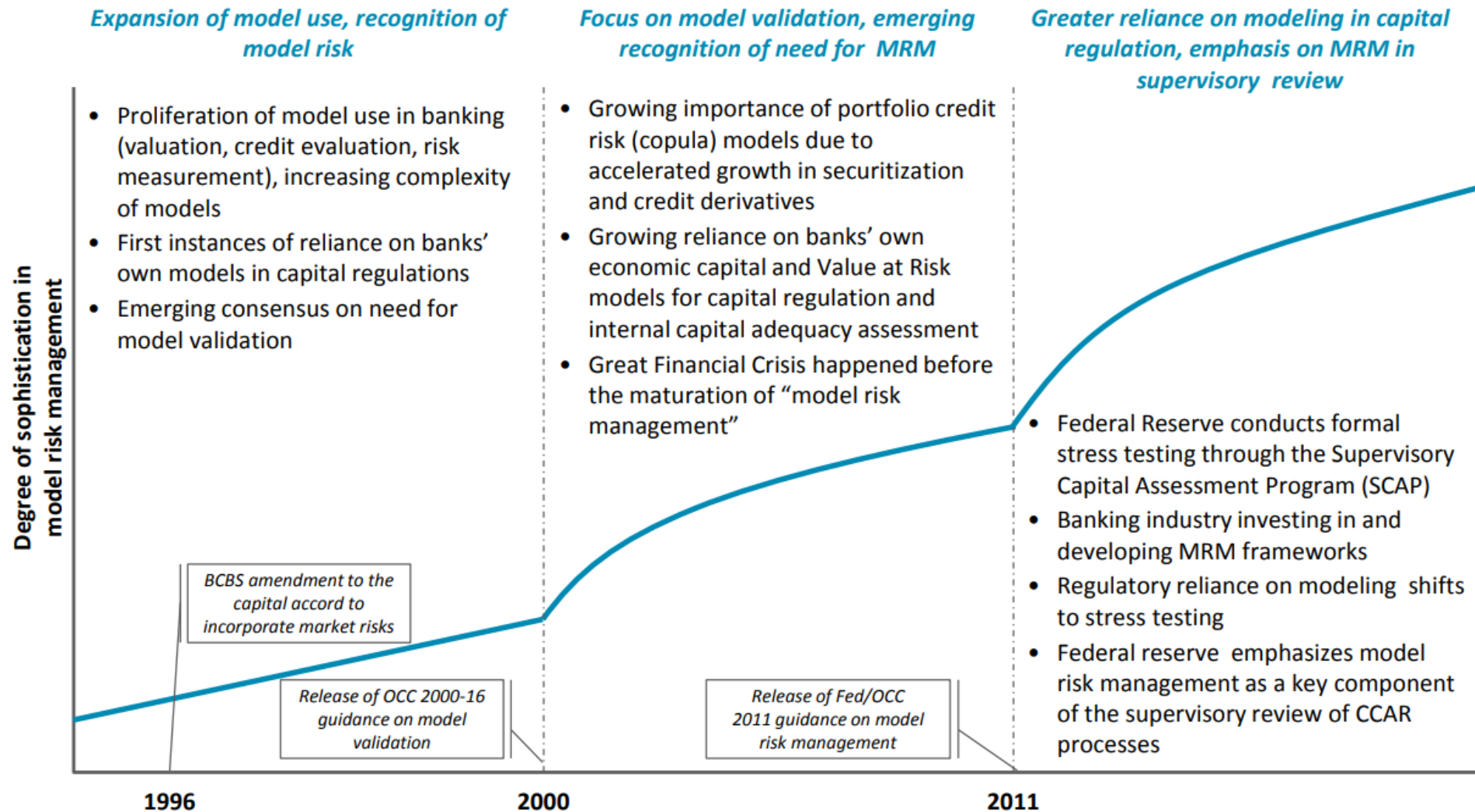# RESPONSIBLE AI

Week 5: AI Regulations

TODAY'S OBJECTIVES

- What precedents are there for AI regulation? Why does AI need special regulation?

- What does the impossibility theorem mean for designing laws around AI? Where is humanity today in terms of AI regulation? What will the regulatory landscape look like as you graduate into data science roles in the workforce?

- How can regulations require AI to be fair and unbiased when those definitions aren't standardized or compatible?

