

Multimodal LLMs in Action: Addressing Cold-Start Recommendation at Xiaohongshu, a Leading Chinese Social Media Platform

Guillaume Salha-Galvan¹

¹Shanghai Jiao Tong University, Shanghai, China

Abstract

This paper presents the invited talk of the same title, to be delivered by Guillaume Salha-Galvan during the Second International Workshop on Data Quality-Aware Multimodal Recommendation (DaQuaMRec 2026), co-located with the Twentieth ACM Conference on Recommender Systems (RecSys 2026), in Minneapolis, Minnesota, on September 28, 2026. This invited talk primarily draws on the RecSys 2026 Industry Track paper *IDProxy: CTR Prediction with Multimodal LLMs for Cold-Start Recommendation at Xiaohongshu*.

Keywords

Multimodal Large Language Model, Cold-Start Recommendation, Click-Through Rate Prediction.

Multimodal LLMs for Cold-Start Recommendation at Xiaohongshu

Xiaohongshu, also known as RedNote, is a leading Chinese content-driven platform with over 300 million monthly active users. The service integrates social networking, content discovery, and e-commerce, with users primarily engaging with short videos and posts across a wide range of topics [1]. Given the massive volume of user-generated content, Xiaohongshu relies heavily on large-scale recommender systems to mitigate information overload and surface the most relevant content to each user, a key factor for engagement and retention [2, 3, 4]. In addition to organic content, Xiaohongshu also delivers personalized advertisements, in which ads are ranked and displayed according to users' interests.

Both recommendation and advertising at Xiaohongshu leverage click-through rate (CTR) prediction models to rank items, such as posts or ads, for each user [5, 6, 7]. These models predominantly rely on item ID embeddings that capture collaborative signals [7, 8]. However, learning reliable ID embeddings requires sufficient interaction history, making such models less effective for new items. This issue, known as the item cold-start problem [9, 10, 11, 12], is particularly severe in an industrial setting at Xiaohongshu's scale, where large volumes of new items are continuously uploaded and must be served immediately to ensure a high-quality user experience.

Exploiting multimodal content signals, such as text and images associated with items, has emerged as an effective way to learn item representations without interaction data [9, 13, 14]. Recent approaches have explored aligning such content representations with collaborative embeddings, either through behavioral co-occurrence patterns [15, 16] or direct content-to-collaborative mapping [17, 18]. While these methods have achieved promising results, their effectiveness may be limited when transferred to industrial settings at Xiaohongshu's scale. In particular, behavioral representations can differ substantially from collaborative ID embeddings, while approaches validated on academic benchmarks with well-structured ID distributions may struggle with the irregular, non-clustered embedding distributions encountered in large-scale industrial systems. As discussed in this talk, an additional challenge lies in integrating such representations into existing ID-centric CTR models while reusing mature architectures, preserving existing ID knowledge and structural priors, remaining aligned with the CTR optimization

DaQuaMRec 2026: Second International Workshop on Data Quality-Aware Multimodal Recommendation, September 28, 2026, Minneapolis, MN, USA

✉ gsalhagalvan@sjtu.edu.cn (G. Salha-Galvan)

🌐 <https://guillaumesalhagalvan.com/> (G. Salha-Galvan)

🆔 0000-0002-2452-1041 (G. Salha-Galvan)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

objective, and avoiding additional deployment costs. This raises a fundamental question: how can we design an alignment mechanism that bridges multimodal semantic representations and collaborative ID embeddings while maximally reusing existing ID knowledge and structural priors?

This talk introduces IDProxy [19], a production-scale system deployed at Xiaohongshu to address these challenges. IDProxy leverages multimodal large language models (MLLMs) [20, 21] to generate proxy item embeddings from rich content signals for cold-start CTR prediction. Through a lightweight coarse-to-fine mechanism, these proxy embeddings are aligned with the collaborative embedding space and jointly optimized end-to-end with the ranking model under the CTR objective, enabling seamless reuse of the existing architecture and continual evolution within production pipelines. IDProxy was evaluated against Xiaohongshu’s highly optimized production CTR baseline through offline experiments and online A/B tests involving millions of users. The online evaluation covered both Content Feed and Display Ads on the Explore Feed, using engagement, user-experience, and business-oriented metrics.

