

# State Governance Considerations on AI

**Sorelle Friedler**

Shibulal Family Associate Professor



**HVERFORD**  
COLLEGE

DEPARTMENT OF COMPUTER SCIENCE







Why is AI important?

## FINANCIAL TIMES

# Nvidia hits \$1tn market cap as chipmaker rides AI wave

Silicon Valley company joins elite group of US-listed companies including Apple, Microsoft, Amazon and Alphabet

 USA

	<b>Apple</b> 1 AAPL	\$2.788 T
	<b>Microsoft</b> 2 MSFT	\$2.462 T
	<b>Alphabet (Google)</b> 3 GOOG	\$1.576 T
	<b>Amazon</b> 4 AMZN	\$1.248 T
	<b>NVIDIA</b> 5 NVDA	\$991.99 B
	<b>Meta Platforms (Facebook)</b> 6 META	\$672.76 B

## FORTUNE

TECH · A.I.

# ChatGPT could rocket Microsoft's valuation another \$300 billion after Nvidia's massive gains, according to analyst Dan Ives

BY TRISTAN BOVE

May 30, 2023 at 2:24 PM EDT



Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.

### *Signatories:*

- AI Scientists     Other Notable Figures

**Geoffrey Hinton**

Emeritus Professor of Computer Science, University of Toronto

**Yoshua Bengio**

Professor of Computer Science, U. Montreal / Mila

**Demis Hassabis**

CEO, Google DeepMind

**Sam Altman**

CEO, OpenAI



Killer robots are not a near-term concern!  
But there are important policy implications of AI as it exists today.



The Washington Post  
Democracy Dies in Darkness

# AI and the future of our food

By Erin Blakemore  
February 28, 2022 at 9:00 a.m. EST



A tractor sprays a soybean field during the spring. (iStock)

[Comment](#) 8 [Save](#) [Gift Article](#)

Robots. Drones. Artificial Intelligence.

All three are touted as potential saviors for farmers, and are already being deployed on large farms, where they assist with such tasks as managing crops, milking cows and helping farmers make decisions about their land.

The potential benefits are huge. Increases in farm productivity could help feed the approximately 2.4 billion people around the world who experience food insecurity and malnutrition and revolutionize the way farmers use their land.

That could come at a cost. The analysis points out potential flaws in the agricultural data that fuels AI-powered systems and the possibility that autonomous systems could place productivity over the environment. That could lead to inadvertent errors causing overfertilization, dangerous pesticide use, inappropriate irrigation or erosion, risking crop yields, water supplies and soil. And wide-scale crop failures could exacerbate food insecurity.





REPORT | APRIL 20, 2023



# AI in Hiring and Evaluating Workers: What Americans Think

*62% believe artificial intelligence will have a major impact on jobholders overall in the next 20 years, but far fewer think it will greatly affect them personally. People are generally wary and uncertain of AI being used in hiring and assessing workers*

BY LEE RAINIE, MONICA ANDERSON, COLLEEN MCCLAIN, EMILY A. VOGELS AND RISA GELLES-WATNICK

## Would you want to apply for a job that uses AI to help make hiring decisions?

% of U.S. adults who say they would or would not want to apply for a job with an employer that uses artificial intelligence to help in hiring decisions

66% say No

32% say Yes

### Americans widely oppose employers using AI to make final hiring decisions, track workers' movements while they work, and analyze their facial expressions

% of U.S. adults who say they \_\_\_ employers' use of artificial intelligence for each of the following



Note: Those who did not give an answer are not shown.  
Source: Survey of U.S. adults conducted Dec. 12-18, 2022.  
"AI in Hiring and Evaluating Workers: What Americans Think"

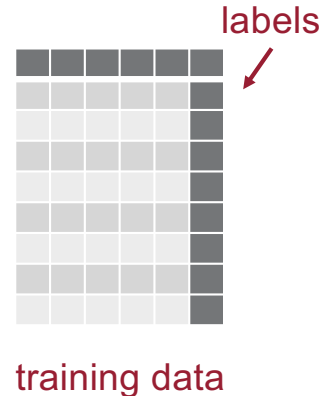
PEW RESEARCH CENTER



What is AI?

