Overview of Relevant Findings and Recommendations from the National Security Commission on AI Final Report

Chuck Howell, howell@mitre.org

1 March 2021

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD™

Background and Caveat

- **Why am I talking about NSCAI?**
  - Supported since September 2019, primarily on the Responsible AI and Ethics line of effort led by Jessica Young
  - A remarkably collegial and collaborative team - opportunities to learn about and engage across the lines of effort are limited only by time and attention

Please Note:
It has been a privilege to support the NSCAI.
I am not a commissioner or government staff for the NSCAI.
I certainly don't speak for the NSCAI; I am providing my personal observations based on sustained engagement with commission.
All NSCAI content that follows is from material freely available at the NSCAI web site: www.nscai.gov

Agenda

1. Provide very high-level overview of NSCAI, the Final Report, and the resources available for more details

2. Highlight selected findings and recommendations of relevance to the Policy and Ethics of Intelligent Autonomous Systems

3. Leave time for Q&A and discussion

NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

https://www.nscai.gov/

COMMISSIONERS

DR. ERIC SCHMIDT, *CHAIRMAN*
SCHMIDT FUTURES

ROBERT O. WORK, *VICE CHAIRMAN*
TEAMWORK, LLC

GILMAN LOUIE
ALSOP-LOUIE PARTNERS
LOE 4,5

DR. ANDREW MOORE
GOOGLE
LOE 1

SAFRA CATZ
ORACLE
LOE 2

ANDY JASSY
AMAZON WEB SERVICES
LOE 2

CHRIS DARBY
IN-Q-TEL
LOE 4,5

KATHARINA MCFARLAND
NATIONAL ACADEMIES OF SCIENCE BOARD FOR ARMY R&D
LOE 2

DR. WILLIAM MARK
SRI
LOE 1,3

DR. KEN FORD
FLORIDA INSTITUTE HUMAN MACHINE COGNITION
LOE 1,2

MIGNON CLYBURN
MLC STRATEGIES, LLC
LOE 3,6

DR. STEVE CHIEN
JET PROPULSION LAB
LOE 1,2

DR. JOSE-MARIE GRIFFITHS
DAKOTA STATE UNIVERSITY
LOE 3,6

DR. JASON MATHENY
GEORGETOWN UNIVERSITY
LOE 4,5,6

DR. ERIC HORVITZ
MICROSOFT RESEARCH
LOE 1,6

CONGRESSIONAL MANDATE

Section 1051 of the Fiscal Year 2019 National Defense Authorization Act (NDAA) established the National Security Commission on Artificial Intelligence as an independent Commission to "consider the methods and means necessary to advance the development of artificial intelligence, machine learning, and associated technologies by the United States to comprehensively address the national security and defense needs of the United States."

LINES OF EFFORT

1. INVEST IN AI RESEARCH & DEVELOPMENT AND SOFTWARE
2. APPLY AI TO NATIONAL SECURITY MISSIONS
3. TRAIN AND RECRUIT AI TALENT
4. PROTECT & BUILD UPON AI TECH ADVANTAGES AND HARDWARE
5. MARSHAL GLOBAL AI COOPERATION
6. ETHICAL AND RESPONSIBLE AI

FINAL REPORT DUE MARCH 1 2021

National Security Commission on Artificial Intelligence Final Report

NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

Part I: Defending America on the AI Era

1. Emerging Threats in the AI Era
2. Foundations of Future Defense
3. AI And Warfare
4. **Autonomous Weapon Systems & Risks Associated with AI-Enabled Warfare**
5. AI and the Future of National Intelligence
6. Technical Talent in Government
7. **Establishing Justified Confidence in AI Systems**
8. **Upholding Democratic Values**

Part II: Winning the Technology Competition

9. A Strategy for Competition and Cooperation
10. The Talent Competition
11. Accelerating AI Innovation
12. Intellectual Property
13. Microelectronics
14. Technology Protection
15. A Favorable International Technology Order
16. Associated Technologies

Strategy Document

FINAL REPORT CHAPTERS + BLUEPRINTS FOR ACTION + APPENDICES

- Draft Legislative Language
- Draft Executive Orders
- Technical Glossary
- **Key Considerations**
- Funding Table

Key Considerations for the Responsible Development & Fielding of AI

- Recommended practices for policymakers and technical practitioners covering the AI lifecycle that should be updated as the technology advances.
- Recommendations for future action, integrated into the Commission's final report.

