

Learning to Comparison-Shop

Jie Tang
Airbnb
San Francisco, USA
jie.tang@airbnb.com

Daochen Zha
Airbnb
San Francisco, USA
daochen.zha@airbnb.com

Xin Liu
Airbnb
San Francisco, USA
xin.liu@airbnb.com

Huiji Gao
Airbnb
San Francisco, USA
huiji.gao@airbnb.com

Liwei He
Airbnb
San Francisco, USA
liwei.he@airbnb.com

Sanjeev Katariya
Airbnb
San Francisco, USA
sanjeev.katariya@airbnb.com

ABSTRACT

In online marketplaces like Airbnb, users frequently engage in comparison shopping before making purchase decisions. Despite the prevalence of this behavior, a significant disconnect persists between mainstream e-commerce search engines and users' comparison needs. Traditional ranking models often evaluate items in isolation, disregarding the context in which users compare multiple items on a search results page. While recent advances in deep learning have sought to improve ranking accuracy, diversity, and fairness by encoding listwise context [7], the challenge of aligning search rankings with user comparison shopping behavior remains inadequately addressed. In this paper, we propose a novel ranking architecture—Learning-to-Comparison-Shop (LTCS) System—that explicitly models and learns users' comparison shopping behaviors. Through extensive offline and online experiments, we demonstrate that our approach yields statistically significant gains in key business metrics—improving NDCG by 1.7% and boosting booking conversion rate by 0.6% in A/B testing—while also enhancing user experience. We also compare our model against state-of-the-art approach and demonstrate that LTCS significantly outperforms it.

KEYWORDS

Search Ranking, learning to rank, comparison shopping, reranking

ACM Reference Format:

Jie Tang, Daochen Zha, Xin Liu, Huiji Gao, Liwei He, and Sanjeev Katariya. 2025. Learning to Comparison-Shop. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (CIKM '25)*. ACM, New York, NY, USA, 8 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 INTRODUCTION

Online marketplaces such as Amazon, eBay, Walmart, and Airbnb provide search functions that help users discover products or services tailored to their preferences. Despite offering different services, from retail goods to food delivery and rental accommodation,

these platforms share a common goal: their ranking models are designed to present the most relevant items to users. However, from the user's perspective, the need for comparison shopping often persists before arriving at a final decision. Comparison shopping helps users achieve an optimal balance of price, quality, and convenience when making purchase decisions.

Despite users' need for comparison shopping, mainstream ranking models are typically designed without accounting for this behavior. Traditional learning-to-rank models are primarily built to learn a scoring function that computes a relevance score for each item independently. In other words, these models lack the ability to directly compare items with one another during inference, even though such comparisons are often incorporated into offline training through pairwise [2] [3] or listwise loss functions [4]. Recent efforts [10] [1] [9] have introduced attention mechanisms to capture cross-item interactions in learning-to-rank models. These approaches encode ranking context, enabling the model to directly compare items during inference. However, merely encoding ranking context is insufficient to accurately capture and learn users' comparison shopping behavior.

To better understand this behavior, we can take a step back and consider how users engage in comparison shopping: initially, users browse the top results returned by a search ranking model, evaluating each item independently to determine whether to click on it (a pointwise approach). After selecting items, users compare the clicked items by examining various details, including descriptions, prices, reviews, images, and more. This process often involves several rounds of back-and-forth deliberation, which may span minutes, hours or days before users reach a final decision.

Encoding ranking context can certainly enhance the comparison stage. In a traditional multi-stage ranking system, one might train a ranking model (initial ranker) to learn users' pointwise clicking behavior, while a separate re-ranker [7] employs ranking context encoding to learn users' comparison behavior. However, this approach is fundamentally distinct from actual user comparison shopping behavior. From the user's perspective, pointwise clicking and comparison behavior are highly interconnected, with the rationale behind pointwise clicking significantly influencing the comparison process. Training the ranker and re-ranker separately risks disrupting this connection; in some instances, they may even conflict with one another.

In this paper, we propose a novel model that efficiently learns users' comparison shopping behavior. To the best of our knowledge, this work represents the first attempt to address the comparison

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '25, November 10–November 14, 2025, Seoul, Korea

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/XXXXXXXX.XXXXXXX>

shopping problem within the learning-to-rank domain. Our major contributions include:

- We decompose users’ comparison shopping behavior into two distinct stages: a pointwise evaluation stage and a set-wise comparison stage.
- We propose a novel framework that models and learns users’ two-stage comparison shopping behavior as an integrated process.
- Behavior-aligned co-training: We propose a novel co-trained architecture that mirrors users’ two-stage comparison behavior by combining a pointwise initial ranker with a setwise self-attention re-ranker.
- We successfully applied this framework to a large-scale system, such as Airbnb’s search ranking system. Additionally, we conducted thorough analyses to validate our design assumptions and choices.

While both self-attention mechanisms and co-training approaches have been individually explored in ranking models, our key contribution lies in their behavior-aligned integration to simulate the two-stage comparison shopping process common in online marketplaces. Specifically, our LTCS framework jointly trains a lightweight, pointwise initial ranker and a self-attention-based setwise re-ranker to reflect how users first evaluate listings independently, then compare top candidates before making a decision. This co-trained architecture not only improves ranking performance but also enhances consistency between initial and final ranking stages—something prior work such as SetRank and other neural re-rankers do not address, as they typically model listwise interactions without an intermediate filtering stage or coordinated training. To our knowledge, LTCS is the first LTR architecture that explicitly models and learns from this comparison-driven user behavior, leading to both significant offline gains and measurable business impact in production.