HISTORY OF REGULATIONS AROUND QUANTITATIVE MODELS

Expansion of model use, recognition of model risk

- Proliferation of model use in banking (valuation, credit evaluation, risk measurement), increasing complexity of models
- First instances of reliance on banks' own models in capital regulations
- Emerging consensus on need for model validation

Focus on model validation, emerging recognition of need for MRM

- Growing importance of portfolio credit risk (copula) models due to accelerated growth in securitization and credit derivatives
- Growing reliance on banks' own economic capital and Value at Risk models for capital regulation and internal capital adequacy assessment
- Great Financial Crisis happened before the maturation of "model risk management"

Greater reliance on modeling in capital regulation, emphasis on MRM in supervisory review

- Federal Reserve conducts formal stress testing through the Supervisory Capital Assessment Program (SCAP)
- Banking industry investing in and developing MRM frameworks
- Regulatory reliance on modeling shifts to stress testing
- Federal reserve emphasizes model risk management as a key component of the supervisory review of CCAR processes

Degree of sophistication in model risk management

BCBS amendment to the capital accord to incorporate market risks

Release of OCC 2000-16 guidance on model validation

Release of Fed/OCC 2011 guidance on model risk management

1996   2000   2011

Source: https://wifpr.wharton.upenn.edu/wp-content/uploads/2021/07/15-01.pdf

WHY DO WE NEED AI-SPECIFIC REGULATION?

- Scale

- Intersection and application across specialized fields
  - Cybersecurity and privacy
  - Medicine
  - Human capital / employment law
  - Content moderation
  - Housing
  - Ethics

- Reliance on data
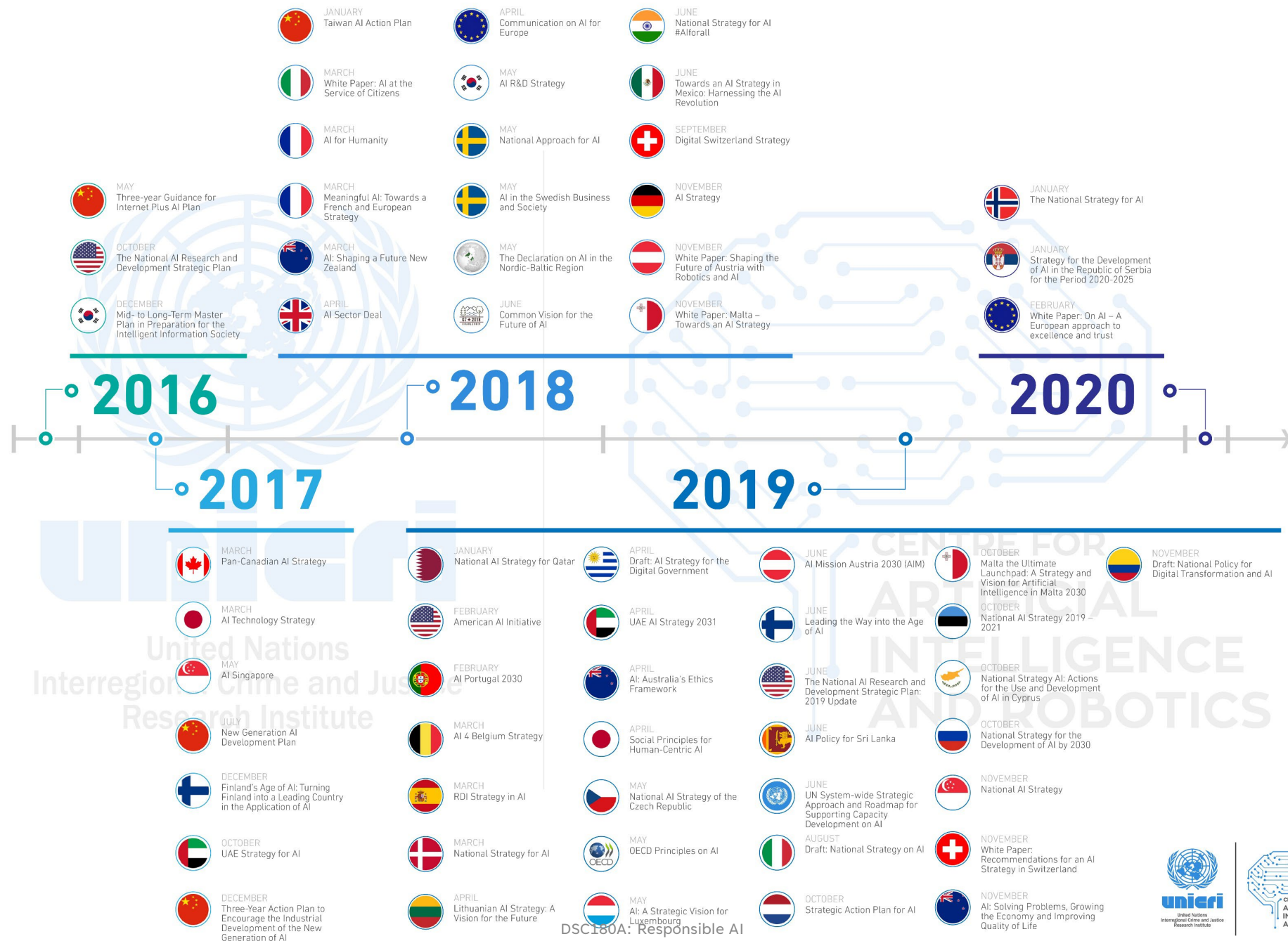 - Bias within different stages of the AI model lifecycle

