NIST AI Update

Information Security Privacy Advisory Board

Elham Tabassi
Information Technology Laboratory

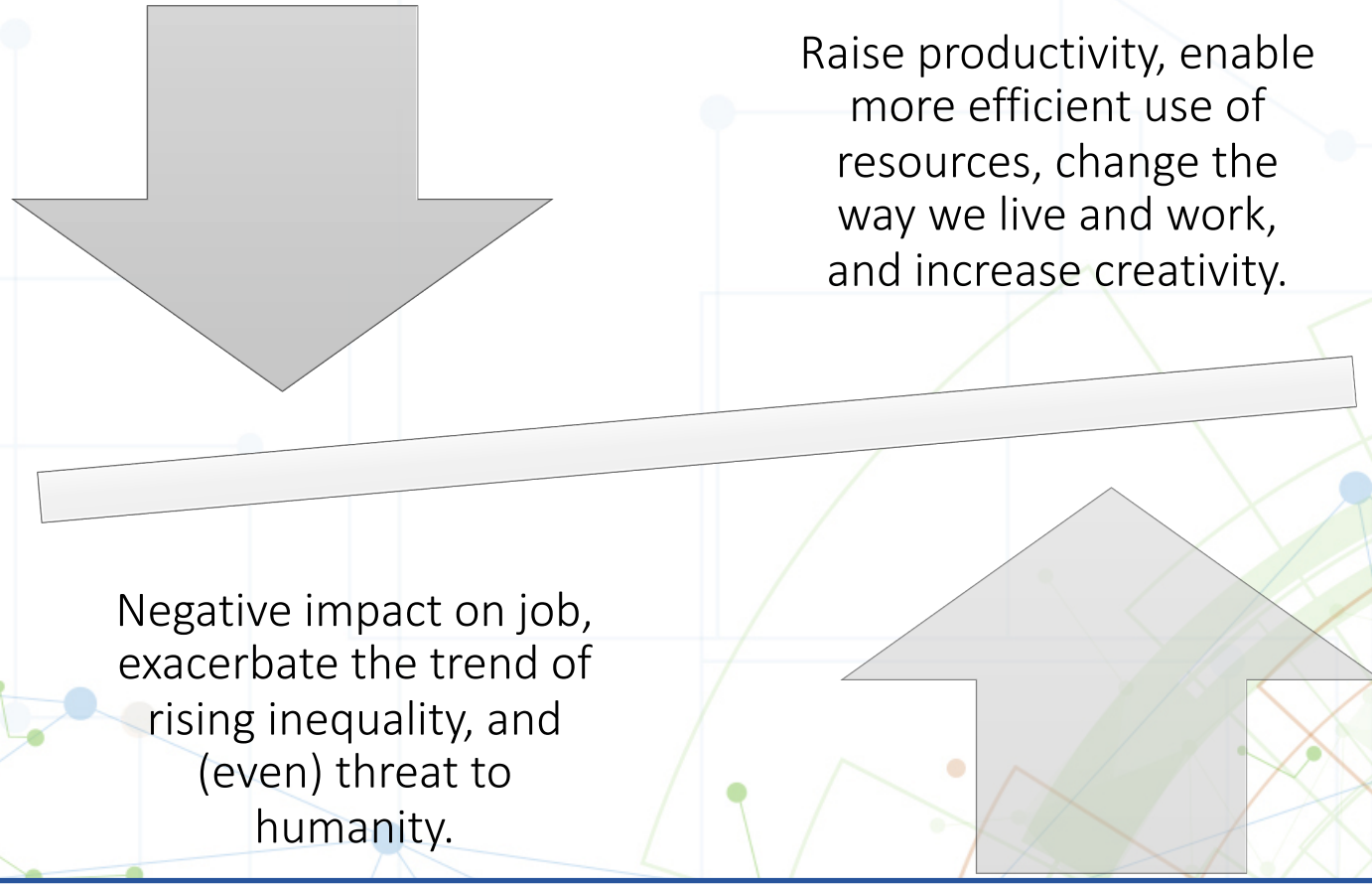
National Institute of Standards and Technology
U.S. Department of Commerce