2 RELATED WORK

Learning to rank (LTR) has been a popular research topic for decades, with one important area of study focused on evolving loss functions. The pointwise loss predicts action probability (e.g. pCTR, pCVR) for each item independently [8]. The pair-wise loss looks at two items each time, and converts ranking to a binary classification problem: whether item A is better than item B. [2, 3]. The list-wise loss considers the whole list of items and try to approximate the optimal order [4].

As demonstrated, researchers have progressively advanced from pointwise to pair-wise and list-wise loss functions, continually refining these methods to more effectively optimize ranking metrics such as NDCG. A key insight is that the more explicitly the relative order of items is modeled, the better the alignment between the loss function and ranking metrics. However, despite these advancements in loss function research, the learned scoring functions remain pointwise, meaning that during inference, the model can only compute relevance scores independently. On another note, it is well established that users’ choices are influenced by both item position (positional bias [6]) and ranking context (contextual bias [5, 18]). Attention mechanisms have been widely adopted to address contextual bias [17, 18] and, more generally, to encode ranking

context. For instance, self-attention networks have been employed to model ranking context or cross-document interactions [1, 9–11]. In this approach, the learned contextual embeddings, combined with each item’s own features, are fed into the scoring function, making the resulting relevance score context-dependent.

The application of self-attention networks to encode ranking context, as demonstrated in [1, 9–11], has advanced learning-to-rank approaches to the next level, resulting in setwise ranking models. However, these efforts have primarily focused on developing standalone rankers or re-rankers aimed solely at improving ranking metrics. None have addressed the comparison shopping problem, which holds the potential to not only enhance business metrics but also improve user experience.

3 PROPOSED FRAMEWORK

3.1 Problem Statement

Given a search query q_i , there are n matched items returned from retrieval stage. The objective of learning to rank is to assign a score s_j to j^{th} item $x_{i,j}$ such that those items could be ranked in descending order of scores. In ML terms, the training example is represented as $\{q_i, X_i, Y_i\}$, where $X_i = \{x_{i,j}\}_{j=1}^n$, with $x_{i,j} \in R^{m \times 1}$ denoting the feature vector. $Y_i = \{y_{i,j}\}_{j=1}^n$ where $y_{i,j} \in \{0, 1\}$ represents the label. In traditional learning-to-rank setup, the goal is to train a model to minimize the loss function $Loss(f) = \frac{1}{N} \sum_i l(f(q_i, X_i), Y_i)$. Here, $f(q_i, X_i)$ is an univariate scoring function that depends solely on individual item. To model users’ comparison shopping behavior, we maintain a similar setup, but modify both the scoring function (model) and the approach used for training and serving the models.

3.2 Introduction to Comparison Shopping

Before diving into the modeling of comparison shopping, it is important to first understand how users engage in comparison shopping. Our logs indicate that users generally follow these steps when making their final purchase decisions:

- Initial search and evaluation (pointwise Thinking)
 - At this stage, users initiate a search query and are presented with a list of items retrieved by the search engine.
 - Users typically evaluate each of the top-ranked items independently, considering key factors such as price, ratings, and descriptions, etc
 - Users may click on individual items to obtain detailed information for subsequent comparisons.
- Comparative Evaluation (Set-wise Thinking)
 - Once a set of potential items has been identified, users enter the comparison stage, where they evaluate the clicked items relative to one another.
 - Factors like price differences, feature variations, and user reviews are weighed against each other.
 - Moreover, the initial impression formed during the pointwise evaluation stage plays a significant role, as users continue to factor in their first impressions when making comparisons.
 - Users typically go back and forth between items, comparing them on various dimensions to determine the best value for their needs.
- Decision Making

