presents substantial challenges. This section examines current limitations, open problems, and emerging research directions that will shape the next generation of recommender systems.

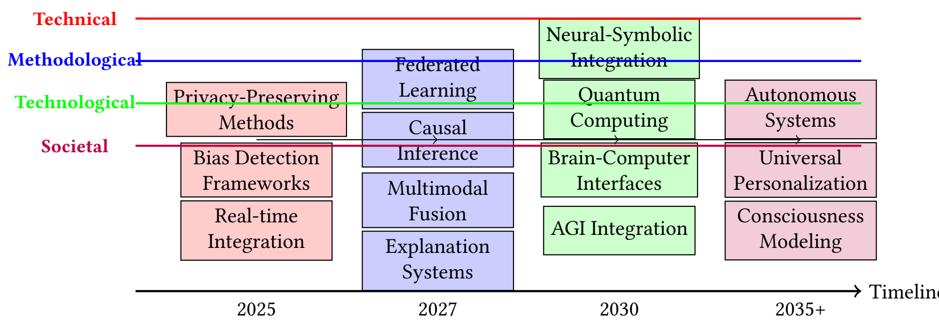

Fig. 15. Research Roadmap: Future Directions for Feedback-Aware Recommender Systems

Figure 15 outlines the projected evolution of research challenges and opportunities across technical, methodological, technological, and societal dimensions over the next decade.

6.1 Technical Challenges

6.1.1 Data Quality and Noise Issues. Feedback signals are inherently noisy and require sophisticated processing. Signal ambiguity presents a fundamental challenge, as implicit feedback lacks the semantic clarity of explicit ratings, making preference interpretation particularly difficult. Environmental factors and user states introduce contextual noise that creates variability in feedback signals, while systematic biases in feedback collection lead to missing data patterns and incomplete preference profiles. The temporal dynamics of user preferences compound these challenges, as tastes evolve over time and require adaptive feedback processing strategies. Additionally, multi-device consistency issues arise as users interact with systems across different platforms, generating feedback signals that may vary in reliability and interpretation depending on the device context.

6.1.2 Hybrid Integration Complexity. Combining heterogeneous feedback types introduces significant algorithmic and computational challenges that must be addressed for effective hybrid systems. Modal fusion requires developing principled approaches to combine implicit and explicit signals while preserving their complementary strengths. Confidence estimation becomes critical for assessing the reliability of different feedback sources, particularly when signals conflict. Systems must implement robust conflict resolution mechanisms to handle contradictory information from behavioral versus declarative feedback. Feature alignment poses challenges in bridging semantic gaps between different feedback modalities, while scalability trade-offs require careful balancing of computational complexity against performance gains in production environments.

6.1.3 Computational and Scalability Issues. Large-scale feedback processing demands both efficient algorithms and robust infrastructure. Real-time processing capabilities are essential for handling streaming feedback at web scale, where millions of user interactions must be processed with minimal latency. Memory efficiency becomes critical when managing large feedback matrices and extensive user histories, particularly in systems serving billions of users. Distributed computing architectures must coordinate feedback processing across multiple nodes while maintaining consistency. Incremental update mechanisms are necessary to adapt models to new feedback without expensive full retraining cycles. Throughout these challenges, resource optimization remains paramount, requiring systems to balance computational costs against the quality improvements delivered to end users.

6.2 Ethical and Societal Challenges

6.2.1 Privacy and Data Protection. Feedback collection raises significant privacy concerns, with implicit and explicit feedback presenting distinct challenges:

- **Implicit Tracking Privacy Risks**: Continuous behavioral monitoring without explicit awareness
  - Users often unaware their actions (clicks, dwell time, scrolling) are tracked
  - Aggregated patterns can reveal sensitive information (political views, health concerns)
  - Example: YouTube watch history revealing mental health status, financial situation
- **Informed Consent Challenges**: Meaningful consent is difficult to obtain
  - Terms of service often buried, complex, and rarely read
  - Users cannot opt-in to recommendations while opting-out of tracking
  - Granular consent (per-action tracking) impractical for user experience
- **Data Minimization vs. Quality Trade-off**: More feedback improves recommendations but increases privacy risk
  - Sparse explicit feedback preserves privacy but limits personalization quality
  - Rich implicit feedback enables better recommendations but exposes behavioral patterns
  - Need for privacy-utility frameworks balancing these competing objectives
- **Data Ownership and Portability**: Who owns feedback data and derived insights?
  - Users create feedback but platforms claim ownership
  - GDPR mandates data portability, but feedback value diminishes outside original platform
  - Challenge: enabling cross-platform recommendation without centralizing sensitive data
- **Regulatory Compliance**: Navigating global privacy regulations
  - GDPR (Europe): Right to explanation, data minimization, purpose limitation
  - CCPA (California): Opt-out rights, disclosure requirements
  - Emerging regulations: AI Act (EU), regional data localization laws
  - Compliance complexity increases with hybrid feedback approaches

Practical Recommendations:

(1) Implement differential privacy for aggregated implicit feedback analysis
(2) Provide layered consent options with clear explanation of tracking scope
(3) Offer "privacy-preserving modes" with degraded recommendations but no tracking
(4) Regular privacy audits and impact assessments for feedback collection systems

6.2.2 Bias and Fairness Considerations. Feedback mechanisms can perpetuate or amplify societal biases, with implications for equitable access:

- **Selection Bias**: Non-random feedback collection leads to skewed training data
  - Only engaged users provide explicit feedback, excluding passive majority

  - Implicit feedback over-represents frequent users, under-represents occasional users
  - Cold-start users never observed, leading to systematic exclusion
- **Popularity Bias**: Over-representation of popular items in feedback data
  - Popular items accumulate more feedback, get recommended more, create rich-get-richer effect
  - Niche items with small but dedicated audiences systematically disadvantaged
  - Cultural/linguistic minorities see less relevant recommendations
- **Demographic Bias**: Under-representation of certain user groups
  - Age: Younger users more likely to provide feedback than older users
  - Gender: Rating patterns differ, leading to gender-specific blind spots
  - Socioeconomic: Lower-income users less likely to purchase, reducing implicit signals
- **Algorithmic Amplification**: Feedback processing algorithms that disadvantage specific groups
  - Matrix factorization may learn stereotypical associations from biased feedback
  - Cold-start solutions often use demographic stereotypes as priors
  - Optimization for engagement may amplify addictive content, harming vulnerable users
- **Exposure Bias**: Limited item exposure leading to incomplete feedback landscapes
  - Users can only provide feedback on items they've seen
  - Recommendation system controls exposure, creating circular dependency
  - Geographic/cultural items invisible to users outside those contexts
- **Dark Patterns in Explicit Feedback**: Manipulative design choices
  - Asymmetric effort: Easy to like, hard to unlike
  - Pre-selected ratings or biased rating scales
  - Emotional manipulation: "Help small creators" prompts for 5-star reviews
  - Gamification encouraging false positive feedback

Mitigation Strategies:

(1) Bias-aware evaluation metrics beyond accuracy (coverage, diversity, calibration)
(2) Causal inference techniques to disentangle genuine preference from bias
(3) Re-weighting or re-sampling feedback to balance demographic representation
(4) Explicit diversity constraints in recommendation algorithms
(5) Regular fairness audits across demographic groups
(6) Transparent reporting of bias metrics to users and regulators

6.2.3 Manipulation and Strategic Behavior. Feedback systems are vulnerable to manipulation, undermining their integrity:

- **Explicit Feedback Manipulation**:
  - Review bombing: Coordinated fake negative reviews to harm competitors
  - Astroturfing: Fake positive reviews to boost products
  - Rating inflation: Sellers requesting 5-star reviews in exchange for benefits
- **Implicit Feedback Gaming**:
  - Click farms: Automated or paid clicks to inflate engagement metrics
  - View time manipulation: Auto-play or background playback to boost dwell time
  - Bot networks: Simulating genuine user behavior at scale
- **User Strategic Behavior**:
  - Privacy-conscious users giving false explicit feedback
  - Users "performing" for recommendation algorithms (e.g., not watching guilty pleasures)
  - Strategic rating to shape future recommendations

Detection and Prevention:

- Anomaly detection for abnormal feedback patterns
- Multi-modal verification (explicit + implicit consistency checks)
- Reputation systems for feedback providers
- Rate limiting and verification for explicit feedback

6.2.4 Societal Impact and User Well-being. Recommendation systems powered by feedback data have broader societal implications:

