Bias and Fairness in Information Retrieval Algorithms on E-commerce Platforms

Synopsis Seminar

Student: Abhisek Dash

Roll No: 17CS91R01

Supervised by

Dr. Saptarshi Ghosh

Dr. Animesh Mukherjee

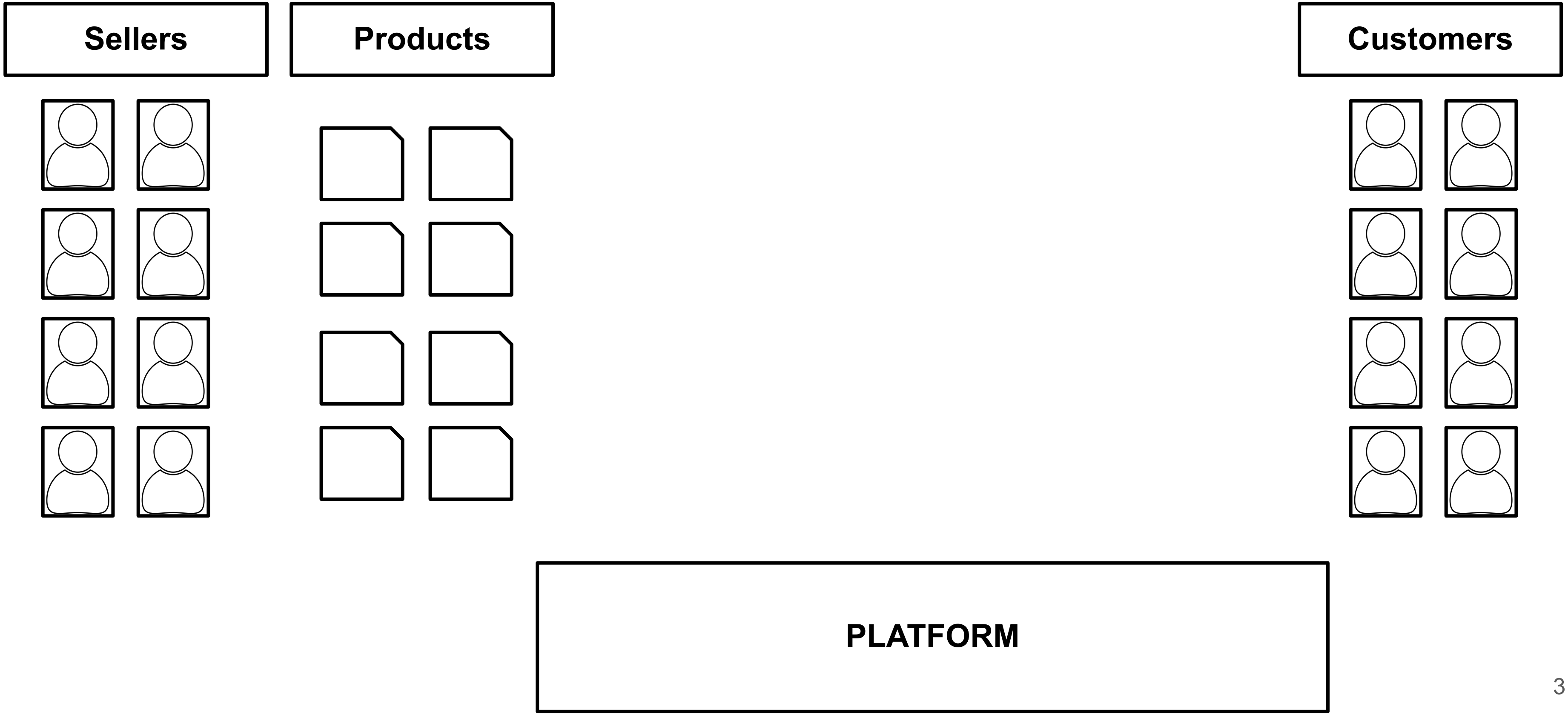
July, 2022

E-commerce platforms

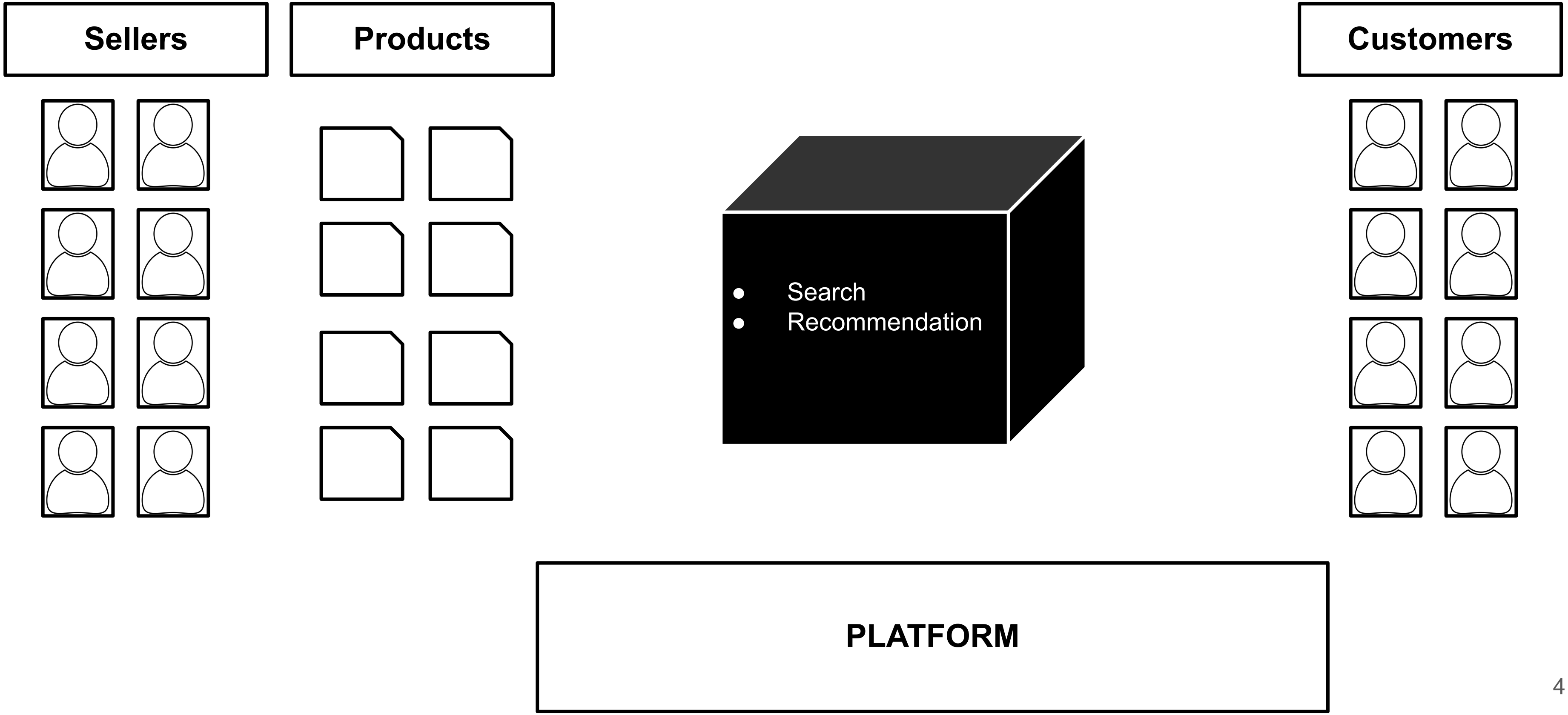
- Products and services are provided by multiple third parties.
- Customers increasingly turning to e-commerce platforms for purchase needs.
- Sellers and producers rely on e-commerce platforms for their livelihood.

I shall use the terms 'platform' and 'marketplace' interchangeably.

Stakeholders on e-commerce platforms



Algorithms mediate interactions between stakeholders



IR research

- Traditionally keyed to
  - Relevance
  - Customer satisfaction etc.

IR research

- ❖ Traditionally keyed to
  - ➢ Relevance
  - ➢ Customer satisfaction etc.

- ❖ Beyond relevance
  - ➢ Fairness to customers (e.g., equal satisfaction across demographics)
  - ➢ Fairness to sellers (e.g., equal exposure across seller groups)

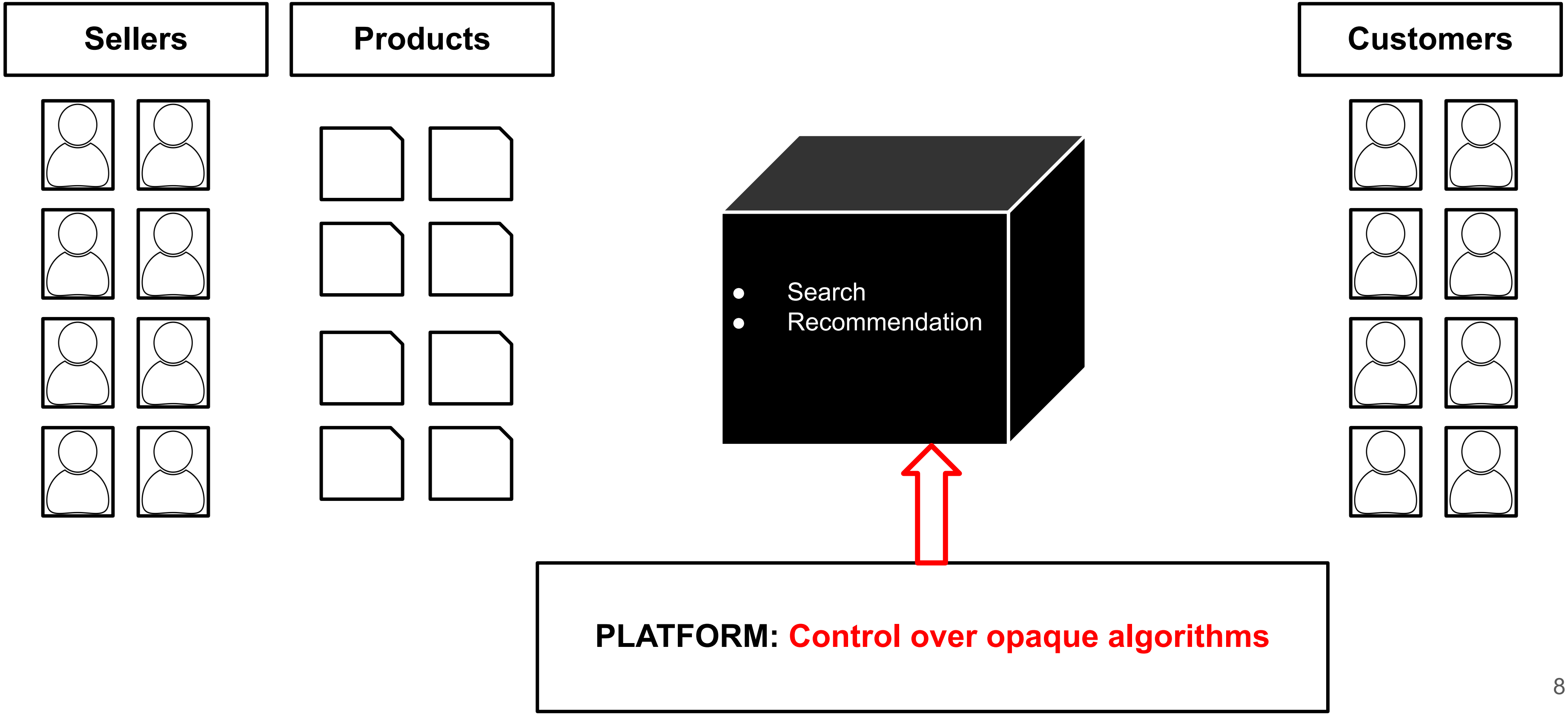
IR research

- Traditionally keyed to
  - Relevance
  - Customer satisfaction etc.

- Beyond relevance
  - Fairness to customers (e.g., equal satisfaction across demographics)
  - Fairness to sellers (e.g., equal exposure across sellers)

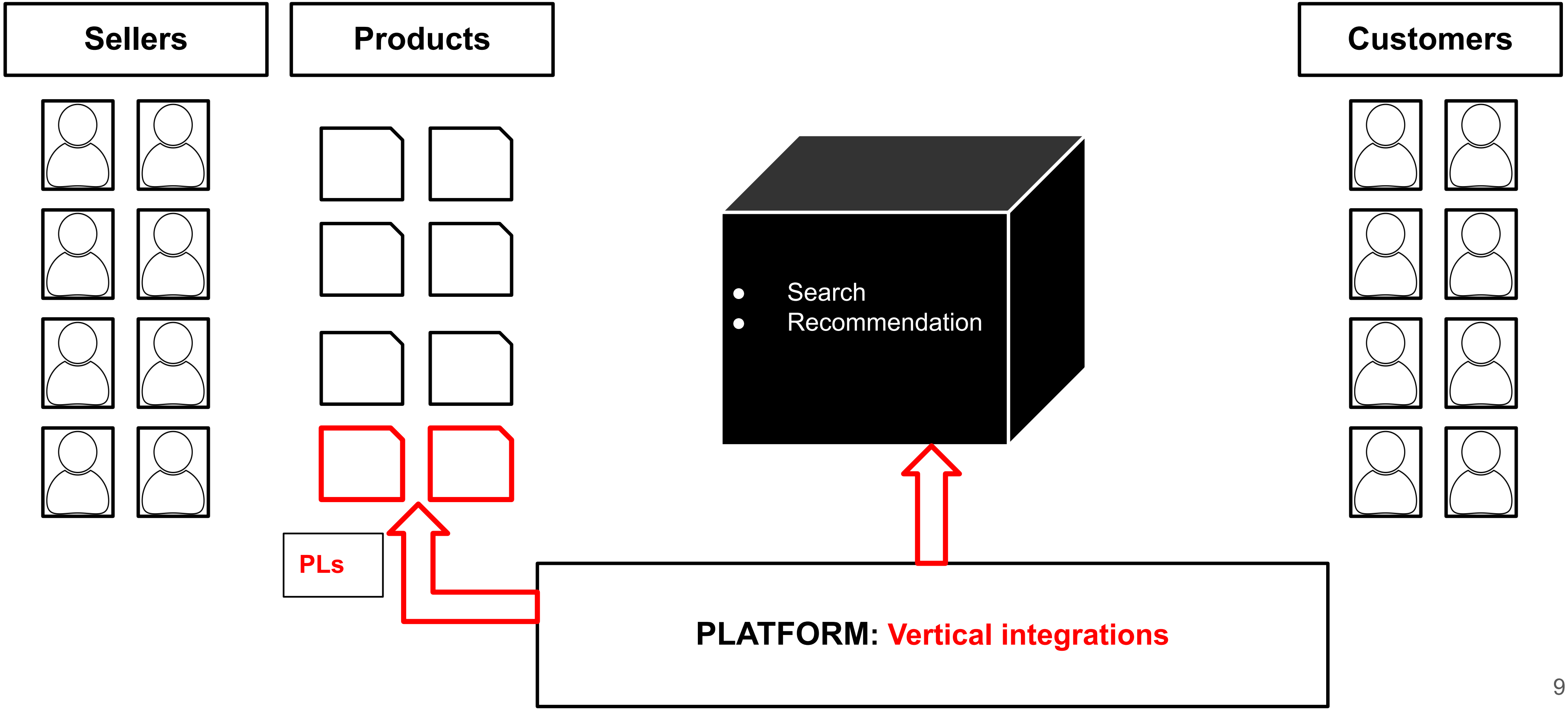
Overlooked the most important stakeholder: Platform / Marketplace

Why is platform important?



Why is platform important?

Manufacture and sell
private label products (PL)



Third party products (3P)

Duracell Ultra Alkaline AA Battery, 8 Pieces

Brand: Duracell

7,145 ratings | 504 answered questions

#1 Best Seller in General Purpose Batteries & Battery Chargers

M.R.P.: ~~₹ 294.00~~
Deal Price: ₹ 279.00 prime
You Save: ₹ 15.00 (5%)
Inclusive of all taxes

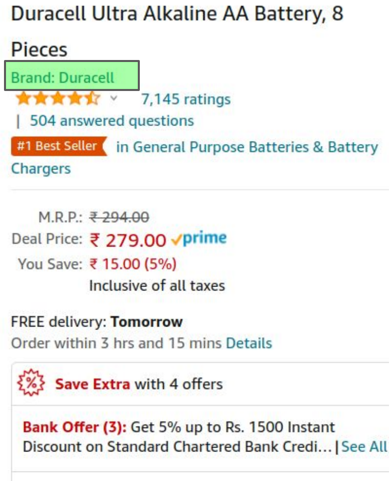
FREE delivery: **Tomorrow**
Order within 3 hrs and 15 mins Details

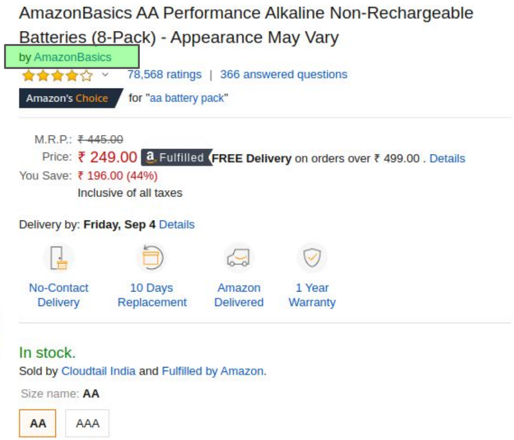
Save Extra with 4 offers

Bank Offer (3): Get 5% up to Rs. 1500 Instant Discount on Standard Chartered Bank Credi... | See All

❖ Many products from different 3P brands are sold on Amazon e.g., Duracell.

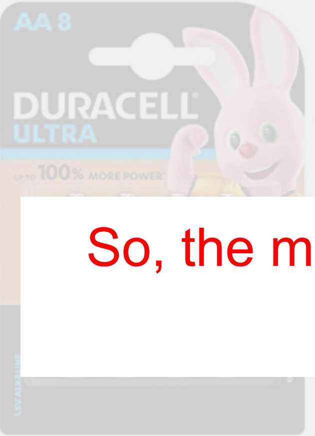
Third party products (3P) Vs. Amazon private labels (PL)

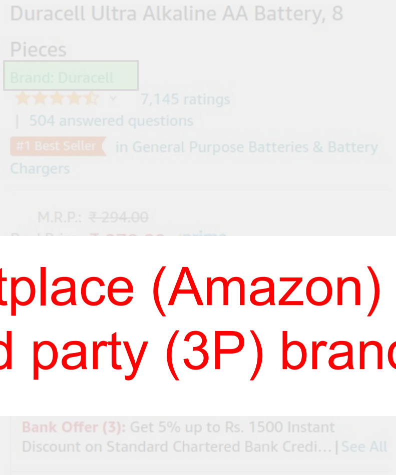


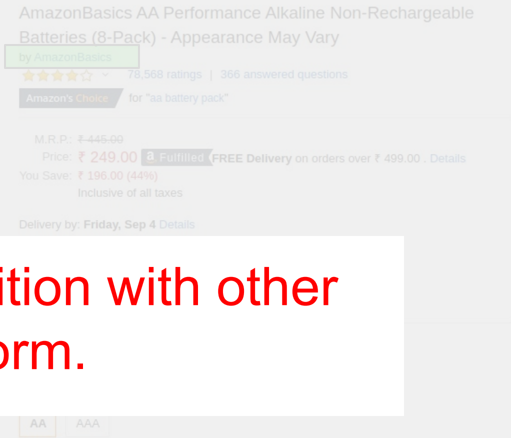


- Many products from different 3P brands are sold on Amazon e.g., Duracell.
- Amazon also produces its own PL products under brand names e.g., AmazonBasics.

