

Improving Economic Welfare in Reputation-based Marketplaces with Truth Warrants



Swapneel Mehta
Postdoctoral Associate

Joint work with Aaron Nichols, Nina Mazar, and Marshall Van Alstyne



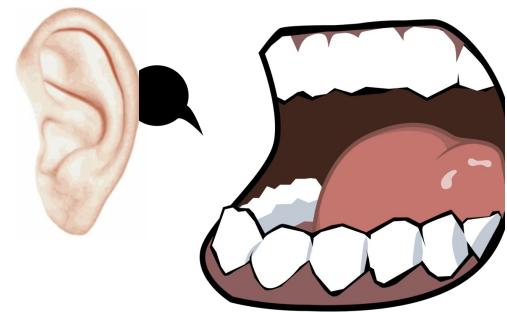
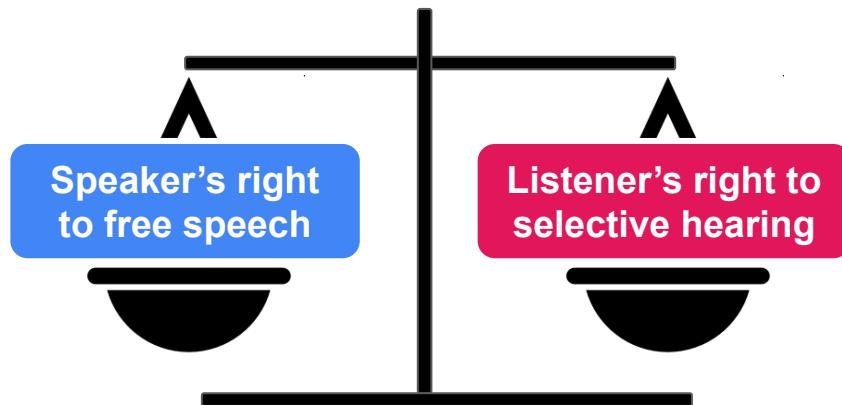
Can our intervention help sell more “honest products” while protecting profits



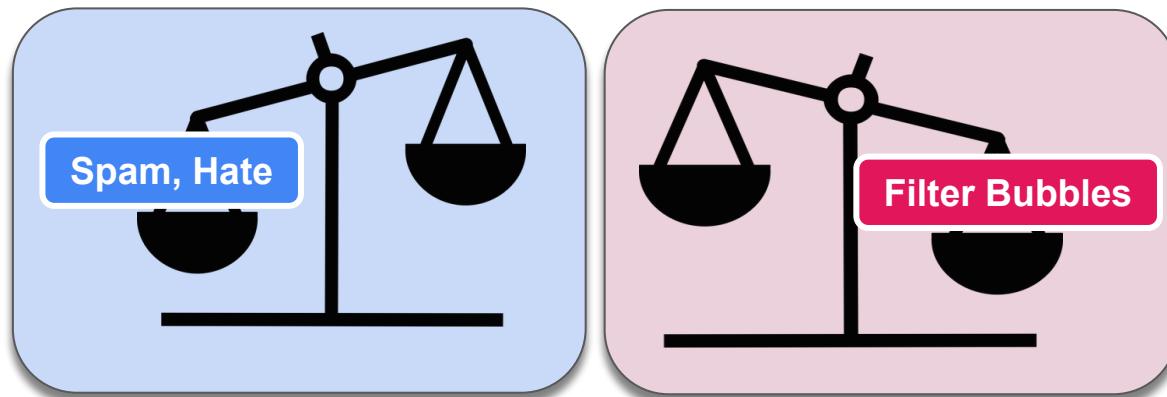
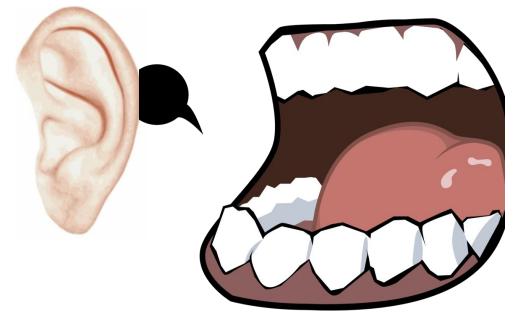
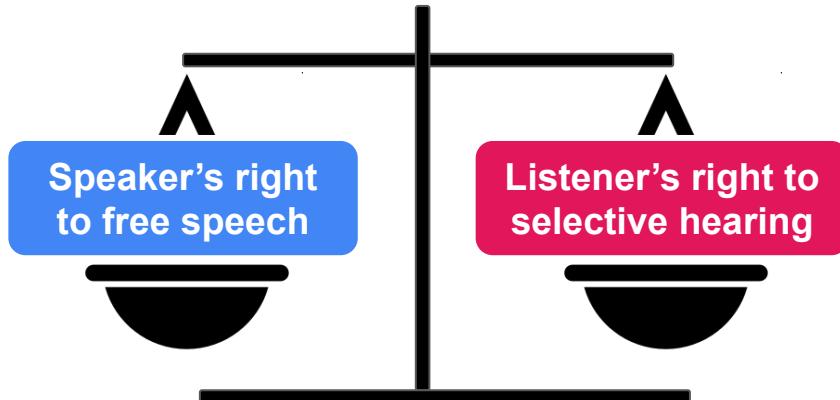
Swapneel Mehta
Postdoctoral Associate

Joint work with Aaron Nichols, Nina Mazar, and Marshall Van Alstyne

Platforms need to balance the rights of speakers and listeners



Platforms need to balance the rights of speakers and listeners



If either one of the two is given preferential treatment, the other suffers!

