

CIVIL RIGHTS IN A DIGITAL SOCIETY: STATE INNOVATIONS TO ADDRESS AI BIAS

YOLANDA N. MELVILLE

DIRECTOR, NEW JERSEY DIVISION ON CIVIL RIGHTS

ONESHIA HERRING

**SENIOR VICE PRESIDENT & COUNSEL, NATIONAL FAIR HOUSING
ALLIANCE**



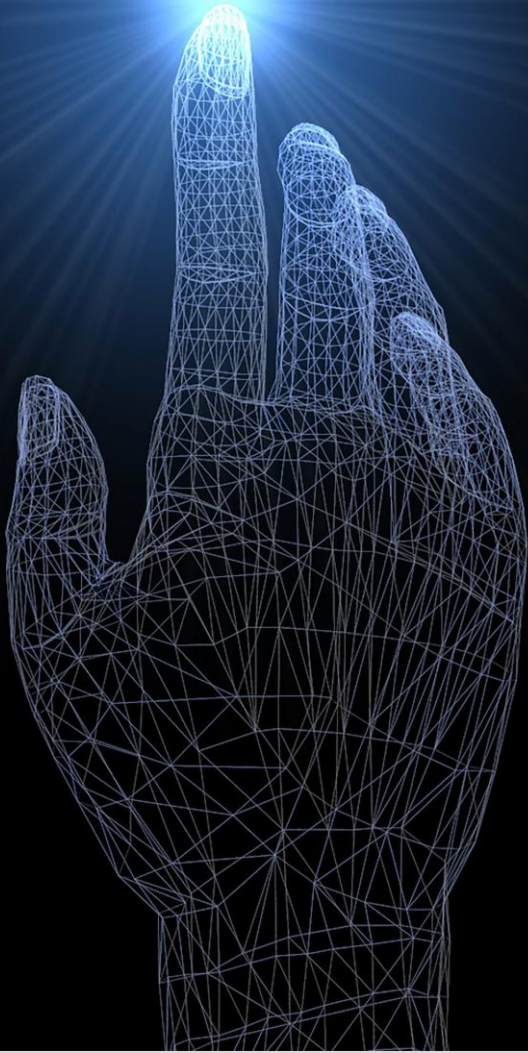


All examples discussed in this session are provided for informational purposes only. The NJ Division on Civil Rights (DCR) is not expressing any legal opinions on any of the cases presented, nor are we endorsing any specific tools or technologies. Because claims of discrimination under the NJ Law Against Discrimination (LAD) require a fact-intensive inquiry based on the specific facts at issue, DCR is not pre-judging the lawfulness of any of the conduct described in the examples and is not commenting on whether the companies' actions would violate the LAD. Rather, we are presenting these for informational purposes and in order to educate the public on different legal concepts.



Agenda

1. Welcome and Introductions
2. What is Algorithmic Discrimination?
3. Why is Algorithmic Discrimination a Civil Rights Issue?
4. The Growing Role of State and Local Civil Rights Agencies
5. New Jersey's AI Guidance and the Evolving Landscape of State-Level Responsible AI Governance
6. From AI Awareness to Enforcement Capacity: NFHA's Responsible AI WorkStrategies for Responsible Use of AI
7. Practical Steps Civil Rights Agencies Can Take Now
8. Q&A



WHAT IS ALGORITHMIC DISCRIMINATION?

