Statistics at the Al Leadership Table

Christopher Steven Marcum, Ph.D., F.G.S.A Senior Statistician and Senior Scientist

Office of the Chief Statistician of the United States
Office of Management and Budget



Al Day for Federal Statistics

Committee on National Statistics, NASEM

May 2, 2024

Outline

- Quick view into M-24-10
 - Pillars
 - Review Federal definition of AI in M-24-10
- OMB M-24-10 Pillars in Statistical Leadership Context
 - Governance
 - Innovation
 - Risk Management
- Summary



OMB M-24-10 and its Pillars



OFFICE OF THE PRESIDENT OFFICE OF MANAGEMENT AND BUDGET WASHINGTON, D.C. 20503

March 28, 2024

M-24-10

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM:

Shalanda D. Young Shalanda D. Yang

SUBJECT: Advanci

Advancing Governance, Innovation, and Risk Management for Agency Use of

Artificial Intelligence



1. STRENGTHENING AI GOVERNANCE



2. ADVANCING RESPONSIBLE AI INNOVATION



3. MANAGING RISKS FROM THE USE OF AI



Federal Definition of Artificial Intelligence

Section 238(g) of the John S. McCain National Defense Authorization Act for FY 2019

The term "artificial intelligence includes the following: An artificial system that...

- Performs tasks under varying and unpredictable circumstances without significant human oversight and can learn from experience and improve performance when exposed to data sets
- Is developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, and neural networks
- Is designed to think or act like a human, including cognitive architectures and neural networks
- Is designed to act rationally, including an intelligent software agent or embodied robot
- Includes a set of techniques, including machine learning, that is designed to approximate a cognitive task

Public Law 115-232 115th Congress

Aug. 13, 2018 [H.R. 5515]

To authorize appropriations for fiscal year 2019 for military activities of the Department of Defense, for military construction, and for defense activities of the Department of Energy, to prescribe military personnel strengths for such fiscal year,

National Defense Act for Fiscal

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

- (a) IN GENERAL.—This Act may be cited as the "John S. McCain National Defense Authorization Act for Fiscal Year 2019".
- (b) REFERENCES.—Any reference in this or any other Act to the "National Defense Authorization Act for Fiscal Year 2019" shall be deemed to be a reference to the "John S. McCain National Defense Authorization Act for Fiscal Year 2019"

SEC. 2. ORGANIZATION OF ACT INTO DIVISIONS; TABLE OF CONTENTS

- - (1) Division A—Department of Defense Authorizations.
- (2) Division B-Military Construction Authorizations. (3) Division C-Department of Energy National Security Authorizations and Other Authorizations.
- (4) Division D-Funding Tables.
- (b) TABLE OF CONTENTS.—The table of contents for this Act
- Sec. 2. Organization of Act into divisions; table of contents.
- Congressional defense committees. Budgetary effects of this Act.

DIVISION A—DEPARTMENT OF DEFENSE AUTHORIZATIONS

TITLE I-PROCUREMENT

Subtitle A-Authorization Of Appropriations

Sec. 101. Authorization of appropriations.

Subtitle B-Army Programs

Sec. 111. National Guard and reserve component equipment report.
Sec. 112. Deployment by the Army of an interim cruise missile defense capability.

Sec. 121. Procurement authority for Ford class aircraft carrier program.
Sec. 122. Full ship shock trial for Ford class aircraft carrier.
Sec. 123. Sense of Congress on accelerated production of aircraft carriers.

Multiyear procurement authority for standard missile-6.

Multiyear procurement authority for E-2D aircraft.

Multiyear procurement authority for F/A-18E/F aircraft and EA-18G air-



Office of the Chief Statistician of the United States

How Should the Statistical System Interpret the Definition?

OMB M-24-10 Guidance

This definition includes...

The AI technical subfields of machine learning, reinforcement learning, transfer learning, expert systems, agent-based systems, and generative AI and systems regardless of their...

- complexity...
- level of autonomy...
- level of human oversight



How Should the Statistical System Interpret the Definition?