- Complexity

- Opacity
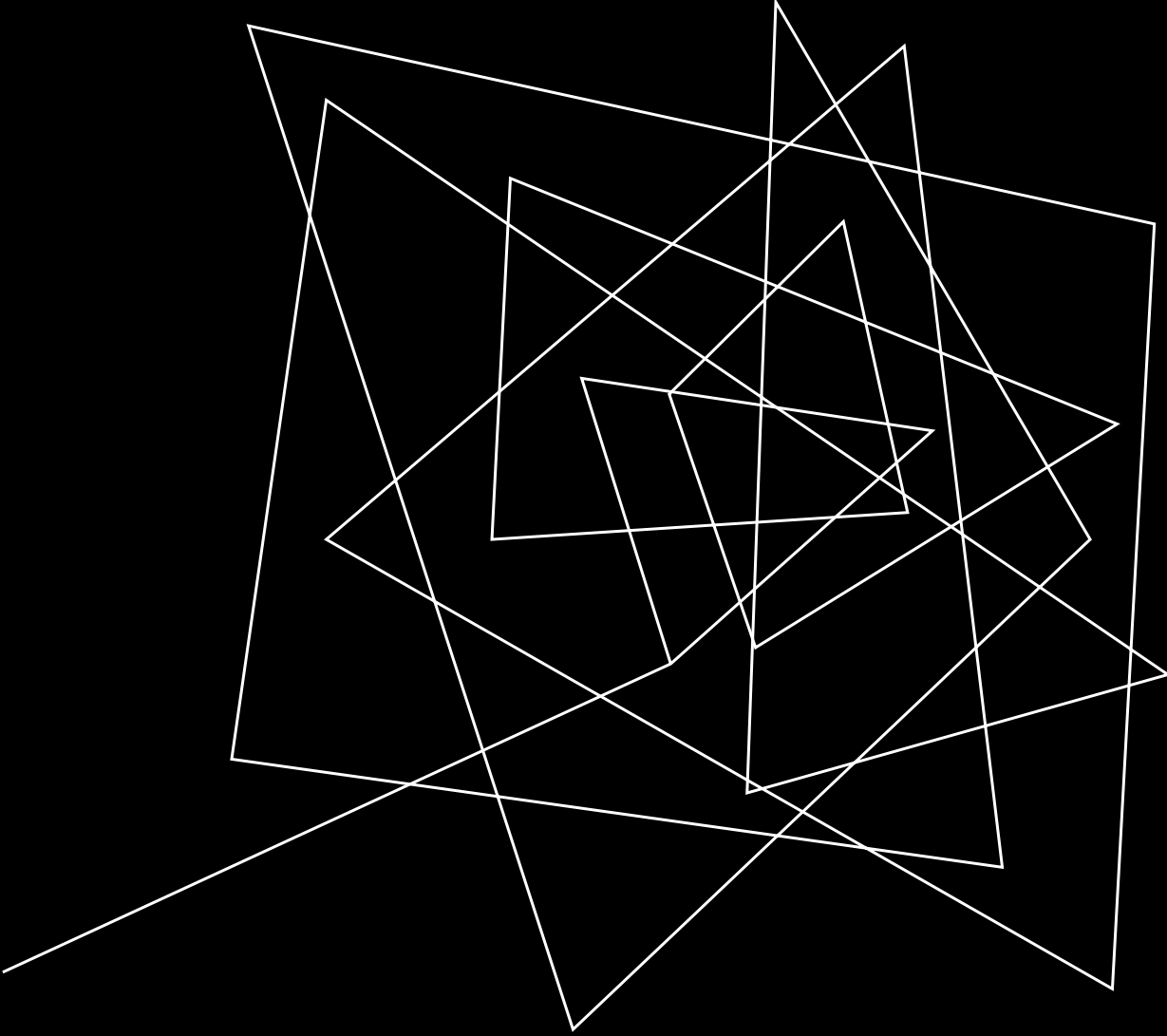
READING PRESENTATION #1

Big Data's Disparate Impact (Barocas and Selbst)

Challenges & Limitations to AI Regulations

- **Impossibility Theorem:** How to choose and balance definitions of fairness?

- **(Lack of) Standardization:** There is no "source of truth" for AI Regulations – how will these regulations be consistently assessed and implemented?

- **False Sense of Security:** Will users "let their guard down" if they believe their AI is regulated?

- **Legal Issues:** Bias Mitigation as Affirmative Action?
    - Equal-protection doctrine may threaten algorithmic fairness. Equal protection doctrine prevents government entities from treating similarly situated individuals differently ("anticlassification"). **This approach prohibits making decisions based on protected attributes** (i.e. hiring someone because of their race or gender). However, leading bias mitigation approaches often **purposefully use protected attributes to promote fair or unbiased outcomes in their algorithms**
    - There is a high level of legal scrutiny required to bypass anticlassification, which is why affirmative action cases are so frequently brought to the courts
    - The courts are likely to consider leading bias mitigation techniques a type of affirmative action, which subjects them to strict scrutiny (= harder to get passed) of the law and even potentially renders them illegal

JANUARY
Taiwan AI Action Plan

APRIL
Communication on AI for Europe

JUNE
National Strategy for AI #AIforall

MARCH
White Paper: AI at the Service of Citizens

MAY
AI R&D Strategy

JUNE
Towards an AI Strategy in Mexico: Harnessing the AI Revolution

