Unified Cross-Domain Semantic Retrieval for Cold-Start Audiobook Discovery

Amit Goyal
Amazon Music
San Francisco, CA, USA
goyalam@amazon.com

ABSTRACT

Recommending a new content type to users who have never tried it is a hard cold-start problem: there is no in-domain history to infer taste from. We study audiobook discovery on a large music streaming platform, where most users have never played an audiobook. Many of them, however, leave behavioral traces in other domains: podcasts, audiobooks elsewhere, videos, and related content. We turn those traces into recommendations with a text-based retrieval model. Each user is described by the titles of the items they recently consumed in each domain, each audiobook by its metadata, and one multilingual text encoder maps both into a single space, so recommendation becomes nearest-neighbor retrieval. A learned gate weights the domains per user and suppresses the ones a user lacks. The prior production model relied on collaborative co-consumption vectors, which carry no information about an item's content. Writing everything as text removes that limit and lets one dense-retrieval model serve both content matching and behavioral recommendation; it also enables title-level transfer across domains, for instance audiobook listening elsewhere informing discovery on the platform. We evaluate on a held-out worldwide cohort covering nine marketplace groups, always against a single production baseline. Music, podcast, and audiobook signals together improve Recall@10 by +27.3%, and adding a fourth signal, video, raises this to +32.8%. A per-cohort breakdown shows each signal helps the users who actually carry it: the audiobook signal is worth roughly +50% for that group, and video helps most for users whose only other signal is music. An online A/B test is underway and will appear in the camera-ready. Per company confidentiality requirements, we report relative numbers only.

KEYWORDS

cross-domain recommendation, cold-start, dense retrieval, text embeddings, gated fusion, unified search and recommendation

1 INTRODUCTION

Content platforms increasingly host several media types under one roof: music, podcasts, and, more recently, audiobooks. Each new media type brings its own cold-start challenge. Almost no one has engaged with it yet, so there is no in-domain history to personalize from. On the music streaming platform we study, only a small share of eligible users have ever played an audiobook. Audiobook discovery still matters: engaged customers are more likely to become and stay paying customers, and good personalization is what gets them engaged in the first place.

In practice, users are rarely truly "cold." A music listener often also follows podcasts, listens to audiobooks elsewhere, watches videos, or engages with other content. Each of these hints at something useful: spoken-word affinity, genre taste, topical interest. The catch is that they live in different domains that share no items with the audiobooks we want to recommend.

Earlier production systems for this task used collaborative co-consumption models over in-domain music and podcast signals, and treated each signal on its own. This leaves two gaps. Co-consumption embeddings cannot score items with little collaborative history, which is the cold-start case itself. And they have no path to an item's content: a co-consumption vector for an audiobook knows nothing about the words in its title, author, or description, nor about how the same user's taste appears in other domains.

Our approach borrows a search idea and applies it to recommendation. We turn every signal into natural-language text, both the item content and a user's behavior in each auxiliary domain, and encode all of it with one multilingual text-embedding model. Retrieval is then dense cosine similarity in that shared space. A user's podcast, audiobook, and video histories become short text passages; an audiobook becomes a passage built from its metadata; and finding recommendations is just nearest-neighbor search between the two. A learned gated-fusion network combines the per-domain user representations, leaning on the informative signals and suppressing the ones a user does not have.

This puts the task at the meeting point of search and recommendation. Two-tower dense retrieval is the shared backbone of modern neural IR and of large-scale candidate generation; once we encode behavior as text, the same machinery handles both content matching, a search primitive, and behavioral personalization, a recommendation primitive, in one embedding space. It also opens up title-level transfer across domains. Because items consumed elsewhere and target audiobooks live in the same text space, a strong semantic match can carry a preference straight across domains, for example when audiobook listening elsewhere informs discovery on the platform. Collaborative embeddings have no way to express that link.

Our contributions are:

(1) A **unified text-embedding retrieval** formulation for cold-start recommendation in which heterogeneous cross-domain behavior and item content share one multilingual dense-retrieval space, bridging search-style and recommendation-style signals.
(2) A **gated fusion over four behavioral domains** (music, podcast, audiobook, and video) with learned graceful degradation to any available signal subset.
(3) A **large-scale evaluation** across nine worldwide marketplace groups analyzing signal contribution by cohort and by market, reporting relative improvements only, and a

qualitative study of how an added domain reshapes recommendations.