Following the success of these tests, IDProxy has been fully deployed on Xiaohongshu’s Explore Feed for both Content Feed and Display Ads, where, at the time of writing, it serves hundreds of millions of users daily. Overall, results demonstrate the benefits of leveraging MLLMs for cold-start CTR prediction by aligning multimodal content representations with collaborative item ID embeddings at industrial scale. Beyond its production impact, this study highlights the importance of end-to-end refinement with CTR models while reusing their core components. The results also demonstrate the potential of MLLM hidden states as content representations. Finally, the talk will discuss current limitations of the approach and promising directions for future research.

Speaker Bio

Guillaume Salha-Galvan is an Associate Professor at Shanghai Jiao Tong University (SJTU). He is a faculty member of the SJTU Paris Elite Institute of Technology, an international institute established at SJTU in partnership with four leading French engineering schools: École Polytechnique, Mines Paris, Télécom Paris, and ENSTA. He is the founder and principal investigator of the Sino-French Data and AI Lab.

His research focuses on recommender systems, LLMs, and their real-world applications. He is a regular author and Program Committee member at RecSys, where three of his past papers were shortlisted for awards in 2020 [22], 2021 [12], and 2024 [23]. Several of his research contributions have also had significant real-world impact, notably by powering recommender systems used by millions of users.

His research is supported by several competitive grants and talent programs, including the NSFC Excellent Young Scientists Fund (Overseas), the Siyuan Young Scholar Award, the Shanghai Magnolia Talent Program, and the Xiaomi Young Scholar Award, for which he was the first non-Chinese recipient.

Prior to joining SJTU, he gained nearly a decade of industry experience, including serving as Director of Research at Kibo Ryoku and as a Research Scientist and Coordinator for Music Recommendation at Deezer. He received his Ph.D. in Computer Science from École Polytechnique in France.

Acknowledgments

This DaQuAMRec invited talk is primarily based on the collective work published in the paper *IDProxy: CTR Prediction with Multimodal LLMs for Cold-Start Recommendation at Xiaohongshu*, accepted for oral presentation in the Industry Track of RecSys 2026 [19]. Guillaume Salha-Galvan is grateful to all co-authors of this work, from Xiaohongshu Inc. and Fudan University, and acknowledges their essential contributions to the RecSys paper and, consequently, to this invited talk. He is especially grateful to Yubin Zhang from Xiaohongshu Inc., the first author and main investigator of the project, whose sustained efforts throughout the past year were instrumental in making this project possible.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