MARCH
AI for Humanity

MAY
National Approach for AI

SEPTEMBER
Digital Switzerland Strategy

MAY
Three-year Guidance for Internet Plus AI Plan

MARCH
Meaningful AI: Towards a French and European Strategy

MAY
AI in the Swedish Business and Society

NOVEMBER
AI Strategy

JANUARY
The National Strategy for AI

OCTOBER
The National AI Research and Development Strategic Plan

MARCH
AI: Shaping a Future New Zealand

MAY
The Declaration on AI in the Nordic-Baltic Region

NOVEMBER
White Paper: Shaping the Future of Austria with Robotics and AI

JANUARY
Strategy for the Development of AI in the Republic of Serbia for the Period 2020-2025

DECEMBER
Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society

APRIL
AI Sector Deal

JUNE
Common Vision for the Future of AI

NOVEMBER
White Paper: Malta – Towards an AI Strategy

FEBRUARY
White Paper: On AI – A European approach to excellence and trust

2016

2018

2020

2017

2019

MARCH
Pan-Canadian AI Strategy

JANUARY
National AI Strategy for Qatar

APRIL
Draft: AI Strategy for the Digital Government

JUNE
AI Mission Austria 2030 (AIM)

OCTOBER
Malta the Ultimate Launchpad: A Strategy and Vision for Artificial Intelligence in Malta 2030

NOVEMBER
Draft: National Policy for Digital Transformation and AI

MARCH
AI Technology Strategy

FEBRUARY
American AI Initiative

APRIL
UAE AI Strategy 2031

JUNE
Leading the Way into the Age of AI

OCTOBER
National AI Strategy 2019 – 2021

MAY
AI Singapore

FEBRUARY
AI Portugal 2030

APRIL
AI: Australia's Ethics Framework

JUNE
The National AI Research and Development Strategic Plan: 2019 Update

OCTOBER
National Strategy AI: Actions for the Use and Development of AI in Cyprus

