ALGORITHMIC JUSTICE:
A MODEL TO IDENTIFY, DETECT, AND REDUCE DISCRIMINATORY BIASES AND IMPROVE THE ALGORITHMIC STATE OF THE NATION

Emile Loza de Siles
Technology & Cybersecurity Law Group
Duquesne University School of Law (Fall 2019)
April 26, 2019
ROAD MAP

- Introduction.
- Problem Statement.
- National Institute of Standards and Technology ("NIST"), Introduced.
- The Algorithmic Justice Model in Three Parts.
- Law and Policy Developments.
- Concluding Remarks.
The ACLU Used Amazon’s Facial Recognition And It Labelled Congress Members As Criminals

By Dave Gershgorn, July 27, 2018

© 2018-19 Emile Loza de Siles. All rights reserved.
ALGORITHMIC SYSTEMS
UBIQUITOUS | @ SCALE | ACROSS DOMAINS

>200 law review articles published in 2017-2018.

- Access to justice
- Administrative law & regulation
- Compliance
- Credit, lending
- Criminal law, bail, policing
- Cryptocurrency
- Cybersecurity
- Discrimination
- Education
- Employment
- Environment
- Ethics
- Family law
- Finance, investment
- Fiduciaries, estates, taxes
- Government, public assistance
- Human rights, international humanitarian law
- Innovation
- Insurance
- Intellectual property, cultural heritage
- Internet of Things
- Judiciary, dispute resolution
- Legal research, scholarship, practice
- Medicine, healthcare, genetics
- News, fake news, weaponized information
- Parole & sentencing
- Personhood
- Smart cities
- Smart contracts
- Torts
- Voting

© 2018-19 Emile Loza de Siles. All rights reserved.
“Algorithms are still made by human beings, and those algorithms are still pegged to basic human assumptions.”

U.S. Representative Alexandria Ocasio-Cortez
Riverside Church, Harlem, New York
Martin Luther King, Jr. Day 2019
Discriminatory algorithmic biases:

- May arise through the design, development, training, testing, or use of machine learning and other artificial and algorithmic “systems.”
- Are likely prevalent in systems operating ubiquitously at scale and across many diverse domains, and are capable of and doing great and irreparable harm to individuals, their communities, and society.
- May result in negative effects that are compounded and propagated over time and across systems.
- Are almost impossible to identify, detect, control, or enforce against within existing legal mechanisms.

Thus, no effective, practical, systematic, and transparent way exists for discriminatory algorithmic biases to be identified, detected, and reduced (or eradicated). There is an urgent need for a workable model.
In response, I propose an **Algorithm Justice Model** to begin in three parts.

The Model harnesses the technological and scientific expertise, demonstrated successful leadership, bipartisanally esteemed “apolitical” reputation of the **National Institute of Standards and Technology** (“NIST”).

NIST’s legal authority exists as least for the Model’s Part 1; analyses pending as to Parts 2 and 3.

Recent law and policy developments support the Model’s adoption.

For these reasons, the Model is:
- Possible;
- Practical;
- From prior NIST experience, likely to succeed; and
- Can be timely commenced and established within reasonable time.
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
AI RESEARCH SINCE 1950s

© 2018-19 Emile Loza de Siles. All rights reserved.
Non-regulatory agency.

Leading science & technology experts, and relevantly experienced staff.

Leading government authority on cybersecurity and other technical standards and consensus-based frameworks, reference data sets, and cryptographic algorithmic testing and validation.

ALGORITHMIC JUSTICE MODEL
BEGINNING IN THREE PARTS
PART 1

ALGORITHMIC JUSTICE STANDARDS
PROOF OF CONCEPT: NIST CYBERSECURITY FRAMEWORK

Framework for Improving Critical Infrastructure Cybersecurity

Version 1.1
National Institute of Standards and Technology
April 16, 2018

Released Version 1.0 within 1 year of Exec. Order


© 2018-19 Emile Loza de Siles. All rights reserved.
WHICH ALGORITHMIC SYSTEMS?

Is the algorithmic system a WMD or have the potential to be, including in operation with others?

