MEETING LOG

SUBECT: CPSC Artificial Intelligence and Machine Learning Forum

LOCATION: Webinar on GoToWebinar (*Virtual*)

DATE: March 2, 2021 9:00AM – 4:00PM

ENTRY DATE: April 7, 2021 submitted

LOG ENTRY SOURCE: Nevin Taylor

ATTENDEES: Contact Nevin Taylor at ntaylor@cpsc.gov for a list of attendees.

MEETING SUMMARY: The U.S. Consumer Product Safety Commission (*CPSC*) staff hosted a webinar forum on Artificial Intelligence (*AI*) and Machine Learning (*ML*) technology and consumer product safety. The objective of the event was to exchange information pertaining to the safety of consumer products that use AI and ML technology. The forum consisted of discussions on AI and ML policy, standards, application, assessments, and safety.

The program included two opening presentations, four panel presentations, and a roundtable exchange of information (*see Appendix A: Agenda and Appendix B: Presentations*).

The two opening presentations outlined the goals and objectives of the federal government in relation to AI and ML, highlighting the policies, principles, and standards.

1. *"Advancing Trustworthy AI," from the White House Office of Science and Technology Policy (OSTP), focused on national level policy, principles, and initiatives, and*
2. *"Trustworthy and Responsible AI," from the National Institute of Standards and Technology (NIST), provided standards frameworks and highlights from the recently published National Security Commission (NSC) AI Final Report.*

There were four AI and ML panels: Policy, Standards, 4I Program, and Consumer Safety.

1. *The Policy panel consisted of presentations from the European Commission, the University of Nevada Las Vegas William S. Boyd School of Law, and Pennsylvania State University Law School, Engineering Department, School of Electrical Engineering and Computer Science. The presenters focused on policy considerations regarding AI and ML in consumer products.*
2. *The Standards panel included UL, American National Standards Institute (ANSI), and the Consumer Technology Association (CTA) with presentations on Traditional Safety Standards, ISO JTC 1/SC 42 AI standard, and AI in Consumer Products.*
3. *The program panel discussed a proposed framework for CPSC to address AI and ML in consumer products with presentations from Worchester Polytech Institute (WPI), American Council for Technology-Industry Advisory Council (ACT-IAC), Underwriters Laboratories (UL), and the National Artificial Intelligence Institute (NAII). The presenters identified AI and ML Components, discussed implications of AI and ML Capabilities, the impacts of AI and ML for Consumer Products, and the concerns for the Iterative and adaptive nature of AI and ML technologies.*
4. *The Consumer Safety panel discussed safety concerns and opportunities relative to AI and ML in consumer products from Kids in Danger, UL, Harvard, Testing Inspections and Certification, and Bureau Veritas.*

The day concluded with a roundtable conversation regarding safety considerations surrounding the design, development, and deployment of AI and ML in consumer products.

APPENDIX:

1. *AI ML Forum Agenda (included)*
2. *AI ML Forum Presentations (included)*

LINKS:

1. *CPSC AI ML Forum Public Notice:* https://www.federalregister.gov/documents/2020/12/01/2020-26441/cpsc-artificial-intelligence-forum
2. *AI ML Forum Event Video:* https://www.youtube.com/watch?v=mltDge0Fmx4

CPSC

AI ML FORUM

APPENDIX 1:

AGENDA

Artificial Intelligence and Machine Learning Consumer Products Forum

U.S. Consumer Product Safety Commission
March 2, 2021

0900 – Opening Remarks: Duane Boniface, CPSC

0910 – Introduction: Nevin Taylor, CPSC Chief Technologist

0915 – Presentation: "Advancing Trustworthy Artificial Intelligence" by Lynne Parker, Ph.D. (*Deputy U.S. Chief Technology Officer and Founding Director ot the National Artificial Intelligence Initiative Office*)

0935 – Policy Panel:

- "Product safety and new technologies: the EU approach" by Pablo Olivares Martinez (*Legal and Policy Officer, European Commission*)
- "Consumer Product Safety Commission (*CPSC*) Artificial Intelligence (*AI*) and Machine Learning (*ML*) Forum" by Dr. Andrea Matwyshyn (*Associate Dean of Innovation and Technology, Pennsylvania State University Law School*) and Dr. Patrick McDaniel (*William L. Weiss Professor of Information and Communications Technology in the School of Electrical Engineering and Computer Science, Pennsylvania State University)*
- "AI Policy and Safety Concerns in Children's Toys" by Professor Ben Edwards and Julia Armendariz *(William S. Boyd School of Law, University of Nevada Las Vegas)*

1030 – Break

1040 – Panel: "Trustworthy and Responsible AI" by Elham Tabassi (*Chief of Staff in the Information Technology Laboratory (ITL), National Institute of Standards and Technology (NIST)*)

1100 – Standards Panel

- "UL Standards Approach to AI" by Deborah Prince (*Standards Program Manager, UL*)
- "Artificial Intelligence: Introduction of JTC 1/SC 42" by Heather Benko (*Senior Manager, ANSI*)
- "CPSC AI ML Forum: CTA Remarks" by Dave Wilson (*Vice President of Technology & Standards, Consumer Technology Association (CTA)*)

1200 – Lunch Break

1300 – 4I Program Panel

- "AI/ML in Consumer Products" by Professor Holly Ault, Austin Master, Benjamin Guerriero, Tyson Wiseman, and Logan Young (*Worchester Polytechnic Institute (WPI*))
- "Implications of AI and ML in Consumer Products" by Swathi Young (*Chief Technology Officer, Integrity Management Services, Inc.)*
- "Impact of AI on Product Safety" by Dr. Mahmood Tabaddor (*Global Head of Predictive Modeling and Analytics: Machine Learning, Data Science, Data Analytics, System Modeling, UL*)
- "AI Iteration: What happens when AI keeps learning?" by Gil Alterovitz, PhD. (*Director, National Artificial Intelligence Institute (NAII)*)

1410 – Break

1420 – AI and ML Consumer Products Discussion Panel

- Karin Athanas *(Executive Director of the Testing, Inspections, and Certification (TIC) Council)*
- Nancy Cowles *(Executive Director, Kids in Danger)*
- Travis Norton *(Director of Technical Services, Americas Region, for Bureau Veritas)*
- Michael Wiklund *(General Manager, EMERGO by UL)*
- Adam Wood *(Collaborator, Harvard Innovation Laboratory)*

1500 – Future of AI/ML Round Table

1600 – Closing

CPSC

AI ML FORUM

APPENDIX 2:

PRESENTATIONS

Advancing Trustworthy Artificial Intelligence

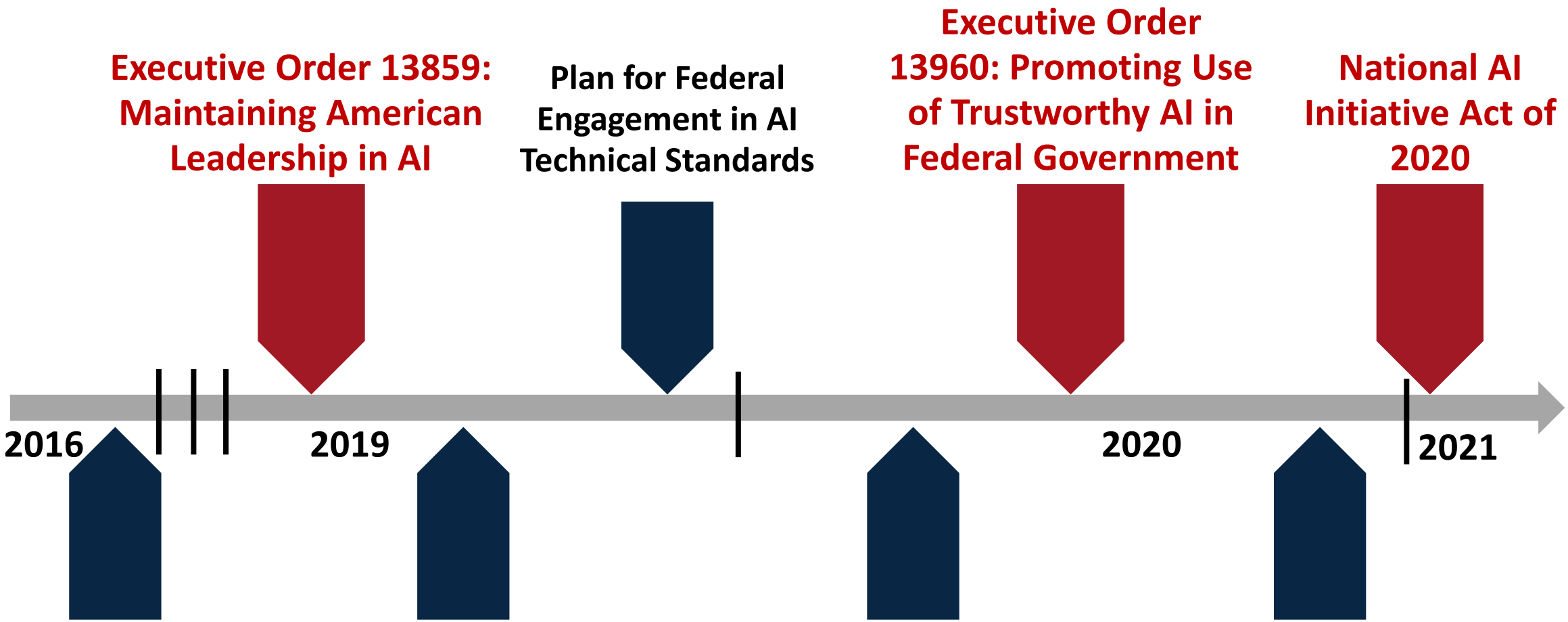
Lynne Parker, Ph.D.

Deputy U.S. Chief Technology Officer

Founding Director, National Artificial Intelligence Initiative Office

March 2021

Numerous Landmark U.S. National AI Policy Actions



(timeline not to scale)

National AI Initiative Act of 2020 (NAIIA)

- Became law on January 1, 2021
  - As part of the "*William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021", H.R. 6395, Division E.*

DIVISION E—NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE ACT OF 2020

SEC. 5001. SHORT TITLE.

This division may be cited as the "National Artificial Intelligence Initiative Act of 2020".

- Bipartisan legislation defining National AI Initiative, with goals of:
  - Ensuring continued U.S. leadership in AI research and development (R&D);
  - ***Leading world in development and use of trustworthy AI systems in public and private sectors;***
  - Preparing present and future U.S. workforce for integration of AI systems across all sectors of economy and society; and
  - Coordinating AI research, development, and demonstration activities among civilian agencies, Department of Defense, and Intelligence Community to ensure that each informs work of the others.

National AI Initiative – Advancing Trustworthy AI

- The United States must foster public trust and confidence in AI in order to fully realize AI's potential.
- Fundamental to public trust is the development of appropriate AI technical standards and risk management frameworks.
- When considering AI regulations or policies, Federal agencies should continue to promote innovation while protecting privacy, civil rights, civil liberties, and other democratic values.

Advancing Trustworthy AI

- *What is trustworthy AI?*
- *How do we best advance trustworthy AI?*

What is trustworthy AI?

- **"Trustworthy AI" includes concepts such as:**
  Explainability, transparency, safety, privacy, security, robustness, fairness, bias, ethics, validation, verification, interpretability, etc.

- **Concepts reflected in several national and international documents, such as:**
  - ***OECD Recommendation on AI*** – May 2019
    - An historic intergovernmental, consensus statement of AI principles, now agreed to by 44 nations. G20 also adopted same principles.

  - ***U.S. Executive Order 13960 – "Promoting the Use of Trustworthy AI in the Federal Government"***
    - *Lawful*
    - *Purposeful and performance-driven*
    - *Accurate, reliable, and effective*
    - *Safe, secure, and resilient*
    - *Understandable*
    - *Responsible and traceable*
    - *Regularly monitored*
    - *Transparent*
    - *Accountable*

How do we *advance* trustworthy AI?

Regulatory Agencies

Per OMB M-21-06: Guidance for regulation of AI in the private sector

OMB Policy Guidance

Per EO 13960: Roadmap of policy guidance for agencies in their use of AI in government

DoD's/IC's implementation of ethical principles for AI

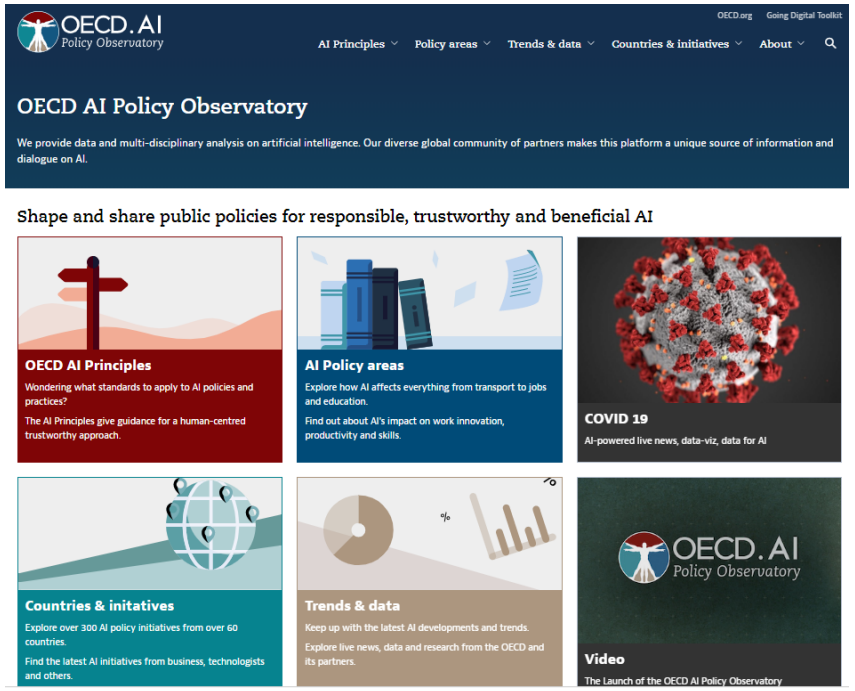
GSA's AI Center of Excellence

ARTIFICIAL INTELLIGENCE

GSA

IT MODERNIZATION Centers of Excellence

OECD AI Policy Observatory

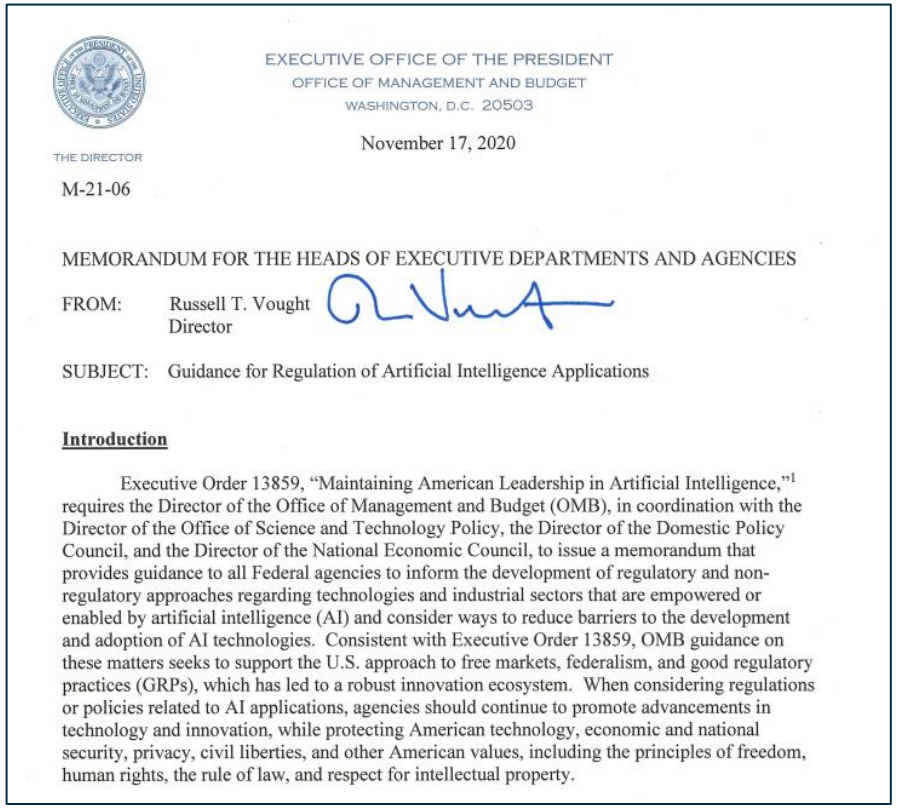


GPAI

THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

OMB Guidance for Regulation of AI in the Private Sector (1)

- Final Guidance Memo (M-21-06) issued Nov. 2020
- Defines guidance for regulation of AI in private sector
- **Principles for stewardship of AI applications:**
  1) Public Trust in AI
  2) Public Participation
  3) Scientific Integrity & Information Quality
  4) Risk Assessment & Management
  5) Benefits & Costs
  6) Flexibility
  7) Fairness & Non-Discrimination
  8) Disclosure & Transparency
  9) Safety & Security
  10) Interagency Coordination



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

November 17, 2020

THE DIRECTOR

M-21-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought
Director

SUBJECT: Guidance for Regulation of Artificial Intelligence Applications

Introduction

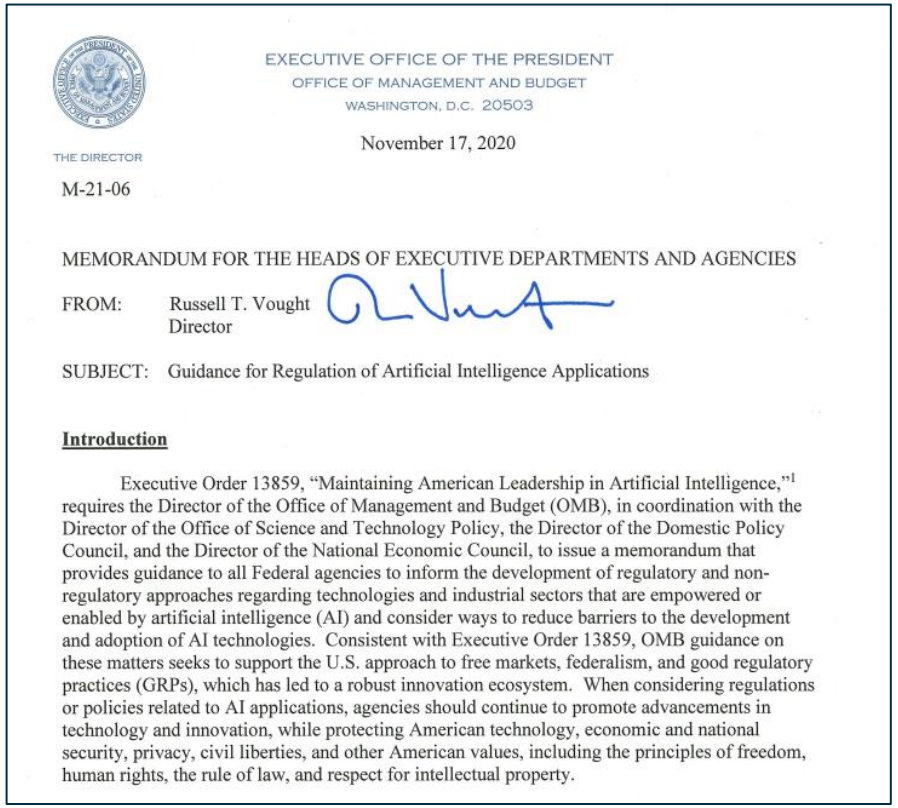
Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence,"[1] requires the Director of the Office of Management and Budget (OMB), in coordination with the Director of the Office of Science and Technology Policy, the Director of the Domestic Policy Council, and the Director of the National Economic Council, to issue a memorandum that provides guidance to all Federal agencies to inform the development of regulatory and non-regulatory approaches regarding technologies and industrial sectors that are empowered or enabled by artificial intelligence (AI) and consider ways to reduce barriers to the development and adoption of AI technologies. Consistent with Executive Order 13859, OMB guidance on these matters seeks to support the U.S. approach to free markets, federalism, and good regulatory practices (GRPs), which has led to a robust innovation ecosystem. When considering regulations or policies related to AI applications, agencies should continue to promote advancements in technology and innovation, while protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property.