# A Basic AI Pipeline

## Training



## Examples:

- breast cancer scans with radiologist highlighted concerns
- resumes with historical hire / no hire decisions from previous company processes
- text prompts with written responses from specialized contractors

## Data takeaways:

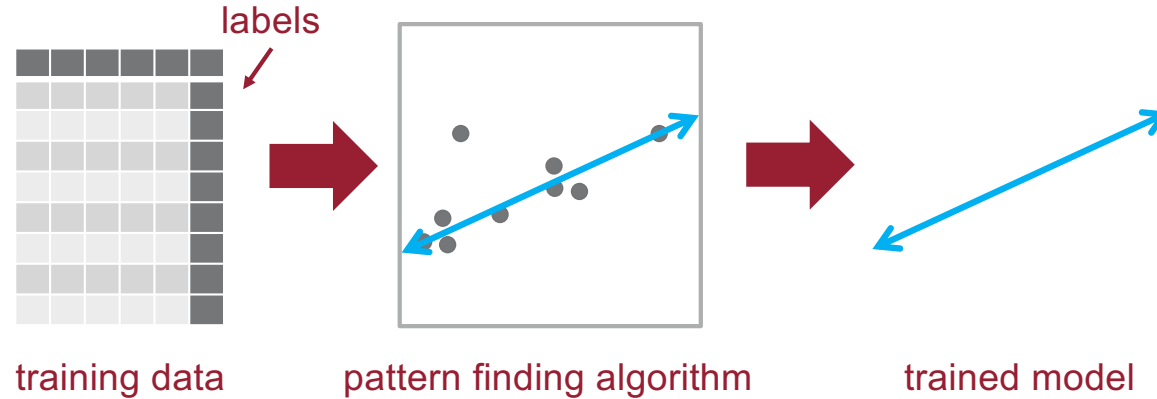
- Requires data that is accurately able to represent the goal – this is **not magic!**
- Uses data collected about people who may have **privacy** concerns with its use.



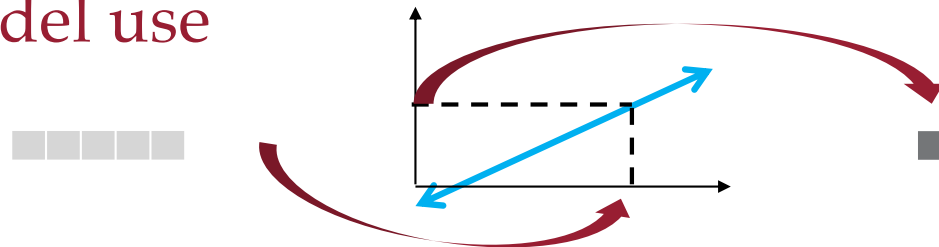


# A Basic AI Pipeline

## Training

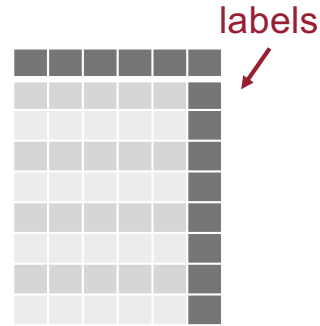


## Model use



# A Basic AI Pipeline

## Training

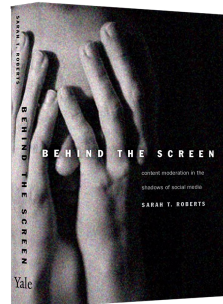
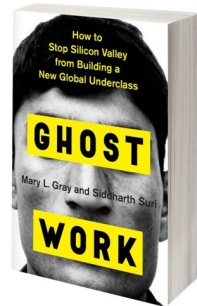


training data

## Examples:

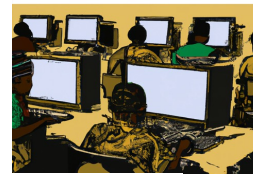
- breast cancer scans with radiologist highlighted concerns
- resumes with historical hire / no hire decisions from previous company processes
- text prompts with written responses from specialized contractors

Manual labor from people makes this possible!



## TIME

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic



BY BILLY PERRIGO  
JANUARY 18, 2023 7:00 AM EST



How can policymakers intervene?