- [1] Xiaohongshu, <https://www.xiaohongshu.com/explore>, 2026.
- [2] S. Zhang, L. Yao, A. Sun, Y. Tay, Deep Learning Based Recommender System: A Survey and New Perspectives, *ACM Computing Surveys* 52 (2019) 1–38.
- [3] L. Briand, G. Salha-Galvan, W. Bendada, M. Morlon, V.-A. Tran, A Semi-Personalized System for User Cold Start Recommendation on Music Streaming Apps, in: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2021, pp. 2601–2609.
- [4] Y. Li, K. Liu, R. Satapathy, S. Wang, E. Cambria, Recent Developments in Recommender Systems: A Survey, *IEEE Computational Intelligence Magazine* 19 (2024) 78–95.
- [5] H. B. McMahan, G. Holt, D. Sculley, M. Young, et al., Ad Click Prediction: A View from the Trenches, in: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2013, pp. 1222–1230.
- [6] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li, K. Gai, Deep Interest Network for Click-Through Rate Prediction, in: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2018, pp. 1059–1068.
- [7] W. Zhang, J. Qin, W. Guo, R. Tang, X. He, Deep Learning for Click-Through Rate Estimation, in: *Proceedings of the 30th International Joint Conference on Artificial Intelligence*, 2021, pp. 4695–4703.
- [8] H. Papadakis, A. Papagrigoriou, C. Panagiotakis, E. Kosmas, P. Fragopoulou, Collaborative Filtering Recommender Systems Taxonomy, *Knowledge and Information Systems* 64 (2022) 35–74.
- [9] W. Zhang, Y. Bei, L. Yang, H. P. Zou, P. Zhou, A. Liu, Y. Li, H. Chen, J. Wang, Y. Wang, et al., Cold-Start Recommendation Towards the Era of Large Language Models (LLMs): A Comprehensive Survey and Roadmap, *arXiv preprint arXiv:2501.01945* (2025).
- [10] W. Zhang, C. Wu, X. Li, Y. Wang, K. Dong, Y. Wang, X. Dai, X. Zhao, H. Guo, R. Tang, LLMTreeRec: Unleashing the Power of Large Language Models for Cold-Start Recommendations, in: *Proceedings of the 31st International Conference on Computational Linguistics*, 2025, pp. 886–896.
- [11] J. Wang, H. Lu, J. Caverlee, E. H. Chi, M. Chen, Large Language Models as Data Augmenters for Cold-Start Item Recommendation, in: *Companion Proceedings of the 2024 ACM Web Conference*, 2024, pp. 726–729.
- [12] G. Salha-Galvan, R. Hennequin, B. Chapus, V.-A. Tran, M. Vazirgiannis, Cold Start Similar Artists Ranking with Gravity-Inspired Graph Autoencoders, in: *Proceedings of the 15th ACM Conference on Recommender Systems*, 2021, pp. 443–452.
- [13] A. Lopez-Avila, J. Du, A Survey on Large Language Models in Multimodal Recommender Systems, *arXiv preprint arXiv:2505.09777* (2025).
- [14] X. Pan, Y. Chen, C. Tian, Z. Lin, J. Wang, H. Hu, W. X. Zhao, Multimodal Meta-Learning for Cold-Start Sequential Recommendation, in: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022, pp. 3421–3430.
- [15] X. Luo, J. Cao, T. Sun, J. Yu, R. Huang, W. Yuan, H. Lin, Y. Zheng, S. Wang, Q. Hu, et al., QARM: Quantitative Alignment Multi-Modal Recommendation at Kuaishou, in: *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, 2025, pp. 5915–5922.
- [16] D. Zhang, C. Fu, Z. Nie, J. Liu, W. Guan, Y. Gao, J. Song, P. Wang, J. Xu, B. Zheng, MOON: Generative MLLM-based Multimodal Representation Learning for E-Commerce Product Understanding, in: *Proceedings of the 19th ACM International Conference on Web Search and Data Mining*, 2026, pp. 924–933.
- [17] O. Barkan, N. Koenigstein, E. Yogev, O. Katz, CB2CF: a Neural Multiview Content-to-Collaborative Filtering Model for Completely Cold Item Recommendations, in: *Proceedings of the 13th ACM Conference on Recommender Systems*, 2019, pp. 228–236.
- [18] Y. Wei, X. Wang, Q. Li, L. Nie, Y. Li, X. Li, T.-S. Chua, Contrastive Learning for Cold-Start Recommendation, in: *Proceedings of the 29th ACM International Conference on Multimedia*, 2021, pp. 5382–5390.
- [19] Y. Zhang, H. Xu, G. Salha-Galvan, R. Han, F. Xiao, Y. Huang, L. Lin, Y. Luo, Y. Hu, IDProxy:

- CTR Prediction with Multimodal LLMs for Cold-Start Recommendation at Xiaohongshu, in: Proceedings of the 20th ACM Conference on Recommender Systems, 2026.
- [20] Y. Ye, Z. Zheng, Y. Shen, T. Wang, H. Zhang, P. Zhu, R. Yu, K. Zhang, H. Xiong, Harnessing Multimodal Large Language Models for Multimodal Sequential Recommendation, in: Proceedings of the 39th AAAI Conference on Artificial Intelligence, volume 39, 2025, pp. 13069–13077.
 - [21] Z. Chen, J. Wu, W. Wang, W. Su, G. Chen, S. Xing, M. Zhong, Q. Zhang, X. Zhu, L. Lu, et al., InternVL: Scaling up Vision Foundation Models and Aligning for Generic Visual-Linguistic Tasks, in: Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 24185–24198.
 - [22] W. Bendada, G. Salha, T. Bontempelli, Carousel Personalization in Music Streaming Apps with Contextual Bandits, in: Proceedings of the 14th ACM Conference on Recommender Systems, 2020, pp. 420–425.
 - [23] V.-A. Tran, G. Salha-Galvan, B. Sguerra, R. Hennequin, Transformers Meet ACT-R: Repeat-Aware and Sequential Listening Session Recommendation, in: Proceedings of the 18th ACM Conference on Recommender Systems, 2024, pp. 486–496.