Third party products (3P) Vs. Amazon private labels (PL)



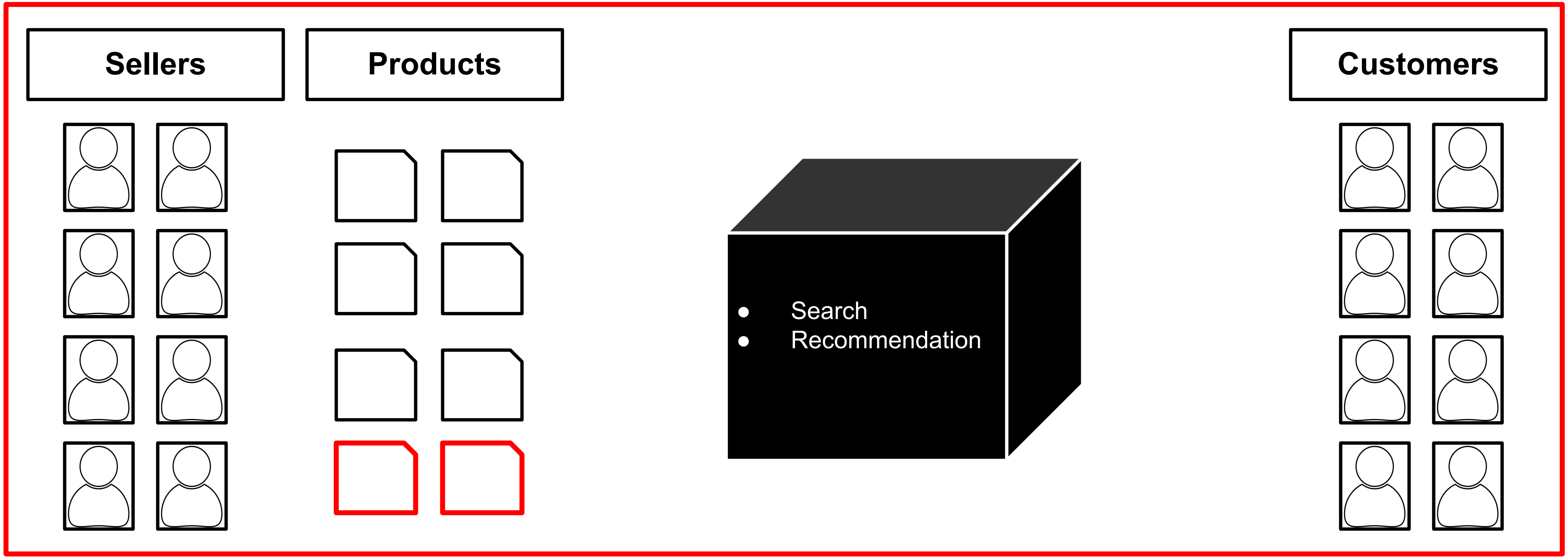




So, the marketplace (Amazon) is in direct competition with other third party (3P) brands on its own platform.

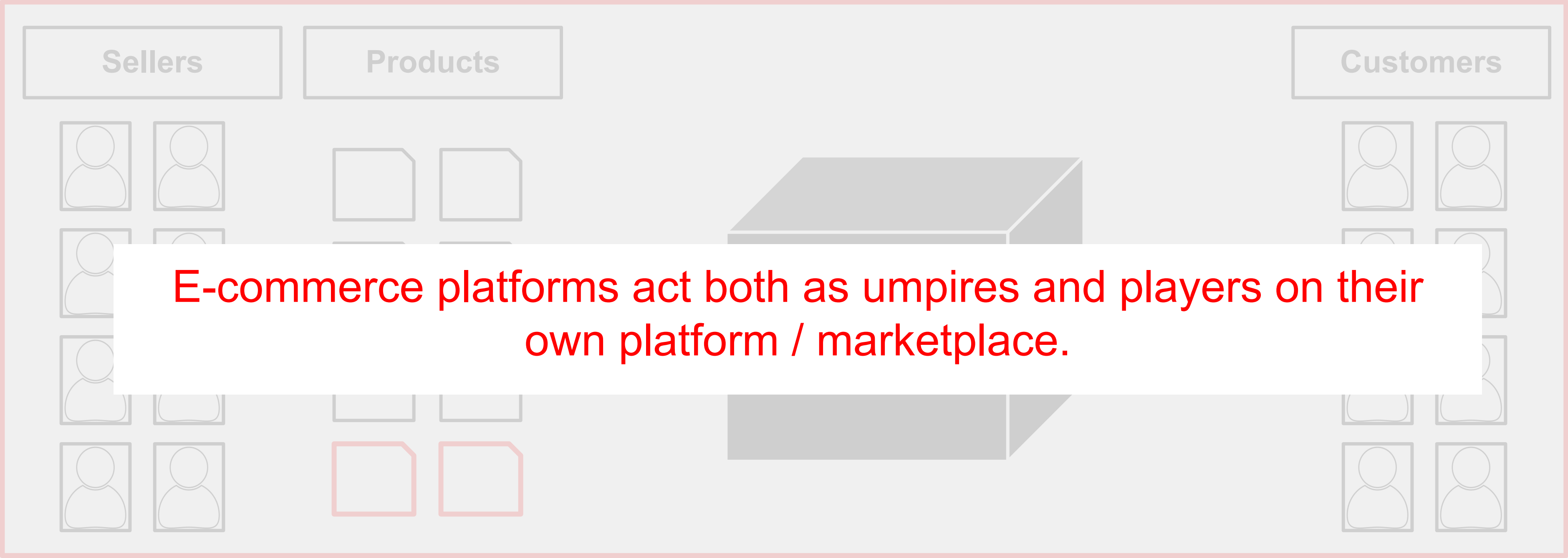
- Many products from different 3P brands are sold on Amazon e.g., Duracell.
- Amazon also produces its own PL products under brand names e.g., AmazonBasics.

Why is platform important?



PLATFORM: **Scale and sophistication**

Why is platform important?



E-commerce platforms act both as umpires and players on their own platform / marketplace.

PLATFORM: Scale and sophistication

Policymakers across the globe are equally concerned

HOME ABOUT COMMITTEE ACTIVITY MEDIA CENTER ISSUES SUBCOMMITTEES LETTERS CONTACT

Digital Markets Investigation

Antitrust Investigation of the Rise and Use of Market Power Online and the Adequacy of Existing Antitrust Laws and Current Enforcement Levels

Ministry of Commerce & Industry

Review of policy on Foreign Direct Investment (FDI) in e-commerce

Posted On: 26 DEC 2018 5:25PM by PIB Delhi

European Commission - Press release

Antitrust: Commission sends Statement of Objections to Amazon for the use of non-public independent seller data and opens second investigation into its e-commerce business practices

Concerns raised in popular press

◆ WSJ NEWS EXCLUSIVE

Amazon Changed Search Algorithm in Ways That Boost Its Own Products

The e-commerce giant overcame internal dissent from engineers and lawyers, people familiar with the move say

The New York Times

How Amazon Steers Shoppers to Its Own Products

REUTERS® INVESTIGATES | The Imitation Game

A REUTERS SPECIAL REPORT

Amazon copied products and rigged search results to promote its own brands, documents show

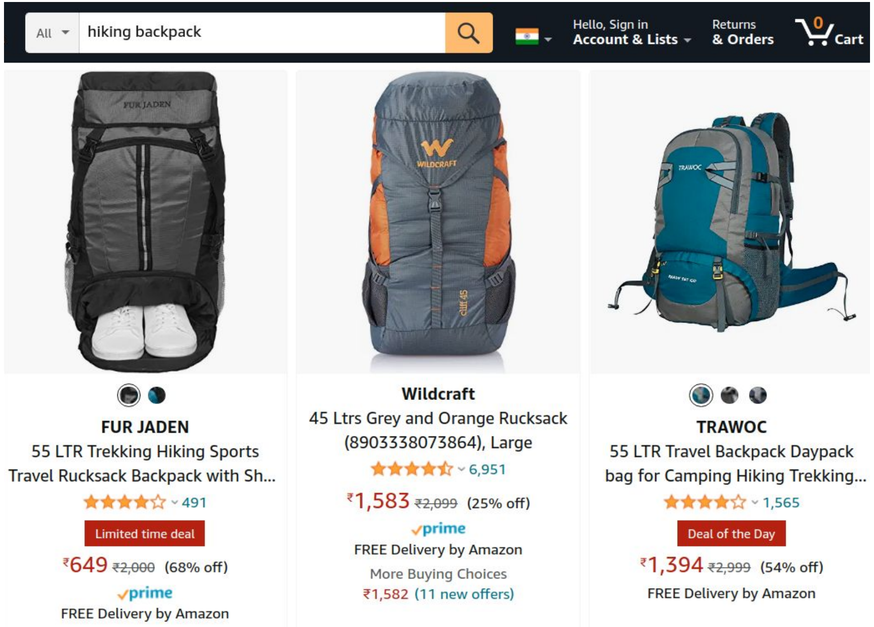
Need of the hour

Except media reports raising concerns based on anecdotal evidences, there is no detailed public scrutiny of these practices.

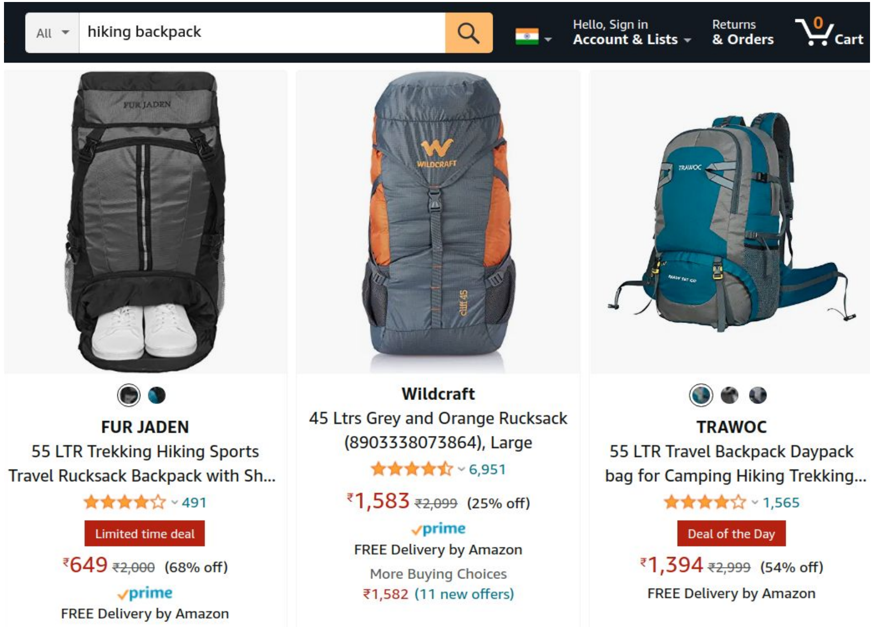
We need a systematic end-to-end audit of the e-commerce ecosystem

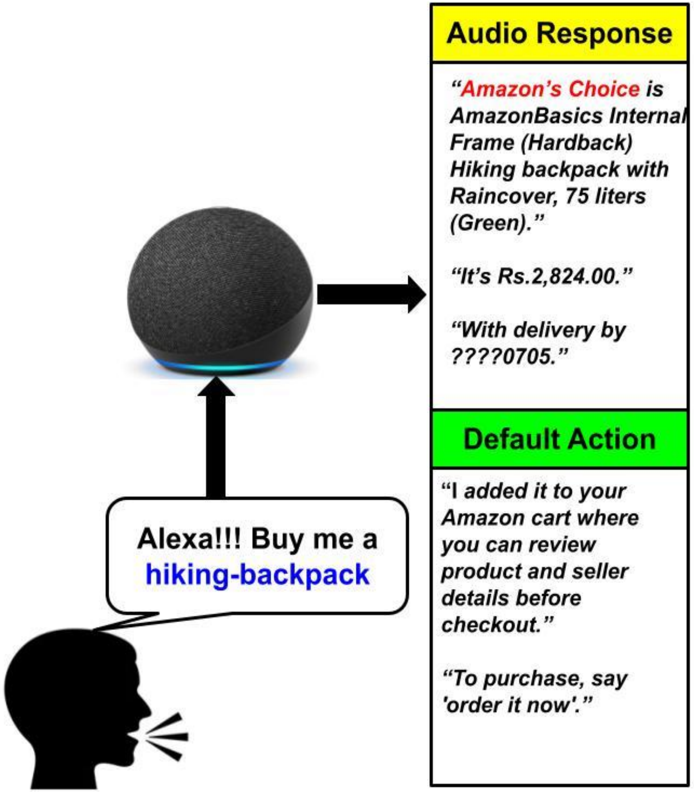
These issues also propagate to other modes of interactions with IR algorithms.

Traditional search



Traditional search Vs. Search through Smart Speakers





Differences in the two modes of interaction

- Availability of choices ↓
- Autonomy of customers ↓

Such restricted autonomy warrants the voice assistants underlying smart speakers to be more responsible.

Disclaimer

- ❖ We do not imply these bias / fairness concerns are intentionally induced by algorithm designers or platform organizations.
- ❖ However, these biases should not go unnoticed.
- ❖ Identifying and acknowledging their existence is the first step toward mitigation.

Disclaimer

- ❖ We do not imply these bias / fairness concerns are intentionally induced by algorithm designers or platform organizations.
- ❖ However, these biases should not go unnoticed.
- ❖ Identifying and acknowledging their existence is the first step toward mitigation.
- ❖ Limitation: Most of the audits done are black box in nature.

Disclaimer

- ❖ We do not imply these bias / fairness concerns are intentionally induced by algorithm designers or platform organizations.
- ❖ However, these biases should not go unnoticed.
- ❖ Identifying and acknowledging their existence is the first step toward mitigation.
- ❖ Limitation: Most of the audits done are black box in nature.
- ❖ Our aim: A marketplace providing level playing field to all stakeholders.

Research questions addressed in this thesis

How does one quantify bias (if any) induced by related item recommendation algorithms on e-commerce platforms?

How does one generate fair related item recommendations?

How fair and interpretable are the response and default action of voice assistants for e-commerce search queries?

Research questions addressed in this thesis

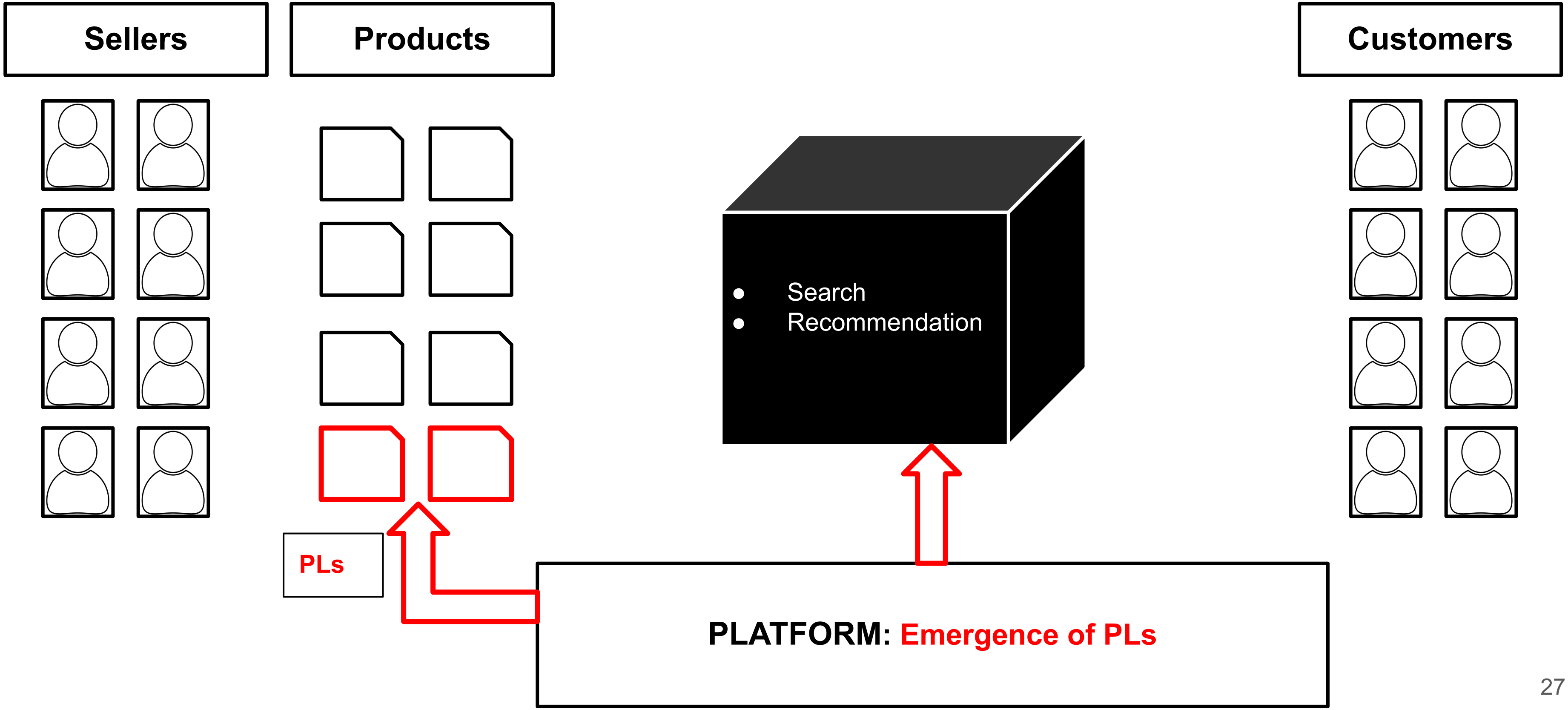
How does one quantify bias (if any) induced by related item recommendation algorithms on e-commerce platforms?

How does one generate fair related item recommendations?

How fair and interpretable are the response and default action of voice assistants for e-commerce search queries?

Why is platform important?