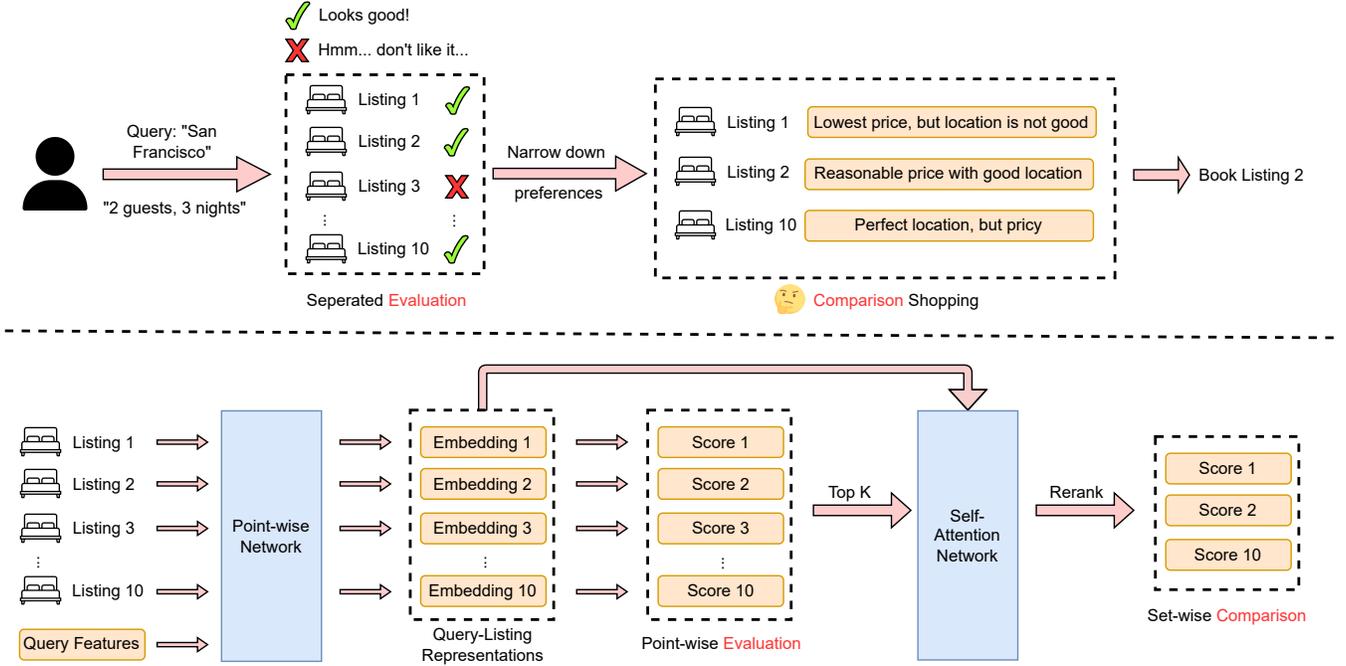


Figure 1: Top: An illustration of how users book listings on Airbnb, encompassing two stages: evaluation and comparison. In the evaluation stage, users quickly browse through the available listings to identify those that pique their interest. During the comparison stage, they compare a small selection of candidate listings across various dimensions to make a final decision. Bottom: We model this comparison shopping behavior using pointwise and set-wise self-attention networks for the above two stages, respectively. These two networks are co-trained to predict users’ booking probabilities.

- After a few rounds of comparisons, users make a final decision based on the information gathered.
- This decision could be influenced by specific preferences (e.g., budget, features) and may take place immediately or after additional rounds of research over multiple days.

Early research efforts [2] [3] [4] [8] primarily focused on modeling users’ pointwise thinking behavior, irrespective of the loss function employed, as the scoring functions remain univariate at inference time. More recent studies [1] [9] [10] [11] have incorporated attention mechanisms to capture item interactions, which can be seen as an attempt to model users’ set-wise thinking. Although attention-based methods often employ a re-ranker built on top of a ranker that learns pointwise behavior[7], these components are trained independently. While this strategy mimics certain aspects of comparison shopping, it treats the processes as separate rather than integrated, ultimately leading to sub-optimal outcomes. The rationale behind this is that, during the comparison shopping stage, users continue to apply reasoning and information gained through pointwise thinking. As a result, decoupling the re-ranker from the initial ranker would create a substantial disconnect between the model’s behavior and the actual decision-making process of users.

In the following section, we will demonstrate this more rigorously through mathematical formalization and introduce our Learning-to-Comparison-Shop (LTCS) framework. Figure 1 offers an overview: the top part illustrates how an Airbnb user engages in comparison shopping to make a booking decision, while the

bottom part depicts how our proposed LTCS framework models this behavior.

3.3 Learning to Comparison-Shop Framework

As demonstrated by [7], re-rankers that leverage self-attention mechanisms to encode ranking context have seen widespread adoption in multi-stage ranking systems. In this approach, an initial ranking model is trained to estimate the purchase probability $P(E_{i,j}|q_i, x_{i,j})$ for each item $x_{i,j}$ retrieved in response to a query q_i , here $E_{i,j}$ denotes the observed instance of a user purchasing the j -th item. Subsequently, a re-ranker is independently trained to predict the purchase probability $P(E_{i,j}|q_i, x_{i,j}, C_{i,j,k})$ for each item $x_{i,j}$ among the top-K items returned by the initial ranking model, with $C_{i,j,k}$ representing the context of $x_{i,j}$ relative to the surrounding items in the top-K items.

To enhance this framework, we first demonstrate that:

$$P(E_{i,j}|q_i, x_{i,j}, C_{i,j,k}) \propto P(E_{i,j}|q_i, x_{i,j})P(E_{i,j}|q_i, C_{i,j,k}) \quad (1)$$

Here, $P(E_{i,j}|q_i, x_{i,j})$ denotes the ranking score produced by the initial ranker, $P(E_{i,j}|q_i, x_{i,j}, C_{i,j,k})$ represents the ranking score generated by the re-ranker, and $P(E_{i,j}|q_i, C_{i,j,k})$ models the contextual ranking bias.

According to Bayesian rule, we could rewrite right side of (1) as

$$P(E_{i,j}|q_i, x_{i,j})P(E_{i,j}|q_i, C_{i,j,k}) = \frac{P(E_{i,j}, x_{i,j}|q_i)P(E_{i,j}, C_{i,j,k}|q_i)}{P(x_{i,j}|q_i)P(C_{i,j,k}|q_i)} \quad (2)$$

Assuming that the items $\{x_{i,j}\}$ are independent of each other under the condition of query q_i can be a reasonable approximation in certain contexts. Retrieved items often share similar features due to the query q_i , which may introduce dependencies among these features. However, when modeling users' comparison-shopping behavior, these common features may not significantly influence the decision-making process. Instead, it is the uncorrelated or unique features that make an item stand out to users. Therefore, assuming independence among items with respect to these distinctive features could be justified, as it emphasizes the attributes most influential in differentiating items during comparison shopping.