https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

OMB Guidance for Regulation of AI in the Private Sector (2)

- **Non-regulatory approaches:**
  1) Sector-specific policy guidance
  2) Pilot programs and experiments
  3) Voluntary consensus standards
  4) Voluntary frameworks

- **Reducing barriers to deployment/use of AI:**
  1) Access to federal data and models for AI R&D
  2) Transparently communicating benefits/risks
  3) Agency participating voluntary consensus standards and conformity assessment
  4) International regulatory cooperation



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

November 17, 2020

THE DIRECTOR

M-21-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought
Director

SUBJECT: Guidance for Regulation of Artificial Intelligence Applications

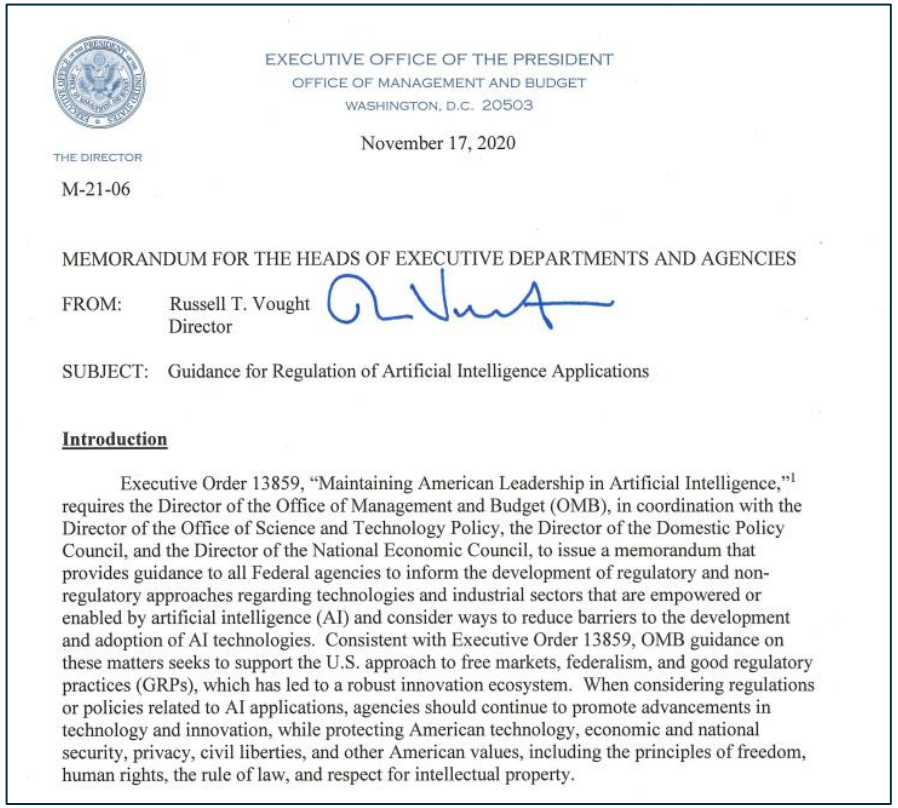
Introduction

Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence,"[1] requires the Director of the Office of Management and Budget (OMB), in coordination with the Director of the Office of Science and Technology Policy, the Director of the Domestic Policy Council, and the Director of the National Economic Council, to issue a memorandum that provides guidance to all Federal agencies to inform the development of regulatory and non-regulatory approaches regarding technologies and industrial sectors that are empowered or enabled by artificial intelligence (AI) and consider ways to reduce barriers to the development and adoption of AI technologies. Consistent with Executive Order 13859, OMB guidance on these matters seeks to support the U.S. approach to free markets, federalism, and good regulatory practices (GRPs), which has led to a robust innovation ecosystem. When considering regulations or policies related to AI applications, agencies should continue to promote advancements in technology and innovation, while protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property.

https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

OMB Guidance for Regulation of AI in the Private Sector (3)

- **Agency plans for consistency with memo:**
  1) Statutory authorities for agency regulation of AI
  2) Active collections of AI-related information
  3) AI use case priorities
  4) AI regulatory barriers
  5) Planned regulatory actions concerning AI applications



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

November 17, 2020

THE DIRECTOR

M-21-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought
Director

SUBJECT: Guidance for Regulation of Artificial Intelligence Applications

Introduction

Executive Order 13859, "Maintaining American Leadership in Artificial Intelligence,"[1] requires the Director of the Office of Management and Budget (OMB), in coordination with the Director of the Office of Science and Technology Policy, the Director of the Domestic Policy Council, and the Director of the National Economic Council, to issue a memorandum that provides guidance to all Federal agencies to inform the development of regulatory and non-regulatory approaches regarding technologies and industrial sectors that are empowered or enabled by artificial intelligence (AI) and consider ways to reduce barriers to the development and adoption of AI technologies. Consistent with Executive Order 13859, OMB guidance on these matters seeks to support the U.S. approach to free markets, federalism, and good regulatory practices (GRPs), which has led to a robust innovation ecosystem. When considering regulations or policies related to AI applications, agencies should continue to promote advancements in technology and innovation, while protecting American technology, economic and national security, privacy, civil liberties, and other American values, including the principles of freedom, human rights, the rule of law, and respect for intellectual property.

https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

National AI Initiative

Prioritize AI R&D
Grow and sustain U.S. research leadership and capacity

Strengthen AI Research Infrastructure
Enhance access to high quality data, models, and computing resources

Advance Trustworthy AI
Modernize governance and technical standards for AI-powered technologies, protecting privacy, civil rights, civil liberties, and other democratic values

Leverage AI for Government and National Security
Apply AI to improve provision of government services and national security

Promote International AI Engagement
Engage with like-minded allies to promote a global AI environment supportive of democratic values

Train AI-Ready Workforce
Provide AI-ready education at all levels: K-12, college, re-training, re-skilling, R&D workforce

Thank you!

Lynne Parker, Ph.D.

Deputy U.S. Chief Technology Officer

Founding Director, National Artificial Intelligence Initiative Office

March 2021

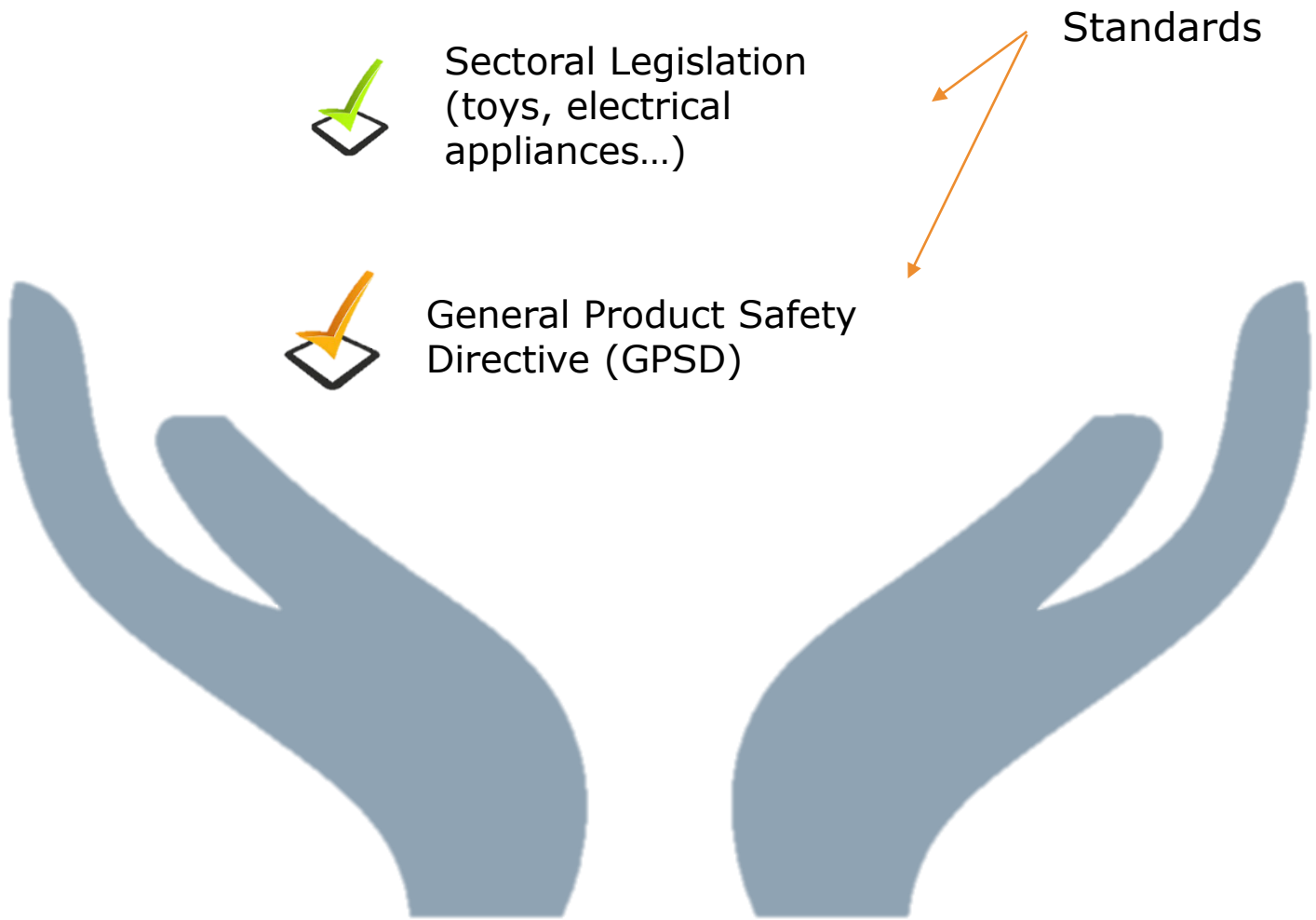
Product safety and new technologies: the EU approach

Artificial Intelligence (AI) and Machine Learning (ML) Forum

2 March 2021

Pablo Olivares Martinez
European Commission
Product Safety and Rapid Alert System Unit

The EU Product Safety framework



European Commission

General Product Safety Directive

2001

General Product Safety Directive (GPSD)

2021

Proposal for a revised General Product Safety legislation

European Commission

Sub-group on AI, connected products and other new challenges in product safety

- Group set in December 2019
- Wide range of stakeholders: businesses, industry associations, consumer organisations, national authorities, academia, ‘ethical hackers’
- Support to the revision process of the General Product Safety Directive
- Opinion published in January 2021

European Commission

Content of the Opinion

New risks

- Cybersecurity
- Personal security
- Mental health

Placing on the market

- Safety during the entire lifecycle of a product

Software

- Software updates
- The concept of ‘substantial modification’

European Commission

Other upcoming initiatives

- Horizontal framework on AI
- Revision of the Machinery Directive
- Initiatives on liability of new technologies

European Commission

Thank you

European Commission

AI Policy and Safety Concerns in Children's Toys

Presented by:

Julia Armendariz
JD Candidate

Kelsey DeLozier
JD Candidate

Pete Reyes
JD Candidate

Research Supervisor:
Benjamin P. Edwards
Associate Professor of Law

On behalf of:
Consumer Federation of America

UNLV | WILLIAM S. BOYD SCHOOL OF LAW

News

Parents upset after Stanford Shopping Center security robot knocks down toddler

Three-hundred pound autonomous machine allegedly ran over boy in 'freak accident'

by Sue Dremann / Palo Alto Weekly

Uploaded: Wed, Jul 13, 2016, 4:41 pm
Time to read: about 4 minutes

36

The parents of a 17-month-old boy say a Stanford Shopping Center security robot knocked down their son and ran over his foot before they were able to get him out of the way.

Tiffany Teng and Eric Cheng of San Jose were shopping at the mall, located on El Camino Real, on July 7 when their son, Harwin, was struck by the autonomous robot and knocked to the ground, they said.

Harwin had been walking in front of them near the Splendid and Armani Exchange shops when he collided with the robot, which was moving toward them. The boy struck his head on the robot and was knocked to the

SLIDESHOW

San Jose parents Tiffany Teng and Eric Cheng say their 17-month-old son, Harwin, was struck by a security robot, made by Mountain View-based Knightscope, at Stanford Shopping Center in Palo Alto last week. Photo courtesy of Tiffany Teng.

Basic Concepts

Artificial Intelligence (AI)

Any method for programming computers to enable them to carry out tasks or behaviors that would require intelligence if performed by humans.

Machine Learning (ML)

Involves improving algorithms as their exposure to more data is extended over time, resulting in a machine or system learning and detecting patterns.

Consumer Products with AI technology

Example: **A product that adapts and conforms to consumer preferences through evolution resulting in outcomes beyond a manufacturer's control.**

Advantages: **Enhance product efficiency, performance, and consumer experience**.

Challenges: **Establishing policies addressing public concerns.**

Policy Challenges and Problems

Unique Challenges with AI Policy

Fostering Appropriate Understanding and Consumer Trust

Pace of Technological Development

Applying Human-Focused Law to Non-Human Decision-Makers

Policy Challenges and Problems

Policy Proposals and Responses

Voluntary Standards

Extending Existing Laws to Cover AI and Machine Learning

Developing New Legal Frameworks

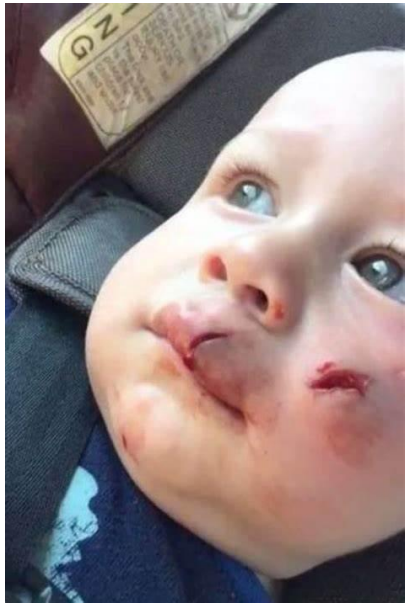
Opt-Out and Information Rights

Enable External Audits

Insurance-Style Solutions

Concerns for Children Interacting AI and ML Products

- Direct Safety Risks
- Indirect Safety Risks
- Circumstantial Dangers
- Developmental Risks



Specific Recommendations for AI and ML Consumer Products for Children

- Standard Guidelines
- Appropriate Datasets for Training
- Consumer Education

UNLV | WILLIAM S. BOYD SCHOOL OF LAW

Conclusion

The real problem is not whether machines think but whether men do. - B.F. Skinner

Consumer Product Safety Commission (CPSC) Artificial Intelligence (AI) and Machine Learning (ML) Forum

March 2, 2021

Prof. Andrea M. Matwyshyn

Assoc. Dean of Innovation and Technology, Penn State Law

Professor, Penn State Engineering

Founding Director, Penn State Policy Innovation Lab of Tomorrow (PILOT)

AI/ML Security = Product Safety

- Direct risks to physical safety from manipulability of AI/ML products
  - AGI isn’t on the near horizon
- Risks are already visible in ML
  - Training data methodology
  - Outcome manipulability
- CPSC is key to AI/ML safety oversight
  - Building safe(r) products
  - Getting unsafe products off the market
- Case study: AI/ML in baby monitors and other health-ish products

Consumer Product Safety Commission (CPSC) Artificial Intelligence (AI) and Machine Learning (ML) Forum

March 2, 2021

Prof. Patrick Drew McDaniel

William L. Weiss Professor of Information and Communications Technology in the School of Electrical Engineering and Computer Science at Pennsylvania State University

AL/ML and Unsafe Products: A technical perspective

- AI - planning - think about Roombas
- ML is statistical machine learning - think spam detection
- AI and ML can both be manipulated
  - Leading to confidentiality, integrity, privacy and availability problems
  - Undermine tasks, putting consumers at risk
- Consumer products with embedded ML that are easily manipulable through attacks are unsafe
- Most builders are not currently taking precautions against attacks on AI/ML.
- Case study : IoT door camera/lock system

Trustworthy and Responsible AI

Elham Tabassi
National Institute of Standards and Technology
March 2, 2021

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

1956 Summer Research at Dartmouth

A PROPSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I.B.M. Corporation
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carries out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely describes that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kind of problems of now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

62 years later …

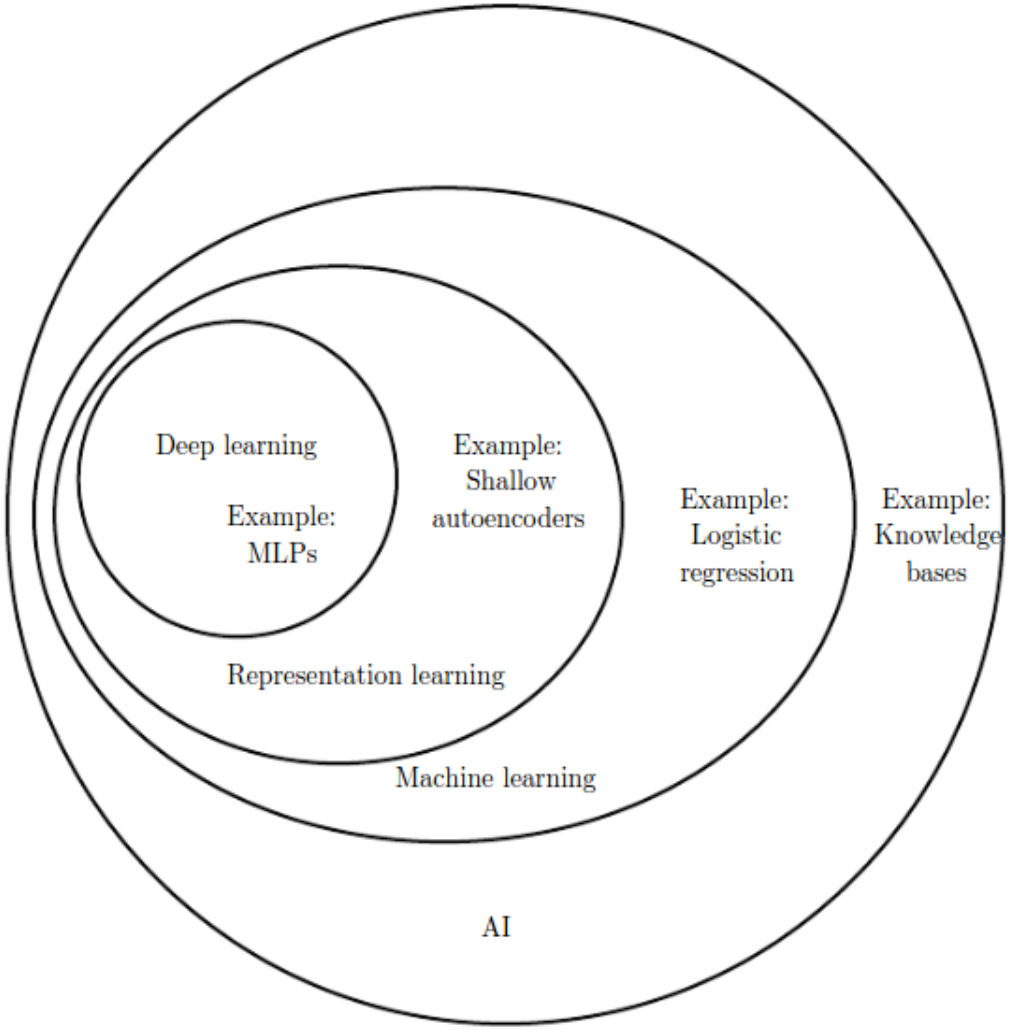


Figure 1.4: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology.

