

Auditing Algorithms

Casualty Actuarial Society

November 2022

WMDs

**Three
Types of
Audits**

**The
Ethical
Matrix**

**Case Study:
Credit**

**Explainable
Fairness**



WMDs

Algorithm

Open Q's

- Widespread & Important
- Mysterious & Opaque
- Destructive

Teachers

Facial
Recognition

Hiring







August 2018 Accuracy on Facial Analysis Pilot Parliaments Benchmark

98.7%

68.6%

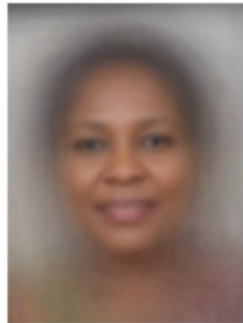
100%

92.9%

amazon



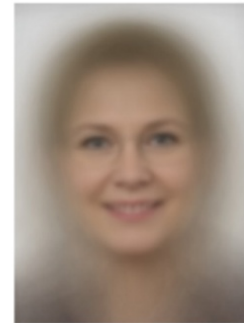
**DARKER
MALES**



**DARKER
FEMALES**



**LIGHTER
MALES**



**LIGHTER
FEMALES**

Amazon Rekognition Performance on Gender Classification

Open Questions

- What exact context?
- How to define fair?
- How to set thresholds?

Auditing Algorithms

Casualty Actuarial Society

November 2022

WMDs

**Three
Types of
Audits**

**The
Ethical
Matrix**

**Case Study:
Credit**

**Explainable
Fairness**



Three Types of Audits

- Adversarial audits (via agencies or regulators)
- Invitational audits
- Third party audits (via laws or regulators)

Auditing Algorithms

Casualty Actuarial Society

November 2022

WMDs

**Three
Types of
Audits**

**The
Ethical
Matrix**

**Case Study:
Credit**

**Explainable
Fairness**



The Ethical Matrix

- For whom is this working?
- What kind of failure?
- Non-technical, values-based
- An artifact of a conversation
- Not a solution!

The matrix

- Rows are stakeholders
- Columns are concerns
- Stakeholders should be represented
- We consider each cell
- We rank concerns
- This is practical ethics

Auditing Algorithms

Casualty Actuarial Society

November 2022

WMDs

**Three
Types of
Audits**

**The
Ethical
Matrix**

**Case Study:
Credit**

**Explainable
Fairness**



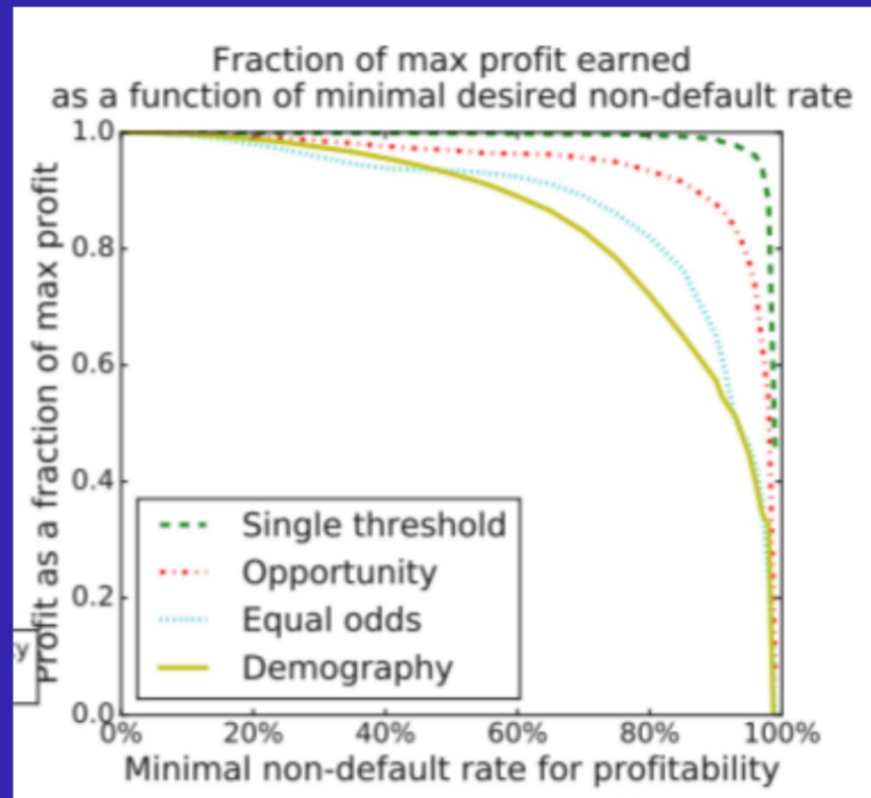
Credit

Hardt et al:
Equality of Opportunity in Supervised Learning

**Fairness
costs \$**

**Ethical
Matrix**

Fairness Costs Money



Ethical Matrix

	Profit	False +'s	False -'s	Data Quality
Company	Yellow	Yellow	Green	Green
White Customers	Green	Green	Yellow	Yellow
Black Customers	Green	Green	Red	Red

Auditing Algorithms

Casualty Actuarial Society

November 2022

WMDs

**Three
Types of
Audits**

**The
Ethical
Matrix**

**Case Study:
Credit**

**Explainable
Fairness**



Explainable Fairness

Currently

**Ex: student
lending**

- From a regulator's viewpoint
- Legal categories of stakeholders and concerns
- "Discrimination" needs translation with nuance

**A Balancing
Act**

**Ex: disability
insurance**

What's Happening Now



Government is starting to demand “unbiased” AI.

- New laws: NYC, DC, CO
- Regulated industries first
- Key definitions left to rulemaking



Companies are improvising since they don't know how to prepare.

- Voluntary “exhibition” audits
- No-action letters
- One-off approaches e.g., [Twitter](#), [Airbnb](#)

Balancing Framework

1. Regulator: There's a difference in outcome
2. Industry: Oh but there's such a good reason for it, i.e. a benefit
3. Regulator: Even accounting for those beneficial factors, we still see a difference, we need to balance.

Explainable Fairness helps create a positive feedback loop

If a regulator does the same balancing test for the entire industry, they get to set and, over time, improve standards.

Use Case: Student Lending

Step 1: Outcomes of interest include loan offer, APR, term, consequences of late payment or default. Choose one.

Step 2: Infer protected class status and categorization buckets.

Step 3: Measure outcomes for various classes.

Step 4: If we see “significant”* differences, consider “accounting factors”** such as: FICO, salary, type of major, rank of college.

* what does significant mean here? ** when is a factor considered legitimate?

Use Case: Disability Insurance

Step 1: Outcomes of interest include: claim approval, length of initial claim, number of total weeks covered, number of total extensions.

Step 2: Infer protected class status and categorization buckets.

Step 3: Measure outcomes for various statuses.

Step 4: If we see “significant”* differences, consider “accounting factors”** such as: type of injury, maternity claim, age, comorbidity, type of work.

*what does significant mean? ** when is it considered legitimate?

Auditing Algorithms

Casualty Actuarial Society

November 2022

WMDs

**Three
Types of
Audits**

**The
Ethical
Matrix**

**Case Study:
Credit**

**Explainable
Fairness**