PRIVATE LABEL PRODUCTS (PLs)



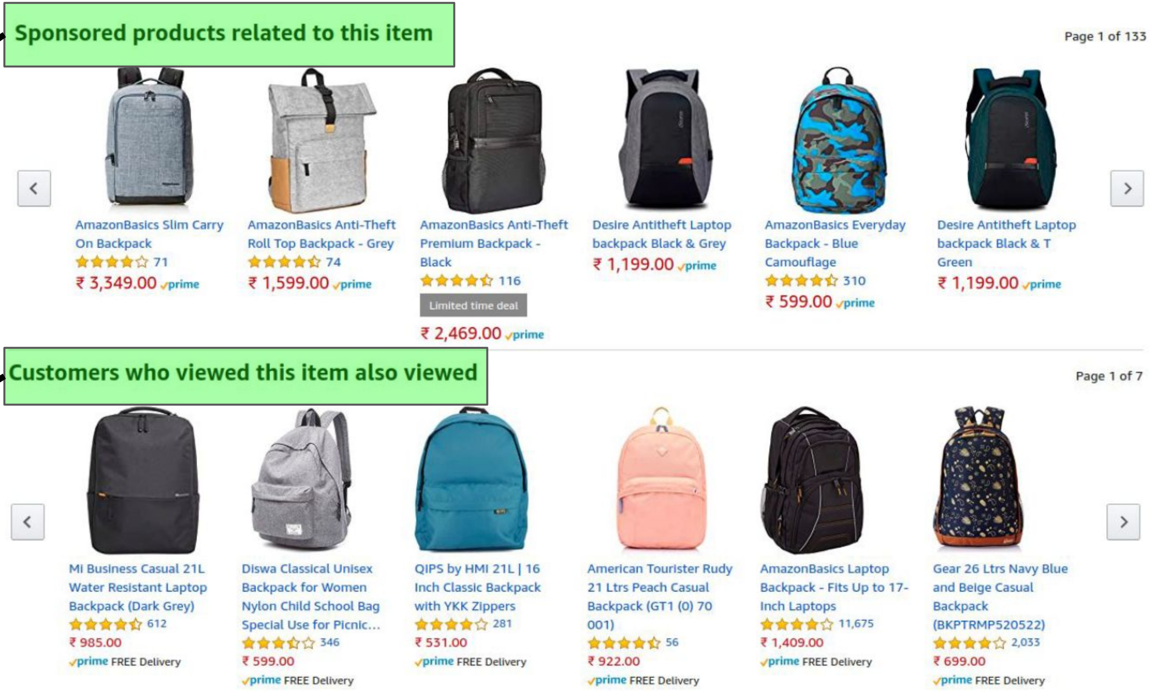
This study

- This is the first systematic audit of the Amazon ecosystem
  - **Related item recommendations on Amazon**
- We propose a number of network centric measures to quantify biases
  - **Toward Amazon private label products**

Related item recommendations (RIRs) on Amazon

Sponsored Recommendations

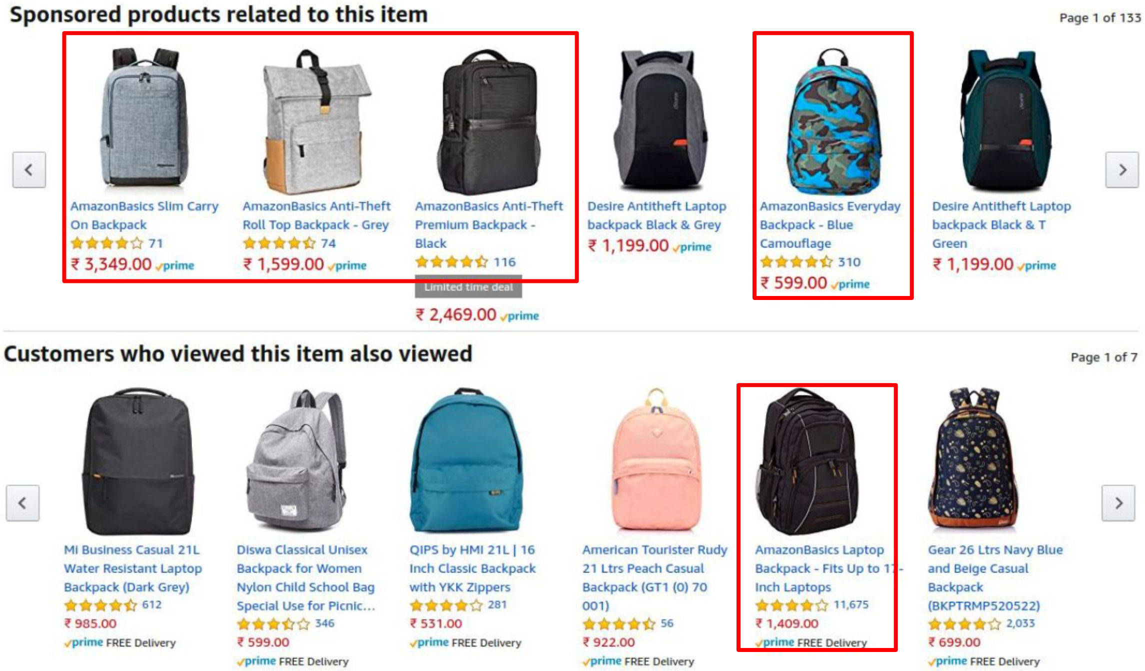
Organic Recommendations



*Screenshot taken from Amazon.in.

Types of products recommended on Amazon

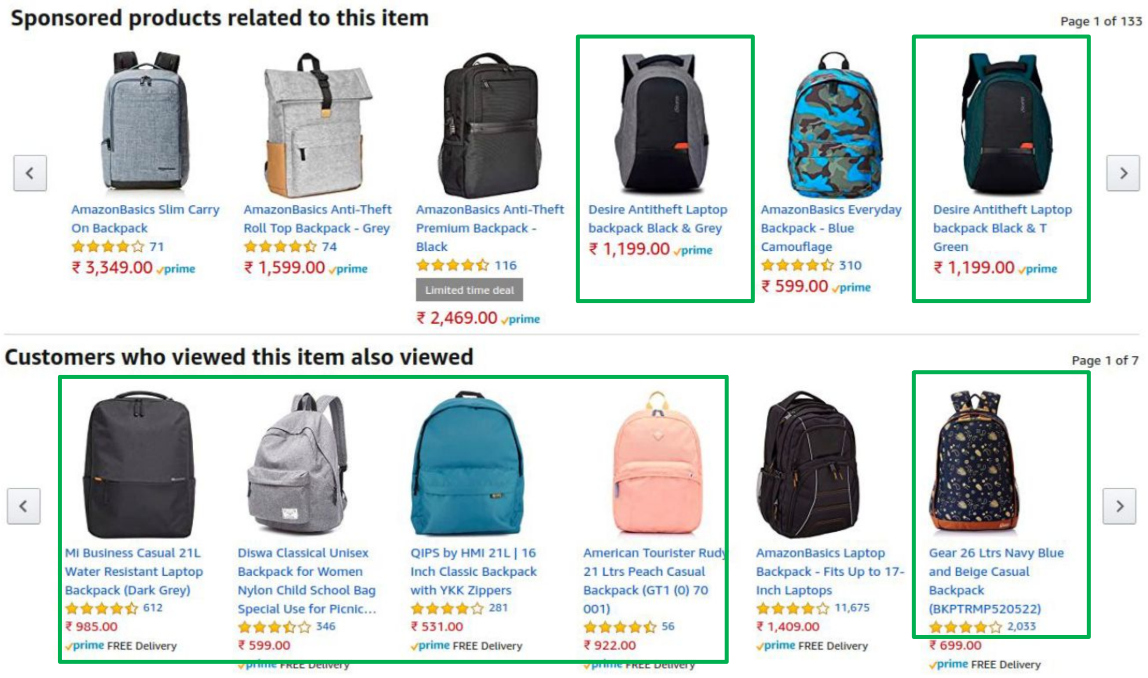
Amazon Private Label products (PL)



*Screenshot taken from Amazon.in.

Types of products recommended on Amazon

Third party products (3P)



*Screenshot taken from Amazon.in.

Concerns regarding sponsored ads

- ❖ Sponsored advertisements replacing organic recommendations

https://www.marketplacepulse.com/articles/amazon-is-replacing-product-suggestions-with-ads

Concerns regarding sponsored ads

- ❖ Sponsored advertisements replacing organic recommendations
- ❖ Survey of 2,000+ Amazon customers show that 50% respondents didn’t even realize being advertised to on Amazon product pages.
- ❖ Sponsored recommendations offer a powerful option to nudge customers.

https://www.cnbc.com/2019/07/17/how-amazon-advertising-works.html

Concerns regarding sponsored ads

- Sponsored advertisements replacing organic recommendations
- Survey of 2,000+ Amazon customers show that 50% respondents didn’t even realize being advertised to on Amazon product pages.
- Sponsored recommendations offer a powerful option to nudge customers.

Thus biases in the sponsored ads can not be neglected

This study

- This is the first systematic audit of the Amazon echo-system
  - **Related item recommendations on Amazon**
- We propose a number of network centric measures to quantify biases
  - **Toward Amazon private label products**
- **Specifically, we investigate for biases (if any) in the sponsored recommendations on Amazon**

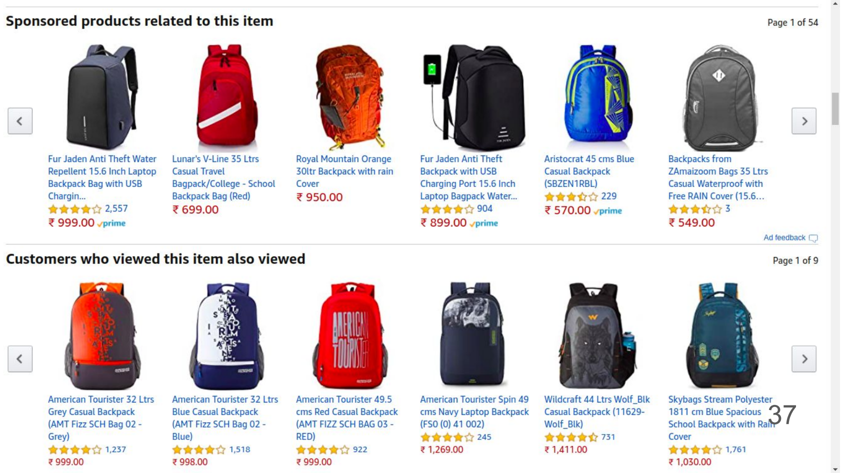
Data collection

- ❖ Categories: Backpack and Battery

Data collection

- Categories: Backpack and Battery
- We crawled the Amazon website for data collection.
- The recommendations and metadata for each products were collected.
- Both organic and sponsored recommendations were collected.

Category	# Items	#PLs
Backpack	10,775	161
Battery	5,352	17

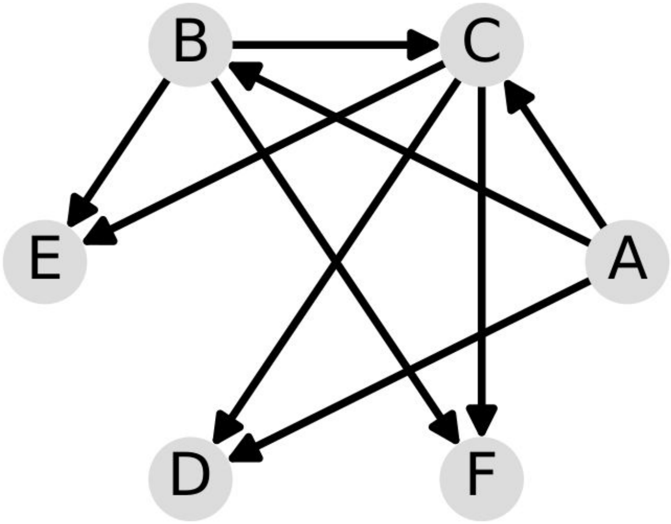


Framework for auditing recommendation systems

The framework

❖ To understand the recommendation ecosystem, we instantiate related item recommendations as Related Item Network (RIN).

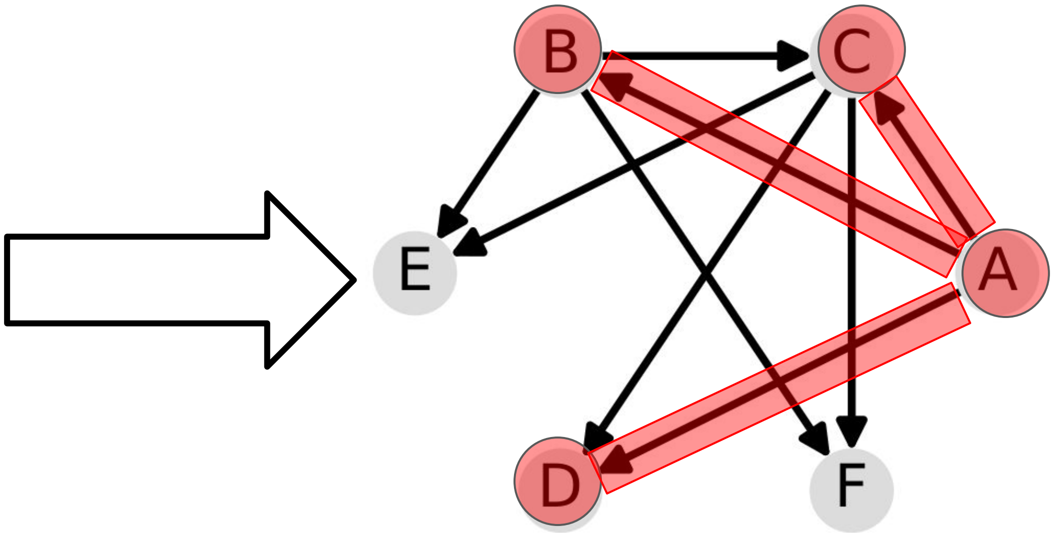
Items	Related items
A	B, C, D
B	E, C, F
C	D, E, F



The framework

❖ To understand the recommendation ecosystem, we instantiate related item recommendations as Related Item Network (RIN).

Items	Related items
A	B, C, D
B	E, C, F
C	D, E, F



Two RINs

- RINs constructed from
  - Sponsored recommendations: Sponsored RIN (*S*)
  - Organic recommendations: Organic RIN (*O*)

Two RINs and the rationale behind it

- RINs constructed from
  - Sponsored recommendations: Sponsored RIN (*S*)
  - Organic recommendations: Organic RIN (*O*)
- Difficult to quantify bias in absolute terms; however it is easier provided an unbiased reference for comparison.

Two RINs and the rationale behind it

- RINs constructed from
  - Sponsored recommendations: Sponsored RIN (*S*)
  - Organic recommendations: Organic RIN (*O*)
- Difficult to quantify bias in absolute terms; however it is easier provided an unbiased reference for comparison.
- Assumption: Organic recommendations are an appropriate representation of user activities.
- Aim: Quantification of relative bias (if any) toward PLs in *S* as compared to *O*.

Methodologies for evaluation of bias

- ❖ Promotion bias
- ❖ Ranking bias
- ❖ Representation in the core of a network
- ❖ Exposure bias
- ❖ Quantifying the influence of the sensitive attribute

Methodologies for evaluation of bias

- **Promotion bias**
- Ranking bias
- Representation in the core of a network
- **Exposure bias**
- Quantifying the influence of the sensitive attribute

Promotion bias

Promotion bias

- Amazon PLs get sponsored recommendations from half of the product space.
  - In organic RIN this percentage drops to 15%.
- Amazon PLs are not similar to many products as per the organic customer behavior.

Comparing in-degree of different type of batteries

Properties	Organic RIN	Sponsored RIN
Avg. in-degree of nodes	11	11
Categorization as per products' relationship with Amazon		
Avg. in-degree of private label products	**46**	**520**
Avg. in-degree of 3P products	11	09

* All the differences were found to be statistically significant.

Comparing in-degree of different type of batteries

Properties	Organic RIN	Sponsored RIN
Avg. in-degree of nodes	11	11
Categorization as per products' relationship with Amazon		
Avg. in-degree of private label products	**46**	**520**
Avg. in-degree of 3P products	11	09

* All the differences were found to be statistically significant.

Comparing in-degree of different type of batteries

Properties	Organic RIN	Sponsored RIN
Avg. in-degree of nodes	11	11
Categorization as per products' relationship with Amazon		
Avg. in-degree of private label products	**46**	**520**
Avg. in-degree of 3P products	11	09

* All the differences were found to be statistically significant.

Comparing in-degree of different type of batteries

The degree of promotion of Amazon PLs is significantly higher in the sponsored RIN as compared to organic RIN.

* All the differences were found to be statistically significant.

Exposure bias

Estimating exposure of an item from a RIN

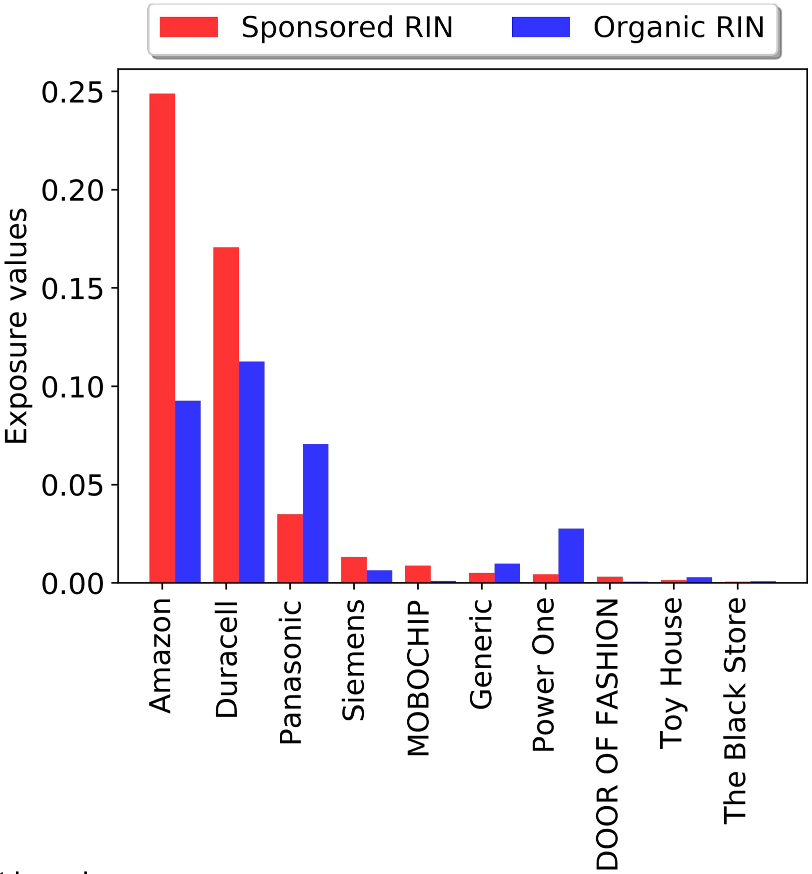
- We use the ‘Random Surfer Model’ for exposure estimation.
- The steady state visit frequency of an item on the RIN is its exposure.
- We account for the following:
  - User propensity to follow recommendations.
  - Installation of the long-tail popularity distribution
  - Presentation bias on the online platform of Amazon.