March 4, 2021

Trustworthy AI @ NIST

National Institute of Standards and Technology
U.S. Department of Commerce

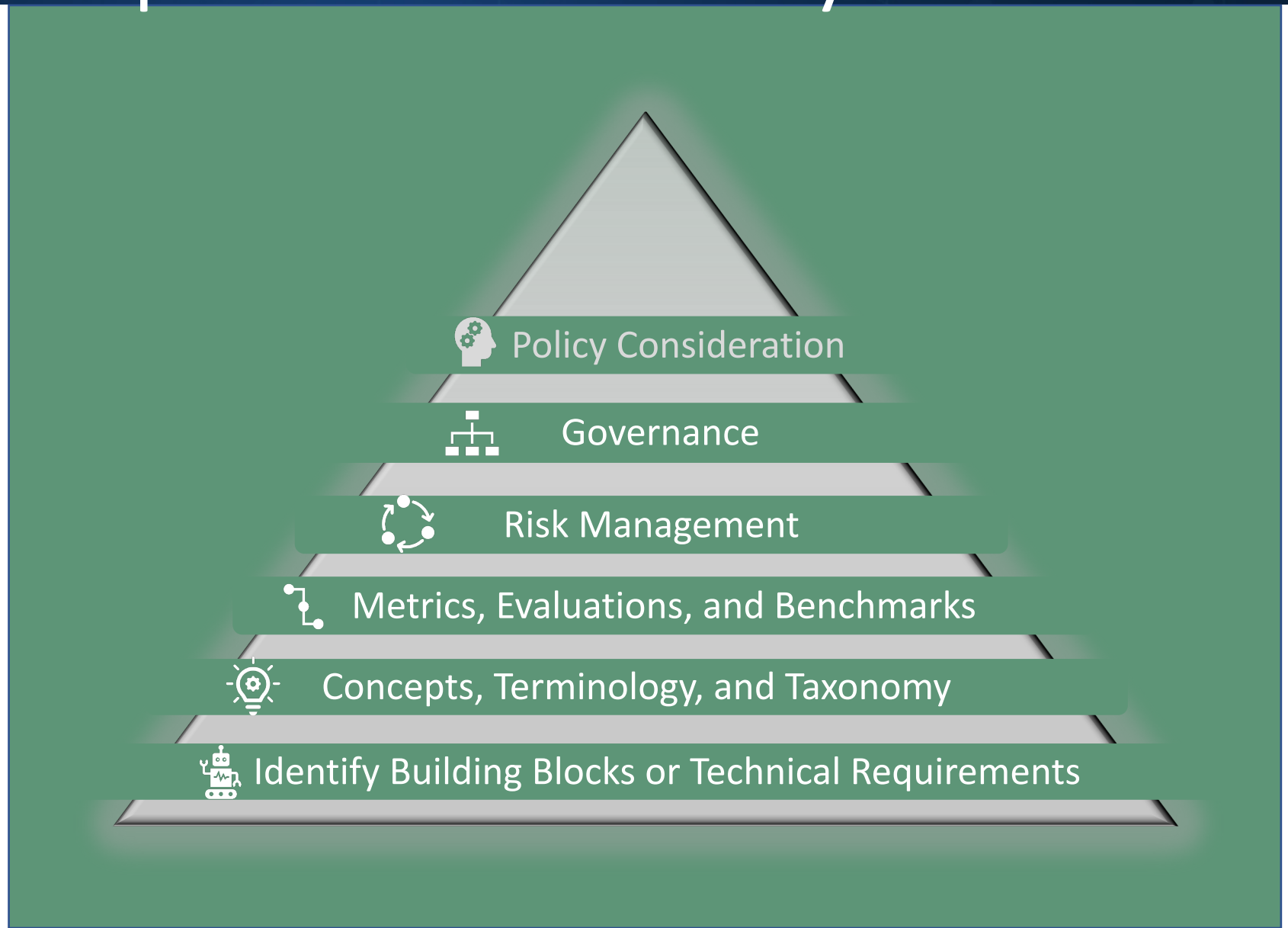
Major Advances in A.I. Continue to Drive Need for Universal Understanding of Risks



Raise productivity, enable more efficient use of resources, change the way we live and work, and increase creativity.