How to find an equilibrium?

Fake News “Interventions” Don’t Work Too Well These Days

Arms Race

Bot Detection

Spam

Phishing

Adversarial Practices



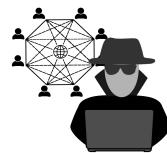
Fake News “Interventions” Don’t Work Too Well These Days

Arms Race

Bot Detection
Spam
Phishing
Adversarial Practices

Discrediting Raters

Flagging
Fact-checking
Debunks
Central Authority



Fake News “Interventions” Don’t Work Too Well These Days

Arms Race

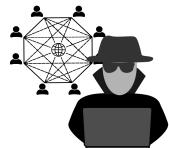
Bot Detection
Spam
Phishing
Adversarial Practices

Discrediting Raters

Flagging
Fact-checking
Debunks
Central Authority

Misplaced Responsibility

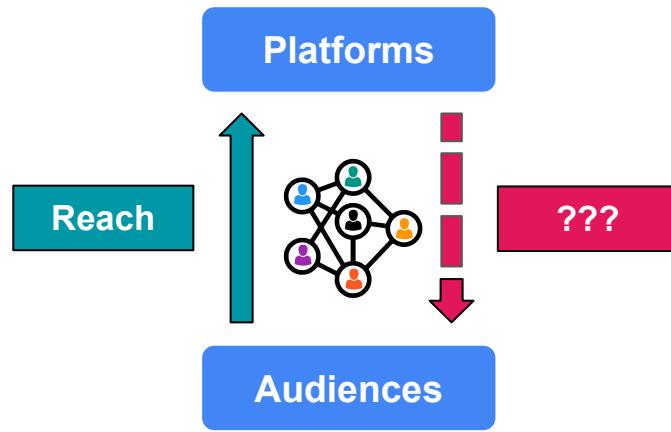
Auditing Algorithms
Deamplification
Prebunks
Burden on Platforms, Listeners



Hate Speech’s Rise on Twitter Is Unprecedented, Researchers Find

What if we introduced consequences to producing fake news?

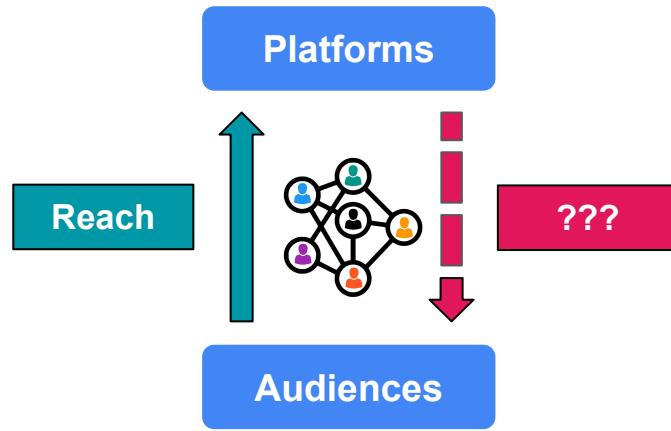
Platforms should support Free Speech, but not *Free Reach*



Audiences provide platforms with *reach*

Platforms should be accountable to protect audiences, at least global regulators think so.

Platforms should support Free Speech, but not *Free Reach*



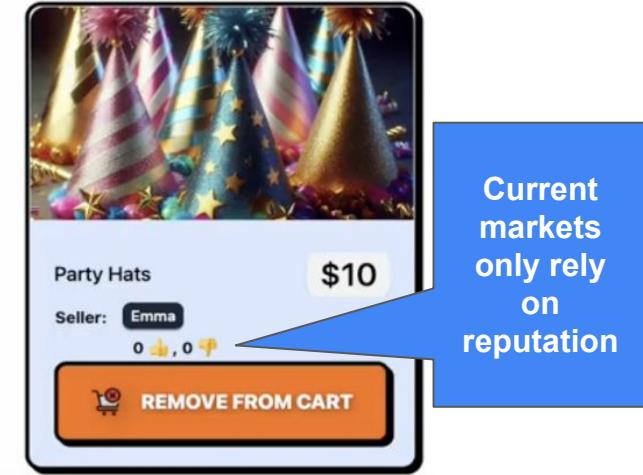
Audiences provide platforms with *reach*

Platforms should be accountable to protect audiences, at least global regulators think so.

Free speech is absolutely vital + legal right.

