Algorithm Audits

CS 347
Danaë Metaxa & Parastoo Abtahi
Announcements

Guest lecture on Monday: John Tang on Accessibility
Today: algorithm audits

What is algorithm auditing and when is it useful?

History: How did it originate?

Basics of algorithm audits

Frontiers: Algorithm audits

Legal and ethical aspects of auditing
The problem

Google Ads were 25% more likely to suggest an arrest record for Black-sounding names than white ones (Sweeney, 2013)
Algorithm audits repeatedly query an algorithm with controlled inputs, analyze outputs to draw inferences.
History of Audits
Audits (without the “algo”)

Auditing is widespread, and used in a wide range of contexts

- Accounting (e.g. tax audits)
- Government audits (e.g. U.S. GAO)
- Quality audits (e.g. ISO 9000)
- Discrimination (e.g. redlining)
Milestones in auditing

1960s: UK Parliament mandated oversight for anti-discrimination legislation (field audits; Daniel, 1968)
Milestones in auditing

1960s: UK Parliament mandated oversight for anti-discrimination legislation (field audits; Daniel, 1968)

1980s: Correspondence audits (e.g. Bertrand and Mullainathan, 2003)
Milestones in auditing

1960s: UK Parliament mandated oversight for anti-discrimination legislation (field audits; Daniel, 1968)

1980s: Correspondence audits (e.g. Bertrand and Mullainathan, 2003)

2000s: Automated correspondence audits (e.g. Oreopoulos, 2011)
Auditing: Now with more technology!

From auditing to algorithm auditing
Auditing digital discrimination

For algorithmic systems, functional platform is not enough

Algorithm audit introduced in this paper: "look inside the black box of the algorithm to pursue knowledge about pressing public problems"

Types of audits: code; noninvasive user audit; scraping; sock puppet; crowdsourced/collaborative

“Regulation towards auditability”
Three components

The overall design of an algorithm audit has three main components:

1. Attribute—along what axis might the disparity exist?
2. Topic—what is the overall category or theme?
3. Platform—which algorithm or platform are we studying?
What should we audit?

What kinds of biases should we be auditing for in algorithms?

- Race
- Sex
- Nationality
- Religion
- Disability

- Gender
- Sexuality
- Political views
- SES
- Education

- Employment
- Health
- Age
- Pregnancy/parent status

How should we select this attribute?

- Legal ramifications
- Personal interest
- Ethical importance
Common topics

Social science audits often focused on housing and employment (areas with legal protection and concrete implications for finding inequality).

Common topics for algorithm audits have included

- Housing
- Employment
- Gig economy
- Heathcare
- Consumer markets
Common platforms

The platform audited may very independently of the domain; if the domain is the type of bias auditors are looking for, the platform is the place where they’re looking.

Common platforms include…

• Social media sites (FB, Twitter, etc.)
• Search engines (Google, Google Images, Bing, etc.)
• Commercial systems (health records, legal systems, etc.)
• Other online platforms (hiring sites, ads, etc.)
Other dimensions to consider

Once you’ve selected an axis of difference, topic, and algorithm, there are several other dimensions to keep in mind:

1. Temporal considerations (longitudinal? two points? single point?)
2. Data collection (manual, API, scraping, etc.) and annotation (Manual, crowdsourcing, ML, etc.)
3. Data analysis (paired t-test, fit model to ground truth, etc.)
4. Communicating findings (academic pub, general audience, direct)
5. Legal and ethical aspects (more later in a few slides!)
Recent frontiers
Criminal risk scores

[Angwin & Larson for ProPublica, 2016]

Analyzed data from a bail/sentencing algorithm to argue that risk scores were unequally assigned to defendants of different races

Identified disparate impact; some debate about whether that’s the right metric, what the root problem is, etc.

Overall very high-profile example
Digital redlining

[Edelman and Luca, 2014]

Redlining: discriminatory practice of avoiding investment in communities seen as undesirable (e.g. banks not lending equally to POC)

Digital redlining: creating/perpetuating inequity through digital tech

Method: using host photos and rental quality/price info, compare rentals by host race
Non-Black hosts get 12% more for same quality listing
Jobs & Ads

[Chen et al., CHI 2018]

Collected resumes from hiring websites (recruiter perspective) and coded names for gender

Finding: slight penalty for female names (controlling for all else)

[Speicher et al., PMLR 2018]

FB bans targeting by race/gender when advertising housing, jobs, etc

They show advertisers can easily still do discriminatory targeting via user attributes, free-form inputs, PII-based custom ads, and look-alike audiences
Healthcare

[Obermayer et al., Science 2019]

US health system uses tech to guide decisions, including a widely-used algo that assigns risk scores to patients

Using heath data, find that Black patients are sicker (more chronic illness) than white patients assigned the same score

Why? Costs as proxy for needs!

Fig. S1. Number of chronic illnesses vs. algorithm-predicted risk, including non-Black, non-White patients. Mean number of chronic conditions by race conditional on algorithm risk score.
Bias in image search results

[Metaxa et al., CSCW 2021]

Do image search results accurately reflect real-world gender and racial diversity? How do biased search results affect users?
Bias in image search results

[Metaxa et al., CSCW 2021]

1. Scrape Image Search results for 100 occupations
2. Crowdsourced race and gender annotations
3. Finalize and validate list of occupations

? % marginalized
Race

2020 Image Search proportion POC

2020 BLS workforce proportion POC
If an occupation is 50% women, images predicted to show 42% women

If an occupation is 22% POC, images predicted to show 16% POC
Ethics and law
Who should conduct audits?

Algorithm auditing is high stakes and increasingly widespread, but no clear model has emerged yet. Who should audit?

- Government agencies?
- Private consultancies?
- Academic researchers?
- Journalists?
- Internal teams?

Gold standard: independent, professional third parties
Law and ethics

Since audits focus on human-facing systems with real-world implications, there are often connections to law and policy, as well as legal and ethical risks to conducting an audit.

• Potentially illegal actions: Terms of Service or CFAA violations (for things like scraping content or using fake accounts)

• Impacts on platform services: Accidental DDoS, showing the service down for legitimate users

These decisions are subjective, sometimes normative, never easy!
What about…

Personalization: what if we expected results to vary a lot by user?

Open access: Should researchers conducting audits make code or data publicly available?

Activism: “efforts to promote, impede, direct, or intervene in social, political, or environmental reform.” Are audits activist?

When *shouldn’t* we audit?
When *shouldn’t* we audit?

HireVue: uses video of people to screen them for hiring

Hired ORCAA (Cathy MacNeil’s audit consultancy) & claimed they found no bias

Anyone downloading the report cannot share it in any way (stymies journalists)

Should they have audited?
P.S. Want to learn more?

If you find this kind of work interesting and exciting, you might consider doing a PhD with a focus in an area like algorithm auditing.

“But Danaë, how could I do that?”

I’m so glad you asked! There are great researchers doing this kind of work at Northeastern (Christo Wilson), CMU (Motahhare Eslami), UIUC (Karrie Karahalios), and starting next year, Penn (me!)