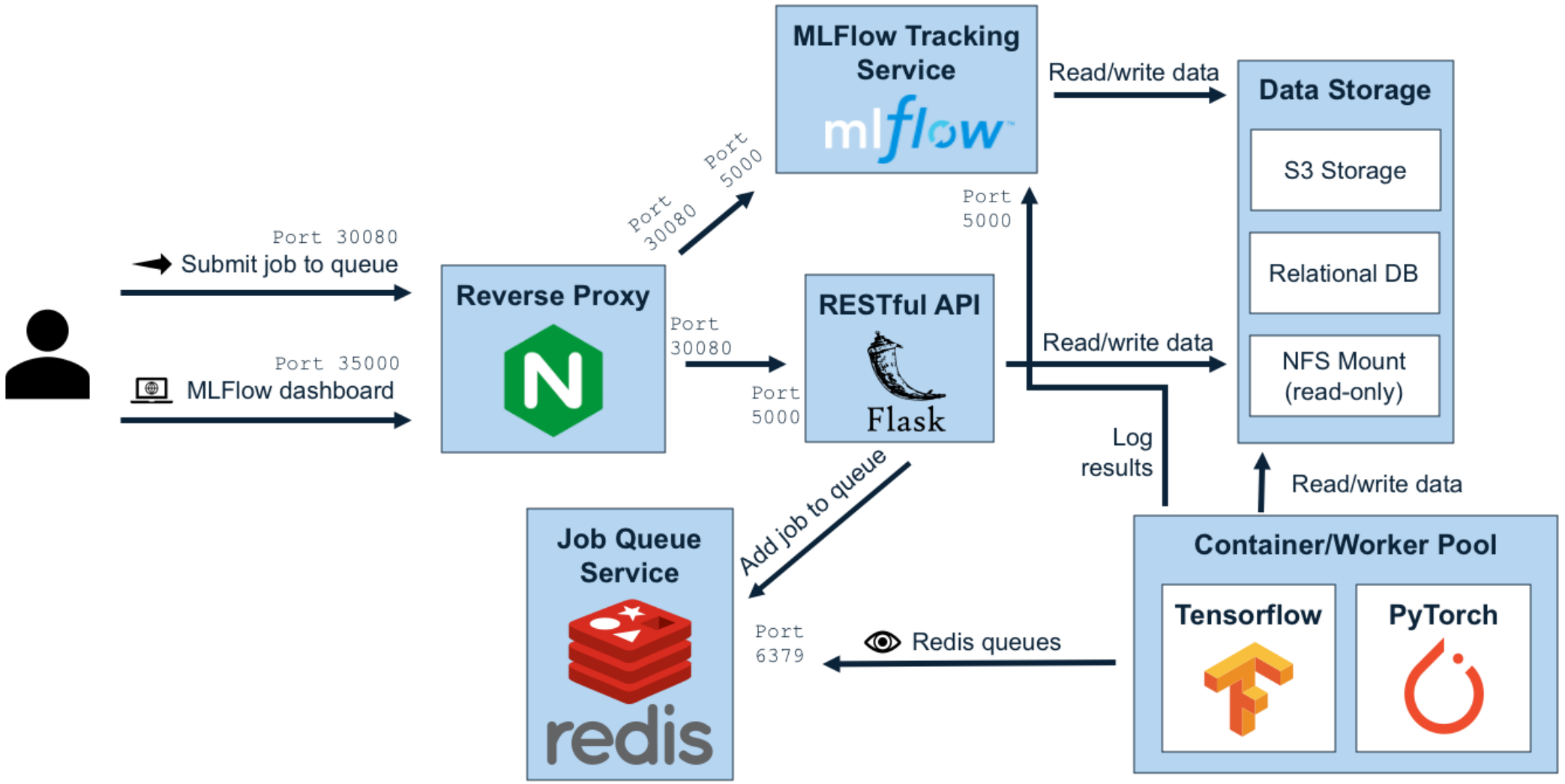
Negative impact on job, exacerbate the trend of rising inequality, and (even) threat to humanity.