JULY
New Generation AI Development Plan

MARCH
AI 4 Belgium Strategy

APRIL
Social Principles for Human-Centric AI

JUNE
AI Policy for Sri Lanka

OCTOBER
National Strategy for the Development of AI by 2030

DECEMBER
Finland's Age of AI: Turning Finland into a Leading Country in the Application of AI

MARCH
RDI Strategy in AI

MAY
National AI Strategy of the Czech Republic

JUNE
UN System-wide Strategic Approach and Roadmap for Supporting Capacity Development on AI

NOVEMBER
National AI Strategy

OCTOBER
UAE Strategy for AI

MARCH
National Strategy for AI

MAY
OECD Principles on AI

AUGUST
Draft: National Strategy on AI

NOVEMBER
White Paper: Recommendations for an AI Strategy in Switzerland

DECEMBER
Three-Year Action Plan to Encourage the Industrial Development of the New Generation of AI

APRIL
Lithuanian AI Strategy: A Vision for the Future

MAY
AI: A Strategic Vision for Luxembourg

OCTOBER
Strategic Action Plan for AI

NOVEMBER
AI: Solving Problems, Growing the Economy and Improving Quality of Life

UNICRI

CENTRE FOR ARTIFICIAL INTELLIGENCE AND ROBOTICS

United Nations Interregional Crime and Justice Research Institute

DSC180A: Responsible AI

7

AI Policy Efforts in the US

FEBRUARY 11, 2019

Executive Order 13859: Maintaining American Leadership in Artificial Intelligence

Announced the American AI Initiative

*"The United States must foster **public trust and confidence in AI technologies** and protect civil liberties, privacy, and American values in their application in order to fully realize the potential of AI technologies for the American people."*

FEBRUARY 26, 2020

American AI Initiative: Year One Annual Report

Provided a summary of progress and long-term vision for the American AI Initiative, emphasizing the need to *"**embrace trustworthy AI** for government services and missions."*

NOVEMBER 17, 2020

Office of Management and Budget (OMB) Memorandum M-21-06: Guidance for Regulation of Artificial Intelligence Applications

Provided guidance to Federal agencies, including considerations for reducing barriers to AI development and adoption

*"The government's regulatory and non-regulatory approaches to AI should contribute to public trust in AI by **promoting reliable, robust, and trustworthy AI applications**."*

DECEMBER 3, 2020

Executive Order 13960: Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government

Outlined a set of principles and actions to accelerate trustworthy AI use and development

JANUARY 1, 2021

National AI Initiative Act of 2020

*"The mission of the National AI Initiative is to ensure continued U.S. leadership in AI research and development, **lead the world in the development and use of trustworthy AI** in the public and private sectors..."*

OCTOBER 4, 2022

Blueprint for the AI Bill of Rights

Outlines five protections Americans should have in the AI age:

1. Safe and Effective Systems
2. Algorithmic Discrimination Protection
3. Data Privacy
4. Notice and Explanation, and
5. Human Alternatives, Consideration, and Fallback

What would binding AI regulation look like?

BRIEFING

EU Legislation in Progress

European Parliament

Artificial intelligence act

OVERVIEW

The European Commission unveiled a new proposal for an EU regulatory framework on artificial intelligence (AI) in April 2021. The draft AI act is the first ever attempt to enact a horizontal regulation of AI. The proposed legal framework focuses on the specific utilisation of AI systems and associated risks. The Commission proposes to establish a technology-neutral definition of AI systems in EU law and to lay down a classification for AI systems with different requirements and obligations tailored on a 'risk-based approach'. Some AI systems presenting 'unacceptable' risks would be prohibited. A wide range of 'high-risk' AI systems would be authorised, but subject to a set of requirements and obligations to gain access to the EU market. Those AI systems presenting only 'limited risk' would be subject to very light transparency obligations. While generally supporting the Commission's proposal, stakeholders and experts call for a number of amendments, including revising the definition of AI systems, broadening the list of prohibited AI systems, strengthening enforcement and redress mechanisms and ensuring proper democratic oversight of the design and implementation of EU AI regulation.

https://artificialintelligenceact.eu/the-act/

Federal labor laws and protections have been found in courts to extend to AI-based decisions...

... to protect candidates regardless of age, gender, race, religion, nationality, disabilities, etc. ...

... throughout the sourcing, contacting, screening, and interview of potential candidates...

... all the way down to the state and local level.