But we need mechanisms to introduce *accountability* in exchange for *reach!*

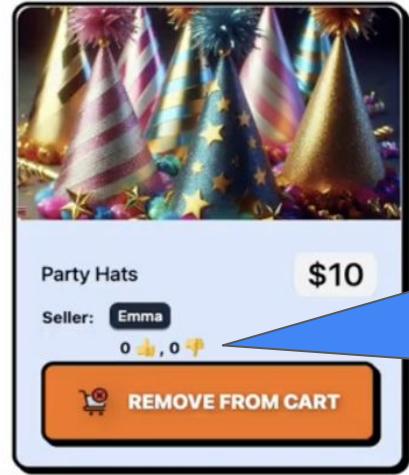
Improving Honesty in Ads Marketplace via *Self-Certification*



Current markets only rely on reputation

Current markets are Reputation-based (👍 / 👎)

Improving Honesty in Ads Marketplace via *Self-Certification*



Current markets only rely on reputation



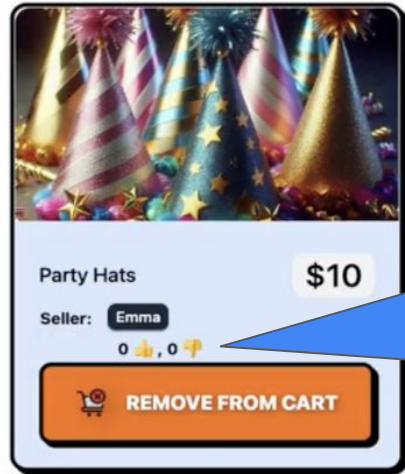
We believe “Warrant” Label can introduce pre-sales seller accountability

Current markets are Reputation-based (👍 / 👎)

Warrant = signaling mechanism.

Seller **escrows** collateral to back claims. If claim is false and is challenged, money is lost!

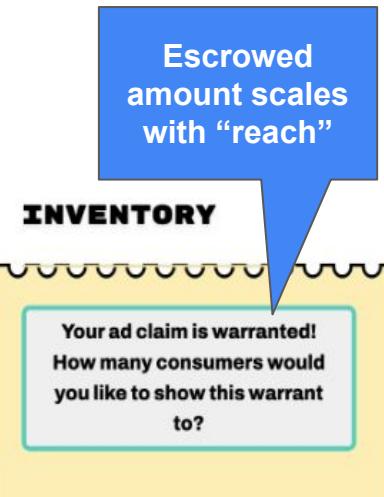
Improving Honesty in Ads Marketplace via *Self-Certification*



Current markets only rely on reputation



We think "Warrant" Label increases credibility



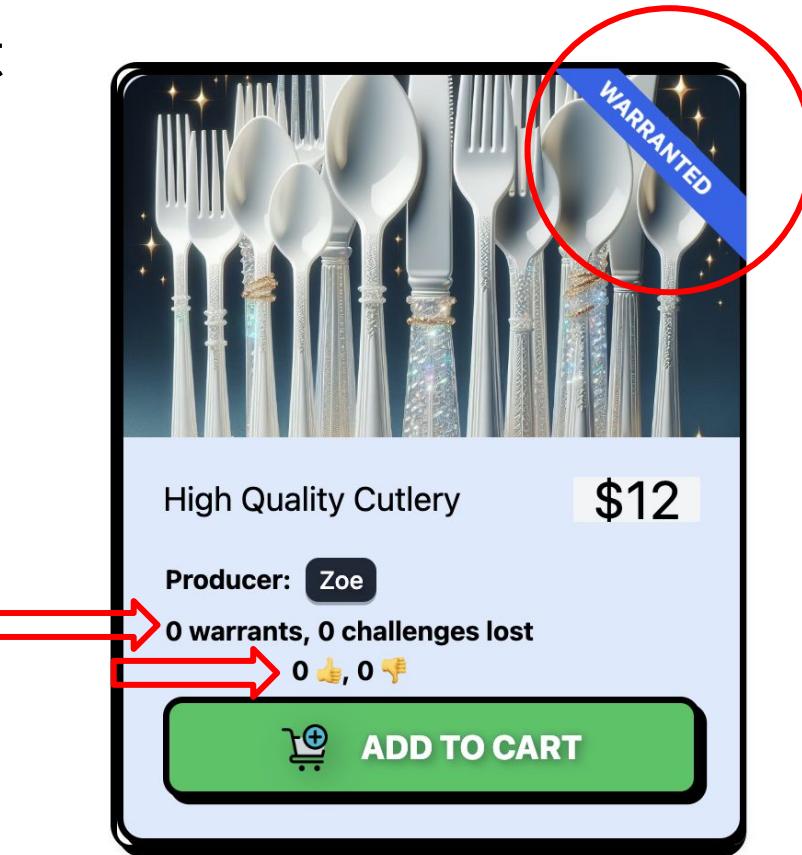
Current markets are Reputation-based (👍 / 👎)

Self-Certification = seller escrows extra money to back claims. If community agrees claim false, money lost!

Warrants are shown to a predetermined no. of people.

Three Signals in an Advertisement

1. Rating - Thumbs Up / Down
2. Self-Certification - Warrant Label Added to Advertisement
3. 'Warrant Histories' - History of self-certification and challenges