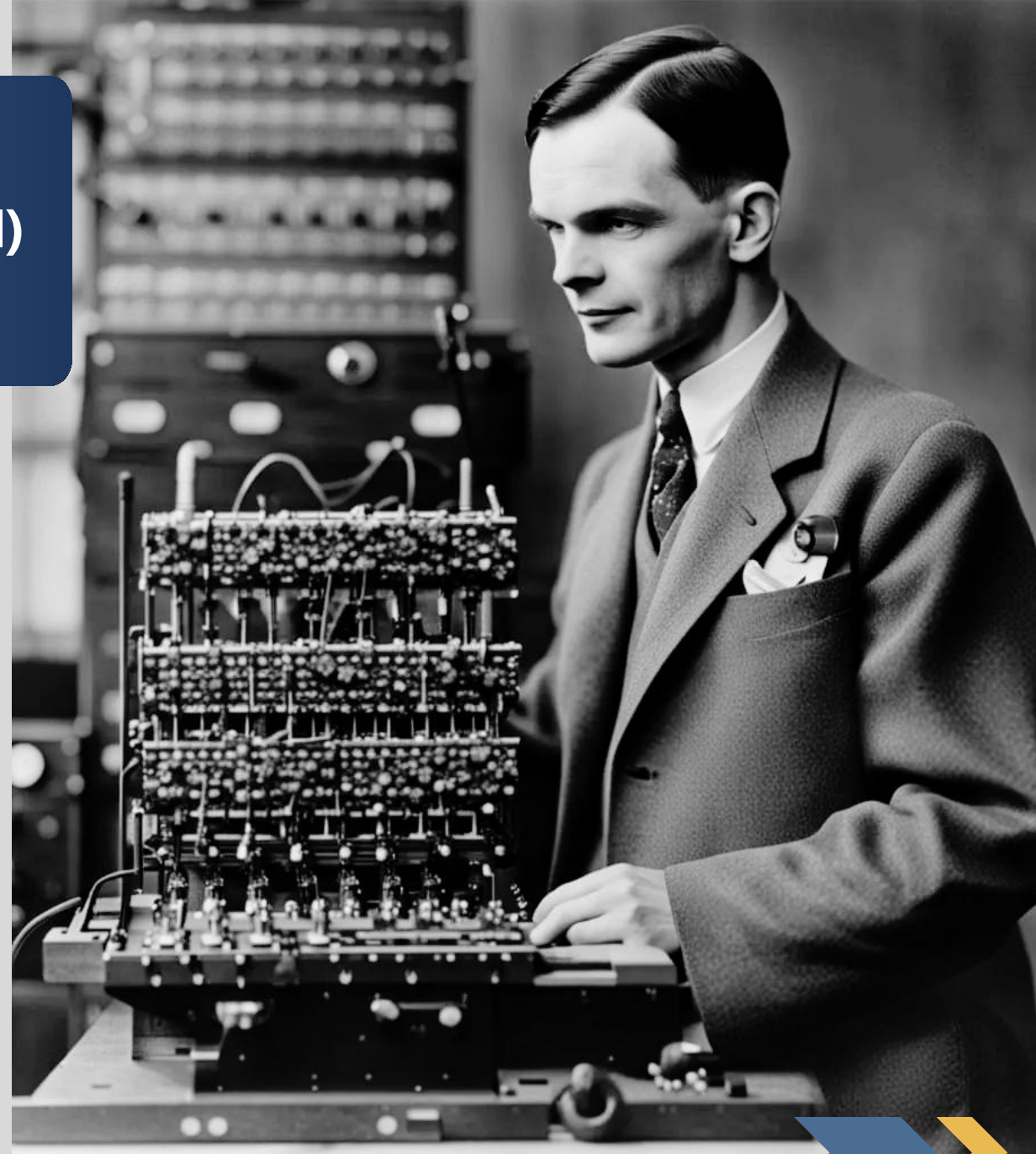




Artificial Intelligence (AI)

AI refers to the ability of machines to mimic human-like intelligence

Generative AI: a type of artificial intelligence that can learn from and mimic large amounts of data to create content such as text, images, music, videos, code, and more, based on inputs or prompts



Alan Turing, the father of AI



AI and its Capabilities, as Conveyed by ChatGPT



Natural Language Processing

Understand and generate human language

- ChatGPT, Siri or Alexa virtual assistants, autocorrect, translation services



Image and Speech Recognition

Identify and interpret images and words

- Google image search and facial recognition software



Data Analysis

Analyze vast amounts of data, find patterns for decision-making and predictions

- Self-driving vehicles



Automation

Automate repetitive tasks, improve efficiency

- Industrial robots used in manufacturing

Machine Learning

Ability to improve performance over time by learning from data

- Email system learning to classify spam

Recommendation Systems

Personalized recommendations in streaming platforms and online shopping

- Netflix recommendations "because you watched," social media feed



Automated Decision-Making Tools

Any technological tool, including but not limited to a software, system, or process that is used to automate all or part of the human decision-making process.