2 RELATED WORK

Unified search and recommendation. Search and recommendation are increasingly modeled jointly, motivated by shared user intent and shared retrieval machinery [16, 17]. A recurring theme is a common representation space in which query-driven and behavior-driven signals can be retrieved uniformly. Our work contributes to this agenda from the recommendation side: by encoding user behavior as text and matching it against text-encoded items via dense retrieval, we obtain a recommender whose core operation is indistinguishable from semantic search, making the two tasks composable in a single space.

Dense retrieval and text embeddings. Two-tower models learn user/query and item/document encoders for efficient approximate-nearest-neighbor serving [3, 15]. Dense passage retrieval [4] and sentence embeddings [8] established text encoders as strong retrievers, and contrastively pre-trained multilingual text embeddings [13] now provide off-the-shelf semantic spaces across languages. We build on this line by treating *behavioral* histories as text and reusing a multilingual text encoder as the shared backbone for a recommendation task.

Cross-domain recommendation. Cross-domain methods transfer knowledge across domains with shared users but disjoint items. Collective Matrix Factorization [10] jointly factorizes coupled matrices; EMCDR [7] learns explicit user-embedding mappings; CoNet [2] couples domains via cross-stitch units; graph methods propagate preferences across user–item graphs [18]; and surveys chart the broader landscape [19]. These assume overlapping items or explicit ratings in both domains. Our auxiliary-domain signals share *no* item space with target audiobooks, so we rely on implicit taste transfer expressed through a shared semantic (text) space rather than collaborative bridging.

Cold-start recommendation. Content-based methods bootstrap cold items or users from audio [11] or metadata [9]; meta-learning adapts from few interactions [5]; and DropoutNet [12] drops features during training to simulate missingness. Production systems incorporate side context [1]. We extend robustness-to-missingness to the *domain* level with learned gating, and use content semantics as the primary cold-start mechanism.

Gating for feature fusion. Mixture-of-experts gating [6] and explicit feature crosses [14] adaptively combine signals. We use a lightweight learned gate over domain sub-towers so that the fused user representation reflects only the signals a given user actually has.

3 PROBLEM FORMULATION

Let $\mathcal{U}$ be the set of users, $\mathcal{A}$ the catalog of audiobooks, and $\mathcal{M}$ the set of marketplaces. Each user u has a marketplace $m_u \in \mathcal{M}$ and behavioral histories across a set of domains $\mathcal{D}$. In our setting $\mathcal{D} = \{\text{music}, \text{podcast}, \text{audiobook}, \text{video}\}$, where music and podcast are in-domain (on the music platform) and the audiobook and video signals come from auxiliary domains available for the user. A user is *audiobook cold-start* if they have no audiobook plays on the music platform.

Objective. Given whatever signal subset a user has, produce a ranked list of K audiobook candidates $\hat{\mathcal{A}}_u = (a_1, \ldots, a_K)$ maximizing overlap with the user's actual future audiobook consumption $\mathcal{G}_u$ on the platform:

$$\text{Recall@}K = \frac{|\hat{\mathcal{A}}_u \cap \mathcal{G}_u|}{|\mathcal{G}_u|}. \tag{1}$$

The system serves as the retrieval stage of a multi-stage pipeline, favoring broad coverage at moderate K over precision at rank 1.

Signal–target separation. For training and the held-out evaluation, the audiobook *input* signal is drawn only from the auxiliary domain, i.e., off-platform audiobook listening, while the *target* is audiobook consumption on the music platform. The two are disjoint and the target is held out, so no target information leaks into the input. At inference time the model may additionally use on-platform audiobook listening as input, since at serving there is no future target to leak into. The gains we report therefore come from the leakage-free held-out protocol, while deployment can draw on all available signals.