Different exposures and exposure bias

- **Organic exposure:** Exposure of an item due to organic recommendations on Amazon (E_o)
- **Sponsored exposure:** Exposure of an item due to sponsored recommendations on Amazon (E_s)
- **Exposure bias** = KLD (E_s || E_o)

Exposure distortion due to sponsored recommendations

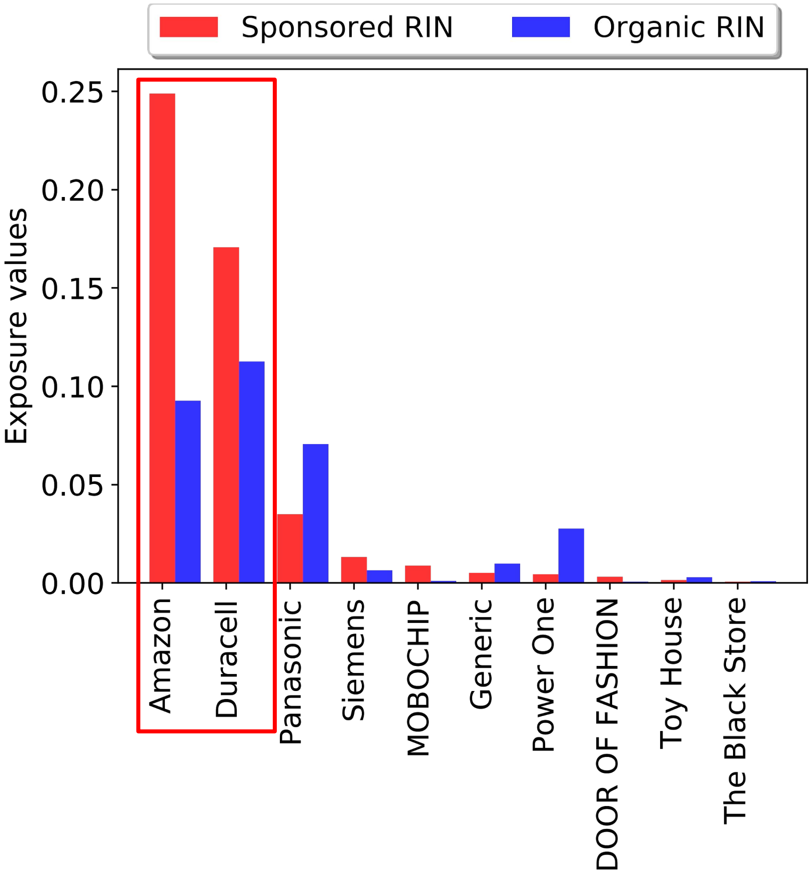
Distortion in exposure for brands on Amazon



Exposure of a brand is evaluated as the sum of exposure of all items of that brand.

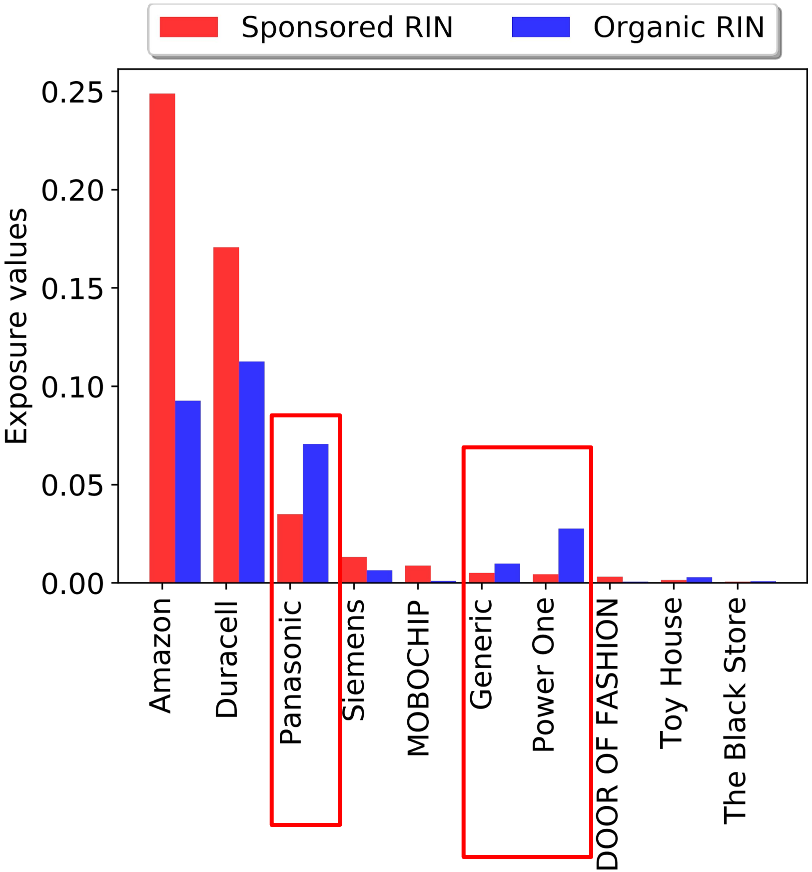
Distortion in exposure for brands on Amazon

- ❖ Amazon private label brands and Duracell saw increase in their exposure in the sponsored RIN.
- ❖ 17 PLs accounted for **25%** of the total exposure in sponsored RIN.



Distortion in exposure for brands on Amazon

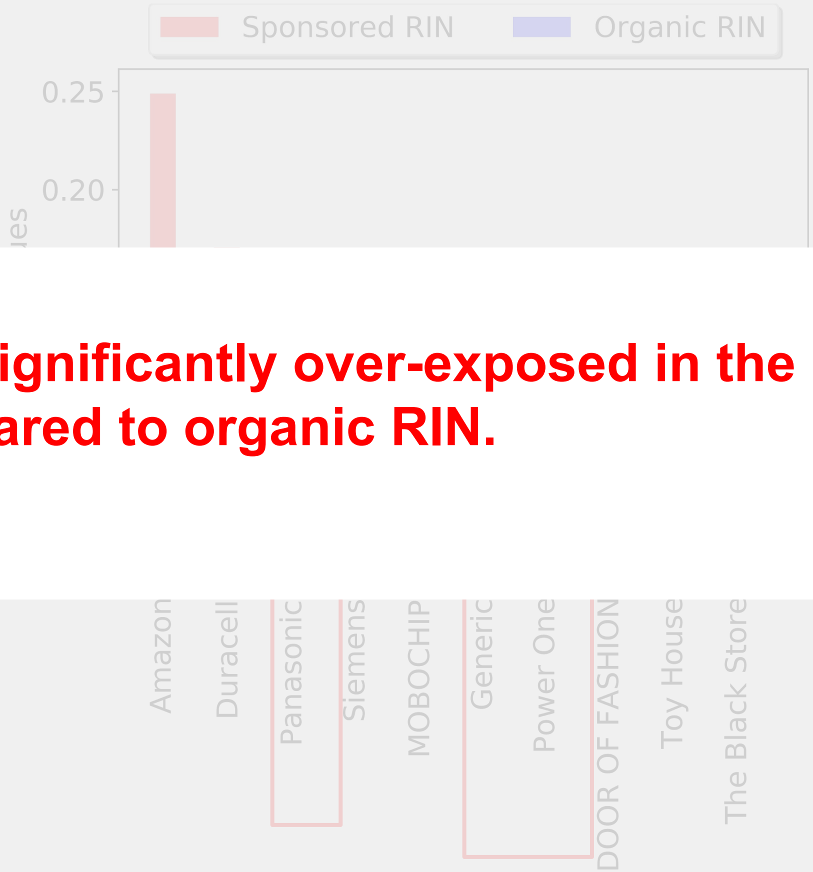
- Top brands like Power One, Generic, and Panasonic saw significant drop in sponsored exposure as compared to their organic exposure.
- 75% of all brands were underexposed.



Distortion in exposure for brands on Amazon

❖ Top brands like Power One,

underexposed.



Amazon private label products are significantly over-exposed in the sponsored RIN as compared to organic RIN.

Methodologies for evaluation of bias

- **Promotion bias**
- Ranking bias
- Representation in the core of a network
- **Exposure bias**
- Quantifying the influence of the sensitive attribute

All the above showed indication of biases toward Amazon private label products.

Why are biases in sponsored recommendations important?

- ❖ Sponsored Vs. Organic recommendations

- ❖ Economic aspect of self-sponsorship

Why are biases in sponsored recommendations important?

- ❖ Sponsored Vs. Organic recommendations
  - ➢ Sponsored recommendations ought to be different from organic ones.
  - ➢ They can have delayed impact on organic recommendations.

- ❖ Economic aspect of self-sponsorship

Why are biases in sponsored recommendations important?

- ❖ Sponsored Vs. Organic recommendations
  - ➢ Sponsored recommendations ought to be different from organic ones.
  - ➢ They can have delayed impact on organic recommendations.

- ❖ Economic aspect of self-sponsorship
  - ➢ Amazon can unilaterally reserve ad-space without being accountable.
  - ➢ This can increase the advertising cost for the remaining ad-space.

Why are biases in sponsored recommendations important?

- ❖ Sponsored Vs. Organic recommendations

- ❖ Hence biases in sponsored recommendations should not go unnoticed.
- ❖ Some clarity regarding the exact practices are also desirable.

  - ➢ Amazon can unilaterally reserve ad-space without being accountable.
  - ➢ This can increase the advertising cost for the remaining ad-space.

Summary: The umpire is also a player

- **Umpire:** Amazon has the control over the algorithmic systems.

- **Player:** Amazon’s PL products also compete on its platform.

Summary: The umpire is also a player

- **Umpire:** Amazon has the control over the algorithmic systems.
- **Player:** Amazon's PL products also compete on its platform.

- **Umpire:** Amazon has the control over ad-space allocation.
- **Player:** Ads of Amazon's PL products are also shown on Amazon.

Publications from this chapter

❖ **A. Dash**, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *When the Umpire is also a Player: Bias in Private Label Product Recommendations on E-commerce Marketplaces*, In Proceedings of the **ACM Conference on Fairness, Accountability, and Transparency**, (ACM FAccT), March, 2021.

❖ **A. Dash**, A Mukherjee, S. Ghosh, *A Network-centric Framework for Auditing Recommendation Systems*, In Proceedings of the **IEEE International Conference on Computer Communications** (IEEE INFOCOM), April, 2019.

Research questions addressed in this thesis

How does one quantify bias (if any) induced by related item recommendation algorithms on e-commerce platforms?

How does one generate fair related item recommendations?

How fair and interpretable are the response and default action of voice assistants for e-commerce search queries?

Exposure bias on Amazon

- ❖ Special relationships on the platform can lead to exposure biases.
- ❖ Potential remedy : Policy changes / more succinct regulations.

How to mitigate exposure bias from RIRs?

- ❖ We should be trying to mitigate exposure bias on Amazon directly.

How to mitigate exposure bias from RIRs?

- We should be trying to mitigate exposure bias on Amazon directly.
- Algorithms deployed on Amazon are proprietary.

How to mitigate exposure bias from RIRs?

- ❖ We should be trying to mitigate exposure bias on Amazon directly
- ❖ Algorithms deployed on Amazon are proprietary.

- ❖ We do the next best thing by
  - ➢ Investigating for exposure bias (if any) in standard RIR algorithms
  - ➢ Trying to mitigate them

Can exposure bias be implicitly induced due to RIR algorithms?

RIR algorithms and Exposure Bias thereof

Some popular related item recommendation algorithms

- ratingSVD
  - Applies SVD to the user-item rating matrix
  - Cosine-similarity for similarity evaluation
  - Notion: people who like this item also like

Some popular related item recommendation algorithms

- ratingSVD
  - Applies SVD to the user-item rating matrix
  - Cosine-similarity for similarity evaluation
  - Notion: people who like this item also like

- item2vec
  - Replication of word2vec
  - Tries to find co-occurrence patterns
  - Notion: people consume these items together

Some popular related item recommendation algorithms

- ratingSVD
  - Applies SVD to the user-item rating matrix
  - Cosine-similarity for similarity evaluation
  - Notion: people who like this item also like

- item2vec
  - Replication of word2vec
  - Tries to find co-occurrence patterns
  - Notion: people consume these items together

Datasets utilised

- MovieLens 10M
  - 10M ratings
  - 71K different users
  - 10K distinct movies

- Amazon review dataset
  - 194K ratings
  - 27K different users
  - 10K different products

Datasets utilised

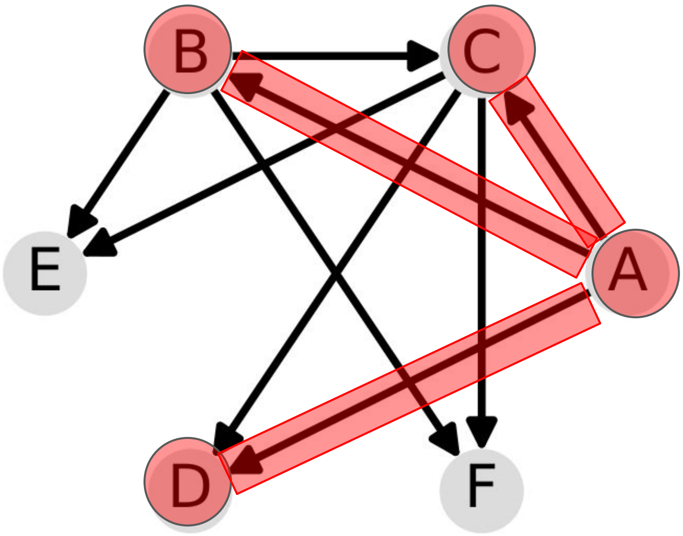
- MovieLens 10M
  - 10M ratings
  - 71K different users
  - 10K distinct movies

- Amazon review dataset
  - 194K ratings
  - 27K different users
  - 10K different products

Related Item Network

❖ To understand the recommendation ecosystem, we instantiate related item recommendations as Related Item Network (RIN).

Items	Related items
A	B, C, D
B	E, C, F
C	D, E, F



Estimating observed exposure of an item from a RIN

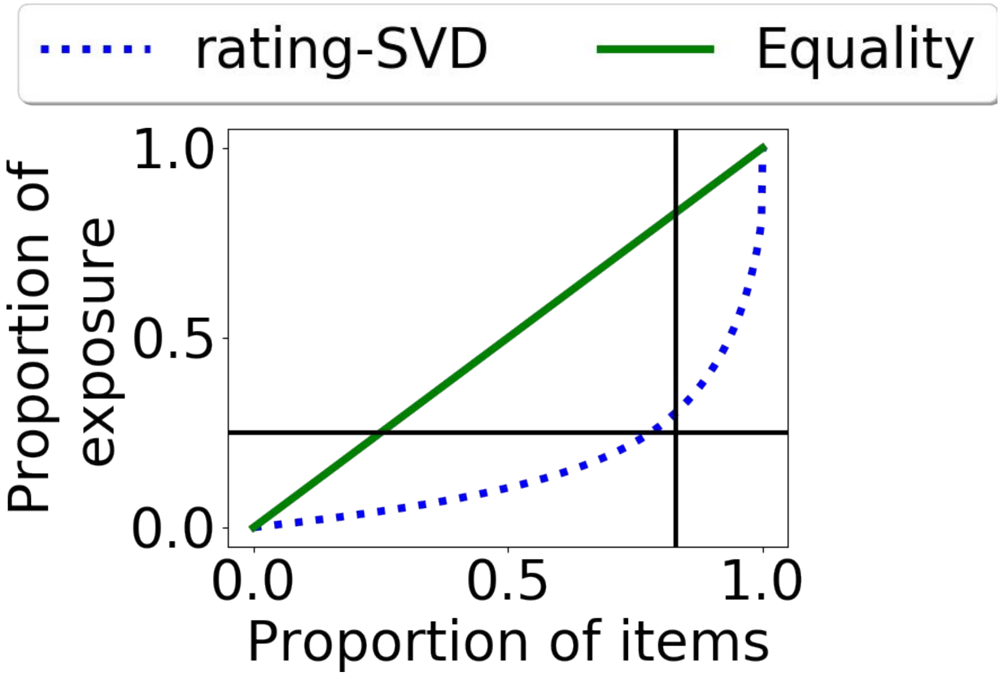
- ❖ We use the ‘Random Surfer Model’ for exposure estimation.

- ❖ The steady state visit frequency of an item on the RIN is its exposure.

Skew in exposure of items

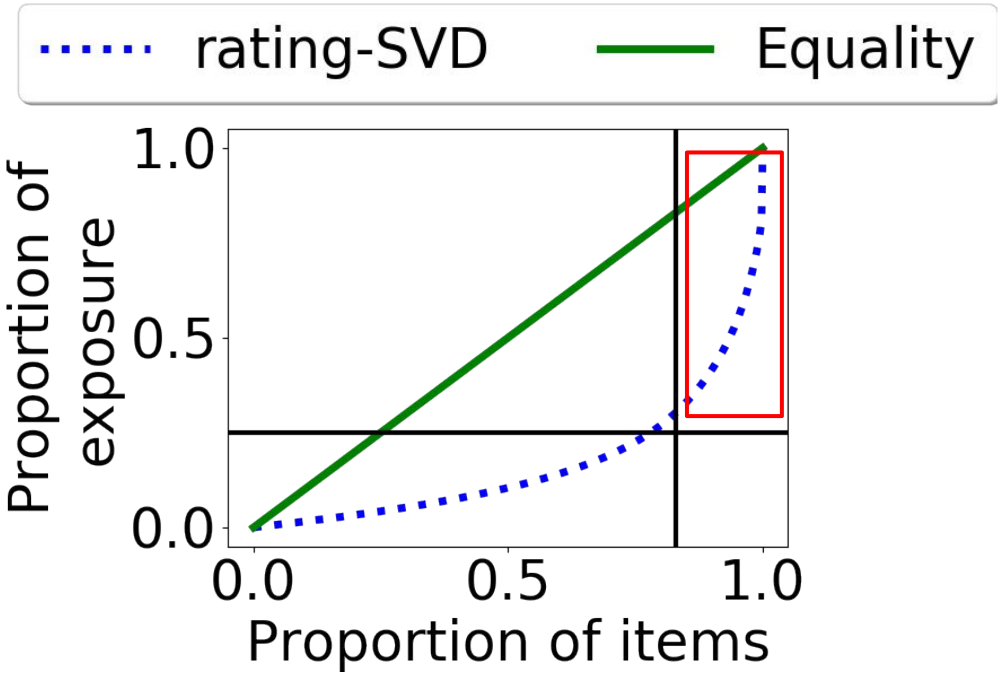
X axis: Cumulative proportion of items in increasing order of their exposure (left to right)

Y axis: Cumulative proportion of exposure distribution



Skew in exposure of items

Top 25% of the items with most exposure in the Amazon dataset account for 75% of the entire exposure.