- Is the “participant” aware of [and provide informed consent to] being algorithmically modeled and, if so, what is the model’s use? [Does the participant remain so aware and consenting as the model or its purveyor’s underlying business or operations evolve?]
- Even if aware [and consenting], is the model opaque, invisible[, or reasonably discoverable]?
- Is the model unfair[, deceptive, or otherwise illegal or unethical]? Does it work against the participant[, community, or society’s interest? Does it destroy or damage?
- Is the model scalable? Is its potential for harm exponential? Does it have the capacity to approximate the power of law? [Is there effective, timely, and accessible means of enforcement against the model?]
PART 2

BIG DATA REFERENCE SETS
NIST produces the Nation’s Standard Reference Data (SRD). These data are assessed by experts and are trustworthy such that people can use the data with confidence and base significant decisions on the data. NIST provides 49 free SRD databases and 41 fee-based SRD databases. SRD must be compliant with rigorous critical evaluation criteria.

Send questions to data@nist.gov or call 1(844) 374-0183 (Toll Free).
PART 3

ALGORITHMIC TESTING & VALIDATION
Project Overview

The Cryptographic Algorithm Validation Program (CAVP) provides validation testing of FIPS-approved and NIST-recommended cryptographic algorithms and their individual components. Cryptographic algorithm validation is a prerequisite of cryptographic module validation.

NIST provides the CAVP to test algorithm implementations. An algorithm implementation is validated by NIST based on an appropriate validation list, which identifies the algorithm, implementation, operational environment, validation data, and algorithm details.

Tests

CAVP tests the following cryptographic algorithms. Follow the links to the algorithm specifications, validation testing requirements, validation testing data, and test vectors:

- **Block Ciphers**
  - AES, Triple DES, Skipjack (Encryption only)
  - Tests for CBC, CFB, OFB, and CTR modes.
- **Block Cipher Modes**
  - CCM, CCMx, GCM, 3GEM, 3PGE, Key Wrap, XTS
- **Digital Signatures**
  - FIPS 186-4: DSA, ECDSA, RSA
  - FIPS 186-2: DSA, ECDSA, RSA
- **Key Derivation Functions**
  - HKDF
- **Key Management**
  - HSM
- **Message Authentication**
  - HMAC (FIPS 199-2)

Cryptographic Module Validation Program

Project Overview

What is the Purpose of the CMVP?

On July 17, 1999, NIST established the Cryptographic Module Validation Program (CMVP) that validates cryptographic modules to Federal Information Processing Standards (FIPS) standards, Security Requirements for Cryptographic Modules, and other FIPS cryptography-based standards. FIPS 140-1, Security Requirements for Cryptographic Modules, was released on May 15, 2001, and superseded FIPS 140-1. The CMVP is a joint effort between NIST and the Canadian Centre for Cyber Security (CCCS), a branch of the Communications Security Establishment (CSE).

Modules validated as conforming to FIPS 140-2 are accepted by the Federal Agencies of both countries for the protection of sensitive information. Vendors of cryptographic modules that meet the requirements of the CMVP are listed on NIST's Computer Security Division (CSD) and CCCS websites.

What is the Applicability of CMVP to the U.S. Government?

FIPS 140-1 became a mandatory standard for the protection of sensitive data when the Secretary of Commerce signed the standard on January 15, 1994. FIPS 140-2 superseded FIPS 140-1 and the standard was signed on May 29, 2002. The applicability statement from FIPS 140-2 (page 4.6)
AI LAW & POLICY DEVELOPMENTS
EXECUTIVE ORDER ON MAINTAINING AMERICAN LEADERSHIP IN ARTIFICIAL INTELLIGENCE
(FEB. 11, 2019)

AI EXECUTIVE ORDER

- Announced **American AI Initiative** coordinated by National Science & Technology Council’s Select Committee on Artificial Intelligence.

- To “unleash America’s AI resources,” subject to controls to foster trusted AI environment and, in turn, spur research, innovation, use, industry, the economy, and national security.

- By August 11, 2019, OMB is to issue memorandum informing agencies re:
  - “Development of regulatory and non-regulatory approaches” to AI-related technologies & industries within their jurisdictions; and
  - Other ways to reduce barriers to AI use and advance the Order’s objectives.
AI ORDER & ALGORITHMIC JUSTICE MODEL

- Stated policy goals to protect civil liberties and privacy and for the government to promote “trust of the American people in the development and deployment of AI-related technologies.”

- As to NIST, by August 11, 2019, the Commerce Secretary, through NIST, is to issue a plan for “Federal engagement in the development of technical standards and related tools in support of reliable, robust, and trustworthy systems that use AI technologies.”
AI ORDER & ALGORITHMIC JUSTICE MODEL

- **As to technical standards**, the Federal government is to drive the development of “appropriate” such standards “and reduce barriers to the safe testing and deployment of AI technologies.”