4 METHOD

4.1 Everything as Text

The central design choice is to represent each domain as a short natural-language passage, so that a single text encoder maps users and items into one space:

- **Podcast / audiobook / video** histories are serialized as concatenated titles (and lightweight metadata such as genre) of a user's most recently consumed items in that domain. We keep only the most recent items per domain, for two reasons: recent activity is a fresher, stronger signal than long-tail history, and the text encoder accepts at most 512 tokens.
- **Audiobook items** are serialized from title, author, narrator, genre, and a truncated description.
- **Music** taste is captured by an averaged artist-preference embedding, projected into the shared space by a dedicated sub-tower (the one non-text domain, retained because it is near-universal).

Text passages are encoded with a frozen multilingual sentence encoder [13] (mean-pooled, L2-normalized), giving native support for the platform's many languages without per-market models.

4.2 Domain Sub-Towers

Each domain d has a sub-tower f_d mapping its input embedding to a shared h-dimensional representation $\mathbf{h}_u^{(d)} = f_d(\mathbf{x}_u^{(d)})$. Each sub-tower applies a linear projection with LayerNorm, a residual MLP block with GELU activation and dropout, and a final LayerNorm. Text-domain sub-towers use a dual-path residual transformation over the frozen encoder output; the music sub-tower projects the averaged artist embedding.

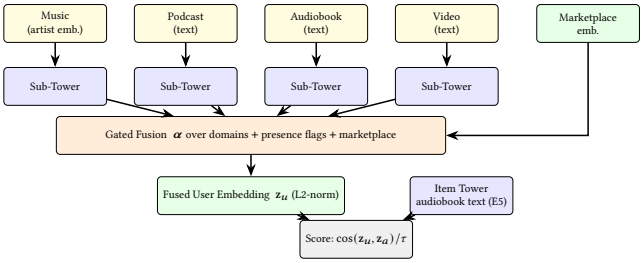


Figure 1: Unified text-embedding retrieval. Behavioral domains are serialized to text (except near-universal music taste) and encoded into a shared multilingual space; a gate fuses them into one user embedding scored against text-encoded audiobook items by cosine similarity.

4.3 Gated Fusion

A learned gate produces per-domain weights $\boldsymbol{\alpha} \in [0,1]^{|\mathcal{D}|}$ from the concatenation of domain representations, a domain-presence indicator $\mathbf{p}_u$, and a marketplace embedding $\mathbf{e}_{m_u}$:

$$\boldsymbol{\alpha}_u = \mathrm{Gate}\big([\,\mathbf{h}_u^{(d)}\ \forall d \,\|\, \mathbf{p}_u \,\|\, \mathbf{e}_{m_u}]\big), \quad \mathbf{z}_u = \mathrm{L2}\Big(\mathrm{LN}\big(\textstyle\sum_d \alpha_{u,d}\, \mathbf{h}_u^{(d)}\big)\Big). \tag{2}$$

The presence indicator and marketplace embedding let the gate learn *market-specific* fusion and route weight away from absent domains. When a domain is missing, its sub-tower output is replaced by a learned fallback (for absent podcast signal we found a most-popular-podcast imputation to be a stable alternative to a random fallback).

4.4 Item Tower and Scoring

Audiobook metadata text is encoded by the same frozen text encoder and passed through an item tower symmetric to the sub-towers, yielding $\mathbf{z}_a$. Scoring uses temperature-scaled cosine similarity $s(u,a) = \cos(\mathbf{z}_u, \mathbf{z}_a)/\tau$. Because both towers land in the same normalized space, serving reduces to approximate-nearest-neighbor retrieval.

4.5 Training

We train with a full-softmax cross-entropy over the catalog with label smoothing, augmented with a KL-divergence regularizer and an entropy term to stabilize the learned distribution; logits are normalized before the softmax. Item embeddings are trainable with a reduced learning-rate multiplier. During training each domain is independently masked with a small probability, replacing its sub-tower output with the fallback, which teaches robust inference from any signal subset at negligible cost on fully-observed users.

5 EXPERIMENTAL SETUP

Data and split. We evaluate on a worldwide dataset spanning nine marketplace groups: seven individual markets (US, UK, DE, FR, JP, IT, ES) and two grouped markets (NZ/AUS and CA/BR/MX). The test cohort consists of on-platform audiobook adopters, so that recommendation quality can be measured against known preferences; at deployment the model serves cold-start non-adopters, for whom we expect similar cross-domain coverage patterns. We use a standard train/dev/test split with disjoint users, select checkpoints by dev Recall@10, and report on the held-out test users.