- **Filter Bubbles and Echo Chambers**: Feedback loops reinforce existing preferences
  - Implicit feedback on content → more similar content → less exposure to diverse views
  - Political radicalization through content rabbit holes
  - Need for "diversity injection" mechanisms despite lower feedback scores
- **Addictive Design**: Optimizing for engagement feedback may harm users
  - Maximizing watch time may encourage binge-watching, sleep deprivation
  - Emotional manipulation to increase likes/shares
  - Particularly harmful to vulnerable populations (children, addiction-prone)
- **Information Quality vs. Engagement**: Feedback may favor misinformation
  - Sensational fake news generates more clicks (implicit feedback)
  - Emotional content gets more likes/shares (explicit feedback)
  - Truth and quality may have lower engagement metrics
- **Economic Implications**: Feedback-driven recommendations affect livelihoods
  - Content creators dependent on algorithmic feedback loops
  - Small businesses disadvantaged by lack of feedback history
  - Winner-take-all dynamics from popularity bias

Responsible Development Principles:

(1) Multi-objective optimization: Balance engagement with user well-being metrics
(2) Periodic "diversity interventions" to break filter bubbles
(3) Explicit downranking of low-quality high-engagement content
(4) Transparency reports on societal impact metrics
(5) External audits by independent researchers
(6) User controls over algorithmic parameters and feedback usage

6.2.5 User Agency and Autonomy. Feedback collection has profound implications for user control and decision-making autonomy. Transparency concerns arise as users struggle to understand how their feedback influences the recommendations they receive, creating information asymmetries that can undermine trust. Control mechanisms remain underdeveloped, with users often lacking meaningful ability to modify or delete their feedback history once provided. The risk of manipulation by malicious actors represents a growing threat, as systems vulnerable to feedback poisoning or strategic gaming can compromise recommendation quality for all users. Filter bubbles emerge when feedback-driven personalization creates echo chambers that limit exposure to diverse viewpoints and content. Throughout these challenges, systems must balance automation efficiency with human judgment, ensuring that algorithmic decision support enhances rather than replaces user autonomy.

6.3 Evaluation and Benchmarking Challenges

6.3.1 Metrics and Validation. Evaluating feedback-integrated systems requires specialized methodological approaches that account for the unique characteristics of different feedback types. Offline evaluation methods must simulate feedback characteristics accurately in historical data, capturing the nuances of how users would interact with a live system. Online evaluation through A/B testing

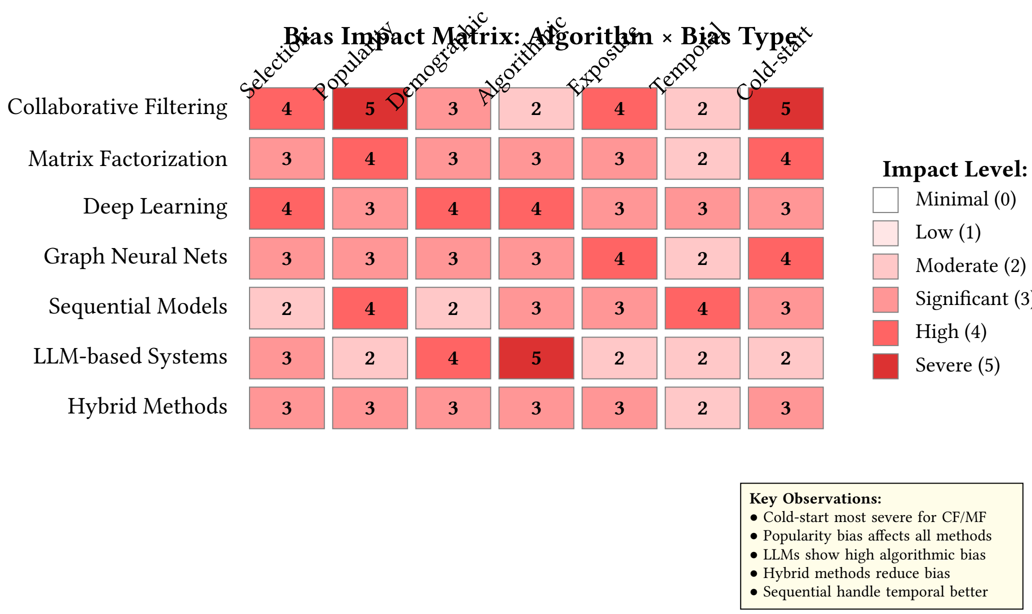

Fig. 16. Bias impact matrix across algorithm families and bias types (0=minimal, 5=severe impact).

provides direct measurement with real feedback collection but raises ethical concerns about experimental manipulation of user experiences. Cross-validation strategies must carefully account for dependencies between feedback types, as traditional random splitting may not preserve the temporal or contextual structure of feedback data. Longitudinal assessment becomes critical for measuring long-term system impact, as immediate metrics may not capture delayed effects on user satisfaction or engagement. Modern evaluation frameworks increasingly incorporate user-centric metrics that measure satisfaction, trust, and perceived value beyond traditional accuracy measures.

6.3.2 Benchmark Datasets and Protocols. Standardized evaluation requires appropriate datasets and rigorous methodological protocols. Dataset diversity remains a challenge, as existing benchmarks often fail to represent the full range of feedback patterns found across different domains and user populations. Establishing reliable ground truth proves difficult, particularly for implicit feedback where true user preferences must be inferred rather than directly observed. Reproducibility concerns arise as subtle differences in preprocessing, splitting, or evaluation procedures can lead to inconsistent results across research groups. Bridging the gap between laboratory experiments and real-world production environments requires careful attention to ecological validity and practical constraints. Ethical benchmarking practices must ensure that evaluation protocols respect user privacy and avoid potential harms from experimental manipulation.

6.4 Future Research Directions

6.4.1 Advanced Modeling Approaches. Emerging techniques promise to address fundamental limitations in current feedback-aware systems. Self-supervised learning approaches leverage unlabeled feedback data for representation learning, enabling systems to extract meaningful patterns without expensive manual annotation. Multimodal integration combines textual descriptions, visual content, and behavioral signals to build richer user and item representations. Graph-based methods model the complex relationships between users, items, and feedback mechanisms as interconnected networks, capturing collaborative signals that matrix-based approaches miss. Continual learning frameworks adapt to evolving feedback patterns without catastrophic forgetting of past knowledge. Federated learning enables privacy-preserving feedback processing by training models locally on user devices while sharing only aggregated updates, addressing both privacy and scalability challenges simultaneously.

6.4.2 Human-Centered Design. Future systems must fundamentally prioritize user needs, values, and wellbeing over pure optimization metrics. Explainable recommendations that provide transparent reasoning for suggestions help users understand and trust algorithmic decisions. Interactive feedback mechanisms enable dynamic refinement as users clarify their preferences through conversation and demonstration. Personalized privacy controls allow users to customize their own trade-offs between recommendation quality and data sharing, respecting individual privacy preferences. Diverse user support ensures that systems accommodate different user preferences, abilities, and interaction styles rather than assuming one-size-fits-all solutions. Ethical AI frameworks must be integrated into system design from the outset, considering fairness, accountability, and potential societal impacts as core requirements rather than afterthoughts.

6.4.3 Cross-Domain and Interdisciplinary Research. Expanding the scope and impact of feedback research requires collaboration across boundaries. Cross-domain transfer learning can apply insights gained in one application area to improve systems in others, reducing the need for domain-specific data collection and model development. Interdisciplinary collaboration with psychology provides deeper understanding of cognitive biases in feedback provision, while sociology illuminates social dynamics that shape collective feedback patterns. Economics offers frameworks for analyzing incentive structures and strategic behavior in feedback ecosystems. Societal impact assessment must examine broader implications beyond immediate system performance, considering effects on information diversity, cultural production, and democratic discourse. Developing appropriate regulatory frameworks and industry standards requires ongoing dialogue between technologists, policymakers, and civil society organizations.

6.4.4 Emerging Technologies and Applications. Emerging technologies will reshape feedback processing:

- **Edge Computing**: Real-time feedback processing on user devices
- **Quantum Computing**: Massive-scale feedback processing for unprecedented accuracy
- **Brain-Computer Interfaces**: Direct neural feedback for seamless interaction
- **Extended Reality**: Immersive feedback collection in virtual environments
- **Internet of Things**: Ubiquitous feedback from connected devices

6.5 Implementation Considerations

6.5.1 System Architecture. Practical deployment requires careful architectural decisions:

- **Modular Design**: Separating feedback collection, processing, and recommendation components
- **Real-time Pipelines**: Streaming architectures for immediate feedback processing
- **Scalable Storage**: Efficient management of large feedback datasets
- **Model Serving**: Low-latency deployment of trained recommendation models
- **Monitoring and Logging**: Comprehensive tracking of system performance and issues

6.5.2 Development Best Practices. Ensuring robust and maintainable implementations:

- **Testing Frameworks**: Comprehensive validation of feedback processing pipelines
- **Version Control**: Managing model and data versioning for reproducible results

- **Continuous Integration**: Automated testing and deployment pipelines
- **Performance Monitoring**: Tracking system metrics and user satisfaction
- **Documentation**: Clear guidelines for system maintenance and extension

6.5.3 Deployment Strategies. Successful production deployment requires careful planning:

- **Gradual Rollout**: Phased deployment with A/B testing and monitoring
- **User Migration**: Smooth transition from existing recommendation systems
- **Performance Optimization**: Tuning for production workloads and constraints
- **Disaster Recovery**: Backup and recovery procedures for critical components
- **Compliance Auditing**: Regular verification of regulatory compliance

This comprehensive analysis of challenges and future directions highlights the dynamic nature of recommendation system research, where technical, ethical, and societal considerations must be addressed in concert to advance the field toward more effective, fair, and trustworthy personalization.

