Recent Challenges and Advances in Industrial Recommender Systems

Liangjie Hong

Invited Talk, **Booking.Com**

LinkedIn

December 2nd, 2020

Recommender Systems, Done Deal?

Agenda

1 Industrial Recommender Systems with Their Ecosystems

2 Case I: Understanding The Interplay between Recommender Systems and Search Systems

3 Case II: The Journey of Long-term Engagement Optimization

Agenda

1 **Industrial Recommender Systems with Their Ecosystems**

2 Case I: Understanding The Interplay between Recommender Systems and Search Systems

3 Case II: The Journey of Long-term Engagement Optimization

Recommender Systems are Matrix Factorization and Neural Models.

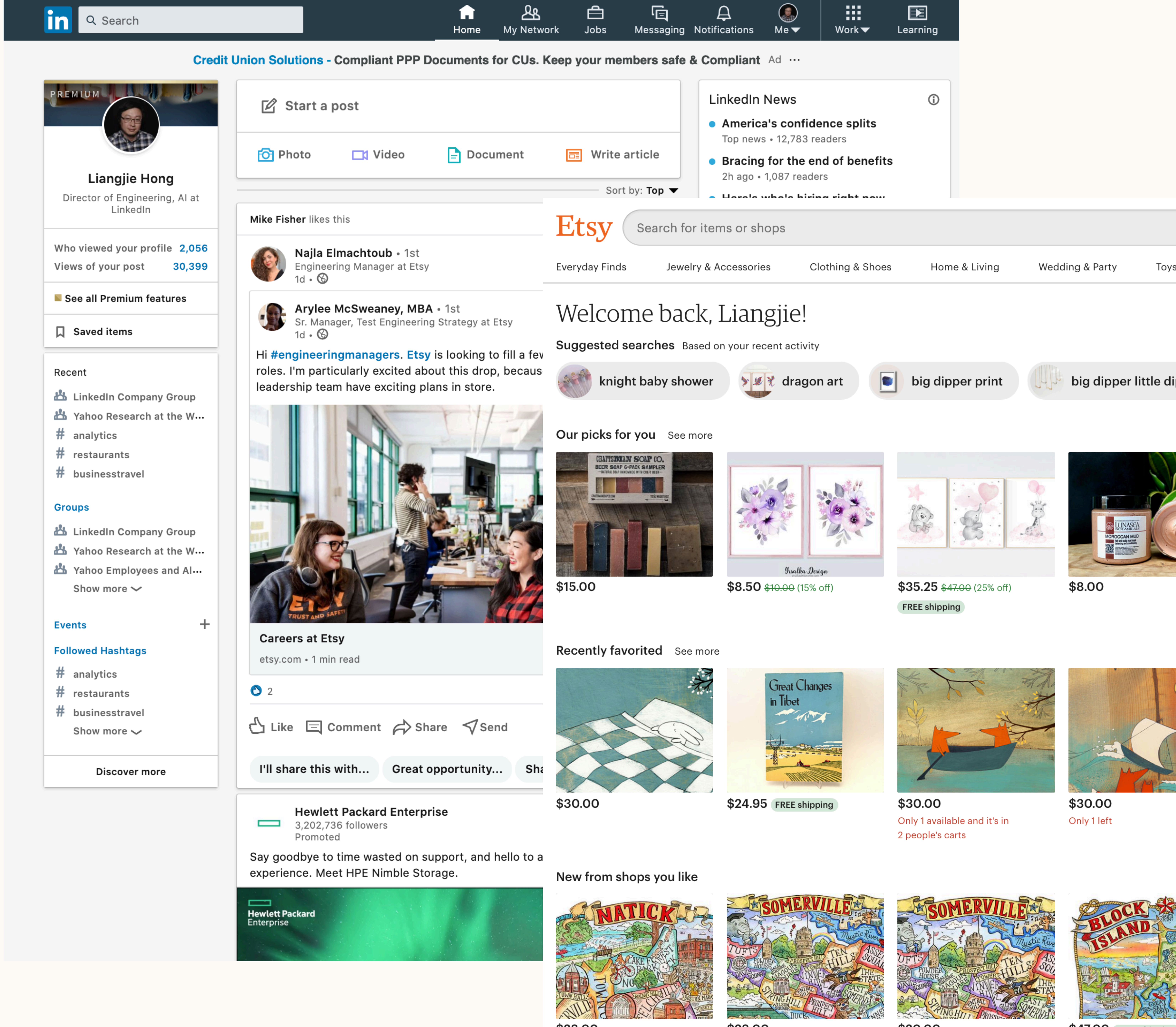
Industrial Recommender Systems are more than Matrix Factorization and Neural Models.

Industrial Recommender Systems are more than Matrix Factorization and Neural Models.

Recommender systems are parts of a larger **ecosystem**, serving the overall **product strategy** for a **business**.

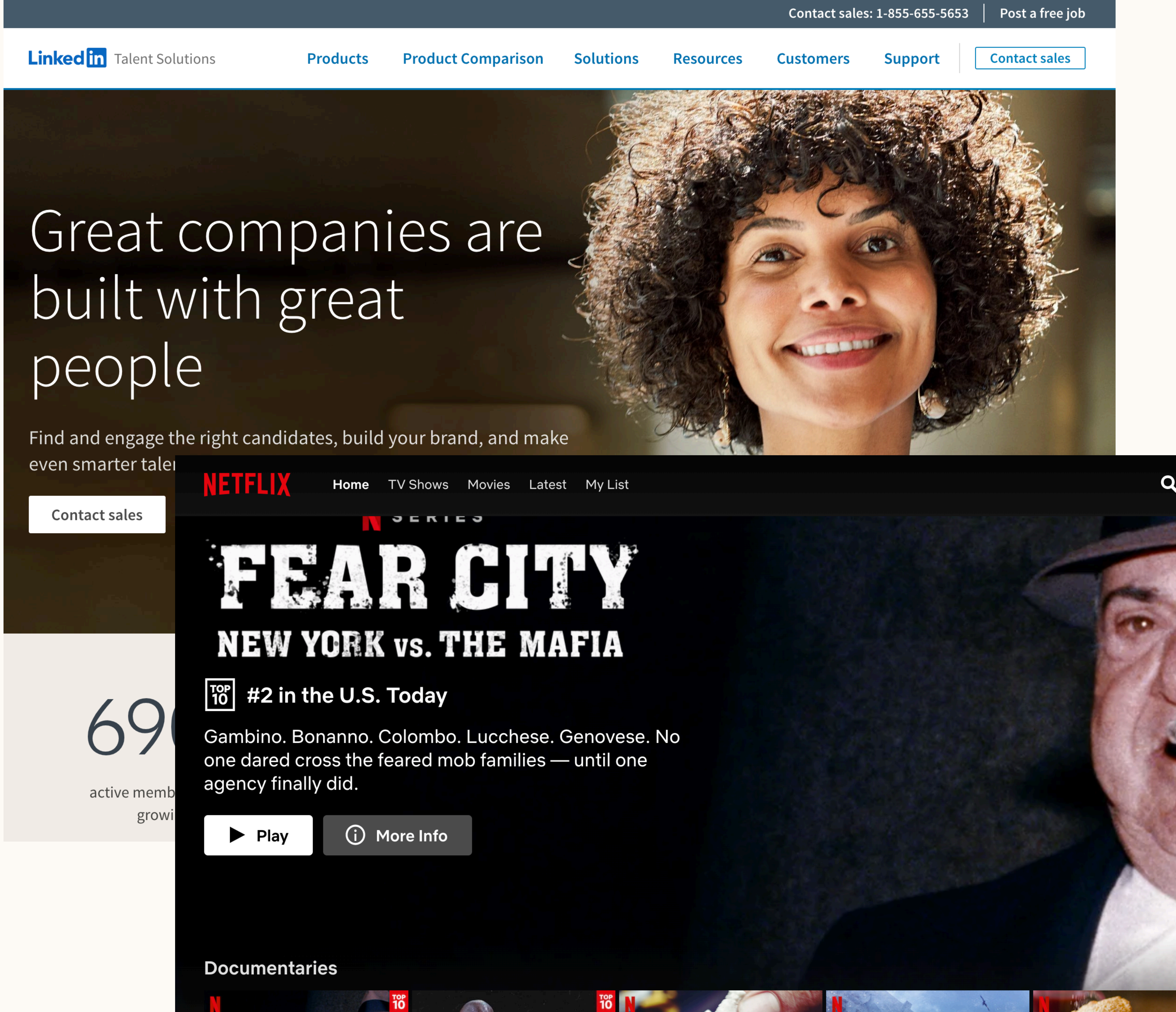
Modules and Pages

Down-funnel decisions (e.g., subscriptions, purchases, conversions and etc.) are outcomes of interactions of multiple modules and pages across sessions.



