Data Science and AI Regulatory Applications Public Workshop

AI Characteristics for
Regulatory Consideration
September 19, 2023

Matt Dennis
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research

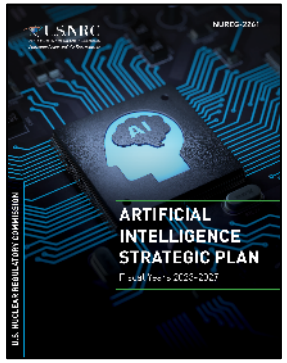
Outline

- Artificial Intelligence (AI) Landscape and the NRC
- AI Strategic Plan Development Background and Overview
- AI Characteristics for Regulatory Consideration
- Moving Forward and Stakeholder Engagement

Artificial Intelligence (AI) Landscape and the NRC

NUCLEAR INDUSTRY (EXTERNAL)

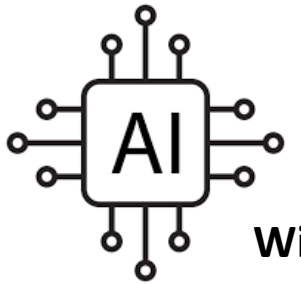
Industry wants to use AI



AI Strategic Plan to prepare staff to review AI

OTHER CONSIDERATIONS AND OPPORTUNITIES (EXTERNAL)

OMB EO 13960 and reporting requirements for implementing agencies



ACTIVITIES

Wide range of AI meetings, conferences, and activities

INTERNAL TO THE NRC

NRC Evidence Building Priority Questions

Internal interest in researching AI-based tools ranging from AI-embedded in commercial applications to custom programming

python™

AI Strategic Plan Development Background

- Formed an interdisciplinary team of AI subject matter experts (2021)
  - Insights gained from Data Science and Artificial Intelligence Regulatory Applications Workshops*
  - Engaged across the agency

- Proactively researching AI usage across the nuclear industry, Federal government, and international counterparts
  - Leveraging MOUs (e.g., EPRI and DOE)
  - Maintaining federal awareness (e.g., FDA and NIST)
  - International collaboration (e.g., CNSC, ONR and IAEA)

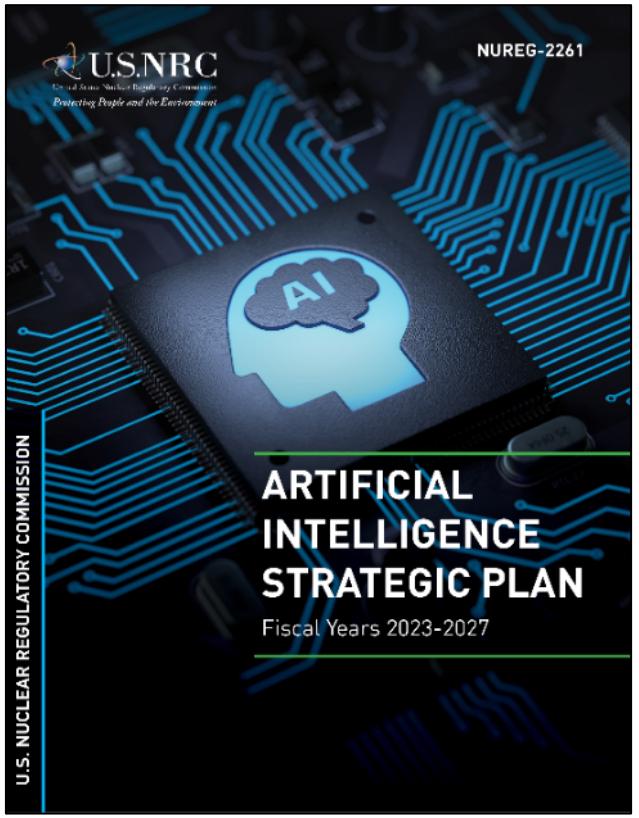
- Early stakeholder engagement and data gathering to execute the AI Strategic Plan
  - AI Strategic Plan comment-gathering public meeting (Summer 2022)
  - Internal seminars and training opportunities
  - Upcoming AI workshops

*Available at https://www.nrc.gov/public-involve/conference-symposia/data-science-ai-reg-workshops.html

NRC is not alone in considering AI and we're taking a proactive approach

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

AI Strategic Plan Overview



Draft Available at ML22175A206
Final available at ML23132A305

Vision and Outcomes

- Continue to keep pace with technological innovations to ensure the safe and secure use of AI in NRC-regulated activities
- AI framework and skilled workforce to review and evaluate the use of AI in NRC-regulated activities