Trustworthy AI's Foundation: From Technical Requirements to Policy Creation

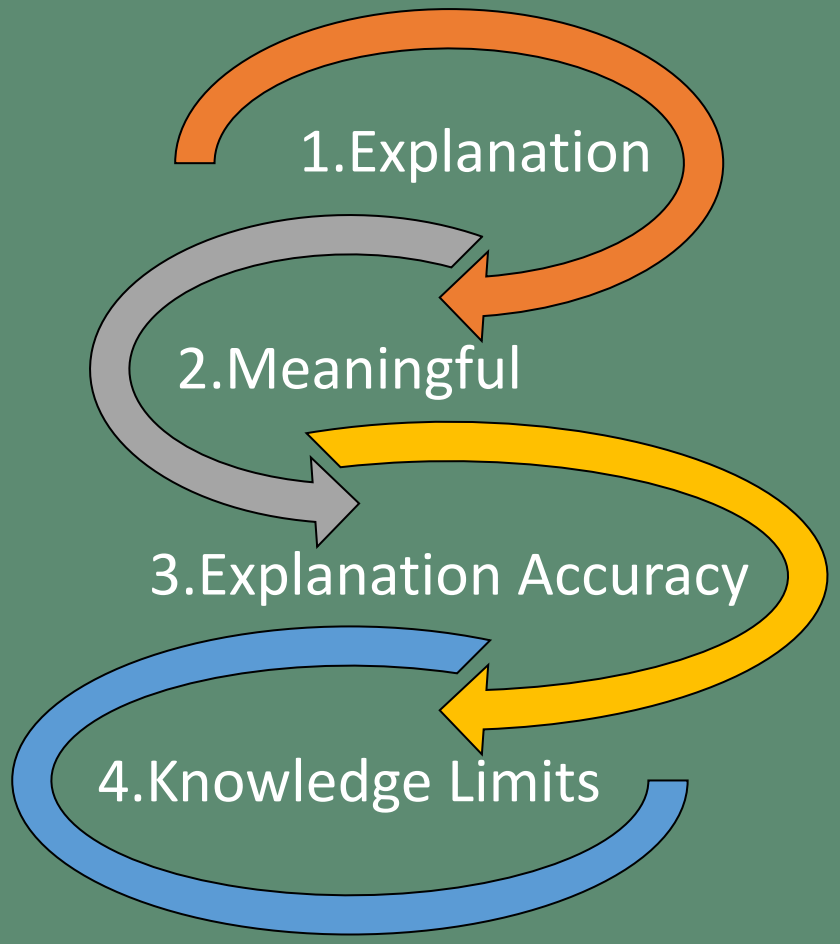


Core Building Blocks of Trustworthy AI

Secure AI Testbed



Four Principles of Explainable AI



NISTIR 8312: Four Principles of Explainable Artificial Intelligence

1. Systems deliver accompanying evidence or reason(s) for all outputs.

2. Systems provide explanations that are understandable to individual users.

3. The explanation correctly reflects the system's process for generating the output.

4. The system only operates under conditions for which it was designed or when has sufficient confidence in its output.

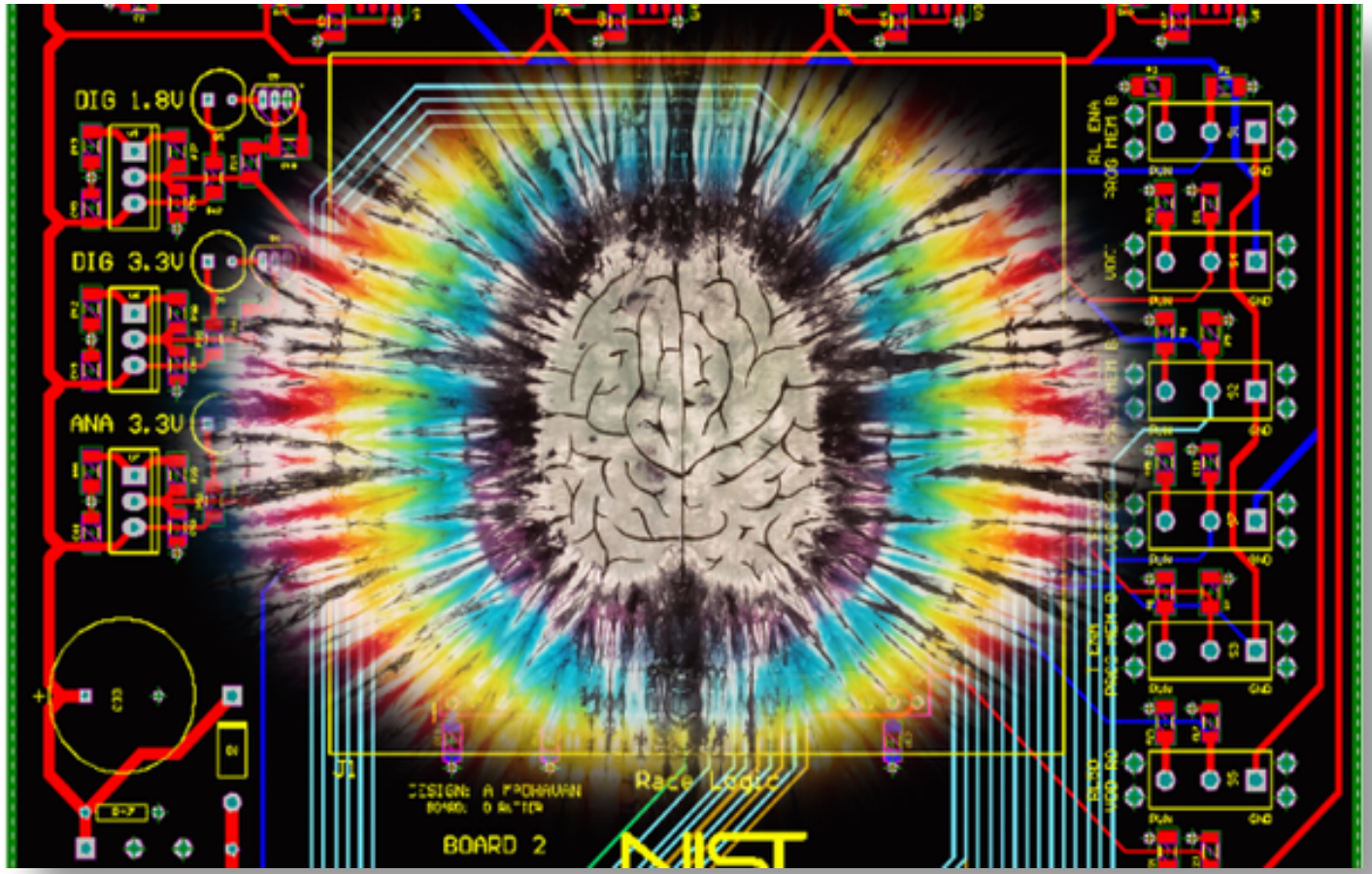
Bias in AI

Take away from the Bias in AI workshop held in August 2020

- Need Consistent Terminology
- Need Standardized Measurement of Bias
- Risk Management of Bias in AI
- Understand datasets & algorithms within their context of use cases
- Include a diverse range of scientific and other scholarly disciplines

https://www.nist.gov/news-events/events/2020/08/bias-ai-workshop

Hardware for AI



Fabricating and measuring new brain-inspired circuits and architectures based on novel devices to enable future generations of AI.