Product and Business Needs

Recommender systems have to fit into the overall product and business strategy. Sometimes, it is hard to link the success of a module to the overall business.



A Collection of Software Services.

In large-scale recommender systems, any of data, features, tools, pipelines, online servings and etc. can be owned and operated by different teams with different goals.

Industrial Recommender Systems are more than Matrix Factorization and Neural Models.

Recommender systems are parts of a larger **ecosystem**, serving the overall **product strategy** for a **business**.

- A typical web application has many pages with different types of modules to serve the product need. The interplay between these modules is complex.
- A product with recommender systems needs to serve a set of business purposes, including long-term and short-term goals.
- A recommender system is a product of large-scale engineering practices involving many teams and organizations. End-to-end optimization is challenging.

Industrial Recommender Systems are more than Matrix Factorization and Neural Models.

Recommender systems are parts of a larger **ecosystem**, serving the overall **product strategy** for a **business**.

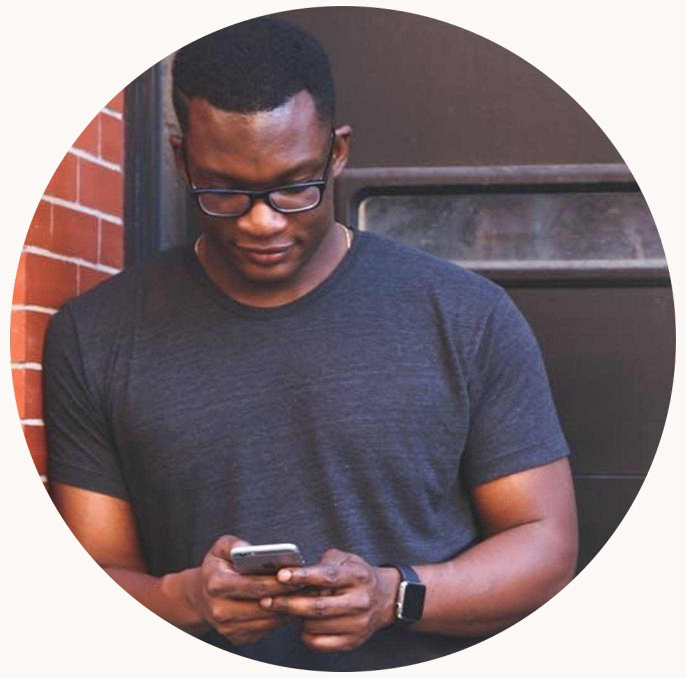
- **A typical web application has many pages with different types of modules to serve the product need. The interplay between these modules is complex.**
- **A product with recommender systems needs to serve a set of business purposes, including long-term and short-term goals.**
- A recommender system is a product of large-scale engineering practices involving many teams and organizations. End-to-end optimization is challenging.