Under this assumption, we have:

$$P(C_{i,j,k}|q_i) = \prod_{l=1, l \neq j}^k P(x_{i,l}|q_i) \quad (3)$$

$$\begin{aligned} & P(x_{i,j}|q_i)P(C_{i,j,k}|q_i) \\ &= P(x_{i,j}|q_i) \prod_{l=1, l \neq j}^k P(x_{i,l}|q_i) = P(x_{i,j}, C_{i,j,k}|q_i) \end{aligned} \quad (4)$$

Similarly, assuming items $x_{i,j}$ are independent of each other under condition of query q_i and purchase event $E_{i,j}$, we have

$$P(x_{i,j}|E_{i,j}, q_i)P(C_{i,j,k}|E_{i,j}, q_i) = P(x_{i,j}, C_{i,j,k}|E_{i,j}, q_i) \quad (5)$$

Thus, applying Bayes' rule and incorporating Eq(4) and Eq(5), Eq (2) could be rewritten as

$$\begin{aligned} & P(E_{i,j}|q_i, x_{i,j})P(E_{i,j}|q_i, C_{i,j,k}) \\ &= \frac{P(E_{i,j}, x_{i,j}|q_i)P(E_{i,j}, C_{i,j,k}|q_i)}{P(x_{i,j}|q_i)P(C_{i,j,k}|q_i)} \\ &= \frac{P(x_{i,j}|E_{i,j}, q_i)P(E_{i,j}|q_i)P(C_{i,j,k}|E_{i,j}, q_i)P(E_{i,j}|q_i)}{P(x_{i,j}, C_{i,j,k}|q_i)} \\ &= \frac{P(x_{i,j}, C_{i,j,k}|E_{i,j}, q_i)P(E_{i,j}|q_i)P(E_{i,j}|q_i)}{P(x_{i,j}, C_{i,j,k}|q_i)} \\ &= \frac{P(x_{i,j}, C_{i,j,k}, E_{i,j}|q_i)P(E_{i,j}|q_i)}{P(x_{i,j}, C_{i,j,k}|q_i)} \\ &= P(E_{i,j}|q_i, x_{i,j}, C_{i,j,k})P(E_{i,j}|q_i) \end{aligned} \quad (6)$$

In Eq (6), $P(E_{i,j}|q_i)$ is a prior which could be considered as constant, so we could prove Eq(1)

$$P(E_{i,j}|q_i, x_{i,j}, C_{i,j,k}) \propto P(E_{i,j}|q_i, x_{i,j})P(E_{i,j}|q_i, C_{i,j,k})$$

From Eq(1), we could infer 1) the re-ranker's score is highly correlated with that of the initial ranker, and 2) the re-ranker's learning objective inherently encompasses the initial ranker's objective. These two points are consistent with our intuition and observations in 3.2, where we argue that users' set-wise thinking is not independent of pointwise thinking; rather, they are integrated components of the decision-making process. These insights suggest a clear direction for improving the modeling of users' comparison behavior: co-training the initial ranker and the set-wise re-ranker. This approach offers several potential benefits:

(1) As demonstrated earlier, the initial ranker and the set-wise re-ranker are highly correlated. Co-training them can allow the models to complement each other, mitigating concerns about negative transfer.

(2) By learning a shared representation of items, reasoning from the pointwise evaluation can be seamlessly transferred to set-wise comparison. Additionally, understanding how users engage in set-wise reasoning enables the initial ranker to more effectively select the top-K items for comparison shopping.

4 LEARNING-TO-COMPARISON-SHOP (LTCS) SYSTEM

4.1 System Overview

As outlined in Section 3.3, we propose an innovative ranking system that co-trains the initial ranker alongside a set-wise re-ranker. This system is specifically designed to emulate users' comparison behavior, with the dual objectives of optimizing key business metrics and improving the overall user experience. The architecture of the model is depicted in Figure 2:

- (1) The initial ranker is designed as a simple Multi-Layer Perceptron (MLP) network.
 - (a) The input to the network comprises item features and query features.
 - (b) The network outputs two components: (i) a logit representing the predicted purchase probability, and (ii) an initial ranking embedding. The initial ranking embedding is derived from the final hidden layer of the initial ranker, encapsulating information from each {query, item} pair.
 - (c) The initial rankers in Figure 2 are actually same ranker, all of them share same parameters.
- (2) The set-wise re-ranker is jointly trained with the initial ranker.
 - (a) The initial ranking embeddings of the top-K items (sort by their initial ranker logit scores) are transmitted from the initial ranker to the re-ranker. Within the re-ranker, these embeddings are input into a self-attention layer to produce a context embedding.
 - (b) Finally, for each item, the re-ranking logit is computed by passing its individual initial ranking embedding and the generated context embedding through the re-ranker's multilayer perceptron (MLP) network.