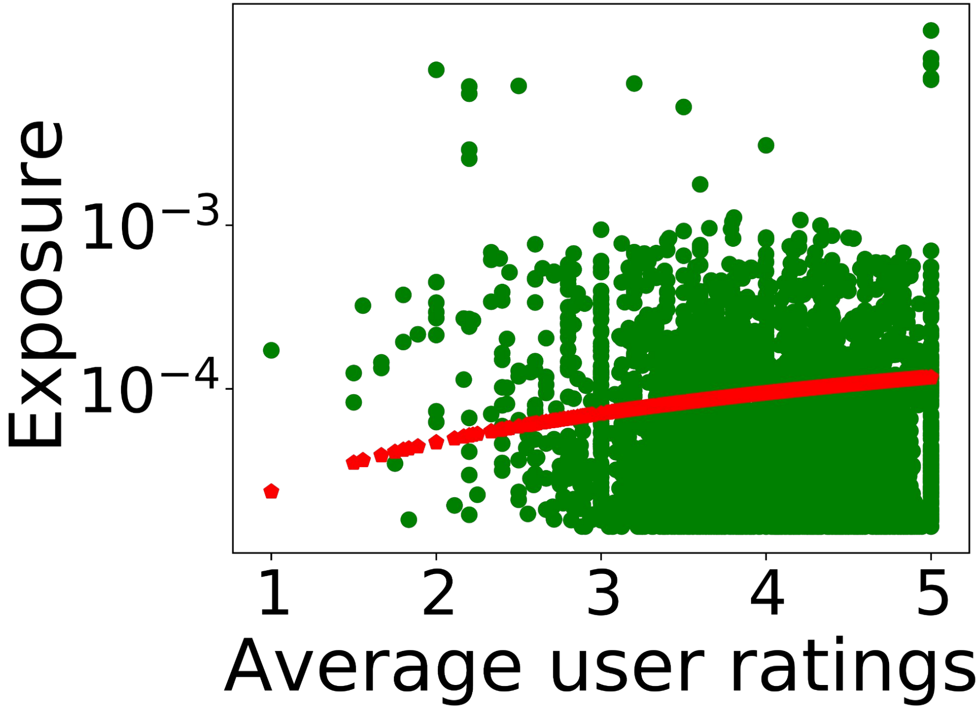


Does gap in quality explain such higher exposure?

X axis: Average user ratings of an item

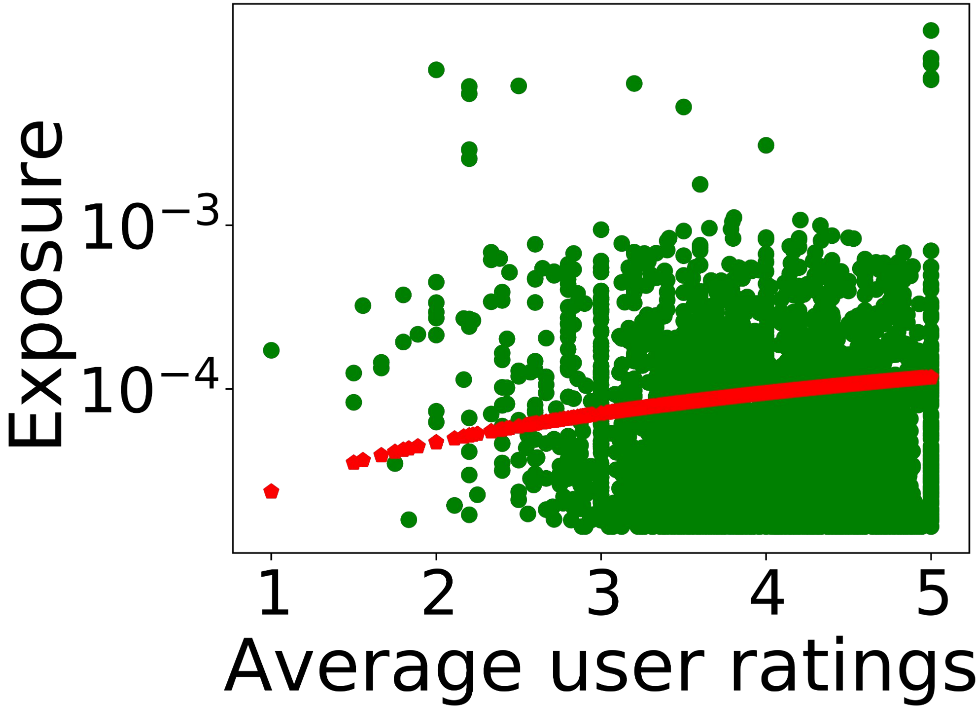
Y axis: Observed exposure (in log scale) of the corresponding item

Red curve: log plot of quality distribution



Does gap in quality explain such higher exposure?

Only 6-7% of the item set has comparable quality and exposure.



What can be potential reasons?

❖ A poor quality item is recommended from the web page of a high quality item. Such items may get some exposure thanks to that of their source items.

❖ This phenomenon may get reinforced over time.

What is a desired exposure of an item?

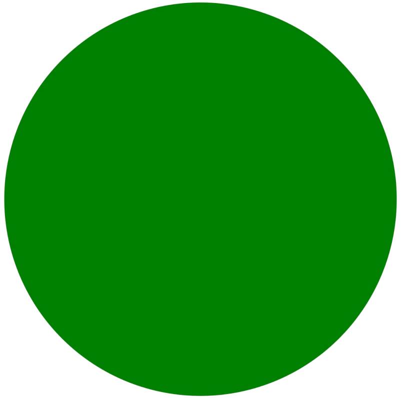
- ❖ Equal for all: Equality of exposure
- ❖ Proportional to quality: Meritocratic fairness
- ❖ Or some middle ground?

What is a desired exposure of an item?

- ❖ Equal for all: Equality of exposure
- ❖ Proportional to quality: Meritocratic fairness
- ❖ Or some middle ground?

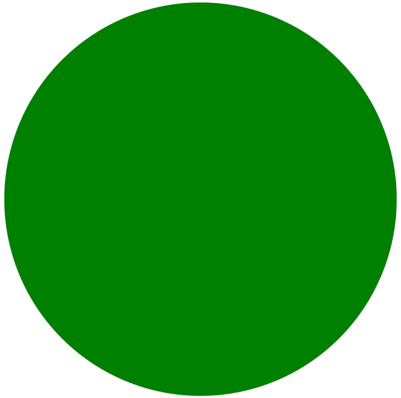
We perceive desired exposure as a necessary controllable knob.

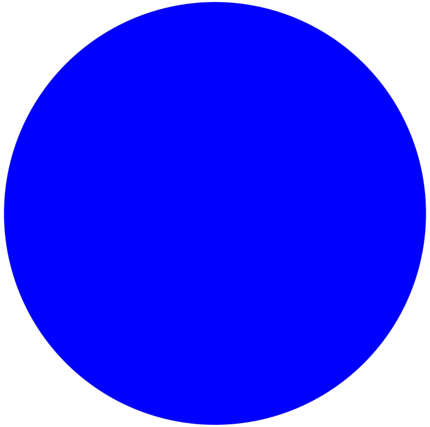
Desired exposure as a control knob



$\boldsymbol{\beta}$ = 1.0→ Equal Exposure

Desired exposure as a control knob

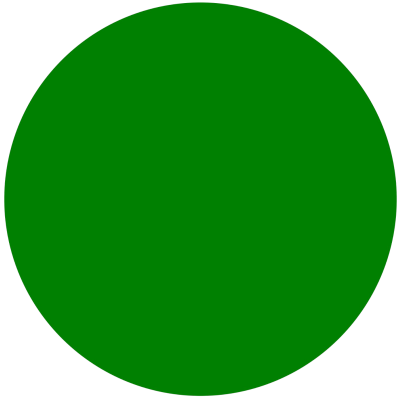




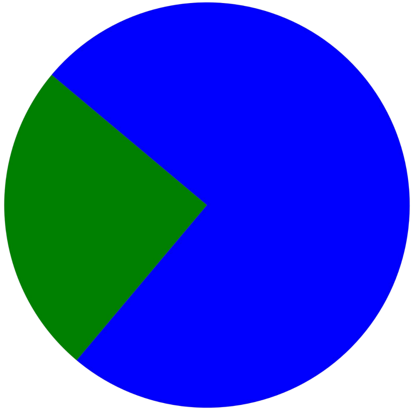
$\boldsymbol{\beta}$ = 1.0→ Equal Exposure

$\boldsymbol{\beta}$ = 0.0→ Meritocracy

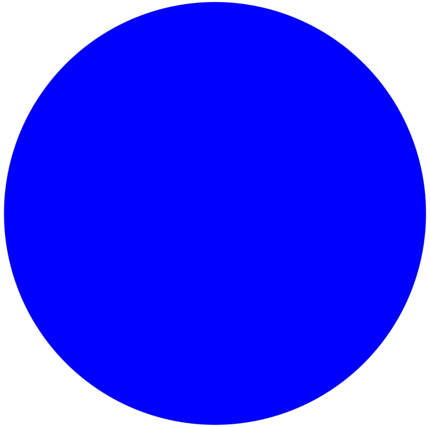
Desired exposure as a control knob



$\boldsymbol{\beta} = 1.0 \rightarrow$ Equal Exposure



$\boldsymbol{\beta} = 0.25 \rightarrow$ 25% of exposure is equally distributed and rest is distributed proportional to merit of the products.



$\boldsymbol{\beta} = 0.0 \rightarrow$ Meritocracy

Desired exposure as a control knob

Two goals for our mitigation strategy

- ❖ Maintain the relatedness of underlying recommendations
- ❖ Minimize / Mitigate the induced exposure bias

β = 1.0→ Equal Exposure

β = 0.25→ 25% of exposure is equally distributed and rest proportional to merit.

β = 0.0→ Meritocracy

Fairness intervention mechanisms

FaiRIR: Suite of fair RIR algorithms

- $FaiRIR_{rl}$: Change the latent space representation such that items with similar desired exposure come closer in the latent space.

- $FaiRIR_{sim}$: Desired exposure based similarity metric

- $FaiRIR_{nbr}$: Select neighboring items based on similarity score and desired exposures.

FaiRIR: Suite of fair RIR algorithms

- *FaiRIR*$_{rl}$: Change the latent space representation such that items with similar desired exposure come closer in the latent space. [Pre-processing]

- *FaiRIR*$_{sim}$: Desired exposure based similarity metric [In-processing]

- *FaiRIR*$_{nbr}$: Select neighboring items based on similarity score and desired exposures. [Post-processing]

FaiRIR: Suite of fair RIR algorithms

- $FaiRIR_{rl}$: Change the latent space representation such that items with similar desired exposure come closer in the latent space. [Pre-processing]

- $FaiRIR_{sim}$: Desired exposure based similarity metric [In-processing]

- $FaiRIR_{nbr}$: Select neighboring items based on similarity score and desired exposures. [Post-processing]

Fair representation learning

- ❖ Inputs:
  - ➢ X : representation learnt from standard RIR algorithms
  - ➢ X : representation encoding desired exposure (Derived from a Desiredness Graph)

- ❖ Goal:
  - ➢ $X^{\sim}$: Fair representation reconciling relatedness and desiredness

Loss function

$$L = \underbrace{\lambda * \sum_{i=1}^{M} \sum_{r=1}^{N} (x_{ir} - \tilde{x}_{ir})^2}_{\text{Relatedness loss}} + \overbrace{\mu * \sum_{i,j \in \{1,\ldots,M\}} [d(\tilde{x}_i, \tilde{x}_j) - d(x_i^*, x_j^*)]^2}^{\text{Desired exposure based similarity loss}}$$

❖ Relatedness loss: Reconstruction loss to be minimized to preserve as much relatedness information learnt as possible.

❖ Desired exposure based similarity loss: Distance between two items *i* and *j* should be preserved in the learnt latent space based on their desired exposure.

Effectiveness in mitigating exposure bias

While only 7% items were getting adequate exposure earlier, now 23% get adequate exposure.

Method	Exposure bias
Vanilla rating-SVD	1.28
FaiRIR_{rl}	0.18

***Exposure bias** = KLD ($E_{observed}$ || $E_{desired}$)

Fair neighbor selection

- We maintain two ranked-list of items based on their
  - Relatedness
  - Desired exposure

- Reconcile between the two: Rank aggregation method based on Borda count

- Intuition: Any item having higher rank in both the ranked lists should be considered the most suitable related item for the given item.

Effectiveness in mitigating exposure bias

While only 7% items were getting adequate exposure earlier, now 100% get adequate exposure.

Method	Exposure bias
Vanilla rating-SVD	1.28
FaiRIR_{nbr}	0.002

***Exposure bias** = KLD ($E_{observed}$ || $E_{desired}$)

Effectiveness in mitigating exposure bias

While only 7% items were getting adequate exposure earlier, now 100% get adequate exposure.

Mitigation strategies are effective in reducing exposure bias.

However, do we have to pay a huge cost in terms of relatedness of the recommendations?

*Exposure bias = KLD ($E_{observed}$ || $E_{desired}$)

Evaluation of Relatedness

Does mitigation of biases come at the cost of relatedness?

- ❖ Genre / Category based similarity
- ❖ Users' propensity to like recommended items

Based on the given datasets

- ❖ User survey to judge the goodness of recommendations

Does mitigation of biases come at the cost of relatedness?

- ❖ Genre / Category based similarity
- ❖ Users' propensity to like recommended items

Based on the given datasets

- ❖ User survey to judge the goodness of recommendations

Design choices of the user survey

- ❖ We consider the top 5% most popular movies (based on #Ratings).
- ❖ We also provide the link to IMDb information page to ensure more reliability.

Design choices of the user survey

- ❖ We consider the top 5% most popular movies (based on #Ratings).
- ❖ We also provide the link to IMDb information page to ensure more reliability.

- ❖ For a particular source movie (*x*)
  - ➢ Generate top-5 recommendations using various algorithms (*y*).
  - ➢ *If your friend likes movie x, how likely are you to recommend movie y?*
  - ➢ Likert scale of [1, 5] (1-- very unlikely, 5-- very likely)
  - ➢ The recommendations were anonymized.

Survey platform

- ❖ We performed the survey on Amazon Mechanical Turk.
- ❖ Recommendations for 100 different source movies were evaluated.
- ❖ 1550 distinct source-recommended movie pairs were evaluated.
- ❖ Each pair was evaluated by at least 10 *AMT master workers.*

Mean relevance score

We evaluate the mean relevance score based on the score each recommended movie received from the AMT workers for the source movies.

Mean relevance score

We evaluate the mean relevance score based on the score each recommended movie received from the AMT workers for the source movies.

Variant of the RIR Algorithm (rating-SVD)	Mean Relevance Score
Vanilla	3.73
$FaiRIR_{rl}$	2.69
$FaiRIR_{sim}$	3.63
$FaiRIR_{nbr}$	3.62

Mean relevance score

We evaluate the mean relevance score based on the score each recommended movie received from the AMT workers for the source movies.

Proposed intervention mechanisms successfully mitigate exposure bias without sacrificing much on relatedness.

FaiRIR_{sim}	3.63
FaiRIR_{nbr}	3.62

Publication from this chapter

A. Dash, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *FaiRIR: Mitigating Exposure Bias from Related Item Recommendations in Two-Sided Platforms*, **IEEE Transactions on Computational Social Systems** (IEEE TCSS), April, 2022.

Research questions addressed in this thesis

How does one quantify bias (if any) induced by related item recommendation algorithms on e-commerce platforms?

How does one generate fair related item recommendations?

How fair and interpretable are the response and default action of voice assistants for e-commerce search queries?

Till now

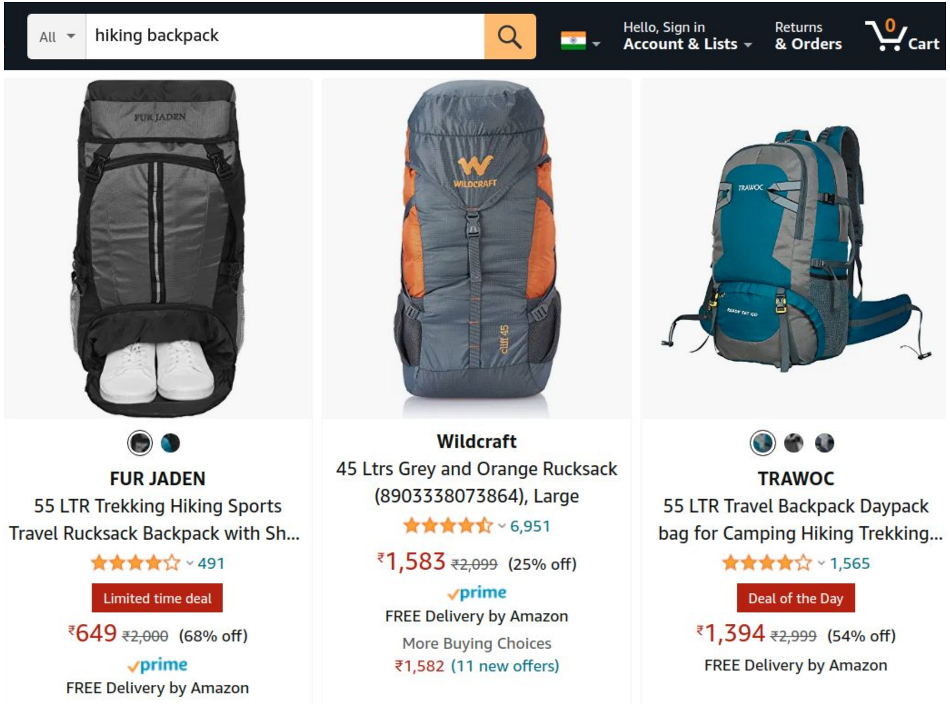
- ❖ We observe biases in environments where choices are served in abundance.
  - ➢ Related Item Recommendation

- ❖ We proposed methodologies to mitigate those biases.

- ❖ The observations are valid for traditional mediums e.g., Desktop search etc..

