

## RESEARCH ARTICLE

# Causal Transformer Networks for Counterfactual Reasoning in Large-Scale Recommendation Systems

Dawen Liang, Peng Cui, Wenjie Wang

Published: 2026-05-18 | FAIDS Vol. 1, No. 1 (2026)

**Abstract:** Modern recommendation systems suffer from popularity bias, filter bubbles, and spurious correlations that degrade long-term user satisfaction. We introduce CausalRec, a Transformer-based architecture that integrates structural causal models into the attention mechanism, enabling counterfactual reasoning at inference time: "Would the user have clicked this item if it were not promoted on the homepage?" Deployed in a 28-day A/B test on a major e-commerce platform (430 million daily active users), CausalRec increases 30-day user retention by 3.8%, reduces popularity bias Gini coefficient by 22%, and improves content diversity by 31% while maintaining gross merchandise value (GMV) parity.

## 1. Introduction

Recommendation systems drive over 35% of e-commerce revenue and 80% of content consumption on streaming platforms. However, training on observational interaction data creates a feedback loop: popular items receive more exposure, generating more clicks, further reinforcing their prominence — the "rich get richer" phenomenon. This popularity bias concentrates recommendations on a narrow subset of items, reducing marketplace fairness for long-tail sellers and degrading user experience through monotonous recommendations.

Causal inference provides a principled framework for disentangling genuine user preferences from confounding factors such as position bias, exposure bias, and promotional effects. However, integrating causal reasoning into large-scale production recommenders with billions of parameters and millisecond latency requirements remains an open challenge.

## 2. Method

CausalRec augments a standard Transformer recommendation model with causal attention layers that encode a structural causal model as an attention mask. The causal graph encodes known confounders: item position → click, promotion status → click, user demographics → item preference. During training, the model learns both factual and

counterfactual attention distributions via a variational inference objective. At inference time, the do-calculus intervention operator removes confounding edges, producing debiased relevance scores.

### 3. Online A/B Test Results

The 28-day A/B test allocated 5% of traffic (21.5M users) to CausalRec. Primary metrics show significant improvements in long-term engagement while maintaining revenue neutrality. The 3.8% improvement in 30-day retention is the largest single-model gain in the platform's history, demonstrating that debiased recommendations improve user satisfaction.

**Table 1. Online A/B test results: CausalRec vs. production baseline (28 days, 21.5M users per group)**

Metric	Baseline	CausalRec	Relative Change	p-value
CTR	4.82%	4.91%	+1.9%	0.003
30-day Retention	62.1%	64.5%	+3.8%	<0.001
GMV per User	¥147.2	¥146.8	-0.3%	0.42
Gini Coefficient	0.78	0.61	-22%	<0.001
Category Diversity	3.2	4.2	+31%	<0.001

### 4. Conclusions

CausalRec demonstrates that causal reasoning can be practically integrated into production-scale recommendation systems with measurable benefits in user retention, content diversity, and marketplace fairness. The approach maintains revenue neutrality while significantly improving the long-term health of the recommendation ecosystem.

### References

- [1] Pearl, J. Causality: Models, Reasoning, and Inference, 2nd ed.; Cambridge University Press: Cambridge, 2009.
- [2] Schnabel, T.; Swaminathan, A.; Singh, A.; Chandak, N.; Joachims, T. Recommendations as Treatments. ICML 2016.
- [3] Wang, W.; Feng, F.; He, X.; Nie, L.; Chua, T.-S. Deconfounded Recommendation for Alleviating Bias Amplification. KDD 2021.
- [4] Zheng, Y.; Gao, C.; Li, X.; He, X.; Jin, D.; Li, Y. Disentangling User Interest and Conformity for Recommendation with Causal Embedding. WWW 2021.
- [5] Chen, J.; Dong, H.; Wang, X.; Feng, F.; Wang, M.; He, X. Bias and Debias in Recommender System: A Survey and Future Directions. ACM TOIS 2023, 41, 1-39.