Deep Learning Book, *Ian Goodfellow and Yoshua Bengio and Aaron Courville*
https://www.deeplearningbook.org/contents/intro.html

NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE

FINAL REPORT: **KEY JUDGMENTS IN THE INTRODUCTION**

WHY DOES AI MATTER?

- AI is ubiquitous in everyday life.
- Deploying and adopting AI remains a hard problem.
- AI tools are diffusing broadly and rapidly.
- AI is changing relationships between humans and machines.

PART I - DEFENDING AMERICA IN THE AI-ERA.

1. AI is the quintessential "dual use" technology—it can be used for civilian and military purposes.
2. We can expect the large-scale proliferation of AI-enabled capabilities.
3. AI-enabled capabilities will be tools of first resort in a new era of conflict.
4. AI will transform all aspects of military affairs.
5. Competitors are actively developing AI concepts and technologies for military use.
6. AI will revolutionize the practice of intelligence.
7. Defending against AI-capable adversaries without employing AI is an invitation to disaster.
8. Compelling logic dictates quick, but careful and responsible AI adoption.
9. There is an emerging consensus on principles for using AI responsibly in the defense and intelligence communities.
10. The U.S. government still operates at human speed not machine speed.

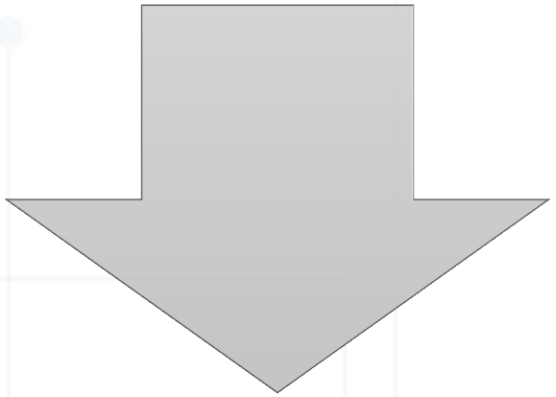
PART II - WINNING THE TECHNOLOGY COMPETITION.

1. China is organized, resourced, and determined to win the technology competition.
2. Advancements in AI are contributing to a broad platform technology competition in e-commerce, search engines, social media, and much else.
3. The AI competition is complicated by deep interconnections.
4. The United States retains advantages in critical areas, but trends are worrisome.
5. The U.S. government must take a hands-on approach to national technology competitiveness.
6. The AI competition will require White House leadership.

AI FOR WHAT ENDS? TECHNOLOGY AND VALUES.

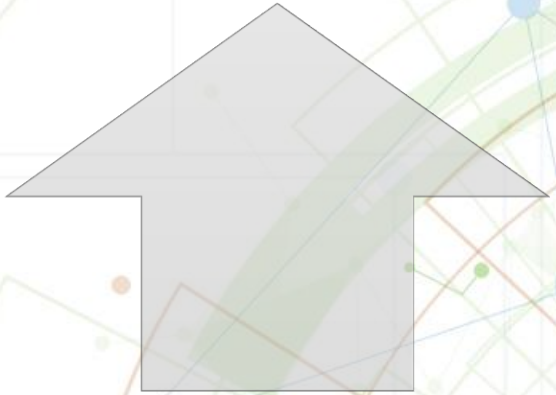
1. The U.S. government should develop and field AI-enabled technologies with adequate transparency, strong oversight, and accountability to protect against misuse.
2. The United States must lead a coalition of democracies.

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

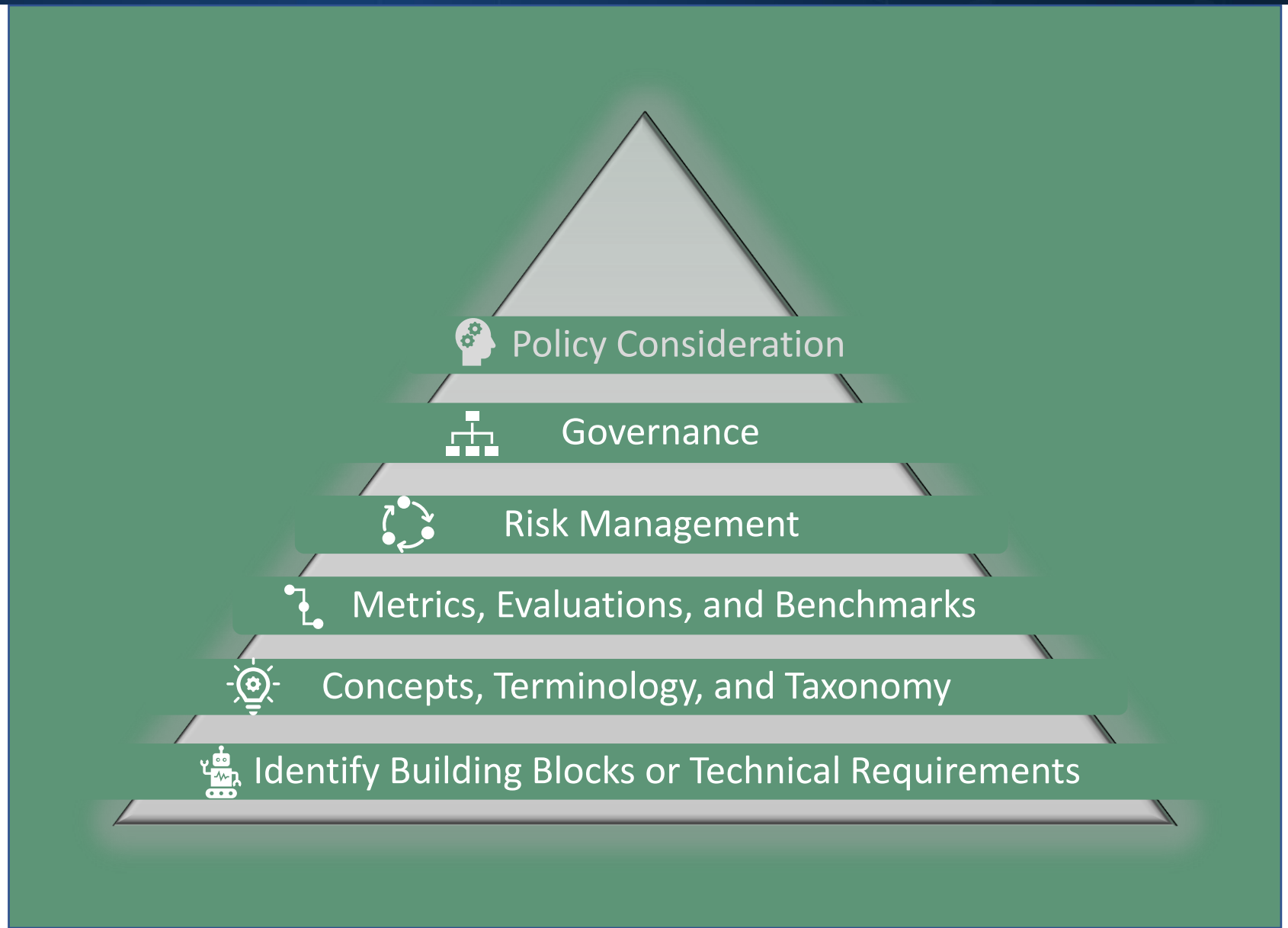


Raise productivity, enable more efficient use of resources, change the way we live and work, and increase creativity.

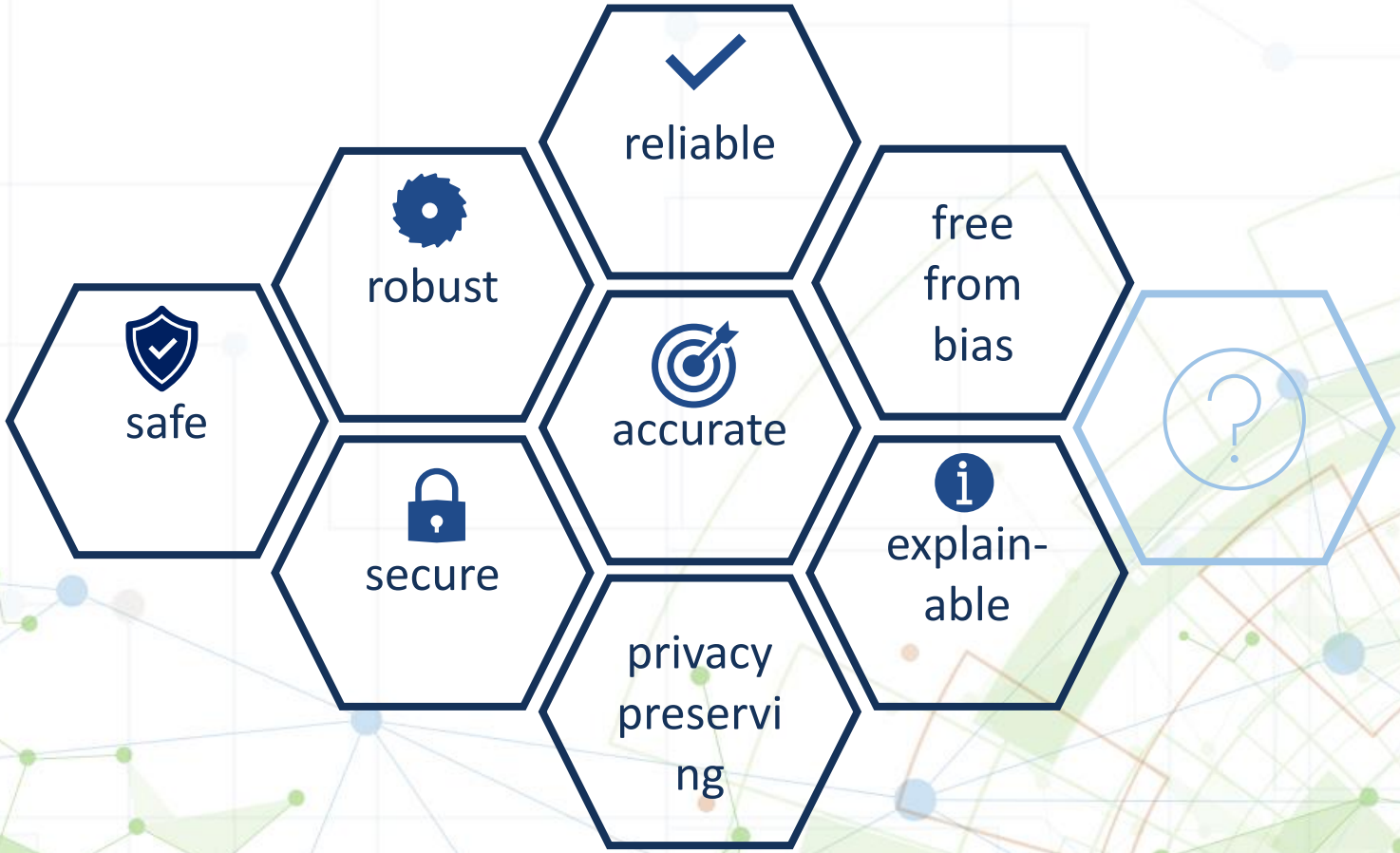
Negative impact on job, exacerbate the trend of rising inequality, and (even) threat to humanity.

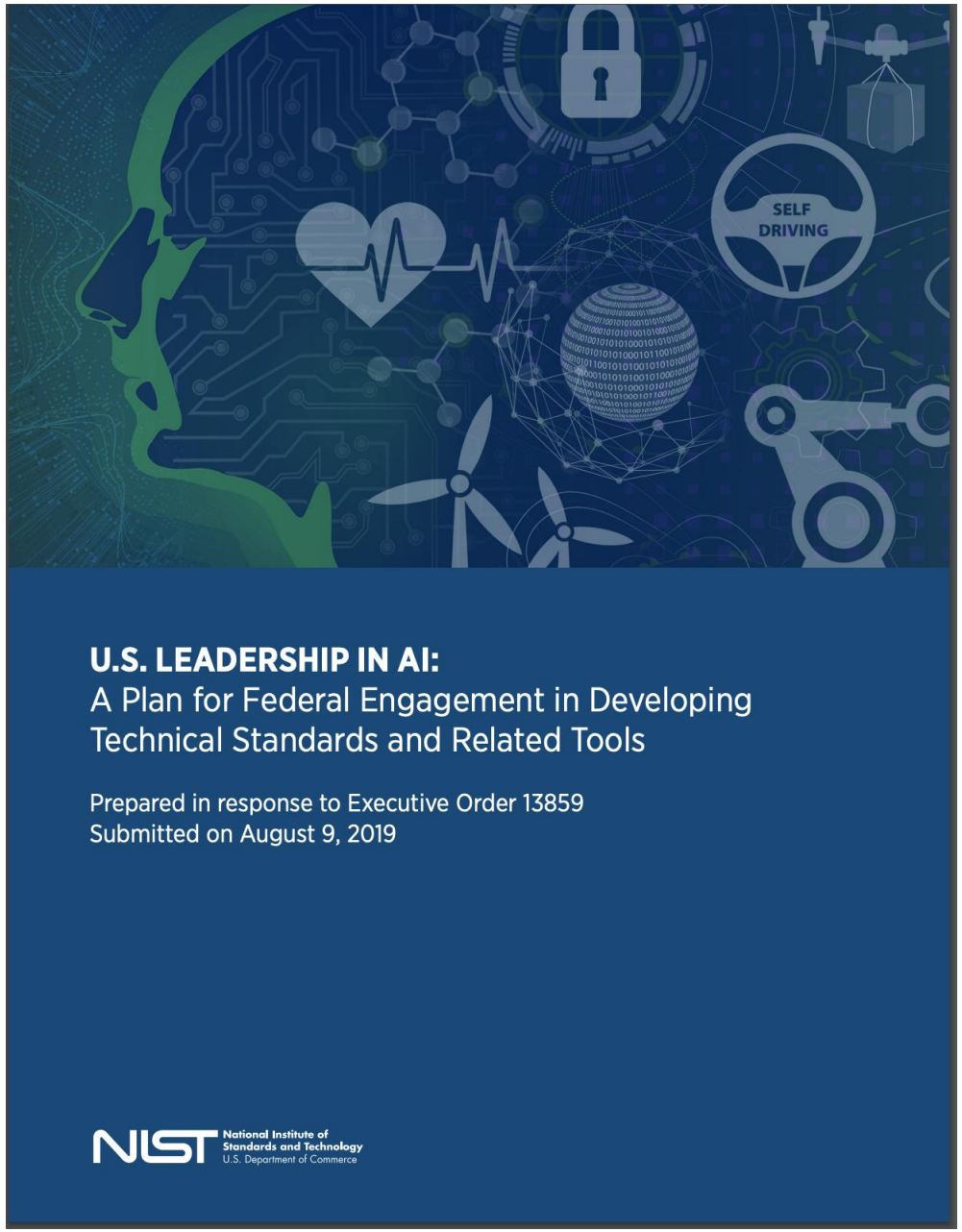


Trustworthy AI's Foundation
From Technical Requirements to Policy Creation



Core Building Blocks of Trustworthy AI





https://www.nist.gov/sites/default/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf

What AI technical standards and related tools are needed?

Risk Management

Safety

Performance Testing Methodology

Human Interaction

Societal & Ethical Consideration

Trustworthiness

Networking

Data and Knowledge

Privacy

Metrics

Governance

Concepts and Terminology

Technically sound,

fit for purpose,

and timely AI standards.

Which standards characteristics are important?