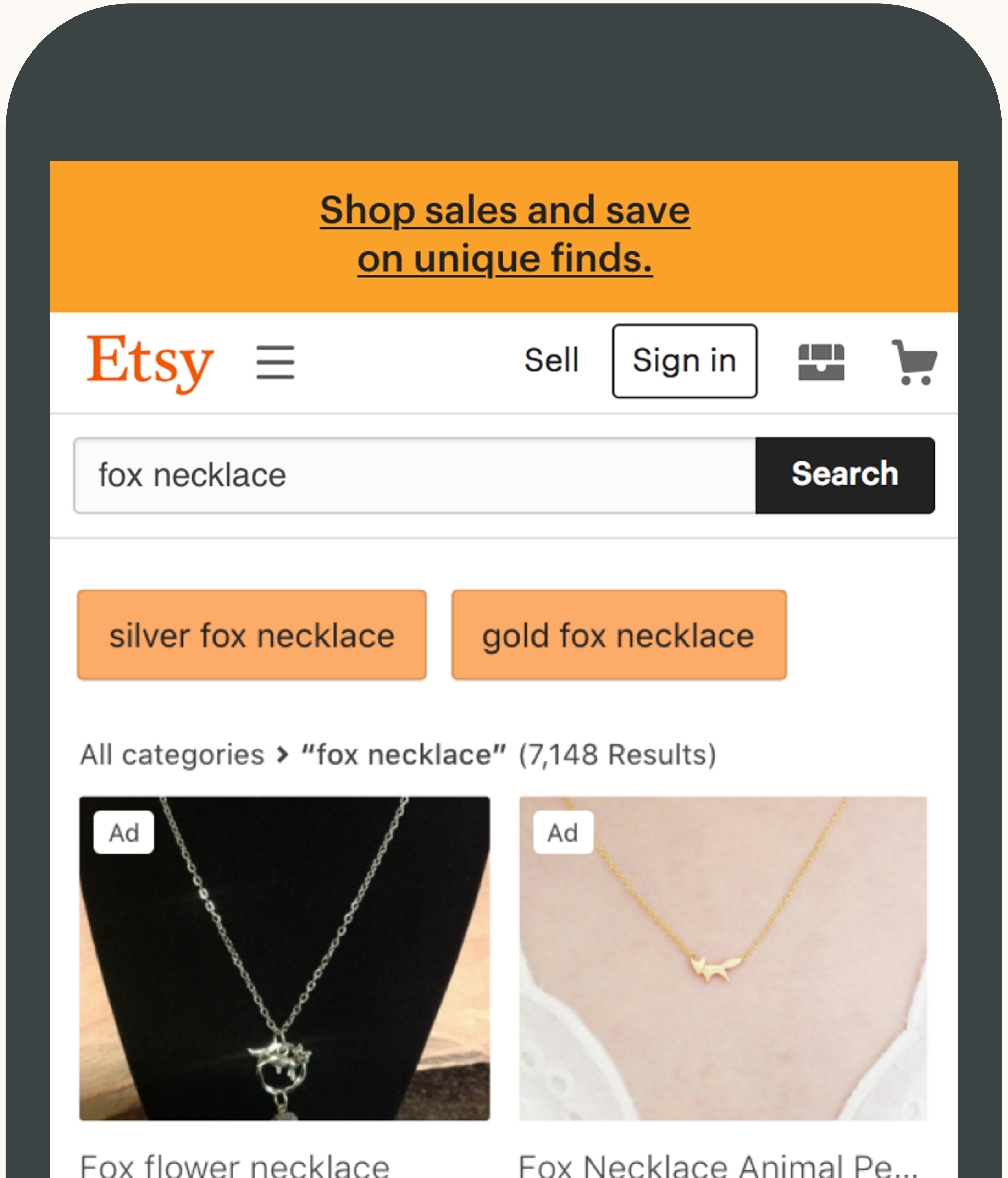
Agenda

1 Industrial Recommender Systems with Their Ecosystems

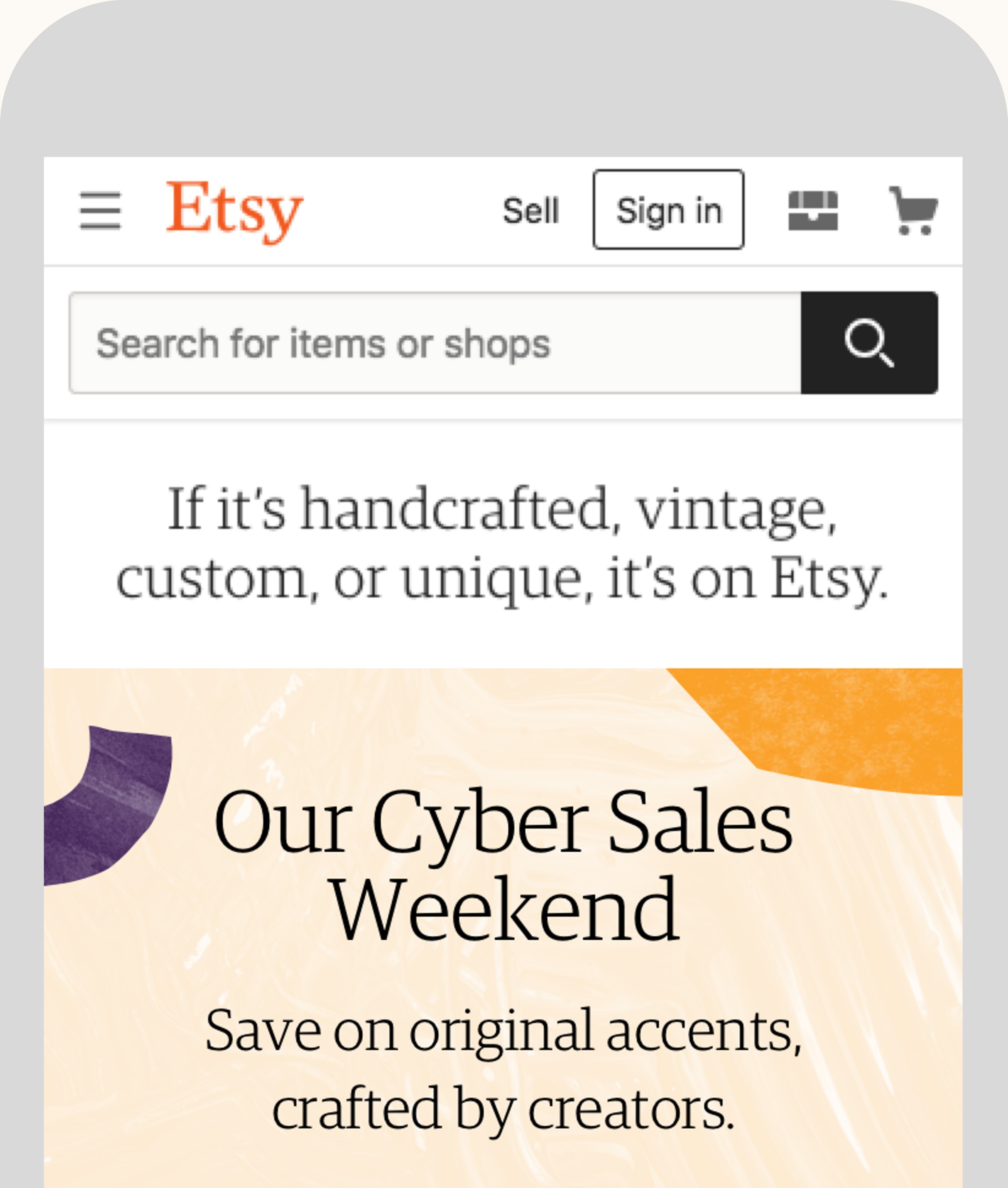
2 **Case I: Understanding The Interplay between Recommender Systems and Search Systems**

3 Case II: The Journey of Long-term Engagement Optimization





Shop sales and save
on unique finds.
Etsy
Sell
Sign in
fox necklace
Search
silver fox necklace
gold fox necklace
All categories > "fox necklace" (7,148 Results)
Ad
Ad
Fox flower necklace
Fox Necklace Animal Pe...



Etsy
Sell
Sign in
Search for items or shops
If it's handcrafted, vintage, custom, or unique, it's on Etsy.
Our Cyber Sales Weekend
Save on original accents, crafted by creators.

Challenges

- We can observe changes of site-wide metrics (e.g., site-wide conversion rate, revenue, success search sessions and etc.) via online A/B tests.
- We can observe changes of module-wise metrics (e.g., CTR, dwell time, swipes and etc.) via online A/B tests.
- We cannot easily observe or obtain contributions from each module to site-wide metrics.

Example I:
An A/B Test Result for A New Recommendation Algorithm

	% Change
Recommendation Clicks	+5%
Search Clicks	-3%
Revenue	~

Example I:
An A/B Test Result for A New Recommendation Algorithm

	% Change
Recommendation Clicks	+5%
Search Clicks	-3%
Revenue	~

Notes:

- Improvements might come as a result of a series of A/B testing results.
- Not shipping early corner-stone results might lead to a sub-optimal user experience in a long run.
- Shipping placebo results might lead to a sub-optimal user experience in a long run.

Example II:
An A/B Test Result for A New Recommendation Algorithm

	% Change
Recommendation Clicks	-10%
Search Clicks	+5%
Revenue	+1%

Example II:
An A/B Test Result for A New Recommendation Algorithm

	% Change
Recommendation Clicks	-10%
Search Clicks	+5%
Revenue	+1%

Notes:

- Deteriorations might come as a result of a series of A/B testing results.
- Once damage is done, it might impact machine learning algorithms in many ways (e.g., training bias).
- Not shipping early corner-stone results might lead to a sub-optimal user experience in a long run.
- Shipping placebo results might lead to a sub-optimal user experience in a long run.

We need to understand the **interplay** between recommendation and search modules as well as their whole **ecosystem** to create a **coherent** user experience and optimize user engagement.

- Opportunity 1:
  Understand experimental results while multiple teams work on different recommendation and search modules.

- Opportunity 2:
  Develop and implement strategies to improve multiple modules and possibly optimize overall user engagement.

- Opportunity 3:
  Develop machine learning models to directly optimize user engagement from a whole user journey perspective.

We need to understand the **interplay** between recommendation and search modules as well as their whole **ecosystem** to create a **coherent** user experience and optimize user engagement.

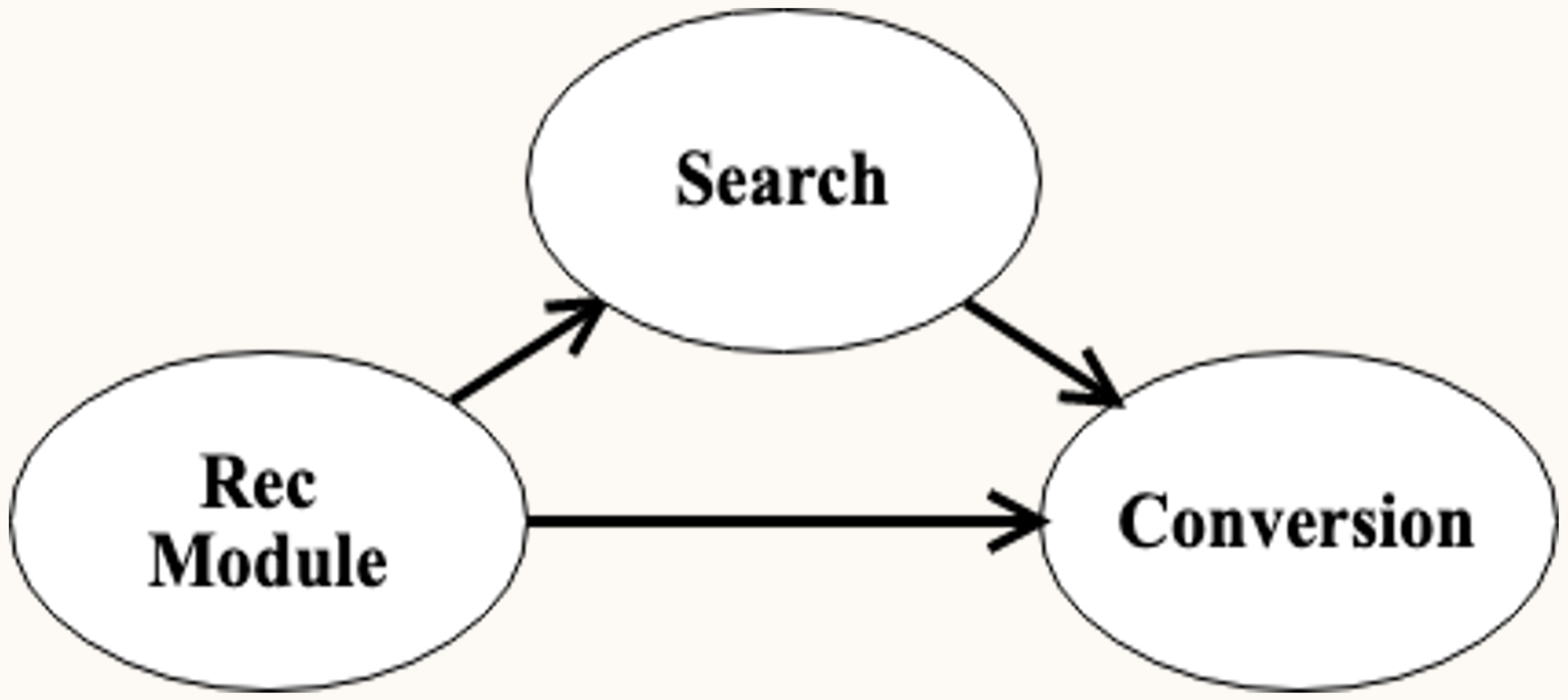
X. Yin and L. Hong. **The Identification and Estimation of Direct and Indirect Effects in A/B Tests through Causal Mediation Analysis**. KDD 2019.

