

Statistics at the AI 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



Office of the Chief Statistician
of the United States

AI 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



THE DIRECTOR

EXECUTIVE 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. Young*

SUBJECT: 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**



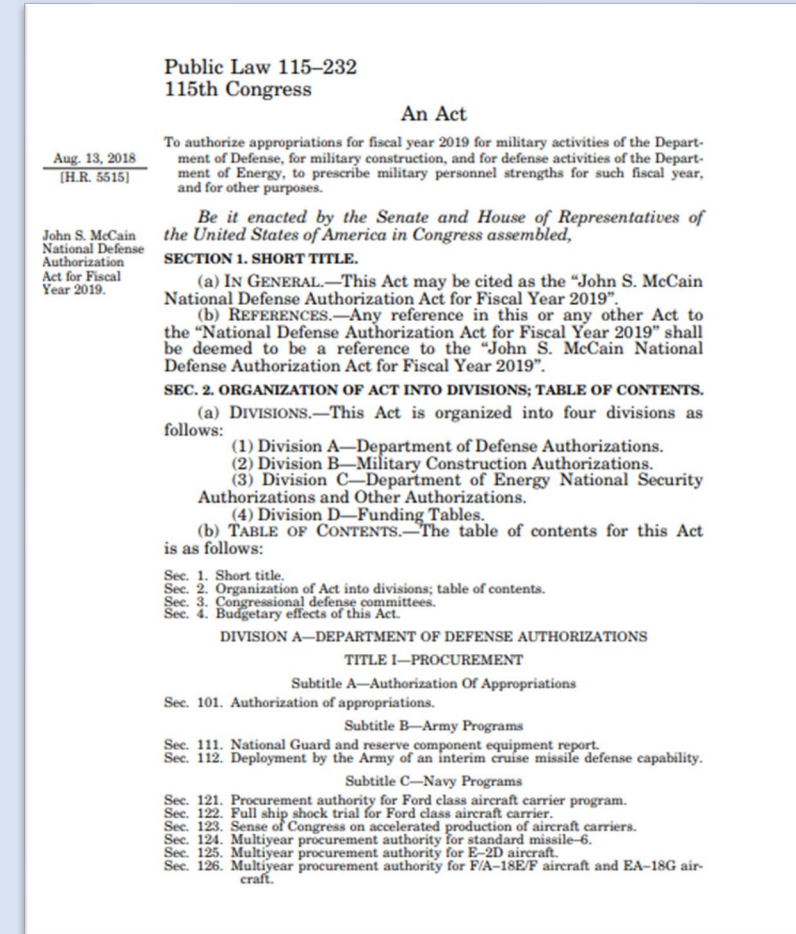
Office of the Chief Statistician
of the United States

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



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



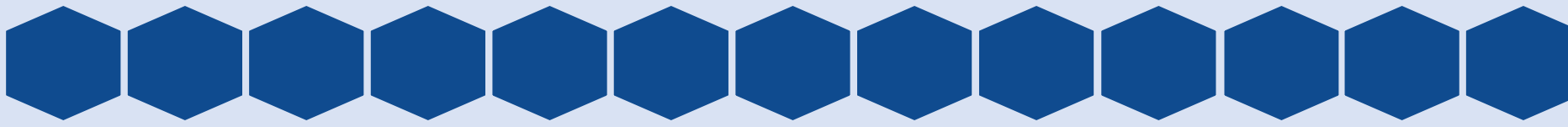
How Should the Statistical System Interpret the Definition?