The AI Strategic Plan consists of five strategic goals

- Goal 1: Ensure NRC Readiness for Regulatory Decisionmaking
- Goal 2: Establish an Organizational Framework to Review AI Applications
- Goal 3: Strengthen and Expand AI Partnerships
- Goal 4: Cultivate an AI-Proficient Workforce
- Goal 5: Pursue Use Cases to Build an AI Foundation Across the NRC

KEEPING THE END IN MIND – DETERMINING THE DEPTH OF REVIEW

Goal 1. Ensure NRC Readiness for Regulatory Decisionmaking

AI Research

- Determine approach to assess AI (e.g., XAI, trustworthiness, etc.)
- Development of AI standards and identify where gaps exists

Framework and Tools

- Clarify the process and procedures for AI regulatory reviews and oversight
- Consider options for long-range changes for AI regulatory reviews and oversight that may require rulemaking

Communications

- Public meetings to inform key activities

- Agency-wide internal communications and coordination to harmonize AI activities

Outcome: Develop an AI framework to review the use of AI in NRC-regulated activities

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Regulatory Considerations for AI Applications

- NRC AI Strategic Plan (NUREG-2261)
  - Table 1, “Notional AI and Autonomy Levels in Commercial Nuclear Activities” - notional framework to consider the levels of human-machine interaction with AI systems
  - Serves as a starting point in this public meeting to further discuss the variety of AI attributes which may affect regulatory considerations at each notional level
- AI Attributes Working Group
  - Formed May 2023 and includes members from agency offices
  - Paul Krohn, Matt Dennis, Trey Hathaway, Jonathan Barr, Reed Anzalone, Josh Kaizer, Dave Desaulniers, Jesse Seymour, Tanvir Siddiky, Joshua Smith, Scott Rutenkroger, David Strickland, and Howard Benowitz

Notional AI and Autonomy Levels in Commercial Nuclear Activities

Human Involvement

Level	Notional AI and Autonomy Levels	Potential Uses of AI and Autonomy in Commercial Nuclear Activities
Level 0	**AI Not Used**	No AI or autonomy integration in systems or processes
Level 1	**Insight** Human decision-making assisted by a machine	AI integration in systems is used for optimization, operational guidance, or business process automation that would not affect plant safety/security and control
Level 2	**Collaboration** Human decision-making augmented by a machine	AI integration in systems where algorithms make recommendations that could affect plant safety/security and control are vetted and carried out by a human decisionmaker
Level 3	**Operation** Machine decision-making supervised by a human	AI and autonomy integration in systems where algorithms make decisions and conduct operations with human oversight that could affect plant safety/security and control
Level 4	**Fully Autonomous** Machine decision-making with no human intervention	Fully autonomous AI in systems where the algorithm is responsible for operation, control, and intelligent adaptation without reliance on human intervention or oversight that could affect plant safety/security and control

Machine Independence

Common Understanding of the Level Key for Regulatory Readiness

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Disclaimer to AI Regulatory Considerations

- Considering NIST AI Risk Management Framework (RMF)* and other frameworks for future alignment
- The following AI characteristics and considerations for developing AI systems does not represent an exhaustive list of categories for consideration
- The following AI characteristics are defined by a range of implementation levels that may impact regulatory decision-making

*NRC has not endorsed using the NIST AI RMF as means to meet current or future regulation

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

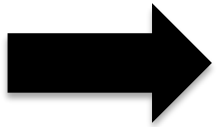
AI Characteristics for Regulatory Consideration

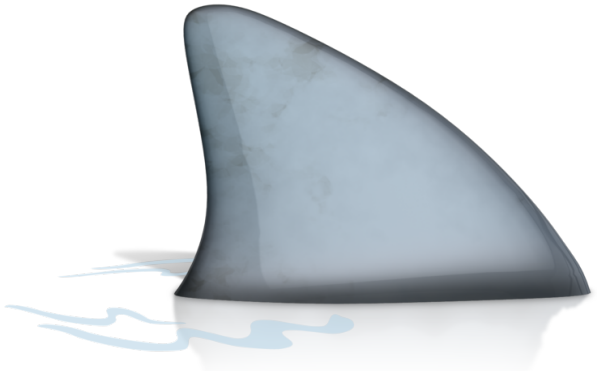
Safety Significance	**AI Autonomy**	**Security**	**Explainability**
Model Lifecycle	**Regulated Activity**	**Regulatory Approval**	**Application Maturity**

Safety Significance

- What is the safety significance of the use of AI?
- Safety Principles using Risk or Determinism – In the absence of the ability to quantify risk, there are good engineering principles (e.g., defense-in-depth) that can be used to guard against unintended consequences.
- Failure and Consequence Identification – A first step as part of AI systems engineering, a formalized process to quantify the hazards and modes of operation can be considered to ensure adequate system design.