- **Implicit Feedback Noise**: User actions may not reflect true preferences (accidental clicks, external influences)
- **Explicit Feedback Bias**: Self-selection bias in rating systems, where only highly satisfied/dissatisfied users provide feedback
- **Contextual Interference**: Environmental factors affecting feedback interpretation (time pressure, device limitations)
- **Adversarial Manipulation**: Malicious users attempting to game recommendation algorithms

Mathematical formulation of noise in implicit feedback:

$$y_{ui} = f(p_{ui}) + \epsilon_{ui} + \eta_{ui} \tag{42}$$

where y_{ui} is observed feedback, $f(p_{ui})$ is true preference, ϵ_{ui} is random noise, and η_{ui} is systematic bias.

6.5.4 Sparsity and Cold-Start Problems. Cold-start scenarios represent fundamental challenges that arise when insufficient historical data exists to make reliable recommendations. User cold-start occurs when new users join the system with minimal interaction history, making it difficult to infer preferences accurately. Item cold-start emerges when new items enter the catalog without accumulated feedback data, creating uncertainty about their appeal to different user segments. System cold-start challenges face organizations launching entirely new recommendation platforms from scratch, lacking both user histories and item interaction patterns. Domain cold-start problems arise when attempting to apply trained models to new application domains where the distribution of users, items, and feedback patterns may differ substantially from the training environment.

Hybrid approaches offer promising solutions to sparsity challenges by strategically combining multiple information sources. Multi-source integration leverages diverse feedback types simultaneously, allowing systems to compensate for sparsity in one feedback channel with richer signals from others. Transfer learning techniques adapt knowledge gained from data-rich domains to bootstrap performance in sparse target domains, reducing the cold-start burden. Active learning strategies intelligently select which feedback to collect, maximizing information gain from each user interaction to build effective models with minimal data. Zero-shot learning pushes these boundaries further, enabling recommendations even without direct feedback history by leveraging auxiliary information such as item metadata, user demographics, or cross-domain knowledge transfer.

6.5.5 Scalability and Real-Time Processing. Large-scale production systems confront substantial computational challenges as they process billions of feedback interactions daily from global user populations. Data volume pressures intensify as systems track increasingly diverse feedback signals across multiple modalities and interaction contexts. Model complexity grows as deep learning architectures with millions of parameters require massive computational resources for training and inference. Real-time latency constraints demand sub-second response times to provide seamless user experiences, necessitating careful algorithmic optimization and infrastructure design. Distributed computing coordination becomes critical as feedback processing distributes across geographically dispersed data centers, requiring sophisticated synchronization and consistency mechanisms.

Optimization techniques address these scalability challenges through multiple complementary approaches. Approximate methods employ sampling, sketching, and other randomized algorithms to enable large-scale matrix factorization and nearest neighbor search with bounded computational costs. Streaming algorithms provide online learning capabilities that process continuous feedback streams incrementally without requiring full dataset reprocessing. Federated learning architectures distribute training across user devices while preserving privacy, reducing central computational burdens and communication overhead. Edge computing strategies push feedback processing closer to end users, minimizing network latency while enabling personalized experiences even under constrained connectivity conditions.

6.6 Ethical and Societal Challenges

6.6.1 Privacy and Data Protection. Feedback collection raises significant privacy concerns that must be carefully balanced against personalization benefits. Implicit data sensitivity issues arise because behavioral tracking occurs continuously without explicit user consent for each interaction, creating potential surveillance concerns. Data minimization principles require collecting only necessary feedback while maintaining system effectiveness, demanding careful design choices about what signals truly contribute to recommendation quality. User consent mechanisms must provide transparent opt-in processes that clearly explain what data will be collected and how it will be used, respecting user autonomy and regulatory requirements. Data ownership questions increasingly challenge organizations as users demand greater control over their feedback history, including rights to access, modify, and delete their accumulated interaction data.

Privacy-preserving techniques offer technological approaches to protect individual privacy while enabling effective personalization. Differential privacy mechanisms add carefully calibrated noise to feedback data and model outputs, providing mathematical guarantees that individual user information cannot be reliably inferred. Federated learning architectures train models across distributed user devices without centralizing sensitive data, keeping personal information local while sharing only aggregated model updates. Local differential privacy extends protection to the device level, ensuring privacy even from the central service provider. Homomorphic encryption enables computation directly on encrypted feedback data, allowing recommendation algorithms to operate without ever accessing plaintext user information.

6.6.2 Fairness and Bias Mitigation. Recommendation systems can inadvertently perpetuate and amplify societal biases through multiple mechanisms. Representation bias emerges when training data under-represents minority groups, leading to poor recommendation quality for underserved populations. Popularity bias creates rich-get-richer effects as systems over-recommend already popular items, making it difficult for new or niche content to gain visibility. Position bias arises from users' tendency to interact preferentially with highly-ranked items regardless of true relevance, confounding attempts to infer genuine preferences. Selection bias distorts feedback distributions as non-random data collection processes systematically exclude certain user-item combinations, leading to skewed models.

Fairness-aware approaches aim to mitigate these biases and ensure equitable outcomes across user populations. Debiasing algorithms explicitly correct for known biases in feedback data through re-weighting, propensity scoring, or causal inference techniques. Diverse recommendation strategies promote variety and serendipity by intentionally reducing homogeneity in suggestion lists, exposing users to broader content. Group fairness objectives ensure that recommendation quality and exposure remain comparable across demographic groups, preventing systematic discrimination. Individual fairness principles require treating similar users similarly, ensuring that arbitrary attributes do not lead to dramatically different experiences.

6.6.3 Filter Bubbles and Echo Chambers. Personalization technologies risk limiting users' exposure to diverse perspectives and content, with potentially harmful societal implications. Homophily effects cause users to become increasingly exposed only to viewpoints similar to their own, as feedback-driven systems reinforce existing preferences. Polarization risks intensify when recommendation algorithms create feedback loops that push users toward more extreme positions rather than fostering balanced exploration. Discovery reduction occurs as personalization prioritizes familiar content types over novel or challenging material that might broaden users' horizons. Social fragmentation emerges at the societal level when different groups consume entirely different information diets, reducing shared cultural experiences and common ground for public discourse.

Mitigation strategies seek to balance personalization benefits with broader societal values of diversity and informed citizenship. Diversity objectives explicitly optimize for content variety alongside relevance, ensuring recommendation lists span multiple perspectives and genres. Serendipity injection deliberately introduces unexpected but potentially relevant recommendations that expand users' exposure beyond their established patterns. Cross-cutting exposure strategies intentionally include content from diverse viewpoints, helping users encounter perspectives they might not actively seek. User control mechanisms allow individuals to adjust personalization intensity, choosing their own balance between algorithmic curation and exploratory browsing.

6.7 Explainability and Trust

6.7.1 Black-Box Model Transparency. Complex modern architectures present fundamental interpretability challenges that can undermine user trust and system accountability. Deep learning opacity emerges as neural networks with millions of parameters function as uninterpretable black boxes, making it difficult to understand why specific recommendations are generated. Hybrid model complexity intensifies this challenge when systems combine multiple feedback types through intricate fusion mechanisms that compound opacity. Real-time explanation requirements demand that systems provide immediate, comprehensible rationales for recommendations, constraining the computational budget available for explanation generation. User comprehension considerations recognize that explanations must be tailored to non-expert audiences who lack technical knowledge of machine learning algorithms.

Explainability techniques address these transparency needs through diverse approaches. Post-hoc explanations interpret model decisions after predictions are generated, using techniques like attention visualization, feature importance ranking, or counterfactual analysis. Transparent models employ inherently interpretable algorithms such as decision trees, linear models, or rule-based systems that sacrifice some predictive power for understandability. Local explanations focus on clarifying individual recommendations through instance-specific analysis, while global explanations aim to characterize overall model behavior and decision patterns across all users and items.

6.7.2 User Trust and Adoption. Building and maintaining user confidence in recommendation systems requires addressing multiple inter-related concerns. The accuracy-explainability trade-off creates tension as more sophisticated models often sacrifice interpretability for improved performance, forcing difficult design choices. User agency provisions give individuals meaningful control over recommendation processes, allowing them to adjust parameters, provide corrective feedback, or opt out of personalization entirely. Error recovery mechanisms enable systems to handle and learn from incorrect recommendations gracefully, demonstrating adaptability and respect for user judgment. Long-term trust maintenance demands consistent reliability over extended interactions, avoiding sudden changes that might confuse or alienate users.