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





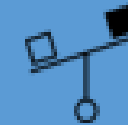
OMB M-24-10 Pillars in Statistical Leadership Context



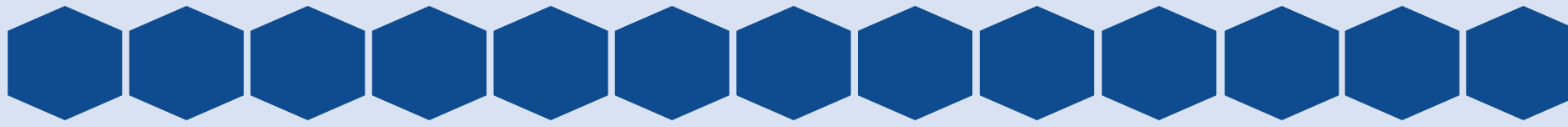
**1. STRENGTHENING AI
GOVERNANCE**



**2. ADVANCING RESPONSIBLE
AI INNOVATION**



**3. MANAGING RISKS
FROM THE USE OF AI**



Statistical Leadership in Strengthening AI Governance

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

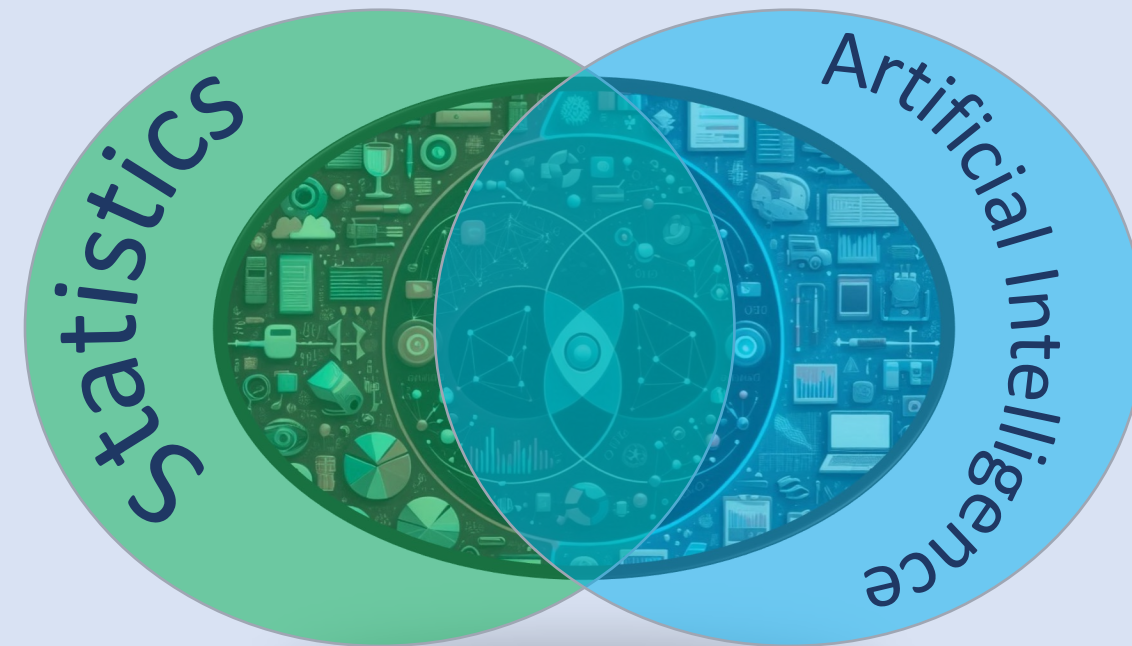
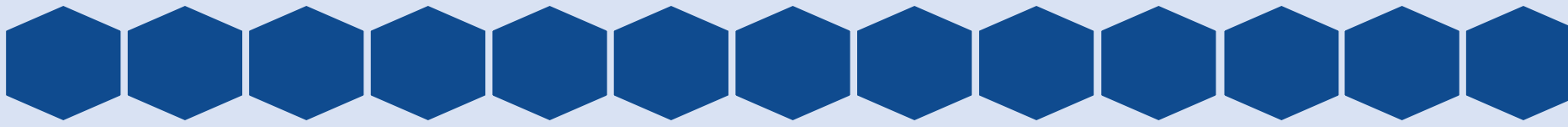
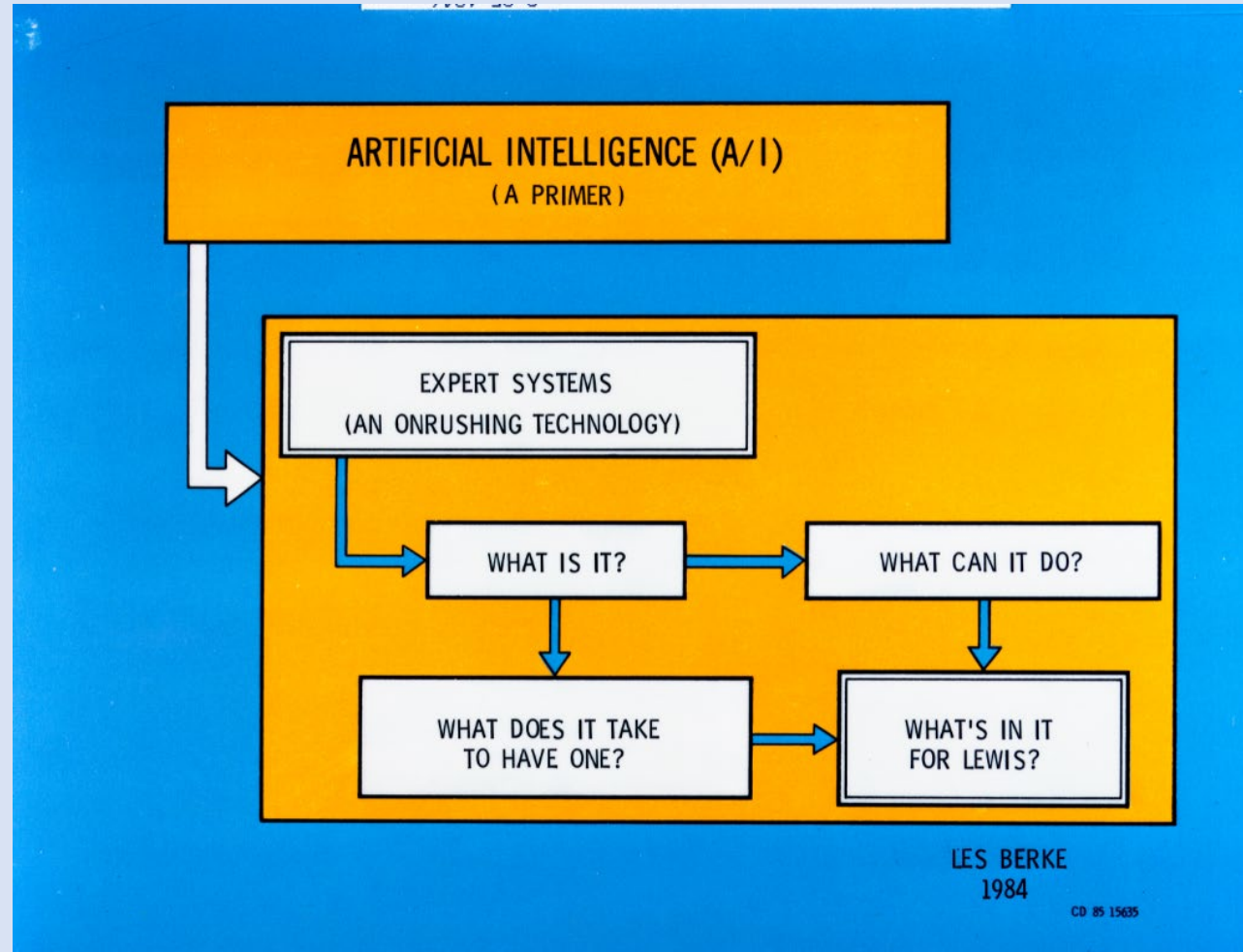