For the purpose of model training, the loss function is defined as follows: Let N denote the number of training examples, where each example corresponds to a single search query. Each training example contains the top K listings returned by the ranking system for that query.

$$Loss_{initial} = - \sum_{i=1}^N \sum_{j=1}^K y_{i,j} \log \left(\frac{e^{logit_{initial,i,j}}}{\sum_{k=1}^K e^{logit_{initial,i,k}}} \right) \quad (7)$$

$$Loss_{rerank} = - \sum_{i=1}^N \sum_{j=1}^K y_{i,j} \log \left(\frac{e^{logit_{rerank,i,j}}}{\sum_{k=1}^K e^{logit_{rerank,i,k}}} \right) \quad (8)$$

$$Loss = (1 - \alpha)Loss_{initial} + \alpha Loss_{rerank}, \quad (9)$$

where

- (1) $logit_{initial,i,j}$ is logit output from initial ranker.
- (2) $logit_{rerank,i,j}$ is logit output from reanker.
- (3) $embedding_{i,j}$ is extracted from last hidden layer of initial ranker.

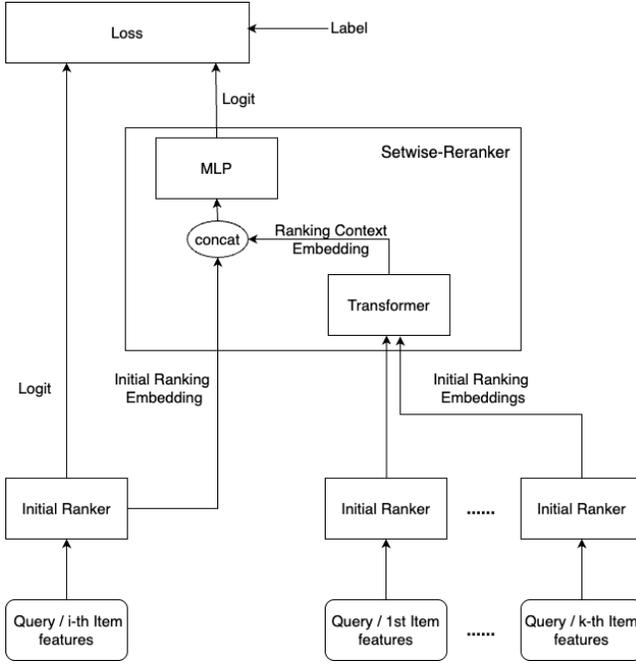


Figure 2: Overall system diagram: For a given candidate item, initial ranker computes its initial ranking logit and also initial ranking embedding based on query and item features. To compute the context embedding, a self-attention network is employed over top-k items’ initial ranking embeddings. Finally the candidate item’s initial embedding and context embedding are concatenated and passed through a MLP to generate re-ranker logit. The loss is computed by weighted sum of initial ranker loss and re-ranker loss.

- (4) $context_embedding_i$ is computed using a self-attention network, which takes as input the top-K item embeddings produced by the initial ranker as illustrated in Figure 2.

As shown in Eq(7), Eq(8), a list-wise loss function [4] is applied to both the initial ranker and the re-ranker. The total loss is then obtained by combining these two components through a weighted sum, as described in Eq(9). Additionally, as previously mentioned, the embedding of the j -th item ($embedding_{i,j}$) is extracted from the final hidden layer of the initial ranker. To compute the re-ranking logit for each item, its embedding $embedding_{i,j}$ is concatenated with ranking context embedding $context_embedding_i$. The concatenated embedding is then passed through the re-ranker MLP for further processing. Consistent with prior set-wise ranking approaches [1, 7, 9, 10], a self-attention network [15] is employed to effectively encode the ranking context as $context_embedding_i$.

It is evident that the proposed model structure and training schema are well-aligned with the learning-to-comparison-shop framework outlined in 3.3:

- (1) The initial ranker and re-ranker are co-trained, facilitating the learning of shared item representations.
- (2) Knowledge gained from the initial ranker can be transferred to the re-ranker through shared embeddings, while the initial

ranker benefits from improved collaboration with the set-wise re-ranker, enabled by these shared embeddings.

4.2 Model Serving

In our Learning-to-Comparison-Shop (LTCS) framework, the initial ranker and re-ranker are exported separately from the training graph (Figure 2) for deployment. To improve serving efficiency, we adopt two key designs:

- **Hierarchical model deployment.** The lightweight, pointwise initial ranker scores all candidates across distributed leaf nodes. The more complex, set-wise re-ranker operates at the master node, refining the top results. This structure enables scalable handling of large candidate pools by expanding leaf nodes.
- **Representation Reuse:** To avoid redundant computation, query-listing representations computed by the initial ranker are reused by the re-ranker via direct transmission from leaf to master nodes, improving efficiency.

5 EXPERIMENTAL RESULTS

5.1 Experiment Setup

The Airbnb ranking system operates as a multi-stage ranking framework. Given a query, the retrieval stage may initially return thousands of listings which are then passed to the initial ranking stage. At this stage, the initial ranker reduces the candidate set, for example, to a few hundred listings. Finally, in the re-ranking stage, few re-rankers further refine this set to optimize for business constraints and objectives.