USG AI standards Coordinator

Outreach to connect with all known federal efforts relating to AI standards development and use with the goal of community participants leveraging and learning from the successes of other participants.

In collaboration with OSTP and the NSTC MLAI Subcommittee plan and execute an international campaign coordinating efforts in AI standards development

Summary of AI Provisions from the National Defense Authorization Act 2021

AI Provisions from the National Defense Authorization Act 2021

Created an executive branch entity within the Office of Science and Technology Policy to coordinate federal support for AI research and development, education and training, research infrastructure, and international engagement in order to achieve national priorities as defined in a regularly updated strategic plan for AI.

Included provisions that established a *National AI Research Resource task force*, formalized the *National AI Research Institute* effort,

assigned the National Institute of Standards and Technology with developing an *AI Risk Management Framework*.

National AI Initiative (Title LI, Sec. 5101)

National AI Initiative Office (Title LI, Sec. 5102)

The office's mission is to serve as the point of contact for Federal AI activities for Federal departments and agencies, as well as other public and private entities that may be involved in the initiative.

Coordination by Interagency AI Committee (Title LI, Sec. 5103)

National AI Advisory Committee (Title LI, Sec. 5104)

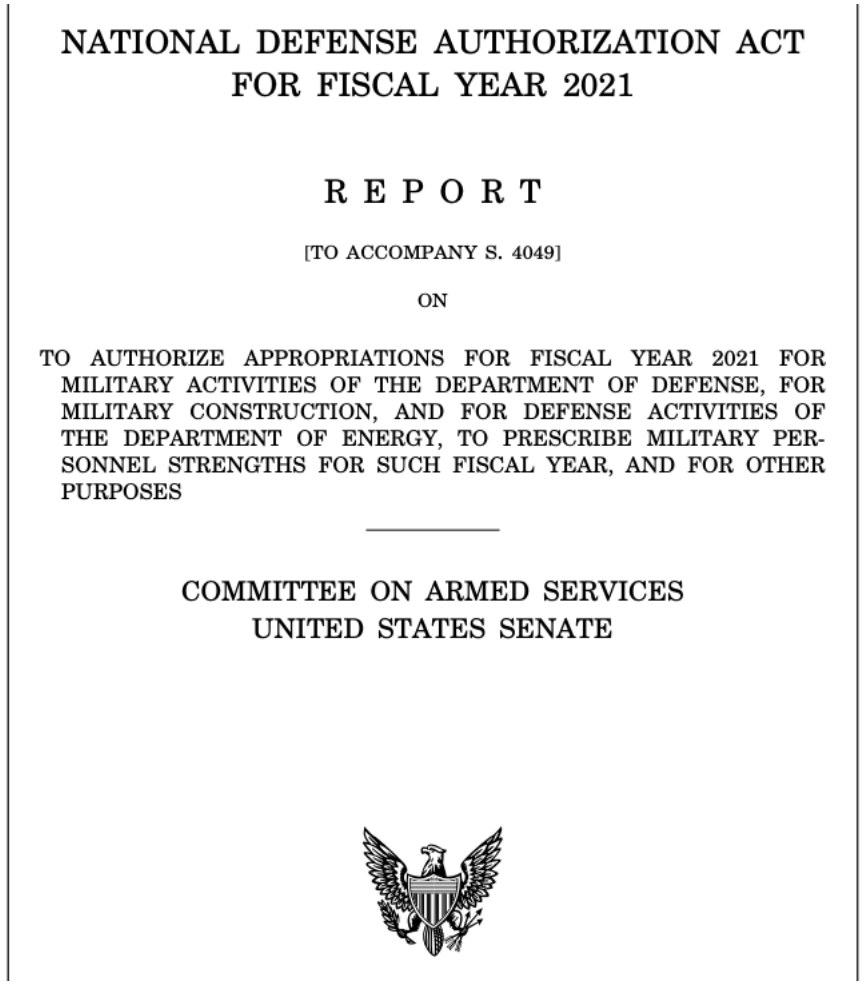
National Institute of Standards and Technology Activities

Title LIII, Sec. 5301

NIST is to expand its mission to include *advancing collaborative frameworks, standards, guidelines for AI, supporting the development of a risk-mitigation framework for AI systems*, and supporting the *development of technical standards and guidelines to promote trustworthy AI systems*.