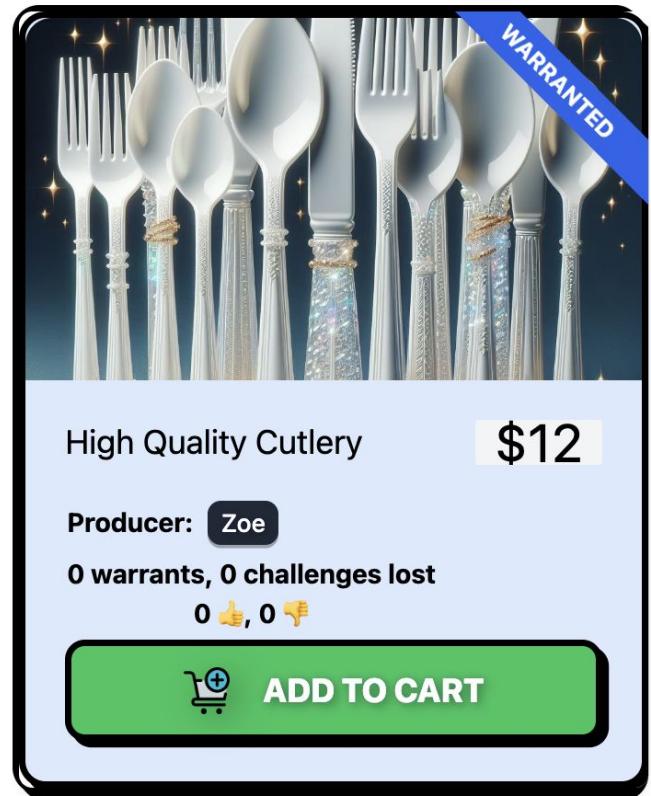


Three Signals in an Advertisement

1. Rating - Thumbs Up / Down
2. Self-Certification - Warrant Label Added to Advertisement
3. Warrant Histories - Warrants and Challenges

Compare economic value of each of these signals!

- ❖ Do they incentivise honest purchases?
- ❖ Do they boost profits?
- ❖ Do they increase consumer utility?



Design a two-sided experiment with buyers and sellers



- ❖ N = 250 players, between-subjects design
- ❖ 7 Rounds of Sales in each Game
- ❖ **1750 Rounds of Gameplay**
- ❖ 6 Sellers vs. 1 Buyer in each Game
- ❖ Buyers may self-certify their claims with collateral
- ❖ Buy from multiple products in marketplace based on capital

Sellers: 6 adaptive ‘Bots’ reflecting variance in honesty

1. **Honest:** Always produce high quality, advertised as high quality.



2. Reformed Cheat: produce low quality until sold, then switch to high quality.

3. Politician: produce high quality until two sales, switch to low quality, and back.

4. Bait-and-Switch: produce high quality until sold, switch to low quality and back.

5. Goldfish: produce low quality until sold, switch to high quality and back.



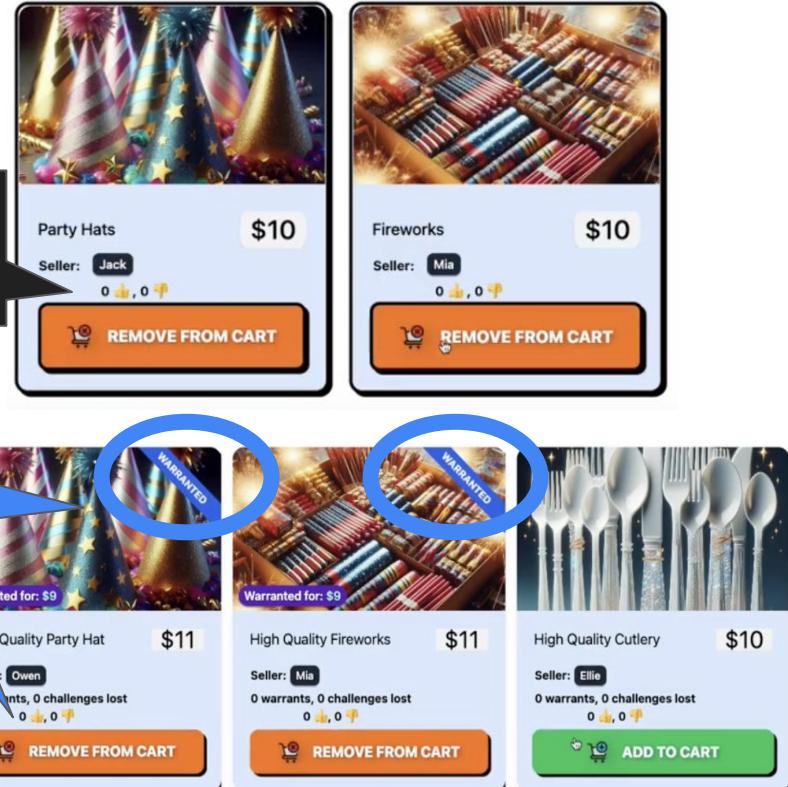
6. **Cheater:** Always produce low quality until sold, advertised as high quality.