**BLUEPRINT FOR AN  
AI BILL OF  
RIGHTS**

**MAKING AUTOMATED  
SYSTEMS WORK FOR  
THE AMERICAN PEOPLE**

**OCTOBER 2022**



# Blueprint for an AI Bill of Rights

THE WHITE HOUSE



## Safe and Effective Systems

*You should be protected from unsafe or ineffective systems.*

## Algorithmic Discrimination Protections

*You should not face discrimination by algorithms and systems should be used and designed in an equitable way.*

## Data Privacy

*You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used.*

## Notice and Explanation

*You should know when an automated system is being used and understand how and why it contributes to outcomes that impact you.*

## Human Alternatives, Consideration, and Fallback

*You should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems you encounter.*



President Biden

@POTUS

United States government official

Artificial Intelligence has enormous potential to tackle some of our toughest challenges.

But we must address its risks.

That's why last year, we proposed an AI Bill of Rights to ensure that important protections for the American people are built into AI systems from the start.

4:05 PM · Apr 4, 2023 · 3.9M Views



President Biden

@POTUS

United States government official

When it comes to AI, we must both support responsible innovation and ensure appropriate guardrails to protect folks' rights and safety.

Our Administration is committed to that balance, from addressing bias in algorithms – to protecting privacy and combating disinformation.

5:05 PM · Apr 4, 2023 · 2.2M Views

<http://www.whitehouse.gov/ostp/ai-bill-of-rights>

# A Technical Companion to the Blueprint for an AI Bill of Rights

## **1** WHY THIS PRINCIPLE IS IMPORTANT:

This section provides a brief summary of the problems that the principle seeks to address and protect against, including illustrative examples.

## **2** WHAT SHOULD BE EXPECTED OF AUTOMATED SYSTEMS:

- The expectations for automated systems are meant to serve as a blueprint for the development of additional technical standards and practices that should be tailored for particular sectors and contexts.
- This section outlines practical steps that can be implemented to realize the vision of the Blueprint for an AI Bill of Rights. The expectations laid out often mirror existing practices for technology development, including pre-deployment testing, ongoing monitoring, and governance structures for automated systems, but also go further to address unmet needs for change and offer concrete directions for how those changes can be made.

## **3** HOW THESE PRINCIPLES CAN MOVE INTO PRACTICE:

This section provides real-life examples of how these guiding principles can become reality, through laws, policies, and practices. It describes practical technical and sociotechnical approaches to protecting rights, opportunities, and access.

But how can we do this, concretely?

Specific recommendations



Identifying systems of concern



# Applying the Blueprint for an AI Bill of Rights

**THIS FRAMEWORK DESCRIBES PROTECTIONS THAT SHOULD BE APPLIED WITH RESPECT TO ALL AUTOMATED SYSTEMS THAT HAVE THE POTENTIAL TO MEANINGFULLY IMPACT INDIVIDUALS' OR COMMUNITIES' EXERCISE OF:**

## **RIGHTS, OPPORTUNITIES, OR ACCESS**

**Civil rights, civil liberties, and privacy**, including freedom of speech, voting, and protections from discrimination, excessive punishment, unlawful surveillance, and violations of privacy and other freedoms in both public and private sector contexts;

**Equal opportunities**, including equitable access to education, housing, credit, employment, and other programs; or,

**Access to critical resources or services**, such as healthcare, financial services, safety, social services, non-deceptive information about goods and services, and government benefits.

# Definitions

**CONSEQUENTIAL DECISION.**— “Consequential decision” means a decision or judgment that has a legal, material, or similarly significant effect on an individual’s life relating to the impact of, access to, or the cost, terms, or availability of, any of the following:

- (1) **Employment**, workers management, or self-employment, including, but not limited to, all of the following: (A) Pay or promotion. (B) Hiring or termination. (C) Automated task allocation.
- (2) **Education** and vocational training, including, but not limited to, all of the following:
  - (A) Assessment, including, but not limited to, detecting student cheating or plagiarism.
  - (B) Accreditation. (C) Certification. (D) Admissions. (E) Financial aid or scholarships.
- (3) **Housing** or lodging, including rental or short-term housing or lodging.
- (4) **Essential utilities**, including electricity, heat, water, internet or telecommunications access, or transportation.
- (5) **Family planning**, including adoption services or reproductive services, as well as assessments related to child protective services.
- (6) **Health care or health insurance**, including mental health care, dental, or vision.
- (7) **Financial services**, including a financial service provided by a mortgage company, mortgage broker, or creditor.
- (8) **The criminal justice system**, including, but not limited to, all of the following: (A) Risk assessments for pretrial hearings. (B) Sentencing. (C) Parole.
- (9) **Legal services**, including private arbitration or mediation.
- (10) **Voting**.
- (11) **Access to benefits or services or assignment of penalties**.