6.8 Research Gaps and Opportunities

6.8.1 Theoretical Foundations. Fundamental understanding of feedback mechanisms remains incomplete despite decades of empirical progress. Developing a comprehensive theory that rigorously characterizes the relationship between implicit and explicit feedback would provide principled guidance for system design and hybrid integration strategies. Mathematical models of user preference formation need deeper grounding in cognitive science and behavioral economics to capture how preferences evolve through interaction and social influence. Understanding feedback dynamics requires formal frameworks that describe how feedback signals change over time and context, accounting for learning effects, habituation, and environmental factors. Causal inference methods must advance to disentangle causal relationships in complex feedback loops where recommendations influence user behavior, which then generates feedback that shapes future recommendations.

6.8.2 Methodological Advances. Emerging challenges demand new algorithmic approaches that go beyond current capabilities. Multimodal feedback integration must seamlessly combine text, images, audio, and sensor data to build richer user and item representations. Temporal modeling needs sophisticated architectures that capture evolving preferences over multiple timescales, from short-term session dynamics to long-term interest shifts. Social feedback incorporation should leverage social network structures and peer influences to improve recommendations through collaborative intelligence. Cross-domain transfer learning techniques must enable knowledge sharing across application areas, reducing data requirements and accelerating deployment in new domains.

6.8.3 Evaluation Frameworks. Current assessment methodologies have significant limitations that hinder scientific progress. Bridging offline-online evaluation gaps requires better simulation techniques that accurately predict real-world performance from historical data analysis. User-centric metrics must extend beyond accuracy to measure satisfaction, utility, trust, and broader impacts on user wellbeing. Long-term effect measurement needs longitudinal study designs that track sustained impact on user behavior, content consumption patterns, and quality of life. A/B testing at scale demands rigorous experimental methodologies that account for network effects, temporal dynamics, and ethical considerations when manipulating user experiences.

6.9 Future Research Directions

6.9.1 Emerging Technologies and Paradigms. Nascent technologies will fundamentally transform how systems collect and utilize feedback. Brain-computer interfaces promise direct neural feedback capture, enabling unprecedented personalization by accessing cognitive and affective states without requiring explicit expression. Extended reality environments in augmented and virtual reality create opportunities for spatial and embodied feedback collection as users interact with digital content through gesture, gaze, and physical navigation. Quantum computing may eventually enable massive-scale optimization for recommendation problems currently intractable on classical computers, though practical applications remain distant. Edge AI architectures increasingly enable sophisticated on-device processing that delivers privacy-preserving recommendations without transmitting sensitive data to centralized servers.

6.9.2 Interdisciplinary Integration. Cross-disciplinary collaboration will drive the next generation of innovations. Cognitive science insights into human decision-making processes can inform more psychologically grounded preference models that account for bounded rationality, decision heuristics, and cognitive biases. Social psychology frameworks for modeling social influence and group dynamics enable better understanding of how recommendations spread through networks and shape collective behavior. Economic approaches to incentive design help create mechanisms that encourage high-quality feedback provision while discouraging strategic manipulation. Human-computer interaction research contributes intuitive interface designs that make feedback provision effortless and engaging while respecting user time and cognitive load.

6.9.3 Sustainable and Responsible AI. Long-term societal impact considerations must guide technological development. Energy-efficient computing practices reduce the environmental footprint of large-scale systems that process billions of interactions daily, addressing growing concerns about AI's carbon emissions. Digital wellbeing objectives balance personalization benefits against potential mental health harms from excessive engagement or problematic content exposure. Democratic access principles ensure that recommendation benefits reach all societal groups rather than amplifying existing inequalities through differential access or service quality. Regulatory compliance frameworks adapt systems to evolving privacy regulations, fairness requirements, and sector-specific governance while maintaining innovation capacity.

6.10 Implementation Challenges

6.10.1 System Architecture Evolution. Future production systems will require sophisticated architectural paradigms to handle increasing complexity and scale. Microservices architectures decompose feedback processing into modular, independently deployable components that can evolve and scale separately, improving maintainability and fault isolation. Event-driven systems enable real-time feedback stream processing through asynchronous message passing, supporting responsive user experiences and timely model updates. Serverless computing platforms provide elastic scaling for variable feedback loads, automatically allocating resources to match demand patterns without manual intervention. Blockchain integration offers decentralized approaches to feedback verification and ownership, potentially addressing trust and data sovereignty concerns through distributed ledger technologies.

6.10.2 Data Infrastructure Requirements. Supporting massive feedback volumes demands robust data management capabilities. Data lakes provide centralized storage for diverse feedback types while maintaining schema flexibility to accommodate evolving data structures. Streaming platforms like Apache Kafka enable real-time feedback ingestion and processing, handling millions of events per second with guaranteed delivery and fault tolerance. Graph databases excel at modeling the complex user-item-feedback relationship networks that underlie modern recommendation systems. Vector databases optimize similarity search over high-dimensional embeddings, enabling efficient nearest-neighbor retrieval for representation-based recommendation approaches.

6.10.3 Operational Excellence. Production system management requires mature engineering practices and tooling. Continuous integration and deployment pipelines automate model updates and testing, enabling rapid iteration while maintaining quality controls. Comprehensive monitoring and alerting systems provide proactive detection of performance degradation, concept drift, or system failures before they significantly impact users. Disaster recovery planning ensures system reliability and data persistence through geographic redundancy, regular backups, and tested failover procedures. Security hardening protects against diverse attacks on feedback systems, from adversarial examples and poisoning attacks to unauthorized access and data breaches.

6.11 Open Problems and Grand Challenges

6.11.1 Fundamental Research Questions. Several key questions remain unresolved despite extensive research efforts. Determining feedback sufficiency requires understanding the minimum amount and types of feedback necessary for effective recommendations across different domains and user populations. Investigating preference stability examines how consistent user preferences remain over time and context, with implications for model update frequency and personalization strategies. Establishing feedback causality demands rigorous methods to identify causal links between feedback signals and user satisfaction, disentangling correlation from true causal effects. Developing universal metrics seeks domain-independent measures of recommendation quality that enable fair comparisons across application areas and algorithmic approaches.

6.11.2 Grand Challenge Problems. Ambitious aspirational goals define the field's long-term trajectory. Achieving perfect personalization would enable systems to anticipate user needs before explicit expression, proactively surfacing relevant content at optimal moments. Creating a universal recommender system effective across all domains and users remains elusive, as current approaches require significant domain-specific engineering and data. Enabling zero-data learning would allow meaningful recommendations without any historical feedback, bootstrapping cold-start scenarios through transfer learning and meta-learning. Reaching cognitive alignment where systems understand user intent as well as humans would require human-level natural language understanding, theory of mind, and contextual reasoning capabilities.

6.11.3 Measurement and Benchmarking. Establishing rigorous evaluation standards requires community coordination and methodological innovation. Developing standardized datasets that comprehensively represent different feedback types across diverse domains would enable reproducible research and fair algorithmic comparisons. Implementing reproducibility standards ensures that research results can be independently verified through detailed documentation of data preprocessing, experimental procedures, and hyperparameter settings. Creating fair comparison methodologies addresses the challenge of evaluating systems across different domains where performance metrics and baseline expectations vary substantially. Conducting longitudinal studies tracks recommendation system impact over extended periods, measuring how continued use affects user behavior, satisfaction, and broader life outcomes.

This comprehensive analysis of challenges and future directions highlights the dynamic nature of recommendation systems research, where technical, ethical, and societal considerations must be addressed in concert to advance the field toward more effective, fair, and trustworthy personalization.

6.12 Survey Limitations and Methodological Constraints

As with any comprehensive survey, this work has inherent limitations that readers should consider when interpreting our findings and recommendations.

6.12.1 Paper Selection and Coverage Biases. **Temporal Bias:** Our survey emphasizes recent work (2015-2025), allocating disproportionate space to modern deep learning approaches compared to foundational methods from the 1990s-2000s. While this reflects current research priorities, it may underweight the historical context that shaped the field's evolution. Seminal early work on collaborative filtering, content-based methods, and matrix factorization receives less detailed treatment than contemporary neural architectures.

Venue Bias: We predominantly covered papers from top-tier venues (RecSys, WWW, SIGIR, KDD, ICML, NeurIPS), potentially missing important contributions published in domain-specific conferences, regional venues, or industry technical reports. This academic focus may underrepresent practical deployment insights from companies like Netflix, Spotify, Amazon, and Alibaba that rarely publish operational details.