Innovation-oriented

Cross sector

Sector-specific

Flexible and regularly updated

Sensitive to ethical considerations

Clearly stated scope and intended use

Monitor and manage AI systems

Using clear language

Standards development efforts that merit federal engagement

Consensus Based

Inclusive and accessible

Multi-path

Open and transparent

Globally relevant

Levels of participation at standards development activities

MONITORING

PARTICIPATING

INFLUENCING

LEADING

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

STAY IN TOUCH

ai_standards@nist.gov

www.nist.gov/topics/artificial-intelligence/ai-standards

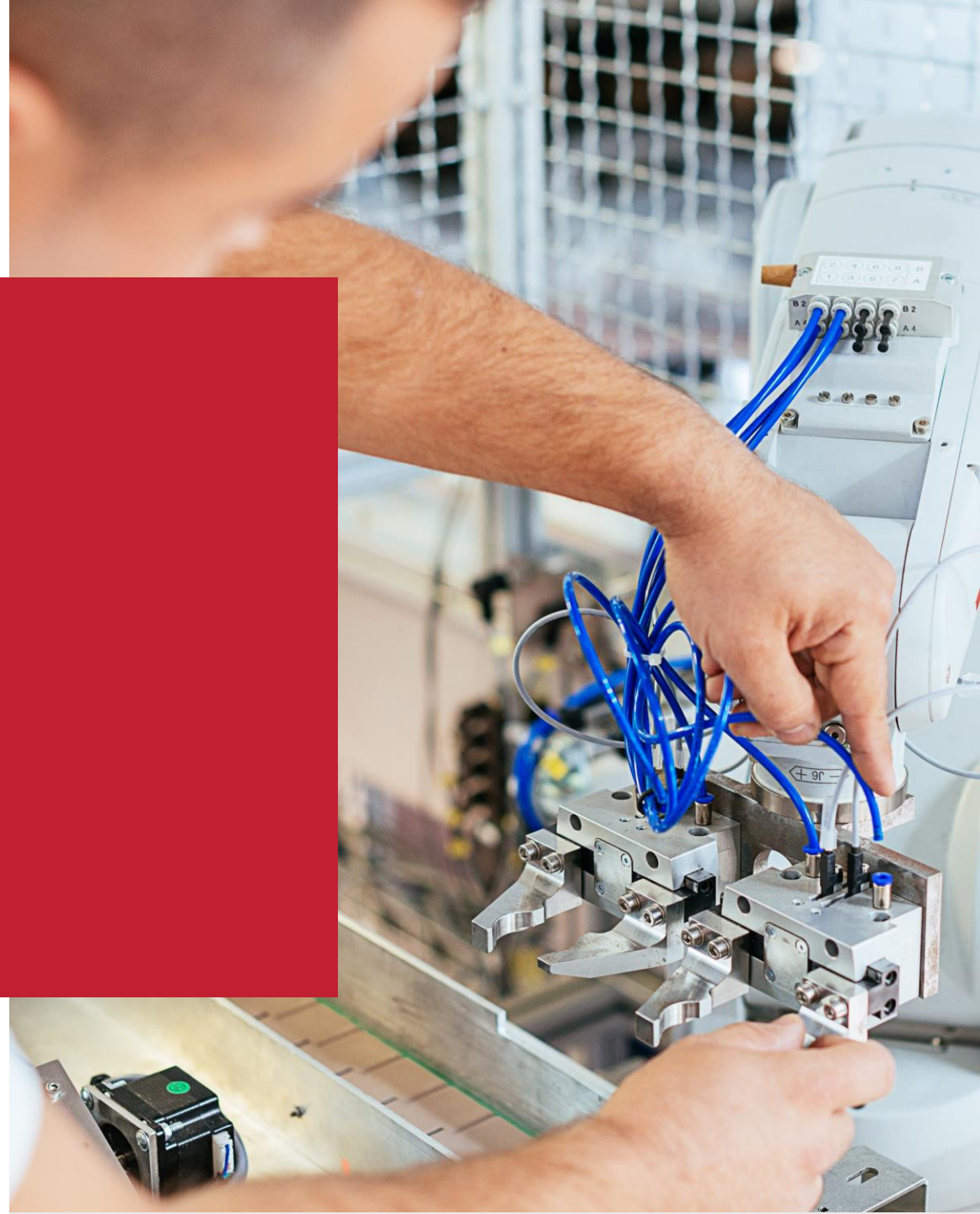
UL Standards Approach to AI

Deborah Prince

Traditional Safety Standards

What are elements of a traditional standard?

- Construction requirements
- Protection against injury to person
- Performance
- Manufacturing and Production Test
- Marking
- Instructions



Adding Software to a Product

Software added to product, now what?

Increasingly products are smart enabled, this requires an additional set of criteria to be considered such as:

- Communication signals
- Functional safety
- Over air updates safety
- Network concerns
- Security

AI/ML Products

AI/ML now added into the product

All the previously mentioned concerns for smart enabled but additional concerns.

Taking the human out of the loop has benefits but it also changes the risk profile

Safe products will depend on defining the exhaustive list of overlapping conditions a product might encounter, what's called the operational design domain

How do you address when a product is out of the ODD?

Addressing safe transfer back to the human if necessary

AI/ML Products

- No longer a single product but a suite of products
- Will the AI/ML of one product negatively effect the safety of another products?

Safety Case Approach

Safety Case Approach

Safety Case is a narrative, a story about the product. Did the manufacturer think about every environment, all scenarios, all potential risks and argue acceptable mitigation. It is a structured argument

“The safety case shall be a structured explanation in the form of claims, supported by argument and evidence, that justifies that the item is acceptably safe for a defined operational design domain, and covers the item’s lifecycle.” *UL 4600, Standard for Safety for Evaluation of Autonomous Products*

Example

Safety Case Defined

A structured argument, supported by a body of evidence, that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given environment.

Claims

A falsifiable statement that contributes to establishing that safety is acceptable

Arguments

Construct a safety case (including claims, arguments, and evidence) that a particular requirement has been met. The resulting argument shows that evidence supports the claims.

Conformance

An independent assessment result indicates that the item’s safety case meets the requirements

UL 4600, Safety for Evaluation of Autonomous Products

UL 4600 is a safety case structured standard. Safety criteria is defined as:

- Mandatory
- Required
- Highly recommended
- Recommended

In addition, UL 4600 provides prompts of “have you considered”, examples as well as known pitfalls

Example of ML clause

The machine learning architecture, training, and V&V approach shall provide acceptable machine learning performance.

8.5.2.2 **REQUIRED:**

a) Description of machine learning techniques used

EXAMPLES: Training technique, use of transfer learning

b) Description of machine learning architecture and hyper-parameters

EXAMPLES: Type of network, number of layers

c) Definition of performance metrics and evaluation against those metrics

EXAMPLES: ROC curves, false positive rate, false negative rate, precision/recall

d) Traceability of performance metrics to argument that performance is acceptable

e) Arguments that V&V procedure follows best practices for machine learning

f) Evidence of suitable engineering rigor in the use of tools and techniques that are safety related

EXAMPLE: Tools supporting collection and analysis of test data, tool support for neural network weight configuration management

g) **Pitfall:** Machine learning techniques are generally prone to overfitting, resulting in lower than expected performance in real world operation.

Example of ML clause

Description of Machine Learning (ML) approach, if any

1) Data selection

2) Data cleaning

3) Algorithm selection

4) ML architecture selection

5) Model training approach

NOTE: It is important to continually revalidate machine learning based functionality that is continually updated based on experience. It is envisioned that revalidation (i.e., update and re-assessment of the safety case) is done after a sufficiently large change to machine learning functionality.

Example of AI clause

8.5.6 **The safety case shall address the acceptability of any other "Artificial Intelligence" ("AI") techniques used beyond machine learning.**

8.5.6.2 **REQUIRED:**

a) Identify and describe other AI techniques being used, if any

b) Argue that each used AI technique provides acceptable capabilities

8.5.6.3 **HIGHLY RECOMMENDED**

:a) Address non-deterministic aspects of AI technique

b) Address validity and coverage of any heuristics used

c) Address adherence to best practices for employing each technique

d) Identify and argue mitigation of potential hazards, and risks

8.5.6.4 **RECOMMENDED:**

a) To maximum extent practicable, rely upon traditional software safety argument approaches

What else?

- As manufacturers gain experience with UL 4600, built in is a recommendation that the knowledge be proposed into the standard expanding the pitfalls or use cases etc. Sharing knowledge will help build a safer products
- Standards must be flexible and nimble as technology evolves

Thank you

Artificial Intelligence

Introduction of JTC 1/SC 42

CPSC Artificial Intelligence Forum

Heather Benko, Committee Manager - SC 42

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS

SC 42 – Artificial Intelligence

ISO/IEC JTC 1 – Information Technology - Overview

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

Created in 1987, JTC 1 continues to be the only joint technical committee of ISO and IEC

Scope –International standardization in the field of Information Technology

ANSI has served as Secretariat since its creation

- Provides management and administration
- Nominates Chair

Philip Wennblom (Intel) serves as JTC 1 Chair

ISO/IEC JTC 1, Information technology –Structure

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

22 Subcommittees

- Areas from coded character sets to IT security and privacy, biometrics and AI
- U.S. participates in 19 Subcommittees
- U.S. holds the Secretariat for 7 JTC 1 Subcommittees

4 Working Groups reporting directly to JTC 1

- Covering Smart cities, 3D printing and scanning, Trustworthiness and **Quantum computing**
- U.S. participates in all 4 Working Groups

11 Advisory Groups

- To study new areas of opportunity and establish practices that apply across JTC 1
- U.S. participates in 10 Advisory Groups

13 consortia approved as JTC 1 PAS Submitters (expedited process)

- Facilitates cooperation and collaboration on ICT standardization
- Fills gaps in the ISO/IEC portfolio of ICT standards where consortia have developed standards with wide market acceptance

Introduction to JTC 1/SC 42

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

Established via JTC 1 Resolution at the end of 2017

- ANSI holds the Secretariat
- Mr. Wael William Diab (US) Chair

Scope

- Standardization in the area of Artificial Intelligence
  - Serve as the focus and proponent for JTC 1's standardization program on Artificial Intelligence
  - Provide guidance to JTC 1, IEC, and ISO committees developing Artificial Intelligence applications

Strong growth of the committee and its program of work

- 30 P-members and 17 O-members. More than **doubled the number of NBs** participating since its creation
- Attendance ~150. Work program grew from 2 initial projects to 23 active projects and 6 published documents
- >50% of work program progressed beyond working drafts
- Recent areas of work: AI MSS, AI Systems Engineering, Data Quality, Explainability of AI, Functional safety, AI Data, Lifecycle

Active liaison relationships to support system integration mission

- Over 30 liaisons established, that include 7 Category A Liaisons. Goal of wide adoption of SC 42 standards

Ecosystem Approach

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

Motivation

- AI is not a single technology but a collection of technologies
- Stakeholders are numerous and diverse
- Stakeholders are not treating AI and other key technologies as separate and disparate technology research areas
- Rather, stakeholders are approaching the deployment of AI systems from a business angle with a focus on customers needs, segments, services, products and regulatory requirements

Considerations for wide adoption

- While technology capability continues to be paramount it is not the only motivator
- Diverse stakeholder ecosystem necessitates industry collaboration across domains (e.g. IT/OT)
  - E.g. application areas such as transportation, medical, financial, robotics, manufacturing etc.
- By considering AI technologies against the backdrop of market segments / needs, additional synergies are being identified e.g. AI, analytics, Big Data, IoT, data ecosystem
- Broad standardization approach that includes and goes beyond traditional interoperability

SC 42 and the Holistic AI Ecosystem

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

A new approach to standardization is needed that

- Takes into account the context of use of the technology by looking at both technology capability and non-technical requirements such as business requirements, regulatory and policy requirements, application domain needs, and ethical and societal concerns
- Translating the above into technical requirements
- Building foundational standards that allow communities to build upon such as terminology, use cases, application guidance and reference architectures
- Linking technology innovation communities such as proprietary implementations, research, SDOs and open source communities

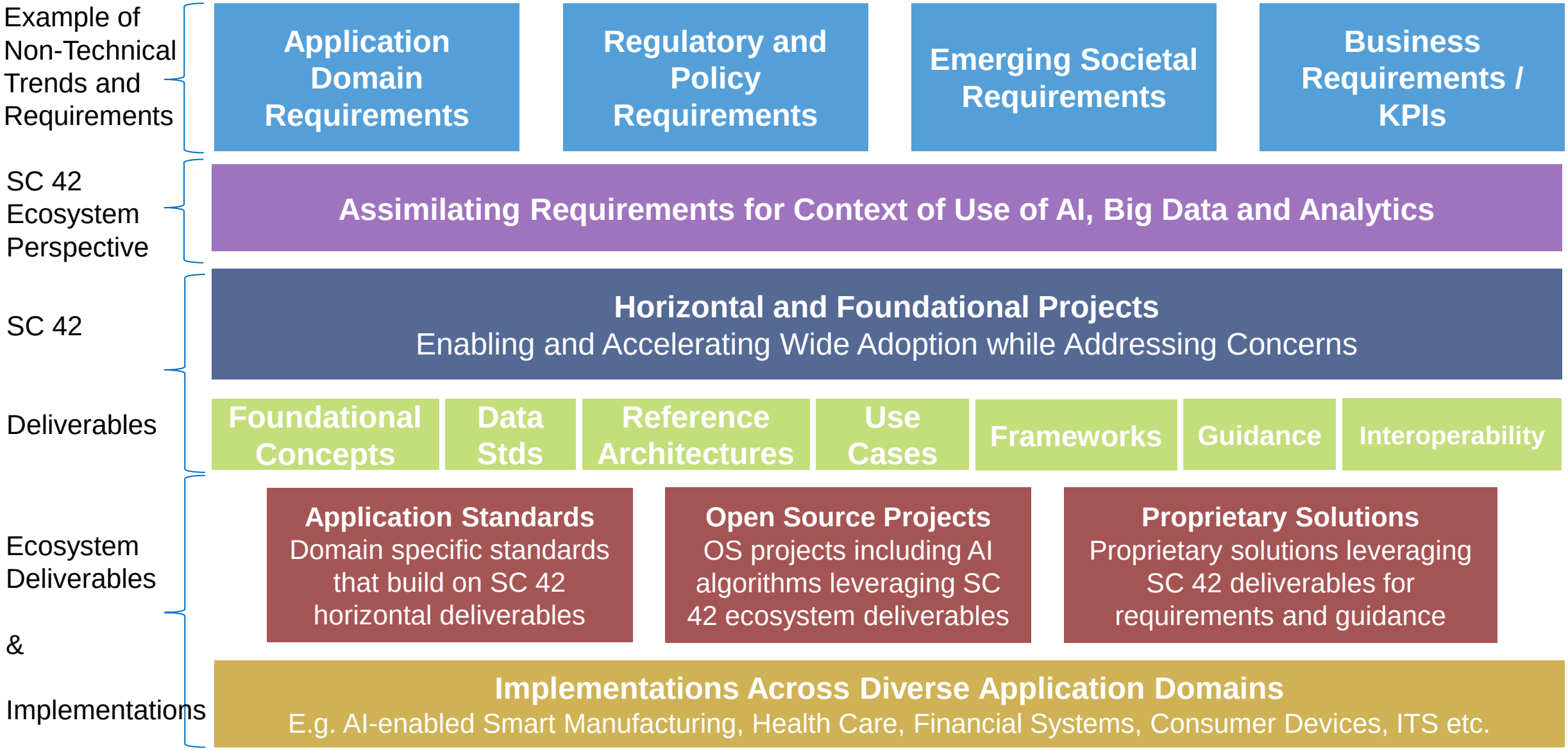
The result not only accelerates technology adoption but it also takes into account its context of use and builds an ecosystem

AI and Big Data are perfect examples of this type of technology, it's use and IT's evolution

SC 42 has adopted this holistic ecosystem approach providing the glue between requirements and technical requirements through the horizontal deliverables the committee develops

Bridging the Gap – An Ecosystem Approach

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence



Recently added projects

AI Management System Standard

Data quality for analytics and ML

- Four part series looking at Terminology, Data Quality measures, Management and guidelines and process frameworks

Functional safety and AI Systems

Objectives and methods for explainability for ML models and AI systems

AI Systems Lifecycle processes

Guidelines for AI applications

Complete Work Programme at:
https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0

AI Management System Standard

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

AI technologies bring AI-specific concerns beyond those of traditional IT systems. For example

- ML based AI system may provide different results depending on the training data used
- The choice of training data when using an AI system is an additional process that an organization needs to perform to ensure the intended overall system performance
- Consumers of AI products and services may lack trust in the AI supplier organization
- Assurance that the organization considered for fairness, inclusiveness, accountability etc. of AI system

MSS containing AI-specific process requirements allows for assessment of conformance or auditability of the processes

- Allows organizations to check how well it meets their objectives in the use of an AI system
- For trusted 3rd party performing a check or audit, a certificate of conformance can be issued

Concluding Remarks

ISO JTC 1 IEC
INFORMATION TECHNOLOGY STANDARDS
SC 42 – Artificial Intelligence

SC 42 is the first of its kind international standards committee looking at the full AI ecosystem

- AI, Big Data and related analytics are key technologies enabling the digital transformation

SC 42 has a rapidly growing work program

- Strong growth and execution on work program. 29 projects (6 published, 23 active), 6 WGs, AHGs on specific topics
- Robust study program for anticipated new work addressing AI ecosystem and system level concerns with AI

SC 42 engaging in extensive outreach and global collaboration

- Tremendous outreach via ISO, IEC and national bodies. Extensive and diverse liaison network

Part of the ISO, IEC and JTC 1 families

- Access to broad, diverse and numerous committees that range from horizontal to vertical areas
- System integration committee providing guidance to ISO, IEC and JTC 1 committees looking at AI applications

Opportunity for international standards to fuel AI market growth and accelerate adoption

Contacts:

Wael William Diab, Chair SC 42 (Artificial Intelligence): wael.diab@gmail.com

Heather Benko, Committee Manager SC 42 (Artificial Intelligence): hbenko@ansi.org

SC 42 – Artificial Intelligence

Consumer Technology Association

CES

CPSC AI ML Forum

CTA remarks by
Dave Wilson
VP, Technology & Standards
Consumer Technology Association

March 2, 2021

About CTA

- World’s leading innovators
  - from startups to global brands
- Members support more than 18 million American jobs
- Owns and produces CES®
  - the most influential tech event in the world

A.I. in consumer products

"children's toys, residential appliances, and recreational products are being marketed touting the use of AI, machine learning, and related technologies to improve product efficacy and consumer experience."

