

Basileia Imana¹, Aleksandra Korolova² and John Heidemann¹

¹ USC/Information Sciences Institute

² Princeton University

Motivation

- Increasing desire for scrutiny of social-media algorithms
 - Digital Services Act (EU) and PATA (USA)
- Concerns regarding algorithmic personalization
 - Bias/discrimination, political polarization, ...
- Tension in value for different stakeholders
 - Users: personalized content => positive experience
 - Businesses: increased engagement => maximize profit
 - Society: fairness => not promote bias/polarization

1) Need for a New Type of Auditing

Policy pushes for access to researchers

- DSA and PATA

Give **actionable** implementation proposal

Conflicting societal vs. business interests

- Provide access vs. protect user/platform privacy

Show access to **relevance estimators** provides a balance

Black-box auditing methods are limited

- Proxies, confounding factors, high cost, ...

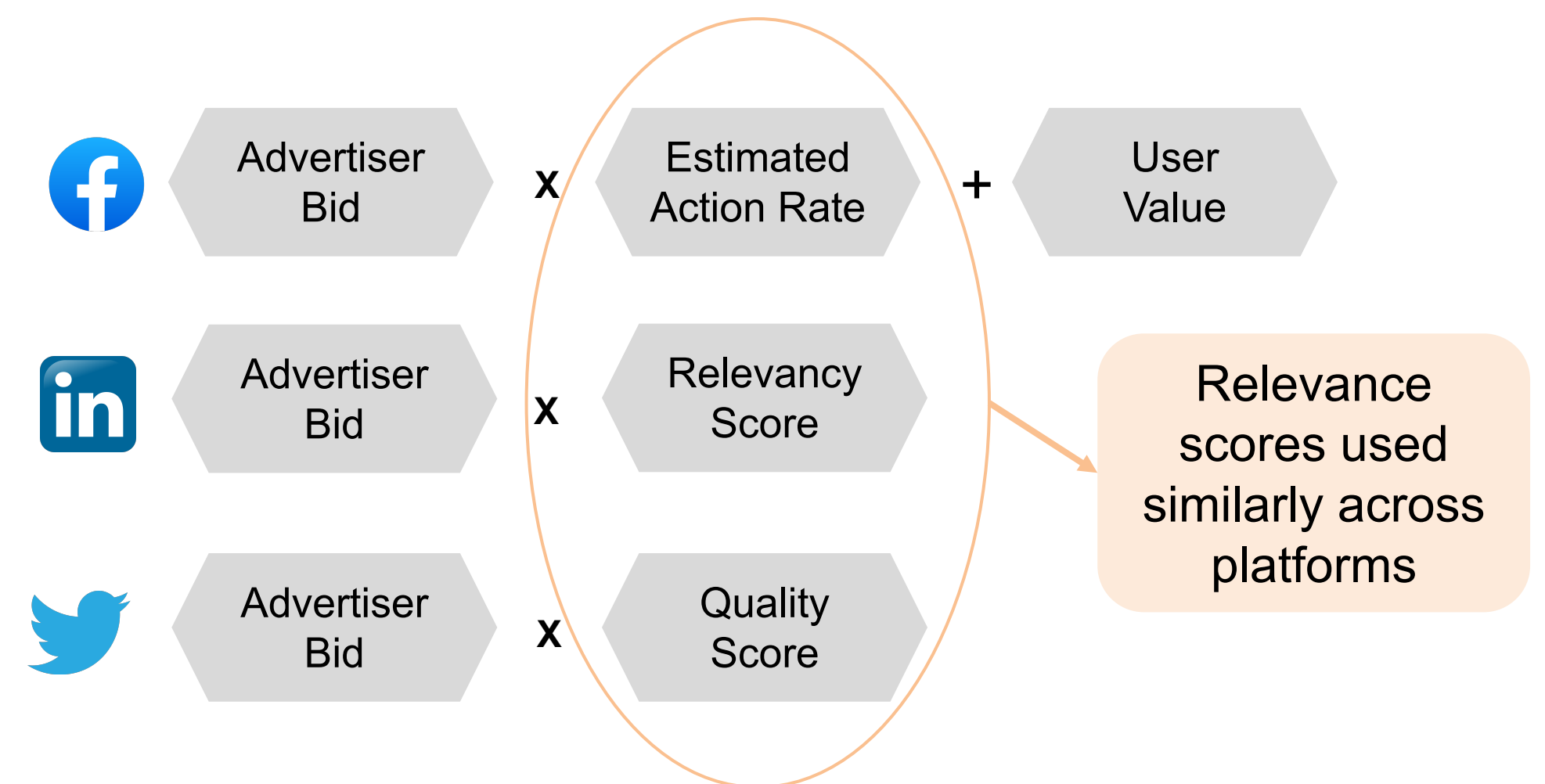
Show **explicit platform support** addresses the limitations