Algorithmic Discrimination

Discrimination that results from the use of an automated decision-making tool



DESIGN, TRAINING, AND DEPLOYMENT



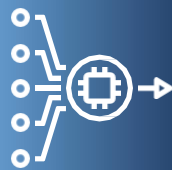
Bias may be coded into AI models in these 3 phases



Design



Developers translate problem into a set of instructions or a mathematical model



Bias can be introduced into the inputs, outputs, or rules used by the algorithm to make decisions



Recent case: iTutorGroup used automated tool to make hiring decisions that considered the birth year of candidates, resulting in age-based discrimination





Training



Automated tools are “trained” through exposure to datasets before use in real world



The tool can become biased if the training data is not diverse, reflects historical bias, is artificially generated, or contains errors



Video, Reuters



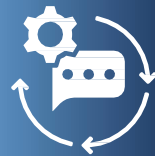
Amazon Scraps AI
Recruiting Tool Showing
Bias Against Women



Deployment



Automated tools can amplify systemic inequities or contribute to a feedback loop



PredPol, no longer in use: predictive policing algorithm targeted locations overrepresented in police data due to racial biases in policing



Created feedback loop when more arrests led to more data fed back into model, predicting more crimes in those locations



WHY IS ALGORITHMIC DISCRIMINATION A CIVIL RIGHTS ISSUE?





Employment

Job advertising; resume review; hiring decisions; performance reviews; promotion, demotion, and firing decisions



Housing

Advertising for housing; tenant screening; mortgage approval and terms; decisions on houses clients will see or tour



Schools

Monitoring students' computer use; filtering content; aiding in disciplinary decisions; predicting whether students graduate or face criminal legal system



Healthcare

Diagnostics; determining quality and quantity of care; determining health risks; allocating resources; prescribing medication



Law Enforcement

Identifying suspects; reading license plates; selecting areas for patrol



Courts

Predicting whether someone will commit another crime; setting bail; deciding eligibility for pre-trial release

Video, WSJ

Artificial Intelligence:
The Robots are Now Hiring



What commonalities or patterns do you see with how automated decision-making tools are used across these settings?

Who might be negatively affected?



Video, CNBC

Rite Aid to be Barred from Using Facial Recognition under Proposed FTC Settlement



What commonalities or patterns do you see with how automated decision-making tools are used across these settings?

Who might be negatively affected?





Algorithmic Discrimination Discussion

What commonalities or patterns do you see with how automated decision-making tools are used across these settings?

Who might be negatively affected?



NEW JERSEY'S AI GUIDANCE AND THE EVOLVING LANDSCAPE OF STATE-LEVEL RESPONSIBLE AI GOVERNANCE





3 Types of Discrimination Claims

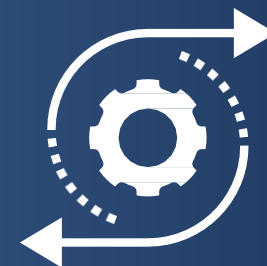
1. Failure to accommodate
2. Disparate treatment discrimination
3. Disparate impact discrimination

The NJ Law Against Discrimination (LAD) prohibits all forms of discrimination, whether the discriminatory conduct was the result of a decision made by an automated decision-making tool or by a human.



Reasonable Accommodations

Changes to policies, procedures, or physical structures that ensure accessibility for people based on disability, religion, pregnancy, or breastfeeding





Accessibility

An accommodation may be necessary if an automated tool is inaccessible

For example, if an automated tool used to measure job applicants' typing speed is unable to do so on a non-traditional keyboard



Training

Automated tools may not be trained on data that includes individuals who use accommodations

For example, an automated tool rejects a job applicant who is blind and requires screen reader software to perform their job



Inclusion

Some automated tools may fail to account for accommodations

For example, if a tool monitors employee productivity and penalizes employees who take additional break time as an accommodation



Disparate Treatment Discrimination

Occurs when a person is treated differently, intentionally or unintentionally, because of their membership in a protected class



Example: AI Detection Software

Tools falsely labeled English Language Learners' writing as being AI generated 61% of the time

Tools were "near perfect" in evaluating native English speakers' essays



Automated Tools and Disparate Treatment

If a school relied on tool and unfairly penalized English Language Learners for using AI, that could be disparate treatment discrimination based on nationality, national origin, or ancestry



Disparate Impact Discrimination

Occurs when policies or practices seem neutral on their face but have a disproportionately negative effect on members of a protected class



Automated Tools and Disparate Impact

Can occur if a tool makes recommendations or decisions that disproportionately harm members of a protected class, unless its use serves a substantial, legitimate, nondiscriminatory interest



Case Example: SafeRent



Class-action discrimination lawsuit against SafeRent, a third party tenant screening vendor

Case alleged that SafeRent's scoring algorithm used credit history to assign disproportionately lower scores to Black and Hispanic applicants