No impact on safety or implemented safety functions

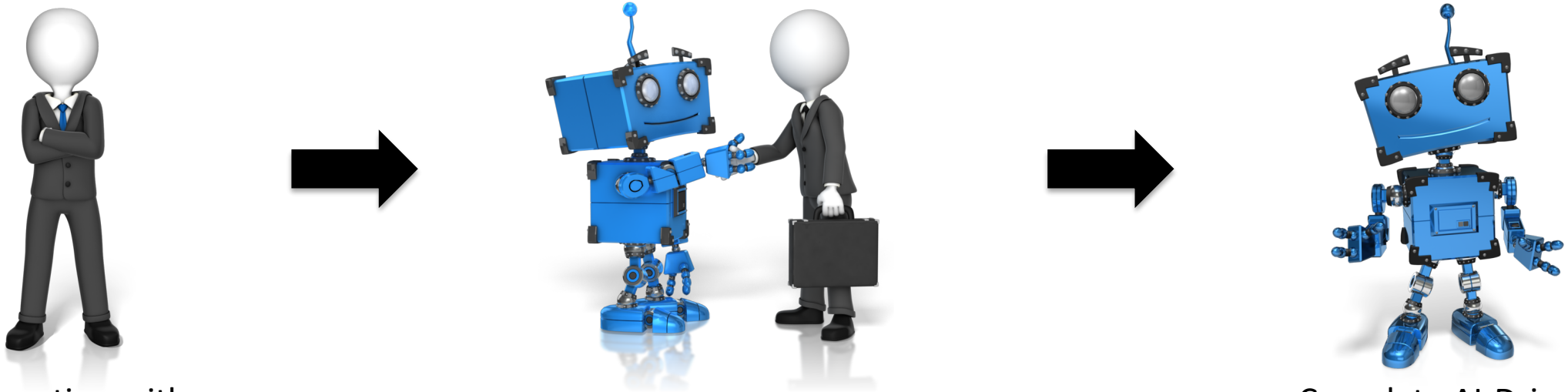




Potential consequences with significant safety implications

AI Autonomy

- Transition point exists where AI controls the process without human intervention
- A graded approach which considers a variety of AI characteristics may determine the level of regulatory review required