Our Contributions

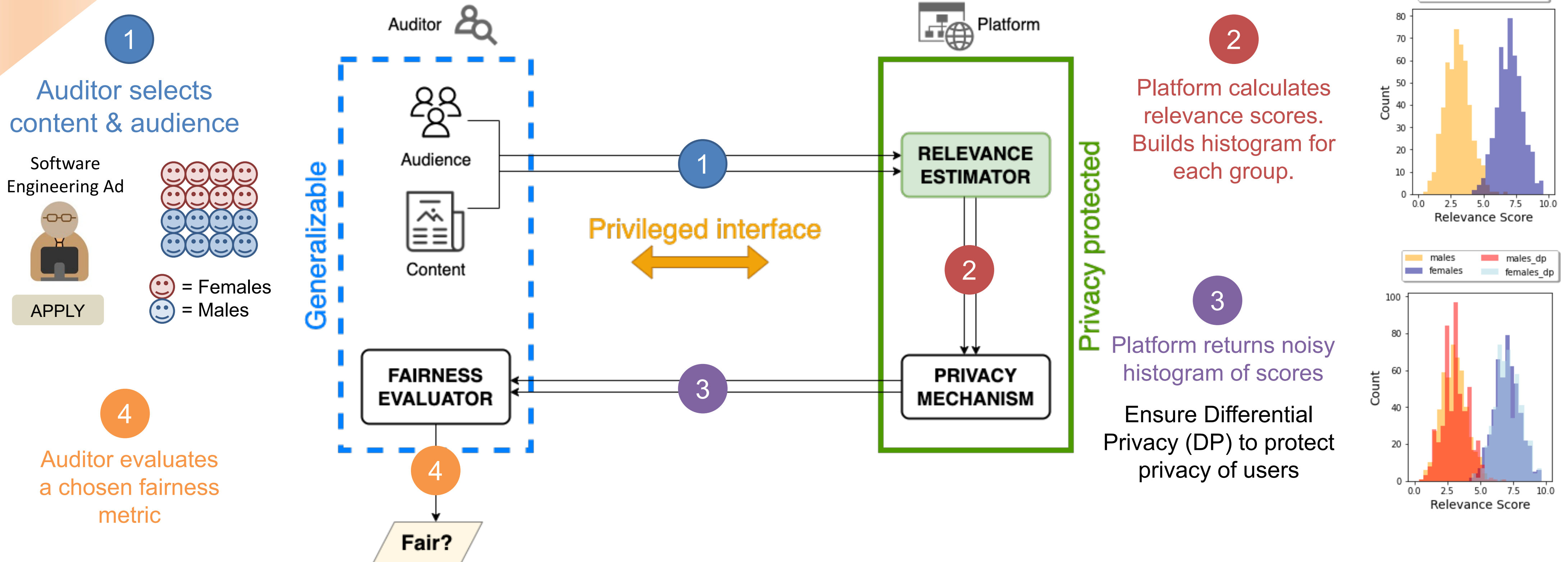
- Discuss the needs for a new type of algorithmic auditing
- Suggest query access to **relevance estimators** is the key for increasing transparency
- Show such access need not risk user privacy nor business interests
- Quantify number of samples needed for auditing with privacy

2) Relevance Estimators: the "Brains"

- We propose **relevance estimators** are the **key to auditing**
- They determine how all content is delivered on platforms
 - Rank **organic content** on a personalized newsfeed
 - Modifier to bids to determine winners of **ad auction**



3) Platform-supported and Privacy-preserving Auditing Framework



4) Sample Size Required for Auditing

- What is the cost of privacy?
- Can we achieve the same statistical confidence using DP?

Adding DP imposes a small constant factor (**4x**) increase to the number of samples

- A lower-bound for number of qualified users needed from each group:

$$\min_{a \in A} n_{a,q} \geq \frac{8}{\alpha^2} \ln \frac{3|A||Y|}{\delta}$$

Level of fairness: α
 Level of privacy: ϵ
 Range of relevance scores: $|Y|$
 Number of demographic groups: $|A|$
 Probability of failure: δ

Conclusion

- A path exists from proposed legislation to realizable auditing system
- Platforms can enable platform-supported auditing without compromising privacy
- Next steps:
 - Collaborate with an ad platform to implement the framework
- Paper pre-print: <https://arxiv.org/abs/2207.08773>
 - To appear in CSCW '23

This work was funded in part by a grant from The Rose Foundation and NSF awards 1916153, 1956435, and 1943584.