Algorithm allegedly didn't consider housing vouchers

Housing providers relied on these scores for housing decisions, rejecting Black and Hispanic renters and those using housing vouchers at higher rates

Court found that the plaintiffs adequately alleged that SafeRent was responsible for the discriminatory effect of their tools under the Fair Housing Act



Amazon Enforcement Action: Key Findings

Automated Systems Failed Workers



- Over 27,000 disability and pregnancy accommodation requests
- Amazon's automated processes did not properly account for these needs

Productivity Algorithms & Discipline



- Rigid productivity metrics enforced through automated discipline
- Workers with pending or approved accommodations were still penalized or terminated

Discriminatory Impact



- Accommodated workers held to the same standards as non-accommodated employees
- Practices disproportionately harmed pregnant workers and workers with disabilities, violating NJ civil rights laws

Mobley v. Workday, Inc. (2023)



Mobley alleges that Workday's software rejected his job applications because he is Black, over the age of 40, and has anxiety and depression

Mobley claims he was screened out due to name of college, graduation year, and personality tests

Workday attempted to have case dismissed, not an "employment agency"

A federal district judge said Workday can be considered an "agent" of the employer in a workplace discrimination case

Later, a federal court allowed the action to move forward as a nationwide class action based on age

“Workday's customers delegate traditional hiring functions, including rejecting applicants, to the algorithmic decision-making tools provided by Workday.

- California federal district judge



Who Should be Held Liable



Who do you think should be held liable in this case?

The employers who used Workday's tools to reject Mobley's job applications

Workday, the third-party vendor whose decision-making tools were used by the employers to reject Mobley's job applications

Both parties

Neither party

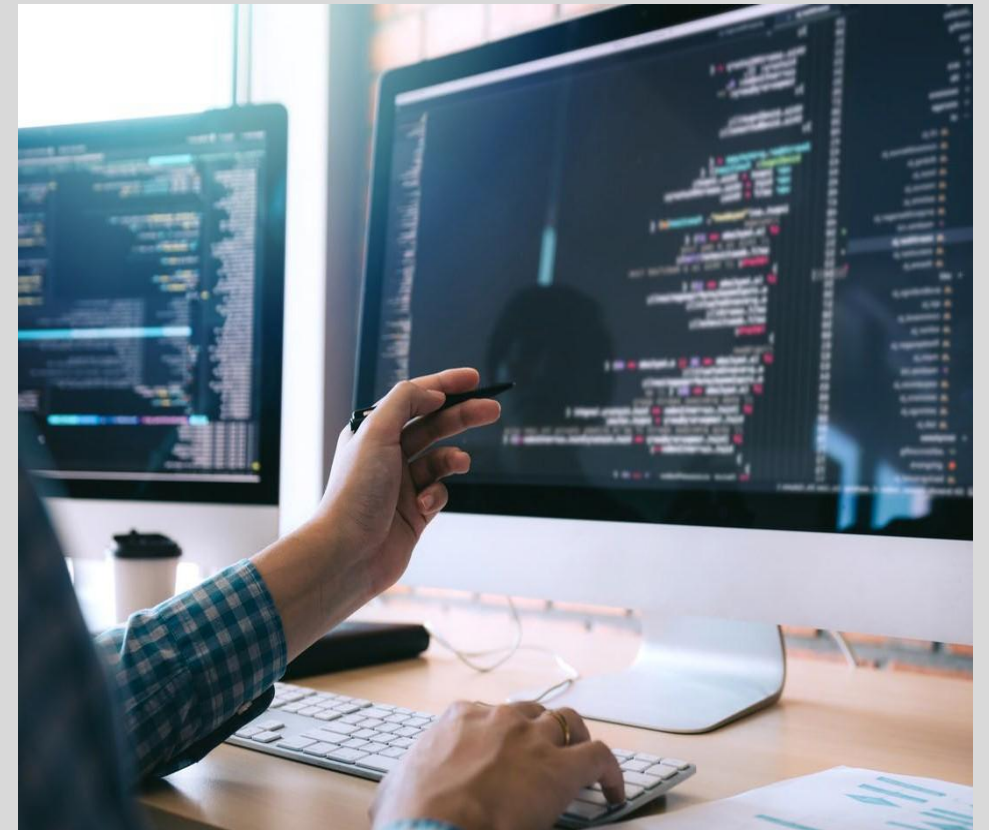




Liability in Algorithmic Discrimination

DCR's guidance on algorithmic discrimination specifies that workplaces, housing providers, or places of public accommodation are liable for algorithmic discrimination that results from their use automated tools