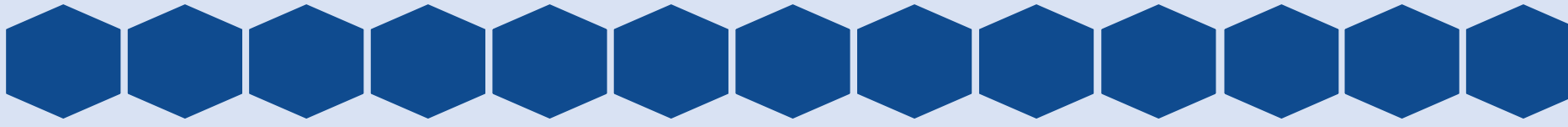
Image contains generative AI



Statistical Leadership in Strengthening AI Governance



*Image Credit: NASA Lewis
Field via National Archives*



Statistical Leadership in Responsible AI Innovation

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

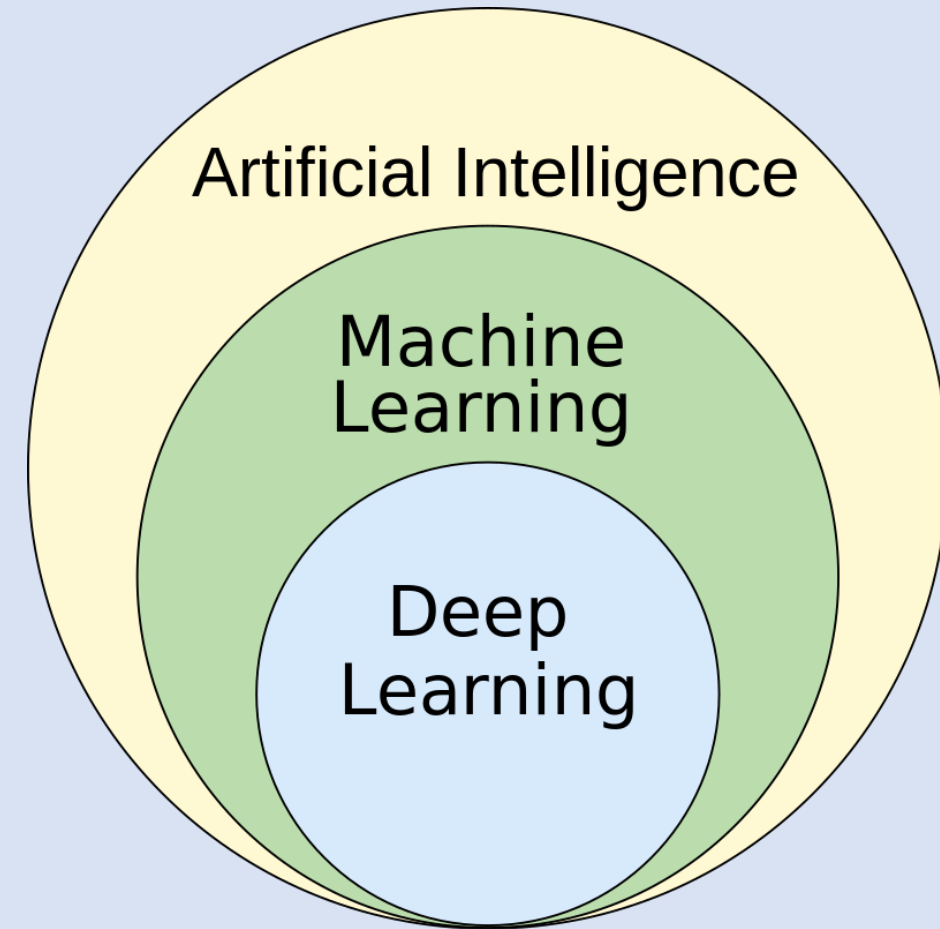
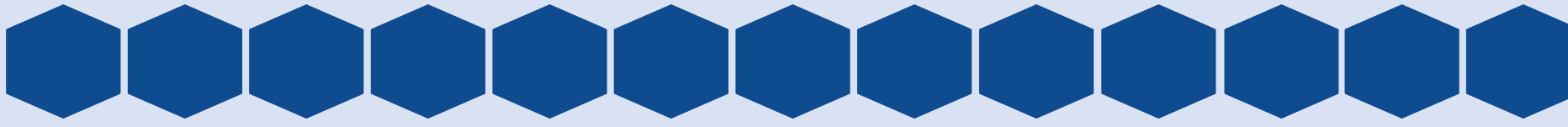