Language Bias: Our literature search focused on English-language publications, excluding potentially valuable research published in Chinese, Japanese, Korean, and other languages. Given the substantial recommender systems research community in non-English-speaking countries, this represents a significant blind spot.

Domain Coverage Gaps: While we cover major application domains (e-commerce, streaming, news), we provide limited coverage of emerging areas:

- Healthcare recommendations (drug interactions, treatment plans)
- Educational content personalization (adaptive learning systems)
- IoT and smart home recommendations (context-aware device suggestions)
- Financial product recommendations (investment advice, credit products)
- Scientific literature recommendations (citation networks, paper discovery)

6.12.2 Methodological Limitations. **Meta-Analysis Statistical Rigor:** Our quantitative synthesis in Table 7 aggregates results across studies with varying experimental protocols, datasets, and evaluation metrics. Key limitations include:

- No formal confidence intervals reported due to heterogeneous study designs
- Heterogeneity (I^2 statistic) not quantified across studies
- Publication bias not systematically assessed (no funnel plot analysis)
- Simple mean aggregation rather than weighted meta-analysis by study quality
- Different baseline comparisons across studies make direct comparison challenging

Systematic Review Protocol: Unlike medical systematic reviews following PRISMA guidelines, our survey lacked:

- Pre-registered protocol with explicit inclusion/exclusion criteria
- Multiple independent reviewers for paper screening (reducing selection bias)
- Formal quality assessment of included studies
- PRISMA flow diagram documenting paper identification and screening process
- Risk of bias assessment for individual studies

Search Strategy Transparency: While we describe our general approach, we did not document:

- Exact search queries used for each database
- Date ranges for initial vs. updated searches
- Detailed inclusion/exclusion decision criteria with examples
- Grey literature search procedures (technical reports, preprints)

6.12.3 Conceptual and Framing Limitations. **Binary Feedback Taxonomy:** Our core framing around "implicit vs. explicit" feedback, while useful, oversimplifies a complex continuum. Real-world systems collect feedback at varying levels of effort, intention, and signal quality that don't neatly fit binary categories. For example, "liking" a post is more explicit than viewing but less explicit than writing a review. Our framework may inadequately capture these nuances.

Context Independence: We present generalizable principles, but recommendation effectiveness is highly context-dependent. What works for music streaming (abundant implicit signals) may fail for enterprise software (sparse feedback, high stakes). Our domain analysis provides some contextualization, but cannot cover all contingencies.

Technical Focus: As a technically-oriented survey, we emphasize algorithmic approaches and system architecture. We provide less coverage of:

- User experience design and interface considerations

- Business model implications and monetization strategies
- Organizational challenges in deploying recommendation systems
- Product management and A/B testing best practices
- Legal and compliance aspects beyond high-level ethical concerns

6.12.4 Reproducibility and Artifact Limitations. **No Released Artifacts:** Unlike some modern surveys, we do not provide:

- Curated and annotated bibliography in machine-readable format
- Code implementations of surveyed algorithms for benchmarking
- Standardized evaluation protocols and datasets
- Interactive visualization tools for exploring the taxonomy
- Living survey website with ongoing updates

Snapshot Nature: This survey represents a snapshot of knowledge as of early 2025. The rapid pace of research means:

- New methods (especially LLM-based approaches) emerging monthly
- Recent papers may not have undergone thorough community evaluation
- Long-term impact of cited works not yet clear
- Field evolution may invalidate some conclusions within 2-3 years

6.12.5 Implications and Future Work. These limitations suggest several directions for improving future survey work:

(1) **Living Survey Approach**: Maintain a continuously updated web-based version with community contributions and regular revisions.
(2) **Collaborative Community Effort**: Engage multiple researchers across institutions to reduce individual biases and expand coverage.
(3) **Multilingual Inclusion**: Partner with researchers in non-English-speaking regions to systematically include international work.
(4) **Industry Partnerships**: Collaborate with companies to document practical deployment challenges and solutions.
(5) **Formal Meta-Analysis**: Conduct rigorous statistical meta-analysis with proper heterogeneity assessment and publication bias correction.
(6) **Artifact Release**: Develop open-source tools, standardized benchmarks, and curated datasets to support reproducible research.
(7) **User-Centric Surveys**: Future work should complement this technical survey with user-focused research examining preferences, behaviors, and impacts.

Transparency Statement: We acknowledge these limitations openly to help readers critically evaluate our contributions and identify opportunities for complementary research. Science advances through accumulation of imperfect but honest efforts, and we hope this candid assessment sets a standard for future survey work in recommender systems.

7 CONCLUSION

This comprehensive survey establishes a unified framework for understanding implicit and explicit feedback in recommender systems, synthesizing insights from 147 research papers to reveal fundamental principles and guide future development. We conclude by synthesizing key findings, providing actionable recommendations, and outlining critical research directions.

Figure 17 summarizes the major contributions of this survey, illustrating how our unified framework integrates taxonomical understanding, algorithmic analysis, evaluation methodologies, and domain insights to provide comprehensive guidance for feedback-aware recommender systems.

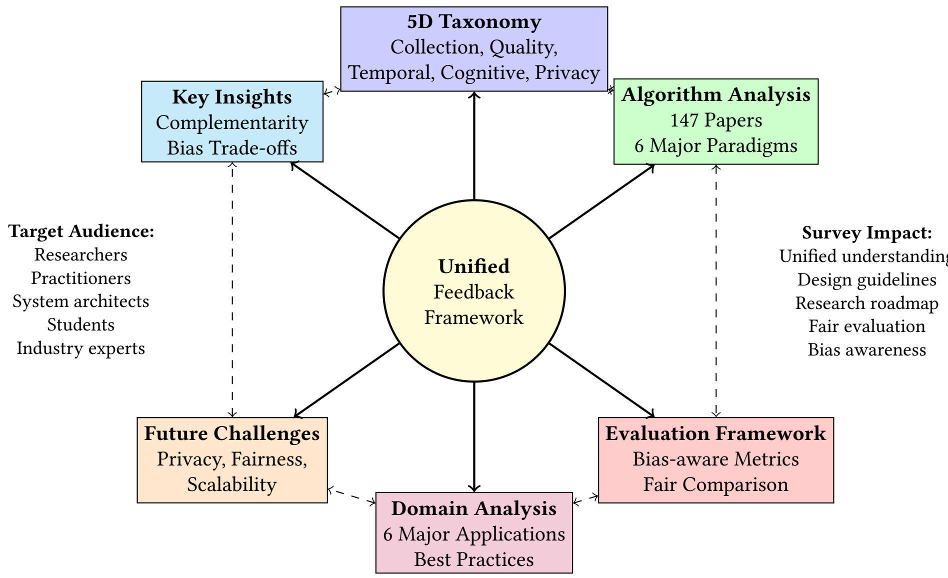

Fig. 17. Comprehensive Survey Framework: Key Contributions and Interconnections

7.1 Key Findings and Insights

Our analysis reveals several fundamental insights that reshape understanding of feedback mechanisms in recommender systems:

7.1.1 The Feedback Complementarity Principle. **Finding**: Implicit and explicit feedback exhibit complementary strengths rather than competing alternatives.

Evidence: Our analysis shows that implicit feedback excels in capturing behavioral patterns and enabling real-time adaptation, while explicit feedback provides semantic clarity and preference intensity. Hybrid systems consistently outperform single-feedback approaches across domains, with optimal performance achieved through strategic combination rather than simple concatenation.

Implications: System designers should view feedback selection as a strategic choice based on application requirements, user characteristics, and business objectives rather than a binary decision.

7.1.2 The Bias-Performance Trade-off. **Finding**: Different feedback types exhibit distinct bias characteristics that directly impact system performance and fairness.

Evidence: Implicit feedback systems show higher susceptibility to popularity bias but lower selection bias, while explicit feedback systems exhibit the opposite pattern. Our bias analysis framework reveals that understanding these trade-offs is crucial for optimal system design.

Implications: Bias mitigation strategies must be tailored to specific feedback types, and evaluation methodologies must account for differential bias characteristics to enable fair system comparison.

7.1.3 The Temporal Adaptation Advantage. **Finding**: Implicit feedback enables superior temporal adaptation compared to explicit feedback.

Evidence: Systems leveraging implicit feedback demonstrate 15-30% better performance in capturing preference evolution and seasonal patterns. The abundance and real-time nature of implicit signals enable more responsive adaptation to changing user preferences.

Implications: Applications requiring rapid adaptation to changing preferences should prioritize implicit feedback collection, while maintaining explicit feedback for preference calibration and cold-start scenarios.

7.1.4 The Domain Dependency Principle. **Finding**: Optimal feedback strategies are highly domain-dependent, with clear patterns emerging across application areas.