- **Opportunity 1:**
  Understand experimental results while multiple teams work on different recommendation and search modules.

- Opportunity 2:
  Develop and implement strategies to improve multiple modules and possibly optimize overall user engagement.

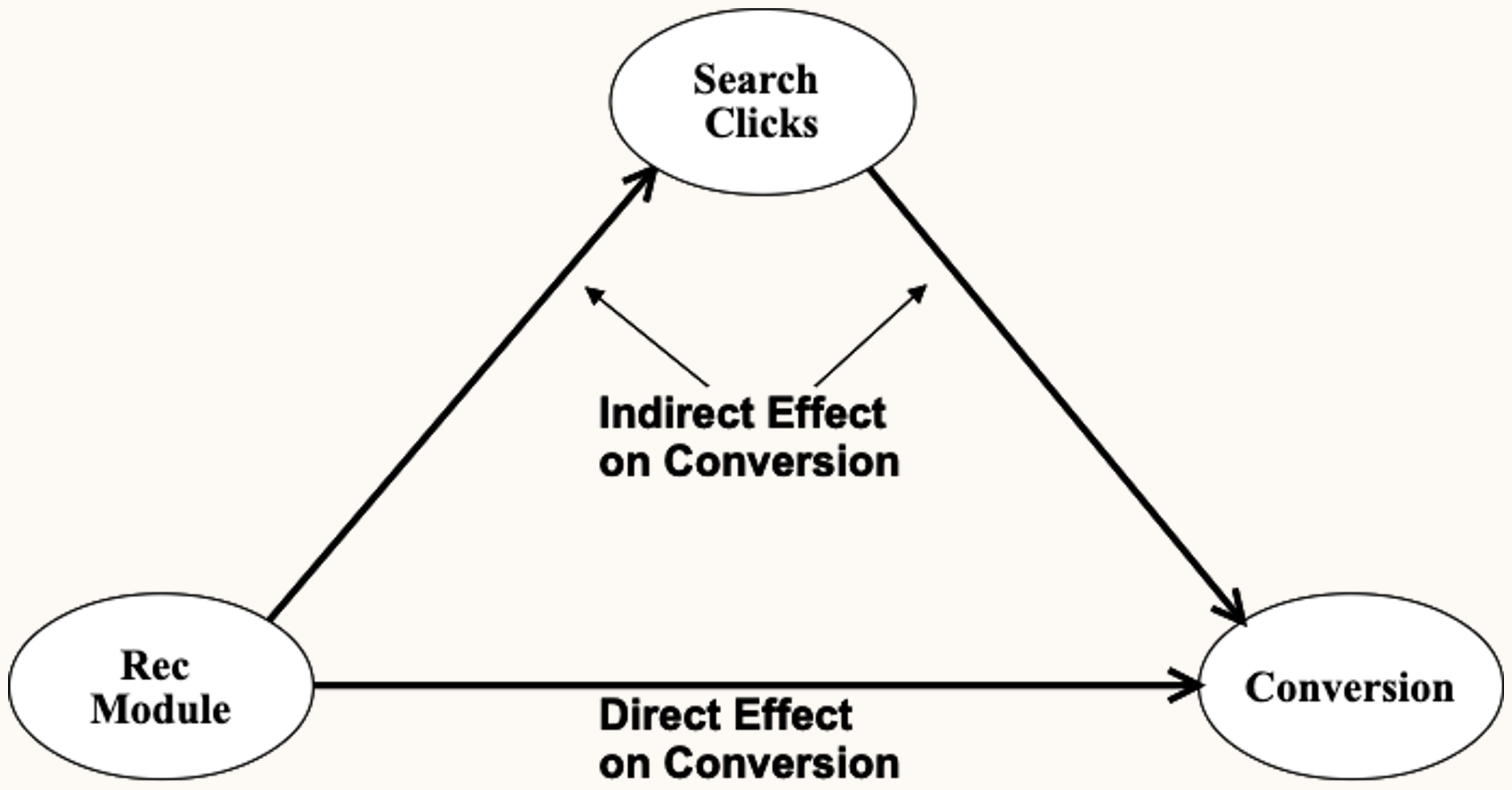
- Opportunity 3:
  Develop machine learning models to directly optimize user engagement from a whole user journey perspective.

A product change could ***induce*** changes in user interaction with other products.