Image Credit: Lollixzc [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

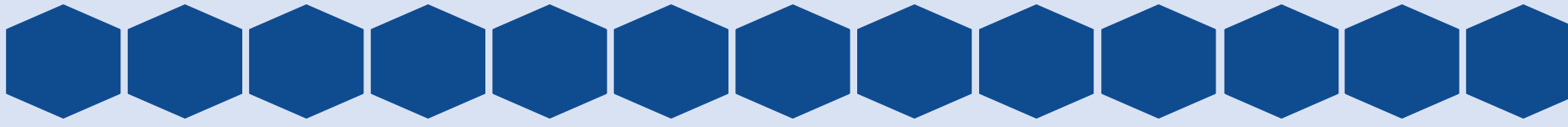


Statistical Leadership in AI Risk Management

- 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 AI-systems (esp. overfitting to training data)



Image Credit: US Census
Disclosure Avoidance for the 2020 Census: An Introduction



Statistical Leadership in AI Risk Management

- Threats to statistical products
- Information quality

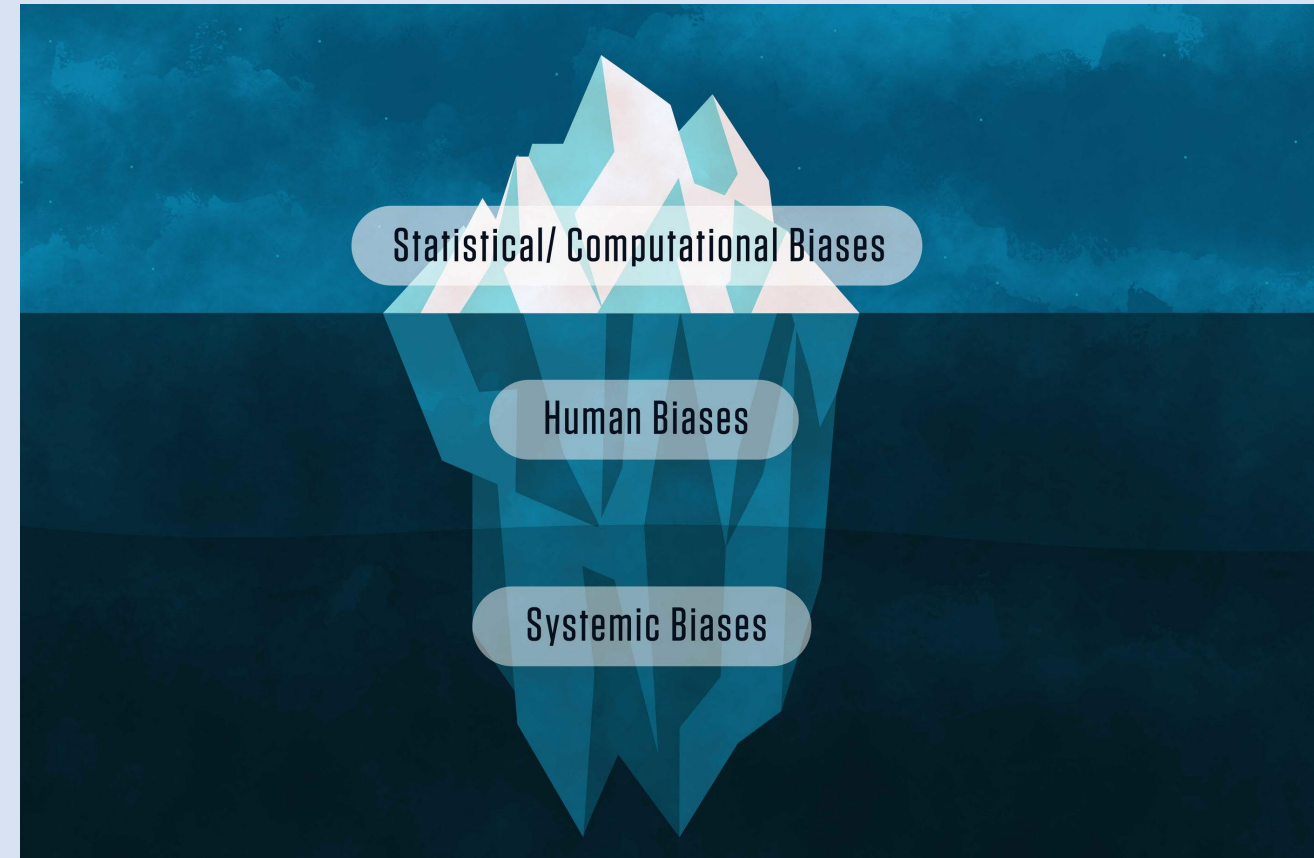
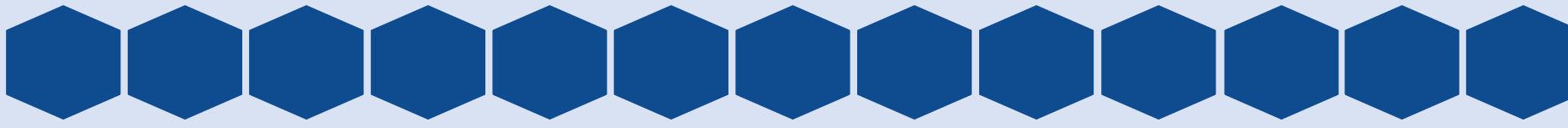