Signal coverage. Within the evaluation cohort, coverage is uneven across domains. Music taste has the highest coverage (available for essentially all users), followed by the video signal (near-universal); the podcast and audiobook signals are considerably sparser. By composition, a large fraction of users have music signal only, with progressively smaller fractions holding one, two, or three additional domains. This coverage structure motivates a fusion mechanism that behaves well under partial observability.

Content embeddings. We use a frozen multilingual sentence encoder [13] (768-d) with mean pooling and L2 normalization, giving native encoding across the platform's languages and enabling a worldwide catalog without per-market content models.

Baseline. We compare against a single **production baseline**: the deployed system, which recommends audiobooks from co-consumption (colocation-popularity) patterns between audiobooks and a user's music and podcast listening—a non-learned heuristic, not a neural model. It uses the podcast co-consumption signal when present, falls back to the music co-consumption signal otherwise, and backs off to marketplace-popular audiobooks when neither is available. To separate the value of the neural architecture from that of the added cross-domain signals, we also evaluate the proposed neural model restricted to the baseline's own signals (music and podcast). Using a single baseline avoids ambiguity when attributing gains to added signals.

Metrics. We report Recall@10, NDCG@10, and MRR@10 for the overall comparison, and Recall@10 for the per-cohort and per-market breakdowns. As a retrieval stage feeding a downstream ranker, we focus on K=10. Per company confidentiality requirements we present *relative* lift over the production baseline rather than absolute values.

6 RESULTS

6.1 Overall

Table 1 reports relative improvements at K=10 over the production baseline as we add signals to the proposed model, keeping the architecture fixed down the rows. The baseline is a co-consumption heuristic, so we first give the neural model the same signals it uses, music and podcast; this measures the architecture on its own and yields a +9.7% Recall@10 gain over the baseline. Adding the audiobook signal then lifts the model to +27.3%, and adding video on top reaches +32.8%, a further +4.4% that comes mostly from video's near-universal coverage. Since the architecture is held fixed, every added signal contributes on top of it. So the +27.3% has two drivers: the architecture on its own and the audiobook signal, which supplies most of the remaining lift. Both matter, with the audiobook signal the larger contributor. NDCG@10 and MRR@10 move with recall (+30.5% and +26.2% for the full model), so the gains are in ranking quality, not only coverage.

6.2 By Signal Cohort

Table 1 varied the signals given to the model. Here we do the opposite: we keep the full all-signal model and split test users into

Table 1: Relative improvement (%) of the proposed neural model over the co-consumption production baseline at K=10, as signals are added under a single architecture. The baseline is a colocation-popularity heuristic; the music+podcast row shares its signals and isolates the architecture gain. Absolute metrics are withheld per company confidentiality requirements.

System	R@10	NDCG@10	MRR@10
Production baseline	—	—	—
Proposed: music + podcast	+9.7%	+6.0%	+1.5%
+ audiobook	+27.3%	+24.8%	+20.9%
+ video (all signals)	+32.8%	+30.5%	+26.2%

Table 2: Relative Recall@10 gain of the full (all-signal) model by user cohort, over the strongest reference within each cohort—a conservative lower bound on the gain over the production baseline (see text).

User cohort (signals present)	Δ Recall@10
Music only	+8.2%
Music + podcast	+23.9%
Music + audiobook	+50.4%
Music + podcast + audiobook	+49.5%

cohorts by the signals they actually have (Table 2). The gains track signal availability. Users with an audiobook history benefit most, roughly +50%, because that signal is a direct spoken-word cue the production baseline never sees. Users with music alone still gain +8.2% from the learned music sub-tower and marketplace conditioning. Grouping by available signals lets us read off each signal's contribution instead of a blended average: the model improves the cohorts that receive a signal, which is what a text-based cross-domain representation should do. One note on how these gains are measured. The per-cohort and per-market numbers (Tables 2 and 4) compare against the strongest non-personalized reference in each segment. The production baseline never beats that reference, so the gains over production are at least this large.