Title VII of the Civil Rights Act of 1964 (Title VII)
- a federal law that protects employees and applicants against discrimination based on certain specified characteristics such as race, color, national origin, sex, and religion
- prohibits discrimination based on disparate treatment and/or disparate impact
- a court could find that an employer faces the same liability for a program exhibiting the unconscious bias of its programmer as it would if the programmer had made the hiring decision him or herself, based on that bias.

Age Discrimination in Employment Act (ADEA)
- forbids age discrimination against people who are age 40 or older. It does not protect workers under the age of 40, although some states have laws that protect younger workers from age discrimination

Americans with Disabilities Act (ADA)
- An A.I.-hiring practice could also implicate the Americans with Disabilities Act ("ADA") if an algorithm discerns an applicant's physical disability, mental health, or clinical diagnosis, all of which are forbidden inquiries in pre-employment candidate assessments. The ADA Amendments Act of 2008 broadened the statutory definition of "disability," increasing the scope of individuals whom the ADA protects.

Equal Employment Opportunity Commission (EEOC)
- enforces disability discrimination laws with respect to employers in the private sector and the federal government
- has issued guidance qualifying the expanded list of personality disorders identified in the psychiatric literature as protected mental impairments
- has already investigated at least two instances of alleged A.I. bias, and has made clear that employers using A.I. hiring practices could face liability for any unintended discrimination.
- in 2017, found reasonable cause to believe an employer violated the ADEA by advertising on Facebook for a position within its company and "limiting the audience for their advertisement to younger applicants."

Uniformed Services Employment and Reemployment Rights Act (USERRA)
- Maintains reemployment, antidiscrimination, and antiretaliation rights for members of uniformed services

State and Local Legislation
- New York City
  - Local law 144 of 2021 prohibits employers from using an automated employment decision tool unless such tool has been subject to a bias audit within one year of the use of the tool. (Laws goes live in Jan 2023).
- Illinois (Artificial Intelligence Video Interview Act)
  - imposes strict limitations on employers who use A.I. to analyze candidate video interviews. Employers must a) notify applicants that A.I. will be used in their interviews; b) obtain consent to use A.I.; c) explain to applicants how the A.I. works and what characteristics the A.I. will track in relation to the position; d) limit sharing of the video interview; and e) comply with requests to destroy the video within 30 days
- Maryland
  - the prohibition of facial/voice recognition software in interviews

DISCUSSION

The NYC Department of Consumer and Worker Protection recently released additional rules and clarification around the NYC hiring law.

New York City Department of Consumer and Worker Protection

Notice of Public Hearing and Opportunity to Comment on Proposed Rules

What are we proposing? The Department of Consumer and Worker Protection ("DCWP" or "Department") is proposing to add rules to implement new legislation related to the use of automated employment decision tools. The proposed rules would clarify the requirements for the use of automated employment decision tools within New York City, the notices to employees and candidates for employment regarding the use of the tool, the bias audit for the tool, and the required published results of the bias audit.

Follow the link provided in chat and discuss these **three** questions in your breakout rooms:

- **What metric of fairness is the Department proposing**? (ctrl+f 'impact ratio', 'selection rate', and 'bias audit')

- **What are the strengths and weaknesses of this approach**? Does this go far enough?
- Unlike familiar financial audits, which have been refined over decades of accounting experience, the AI audit process is new and without clearly established guidelines:
  - … the Surveillance Technology Oversight Project, New York Civil Liberties Union and other organizations has noted the lack of standards for bias audits
  - …. businesses and their service providers are grappling with how to comply with New York City's mandate for AI audits used in hiring

  What does challenges could this pose for different stakeholders (think through the different stakeholders we discussed in previous lectures!)**?**

Make sure to nominate someone to represent your group and share your thoughts with the room when we reconvene!

FOR NEXT WEEK

- Complete next week's **readings**
  - If you signed up to present **Ethical Machine Learning in Health Care (Chen et al. 2021)** come prepared to present next week and submit your presentation to Gradescope by 10 AM PT, Friday, November 4th

- Submit your answers to next week's participation questions on Gradescope by 10 PM PT, Thursday, November 3rd

- **Replication Part #2:** Notebook and writeup on model development and fairness metric assessments
  - Primary contact for replication project: Nandita Rahman (nanrahman@deloitte.com)
  - Office hours: Mondays 1-2pm PST

- First replication project checkpoint is this weekend: submit EDA notebook