We have successfully deployed the Learning-to-Comparison-Shop (LTCS) framework into our ranking system. However, since the re-ranker employs a self-attention network to encode the ranking context, its time complexity is $O(K^2)$, where K represents the number of listings fed into the re-ranker. As a result, K cannot be set too large, as it would substantially increase the search engine’s latency. Therefore, we limit the re-ranking process to the top 40 listings, meaning that only the top 40 results are re-ranked, while the remaining listings retain their original order.

The proposed LTCS models were trained on approximately 360 million examples collected over last one year, using a dataset consisting exclusively of booking labels. The structures of both the initial ranker and re-ranker were defined based on careful hyperparameter tuning to select optimal primary parameters (details provided in Section 5.5).

- As previously mentioned, the initial ranker is implemented as a Multi-Layer Perceptron (MLP) network, with hidden layers consisting of 2048, 1024, 512, 256, and 64 neurons, respectively. The activation function used is Smelu [13]. The model is fed with de-identified set of hundreds of listing, query, and user features, and is trained to optimize multiple objectives, as outlined in [14]. Under the LTCS framework, the model produces two outputs: a booking probability logit and an initial ranking embedding, which is extracted from the final hidden layer.
- As depicted in Fig 2, re-ranker’s input consists of the initial ranking embeddings of the top-K listings returned from the initial ranking stage. These embeddings are fed into a self-attention

network to generate context embedding. The self-attention network consists of 30 stacked layers, each incorporating residual connections and an intermediate normalization layer. Each layer includes 4 attention heads. The input to the network is the embedding sequence of the top 40 listings generated by the initial ranker. Since, at the comparison-shopping stage, users often browse through all top listings irrespective of their initial ranking order, positional embeddings are not required in this context. For each listing, the initial ranking embedding and the context embedding are concatenated and passed through the re-ranker MLP network with hidden layers consisting of 256, 128, and 64 neurons. The activation function used is also Smelu [13]. The final order of the top40 listings is determined based on the re-ranker’s logits.

- The training loss is computed as the average of the initial ranker loss and the re-ranker loss: $Loss = 0.5Loss_{initial} + 0.5Loss_{rerank}$.

The model is trained in a distributed manner using 10 A10G GPUs, with a total training time of approximately 20 hours. Prior to deploying LTCS into our ranking system, a three-week online A/B test was conducted, with the control model being our existing multi-objective learning-to-rank model [14].

5.2 Baselines

To the best of our knowledge, LTCS is the first attempt to explicitly address user comparison shopping behavior in an industrial ranking system. As such, there are no direct existing baselines tailored to this specific objective. Therefore, in this study, we compare LTCS against the following strong alternative baselines:

- **attn-DIN** [10] is a state-of-the-art neural re-ranking model[7]. It has been shown to outperform several influential neural re-rankers, including SetRank [9], Deep Listwise Context Model (DLCM) [12], and Groupwise Scoring Functions (GSF) [1]. These methods typically treat reranking as an isolated stage applied to a fixed candidate set, optimizing for relevance within that subset.
- To ensure a fairer comparison to our LTCS framework—which first uses a pointwise initial ranker (identical to that in LTCS) to filter top-K candidates before applying a setwise re-ranker—we additionally implement a modified version—**attn-DIN+**—which includes a separately trained initial ranker to simulate a two-stage pipeline. However, unlike LTCS, attn-DIN+ trains the rankers independently, without any shared parameters or coordinated optimization, and does not model the transition from broad item evaluation to focused comparison.
- **MO-LTR** [14], our current production model, is a multi-objective learning-to-rank framework designed for one-step, end-to-end optimization across multiple business goals. While it has largely replaced traditional re-rankers in our existing ranking stack by learning a unified objective, our evaluation demonstrates that the additional value provided by the LTCS re-ranker cannot be fully captured by MO-LTR alone.

5.3 Offline evaluation

Table 1 presents the offline evaluation results in terms of NDCG and its standard deviation across four ranking models. Among the baselines, attn-DIN and MO-LTR achieve similar NDCG scores

Model	NDCG	NDCG STDEV
attn-DIN	0.6856	0.01
attn-DIN+	0.6918	0.008
MO-LTR	0.6860	0.0004
LTCS	0.6982	0.0003

Table 1: Offline evaluation results comparing NDCG and its standard deviation across four ranking models

(0.6670 and 0.6667 respectively), while attn-DIN+—a two-stage variant with a separately trained initial ranker—improves performance slightly to 0.6757. In contrast, LTCS achieves the highest NDCG score of **0.6909** (+1.76% comparing to MO-LTR), outperforming all baselines by a clear margin. Notably, LTCS also exhibits the lowest standard deviation (**0.0003**), indicating highly stable training dynamics. These results demonstrate that co-training a pointwise initial ranker with a setwise re-ranker not only yields higher ranking effectiveness, but also leads to significantly improved robustness over independently trained or single-stage models.

5.4 Online Evaluation

We evaluate the performance of our new system, Learning-to-Comparison-Shop (LTCS) System with online A/B tests. We conducted a 3-week A/B test, where the control model was multi-objective learning to ranking model (MO-LTR) [14], and the treatment model was the proposed LTCS system. This A/B test demonstrated a **+0.6%** increase in booking conversion rate (CVR) with a p-value of 0.0, marking one of the largest CVR gains for a core ranking model that we have observed in recent years.