As national security organizations develop and field AI systems, they should consider:

(1) the encoding or alignment of human values within AI systems
(2) responsible engineering practices
(3) system performance or capabilities
(4) designs and effectiveness for human-AI interaction
(5) accountability and governance

An abridged version of the Key Considerations is an appendix in the Final Report. The full Key Considerations document with more detail and references will be available at www.nscai.gov

NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

CHAPTER 4: AUTONOMOUS WEAPON SYSTEMS & RISKS ASSOCIATED WITH AI-ENABLED WARFARE

Properly designed, tested, and utilized AI-enabled and autonomous weapons systems will bring substantial military and potentially humanitarian benefit, and existing, rigorous DoD procedures are capable of ensuring such systems are developed and used in a safe, legal, and ethical manner.

Countries must also take actions to reduce potential risks associated with such systems, including by promoting their responsible development and use, fostering communication between competitors about their impact on stability, and forswearing certain specific capabilities. The United States must lead such efforts.

JUDGMENTS

1. **Provided their use is authorized by a human commander or operator, properly designed and tested AI-enabled and autonomous weapons systems have been and can continue to be used in ways which are consistent with IHL.**

2. **Existing DoD procedures are capable of ensuring that the United States will field safe and reliable AI-enabled and autonomous systems and use them in a manner that is consistent with IHL.**

3. **There is little evidence that U.S. competitors have equivalent rigorous procedures to ensure their AI-enabled and autonomous weapons systems will be responsibly designed and lawfully used.**

4. **The Commission does not support a global prohibition of AI-enabled and autonomous weapons systems.**

RECOMMENDATIONS

- **Clearly and publicly affirm existing U.S. policy that only human beings can authorize employment of nuclear weapons and seek similar commitments from Russia and China.**

- **Discuss AI's impact on crisis stability in the existing U.S.-Russia Strategic Security Dialogue and create an equivalent meaningful dialogue with China.**

- **Work with allies to develop international standards of practice for the development, testing, and use of AI-enabled and autonomous weapons systems.**

- **Pursue technical means to verify compliance with future arms control agreements pertaining to AI-enabled weapons systems.**

- **Fund research on technical means to prevent proliferation of AI-enabled and autonomous weapons systems.**

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.

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.

Research and Development Funding Recommendation

- Increasing federal funding for non-defense AI R&D at compounding levels, doubling annually to reach $32 billion per year by Fiscal Year 2026; and that

- The Office of Science and Technology Policy should balance interagency AI R&D investment portfolios and prioritize critical AI research areas. With a lens for responsible development and fielding, these include an emphasis on (subset selected by me):
    - TEVV
    - Robust and resilient AI
    - Complex multi-agent scenarios
    - AI system risk assessment
    - Human-AI interaction and teaming

Summary

- **NSCAI positioned responsible AI as a fundamental enabler of AI adoption**
- **There are extensive procedures, policies, and experiences in responsible technology adoption and compliance with ethical and legal standards**
  - For AI-enabled systems, some can be adopted as-is, some can be adapted and extended, some technology and policy gaps exist and must be closed
  - AI in consequential systems will always involve decisions about risk tradeoffs; those decisions are better if explicit and deliberate vs. implicit and ad hoc
- **The NSCAI web site is a significant resource for responsible AI adoption**
  - The final report contains high-level findings, top-line recommendations, more detailed blueprints for implementation, abridged Key Consideration, lots of footnotes and references
  - Full Key considerations, video of plenary sessions, podcasts w/ commissioners

From the Introduction to the Final Report

Letter from the Executive Director

“We found consensus among nearly all our partners on three points:

- The conviction that AI is an enormously powerful technology
- Acknowledgement of the urgency to invest more in AI innovation
- Responsibility to develop and use AI guided by democratic principles.”