- Even when developed by a third party
- Even if they do not understand its inner workings



AI and Bias Prevention in Housing



Stanford University's AI tool combs through historic property deeds to find racially restrictive language - reviewed 24 million records in 6 days

other outbuildings incident to the residential use of the plot.

b.No structure shall be erected on any building plot, which plot has an area of less than 5000 square feet nor a width and a frontage of less than 50 feet.

c.No structure shall be erected on any building plot nearer than 20 feet to the front line, nor nearer than 5 feet to any side plot line. The side line restriction shall not apply to a garage located on the rear one-half of plot but no building shall be erected closer than 10 feet to a side street line.

d.No commercial or manufacturing enterprise shall be carried on upon any plot nor shall anything be done which may be or become an annoyance or nuisance to the neighborhood.

e.No persons not of the Caucasian Race shall be allowed to occupy, except as servants of residents, said real property or any part thereof.

f.No trailer, basement, tent, shack, garage, barn or other outbuildings shall at any time be used as a residence temporarily or permanently, nor shall any residence of a temporary character be permitted.

Portions of a racially restrictive 1940 real property deed from Santa Clara County includes the words: "No persons not of the Caucasian Race shall be allowed to occupy, except as servants of residents, said real property or any part thereof."



Algorithmic Discrimination in Housing: The FAIR Act

How New Jersey Is
Regulating Algorithmic
Rent-Setting



What the FAIR Act Does

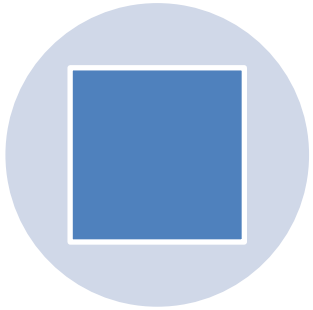
Bans algorithmic rent-setting software that coordinates pricing.

Prevents landlord collusion through shared data-driven platforms.

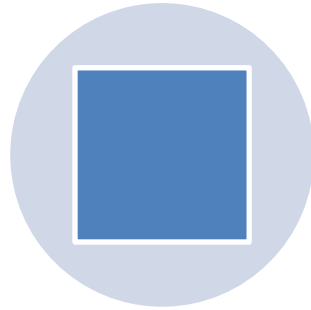
Protects renters from automated price inflation.

Creates clear reporting and enforcement mechanisms.

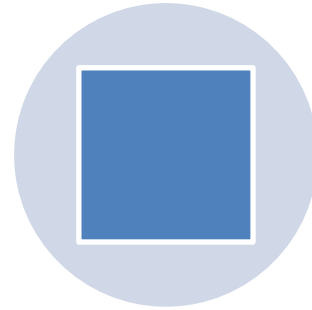
Impacts on Renters



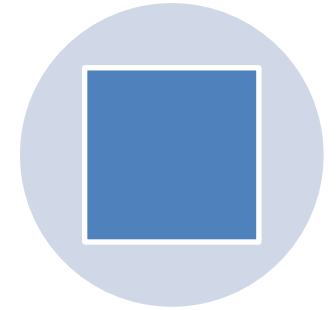
STOPS COORDINATED
ALGORITHMIC PRICING
AMONG LANDLORDS.



PROMOTES FAIRER AND
MORE COMPETITIVE RENTAL
MARKETS.



HELPS ENSURE RENTS
REFLECT ACTUAL MARKET
CONDITIONS.



AIMS TO REDUCE HARMFUL
PRICE HIKES AND STABILIZE
COMMUNITIES.

Voices Supporting the FAIR Act

Housing advocates highlighted rent-setting inequities.

Community organizations supported renter protections.

Economic justice groups praised anti-collusion measures.

Legislators emphasized affordability goals.

Broader Housing Strategy

Part of NJ's larger affordability agenda.

Increasing statewide housing production.

Supporting transit-oriented development.

Expanding affordable housing opportunities.

Housing Governing Council coordination.