Image Source: NIST AI Risk Management Framework



Statistical Leadership in AI Risk Management



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

Show drafts



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

1. Define Parameters:

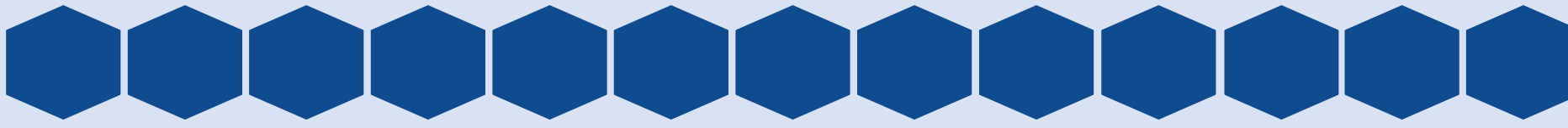
- **Population Median (M):** According to the US Census Bureau <https://www.census.gov/library/publications/2022/demo/p60-276.html>, the median household income in 2021 was \$67,521. We'll use this as our target median.
- **Shape Parameter (alpha):** Let's choose $\alpha = 2.0$. This is a common choice 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

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

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



Statistical Leadership in AI Risk Management



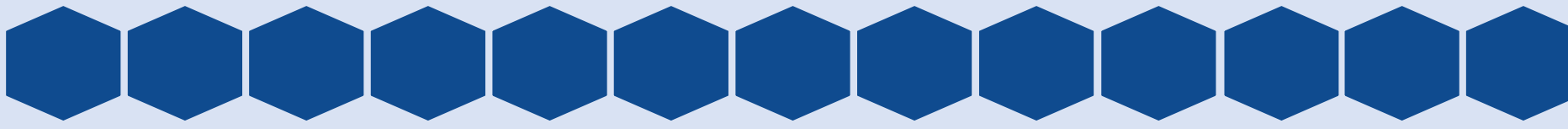
you have cited the wrong statistic, the census says that the median household income in 2021 is \$70,784