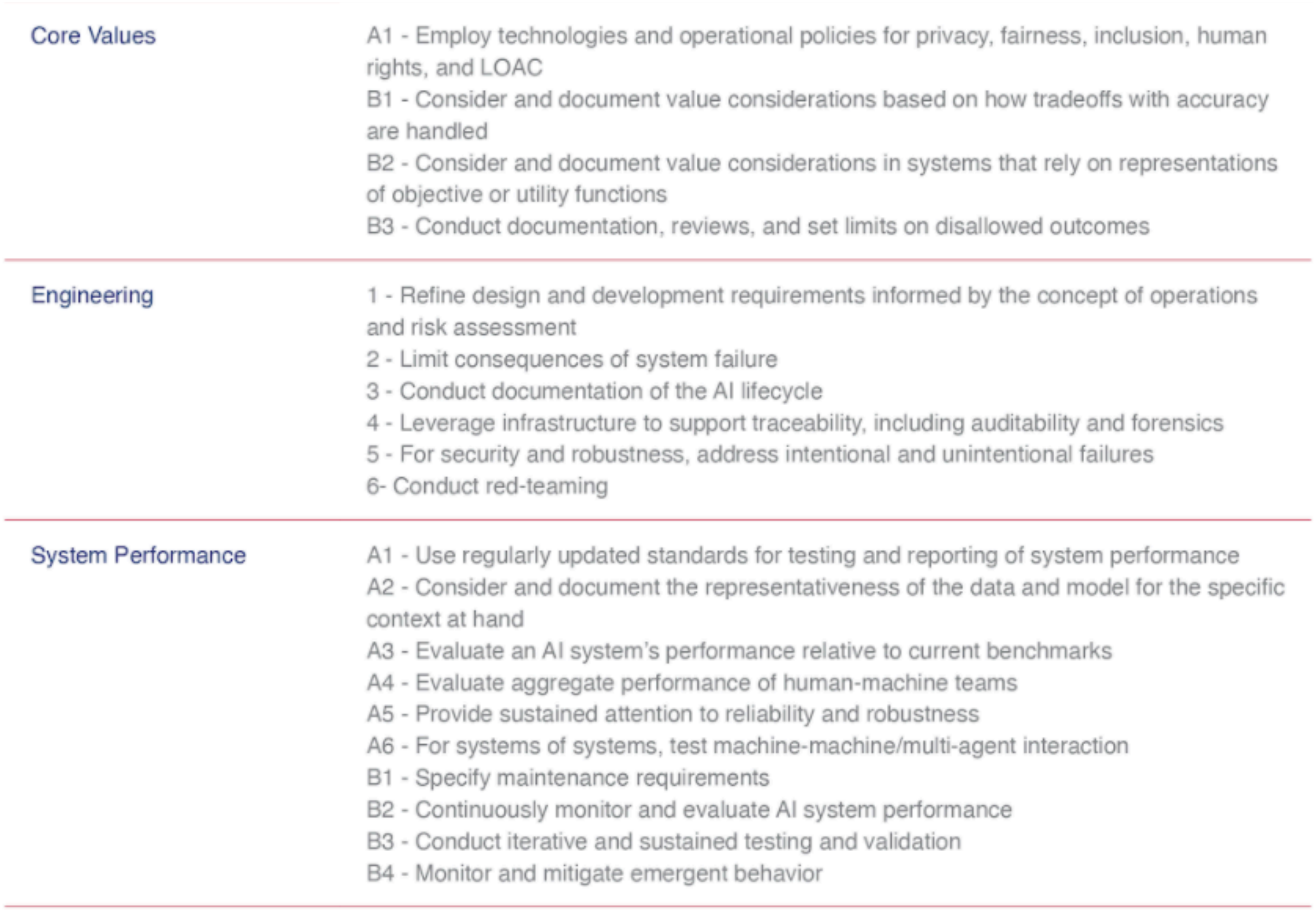
Thank you

Questions?

Backup content

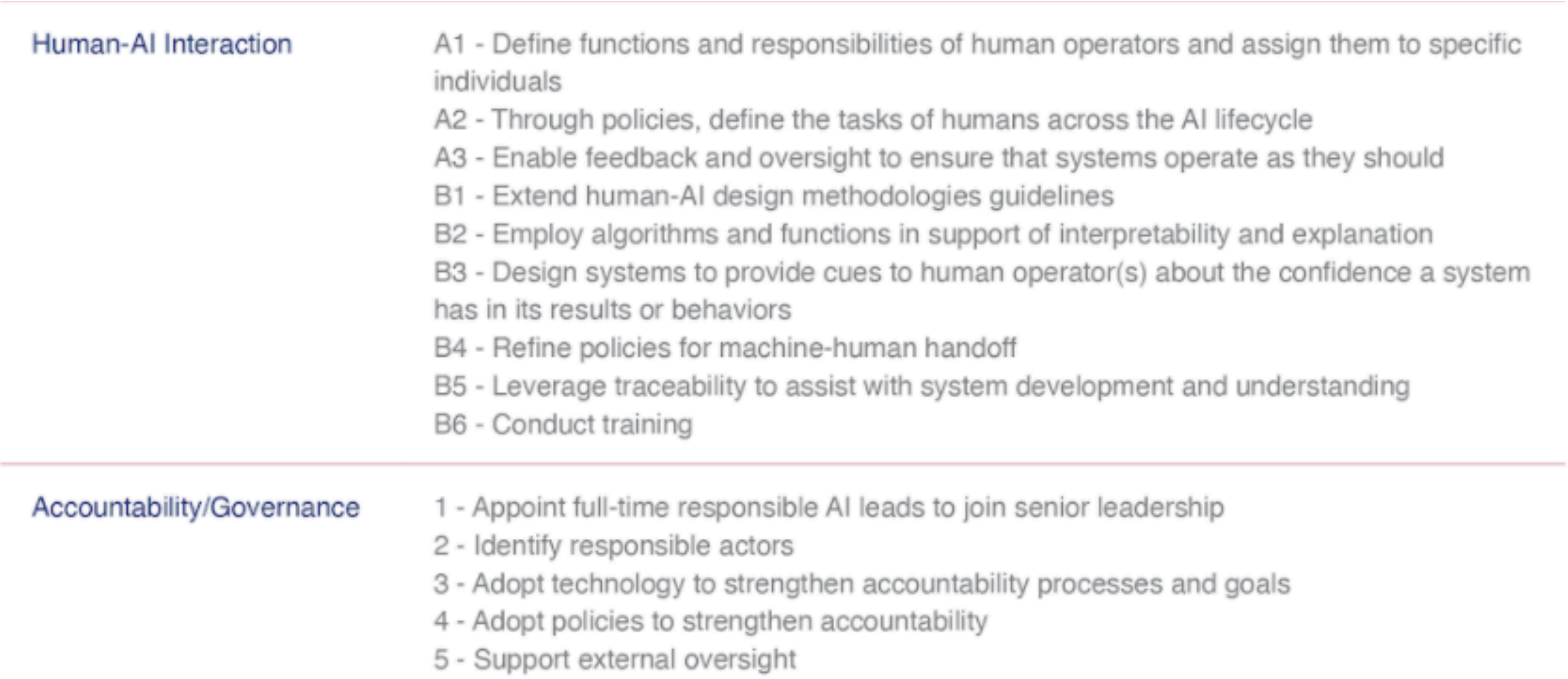
NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