- CPSC AI Forum announcement, 85 Fed. Reg. at 77183

A.I. in consumer products

- Mapping software
- Ridesharing apps
- Email spam filters
- Personal digital assistants
- Streaming services
- Smart thermostats
- Toy dolls

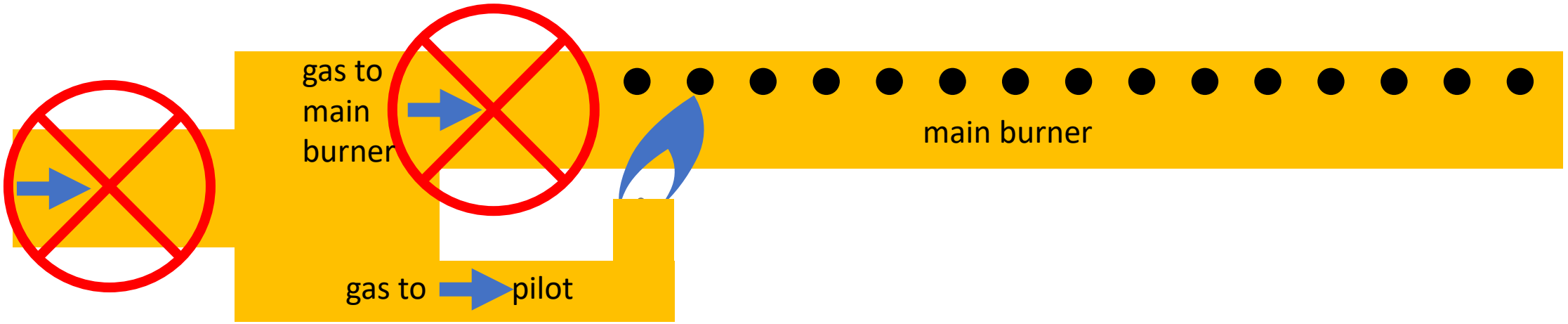
Built-in safety

- A product designed to be safe when operated by human intelligence is likely to also be safe when operated by artificial intelligence

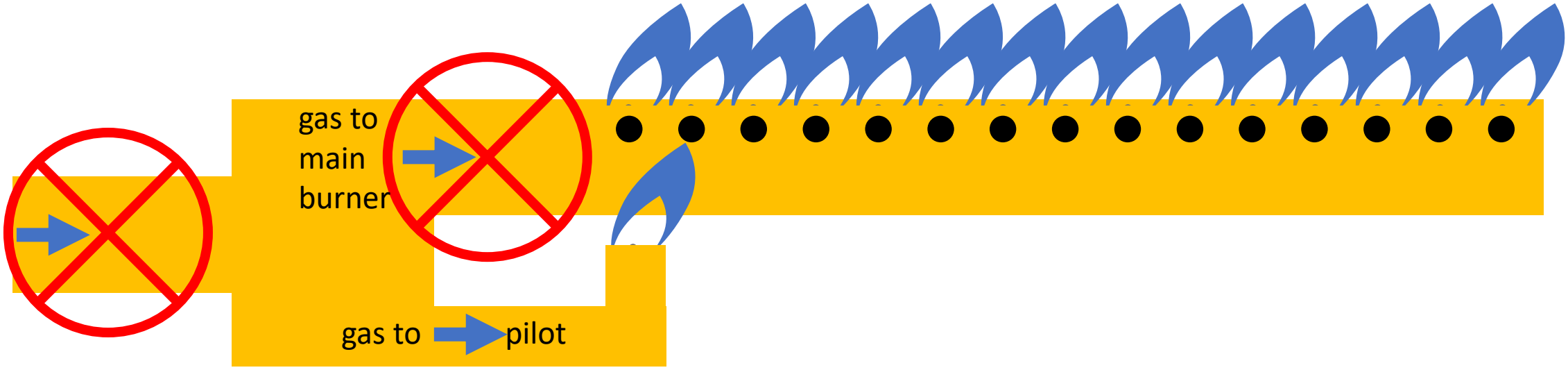
A.I. in consumer products

- Mapping software
- Ridesharing apps
- Email spam filters
- Personal digital assistants
- Streaming services
- **Smart thermostats**
- Toy dolls

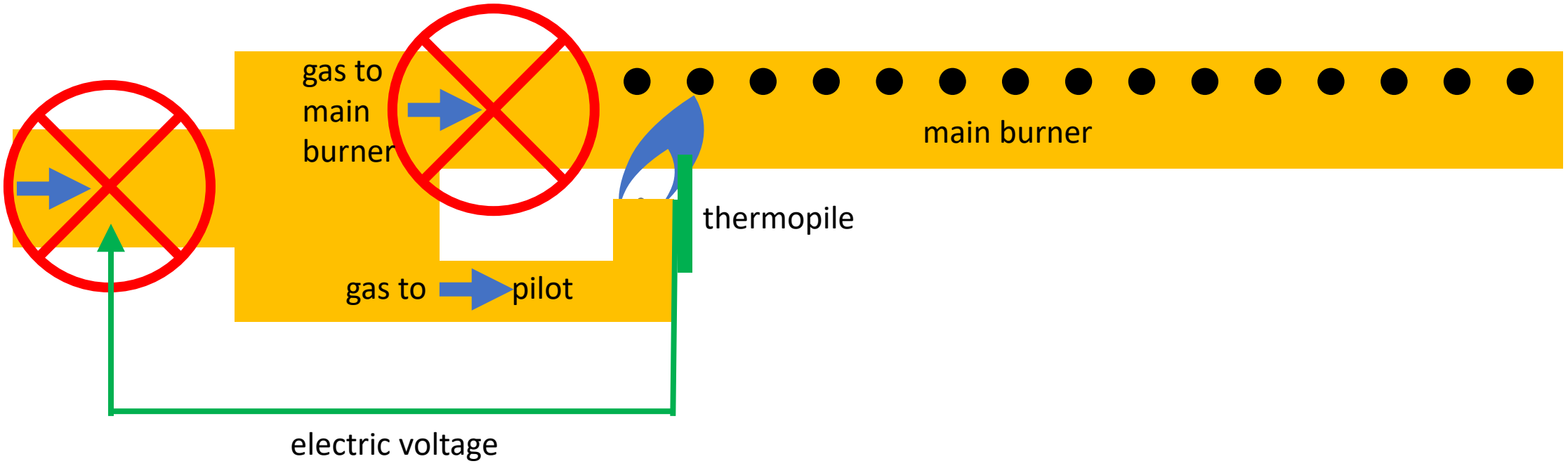
Built-in safety



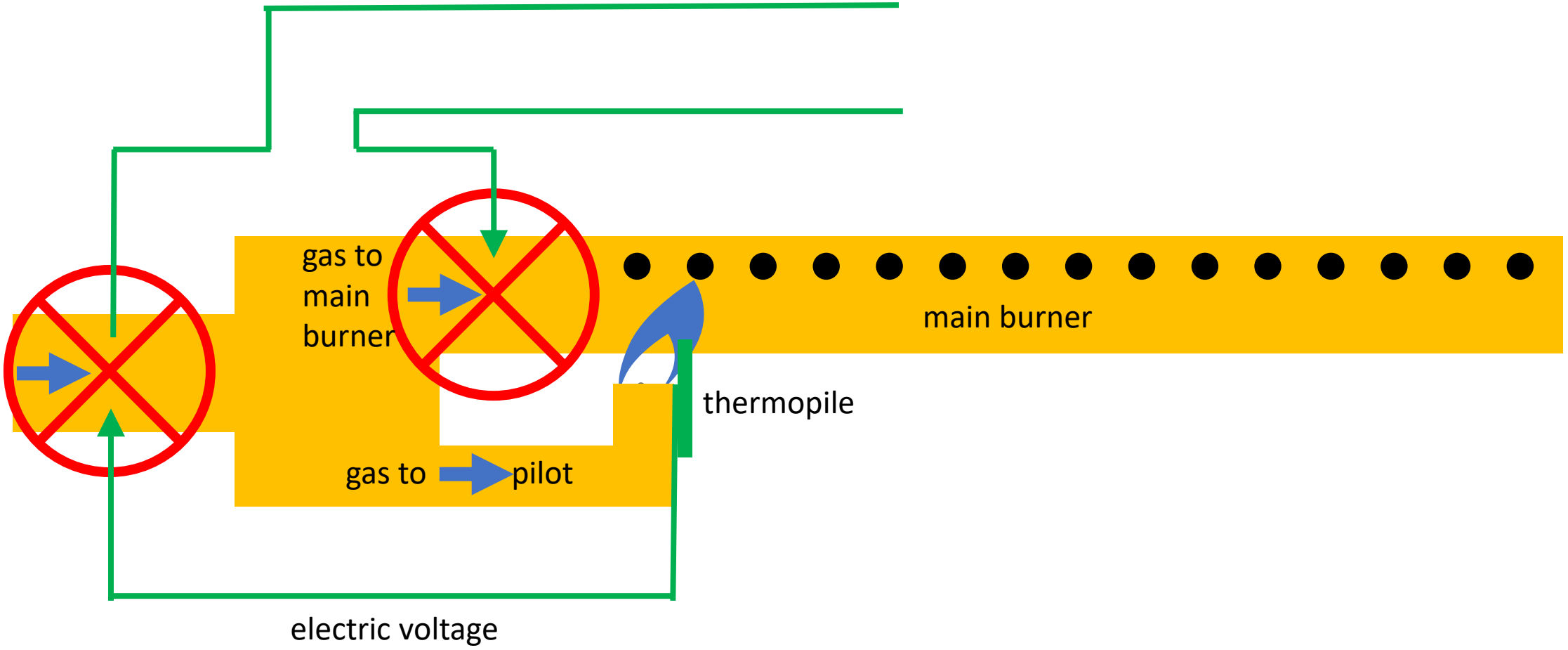
Built-in safety



Built-in safety



Built-in safety



Built-in safety

CPSC should …

- Coordinate with other federal agencies/stakeholders
- Support industry-led, consensus-based standards

AI in Health Care

- ANSI/CTA-2090
- *The Use of Artificial Intelligence in Health Care: Trustworthiness*
- Core requirements for AI-powered health care solution if it wants to be considered trustworthy

https://shop.cta.tech/collections/standards/products/the-use-of-artificial-intelligence-in-healthcare-trustworthiness-cta-2090

Or search “CTA-2090”

Conclusion

- A product designed to be safe when operated by human intelligence is likely to also be safe when operated by artificial intelligence

Thanks.

Dave Wilson
VP, Technology & Standards
Consumer Technology Association

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN CONSUMER PRODUCTS

BEN GUERRIERO, AUSTIN MASTER, TYSON WISEMAN, LOGAN YOUNG

Problem Description

Artificial Intelligence (AI) and Machine Learning (ML) are emerging fields and the CPSC needs a way to identify if consumer products are AI and ML capable

Project Goal

Identify if a consumer product has AI or ML capability

Artificial Intelligence

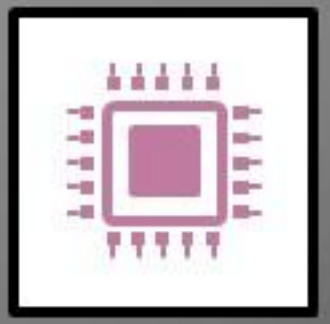
Applied Artificial Intelligence is …

- *Field of computer science*
- *System resembles human decision – making*
- *Processes information to provide outputs*
- *Systems role and support role*

Components for AI Capability

Data Source

Algorithm



Computation

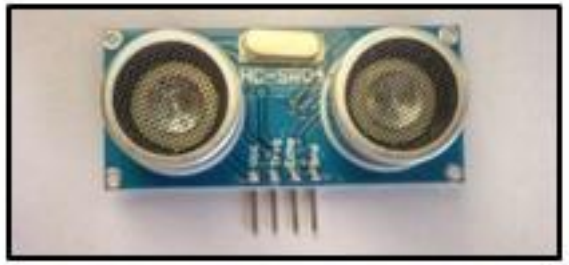
Connection

Data Source

A way to collect information

- *Sensor or human input used to gather data*

Photocell



Ultrasonic Sensor

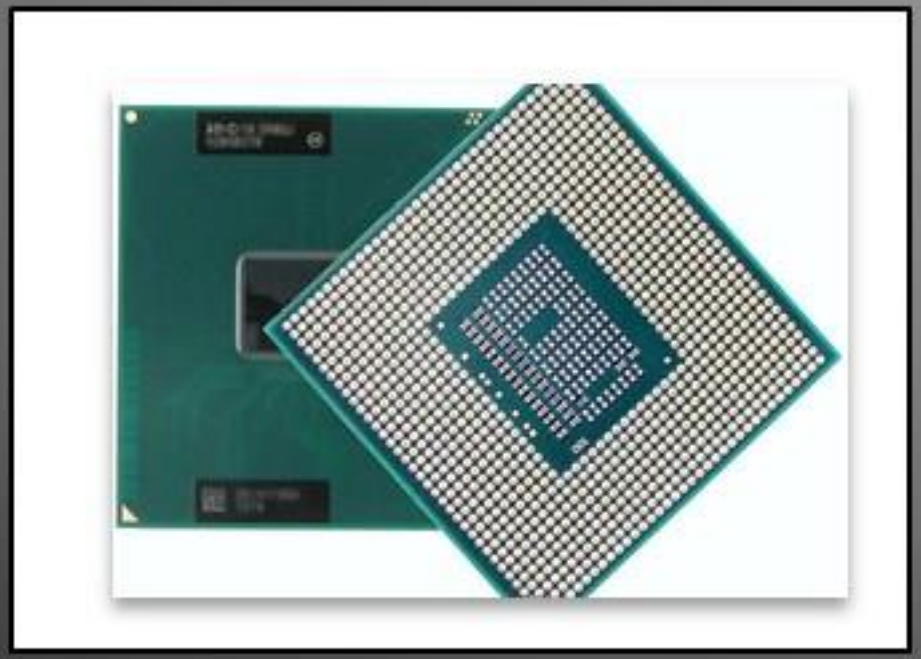
Keyboard

Algorithm

Formulas that generate outputs

- *Software in the system*

Computational Capability



System processes the given data

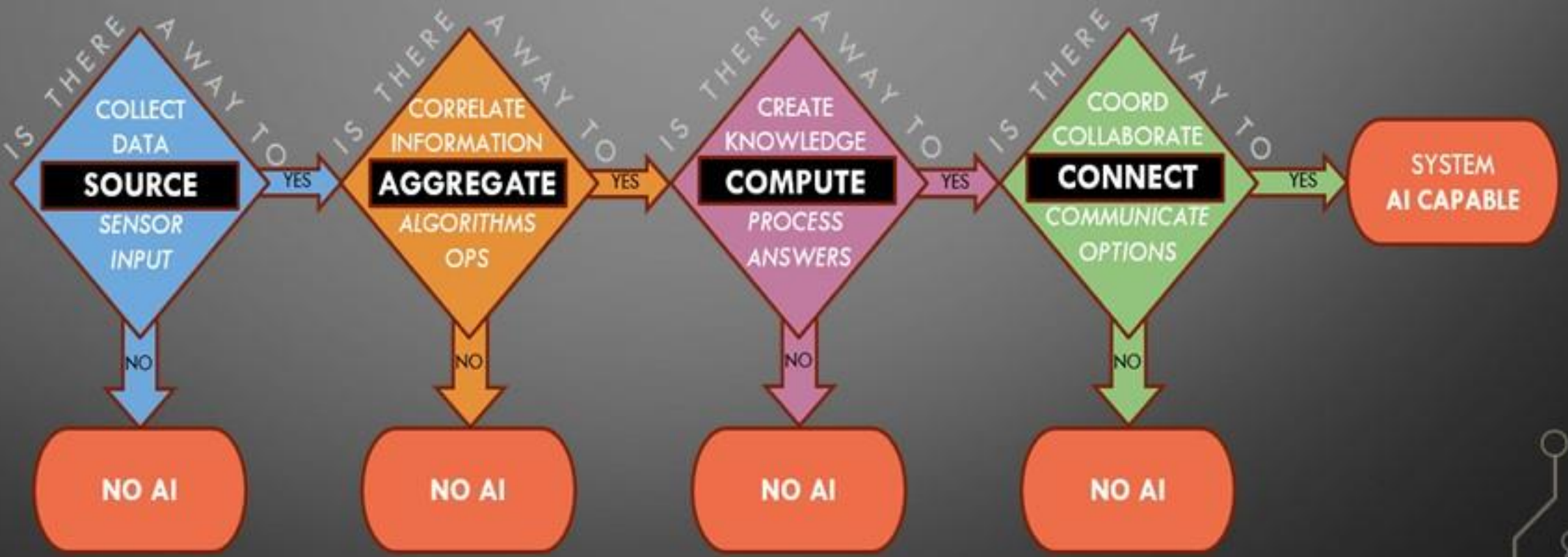
- *Aggregates data collected to allow the system to provide informed answers*
- *CPU*

Connection

Allows for the flow of information

- *Connections*
- *Communication*
- *Address Busing*

AI Screening Process

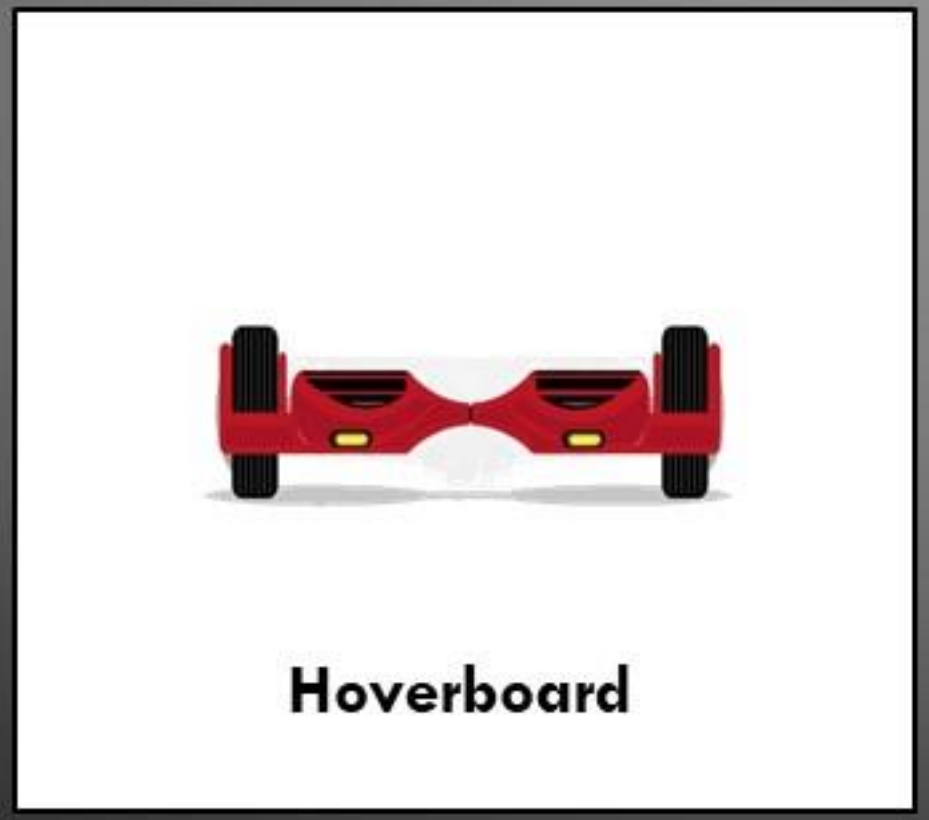


AI Capable Examples

Home Security System

- Data Source ✓
- Algorithm ✓
- Computation ✓
- Connection ✓

AI Capable Examples



Data Source ✓

Algorithm ✓

Computation ✓

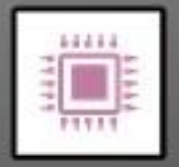
Connection ✓

Non-AI Capable Examples

Christmas Lights Timer

Data Source X

Algorithm X



Computation X

Connection X

Non-AI Capable Examples

Headphones with Microphone

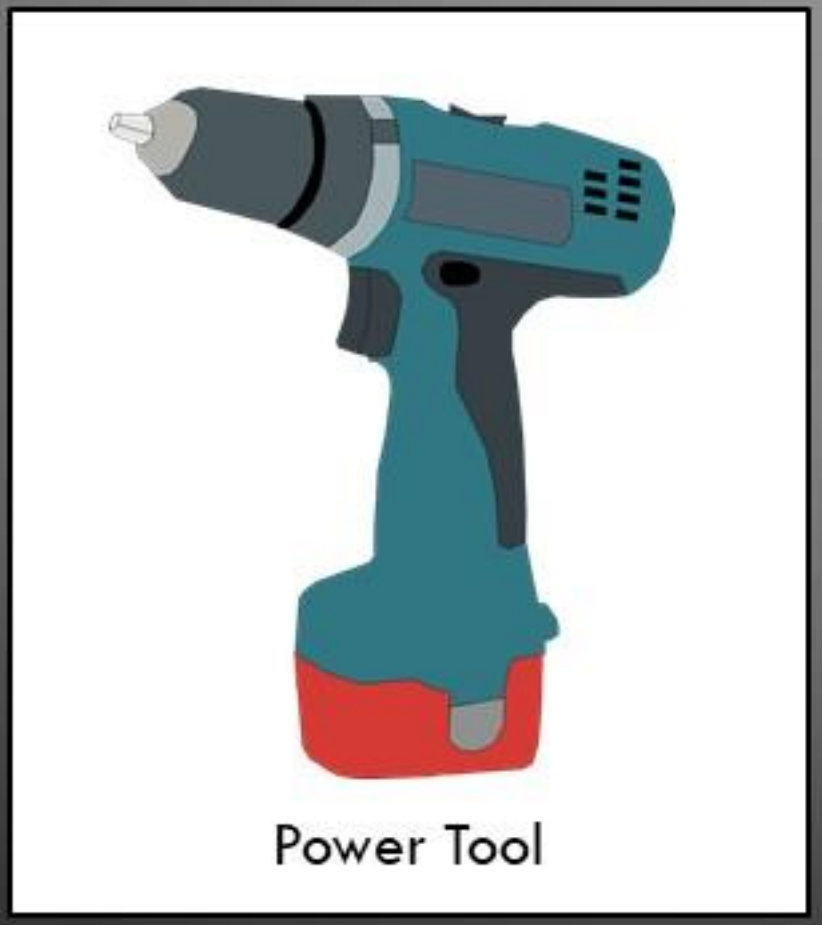
Data Source ✓

Algorithm X

Computation X

Connection ✓

Non-AI Capable Examples



Power Tool

Data Source ✓

Algorithm X

Computation X

Connection X

Machine Learning

Machine Learning

Working Definition

Machine Learning is …

- *Process of gaining knowledge or experience though given data*
- *Assesses outputs*
- *Finds differences*
- *Characterizes outcomes*
- *3D's – Distinctions, Differences, Differentials*