OMB M-24-10 Guidance

Furthermore, this definition does not include:

- Systems whose behavior is defined only by human-defined rules or that learn solely by repeating an observed practice exactly as it was conducted
- Robotic Process Automation
- Classical statistical methods relying on human deductive reasoning for inference



OMB M-24-10 Pillars in Statistical Leadership Context



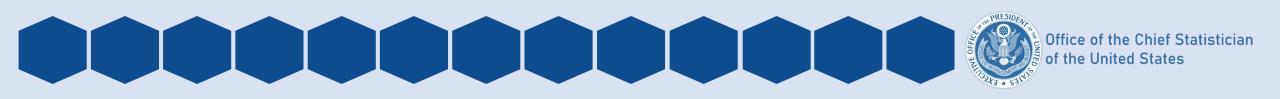
1. STRENGTHENING AI GOVERNANCE



2. ADVANCING RESPONSIBLE AI INNOVATION

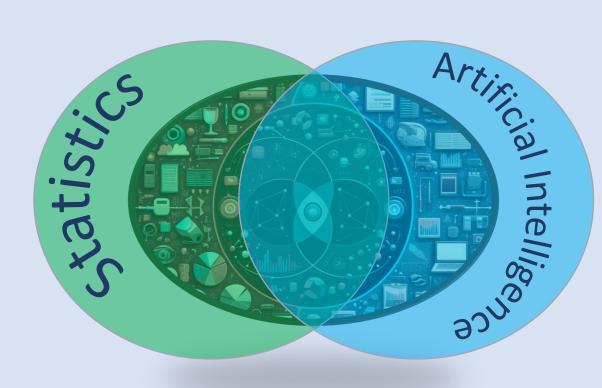


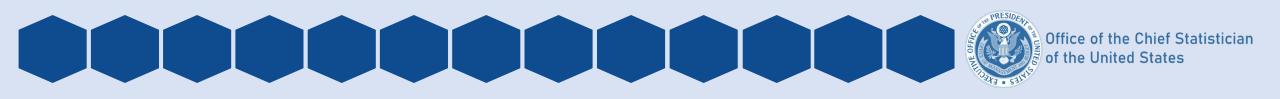
3. MANAGING RISKS FROM THE USE OF AI



Statistical Leadership in Strengthening Al Governance

- Statistical Officials may be appointed as the CAIO
- Al Governance Boards must have senior agency officials with statistical expertise
- Help agencies understand what's in the Al-use case set





Statistical Leadership in Strengthening Al Governance

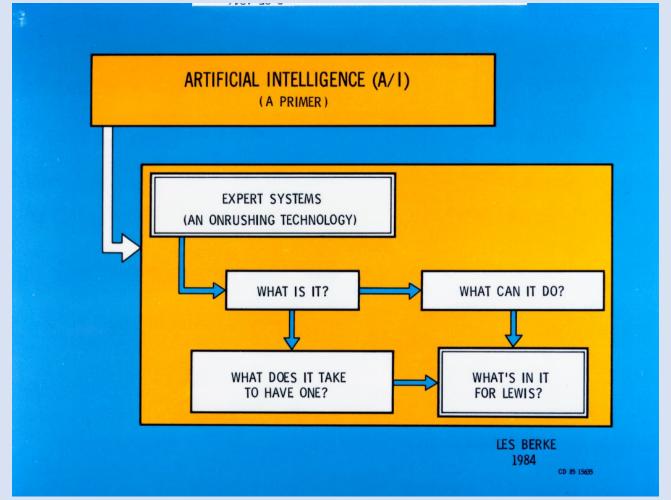


Image Credit: NASA Lewis Field via National Archives



Statistical Leadership in Responsible Al Innovation

- Recognition that statistics underlies both history and models of AI
- Economic and social monitoring of Al
- Practical uses of AI within the statistical system

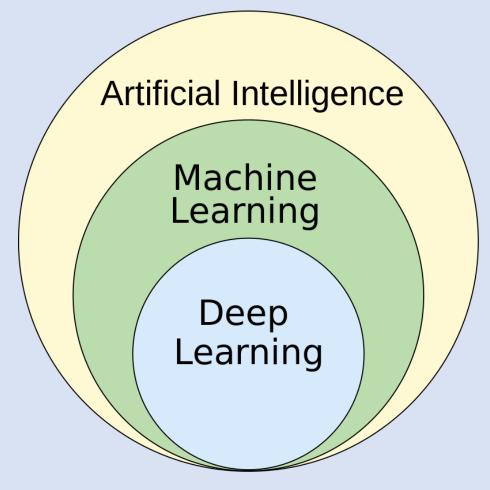


Image Credit: Lollixzc CC BY-SA 4.0



- The statistical system knows a thing or two about risk assessment (esp. disclosure risk)
- Baseline data needed for assessing risks of protected status-based harm (SOGI, race, veteran's status, etc)
- Statistical assessment of Al-systems (esp. overfitting to training data)