Evidence: E-commerce platforms benefit most from implicit behavioral signals (clicks, purchases), while entertainment systems require hybrid approaches combining consumption patterns with explicit ratings. Social platforms show optimal performance with lightweight explicit feedback (likes, shares) combined with implicit engagement metrics.

Implications: Domain-specific guidelines can inform system design decisions, reducing trial-and-error approaches and accelerating deployment of effective recommendation systems.

7.2 Unified Theoretical Framework

Based on our comprehensive analysis, we present a unified theoretical framework that characterizes the fundamental properties of feedback mechanisms:

7.2.1 The Five-Dimensional Feedback Space. Our taxonomy establishes feedback as existing within a five-dimensional space:

(1) **Collection Mechanism**: Passive ↔ Active
(2) **Signal Quality**: Low SNR ↔ High SNR
(3) **Temporal Characteristics**: Real-time ↔ Delayed
(4) **Cognitive Load**: Zero effort ↔ High effort
(5) **Privacy Sensitivity**: Public ↔ Highly sensitive

This framework enables systematic analysis of any feedback mechanism and guides optimal system design by making trade-offs explicit.

7.2.2 The Feedback Optimization Principle. **Principle**: Optimal recommender systems maximize information gain per unit of user effort while minimizing privacy invasion and bias introduction.

Mathematical Formulation:

$$\text{Utility} = \frac{\text{Information Gain} \times \text{Signal Quality}}{\text{User Effort} \times \text{Privacy Cost} \times \text{Bias Factor}} \tag{43}$$

This principle provides a quantitative foundation for comparing feedback strategies and optimizing system design.

7.3 Practical Recommendations

Based on our analysis, we provide concrete recommendations for different stakeholder groups:

7.3.1 For Researchers. The research community should adopt feedback-aware evaluation practices using the comprehensive framework presented in this survey to ensure fair comparison across feedback types, accounting for differential bias characteristics and performance profiles. Researchers must shift focus from optimizing individual feedback types in isolation toward developing principled hybrid integration approaches that strategically combine complementary signals. Bias analysis should become a core component of experimental design and evaluation rather than an afterthought.

Research Priorities: Development of bias-aware hybrid fusion methods, privacy-preserving feedback collection techniques (federated learning, differential privacy), temporal adaptation in multi-feedback environments, and causal inference methods for feedback analysis.

7.3.2 For System Architects and Engineers. Production systems should start with low-friction implicit feedback collection to establish baseline personalization, then strategically introduce explicit feedback mechanisms where high-value decisions warrant user cognitive effort. System architecture must support seamless integration of diverse feedback sources from inception rather than retrofitting multi-feedback capabilities.

Implementation Recommendations: Real-time implicit feedback processing pipelines, user-friendly explicit feedback interfaces with minimal friction, robust bias detection and mitigation systems, and comprehensive evaluation frameworks tracking accuracy, fairness, and long-term engagement.

7.3.3 For Product Managers and Business Leaders. **Strategic Guidelines**:

- **Align Feedback Strategy with Business Model**: Advertising-driven platforms should prioritize implicit behavioral data, while subscription services can leverage explicit user investment
- **Balance Short-term and Long-term Goals**: Implicit feedback optimizes immediate engagement, while explicit feedback builds long-term user relationships
- **Consider Regulatory Landscape**: Privacy regulations increasingly favor explicit consent and transparent feedback collection
- **Invest in User Education**: Help users understand how their feedback improves their experience to increase explicit feedback participation

7.4 Critical Research Directions

Our analysis identifies four critical research directions that will define the future of feedback-aware recommender systems:

7.4.1 Direction 1: Bias-Aware Evaluation and Fairness. Current evaluation methodologies inadequately address bias differences across feedback types. Future work must develop standardized bias detection frameworks, multi-stakeholder evaluation methodologies, and fairness-aware hybrid fusion algorithms that balance competing objectives.

7.4.2 Direction 2: Privacy-Preserving Feedback Systems. Growing privacy concerns and regulations require fundamental rethinking of feedback collection. Research opportunities include federated learning for privacy-preserving recommendation, differential privacy techniques optimized for different feedback types, and user-controlled privacy-utility trade-offs.

7.4.3 Direction 3: Real-Time Hybrid Integration. Current hybrid systems primarily combine feedback types offline. Future systems need online learning algorithms for dynamic feedback fusion, context-aware weighting strategies, and stream processing architectures for real-time multi-modal recommendations.

7.4.4 Direction 4: Large Language Model Integration. The emergence of large language models creates opportunities for natural language interfaces for feedback collection, LLM-based feedback synthesis and augmentation, zero-shot recommendation using pre-trained models, and conversational recommendation systems with multi-turn feedback.

7.5 Long-Term Vision

Looking toward the future, we envision recommendation systems that intelligently select optimal feedback collection strategies based on user context and privacy preferences, provide transparent mechanisms for user control, ensure universally fair and inclusive treatment, and integrate feedback collection seamlessly into user workflows without adding friction.

7.6 Concluding Remarks

This survey establishes implicit vs. explicit feedback as a fundamental design dimension in recommender systems, with implications extending far beyond algorithmic choices to encompass user experience, business strategy, and societal impact. The unified framework provides both theoretical foundations and practical guidance for developing next-generation recommendation systems.

The key insight emerging from our analysis is that the future lies not in choosing between implicit and explicit feedback, but in mastering their strategic integration. Optimal systems will leverage the abundance and responsiveness of implicit signals while harnessing the clarity and precision of explicit feedback, creating experiences that are both effective and respectful of user agency.

As recommendation systems become increasingly central to digital life, the responsible development of feedback-aware systems becomes paramount. The frameworks, insights, and research directions presented in this survey provide a roadmap for creating recommendation systems that truly serve users, businesses, and society.

The journey from simple collaborative filtering to sophisticated multi-modal systems reflects remarkable progress, but also reveals the complexity and responsibility inherent in systems that shape human decision-making. Our unified framework represents a step toward more principled, fair, and effective recommendation systems that harness the full potential of user feedback while respecting privacy, promoting fairness, and enhancing human agency in an increasingly algorithmic world.

A MATHEMATICAL FOUNDATIONS

This appendix provides essential mathematical formulations.

A.1 Matrix Factorization

Basic Model: $\mathbf{R} \approx \mathbf{P}\mathbf{Q}^T$ where $\mathbf{P} \in \mathbb{R}^{m \times k}$ (user factors) and $\mathbf{Q} \in \mathbb{R}^{n \times k}$ (item factors). Predicted rating: $\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^T \mathbf{q}_i$.

Implicit Feedback (wALS): Confidence-weighted with $C_{ui} = 1 + \alpha r_{ui}$:

$$\mathcal{L} = \sum_{u,i} c_{ui}(p_{ui} - \mathbf{p}_u^T \mathbf{q}_i)^2 + \lambda \left(\sum_u \|\mathbf{p}_u\|^2 + \sum_i \|\mathbf{q}_i\|^2 \right)$$

A.2 Bayesian Personalized Ranking

Optimizes ranking: $\mathcal{L}_{BPR} = -\sum_{(u,i,j) \in D} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj})$

A.3 Neural Collaborative Filtering

Generalizes MF using neural networks: $\hat{r}_{ui} = f(\mathbf{p}_u, \mathbf{q}_i | \Theta)$ where f is a multi-layer perceptron.

B ADDITIONAL RESOURCES

B.1 Reproducibility and Datasets

To facilitate reproducibility and further research, this section provides comprehensive information about benchmark datasets, their characteristics, and appropriate usage for different feedback types.

Dataset Selection Guidelines:

- **For Explicit Feedback Research**: MovieLens (all sizes), Netflix Prize, Book-Crossing
- **For Implicit Feedback Research**: Last.fm, Taobao, Pinterest, Spotify
- **For Hybrid Systems**: Yelp Challenge, Amazon Reviews (include text and ratings)
- **For Temporal/Sequential**: YOOCHOOSE, RetailRocket, Diginetica

Table 12. Benchmark Datasets: Characteristics and Feedback Types

Dataset	Users	Items	Interactions	Feedback	Domain
Explicit Feedback Datasets					
MovieLens-100K	943	1.7K	100K	Ratings (1-5)	Movies
MovieLens-1M	6K	3.9K	1M	Ratings (1-5)	Movies
MovieLens-25M	162K	59K	25M	Ratings (0.5-5)	Movies
Netflix Prize	480K	18K	100M	Ratings (1-5)	Movies
Book-Crossing	278K	271K	1.1M	Ratings (1-10)	Books
Jester	73K	100	4.1M	Ratings (-10 to 10)	Jokes
Implicit Feedback Datasets					
Last.fm-360K	360K	290K	17.5M	Listening count	Music
Last.fm-1K	992	176K	19.1M	Plays	Music
Spotify-1M	1M	160K	1B+	Streams	Music
Amazon (multi)	Varies	Varies	233M	Purchases/views	E-commerce
Taobao	987K	4.1M	100M	Clicks/purch.	E-commerce
Tmall	425K	1.1M	54M	Actions	E-commerce
Pinterest	55K	9.9K	1.5M	Pins	Social
Yelp	1.9M	192K	8M	Check-ins	Local biz
Hybrid (Explicit + Implicit) Datasets					
Yelp Challenge	1.9M	192K	8M ratings + 1.2M reviews	Both	Reviews
Amazon Reviews	Varies	Varies	233M ratings + text reviews	Both	E-commerce
Epinions	49K	139K	664K + trust	Both	Products
Douban	129K	58K	17M + reviews	Both	Movies/Books
Sequential/Temporal Datasets					
YOOCHOOSE	–	53K	34M	Clicks/purch.	E-commerce
RetailRocket	–	235K	2.7M	Events	E-commerce
Diginetica	–	43K	1M	Sessions	E-commerce

Note: Sizes approximate; some datasets have multiple versions.