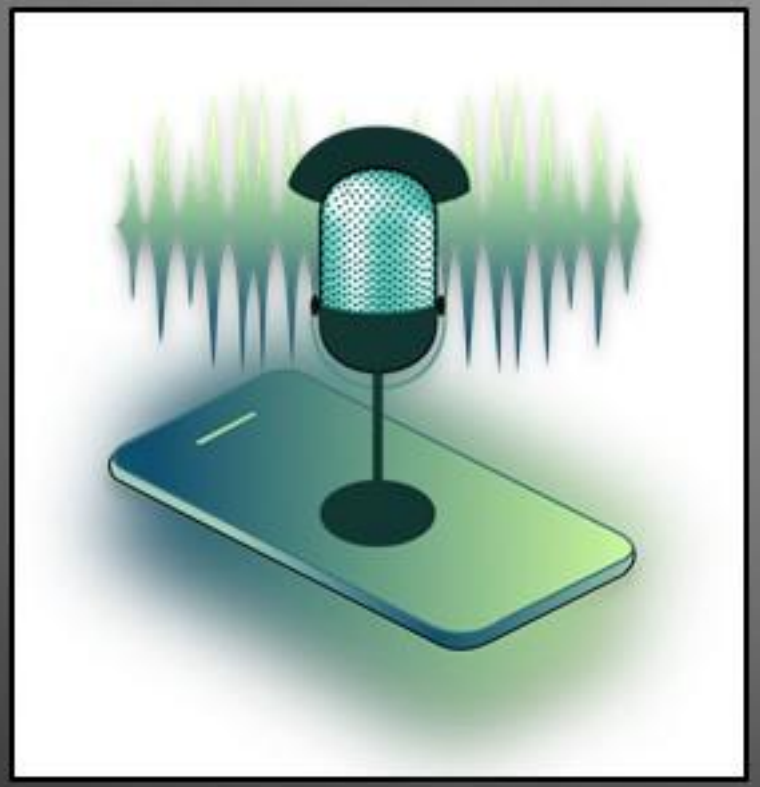
Components for ML Capability

Monitoring

Measuring

Modeling

Monitoring



Assess outputs to observe information

- *Distinctions*

Measuring

Analyze adaptations

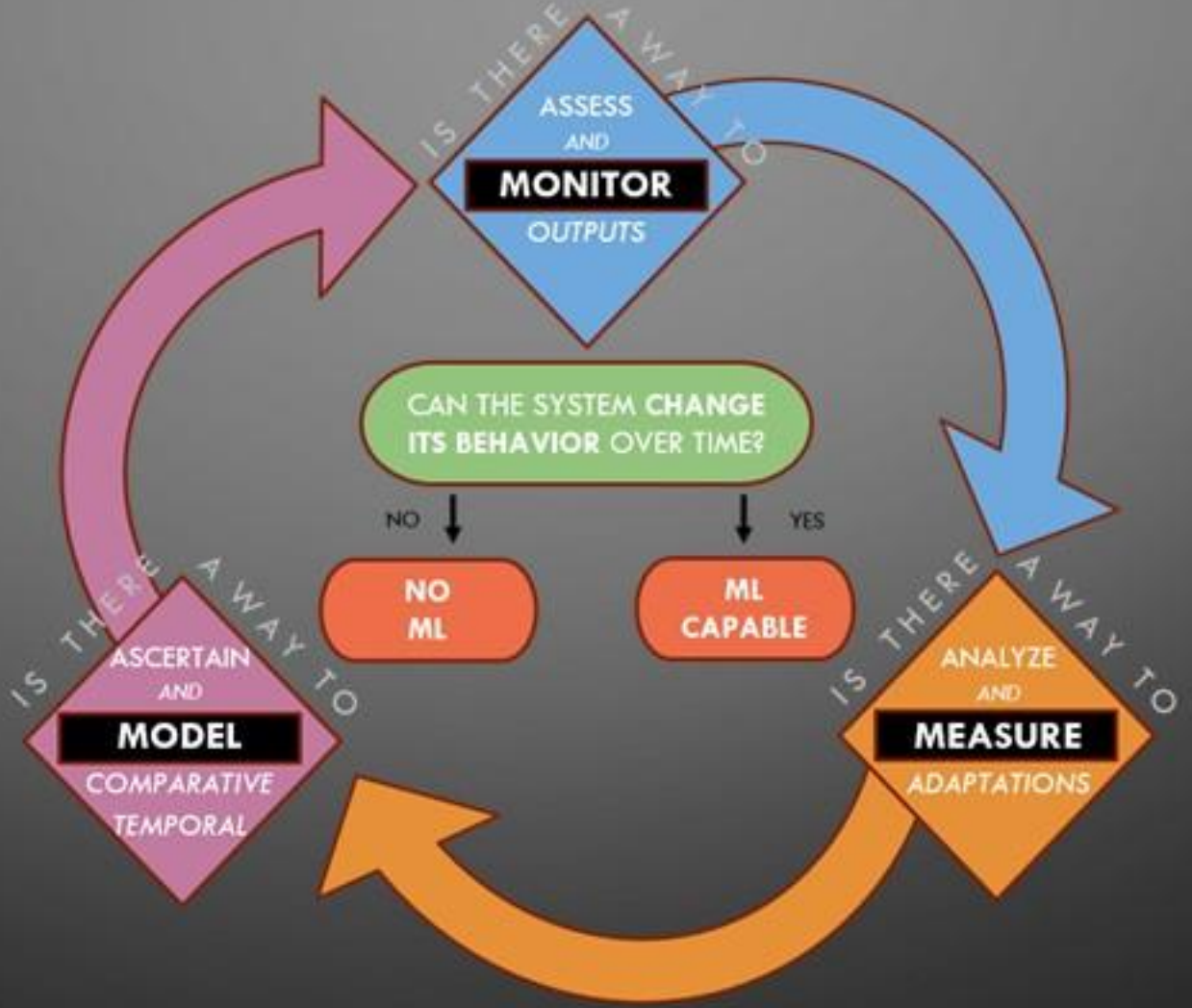
- *Differences*

Modeling

3E's – Enlighten, Empower, Evolve

- *Differentials*

ML Screening Process



ML Capable Examples

Smart Fridge

Automatic Vacuum

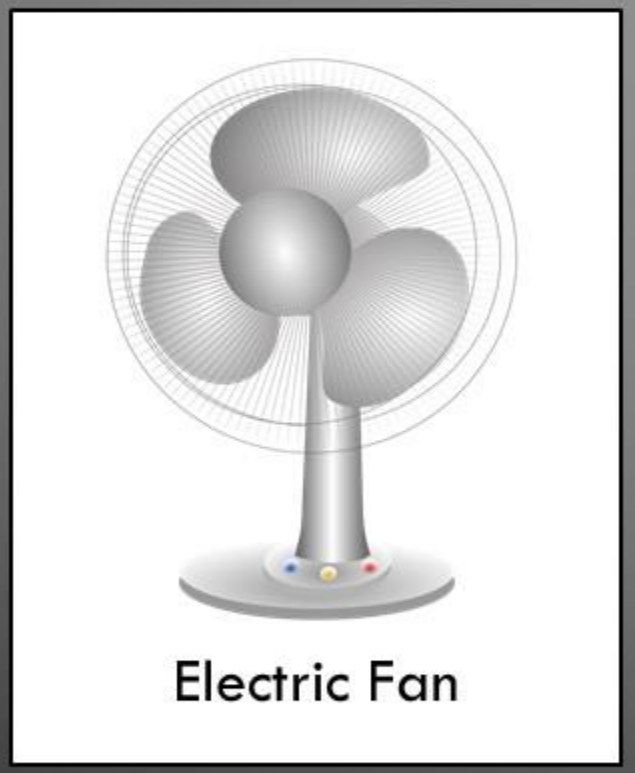
Change Behavior? ✓

Monitoring ✓

Measuring ✓

Modeling ✓

Non-ML Capable Examples



Electric Fan

Change Behavior? X

Monitoring X

Measuring X

Modeling X

Non-ML Capable Examples

Change Behavior? X

Monitoring X

Measuring X

Modeling X

WPI Project Summary

- Working Definitions
- AI Components and screening process
  - *Data source*
  - *Algorithm*
  - *Computation*
  - *Connection*
- ML Components and screening process
  - *Monitoring*
  - *Measuring*
  - *Modeling*

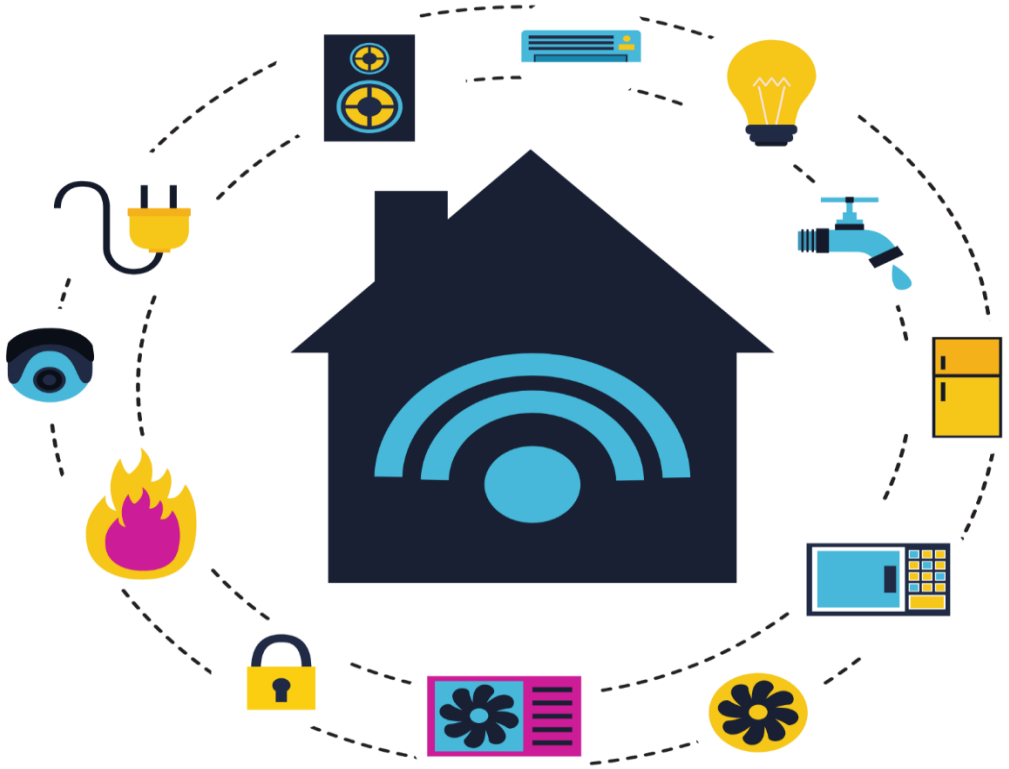
Implications of AI/ML in Consumer Products

Swathi Young
CTO, Integrity Management Services, Inc.
Women in AI Washington, D.C. Ambassador
Forbes Technology Council Member

Mission

“to protect the public from unreasonable risks of injury and death from consumer products.”

The future – Connected Home



“AI is the new electricity”

– Andrew Ng

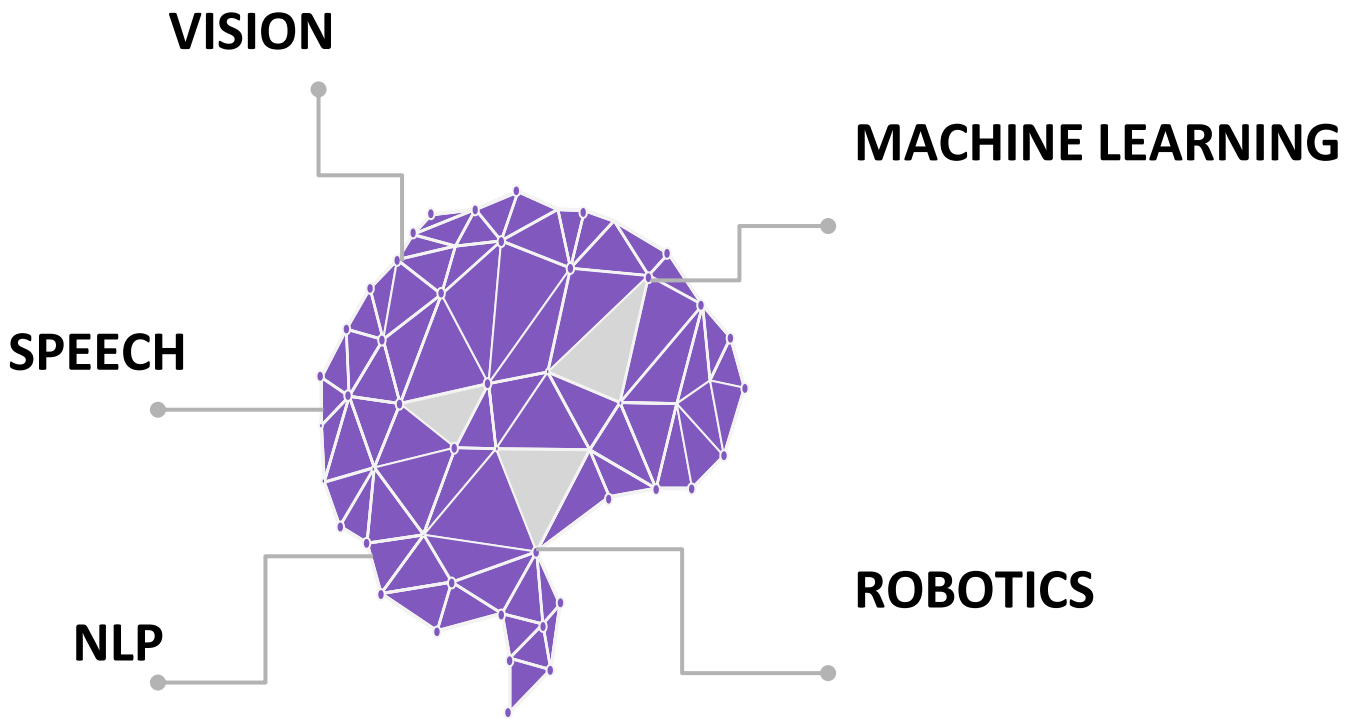
(Entrepreneur and Co-founder of Google Brain)

“WoW” experience for Consumers

What is Artificial Intelligence ?

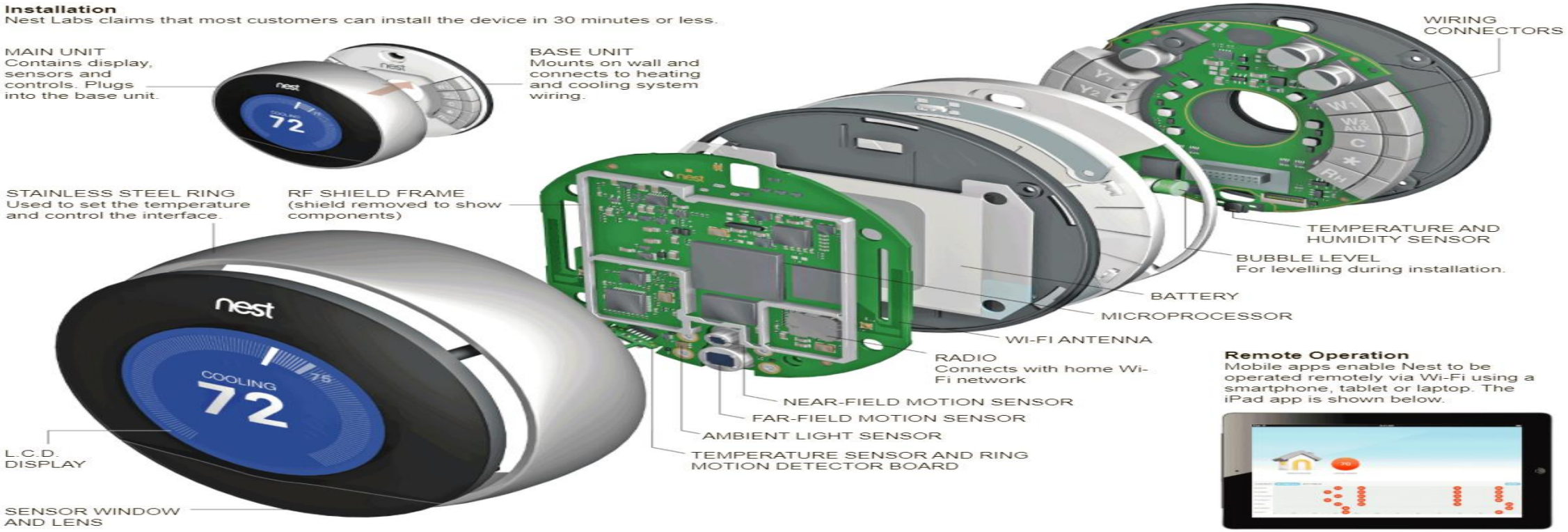
The term was coined by John McCarthy in 1956

"is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. It is an umbrella term whose sub-fields range from robotics to machine learning."