Image Credit: US Census

Disclosure Avoidance for the 2020

Census: An Introduction



- Threats to statistical products
- Information quality

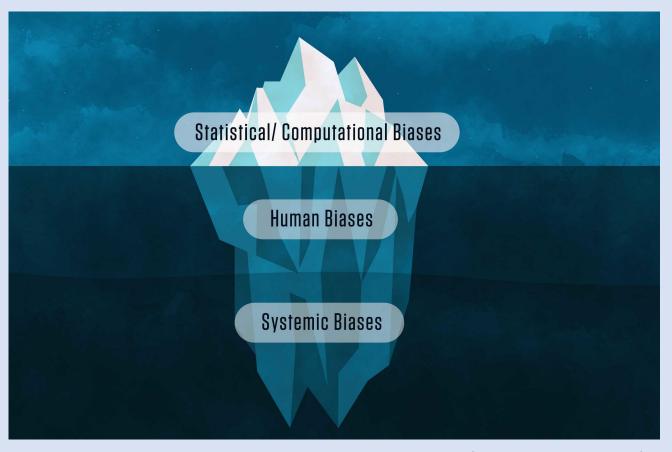
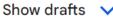


Image Source: NIST AI Risk Management Framework



simulate a sample of incomes from a pareto distribution representative of that within the US population





Here's a sample of incomes simulated from a Pareto distribution, representative of the US population:

- 1. Define Parameters:
- Population Median (M): According to the US Census Bureau https://www.census.gov/library/publications/2022/demo/p60-276.html, the median household income in 2021 was \$67,521. We'll use this as our target median.
- Shape Parameter (alpha): Let's choose alpha = 2.0. This is a common choice for income data, representing a relatively high concentration of low-to-middle

That's the 2020 estimate based on 2010 Census data!¹

The actual website cites a value of \$70,784 in 2021!²

²https://www.census.gov/library/publications/2022/demo/p60-276.html

¹ https://www.census.gov/content/dam/Census/library/publications/2022/demo/p60-276.pdf



you have cited the wrong statistic, the census says that the median household income in 2021 is \$70,784

Show drafts \



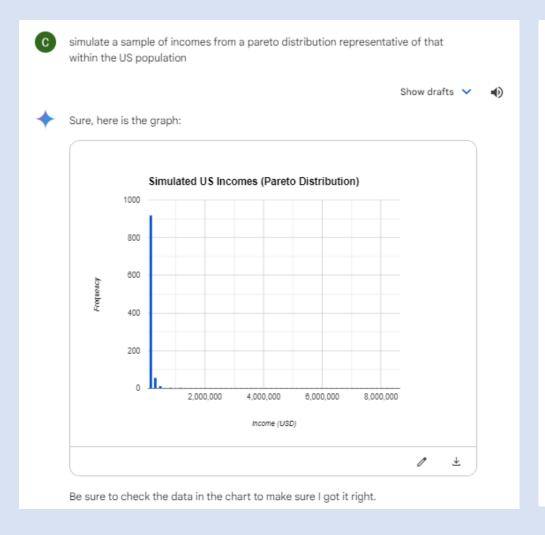
You're absolutely right, I apologize for the mistake. The median household income in the US for 2021 was \$70,784 according to the US Census Bureau

https://www.census.gov/library/publications/2022/demo/p60-276.html. Here's an

updated version of the income simulation using the corrected median:

Gemini





```
Python
                                           IT CAN BE TAUGHT!
import numpy as np
import matplotlib.pyplot as plt
# Define parameters
population median = 70784 Median household income in US (2021)
alpha = 2.0 # Shape parameter for Pareto distribution
sample size = 1000
# Generate random incomes using inverse transform sampling
u = np.random.uniform(size=sample size)
y = population median / (1 - u) ** (1 / alpha)
# Plot the distribution
plt.hist(y, bins=50, edgecolor='black')
plt.xlabel('Income (USD)')
plt.ylabel('Frequency')
plt.title('Simulated US Incomes (Pareto Distribution)')
plt.grid(True)
plt.show()
# Print a sample of 10 incomes
print("Sample of 10 incomes:")
print(y[:10])
Use code with caution.
```

Gemini

Summary

- The Federal Statistical System has a seat at the AI leadership table
- Federal definition of AI needs a statistical lens to ensure there is no scope creep
- Statistical expertise should be leveraged for responsible governance, innovation, and risk management in Al
- More use cases coming soon (some today!)





Office of the Chief Statistician of the United States



Office of the Chief Statistician of the United States

"Statistics at the Al Leadership Table" According to AI:







Thank you and Happy Al Day!

Questions?

Christopher.S.Marcum@omb.eop.gov