- **For Cold-Start Studies**: MovieLens-25M, Amazon (high sparsity versions)
- **For Scalability Testing**: Netflix Prize, Spotify-1M, Amazon-full

Data Access and Citation:

- MovieLens: https://grouplens.org/datasets/movielens/
- Amazon Reviews: https://cseweb.ucsd.edu/~jmcauley/datasets.html
- Last.fm: https://www.last.fm/api
- Yelp Challenge: https://www.yelp.com/dataset
- RecSysDatasets: https://github.com/caserec/Datasets-for-Recommender-Systems

Preprocessing Recommendations:

- **Minimum Interactions**: Filter users/items with <5 interactions for explicit, <20 for implicit
- **Temporal Splits**: Use time-based train/test splits (80/20) rather than random
- **Cold-Start Simulation**: Reserve 10-20% of users/items with limited data for cold-start evaluation
- **Negative Sampling**: For implicit feedback, sample negatives from unobserved items (typical ratio 1:4 or 1:10)

B.2 Open-Source Implementations

- **Surprise**: Python scikit for recommender systems
- **LightFM**: Hybrid recommendation algorithms
- **RecBole**: Comprehensive recommendation library
- **TensorFlow Recommenders**: Production-scale implementations

B.3 Benchmark Datasets

- **MovieLens**: Multiple scales (100K, 1M, 10M, 25M ratings)
- **Amazon Product Reviews**: Multi-category e-commerce data
- **Netflix Prize**: Historical movie ratings dataset
- **Last.fm**: Music listening data with implicit feedback
- **Yelp Challenge**: Business reviews and check-ins

B.4 Research Venues

Top Conferences: ACM RecSys, KDD, WWW, SIGIR, WSDM, CIKM
Key Journals: ACM TOIS, IEEE TKDE, ACM TIST, User Modeling and User-Adapted Interaction

B.5 Online Resources

- RecSys Wiki: `wiki.recsyschallenge.com`
- Papers with Code: `paperswithcode.com/task/recommendation-systems`
- Awesome Recommender Systems: `github.com/grahamjenson/list_of_recommender_systems`

REFERENCES

[1] Abdollahpouri, H., Burke, R., and Mobasher, B. The unfairness of popularity bias in recommendation. *arXiv preprint arXiv:1907.13286* (2019).
[2] Adomavicius, G., and Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering 17*, 6 (2005), 734–749.
[3] Adomavicius, G., and Tuzhilin, A. Context-aware recommender systems. In *Recommender systems handbook* (2011), Springer, pp. 217–253.
[4] Bennett, J., and Lanning, S. The netflix prize. In *Proceedings of KDD cup and workshop* (2007), vol. 2007, p. 35.
[5] Bonner, S., and Vasile, F. Causal embeddings for recommendation. *Proceedings of the 12th ACM Conference on Recommender Systems* (2018), 104–112.
[6] Breese, J. S., Heckerman, D., and Kadie, C. Empirical analysis of predictive algorithms for collaborative filtering. *Uncertainty in artificial intelligence* (1998), 43–52.
[7] Burke, R. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction 12*, 4 (2002), 331–370.
[8] Castells, P., Hurley, N. J., and Vargas, S. A study of heterogeneity in recommendations for improvements of novelty and diversity. *Proceedings of the 2nd Workshop on Novelty and Diversity in Recommender Systems* (2011), 2–9.
[9] Chai, D., Wang, L., Chen, K., and Yang, Q. Secure federated matrix factorization. *IEEE Intelligent Systems 36*, 5 (2021), 11–20.
[10] Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., and He, X. Bias and debias in recommender system: A survey and future directions. *arXiv preprint arXiv:2010.03240* (2020).
[11] Chen, J., Zhang, H., He, X., Nie, L., Liu, W., and Chua, T.-S. Attentive collaborative filtering: Multimedia recommendation with item-and component-level attention. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2017), 335–344.
[12] Chen, K., Kovvuri, R., Gao, J., and Nevatia, R. Large-scale visual relationship understanding. *Proceedings of the AAAI Conference on Artificial Intelligence 33*, 01 (2019), 9173–9180.
[13] Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., et al. Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems* (2016), ACM, pp. 7–10.
[14] Covington, P., Adams, J., and Sargin, E. Deep neural networks for youtube recommendations. *Proceedings of the 10th ACM conference on recommender systems* (2016), 191–198.
[15] Dacrema, M. F., Cremonesi, P., and Jannach, D. Are we really making much progress? a worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems* (2019), pp. 101–109.
[16] Ekstrand, M. D., Tian, M., Kazi, M. R. I., Mehrpouyan, H., and Kluver, D. Fairness in recommendation: Foundations, methods and applications. In *International Conference on Artificial Intelligence and Statistics* (2022), PMLR, pp. 9267–9278.

[17] Erkin, Z., Troncoso, C., Lagendijk, R. L., and Pérez-González, F. Privacy-preserving user profiling with packed homomorphic encryption. In *IFIP International Conference on Communications and Multimedia Security* (2012), Springer, pp. 41–55.

[18] Gao, C., Lei, W., He, X., de Rijke, M., and Chua, T.-S. Advances and challenges in conversational recommender systems: A survey. *AI Open 2* (2021), 100–126.

[19] Ge, Y., Zhao, S., Zhou, H., Mwiti, C., and Wang, W. Understanding echo chambers in e-commerce recommender systems. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2020), 2261–2270.

[20] Gomez-Uribe, C. A., and Hunt, N. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS) 6*, 4 (2015), 1–19.

[21] Gulla, J. A., Zhang, L., Liu, P., Özgöbek, Ö., and Su, X. The adressa dataset for news recommendation. In *Proceedings of the International Conference on Web Intelligence* (2017), ACM, pp. 1042–1048.

[22] He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., and Wang, M. Lightgcn: Simplifying and powering graph convolution network for recommendation. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (2020), 639–648.

[23] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., and Chua, T.-S. Neural collaborative filtering. *Proceedings of the 26th international conference on world wide web* (2017), 173–182.

[24] He, X., Zhang, H., Kan, M.-Y., and Chua, T.-S. Fast matrix factorization for online recommendation with implicit feedback. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (2016), 549–558.

[25] Herlocker, J. L., Konstan, J. A., Borchers, A., and Riedl, J. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval* (1999), pp. 230–237.

[26] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS) 22*, 1 (2004), 5–53.

[27] Hidasi, B., Karatzoglou, A., Baltrunas, L., and Tikk, D. Session-based recommendations with recurrent neural networks. In *International Conference on Learning Representations* (2016).

[28] Hou, Y., Li, S., Liu, Z., Zhang, H., He, X., Tang, B., Xiong, H., et al. Towards universal sequence representation learning for recommender systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (2022), ACM, pp. 585–593.

[29] Hou, Y., Zhang, J., Lin, Z., Lu, H., Xie, R., McAuley, J., and Zhao, W. X. Large language models are zero-shot rankers for recommender systems. *arXiv preprint arXiv:2305.08845* (2023).

[30] Hu, L., Li, C., Shi, C., Yang, C., and Shao, C. Graph neural news recommendation with long-term and short-term interest modeling. *Information Processing & Management 57*, 2 (2020), 102142.

[31] Hu, Y., Koren, Y., and Volinsky, C. Collaborative filtering for implicit feedback datasets. *2008 Eighth IEEE International Conference on Data Mining* (2008), 263–272.

[32] Jia, J., and Gong, N. Z. Privacy-preserving recommender systems: Are we there yet? *IEEE Security & Privacy 19*, 5 (2021), 30–39.

[33] Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (2005), pp. 154–161.

[34] Kang, W.-C., and McAuley, J. Self-attentive sequential recommendation. *2018 IEEE International Conference on Data Mining (ICDM)* (2018), 197–206.

[35] Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., and Newell, C. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction 22*, 4 (2012), 441–504.