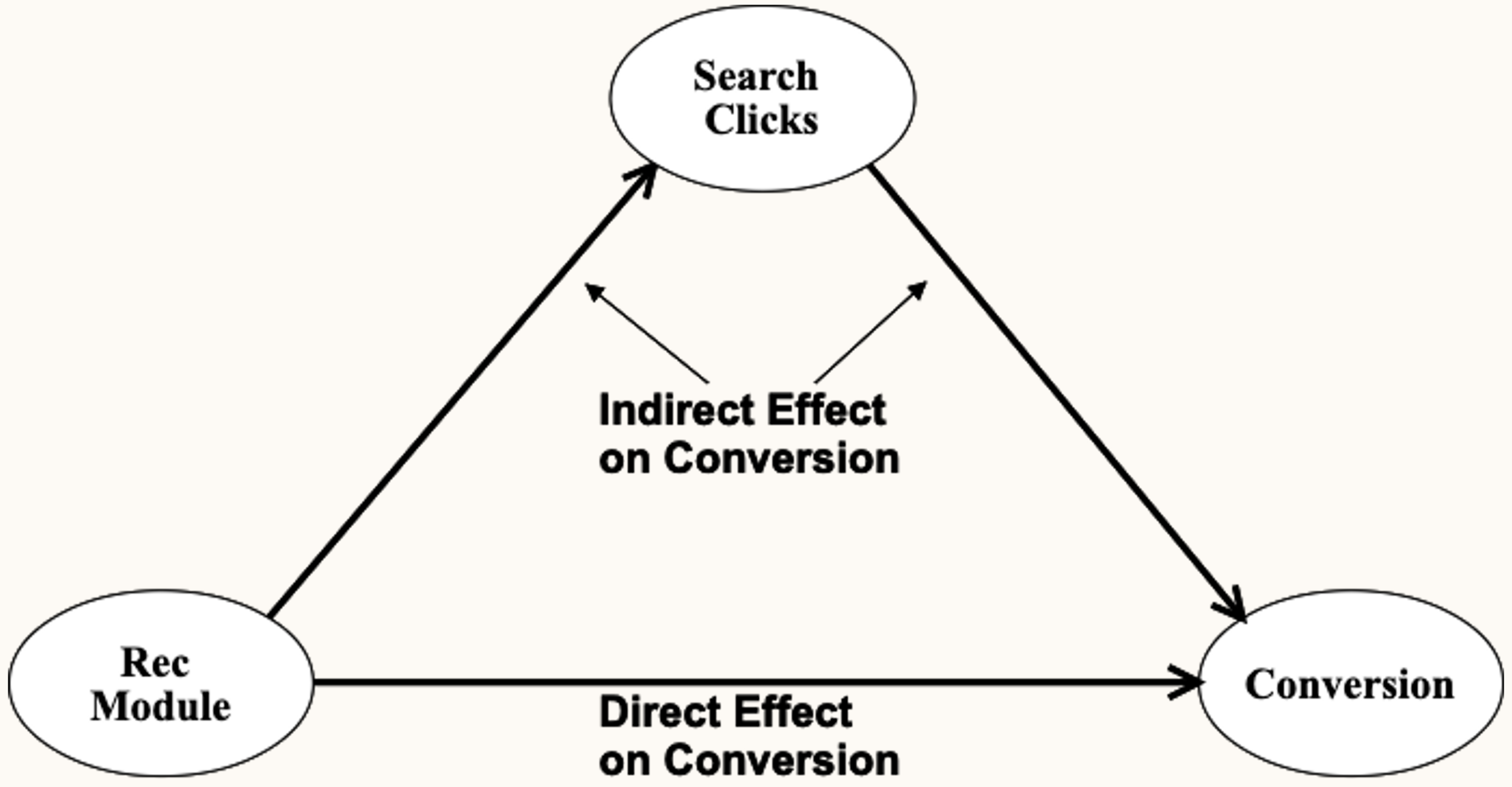


- An improved recommendation module could effectively suggest items that satisfy users' needs so that users don't need to search as much as usual.
- The overall performance of an improved recommendation module could be cannibalized by the induced reduction of user engagement in search.
- The performance of search could be cannibalized by an improved recommendation module.

Common Solution
Splitting Average Treatment Effect (ATE) into Two Parts:
Direct Effect and Indirect Effect



Common Solution
Splitting Average Treatment Effect (ATE) into Two Parts: Direct Effect and Indirect Effect



Notes:

- Causal Mediation Analysis (CMA) is a formal statistical framework to conduct such analysis.
- Average Direct Effect (ADE) is the direct impact of new recommendations while keeping search behavior fixed.
- Average Causal Mediation Effect (ACME) is the impact of induced changes in search behavior due to changes in recommendation algorithm.

Common Solution
Splitting Average Treatment Effect (ATE) into Two Parts: Direct Effect and Indirect Effect

Notes:

- ATE, ADE and ACME has been studied extensively in the literature.
- Existing methodologies cannot be easily utilized due to violations of the key assumptions in the literature: no unmeasured causally-dependent mediator.
- A typical E-commerce site could have hundreds of web-pages and modules, and all of them could be mediators. It is difficult to measure all of them.
- We extended ADE and ACME to Generalized ADE (GADE) and Generalized ACME (GACME) respectively.
- It is easy to implement and only requires solving two linear regression equations simultaneously.
- Git Repo: https://github.com/xuanyin/causal-mediation-analysis-for-ab-tests

Case I:
RecSys Listing Page Same-Shop Experiment

	% Change	
Effect	Conversion Rate	GMV
GADE Direct Effect of the Change of Rec Module	**0.4959%***	0.1681%
GACME The Effect of the Induced Change of Search	**-0.2757%*****	**-0.4200%*****
ATE	0.2202%	-0.2518%

Notes:

1. % Change = Effect/Mean of Control
2. '***' $p<0.001$, '**' $p<0.01$, '*' $p<0.05$, '.' $p<0.1$. Two-tailed p-value is derived from z-test for H_0: the effect is zero, which is based on asymptotic normality.

Case II:
RecSys Listing Page Internal-Bottom Desktop Experiment

	% Change	
Effect	Conversion Rate	GMV
GADE Direct Effect of the Change of Rec Module	**0.3448%***	0.0659%
GACME The Effect of the Induced Change of Search	**-0.0570%**.	**-0.0926%**.
ATE	**0.2878%**.	-0.0267%

Notes:

1. % Change = Effect/Mean of Control
2. '***' $p<0.001$, '**' $p<0.01$, '*' $p<0.05$, '.' $p<0.1$. Two-tailed p-value is derived from z-test for H_0: the effect is zero, which is based on asymptotic normality.

Takeaways

- Understanding direct vs. indirect effects enables us to understand the competition between recommendation modules and search results; and give more informed decisions during roll-outs
- Develop better recommendation strategies such as suggesting items and categories not searched organically or diverse information shown in different surfaces.
- Develop better offline evaluation framework to incorporate both search and recommendation results.

Agenda

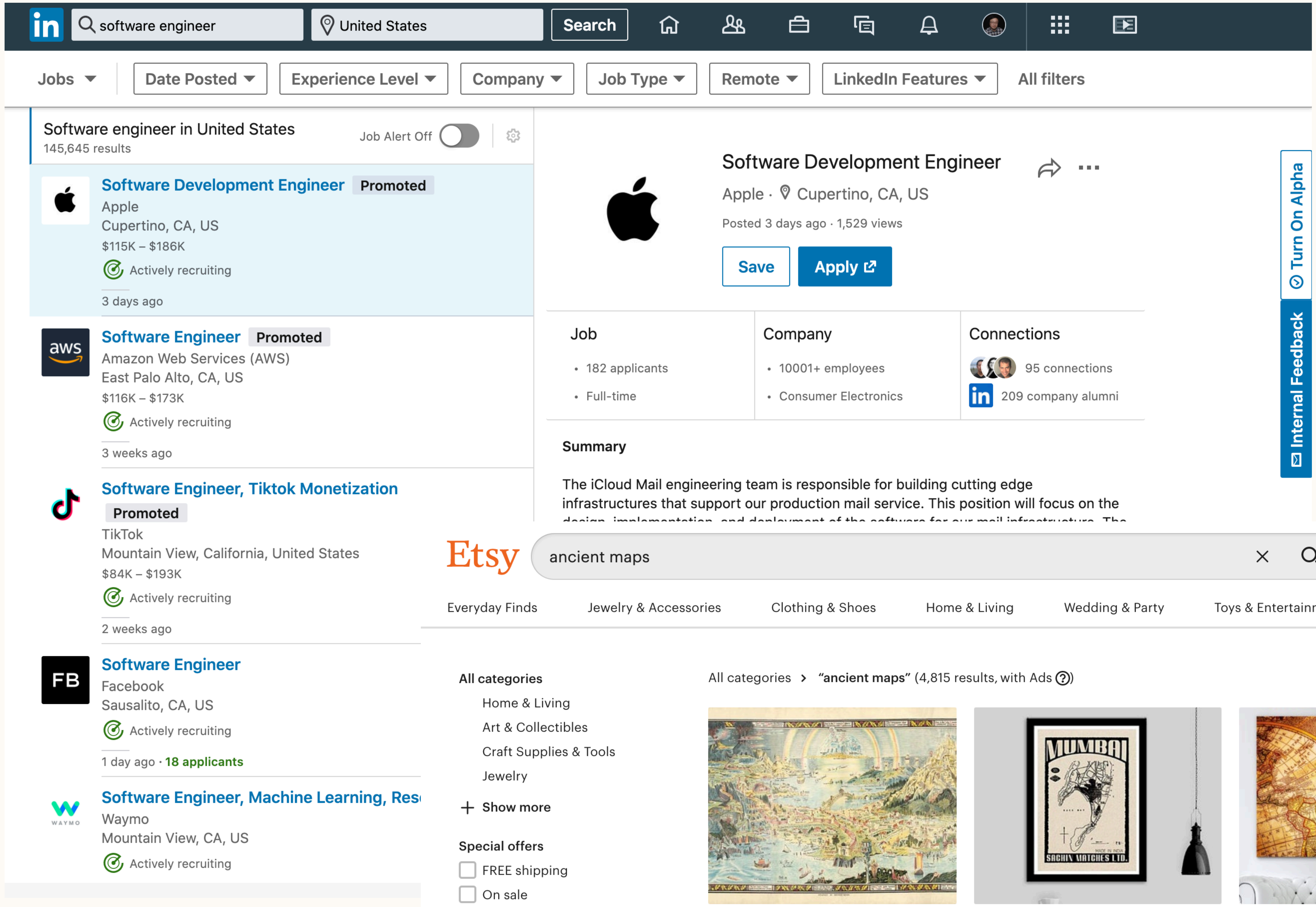
1 Industrial Recommender Systems with Their Ecosystems