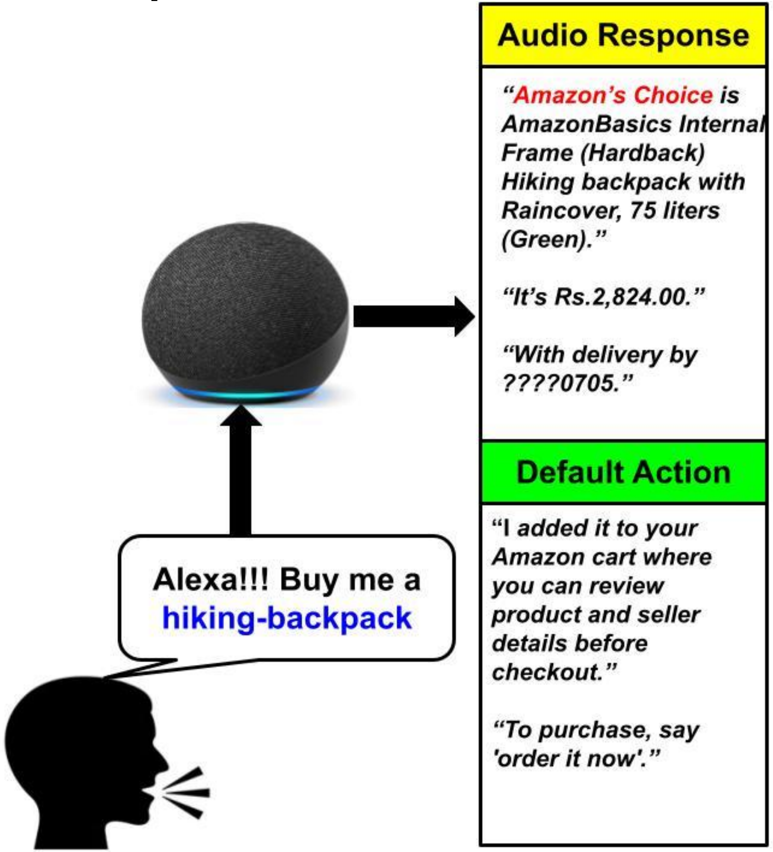
Traditional e-commerce search

- Customers enter query strings
- A ranked **list of products** appear
  - Decreasing order of relevance
  - Metadata of the products shown
- Customers have multiple options to choose from



E-commerce search through Smart Speakers

- ❖ Smart speakers are powered by voice assistants (VAs)
- ❖ Customers utter a query string
- ❖ The details of a **single product** is told
  - ➢ Audio response with a brief explanation
  - ➢ Default action of adding the product to cart



https://mindstreammediagroup.com/introduction-smart-speakers-voice-search-brand-advertisers/

Audio responses from Alexa

"buy me a hiking backpack"

"Amazon's Choice is AmazonBasics Internal Frame (Hardback) Hiking Backpack with Raincover, 75Liters (Green)."

"It's Rs.2,824.00."

"With delivery by ????0705."

"I added it to your Amazon cart where you can review product and seller details before checkout."

"To purchase, say 'order it now'."

Audio responses from Alexa

"buy me a hiking backpack" → **Query string**

"Amazon's Choice is AmazonBasics Internal Frame (Hardback) Hiking Backpack with Raincover, 75Liters (Green)."

"It's Rs.2,824.00."

"With delivery by ????0705."

"I added it to your Amazon cart where you can review product and seller details before checkout."

"To purchase, say 'order it now'."

Audio responses from Alexa

"buy me a hiking backpack" → **Query string**

"Amazon's Choice is AmazonBasics Internal Frame (Hardback) Hiking Backpack with Raincover, 75Liters (Green)."

"It's Rs.2,824.00."

"With delivery by ????0705."

→ **Product details**

"I added it to your Amazon cart where you can review product and seller details before checkout."

"To purchase, say 'order it now'."

Audio responses from Alexa

"buy me a hiking backpack"

"Amazon's Choice is AmazonBasics Internal Frame (Hardback) Hiking Backpack with Raincover, 75Liters (Green)."

"It's Rs.2,824.00."

"With delivery by ????0705."

"I added it to your Amazon cart where you can review product and seller details before checkout."

"To purchase, say 'order it now'."

Query string

Brief explanation

Product details

Audio responses from Alexa

"buy me a hiking backpack"

Query string

"Amazon's Choice is AmazonBasics Internal Frame (Hardback) Hiking Backpack with Raincover, 75Liters (Green)."

Brief explanation

"It's Rs.2,824.00."

"With delivery by ????0705."

Product details

"I added it to your Amazon cart where you can review product and seller details before checkout."

Default action

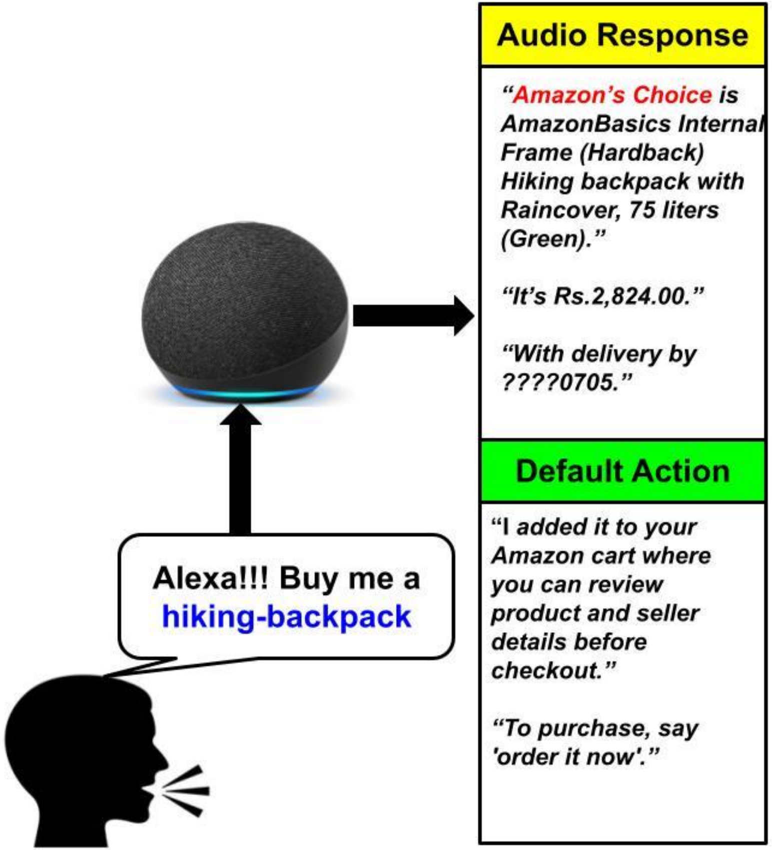
"To purchase, say 'order it now'."

E-commerce search through Smart Speakers

- ❖ Smart speakers are powered by voice assistants (VAs)
- ❖ Customers utter a query string
- ❖ The details of a **single product** is told
  - ➢ Audio response with a brief explanation
  - ➢ Default action of adding the product to cart
- ❖ Unlike traditional product search, customers do not have many options to choose from.

https://mindstreammediagroup.com/introduction-smart-speakers-voice-search-brand-advertisers/

In this paper



RQ 1: How do customers interpret the explanations given in the audio response by the VA?

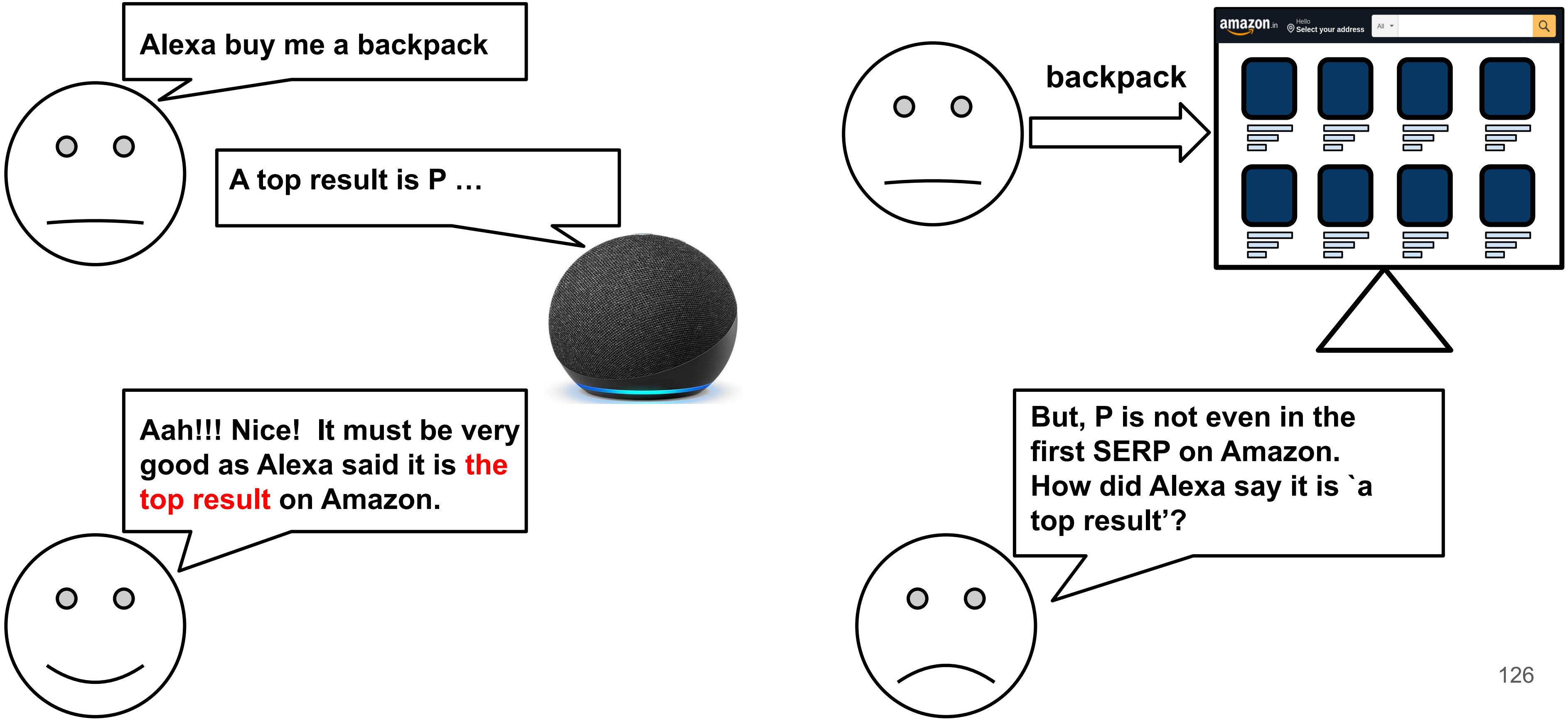
RQ 2: How fair is the default action by the VA?

Why are these questions important?

Gap in explanation and Interpretation of customers

Gap in explanation and Interpretation of customers

Gap in explanation and Interpretation of customers



Fairness in the default action by VA

- ❖ VA selects one specific product as part of its default action
- ❖ Customers have a tendency to take the path of less effort.
- ❖ Default option often comes as an endorsement from the choice architect.
- ❖ Hence, likelihood of choosing / purchasing the default option is generally high.

Thaler, Richard H., and Cass R. Sunstein. "Nudge: Improving decisions about health, wealth, and happiness."

Fairness in the default action by VA

- VA selects one specific product as part of its default action
- Customers have a tendency to take the path of less effort.
- Default option often comes as an endorsement from the choice architect.
- Hence, likelihood of choosing / purchasing the default option is generally high.

- Non-selection of the most relevant product as the default choice
  - may deny its producers sales and revenue opportunities
  - may mislead customers to (possibly) less relevant products

Research questions

RQ 1: How do customers interpret the explanations given in the audio response by the VA?

RQ 2: How fair is the default action by the VA?

Challenges in understanding customers interpretation

- These explanations have several semantics and nuances attached
- Customers are more conversant with traditional mediums e.g., Desktop e-commerce search.

https://tinuiti.com/wp-content/uploads/2020/03/2020_Amazon_Shopper_Survey_02.pdf

Challenges in understanding customers interpretation

- These explanations have several semantics and nuances attached
- Customers are more conversant with traditional mediums e.g., Desktop e-commerce search.
- One way is to contrast the customers’ interpretation of the explanations with observations on desktop search.

https://tinuiti.com/wp-content/uploads/2020/03/2020_Amazon_Shopper_Survey_02.pdf

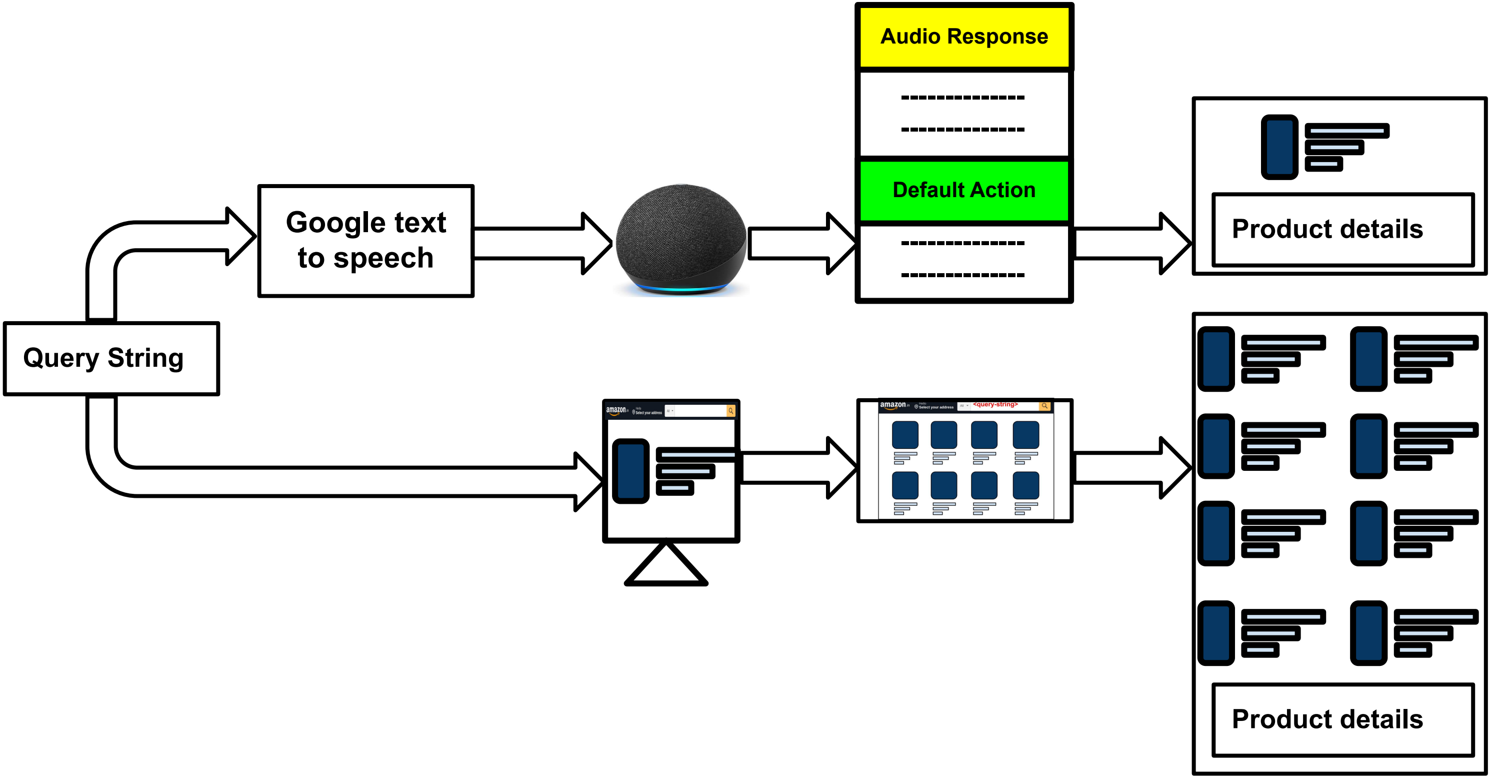
Challenges in understanding customers interpretation

- ❖ These explanations have several semantics and nuances attached
- ❖ Customers are more conversant with traditional mediums e.g., Desktop e-commerce search.
- ❖ One way is to contrast the customers' interpretation of the explanations with observations on desktop search.

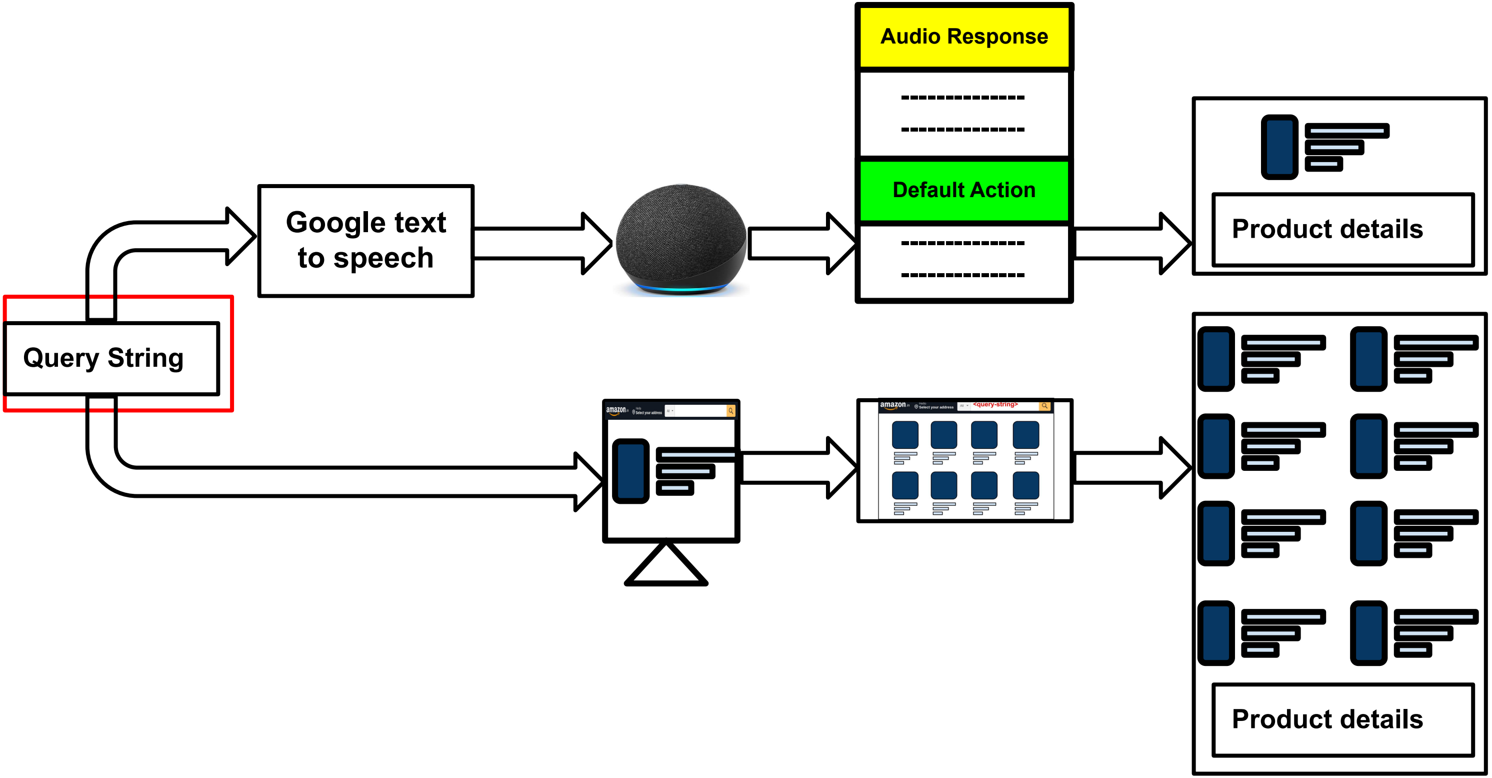
How to design an effective data collection framework for such analysis?