Designing a two-sided online ads market with humans and AI



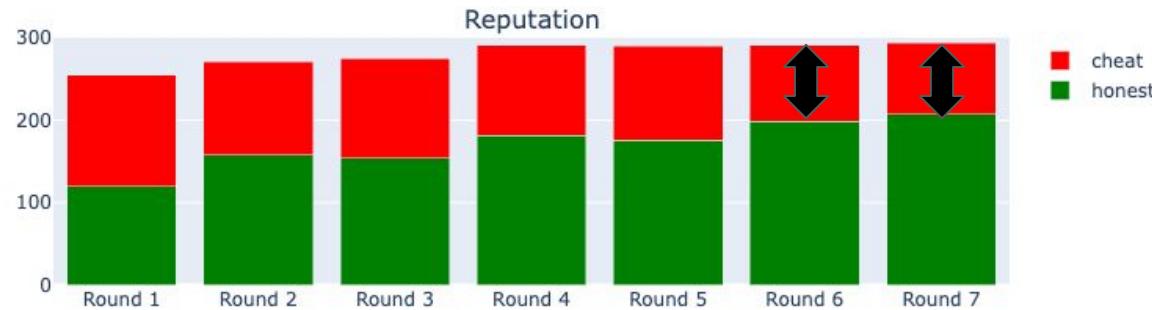
Control
“Reputation”
Market

Treatment
“Warrants”
Market



Warrants markets are less likely to cheat the consumer

Honest and Cheating Sales by Round