1. Development of voluntary risk management framework for trustworthy AI systems
2. Participation in standard setting organizations
3. Data sharing best practices and best practices for documentation of data sets
4. Stakeholder outreach (Title LIII, Sec. 5302)

Other AI provisions from NDAA FY 21



NATIONAL DEFENSE AUTHORIZATION ACT
FOR FISCAL YEAR 2021

REPORT

[TO ACCOMPANY S. 4049]

ON

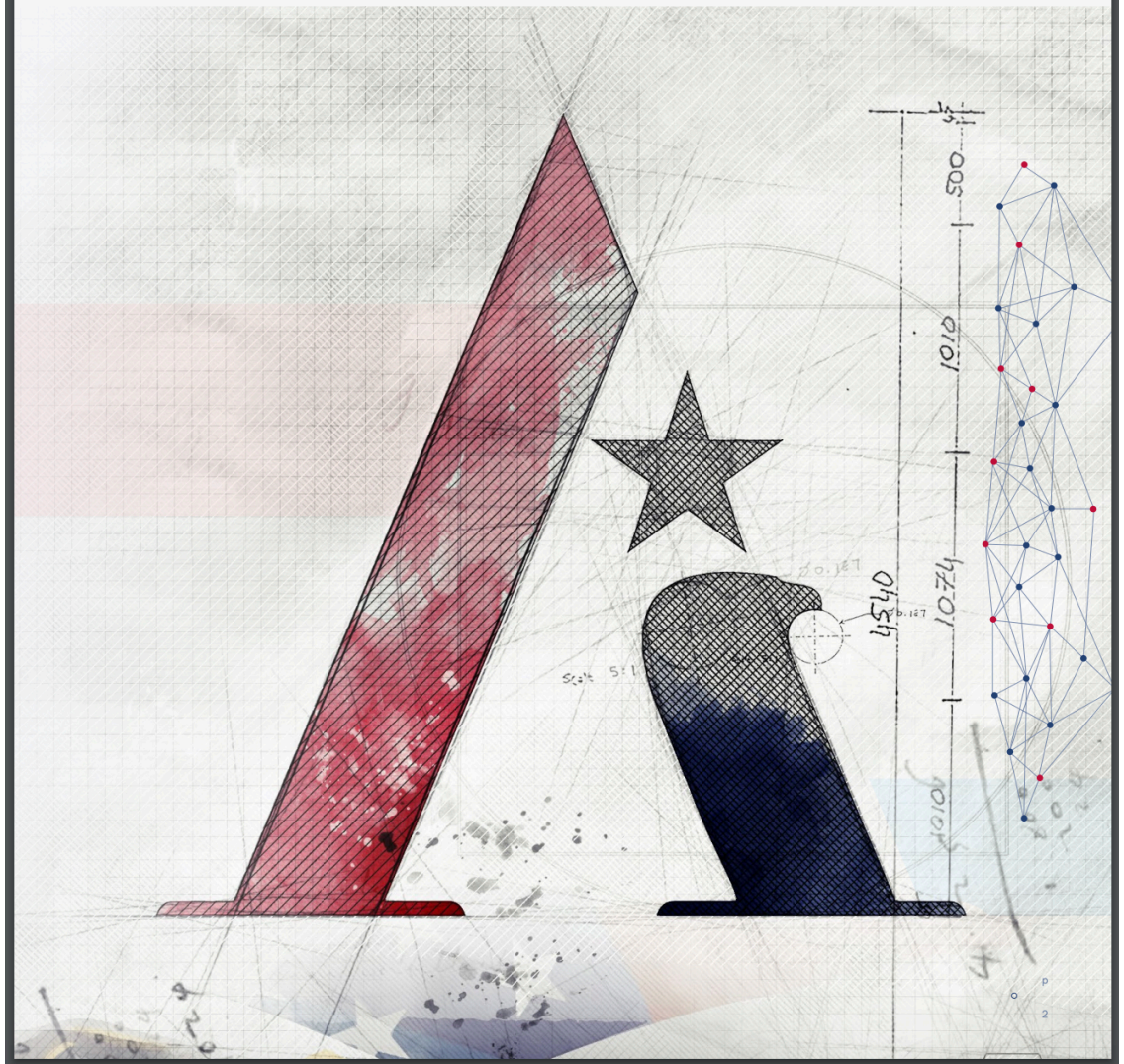
TO AUTHORIZE APPROPRIATIONS FOR FISCAL YEAR 2021 FOR MILITARY ACTIVITIES OF THE DEPARTMENT OF DEFENSE, FOR MILITARY CONSTRUCTION, AND FOR DEFENSE ACTIVITIES OF THE DEPARTMENT OF ENERGY, TO PRESCRIBE MILITARY PERSONNEL STRENGTHS FOR SUCH FISCAL YEAR, AND FOR OTHER PURPOSES

COMMITTEE ON ARMED SERVICES
UNITED STATES SENATE

- National Academies Artificial Intelligence Impact Study on Workforce (Title LI, Sec. 5105)
- National AI Research Resource Task Force (Title LI, Sec. 5106)
- National AI Research Institutes (Title LII, Sec. 5201)

Final Report

National Security Commission on Artificial Intelligence



National Security Commission on Artificial Intelligence

Mandate

Recommendations to the President and Congress to “advance the development of artificial intelligence, machine learning, and associated technologies to comprehensively address the national security and defense needs of the United States.”