- Further, those standards are to **minimize cybervulnerabilities** and “**reflect Federal priorities for innovation, public trust, and public confidence in AI-employing systems.**”
As to Big Data, agencies must “[e]nhance access to high-quality and fully traceable Federal data, models, and computing resources to increase the value of such resources for AI R&D, while maintaining safety, security, privacy, and confidentiality protections consistent with applicable laws and policies.”

They also must identify barriers and requirements as to “increased access to and use of such data and models,” including, among others:

- “Privacy and civil liberty protections for individuals who may be affected by increased access and use, as well as confidentiality protections for individuals and other data providers; and
- Safety and security concerns, including those related to the association or compilation of data and models[.]”
NEW YORK CITY LOCAL LAW 49

- “A Local Law in Relation to Automated Decision Systems used by Agencies”

- **Automated decision system**: “computerized implementations of algorithms, including those derived from machine learning or other data processing or artificial intelligence techniques, which are used to make or assist in making decisions.”

- Requires creation of task force to review & provide recommendations on how information on agency ADSs may be shared with the public and how agencies may address instances where people are disproportionately impacted by agency ADSs.
LOCAL LAW 49
DISPROPORTIONATE IMPACT CLASSES

- Age
- Race
- Creed
- Color
- Religion
- National origin
- Gender
- Disability
- Marital status
- Partnership status
- Caregiver status
- Sexual orientation
- Alienage
- Citizenship status
UPDATE: LOCAL LAW 49

- Jan. 11, 2018 – Law enacted.
- May 2018 – Mayor DeBlasio appoints task force.
- Apr. 4, 2019 - New York City Council Committee on Technology released its report and holds Oversight Hearing.
- Apr. 30, 2019 – Task force plans first public meeting.
Testimony at April 4, 2019 Oversight Hearing of Committee on Technology

Jeff Thamkittikasem, ADS task force chair: Took significant time to determine which agency systems fit within ADS definition, but do not plan to provide list of which such “algorithms”

Other testimony, including from task force members: Non-cooperation, a lack of transparency in decision-making, including City directing halt of agency ADS questionnaire use, a lack of access and information about city agencies’ ADSs, limited public engagement, and other serious problems.
FEDERAL LEGISLATIVE INITIATIVES

- S1108 & HR 2231: Algorithmic Accountability Act of 2019
  Senator Wyden & Booker, Rep. Yvette Clark

- HR 2202: Growing Artificial Intelligence Through Research Act (GrAITR Act)
  Rep. Daniel Lipinski

Note: HR 6629, 2018 NIST Reauthorization Bill
Automated decision system: “computational process, including one derived from machine learning, statistics, or other data processing or artificial intelligence techniques, that makes a decision or facilitates human decision making, that impacts consumers.”

Federal Trade Commission

Covered entity: $50M annual revenues, each of last 3 years & 1M+ consumers or devices; ownership or control; data broker.
HIGH RISK ADS
UNDER ALGORITHMIC ACCOUNTABILITY ACT AS INTRODUCED
SURVEILLANCE AND OTHER FEATURES CLASSIFY ADS AS HIGH RISK
INCLUDING

- Race
- Color
- National origin
- Political opinions
- Religion
- Trade union membership
- Genetic or biometric data
- Health
- Gender
- Gender identity
- Sexuality
- Sexual orientation
- Criminal convictions
- Arrests

© 2018-19 Emile Loza de Siles. All rights reserved.
The proposed **Algorithm Justice Model** consists of three parts by which to effectively identify, detect, and reduce discriminatory algorithmic biases by:

- Assigning leadership and responsibility to the **National Institute of Standards and Technology** to:
  - Part 1: Develop Algorithmic Justice Standards under its current legal authorities and following the consensus-driven processes that those it used to successfully and quickly establish the widely-praised and -adopted Cybersecurity Framework;
  - Part 2: Use its capacities and expertise in reference standards data programs to develop and curate Big Data for training and testing of algorithmic systems; and
  - Part 3: Use its cryptographic algorithm and module testing and validation approach, capacities, and expertise to develop such systems for algorithmic justice purposes.

The Algorithmic Justice Model is **supported by AI law and policy developments and can be further incorporated** therein. The Model can be timely commenced and established within reasonable time.
ALGORITHMIC JUSTICE: A MODEL TO IDENTIFY, DETECT, AND REDUCE DISCRIMINATORY BIASES TO IMPROVE THE ALGORITHMIC STATE OF THE NATION

Emile Loza de Siles
Technology & Cybersecurity Law Group
Duquesne University School of Law (Fall 2019)
April 26, 2019