In addition to optimizing core business metrics, such as booking conversion rate (CVR), LTCS aims to enhance user experience by learning and modeling users’ comparison behaviors during the shopping process. Our A/B tests confirmed improvements in key user engagement metrics. Specifically, the *Average Unique Listing Detail Pages Viewed in the Same Location Before Booking Request* metric decreased by **0.88%** with a p-value of 0.001. This finding indicates that, compared to the control group, users in the treatment group viewed significantly fewer listing detail pages before submitting a booking request than users in the control group. Since this metric is often regarded as an indicator of search efficiency, the findings imply that the ranking model improves the comparison-shopping process, reducing the time and effort required to evaluate multiple listings and ultimately enhancing the user experience.

booking conversion rate	+0.6%
Listings viewed before booking	-0.88%

Table 2: Online Evaluation Result

5.5 Hyper-parameter tuning

The primary hyperparameters of the LTCS framework are the re-ranker input length and the number of stacked self-attention layers. We conducted the following hyperparameter tuning process to select optimal values for these parameters in our deployed model.

- **Impact of Re-Ranker Input Length:** We first examined how varying the re-ranker input length affects model performance. As

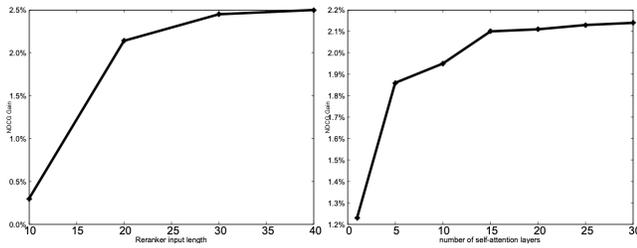


Figure 3: Left: Impact of Re-Ranker Input Length on NDCG Gain: Increasing Input Length Leads to Higher NDCG Gains. Right: Effect of Self-Attention Layers on NDCG Gain: Increasing Layers Improves Performance.

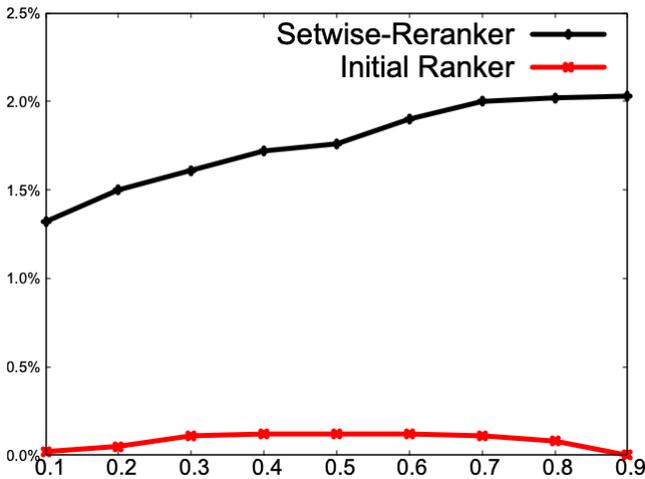


Figure 4: Effect of Re-Ranker Loss Weight on NDCG Gain: Higher Weights Improve Re-Ranker Performance and Benefit the Initial Ranker. The X-axis represents the re-ranker loss weight α , while the Y-axis shows the NDCG gain relative to a ranking system using only the initial ranker.

shown in the left pf Figure 3, increasing the input length from 10 to 40, while keeping the number of self-attention layers fixed at 30, leads to a consistent improvement in NDCG gain (Unless otherwise specified, NDCG gain in this paper refers to the end-to-end NDCG of the full ranking system, encompassing both the initial ranker and the re-ranker). However, beyond an input length of 40, the gain plateaus, indicating diminishing returns. Based on this observation, we selected top-40 as the re-ranking threshold for our deployed model, striking a balance between performance gains and computational efficiency.

- **Impact of Self-Attention Layers:** Next, we investigated the effect of stacking more self-attention layers in the re-ranker. The right-hand side of Figure 3 illustrates that increasing the number of layers from 1 to 30, while keeping the input length fixed at 20, results in a steady NDCG improvement. However, beyond 30 layers, the performance gains saturate, suggesting that additional layers provide limited benefit. Consequently, we selected 30 self-attention layers for the deployed model to optimize performance while maintaining computational efficiency.

- **Impact of Re-Ranker Loss Weight α :** Lastly, we evaluated the role of re-ranker loss weight α (introduced in Equation 9) in model performance. As shown in Figure 4, where we consider $\alpha = 0$ as the baseline (equivalent to a ranking system using only the initial ranker), increasing α from 0.1 to 0.9 improves the setwise-re-ranker NDCG. Interestingly, co-training also benefits the initial ranker due to the correlation between the two tasks. The initial ranker achieves its peak gain at $\alpha = 0.5$, but further increasing α causes the gain to decline. At $\alpha = 0.9$, the initial ranker no longer shows any improvement, even though the setwise re-ranker reaches its peak. This analysis highlights the trade-off in co-training: further increasing α enhances setwise re-ranker performance but reduces the initial ranker’s effectiveness as the model prioritizes setwise learning. Based on Figure 4, $\alpha = 0.7$ appears optimal in balancing the two models. However, in our online A/B test, we found that $\alpha = 0.5$ performed better when considering other business metrics (This falls outside the scope of this paper and will not be further discussed). Therefore, in our deployed model, we set $\alpha = 0.5$. A notable exception occurs when $\alpha = 1.0$, where only the setwise re-ranker is trained. In this scenario, we observed a significant NDCG drop and highly unstable training dynamics when the same model was trained multiple times. On average, the NDCG gain is 0% compared to the baseline $\alpha = 0$, and the standard deviation of NDCG change increased sharply: 0.01 for $\alpha = 1$ vs. 0.0004 for $\alpha < 1$. This further demonstrates an additional benefit of co-training the initial ranker and re-ranker: improved training stability. One possible explanation is that during co-training, the initial ranker loss helps regularize the re-ranker’s updates, and vice versa, leading to a more stable optimization process.[16]