https://tinuiti.com/wp-content/uploads/2020/03/2020_Amazon_Shopper_Survey_02.pdf

Data collection for a meaningful comparison

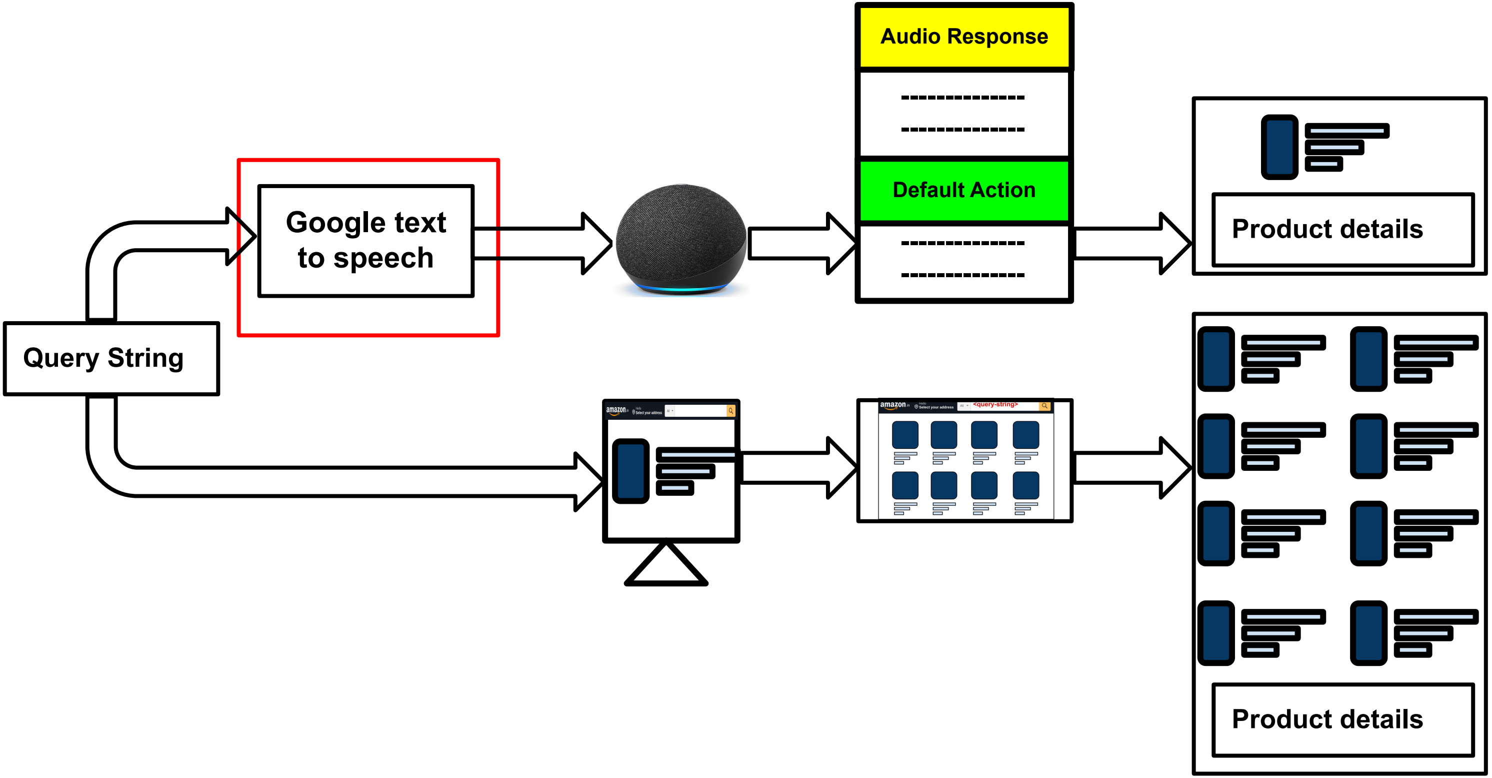


Data collection for a meaningful comparison



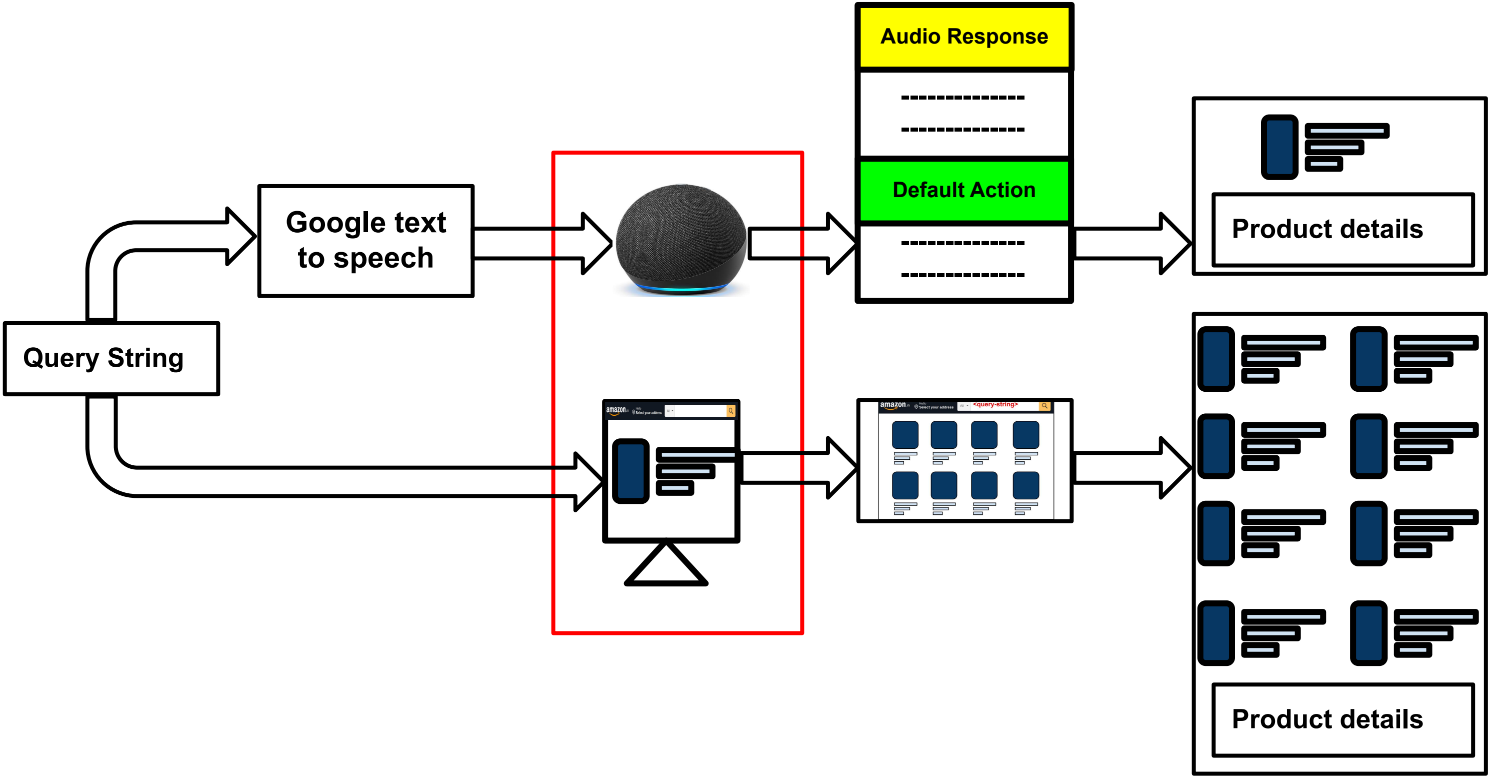
- ➔ Query string

Data collection for a meaningful comparison



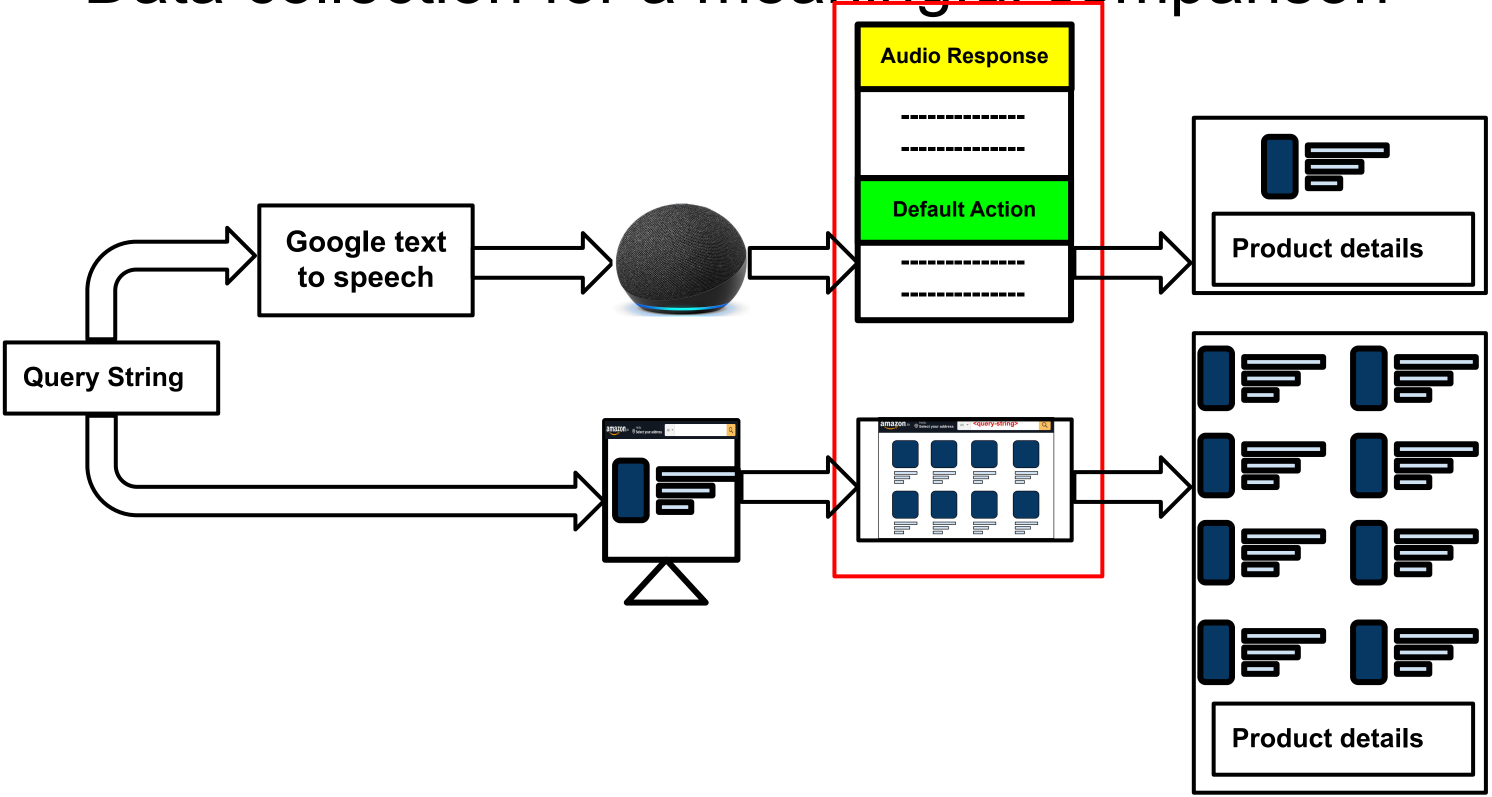
- ➔ Query string
- ➔ gTTS to generate audio command

Data collection for a meaningful comparison



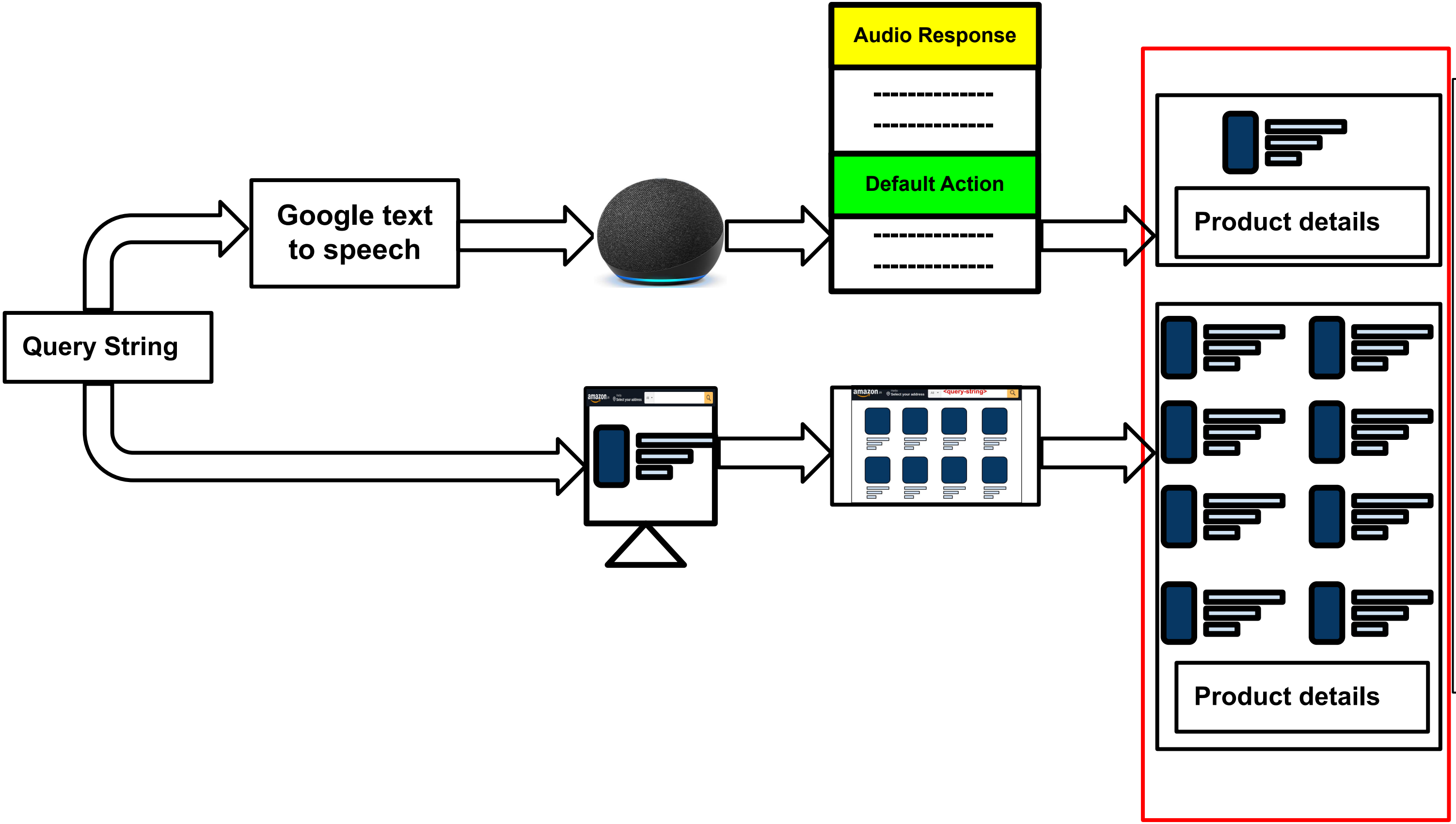
- ➔ Query string
- ➔ gTTS to generate audio command
- ➔ At the same time search on Alexa and desktop

Data collection for a meaningful comparison



- ➔ Query string
- ➔ gTTS to generate audio command
- ➔ At the same time search on Alexa and desktop
- ➔ Get the audio response and desktop SERP

Data collection for a meaningful comparison



- ➔ Query string
- ➔ gTTS to generate audio command
- ➔ At the same time search on Alexa and desktop
- ➔ Get the audio response and desktop SERP
- ➔ Collect the product details

Some statistics of the dataset collected

- ❖ We selected the top-100 keywords searched on Amazon and extended it to a set of 1000 queries for our data collection.
- ❖ We collected data for 1000 query strings on Amazon.
- ❖ We also collected 14 temporal snapshots for top-100 queries on Amazon.
- ❖ The queries cover 10 popular product categories on Amazon.

Some statistics of the dataset collected

- We selected the top-100 keywords searched on Amazon and extended it to a set of 1000 queries for our data collection.
- We collected data for 1000 query strings on Amazon.
- We also collected 14 temporal snapshots for top-100 queries on Amazon.
- The queries cover 10 popular product categories on Amazon.

For access to dataset, please refer to the form at this link: https://forms.gle/aEG2n84Ay82QkVD19

Most prevalent explanations

"buy me a hiking backpack"

"Amazon's Choice is AmazonBasics Internal Frame (Hardback) Hiking Backpack with Raincover, 75Liters (Green)."

"It's Rs.2,824.00."

"With delivery by ????0705."

"I added it to your Amazon cart where you can review product and seller details before checkout."

"To purchase, say 'order it now'."

66% queries out of 1000

Most prevalent explanations

"buy me a hiking backpack"

"Amazon's Choice is AmazonBasics Internal Frame (Hardback) Hiking Backpack with Raincover, 75Liters (Green)."

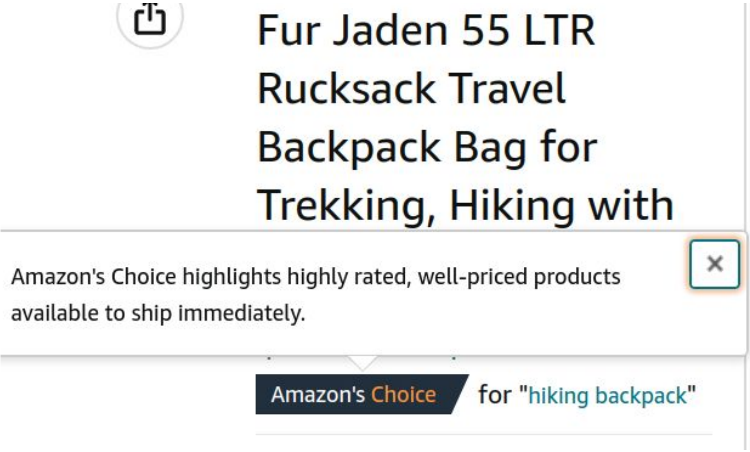
"It's Rs.2,824.00."

"With delivery by ????0705."

"I added it to your Amazon cart where you can review product and seller details before checkout."

"To purchase, say 'order it now'."

66% queries out of 1000



*Amazon's choice highlights **highly rated**, **well-priced** products available to ship immediately.*

User survey for understanding interpretation

- ❖ We conducted a survey among 100 participants.
- ❖ Most of them are conversant with Amazon platform.