Final Report

16 chapters in the Main Report provide topline conclusions and recommendations. The accompanying Blueprints for Action outline detailed steps that the U.S. Government should take to implement the recommendations.

NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

CHAPTER 7: ESTABLISHING JUSTIFIED CONFIDENCE IN AI SYSTEMS

Artificial intelligence (AI) systems must be developed and fielded with justified confidence. The recommendations cover five issue areas:

ROBUST AND RELIABLE AI

- FOCUS MORE FEDERAL RESEARCH AND DEVELOPMENT (R&D) INVESTMENTS ON ADVANCING AI SECURITY AND ROBUSTNESS.
- CONSULT INTERDISCIPLINARY GROUPS OF EXPERTS TO CONDUCT RISK ASSESSMENTS, IMPROVE DOCUMENTATION PRACTICES, AND BUILD OVERALL SYSTEM ARCHITECTURES TO LIMIT THE WORST-CASE CONSEQUENCES OF SYSTEM FAILURE.

TESTING AND EVALUATION, VERIFICATION, AND VALIDATION (TEVV)

- DOD SHOULD ADOPT A SWEEPING PACKAGE OF TESTING AND EVALUATION PROCESSES, METHODS, AND RESOURCES FOR AI SYSTEMS.
- NIST SHOULD PROVIDE A SET OF STANDARDS, PERFORMANCE METRICS, AND TOOLS FOR QUALIFIED CONFIDENCE IN AI MODELS, DATA, AND TRAINING ENVIRONMENTS, AND PREDICTED OUTCOMES.

HUMAN-AI INTERACTION AND TEAMING

- PURSUE A SUSTAINED, MULTI-DISCIPLINARY INITIATIVE THROUGH NATIONAL SECURITY RESEARCH LABS TO ENHANCE HUMAN-AI TEAMING.
- CLARIFY POLICIES ON HUMAN ROLES AND FUNCTIONS, DEVELOP DESIGNS THAT OPTIMIZE HUMAN-MACHINE INTERACTION, AND PROVIDE ONGOING AND ORGANIZATION-WIDE AI TRAINING.

LEADERSHIP

- APPOINT A FULL-TIME, SENIOR-LEVEL RESPONSIBLE AI LEAD IN EACH NATIONAL SECURITY AGENCY AND EACH BRANCH OF THE ARMED SERVICES.
- CREATE A STANDING BODY OF MULTI-DISCIPLINARY EXPERTS IN THE NATIONAL AI INITIATIVE OFFICE.

ACCOUNTABILITY AND GOVERNANCE

- ADAPT AND EXTEND EXISTING ACCOUNTABILITY POLICIES TO COVER THE FULL LIFECYCLE OF AI SYSTEMS AND THEIR COMPONENTS.
- ESTABLISH POLICIES THAT ALLOW INDIVIDUALS TO RAISE CONCERNS ABOUT IRRESPONSIBLE AI DEVELOPMENT, AND INSTITUTE COMPREHENSIVE OVERSIGHT AND ENFORCEMENT PRACTICES.

