Quantitative Auditing of AI Fairness with Differentially Private Synthetic Data

袁至誠
Chih-cheng Rex Yuan
rexyuan.com

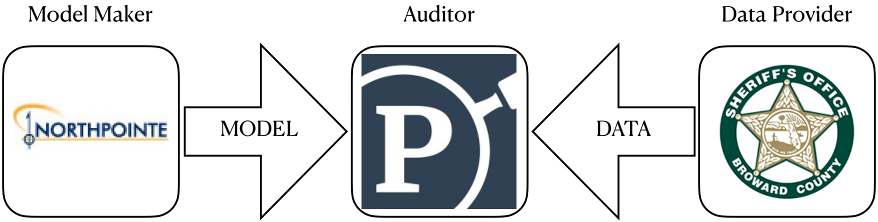
Institute of Information Science, Academia Sinica

Tuesday 3rd June, 2025

Why Fairness Audits Matter

- AI systems are influencing justice, health, finance.
- Bias in AI means real-world discrimination.
- Example: COMPAS audit by ProPublica.

Auditing Framework Overview



Model Maker

Auditor

Data Provider

MODEL

DATA

The Privacy Problem

- Auditors need sensitive data to test fairness.
- But holding this data introduces security and privacy risks.
- Security risk: hackers could steal the data.
- Privacy risk: published stats could expose insights.

Solution: Privacy-Preserving Audits

- Use synthetic data generated from real data.
- Apply differentially private synthetic data to ensure individual info stays hidden.
- Auditors only keep the synthetic data and not the real data.

What Is Differentially Private Synthetic Data?

- Fake data that mimic the real data's patterns.
- Generated with differential privacy, which adds noise to protect privacy.
- Lets auditors analyze fairness without exposing personal data.

How The Synthetic Data Are Generated

- We use a proven method that won a U.S. government competition (NIST, 2018).
- It adds noise to protect privacy while preserving overall patterns.
- The result: data that looks real but contains no real individuals.

Does It Work?

- Compared fairness metrics on real and synthetic data.
- Datasets: Adult, COMPAS, Diabetes.
- Most metrics are within negligible difference.

Policy Implications

- Enables safer third-party audits under privacy guarantees.
- Avoids liability for storing sensitive datasets.

Conclusion

- Synthetic data can support fairness audits with privacy.
- Our framework is practical and provably private.
- Opens new pathways for legal oversight of AI systems.

Slides: `https://github.com/RexYuan/Eunectes`.