6.3 Marginal Value of the Video Domain

Isolating the video signal (Table 3) shows where it helps: mostly where a user has no other cross-domain signal. Music-only users gain +9.6% when video is added, while users who already have a podcast or audiobook signal barely move. Video's value is therefore mainly reach. It covers the large group with no other cross-domain signal, and it overlaps with podcast for users who have one.

6.4 By Marketplace

The music+podcast+audiobook model wins in eight of nine marketplace groups (Table 4), with the largest relative gains in the highest-volume English- and German-language markets. Adding video is especially valuable in smaller markets where podcast and audiobook coverage is lowest. The single regression is a very small market group where a concentrated catalog makes the production

Table 3: Relative Recall@10 lift from adding the video domain, by the signals a user already had.

User already had	Δ from +video
Music only	+9.6%
Music + audiobook	+1.3%
Music + podcast	∼0%
Music + podcast + audiobook	∼0%

Table 4: Relative Recall@10 gain by marketplace: music+podcast+audiobook model vs. the strongest per-market reference (a conservative proxy for the production baseline; see text), and additional lift from adding video.

Market	vs. production	+video (Δ)
DE	+30.6%	+4.6%
US	+27.4%	+2.1%
CA/BR/MX (grouped)	+26.7%	+9.9%
UK	+20.0%	+8.6%
IT	+15.9%	+6.9%
FR	+7.7%	+4.8%
ES	+6.2%	+9.1%
JP	+6.2%	+0.5%
NZ/AUS (grouped)	−22.0%	+41.7%

baseline's popularity backoff unusually strong and the sample too small for reliable estimates; adding the video signal reverses this into a large gain.

6.5 Adaptive Fusion Behavior

The gate behaves as intended. It leans on video when podcast is missing and backs off when podcast already carries genre information. The audiobook signal acts as high-precision but low-reach, and video as broad-reach genre. This lines up with the coverage-driven gains above: the fusion adapts to each user instead of applying a fixed blend.

6.6 Qualitative Analysis

Table 5 contrasts each user's top recommendations before and after video is added, for anonymized users in different markets. A few things stand out. Locale-appropriate content stays put; a German-market user keeps getting German-language titles. Where a user's other signals were generic, video pulls in genre-adjacent titles that match likely viewing taste, such as science-fiction and fantasy. And for users who already had strong signals, the top ranks barely move: video refines the list without disrupting it.

7 DISCUSSION

Why unification helps. Casting recommendation as text-based dense retrieval helps in a few connected ways, each of them tied to search. Cold-start becomes workable, because scoring depends on content semantics instead of collaborative history, so unseen and long-tail audiobooks stay reachable. Behavior from different domains becomes comparable once it is written as text, so one encoder

Table 5: Anonymized top recommendations without vs. with the video signal (music+podcast+audiobook model vs. all signals). Titles are broadly published public works (no platform exclusives); users and markets are anonymized.

User	Without video (top-3)	With video (top-3)
User A (Market 1)	*Atomic Habits*; *Sapiens*; *Thinking, Fast and Slow*	*Project Hail Mary*; *Dune*; *Atomic Habits*
User B (Market 2)	*The Hobbit*; *The Name of the Wind*; *Mythos*	*A Game of Thrones*; *The Name of the Wind*; *The Hobbit*
User C (Market 3, DE)	*Die 1%-Methode*; *Der Herr der Ringe*; *Sapiens*	*Der Wüstenplanet*; *Die 1%-Methode*; *Der Herr der Ringe*

and one index cover all of them. And the shared space gives title-level transfer for free: a semantic match between a title a user consumed elsewhere and an audiobook is really a search-style content match doing recommendation work. Adding an explicit query tower would fit the same space, so a joint search-and-recommendation system is an extension of this design, not a rebuild. We do not evaluate a query workload in this paper; our claim is about the shared representation that makes such unification straightforward, and we leave an explicit query tower and its evaluation to future work.