FROM AI AWARENESS TO ENFORCEMENT CAPACITY: NFHA'S RESPONSIBLE AI WORK





NFHA Is Not New to Responsible AI Work

2017 — Began investigating discriminatory digital housing advertising

2018–2019 — Sued Facebook and secured major platform changes

2021 — Established NFHA's Responsible AI Lab

2022–Present — Developed assessment tools, conducted technical research, studied state legislation, and expanded agency support

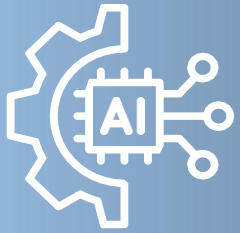


Civil-rights-centered technology work since 2017



Three Cases, Three Enforcement Lessons

Technology changes the method—not the duty to comply with civil-rights law.



Facebook

Protected groups could be excluded from housing ads.

Lesson: Discrimination can occur in targeting and delivery.



Redfin

Minimum-price policies limited services in communities of color.

Lesson: Neutral rules can reproduce redlining.



SafeRent

Screening scores allegedly disadvantaged Black and Hispanic applicants.

Lesson: Vendors and organizations using their systems may both face liability.



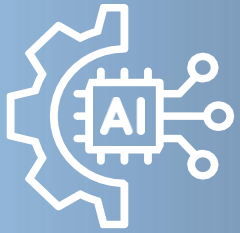


Five Questions for Investigating an Automated Decision

1. What happened?
2. Who built, bought, configured, and used the system?
3. What data, rules, and cutoffs shaped the result?
4. What did the system produce?
5. What evidence will explain the result?

Start with the decision—then trace the system behind it.

Before Your Agency Uses AI



Purpose

What problem are we trying to solve?

Should this task be automated?



Process

What tool and information will be used?

What safeguards and human review are required?

What must the vendor disclose?



Monitoring

Is the tool working as intended?
Are errors or harms affecting certain groups?
Can we correct or stop its use?

Never put confidential information into a public or unapproved AI tool.





What Agencies Can Do Now

1. Identify the automated tools affecting agency work and regulated decisions
2. Add technology questions to intake and investigations
3. Create procedures for complex cases and evidence preservation
4. Adopt rules for the agency's own use of AI
5. Train staff through realistic case scenarios

NFHA can help agencies assess their readiness and build these tools.

STRATEGIES FOR RESPONSIBLE USE OF AI





5 Strategies for Responsible AI Use

1. AI Governance
2. Risk Management Framework
3. Bias Auditing
4. Transparency
5. Humans-in-the-Loop



Strategy 1. AI Governance

Overarching principles or structures that:

- Manage risks
- Build trust in AI systems
- Align AI systems with federal and state regulations





Strategy 2. Risk Management Framework

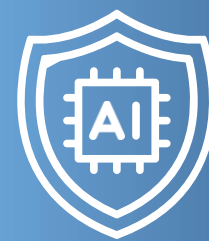
Define and manage risks and negative impacts of AI systems

Enhance positive contributions of AI



Risks for People, Company, and Environment

- Algorithmic bias and threats to civil liberties
- Environmental impacts
- Economic risks, including job loss
- Privacy and security



NIST Framework for AI Risk Management

Key characteristics of trustworthy AI systems:

- Valid and reliable
- Safe and secure
- Accountable and transparent
- Explainable
- Fair; bias mitigation

Strategy 3. Bias Audits



The practice of assessing, mitigating, and assuring an algorithm's safety, legality, and ethics, including:

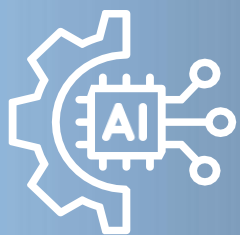
- Diversity of development team
- Datasets used in training
- Rules used to make decisions
- Measuring outputs for fairness and bias

Audit both pre-deployment and post-deployment

Best if performed by an independent party



Strategy 4. Transparency



Explainability of the Technical Components

Can the internal mechanics or rules of the algorithm be easily explained and understood?



Transparency of Use

Are those subjected to the algorithm aware of its use and do they have the opportunity to request reasonable accommodations?



Transparency of Impact

Are the impacts of the algorithm and its capabilities clearly communicated to those who are directly and indirectly affected by it?