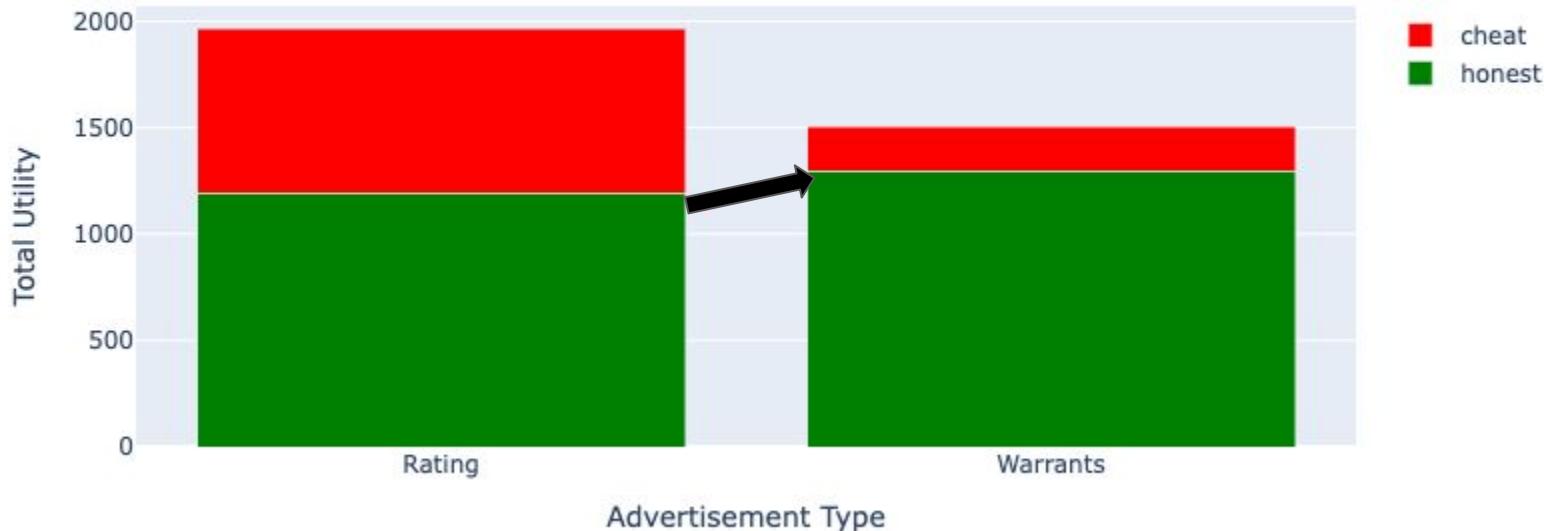


Warrant

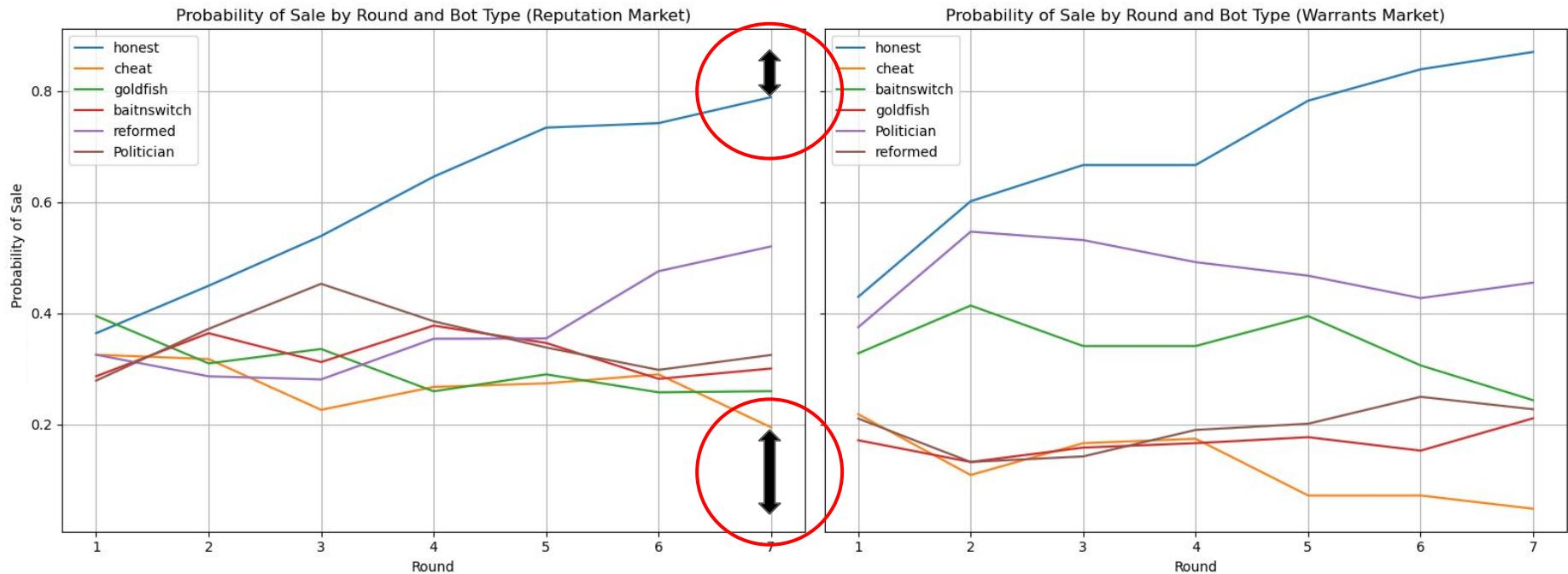


Self-certification increases consumer utility from honest products

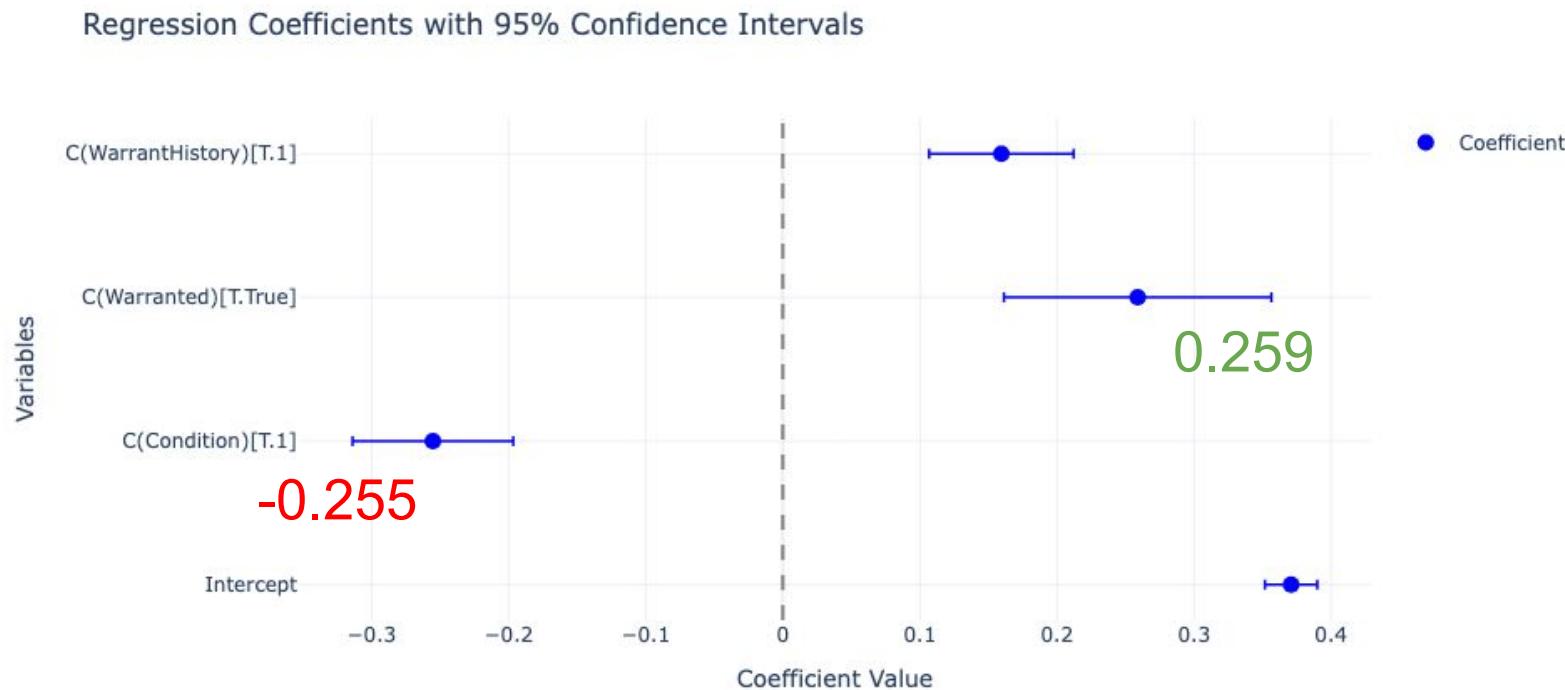
Honest and Cheating Utility by Market Type