# Options

- **Sector-specific scoping**

- **Example:** “Health and health insurance technologies such as medical AI systems and devices, AI-assisted diagnostic tools, algorithms or predictive models used to support clinical decision making, medical or insurance health risk assessments, drug addiction risk assessments and associated access algorithms, wearable technologies, wellness apps, insurance care allocation algorithms, and health insurance cost and underwriting algorithms.”

list from: White House AI Bill of Rights: Examples of Automated Systems

<https://www.whitehouse.gov/ostp/ai-bill-of-rights/examples-of-automated-systems/>

- **Regulatory refinement**

- Identify “consequential decisions” and staff a state agency to update a list of covered algorithms in those areas.



Specific recommendations

↳ Ensuring each principle

# Safety and Efficacy



**Prediction: Bias**

## **Predictive Policing Software Terrible At Predicting Crimes**

A software company sold a New Jersey police department an algorithm that was right less than 1% of the time

By [Aaron Sankin](#) and [Surya Mattu](#)

October 2, 2023 10:00 ET

Photo collage by Gabriel Hongsdusit, Getty Image by by Steve Skinner Photography

<https://themarkup.org/prediction-bias/2023/10/02/predictive-policing-software-terrible-at-predicting-crimes>



# Safety and Efficacy

- **Preemptive and ongoing requirements**
  - Sector-specific and/or regulations from a Tech-focused agency
    - e.g., requirements that policing technology be shown to work
  - Set up a mechanism where concentrated technical talent can work with sector-specific agencies
- **Create narrow and specific red lines**
  - Ban on affective AI in law enforcement



# Sector-specific approaches

## Example: employment

- Americans don't want employers to track movements or facial expressions
- Americans want to know that a final hiring decision is made by a person

### Options:

- Define a list of employment-specific algorithms
- Set out principles / goals
- Have the state Department of Labor issue guidance on meeting these principles

### Americans widely oppose employers using AI to make final hiring decisions, track workers' movements while they work, and analyze their facial expressions

% of U.S. adults who say they \_\_\_ employers' use of artificial intelligence for each of the following



Note: Those who did not give an answer are not shown.  
Source: Survey of U.S. adults conducted Dec. 12-18, 2022.  
"AI in Hiring and Evaluating Workers: What Americans Think"

PEW RESEARCH CENTER



# Preemptive requirements

## Example: employment

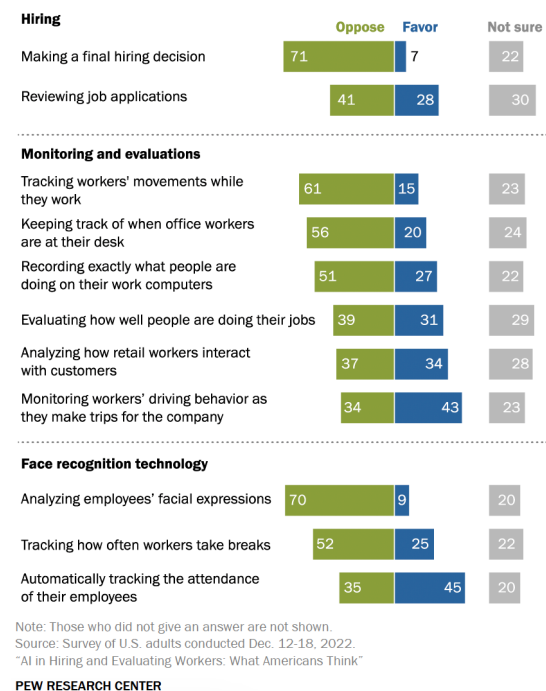
- Americans don't want employers to track movements or facial expressions
- Americans want to know that a final hiring decision is made by a person

### Options:

- Define a list of employment-specific algorithms
- Set out principles / goals
- Have the state Department of Labor issue guidance on meeting these principles
- **Require that this guidance is met *before* any such system can be used in the state**