These studies demonstrate that a larger context window and a deeper self-attention network enable the model to more effectively learn comparison shopping behaviors. This finding aligns with similar observations in the deep learning community and supports our intuition: by reviewing more listings and engaging in deeper cognitive processing, users are likely to make more informed choices or better trade-offs between price and quality. However, it is important to note that users may lose patience with an overly extensive comparison process, the model does not face such limitations.

6 CONCLUSION

In this study, we introduce the Learning-To-Comparison-Shop (LTCS) system, practiced in Airbnb to model comparison shopping behaviors. First, we ground our design in observations from Airbnb, emphasizing the importance of user initial preferences in comparison shopping. Following this, we detail the efficient implementation of LTCS within a production environment. Through comprehensive experiments and analysis of hyperparameters, we demonstrate that LTCS significantly enhances both ranking and core business metrics in online A/B testing.

REFERENCES

- [1] Qingyao Ai, Xuanhui Wang, and Sebastian Bruch etc. 2019. Learning Groupwise Multivariate Scoring Functions Using Deep Neural Networks. In *ICTIR '19: Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*. ACM, New York, NY, 85–92.
- [2] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. 2005. Learning to rank using gradient descent. In *ICML '05: Proceedings of the 22nd International Conference on Machine learning*. 89–96.

- [3] Chris J.C. Burges. 2010. *From RankNet to LambdaRank to LambdaMART: An Overview*. Technical Report MSR-TR-2010-82.
- [4] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, , and H. Li. 2007. Learning to rank: from pairwise approach to listwise approach. In *ICML '07: Proceedings of the 24th International Conference on Machine Learning*. 129–136.
- [5] Shashank Gupta, Philipp Hager, and Jin Huang etc. 2024. Unbiased Learning to Rank: On Recent Advances and Practical Applications. In *WSDM '24: Proceedings of the 17th ACM International Conference on Web Search and Data Mining*. 1118–1121.
- [6] Thorsten Joachims, Adith Swaminathan, and Tobias Schnable. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *WSDM '17: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 781–789.
- [7] Weiwen Liu, Yuanjia Xi, Jiarui Qin, and Fei Sun etc. 2022. Neural Re-ranking in Multi-stage Recommender Systems: A Review. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*. 5512–5520.
- [8] H. Brendan McMahan, Gary Holt, D.Sculley, and Michael Young etc. 2013. Ad Click Prediction: a View from the Trenches. In *KDD '13: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, New York, NY, 1222–1230.
- [9] Liang Pang, Jun Xu, and Qingyao Ai etc. 2020. SetRank: Learning a Permutation-Invariant Ranking Model for Information Retrieval. In *SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, 499–508.
- [10] Rama Kumar Pasumarthi, Honglei Zhuang, and Xuanhui Wang etc. 2020. Permutation Equivariant Document Interaction Network for Neural Learning-to-Rank. In *ICTIR '20: Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*. 145–148.
- [11] Changhua Pei, Yi Zhang, and Yongfeng Zhang etc. 2019. Personalized re-ranking for recommendation. In *RecSys '19: Proceedings of the 13th ACM Conference on Recommender Systems*. ACM, New York, NY, 3–11.
- [12] Jiafeng Guo Qingyao Ai, Keping Bi and W Bruce Croft. 2018. Learning a deep listwise context model for ranking refinement. In *SIGIR*.
- [13] Da Huang Rohan Anil, Sandra Gadhanho and Nijith Jacob etc. 2022. On the Factory Floor: ML Engineering for Industrial-Scale Ads Recommendation Models. In *Recsys 2022 Workshop on Online Recommender Systems and User Modeling*.
- [14] Jie Tang, Huiji Gao, Liwei He, and Sanjeev Katariya. 2024. Multi-objective Learning to Rank by Model Distillation. In *KDD '24: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*.
- [15] Ashish Vaswani, Noam Shazeer, and Niki Parmar etc. 2017. Attention is all you need. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*. 6000–6010.
- [16] Hongxin Wei, Lei Feng, Xiangyu Chen, and Bo An. 2020. Combating Noisy Labels by Agreement: A Joint Training Method with Co-Regularization. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [17] Zhi Zheng, Zhaopeng Qiu, and Tong Xu etc. 2022. CBR: Context Bias aware Recommendation for Debiasing User Modeling and Click Prediction. In *WWW '22: Proceedings of the ACM Web Conference 2022*. 2268–2276.
- [18] Honglei Zhuang, Zhen Qin, and Xuanhui Wang etc. 2021. Cross-Positional Attention for Debiasing Clicks. In *WWW '21: Proceedings of the Web Conference 2021*. 788–797.