What is Machine Learning?

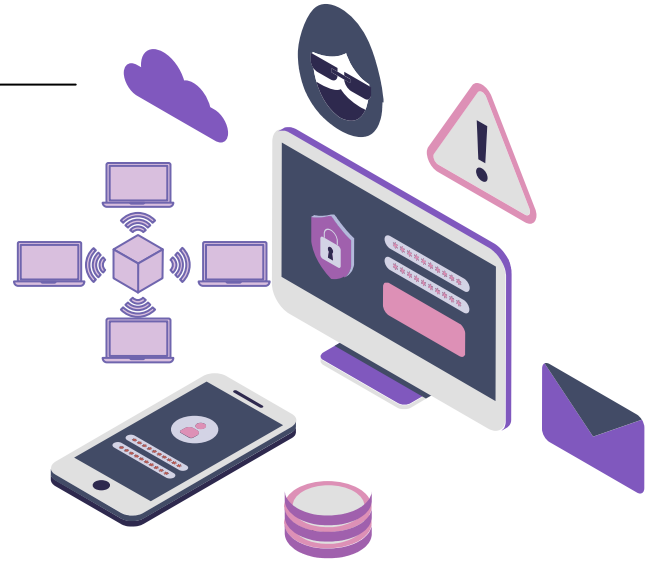
"AI that uses algorithms that automatically "learns" without being explicitly programmed. Machine learning is the application of AI techniques using statistical methods"



Source: NYTimes

Features of AI/ML Products

DATA



SENSORS

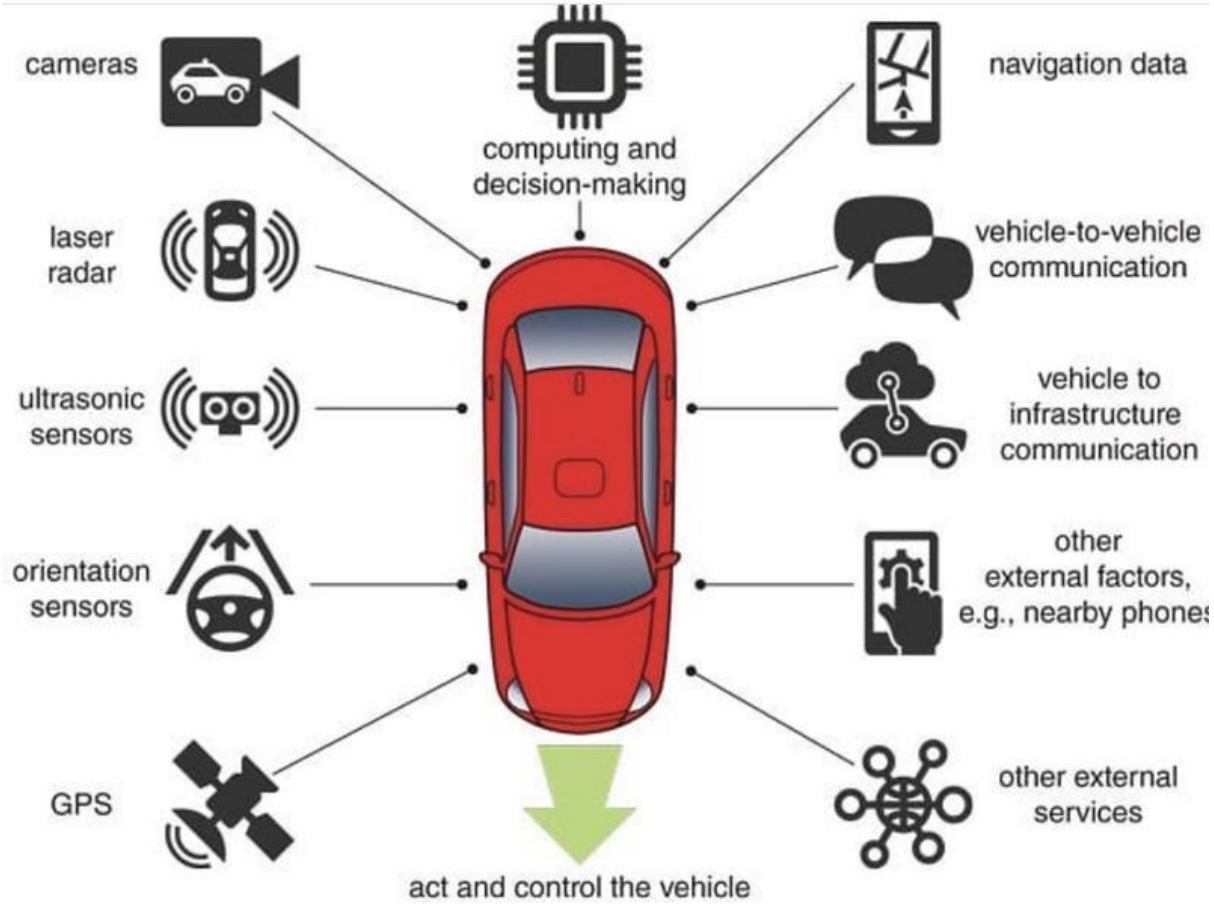
MICROPROCESSORS

WI-FI CONNECTIVITY

SOFTWARE

REMOTE OPERATIONS

AI driven functional Products

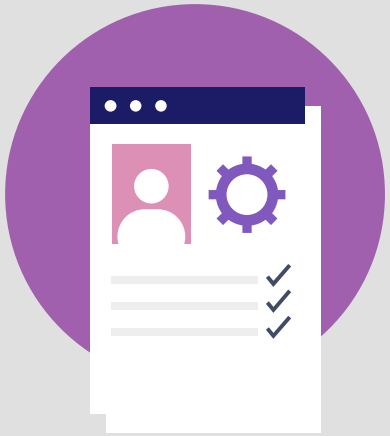


Autonomous Vehicle

Autonomous Vacuum Cleaner

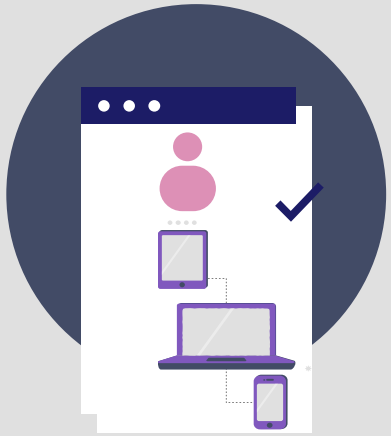
IMPLICATIONS OF AI/ML Products

PROVIDE a PERSONALIZED EXPERIENCE



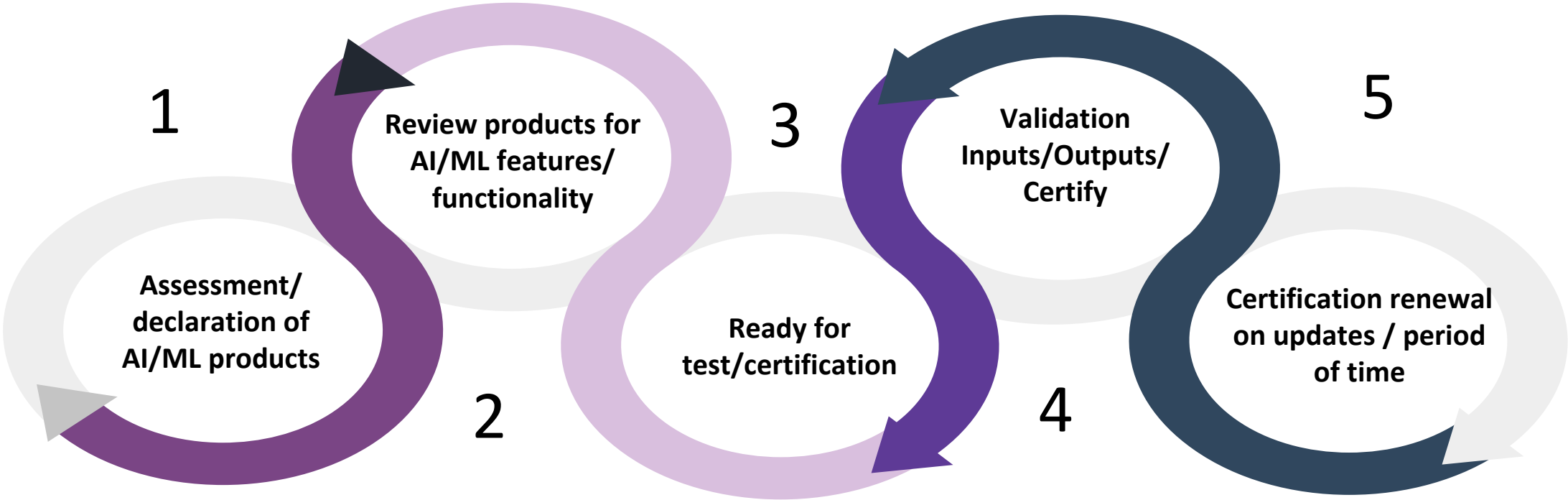
OPTIMIZE PERFORMANCE in REAL TIME

PROVIDE a PERSONALIZED EXPERIENCE



REMOTE OPERATIONS of PRODUCTS

STEPS FOR VALIDATING AI/ML PRODUCTS/FUNCTIONALITY



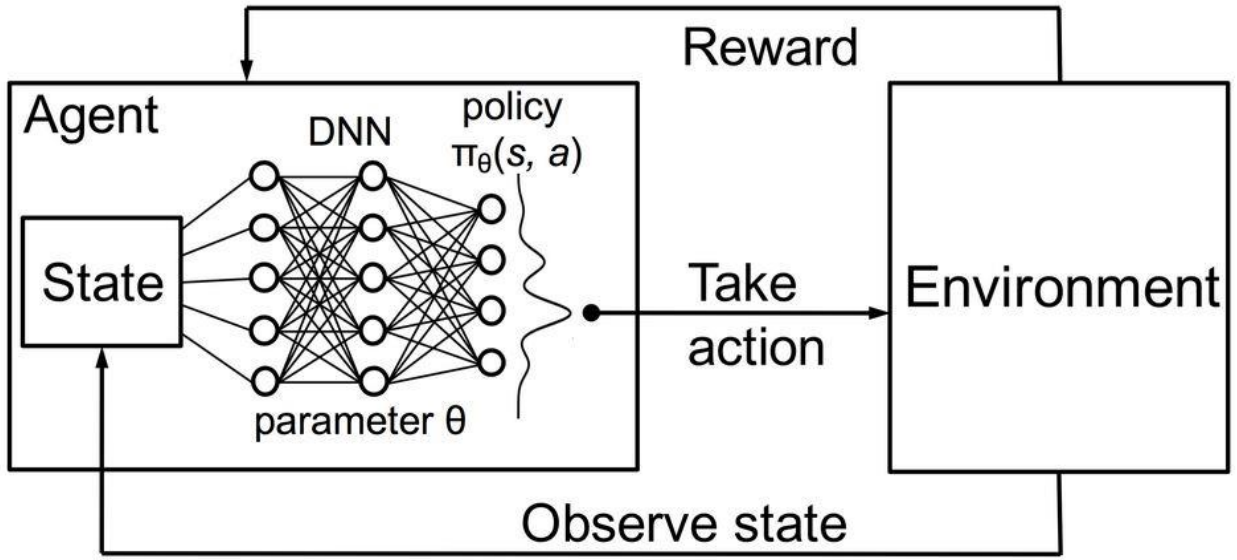
IntegrityM.com
LinkedIn and Twitter
@Swathiyoung

The AI/ML Show

THANK YOU!

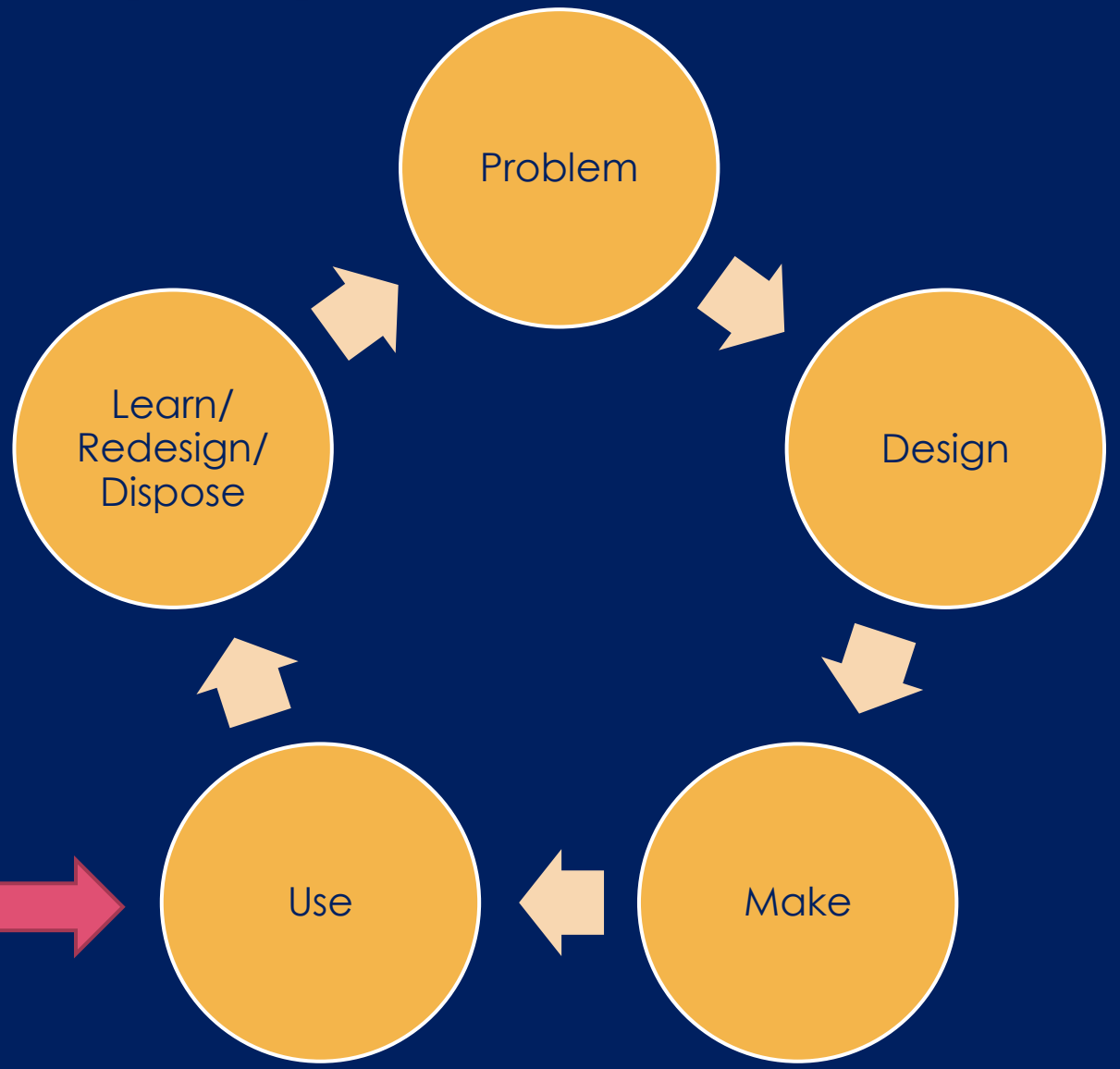
Impact of AI on Product Safety

Mahmood Tabaddor, PhD | CPSC Artificial Intelligence Forum | March 2, 2021



AGENT LEARNING MODEL

TECHNOLOGY INTELLIGENCE



MOST AI PRODUCTS TODAY

INTELLIGENCE OF TECHNOLOGY USAGE

CONSORTIUM FOR SAFER AI

Source: UL funded 2018 report by Dr. R Yampolskiy

SOME EXAMPLES OF AI PRODUCT FAILURES

1958 Advice software deduced inconsistent sentences using logical programming [13].
1959 AI designed to be a General Problem Solver failed to solve real world problems.
1977 Story writing software with limited common sense produced "wrong" stories [14].
1982 Software designed to make discoveries, discovered how to cheat instead.
1983 Nuclear attack early warning system falsely claimed that an attack is taking place.
1984 The National Resident Match program was biased in placement of married couples [15].
1988 Admissions software discriminated against women and minorities [16].
1994 Agents learned to "walk" quickly by becoming taller and falling over [17].
2001 AI agents learned to associate other AIs to food and virtually cannibalized them.
2005 Personal assistant AI rescheduled a meeting 50 times, each time by 5 minutes [18].
2006 Insider threat detection system classified normal activities as outliers [19].
2006 Investment advising software was losing money in real trading [20].
2007 Google search engine returns unrelated results for some keywords.
2010 Complex AI stock trading software caused a trillion dollar flash crash.
2011 E-Assistant told to "call me an ambulance" began to refer to the user as Ambulance.
2013 Object recognition neural networks saw phantom objects in particular noisy images [21].
2013 Google software engaged in name-based discrimination in online ad delivery [22].
2014 Search engine autocomplete made bigoted associations about groups of users [23].
2014 Smart fire alarm failed to sound alarm during fire.
2015 Automated email reply generator created inappropriate responses.
2015 A robot for grabbing auto parts grabbed and killed a man.
2015 Image tagging software classified black people as gorillas.
2015 Medical Expert AI classified patients with asthma as lower risk [24].
2015 Adult content filtering software failed to remove inappropriate content.
2015 Amazon's Echo responded to commands from TV voices.
2016 Linkedin's name lookup suggests male names in place of female ones.
2016 AI designed to predict recidivism acted racist.
2016 AI agent exploited reward signal to win without completing the game course.
2016 Passport picture checking system flagged Asian user as having closed eyes.
2016 Game NPCs designed unauthorized superweapons.
2016 AI judged a beauty contest and rated dark-skinned contestants lower.
2016 Smart contract permitted syphoning of funds from the DAO.
2016 Patrol robot collided with a child.
2016 World champion-level Go playing AI lost a game.
2016 Self driving car had a deadly accident.
2016 AI designed to converse with users on Twitter became verbally abusive.
2016 Google image search returned racists results.
2016 Artificial applicant failed to pass university entrance exam.
2016 Predictive policing system disproportionately targets minority neighborhoods.
2016 Text subject classifier failed to learn relevant features for topic assignment [25].
2017 AI for making inspirational quotes failed to inspire with gems like "Keep Panicking".
2017 Alexa played adult content instead of song for kids.
2017 Cellphone case designing AI utilized inappropriate images.
2017 Pattern recognition software failed to recognize certain types of inputs.
2017 Debt recovery system miscalculated amounts owed.
2017 Russian language chatbot shared pro-Stalinist, pro-abuse and pro-suicide views.
2017 Translation AI stereotyped careers to specific genders [26].
2017 Face beautifying AI made black people look white.
2017 Google's sentiment analyzer became homophobic and anti-Semitic.
2017 Fish recognition program learned to recognize boat IDs instead.
2017 Billing software sent an electrical bill for 284 billion dollars.
2017 Alexa turned on loud music at night without being prompted to do so.
2017 AI for writing Christmas carols produced nonsense.
2017 Autonomous cars had double the number of "fender-benders" of conventional cars [27].
2017 Apple's face recognition system failed to distinguish Asian users.
2017 Facebook's translation software changed Yampolskiy to Polanski
2018 Tesla Autopilot car crash
2018 Uber Self driving car kills pedestrian