Limitations. (1) Evaluation here is offline; an online A/B test is in progress and its results will be included in the camera-ready version. Online behavior may differ from offline estimates due to presentation and exploration effects, though the retrieval stage feeds a downstream ranker. (2) The text encoder is frozen; end-to-end fine-tuning could improve transfer at higher serving cost. (3) Serializing behavior to text discards order and intensity; sequence-aware encoders may capture recency and dosage better. (4) Text quality varies by language, which likely explains smaller gains in some non-English markets. (5) We report relative metrics only, so absolute quality cannot be read from this paper.

8 CONCLUSION

We presented a text-embedding retrieval approach to cold-start audiobook discovery. It writes item content and four behavioral domains as text, encodes them in one multilingual space, and fuses the domains with a learned gate. Against a single production baseline, it improves Recall@10 by +27.3% from music, podcast, and audiobook signals, and +32.8% once video is added, with per-cohort gains near +50% where an audiobook signal is present. Since the core operation is semantic retrieval, the same space that retrieves from behavior can retrieve from queries, which is what makes the method a bridge between search and recommendation. Next steps include an explicit query tower for joint search and recommendation, more auxiliary domains for title-level transfer, sequence-aware encoding of behavior, and online validation.

REFERENCES

[1] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep Neural Networks for YouTube Recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys)*. 191–198.

[2] Guangneng Hu, Yu Zhang, and Qiang Yang. 2018. CoNet: Collaborative Cross Networks for Cross-Domain Recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM)*. 667–676.

[3] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning Deep Structured Semantic Models for Web Search Using Clickthrough Data. In *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM)*. 2333–2338.

[4] Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen tau Yih. 2020. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 6769–6781.

[5] Yueqi Li et al. 2020. MetaSelector: Meta-Learning for Recommendation with User-Level Adaptive Model Selection. In *Proceedings of The Web Conference (WWW)*. 154–164.

[6] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H. Chi. 2018. 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 and Data Mining*. 1930–1939.

[7] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. 2464–2470.

[8] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP-IJCNLP)*. 3982–3992.

[9] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current Challenges and Visions in Music Recommender Systems Research. *International Journal of Multimedia Information Retrieval* 7, 2 (2018), 95–116.

[10] Ajit P. Singh and Geoffrey J. Gordon. 2008. Relational Learning via Collective Matrix Factorization. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 650–658.

[11] Aäron van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep Content-Based Music Recommendation. In *Advances in Neural Information Processing Systems (NeurIPS)*. 2643–2651.

[12] Maksims Volkovs, Guangwei Yu, and Tomi Poutanen. 2017. DropoutNet: Addressing Cold Start in Recommender Systems. In *Advances in Neural Information Processing Systems (NeurIPS)*. 4957–4966.

[13] Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2024. Multilingual E5 Text Embeddings: A Technical Report. *arXiv preprint arXiv:2402.05672* (2024).

[14] Ruoxi Wang, Bin Fu, Gang Fu, and Mingliang Wang. 2017. Deep & Cross Network for Ad Click Predictions. In *Proceedings of the ADKDD Workshop*.

[15] Xinyang Yi, Ji Yang, Lichan Hong, Derek Zhiyuan Cheng, Lukasz Heldt, Aditee Kumthekar, Zhe Zhao, Li Wei, and Ed H. Chi. 2019. Sampling-Bias-Corrected Neural Modeling for Large Corpus Item Recommendations. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys)*. 269–277.

[16] Hamed Zamani and W. Bruce Croft. 2018. Joint Modeling and Optimization of Search and Recommendation. In *Proceedings of the First Biennial Conference on Design of Experimental Search and Information Retrieval Systems (DESIRES)*. 36–41.

[17] Hamed Zamani and W. Bruce Croft. 2020. Learning a Joint Search and Recommendation Model from User-Item Interactions. *Proceedings of the 13th ACM International Conference on Web Search and Data Mining (WSDM)* (2020), 717–725.

[18] Cheng Zhao, Chenliang Li, and Cong Fu. 2019. Cross-Domain Recommendation via Preference Propagation GraphNet. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*. 2165–2168.

[19] Feng Zhu, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. 2021. Cross-Domain Recommendation: Challenges, Progress, and Prospects. *arXiv preprint arXiv:2103.01696* (2021).