https://www.nscai.gov/wp-content/uploads/2021/01/Final-Report-Slides.pdf

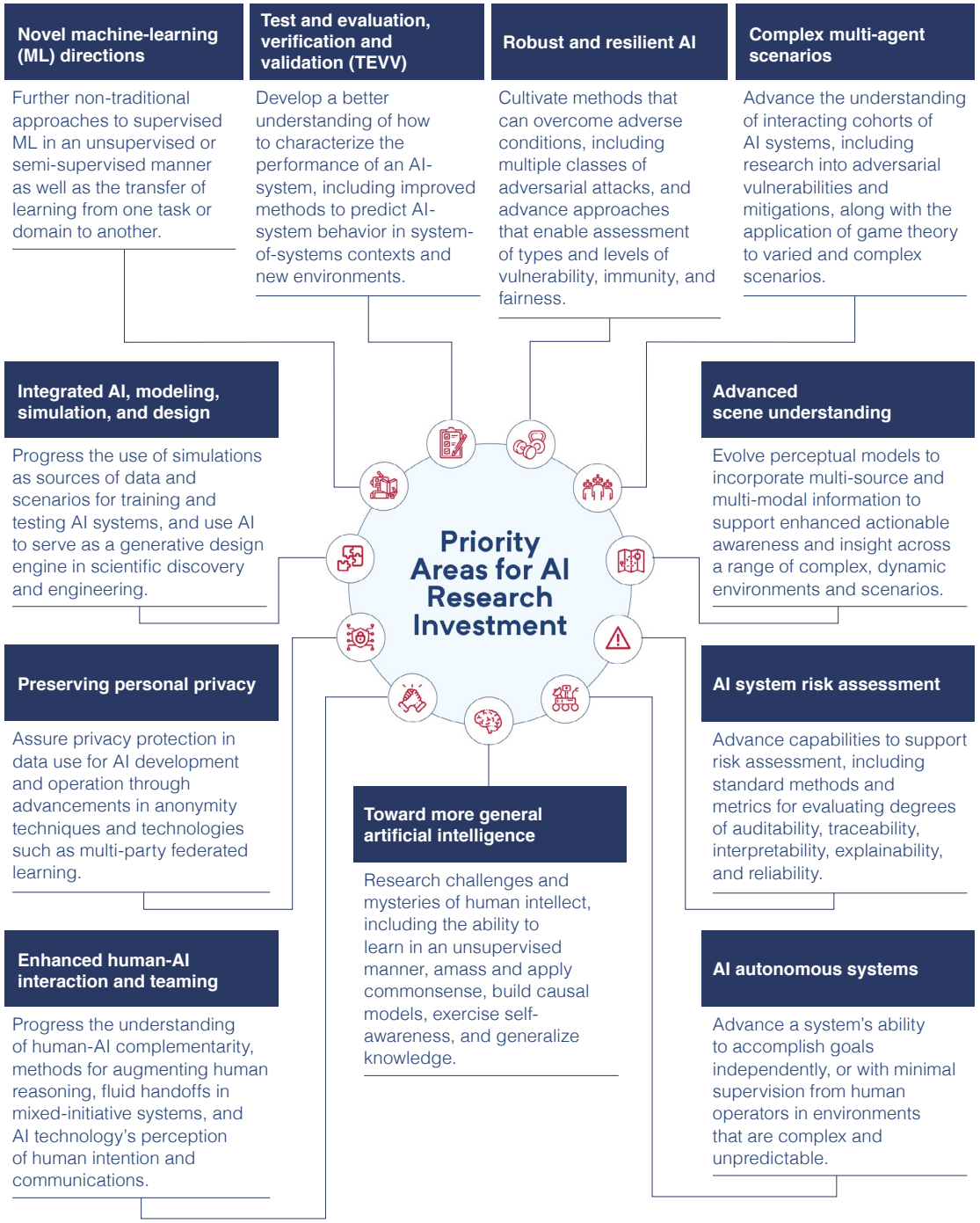
NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

CHAPTER 8: UPHOLDING DEMOCRATIC VALUES: PRIVACY, CIVIL LIBERTIES, AND CIVIL RIGHTS IN USES OF AI FOR NATIONAL SECURITY

With new models of techno-authoritarian governance gaining traction abroad, the United States must continue to serve as a beacon of democratic values.

- **INVEST IN AND ADOPT AI TOOLS TO ENHANCE OVERSIGHT AND AUDITING IN SUPPORT OF PRIVACY AND CIVIL LIBERTIES.**
- **IMPROVE PUBLIC TRANSPARENCY ABOUT HOW THE GOVERNMENT USES AI.**
- **DEVELOP AND TEST SYSTEMS WITH THE GOAL OF ADVANCING PRIVACY PRESERVATION AND FAIRNESS.**
  - Assess risks in the design, development, and testing of AI systems.
  - Identify an office, committee, or team in each agency that can conduct a pre-deployment review of AI technologies that will impact privacy, civil liberties, and civil rights.
  - Establish third-party testing centers for national security-related AI systems that could impact U.S. persons.
- **STRENGTHEN THE ABILITY OF THOSE IMPACTED BY GOVERNMENT ACTIONS INVOLVING AI TO SEEK REDRESS AND HAVE DUE PROCESS.**
  - Review DHS and FBI policies and practices that may impact due process and the ability to seek redress.
  - Issue Attorney General guidance on AI and due process.
- **STRENGTHEN OVERSIGHT MECHANISMS TO ADDRESS CURRENT AND EVOLVING CONCERNS.**
  - Establish a task force to assess the privacy and civil liberties implications of AI and emerging technologies.
  - Strengthen the ability of the Privacy and Civil Liberties Oversight Board to provide meaningful oversight and advice on AI use for national security.
  - Empower DHS Offices of Privacy and Civil Rights and Civil Liberties.
  - Require stronger coordination and alignment among federal oversight and audit organizations.

https://www.nscai.gov/wp-content/uploads/2021/01/Final-Report-Slides.pdf



Priority Areas for AI Research Investment.

Future Work to Improve Assessment of AI Systems

Page 190. NSCAI Final Report

Recommendations for NIST

1. Continue to support the development of best practices for data, model and system documentation.

2. Provide a set of standards, performance metrics, and tools for qualified confidence in AI models, data, training environments, and predicted outcomes.

3. Facilitate third-party test centers for AI systems.

National Institute of Standards and Technology
U.S. Department of Commerce

DISCUSSION