User survey for understanding interpretation

- ❖ We conducted a survey among 100 participants.
- ❖ Most of them are conversant with Amazon platform.

- ❖ We asked them different questions pertaining to
  - ➢ What do customers interpret by `highly rated’ (or `well priced’) products?
  - ➢ Where do customers expect Amazon’s Choice products to appear?
  - ➢ How likely are you to buy the product which is explained as “Amazon’s Choice”?

What do customers interpret by `highly rated’ product?

❖ According to 59% respondents, a product with avg. rating greater than 4.0 out of 5.0 can be considered highly rated.

What do customers interpret by `highly rated’ product?

- ❖ According to 59% respondents, a product with avg. rating greater than 4.0 out of 5.0 can be considered highly rated.

- ❖ In all the 662 queries, the product selected with Amazon’s choice explanation has an average rating greater than 4.0.

What do customers interpret by `well priced’ product?

❖ 61% respondents consider a product with price among the least 5 prices among all the products shown in SERP to be well priced.

What do customers interpret by `well priced' product?

- 61% respondents consider a product with price among the least 5 prices among all the products shown in SERP to be well priced.

- In merely 23% cases, product added to cart adhered to the most common interpretation as mentioned above.

Major takeaways from the survey

Explanation type	Statement	Interpretation (% votes)	Match
Amazon's Choice	Highly rated	Avg. user rating ≥ 4.0 (59%)	✓ (100%) ✗ (00%)
	Well priced	Least-5 price (61%)	✓ (23%) ✗ (77%)
	Expected position	Top-5 in SERP (54%)	✓ (74%) ✗ (26%)
A top result	Expected position	Top result (position 1) (62%)	✓ (19%) ✗ (81%)

Major takeaways from the survey

Respondents' interpretations of the VA's explanation do not align with observations from desktop search results in majority of the cases.

Major takeaways from the survey

Respondents' interpretations of the VA's explanation do not align with observations from desktop search results in majority of the cases.

Additionally, 56% customers answered that they are likely or very likely to purchase products with such explanations.

Research questions

RQ 1: How do customers interpret the explanations given in the audio response by the VA?

RQ 2: How fair is the default action by the VA?

(Un)Fairness toward customers

- Is the product added to cart by Alexa also preferred by customers?
- We conducted a user survey among the same 100 respondents
- We showed the customers:
  - Product that was added to cart by Alexa
  - Product that appeared as the top result in the corresponding desktop search
- Which of the following two products would you prefer to buy?

Observation

- For 22 out of the 30 queries (73.3%), the majority preference of respondents did not match with the products selected by Alexa.

- Out of the 1000 evaluations, we find that in 73.2% cases, respondents preferred the top desktop search result to the product selected by Alexa.

Observation

❖ For 22 out of the 30 queries (73.3%), the majority preference of respondents

These observations further underpin the potential unfairness and customer dissatisfaction concerns due to default product selection of Alexa.

preferred the top desktop search result to the product selected by Alexa.

Summary

- There exists significant gap in customers’ interpretation of Alexa’s explanations and observation in traditional desktop search.
- The default product selection of Alexa has potential unfairness concerns for both customers and producers / sellers.
- Since customers cede complete autonomy to smart speakers (and the VAs therein), VAs should be more responsible during such interactions.

Publication from this chapter

A. Dash, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *Alexa, in you, I trust! Fairness and Interpretability Issues in E-commerce Search through Smart Speakers*, In Proceedings of **The ACM Web Conference** (WWW), April, 2022.

Future directions

- ❖ Special relationships beyond the private labels need thorough investigation.
- ❖ More robust fair recommender systems need to be designed which consider
  - ➢ Customers
  - ➢ Producers / Sellers
  - ➢ Platform
  - ➢ And their special relationships
- ❖ Design choices and algorithms need to be improved when customers interact with digital systems esp., when customers' autonomy is reduced.

Publications from the thesis (reverse chronology)

- **A. Dash**, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *FaiRIR: Mitigating Exposure Bias from Related Item Recommendations in Two-Sided Platforms*, IEEE Transactions on Computational Social Systems (IEEE TCSS), April, 2022.

- **A. Dash**, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *Alexa, in you, I trust! Fairness and Interpretability Issues in E-commerce Search through Smart Speakers*, In Proceedings of The ACM Web Conference (WWW), April, 2022.

- **A. Dash**, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *When the Umpire is also a Player: Bias in Private Label Product Recommendations on E-commerce Marketplaces*, In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, (ACM FAccT), March, 2021.

- **A. Dash**, A Mukherjee, S. Ghosh, *A Network-centric Framework for Auditing Recommendation Systems*, In Proceedings of the IEEE International Conference on Computer Communications (IEEE INFOCOM), April, 2019.

Thesis organization

- ❖ Chapter 1 : Introduce and motivate the line of work in the dissertation
- ❖ Chapter 2 : Surveys the relevant related works
- ❖ Chapter 3 : Introduces RIN framework and performs an audit of Amazon RIN
- ❖ Chapter 4 : Proposes fair related item recommendation algorithms (FaiRIR)
- ❖ Chapter 5 : Introduces fairness and interpretability concerns in conversational e-commerce search via smart speakers
- ❖ Chapter 6 : Summarizes the contributions and future directions

Acknowledgement

- Collaborators
  - Prof. Krishna P. Gummadi (MPI-SWS)
  - Dr. Abhijnan Chakraborty (IIT Delhi))

- TCS RSP

- Microsoft, ACM IARCS, SIGCHI Gary Marsden Student Development Fund

- CNeRG, IIT Kharagpur

All publications (reverse chronology)

- **A. Dash**, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *FaiRIR: Mitigating Exposure Bias from Related Item Recommendations in Two-Sided Platforms*, IEEE Transactions on Computational Social Systems (IEEE TCSS), April, 2022.

- **A. Dash**, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *Alexa, in you, I trust! Fairness and Interpretability Issues in E-commerce Search through Smart Speakers*, In Proceedings of The ACM Web Conference (WWW), April, 2022.

- S. D. Jaiswal, K Duggirala, **A. Dash**, A. Mukherjee, *Two-Face: Adversarial Audit of Commercial Face Recognition Systems*, To be published in Proceedings of the International Conference On Web and Social Media (AAAI ICWSM), June, 2022.

- **A. Dash**, A. Chakraborty, S. Ghosh, A. Mukherjee, K. P. Gummadi, *When the Umpire is also a Player: Bias in Private Label Product Recommendations on E-commerce Marketplaces*, In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, (ACM FAccT), March, 2021.

- A. Shandilya, **A.Dash**, A. Chakraborty, K. Ghosh and S. Ghosh, *Fairness for Whom? Understanding the Reader's Perception of Fairness in Text Summarization*, International Workshop on Fair and Interpretable Learning Algorithms (FILA 2020), In conjunction with IEEE BigData, December, 2020.

- **A. Dash**, A Shandilya, A. Biswas, K. Ghosh, S. Ghosh, A. Chakraborty, *Summarizing User-generated Textual Content: Motivation and Methods for Fairness in Algorithmic Summaries*, In Proceedings of the ACM on Human-Computer Interaction, vol. 3, No. CSCW, Article 172, November 2019.

- **A. Dash**, A Mukherjee, S. Ghosh, *A Network-centric Framework for Auditing Recommendation Systems*, In Proceedings of the IEEE International Conference on Computer Communications (IEEE INFOCOM), April, 2019.

Thank You

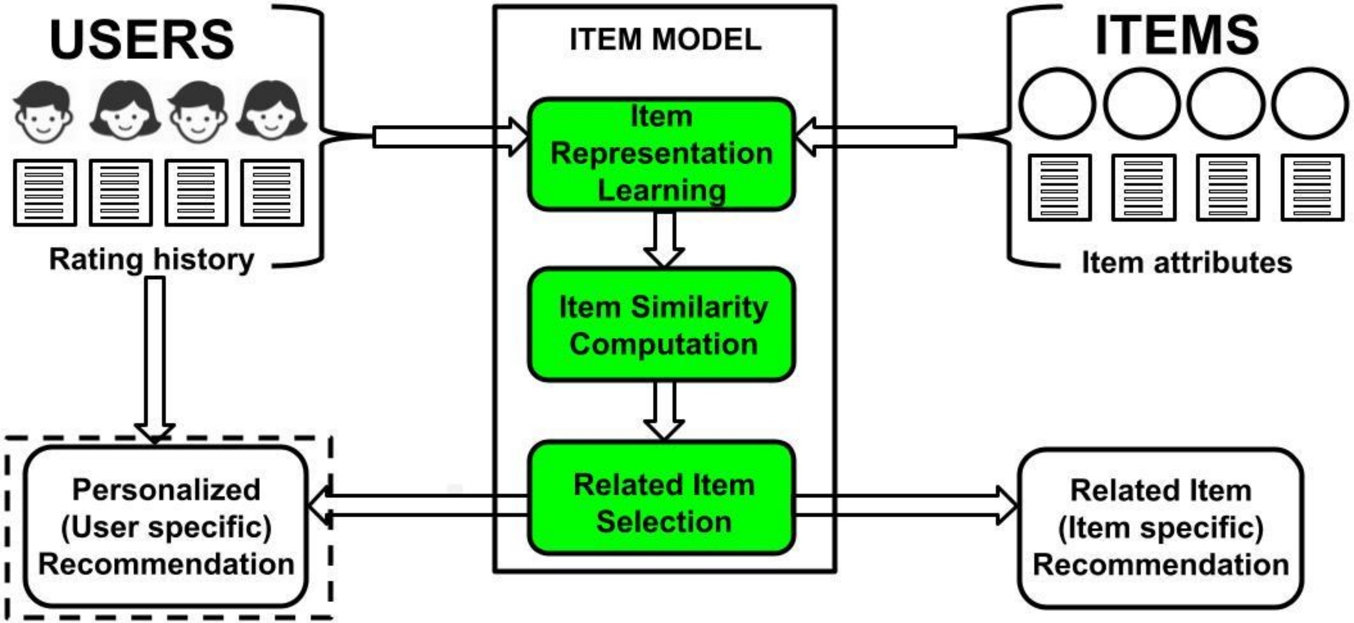
`Amazon's Choice' and special relationships

- ❖ When Amazon PL products appear in search results, for 45% queries they are added to cart with *Amazon's Choice* explanation.

- ❖ 78% products added to cart with Amazon's choice explanation were sold by Amazon Special Merchants. (Cloudtail or Appario)

Future work: The intricacies of conversational medium and effect of special relationships in response and product selection.

Block diagram of RIR algorithms



Why post-processing method is so effective?

- ❖ #Recommendations an item gets become proportional to its desired exposure.
- ❖ Post-processing method does not alter the learnt representation of vanilla RIRs.
- ❖ Pre-processing methods are often application agnostic.

Where do customers expect Amazon’s Choice products to appear?

- 54% respondents expect them to appear in top-5 positions; while 30% respondents expect them to appear as the top search result.

Where do customers expect Amazon's Choice products to appear?

- ❖ 54% respondents expect them to appear in top-5 positions; while 30% respondents expect them to appear as the top search result.
- ❖ In 74% cases, Amazon's Choice product appeared in the top-5 positions.

Where do customers expect Amazon’s Choice products to appear?

- 54% respondents expect them to appear in top-5 positions; while 30% respondents expect them to appear as the top search result.
- In 74% cases, Amazon’s Choice product appeared in the top-5 positions.

- The contribution from position 1 is merely 39%.
- Worse, nearly 8% of the times, it does not even appear in the first SERP.

(Un)Fairness toward producers

- Exposure due to Alexa:
  - 1 if product is added to cart, 0 otherwise
- Exposure due to ranked results of desktop search
  - Can be evaluated using any standard attention distribution mechanism
  - We assume attention / exposure are distributed geometrically
- Exposure bias is the difference between the two exposures
  - Ideally, Exposure bias ≃ 0.

Observation

In 68% of all 1000 queries, exposure bias is non-zero →most relevant product according to desktop search was not added to cart by Alexa.

Observation

Less relevant products were added to cart; thus, depriving producers of more relevant products an opportunity to potential sales and revenue.

Summary

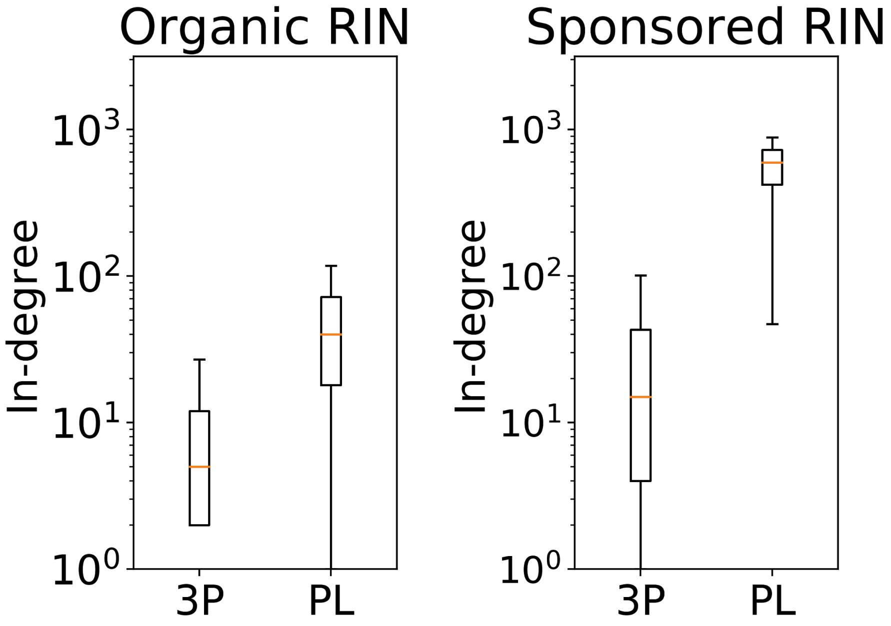
- ❖ We show how sponsored recommendations on Amazon are being used to nudge users toward Amazon private labels.
- ❖ Promotion of private labels is by no means illegal.
- ❖ It is concerning because marketplaces e.g., Amazon are involved in both production and dissemination of products.

Summary

- ❖ We show how existing RIR algorithms can implicitly induce exposure bias.
- ❖ Most often the exposure bias can ***not*** be explained by the quality of the products.
- ❖ We propose three mitigation strategies as interventions in the RIR pipeline.
- ❖ Our proposed mechanisms reduces exposure bias without sacrificing much on the relatedness of the recommendations.

Promotion bias

- Comparison of **in-degree distributions** of different kind of products.

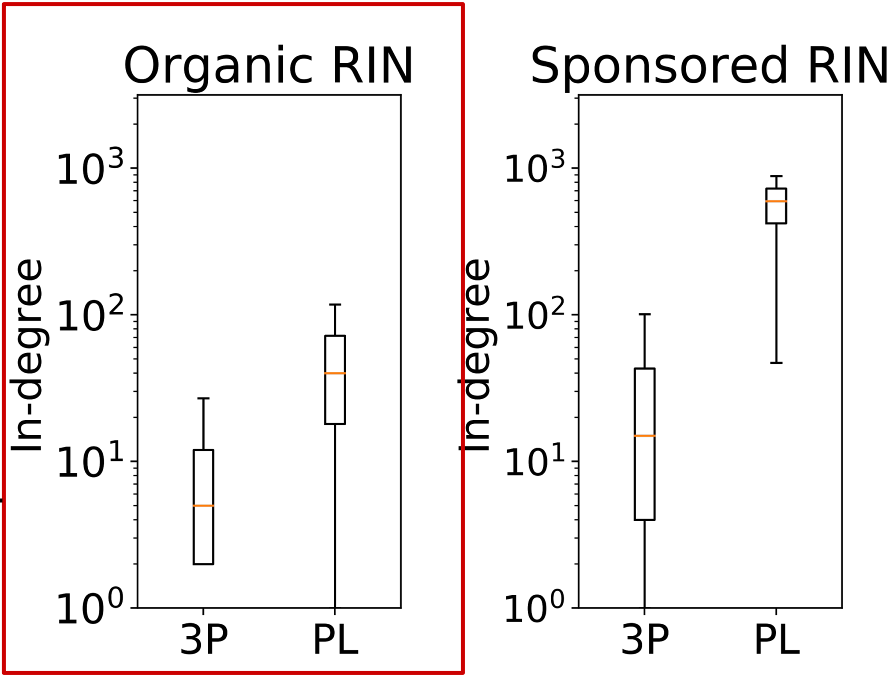


*Y-axes are in log-scale.

Promotion bias

❖ Comparison of **in-degree distributions** of different kind of products.

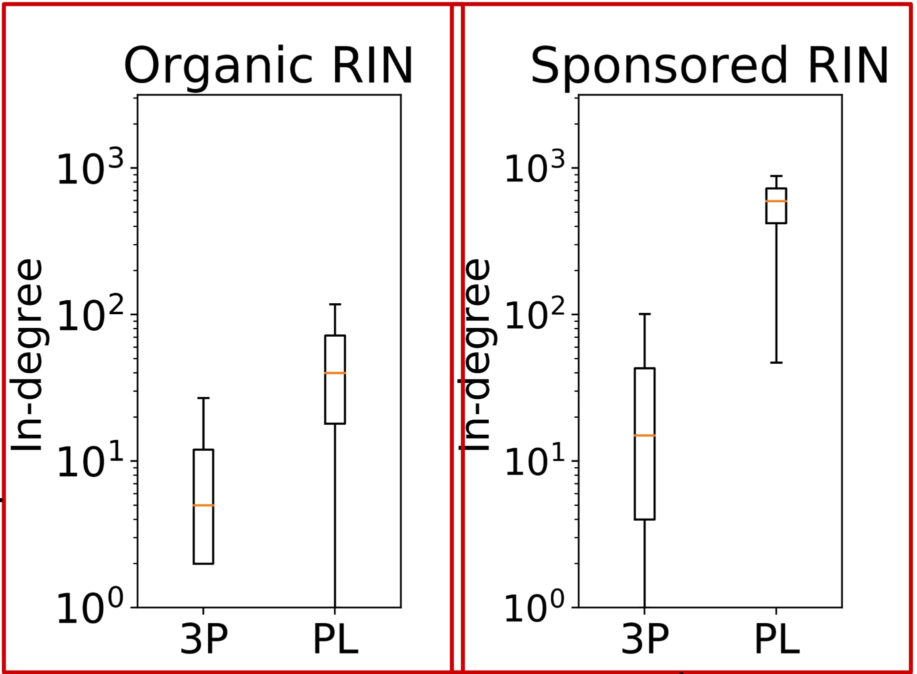
Considerable disparity in the in-degree distributions in the Organic RIN.



*Y-axes are in log-scale.

Promotion bias

❖ Comparison of **in-degree distributions** of different kind of products.



Considerable disparity in the in-degree distributions in the Organic RIN.

Disparity further increases significantly in the Sponsored RIN.

*Y-axes are in log-scale.