GENERAL EMBEDDED SOFTWARE SYSTEM SAFETY CONCERNS

Hacking of safety critical functions (cybersecurity)

Loss of network connectivity affecting safety critical functions

Hardware (Sensor) failures

AI specific failures

Software updates

PRACTICAL AI SAFETY CONCERNS

- User is part of the product
- Algorithm method related failures
- Unintended Consequences
- Safety Transition Stage
- No Worst Case for Standards Testing
- Learning from Failures – the Black Box
- Scaling of Risk Profile

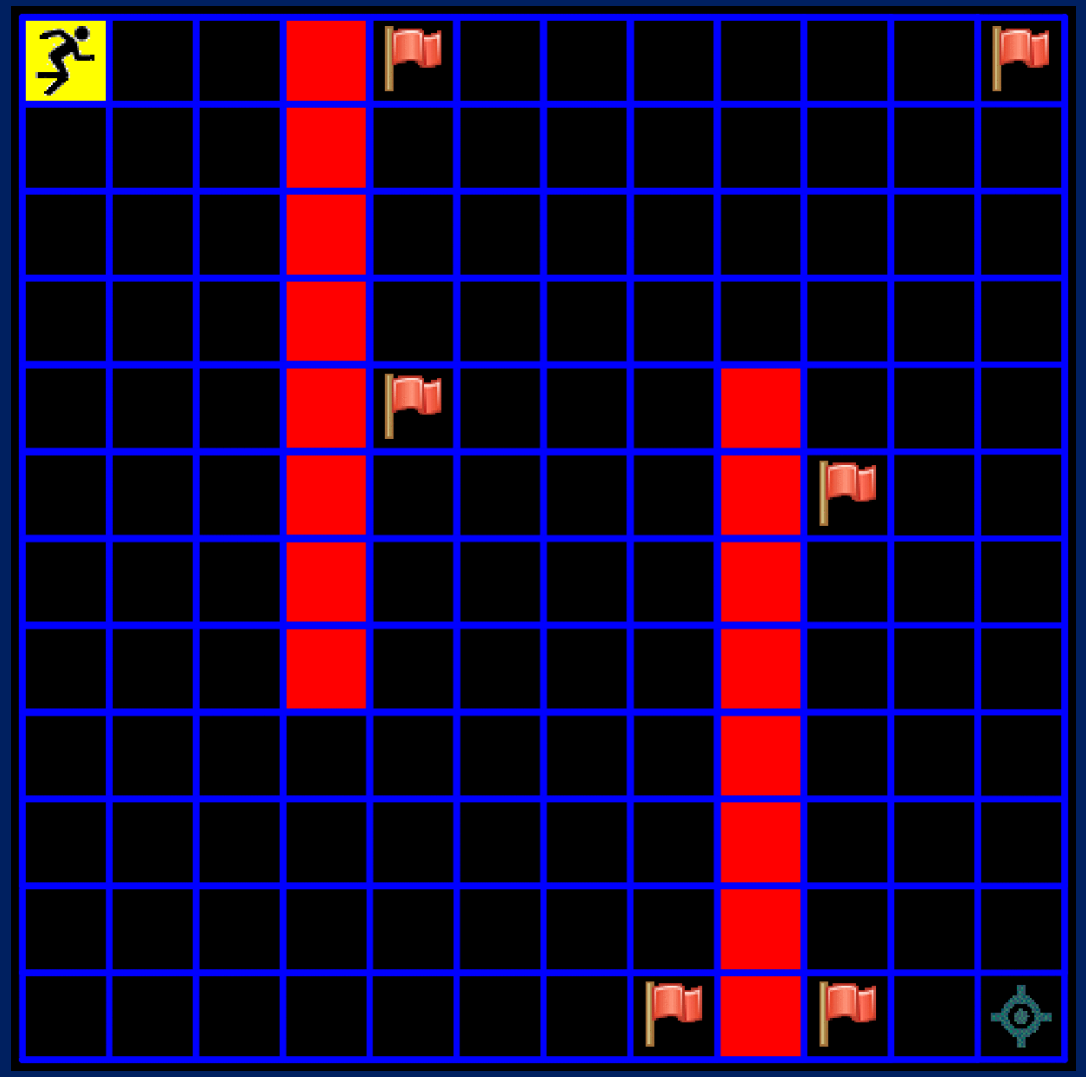
TESTING CHALLENGE

CHALLENGE FOR STANDARDS FOR AI

Operational Space

Learning on the Job

Sensing and Perception



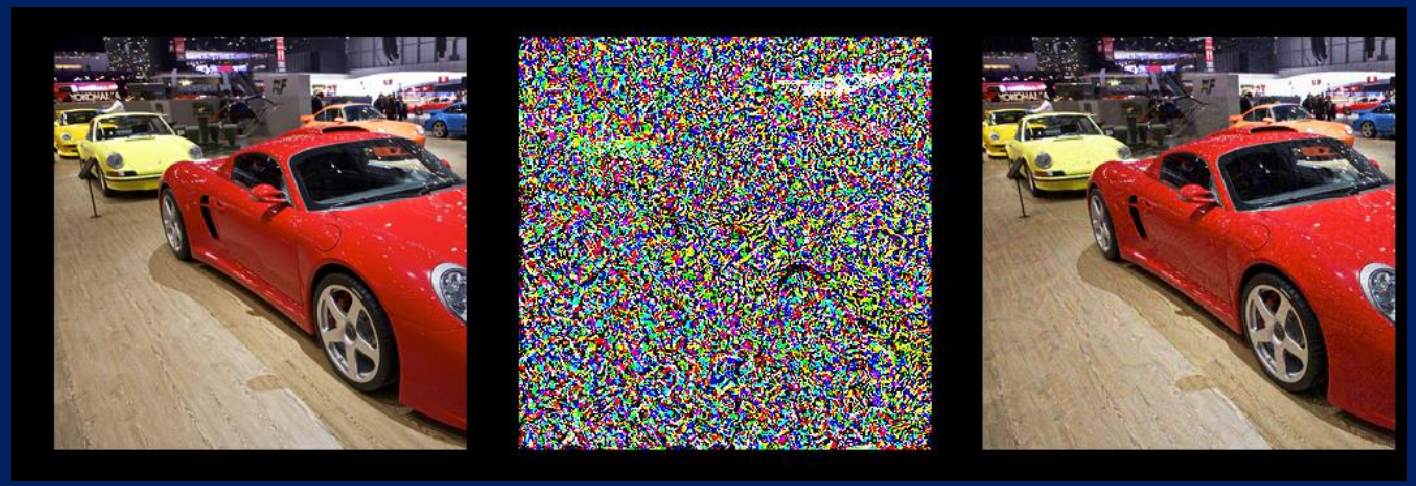
EXPLOITATION

VERSUS

SAFE
EXPLORATION

BRITTLENESS OF MACHINE LEARNING

Sports Car



Toaster

BRITTLENESS AS A ROUTE TO HACKING

RISK PROFILE

RISK = SEVERITY x PROBABILITY

RISK = f(Scaling in # of products and/or connectivity level and learning rate)

CHALLENGES OF AI SAFETY

Safety of AI enabled products will be susceptible to both traditional embedded software system safety issues along with AI specific risks

Interpretability of AI
Bias and Brittleness
Scaling and Emergent Behavior of AI - Multiagent
Augment vs. Automate

A NEW APPROACH

Develop a “narrative”, or safety case, to document and detail rationale to determine the DOMAIN OF TRUST for an intelligent product.

One possible model for product safety standards is **UL 4600: Standard for Autonomous Products**

THANK YOU

MAHMOOD.TABADDOR@UL.COM

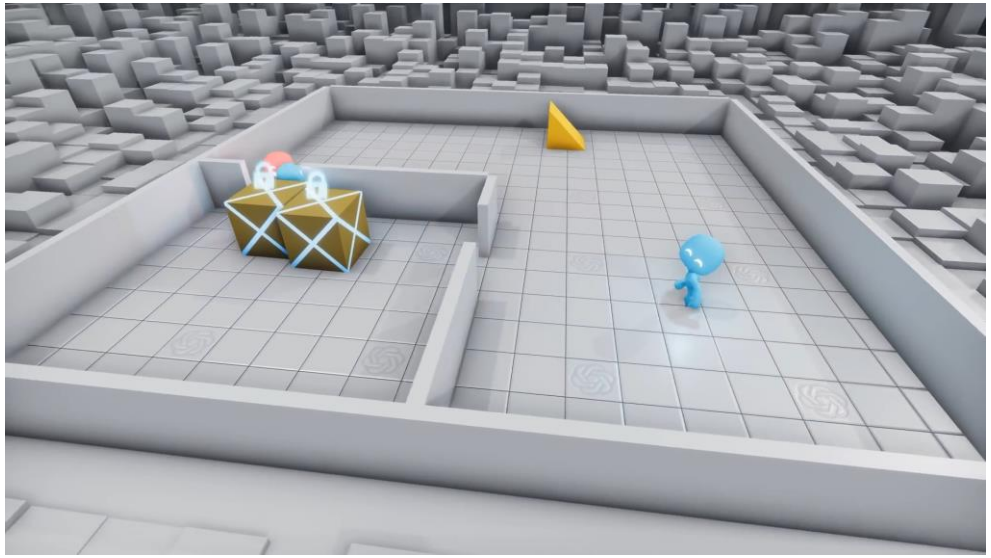
AI Iteration: What happens when AI keeps learning?

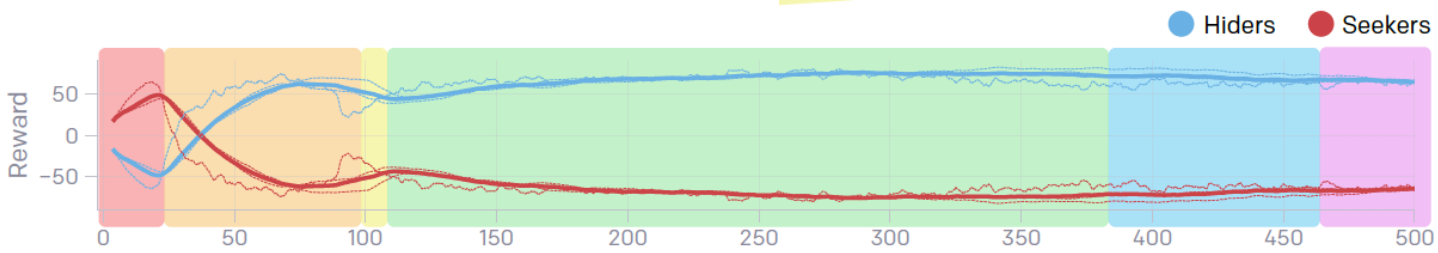
Gil Alterovitz, PhD, FACMI, FAMIA
Director, National Artificial Intelligence Institute

VA | National Artificial Intelligence Institute

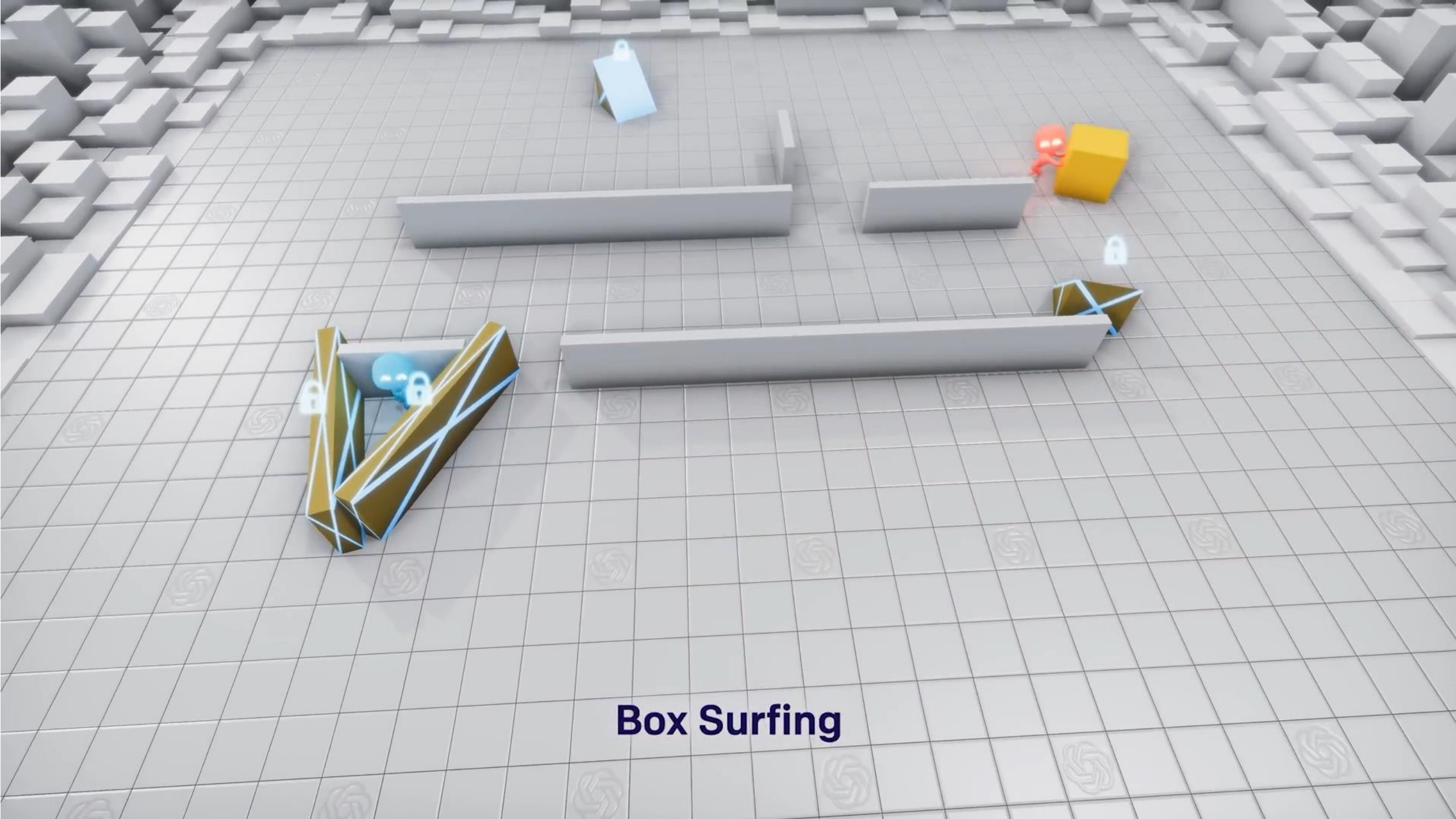
Multi-agent hide and seek example (Open AI)

- Normal hide and seek gives way to unusual iterations.
- Without regular monitoring, unusual situations can arise.
- After a while, unexpected‘‘ the ”seekers” can never find the “hiders.”






Shelter Construction



Box Surfing

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

November 17, 2020

THE DIRECTOR

M-21-06

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought
Director

SUBJECT: Guidance for Regulation of Artificial Intelligence Applications

https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

Executive Order – Promoting the Use of Trustworthy Artificial Intelligence in Federal Government

When designing, developing, acquiring, and using AI in the Federal Government, agencies shall adhere to the following Principles:

(a) Lawful and respectful of our Nation's values
(b) Purposeful and performance-driven
(c) Accurate, reliable, and effective
(d) Safe, secure, and resilient
(e) Understandable
(f) Responsible and traceable
(g) Regularly monitored
(h) Transparent
(i) Accountable

Iteration of AI and ML Capabilities

QUALIFY: *Baseline the origin to assess the current state*

QUANTIFY: *Benchmark the references to ascertain future state*

MONITOR: *Distinguish evolution of iterations to gauge progress*

MEASURE: *Discern transformation to grade the potential for safety mishap*

Is AI and ML consistently safe and reliable?

- **ITERATION:** *Does ML evolve and transform*
  - **DISCERN**: *Develop an Index to Monitoring/Measuring*
  - **DIFFERENTIATE**: *quantify and qualify safe capabilities*
    - Ascertain the evolution of AI and ML in consumer products
      - *How do they transform*
      - *When do they evolve*
    - Monitor/Measure maturation of AI & ML in Consumer Products
      - *Track the transformative effect of AI and ML in consumer products*
      - *Determine when the evolution of products evolves beyond safe operations*

Case study: FDA Proposed Regulatory Framework for Modifications to AI/Machine Learning-based Software as a Medical Device (SaMD)

- Specifically focused on AI/machine learning models where modifications can occur over time.
- AI/ML software can learn continuously, which necessarily induces algorithmic changes that are not part of the initial regulatory approval. The FDA thus seeks to develop a framework for approving such SaMDs and ensuring their ongoing safety as the software ingests additional patient data and subsequently alters it services.
- The FDA has only approved algorithms that are “locked” prior to marketing, meaning that the algorithms are fixed and do not adapt, e.g. as more data is accumulated.

FDA proposed a four-step process that builds on their Pre-Cert Program

1. **Qualify:** Maintain a culture of quality and organizational excellence where manufacturers of AI/ML-based SaMD adhere to Good Machine Learning Practices.
2. **Quantify:** Require initial premarket assurance of safety and effectiveness whereby a manufacturer would submit a "predetermined change control plan" for the FDA's initial premarket review, a description of the algorithm change protocol (ACP) that explains how the algorithm will learn and change throughout the lifecycle of a device, and eventual approval of the SaMD pre-specifications (SPS) and ACP.

FDA proposed a four-step process that builds on their Pre-Cert Program

3. **Monitor:** Allow for modifications after the initial review that takes into account risks to patients. If a modification is outside of the agreed upon SPS and ACP, but does not lead to a new intended use, then the FDA conducts a "focused review" of the proposed SPS and ACP. If a modification is beyond the initially authorized SaMD intended use, manufacturers may submit a new premarket submission.
4. **Measure:** Require manufacturers to report updates that were implemented as part of an approved SPS and ACP, as well as performance metrics for the SaMD. Manufacturers should also be transparent about notifying users about updates and monitoring the performance of their AI/ML-based SaMD.

Contact and More Information

Contact:

NAII@va.gov

Actionable AI in Health and Wellness – From Research to Practice:

https://www.research.va.gov/naii/Actionable-Health-AI.cfm