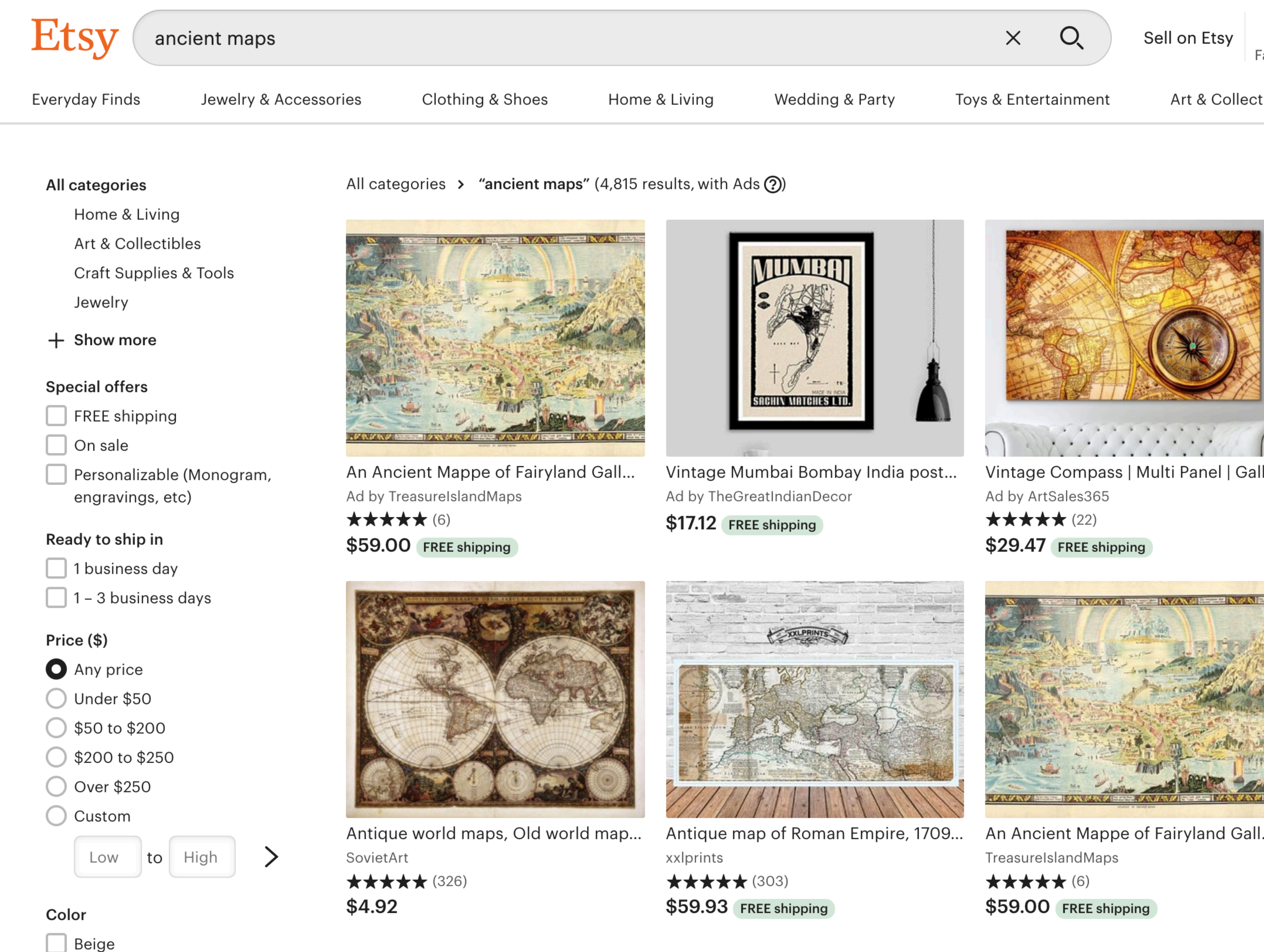
2 Case I: Understanding The Interplay between Recommender Systems and Search Systems

3 **Case II: The Journey of Long-term Engagement Optimization**

Measuring Product and Business Success

Business tends to not focus on immediate user engagement improvements but longer term successes (e.g., confirmed hires, GMS, user retention and etc.).





There are non-trivial gaps between what we could optimize and what we should optimize.

- Opportunity 1:
  Optimize short-term metrics and seek to establish relationships between short-term metrics and long-term metrics.
- Opportunity 2:
  Directly optimize long-term metrics.
- Opportunity 3:
  Accelerate experimentation and directly optimize long-term metrics.

There are non-trivial gaps between what we could optimize and what we should optimize.

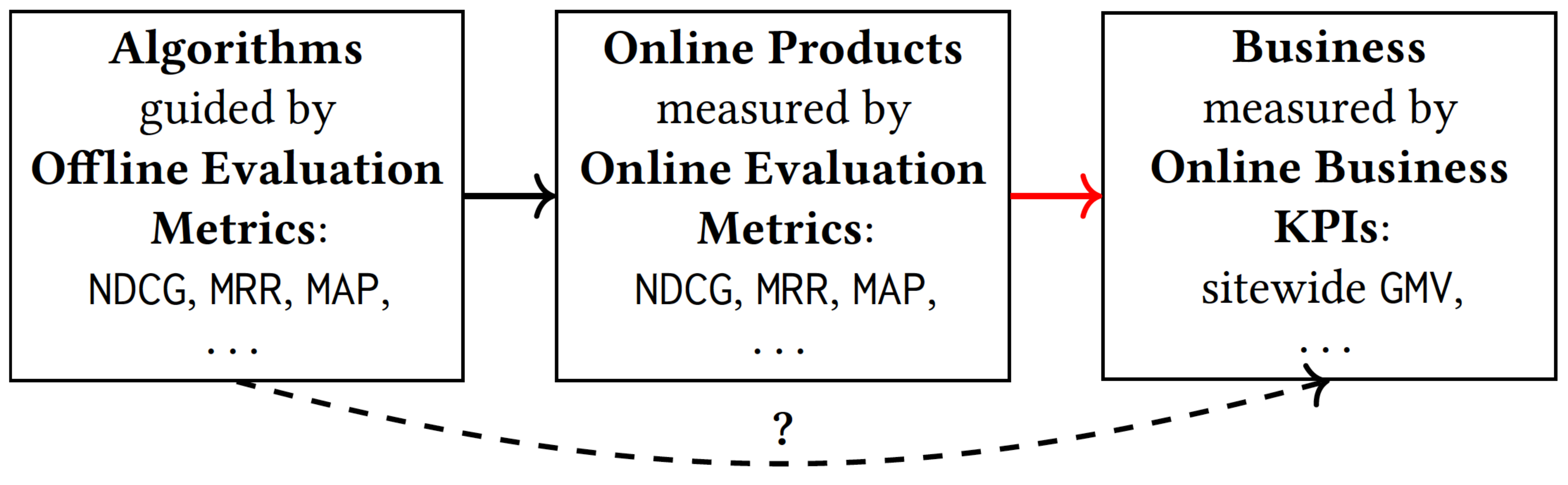
Z. Wang, X. Yin, T. L and L. Hong. **Causal Meta-Mediation Analysis: Inferring Dose-Response Function From Summary Statistics of Many Randomized Experiments**. KDD 2020.

- **Opportunity 1:**
  Optimize short-term metrics and seek to establish relationships between short-term metrics and long-term metrics.

- Opportunity 2:
  Directly optimize long-term metrics.

- Opportunity 3:
  Accelerate experimentation and directly optimize long-term metrics.