Honest Bots Sell, Dishonest Fail in Warrants Market

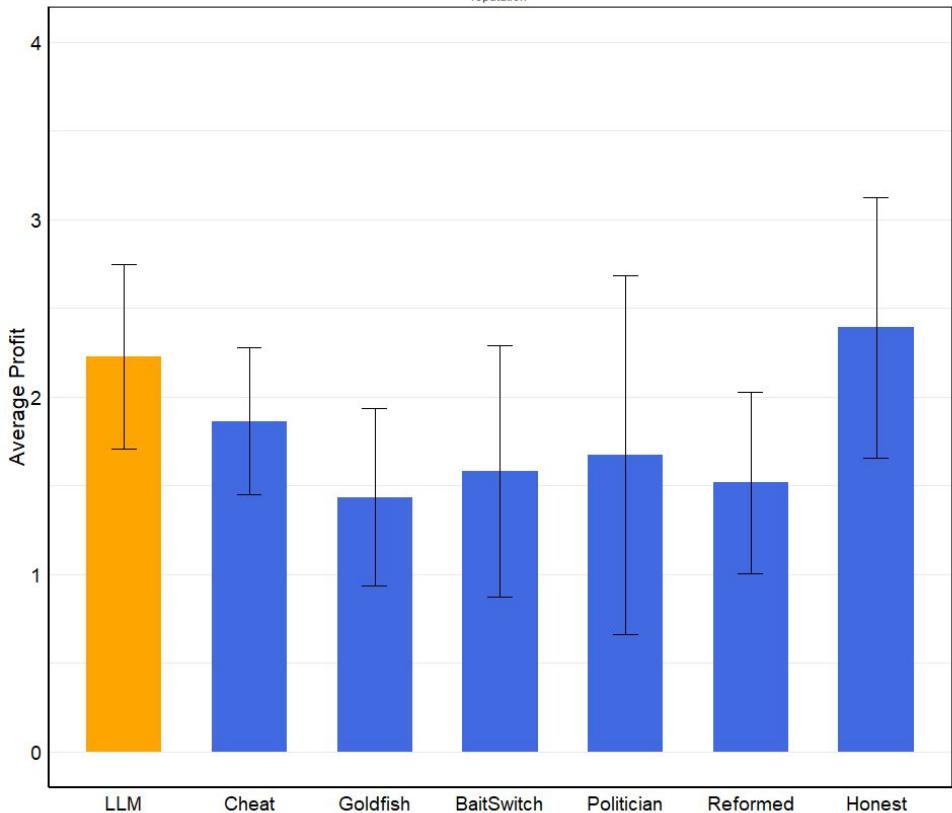


Self-certification and ‘Histories’ Significantly Increase Sold Products

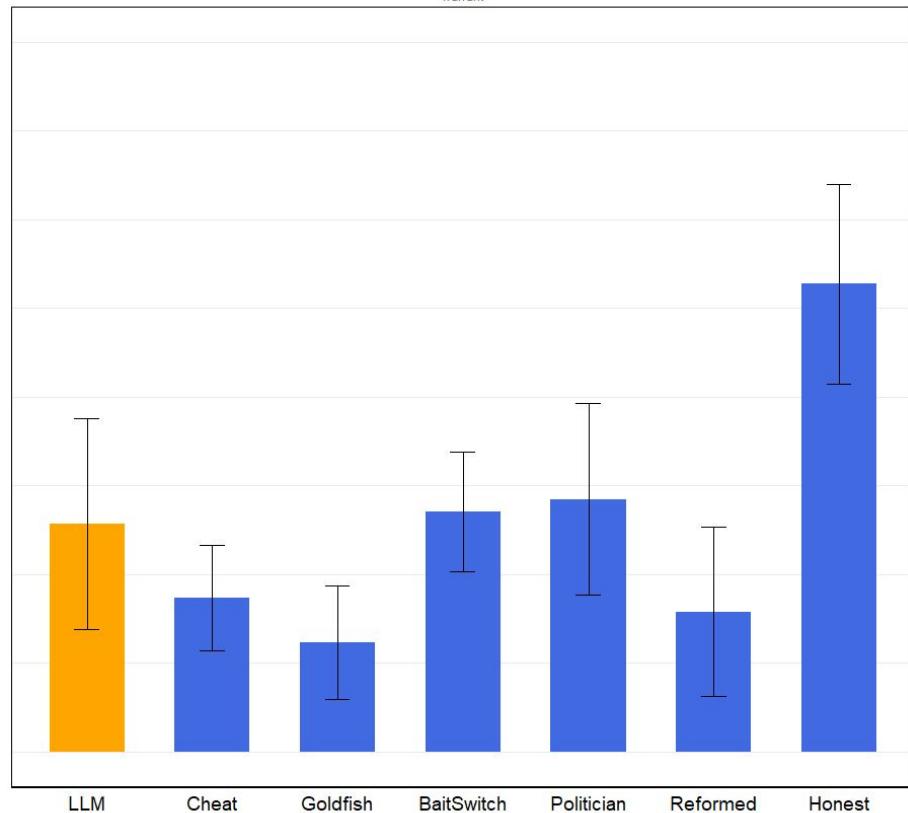


Yes, LLM agents can also sell in this marketplace!