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:

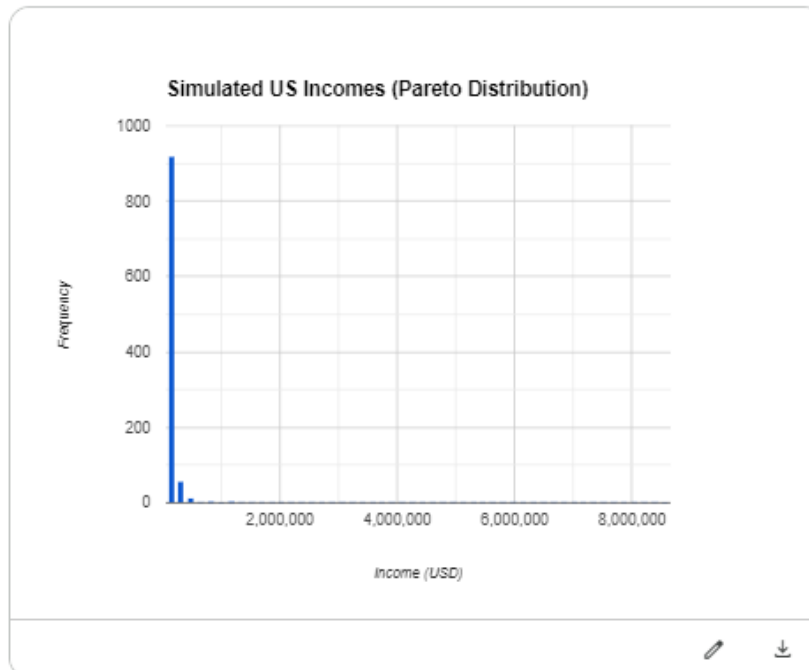


Statistical Leadership in AI Risk Management

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

Show drafts  

◆ Sure, here is the graph:



Be sure to check the data in the chart to make sure I got it right.

Python

```
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.](#)

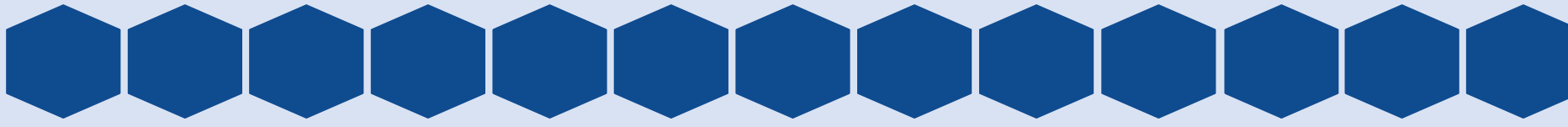
IT CAN BE TAUGHT!

Gemini

-

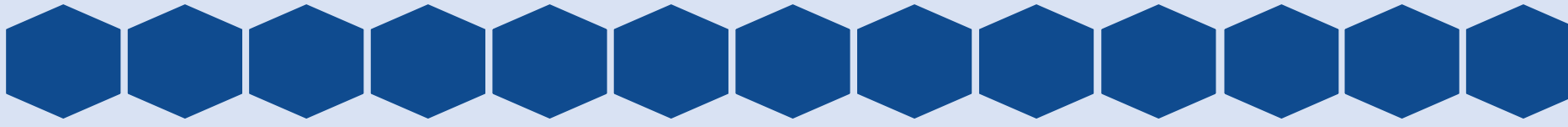


Office of the Chief Statistician
of the United States



“Statistics at the AI Leadership Table” According to AI:





Thank you and Happy AI Day!

Questions?

Christopher.S.Marcum@omb.eop.gov