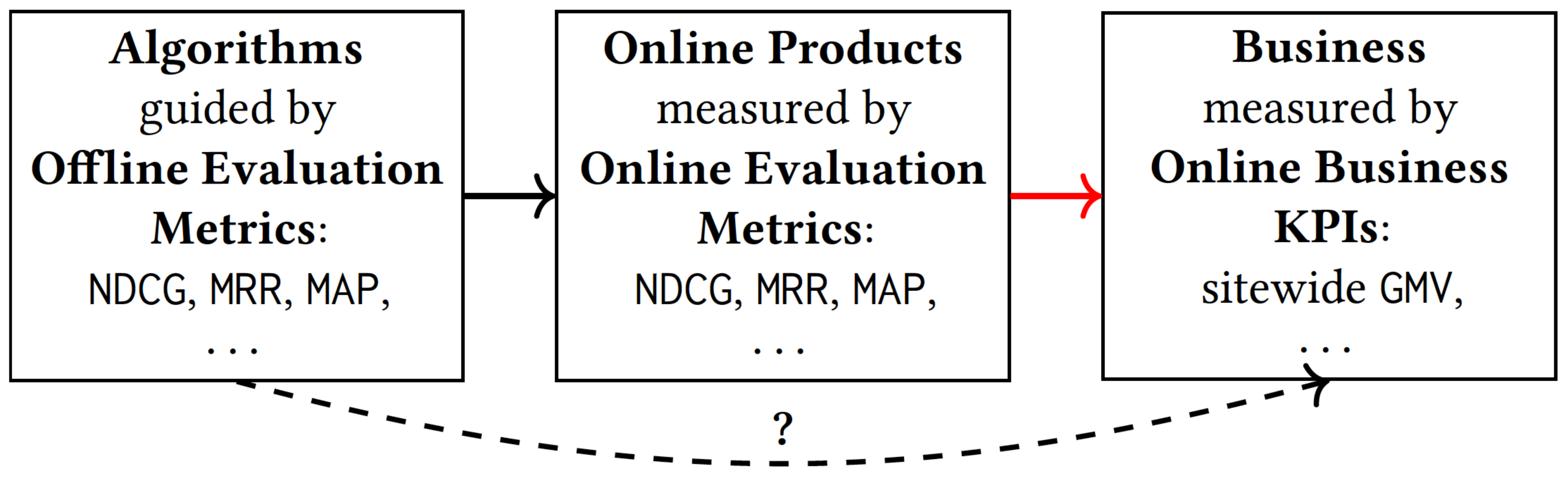
Causal Meta-Mediation Analysis
The Causal Path from Algorithms to Business



Notes:

- First part (the black arrow): counterfactual estimators of offline evaluation metrics to bridge the inconsistency between changes of offline and online evaluation metrics.
- Second part (the red arrow): the causality between online products (assessed by online evaluation metrics) and the business (assessed by online business KPIs).
  e.g. how business KPIs would change for a 10% increase in an online evaluation metric.

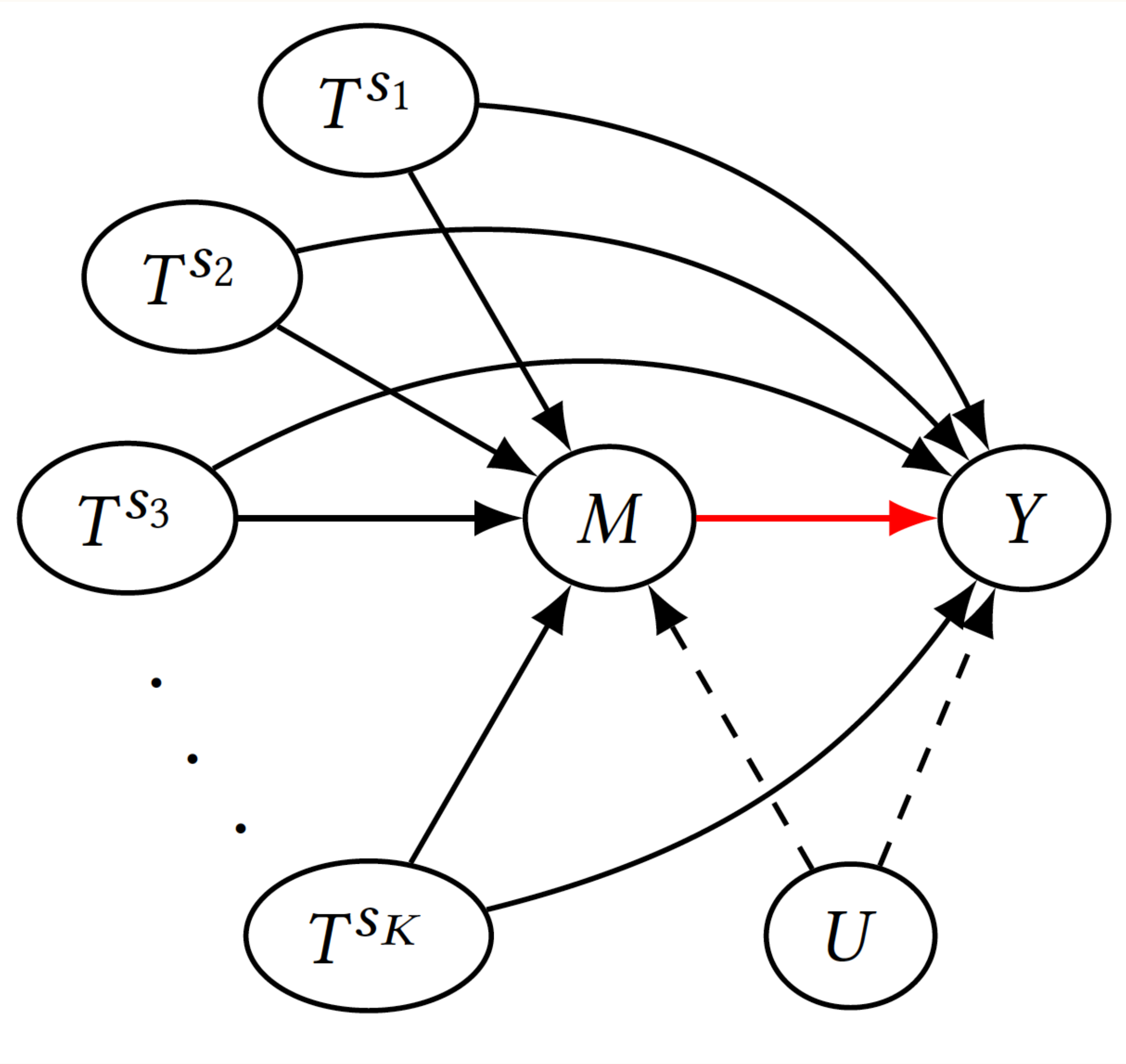
Causal Meta-Mediation Analysis
The Causal Path from Algorithms to Business



Key Ideas:

- We model the causality between online evaluation metrics and business KPIs by *dose-response function* (DRF) in potential outcome framework.
- Instead of conducting online tests, we use results from historical A/B experiments to conduct Meta-Analysis.
- Online evaluation metrics could be mediators that (partially) transmit causal effects of treatments on business KPIs in experiments where treatments are not necessarily algorithm-related.