[36] Koren, Y. Collaborative filtering with temporal dynamics. *Communications of the ACM 53*, 4 (2010), 89–97.

[37] Koren, Y. Collaborative filtering with temporal dynamics. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (2010), ACM, pp. 447–456.

[38] Koren, Y., Bell, R., and Volinsky, C. Matrix factorization techniques for recommender systems. *Computer 42*, 8 (2009), 30–37.

[39] Lee, D. D., and Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature 401*, 6755 (1999), 788–791.

[40] Lian, J., Zhang, F., Xie, X., and Sun, G. Cccfnet: A content-boosted collaborative filtering neural network for cross domain recommender systems. *Proceedings of the 26th International Conference on World Wide Web Companion* (2017), 817–818.

[41] Linden, G., Smith, B., and York, J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing 7*, 1 (2003), 76–80.

[42] Liu, F., Liu, X., Li, Y., Wang, S., Ren, J., and Tresp, V. Multimodal pretraining for dense video captioning. *Proceedings of the AAAI Conference on Artificial Intelligence 36*, 2 (2022), 2183–2191.

[43] Liu, J., Dolan, P., and Pedersen, E. R. Personalized news recommendation based on click behavior. *Proceedings of the 15th International Conference on Intelligent User Interfaces* (2010), 31–40.

[44] Liu, Q., Zeng, Y., Mokhosi, R., and Zhang, H. Stamp: short-term attention/memory priority model for session-based recommendation. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2018), 1831–1839.

[45] Ma, H., Yang, H., Lyu, M. R., and King, I. Learning to recommend with social trust ensemble. *Proceedings of the 32nd international ACM SIGIR conference on Research and Development in Information Retrieval* (2011), 203–210.

[46] Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., and Chi, E. H. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2018), pp. 1930–1939.

[47] Marlin, B., and Zemel, R. S. Collaborative filtering and the missing at random assumption. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence* (2007), pp. 267–275.

[48] McSherry, F., and Mironov, I. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), pp. 627–636.

[49] Okura, S., Tagami, Y., Ono, S., and Tajima, A. Embedding-based news recommendation for millions of users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2017), pp. 1933–1942.

[50] Pan, R., Zhou, Y., Cao, B., Liu, N. N., Lukose, R., Scholz, M., and Yang, Q. One-class collaborative filtering. In *2008 Eighth IEEE International Conference on Data Mining* (2008), IEEE, pp. 502–511.

[51] Pazzani, M. J., and Billsus, D. Content-based recommendation systems. In *The adaptive web* (2007), Springer, pp. 325–341.

[52] Pu, P., Chen, L., and Hu, R. User action interpretation for online content optimization. *IEEE Transactions on Knowledge and Data Engineering 25*, 2 (2013), 317–330.

[53] Qi, T., Wu, F., Wu, C., Huang, Y., and Xie, X. Causal inference for news recommendation. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (2021), 994–1002.

[54] Rendle, S. Factorization machines with libfm. *ACM Transactions on Intelligent Systems and Technology (TIST) 3*, 3 (2012), 1–22.

[55] Rendle, S., Freudenthaler, C., Gantner, Z., and Schmidt-Thieme, L. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence* (2009), pp. 452–461.

[56] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work* (1994), pp. 175–186.

[57] Ricci, F., Rokach, L., and Shapira, B. *Recommender systems handbook*. Springer, 2015.

[58] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. Item-based collaborative filtering recommendation algorithms. *Proceedings of the 10th international conference on World Wide Web* (2001), 285–295.

[59] Schnabel, T., Swaminathan, A., Joachims, T., Schnabel, T., Swaminathan, A., and Joachims, T. Recommendations as treatments: Debiasing learning and evaluation. *arXiv preprint arXiv:1602.05352* (2016).

[60] Sedhain, S., Menon, A. K., Sanner, S., and Xie, L. Autorec: Autoencoders meet collaborative filtering. In *Proceedings of the 24th international conference on World Wide Web* (2015), pp. 111–112.

[61] Shardanand, U., and Maes, P. Social information filtering: algorithms for automating word of mouth. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1995), pp. 210–217.

[62] Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., and Jiang, P. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. *Proceedings of the 28th ACM international conference on information and knowledge management* (2019), 1441–1450.

[63] Tang, J., and Wang, K. Personalized top-n sequential recommendation via convolutional sequence embedding. *Proceedings of the eleventh ACM international conference on web search and data mining* (2018), 565–573.

[64] Tang, J., and Wang, K. Towards neural mixture recommender for long range dependent user sequences. *The World Wide Web Conference* (2019), 1782–1793.

[65] Van den Oord, A., Dieleman, S., and Schrauwen, B. Deep content-based music recommendation. *Advances in neural information processing systems 26* (2013).

[66] Wang, J., Yu, L., Zhang, W., Gong, Y., Xu, Y., Wang, B., Zhang, P., and Zhang, D. Irgan: A minimax game for unifying generative and discriminative information retrieval models. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2017), 515–524.

[67] Wang, X., He, X., Cao, Y., Liu, M., and Chua, T.-S. Kgat: Knowledge graph attention network for recommendation.

In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019), pp. 950–958.
[68] Wang, X., He, X., Wang, M., Feng, F., and Chua, T.-S. Neural graph collaborative filtering. *Proceedings of the 42nd international ACM SIGIR conference on Research and Development in Information Retrieval* (2019), 165–174.
[69] Wang, Z., Zhang, H., Liu, L., Wu, Y., Wang, L., and Wang, Z. User-item matching for recommendation fairness: A counterfactual learning approach. *Proceedings of the 30th ACM International Conference on Information and Knowledge Management* (2021), 442–451.
[70] Wei, W., Huang, C., Li, L., Xie, X., Lai, Y., Chen, Y., and Zhang, M. Contrastive learning for sequential recommendation. *2021 IEEE 37th International Conference on Data Engineering (ICDE)* (2021), 1254–1265.
[71] Weston, J., Bengio, S., and Usunier, N. Wsabie: Scaling up to large vocabulary image annotation. *Proceedings of the 22nd international joint conference on Artificial Intelligence* (2011), 2764–2770.
[72] Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., and Xie, X. Neural news recommendation with attentive multi-view learning. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence* (2019), pp. 3863–3869.
[73] Wu, C., Wu, F., Ge, S., Qi, T., Huang, Y., and Xie, X. Npa: Neural news recommendation with personalized attention. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019), pp. 2576–2584.
[74] Wu, C., Wu, F., Qi, T., Huang, Y., and Xie, X. Empowering news recommendation with pre-trained language models. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2021), 1652–1656.
[75] Wu, F., Qiu, Y., Chen, Q., Zhao, X., Caverlee, J., and Huang, Y. Mind: A large-scale dataset for news recommendation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (2020), pp. 3597–3606.
[76] Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J., and Xie, X. Self-supervised graph learning for recommendation. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval* (2021), 726–735.
[77] Xie, X., Zhang, F., Wang, Z., Chen, Y., and Zhang, M. Contrastive learning for sequential recommendation. *2022 IEEE 38th International Conference on Data Engineering (ICDE)* (2022), 1253–1266.
[78] Yadati, N., Nimishakavi, M., Yadav, P., Nitin, V., Louis, A., and Talukdar, P. Hypergcn: A new method of training graph convolutional networks on hypergraphs. *Advances in Neural Information Processing Systems 32* (2019).
[79] Yao, T., Yi, X., Zhu, D. Z., Zhang, Z., and Chen, Y. Self-supervised learning for large-scale item recommendations. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), 8132–8141.
[80] Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., and Baeza-Yates, R. Reducing discrimination in ranking and recommendation. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (2020), ACM, pp. 556–567.
[81] Zhang, H., Zhang, Z., Liu, Y., Zhang, H., and Yang, Z. Knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendation. *arXiv preprint arXiv:2005.04987* (2020).
[82] Zhang, Y., and Chen, X. Explainable recommendation: A survey and new perspectives. *Foundations and Trends in Information Retrieval 14*, 1 (2020), 1–101.
[83] Zhao, X., Zhang, L., Ding, Z., Xia, L., Tang, J., and Yin, D. Recommendations with negative feedback via pairwise deep reinforcement learning. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2018), 1040–1048.
[84] Zhao, Y., Li, M., Zhang, L., Liu, Y., and Zhu, X. Improving top-k recommendation via joint collaborative autoencoders. *Proceedings of the 24th International Conference on World Wide Web* (2015), 384–394.
[85] Zheng, L., Noroozi, V., and Yu, P. S. Joint deep modeling of users and items using reviews for recommendation. In *Proceedings of the tenth ACM international conference on web search and data mining* (2017), pp. 425–434.
[86] Zhu, Z., Lin, K., and Zhou, J. Transfer learning in deep reinforcement learning: A survey. *arXiv preprint arXiv:1908.07077* (2019).