### Americans widely oppose employers using AI to make final hiring decisions, track workers' movements while they work, and analyze their facial expressions

% of U.S. adults who say they \_\_\_ employers' use of artificial intelligence for each of the following





# Prohibit Algorithmic Discrimination

- **Why? Examples:**

- Loan underwriting and pricing model charged **HBCU alums** more
- Hiring tool rejected applicants with “**women’s**” on their resume
- Statements “I’m **gay**” and “I’m a **Jew**” were marked as toxic
- Remote exam proctoring systems incorrectly marked **disabled students** as cheating
- Healthcare risk assessment incorrectly marked **Black patients** as needing less care



# Prohibit Algorithmic Discrimination

- **Definition:**

- The term “algorithmic discrimination” refers to instances when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their actual or perceived race, color, ethnicity, sex (including based on pregnancy, childbirth, and related conditions; gender identity; intersex status; and sexual orientation), religion, age, national origin, limited English proficiency, disability, veteran status, genetic information, or any other classification protected by law. **EO 14091**



# Prohibit Algorithmic Discrimination

- **How:**
  - Private right of action (e.g.,: CA AB 331)
  - Sector-specific requirements and oversight
  - Impact assessments



---

# Impact Assessments

- Why?
  - Safety and Efficacy Protections
  - Algorithmic Discrimination Tests
  - Transparency
  - Oversight and Accountability



# Impact Assessments

- **What:**
  - Detailed, specific questions about the assessment process and results of an algorithmic system
  - Important: public consultation component
  - **Example:** Algorithmic Accountability Act of 2022
- **How:**
  - pre-release and ongoing
  - kept in private company records versus submitted to a state agency



# Transparency

- Impact assessments
- Notice – to people impacted *before* use
- Explanation – how and why was a decision made
  - such adverse action notices already required for financial decisions
- Environmental impact (kWh)
  - targeted requirement to report on the kWh used for AI



# Data-focused Interventions

- **Data Privacy Protections**
  - Data minimization
  - See, e.g.,: American Data Privacy and Protection Act of 2022 (ADPPA)
- **Intellectual Property Protections**
  - E.g., permission / contract required to use a song as part of training data



---

# Labor

- **Ensuring safety and efficacy**
  - Require human review for consequential decision systems
- **Providing human alternatives**
  - Allow people to opt-out and use a provided human alternative
- **Protecting jobs**
  - Require that AI augments, not replaces, the existing workforce





Specific recommendations



Places to start

# Recommendations

- **Don't set up a task force! Pick something specific instead.**
  - workplace surveillance limits, ban affective AI for law enforcement – **are there AI uses you think should be banned in the state?**
  - state agencies may already have relevant authorities they can use if given encouragement and resources
- **Focus on impacts, not technical details**
  - craft AI definitions that are limited based on impact
  - start with the private and public sector impacts you are most concerned with – **what are these priority areas?**
    - algorithmic discrimination, privacy
    - housing, government benefits



# Recommendations

- **Make use of the sector-specific expertise in state agencies and add (shared) technical expertise as necessary**
  - sector-specific regulation can be owned by the relevant existing agency
  - a centralized team can help agencies with technical expertise
- **Build governance across state agencies**
  - determine who is responsible for AI use/procurement by each agency
    - Chief AI Officer
  - determine how oversight and public accountability will be achieved across agencies
    - Advisory Council with public membership



# Recommendations

- **Be specific when crafting transparency requirements**
  - How is AI being used by state agencies?
    - Make a public inventory.
  - What checks are performed as part of procurement or grant funding?
    - Add specific testing, privacy, and transparency requirements to contracts.



# Resources

- White House AI Bill of Rights
  - [www.whitehouse.gov/ostp/ai-bill-of-rights](http://www.whitehouse.gov/ostp/ai-bill-of-rights)
  - “What should be expected” sections include specific actionable safeguards
  - Appendix includes examples of consequential automated systems
- American Data Privacy and Protection Act (2022)
  - bipartisan enforcement framework
- Algorithmic Accountability Act (2022)
  - useful list of specific questions to ask
- CA AB 331 Automated Decision Tools (2023)
  - consequential decision definition including specific domains
- (soon) Executive Order on AI and OMB memo



Thanks!

`sorelle@cs.haverford.edu`  
Sorelle Friedler, Haverford College