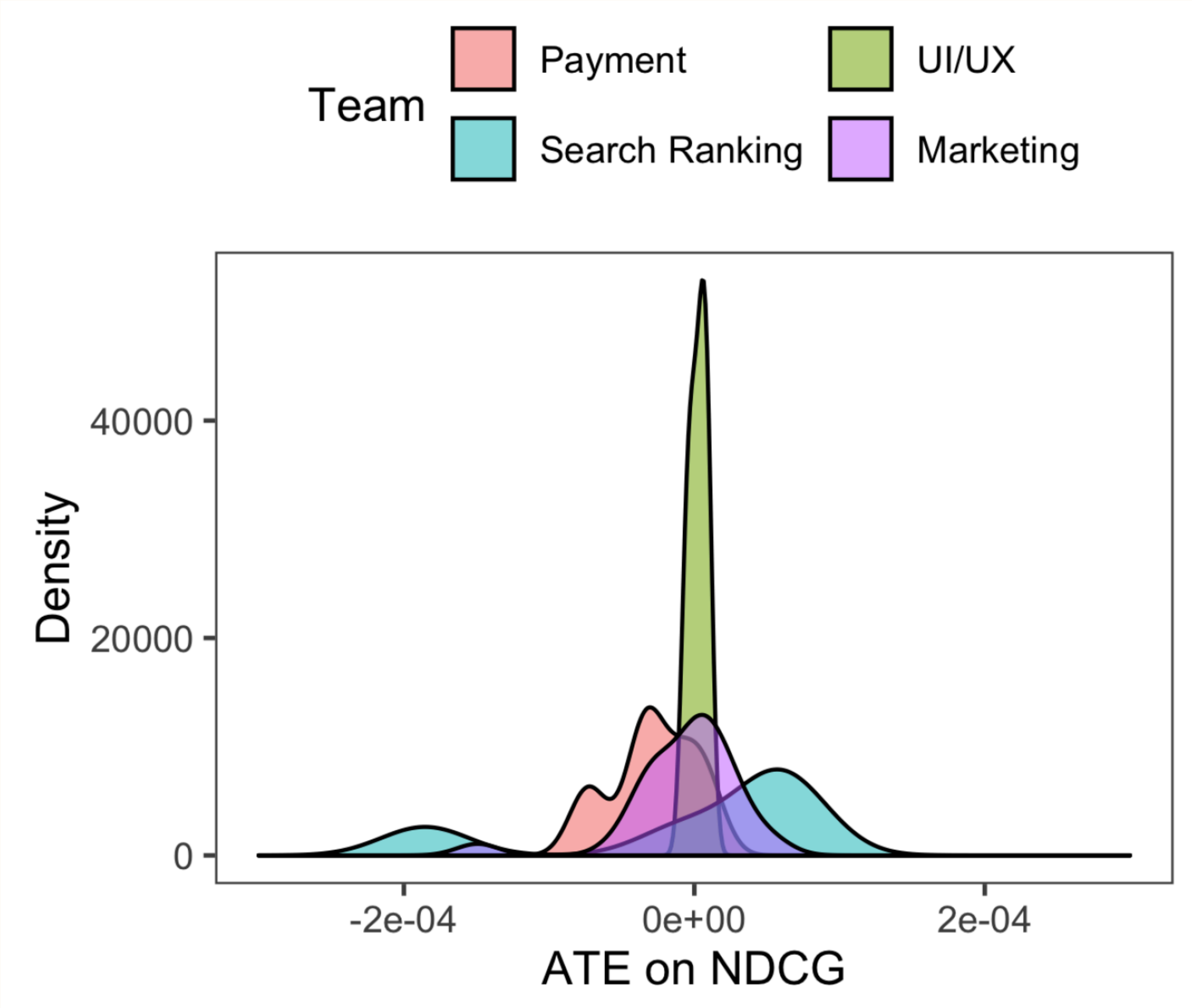
Causal Meta-Mediation Analysis
The Causal Path from Algorithms to Business



Key Ideas:

- Ts are treatments; M is a mediator; Y is a outcome; U is unobserved and unmeasured.
- M is online evaluation metric. Y is online business KPI.

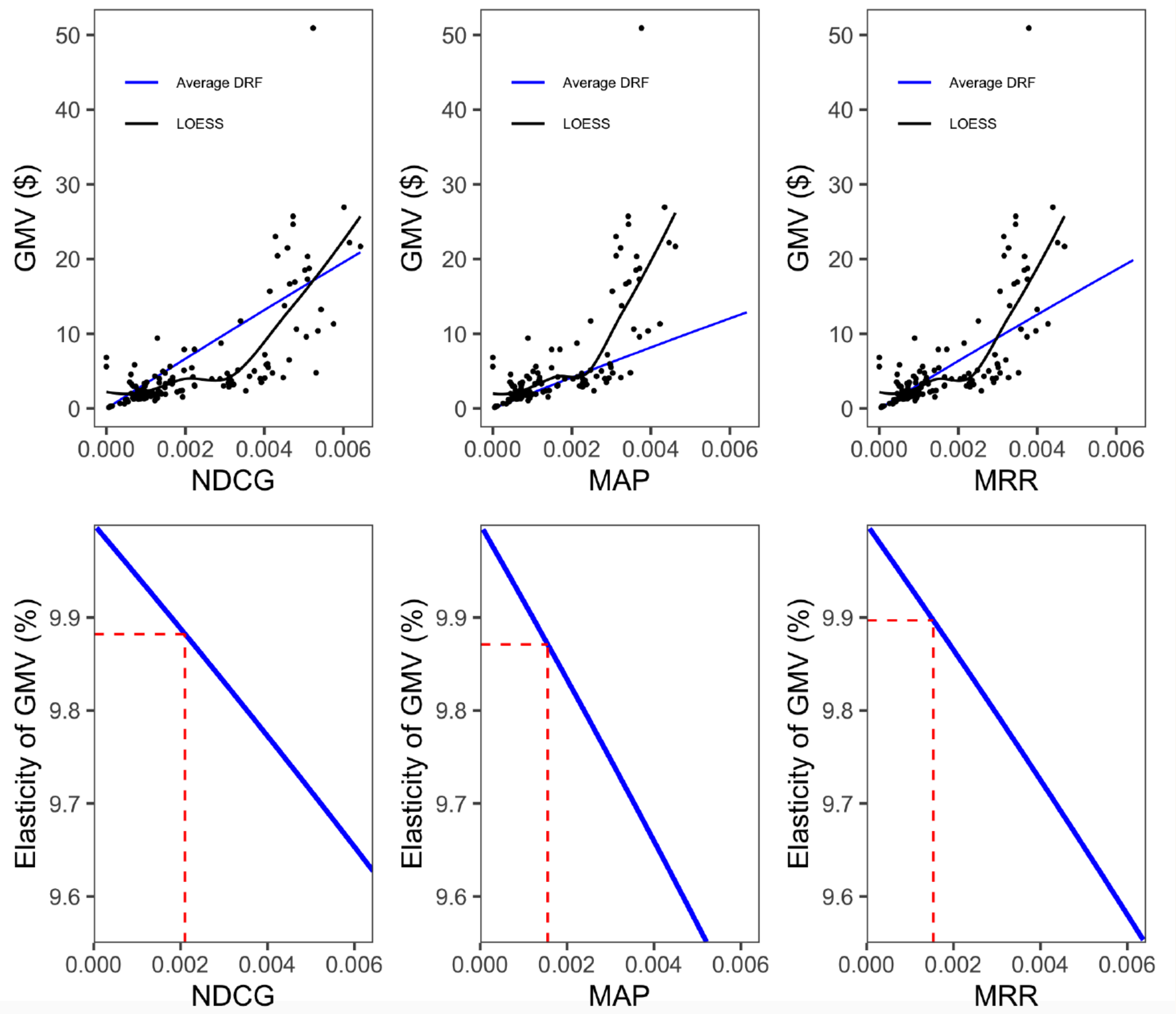
Causal Meta-Mediation Analysis
The Causal Path from Algorithms to Business



Data:

- 190 experiments from different teams.
- The figure shows that basic assumptions used by the method holds: enough variations.

Causal Meta-Mediation Analysis
The Causal Path from Algorithms to Business



Results:

- NDCG, MAP, MRR all have positive casual relationships with GMV.
- We could utilize *elasticity* to choose a better metric.

Takeaways

- We model the causality between online evaluation metrics and business KPIs by dose-response function (DRF) in potential outcome framework.
- Instead of conducting online tests, we use results from historical A/B experiments to conduct Meta-Analysis.
- From 190 experiments' data, we have established positive causal relationships between offline metrics and business KPIs and also could choose which metric is better.

Agenda

1. Industrial Recommender Systems with Their Ecosystems
2. Case I: Understanding The Interplay between Recommender Systems and Search Systems
3. Case II: The Journey of Long-term Engagement Optimization

Thank you

LinkedIn