Strategy 5. Human Oversight or “Humans-in-the-Loop”

Automation bias: the over-reliance by humans on the outputs of AI systems

Algorithm’s outputs are recommendations, not final word

People should:

- Review outputs of automated tools
- Easily override flawed or biased results





[File A Discrimination Complaint With Us](#)

Computer-driven technologies are rapidly reshaping modern society. Advanced technologies have the potential to reduce discrimination. These same technologies, when not used responsibly, can also amplify discrimination in ways that are hard to detect. The New Jersey Division on Civil Rights (DCR) is committed to supporting innovation by helping to prevent and eliminate discrimination in New Jersey and addressing the risks that automated decision-making tools can pose. That is why DCR has launched its Civil Rights and Technology Initiative, which houses DCR's Innovation Lab and DCR's Algorithmic Fairness Project.

Rights Innovation Lab

Through the Civil Rights Innovation Lab, DCR is leveraging innovation responsibly within our own work to advance DCR's mission to prevent, address, and remedy discrimination across New Jersey. The Innovation Lab will be working on developing technology and tools to enhance our enforcement, outreach, and public education work across the State. This includes improving the complaint process, aiding in investigations and enforcement work, updating internal policies and procedures to help the Division better serve all New Jerseyans.

Algorithmic Fairness Project

Through the Algorithmic Fairness Project, DCR is addressing the risks of discrimination and bias-based harassment from the use of advanced technologies. DCR will address how the use of different automated decision-making tools in housing, employment, places of public accommodation, credit, and contracting may violate New Jersey's civil rights laws, and how covered entities can use these tools responsibly and mitigate that risk.



Introduction to Algorithmic Fairness

Algorithmic discrimination occurs when automated decision-making tools are used in a way that illegally disadvantages individuals because of their protected characteristics.

Protected characteristics under the New Jersey Law Against Discrimination (LAD) include race, religion, color, national origin, nationality, sexual orientation, pregnancy, breastfeeding, sex, gender identity, gender expression, disability, and more.

Automated decision-making tools are technological tools that are used to automate all or part of the human decision-making process. They include both emerging technologies like artificial intelligence (AI) and familiar technologies like traditional statistical tools and decision trees.

The LAD prohibits algorithmic discrimination in the same way it prohibits other forms of discrimination in housing, employment, credit, contracting, and places of public accommodation—including schools, banks, police departments, medical facilities, and other places open to the public.

Not every use of automated decision-making tools leads to algorithmic discrimination. Whether these tools result in discrimination depends on their design, development, and application.



Automated Decision-Making Tools in Housing, Work, Schools, and Health Care



Housing Providers may use these tools to:

- screen tenants
- set mortgage terms
- value homes
- approve rentals
- decide which houses to show people
- advertise housing
- approve mortgages

Employers may use these tools to:

- make hiring and firing decisions
- assess interviews
- monitor employees
- review resumes
- make promotion and demotion decisions
- advertise jobs

Schools may use these tools to:

- monitor attendance
- monitor students' computer use
- predict students' likelihood of graduation
- tutor students
- filter and block content available to students
- grade assignments
- aid in disciplinary decisions

Health Care Providers may use these tools to:

- make diagnoses
- prescribe medications
- determine patients' health risks
- decide the type of care a patient receives
- allocate resources between patients



To learn more about automated decision-making tools and the LAD, see our guidance document.



Applying the Strategies

NarxCare, algorithm developed to score patients' risk for developing opioid addiction

Designed to flag potential "doctor shoppers," but 20% of those flagged receive coordinated care for cancer

Scenario:

After surgery, a doctor is prescribing opioids for pain management. She checks the patient's NarxCare score, and notices that it is particularly high. The patient's medical history does not indicate any substance use disorder.



Applying the Strategies



Which strategy would you advise this doctor and hospital to undertake in applying NarxCare scores so that patient care is unbiased and responsible?

How would you apply it?

AI governance (overarching principles for AI use)

Risk management framework (practices that minimize AI harm)

Bias audits (assessment of AI's safety, legality, and ethics)

Transparency (clarity of use and impact)

Humans-in-the-loop (human oversight and ability to override results)



Q&A

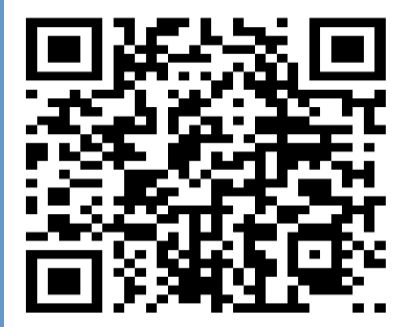




Contact Info:

Oneshia Herring

Senior Vice President and Counsel
National Fair Housing Alliance



Contact Info:

Yolanda N. Melville

Director
New Jersey Division on Civil Rights