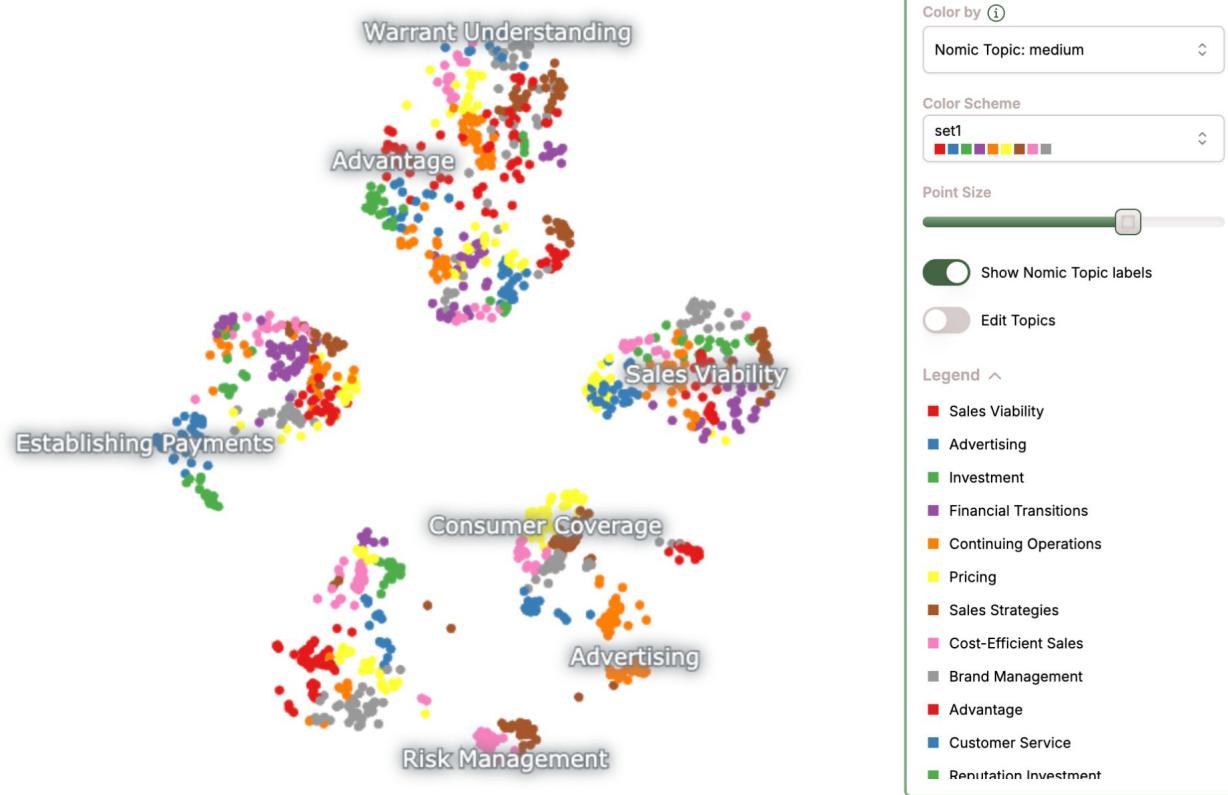
Reputation



Warrant



And their explanations actually make sense!



Early IS Job Market Pitch – Feedback welcome!!



28 y/o

4 year Ph.D.
1 year Postdoc
8x tech intern
1x Nonprofit
1x Startup

Digital Trust,
Responsible AI,
Behavioral
Economics

Papers (in prep)

PNAS x 2, PNAS Nexus,
ISR, Mgmt. Sci.
CHI, ICWSM x 2

Papers (Workshop Oral)

3 x ICWSM, AAAI, 2 x ICML,
2 x NeurIPS, 4 x Stanford
T&S Conf., IC2S2

Yale AI/ML, Columbia
MarkTech

Fellowships / Awards

Google Research Innovator, Belfer Fellow,
FTI People to Watch, Responsible Tech
Affiliate, Atlantic Dialogues Emerging Leader

Independent Grants: \$275,000

What makes me unique

Presented to PMs, Presidents, UN, World
Bank; ongoing work in 6+ countries
Mentored 130+ students @ MIT, NYU, BU,
India; awards 2xGoogle, Mozilla, Wiki Fdn
Collected 0.5B+ Social Media Posts

Come meet me and see a live demo!

swapneel@bu.edu; swapneel@mit.edu

mehtaver.se

0 → 15 undergrads + 5 grads

Undergraduate Students



Quang Nguyen



Laerk Ndrek



Oscar Fang



Abhishek Shah



Jiazheng Su



Vedant Kejariwal



Hang Ji



Jenny Li



Sverre Wiedswang



Tejovan Parker



Zou Yang



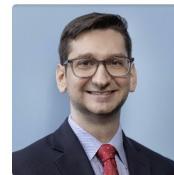
Seamus Dallman



Changxuan Fan



Daian Liu



Aaron Nichols



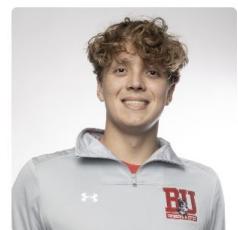
Pratyay Banerjee



Isa Alsafwah



Lauren Ji



Tony Ponomarev

Founded Nonprofit working in 6 countries with 15+ partners

Technology

Information systems research to build products for newsrooms, nonprofits, fact-checkers, and inter-governmental agencies.

Policy

Publish research and engage with policymakers and development agencies in advisory capacity to solve global governance issues.

Education

Deliver internationally acclaimed responsible computing programs to advance AI education for youth in the global majority.

Supporting UN Sustainable Development Goals



4
QUALITY
EDUCATION



5
GENDER
EQUALITY



9
INDUSTRY,
INNOVATION
AND
INFRASTRUCTURE



17
PARTNERSHIPS
FOR THE GOALS

Producer : **Jessie**

8 warrants, 0 challenges left

B 4 L 0 Y

 ADD TO CART



High Quality Cutlery

\$10

Producer : **Jessie**

8 warrants, 0 challenges left

B 4 L 0 Y

 ADD TO CART

Producer : **Mark**

0 warrants, 0 challenges left

B 4 L 0 Y

 REMOVE FROM CART



High Quality Fireworks

\$12

Producer : **Mark**

0 warrants, 0 challenges left

B 4 L 0 Y

 REMOVE FROM CART

Producer : **David**

8 warrants, 0 challenges left

B 4 L 0 Y

 ADD TO CART



High Quality Fireworks

\$10

Producer : **David**

8 warrants, 0 challenges left

B 4 L 0 Y

 ADD TO CART