Key Considerations for the Responsible Development & Fielding of AI



Core Values	A1 - Employ technologies and operational policies for privacy, fairness, inclusion, human rights, and LOAC B1 - Consider and document value considerations based on how tradeoffs with accuracy are handled B2 - Consider and document value considerations in systems that rely on representations of objective or utility functions B3 - Conduct documentation, reviews, and set limits on disallowed outcomes
Engineering	1 - Refine design and development requirements informed by the concept of operations and risk assessment 2 - Limit consequences of system failure 3 - Conduct documentation of the AI lifecycle 4 - Leverage infrastructure to support traceability, including auditability and forensics 5 - For security and robustness, address intentional and unintentional failures 6- Conduct red-teaming
System Performance	A1 - Use regularly updated standards for testing and reporting of system performance A2 - Consider and document the representativeness of the data and model for the specific context at hand A3 - Evaluate an AI system's performance relative to current benchmarks A4 - Evaluate aggregate performance of human-machine teams A5 - Provide sustained attention to reliability and robustness A6 - For systems of systems, test machine-machine/multi-agent interaction B1 - Specify maintenance requirements B2 - Continuously monitor and evaluate AI system performance B3 - Conduct iterative and sustained testing and validation B4 - Monitor and mitigate emergent behavior

Key Considerations for the Responsible Development & Fielding of AI



Human-AI Interaction	A1 - Define functions and responsibilities of human operators and assign them to specific individuals A2 - Through policies, define the tasks of humans across the AI lifecycle A3 - Enable feedback and oversight to ensure that systems operate as they should B1 - Extend human-AI design methodologies guidelines B2 - Employ algorithms and functions in support of interpretability and explanation B3 - Design systems to provide cues to human operator(s) about the confidence a system has in its results or behaviors B4 - Refine policies for machine-human handoff B5 - Leverage traceability to assist with system development and understanding B6 - Conduct training
Accountability/Governance	1 - Appoint full-time responsible AI leads to join senior leadership 2 - Identify responsible actors 3 - Adopt technology to strengthen accountability processes and goals 4 - Adopt policies to strengthen accountability 5 - Support external oversight

NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

NSCAI Recommended Practices:		DOD PRINCIPLES OF AI ETHICS				
		Responsible	Equitable	Traceable	Reliable	Governable
Core Values	A1	X	X			X
	B1	X	X	X	X	
	B2	X	X	X	X	
	B3	X	X	X	X	X
Engineering	1	X	X	X	X	X
	2			X		
	3		X	X		
	4				X	X
	5				X	
System Performance	A1	X	X	X	X	
	A2	X	X	X	X	
	A3			X		X
	A4	X				
	A5	X		X	X	
	A6	X			X	
	B1	X	X	X	X	
	B2	X	X	X	X	X
	B3	X			X	X
	B4	X			X	X
Human-AI Interaction	A1	X		X		
	A2	X				
	A3	X			X	
	B1	X	X	X		X
	B2	X		X		X
	B3	X		X		X
	B4	X		X		X
	B5	X		X	X	X
	B6	X	X	X	X	X
Accountability & Governance	1	X		X		X
	2	X		X		X
	3	X		X		
	4			X		