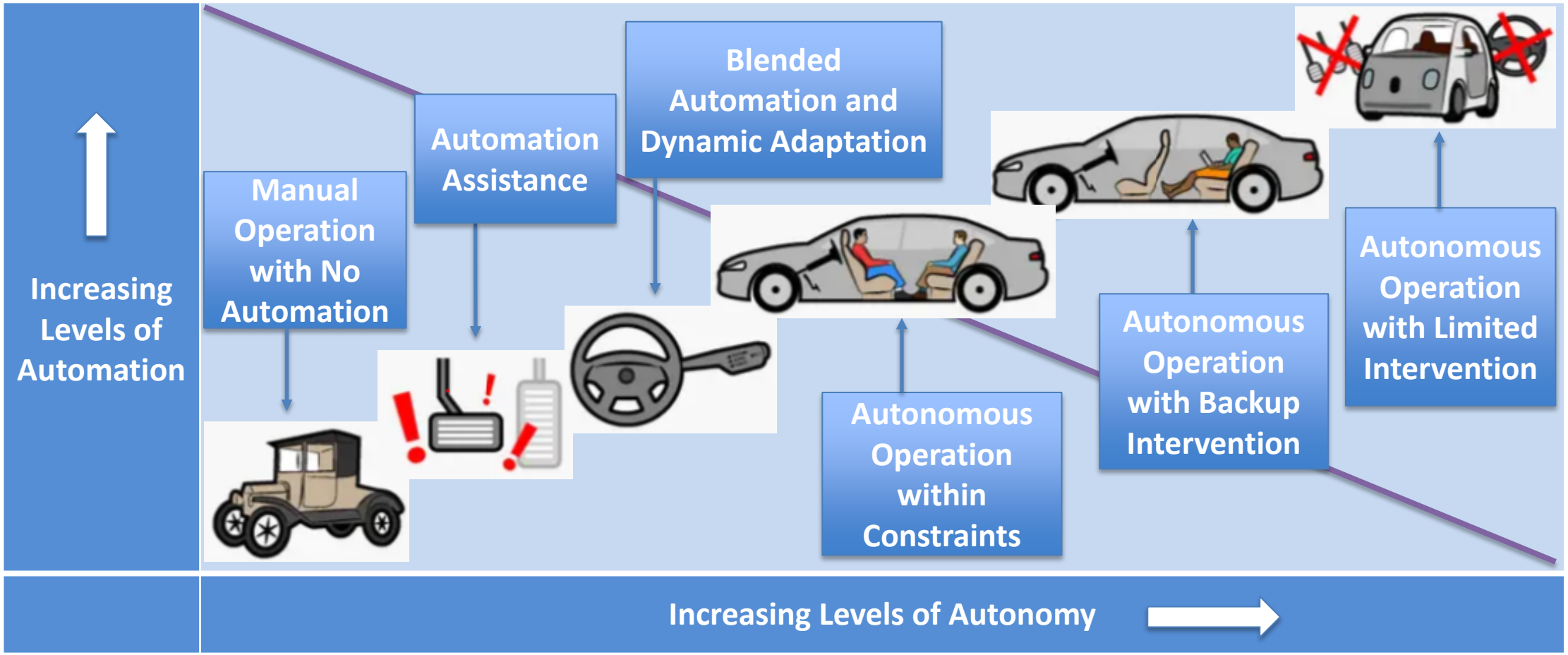
Automation with No AI Utilized

Complete AI-Driven Autonomy

Clarifying Automation, Autonomy, and AI

- AI technologies can enable autonomous systems
  - Not all uses of AI are fully autonomous as many may be used to augment human decision-making rather than replace it.
  - Higher autonomy levels indicate less reliance on human intervention or oversight and, therefore, may require greater regulatory scrutiny of the AI system.
- Multiple definitions exist; however, it is important to have a clear understanding of the differences between automation and autonomy
  - Automation - considered to be a system that automatically acts on a specific task according to *pre-defined, prescriptive rules*. For example, reactor protection systems are automatically actuated when process parameters exceed certain defined limits.
  - Autonomy - a set of intelligence-based capabilities that *allows the system to respond to situations that were not pre-programmed or anticipated* (i.e., decision-based responses) prior to system deployment. Autonomous systems have a degree of self-governance and self-directed behavior resulting in the ability to compensate for system failures without external intervention.

AI applied to Automation and Autonomy



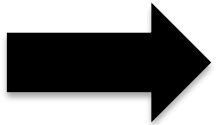
Graphics source: https://www.businessinsider.com/what-are-the-different-levels-of-driverless-cars-2016-10

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Security

- Can others influence the AI?
- Open-Source Tools – Use of open-source tools are not precluded, but using non-specialized software solutions means that there are steps taken to rigorously confirm the safety and security of the implemented solution.

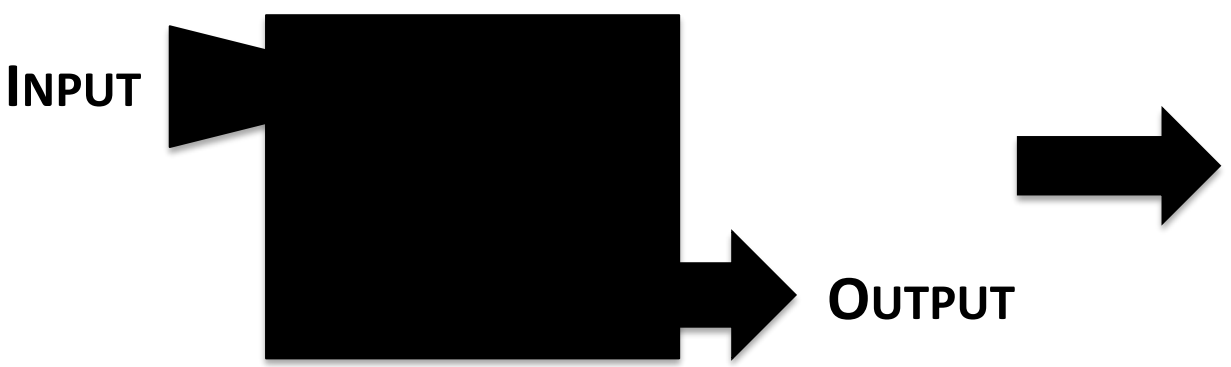
Open access to model, data, and code



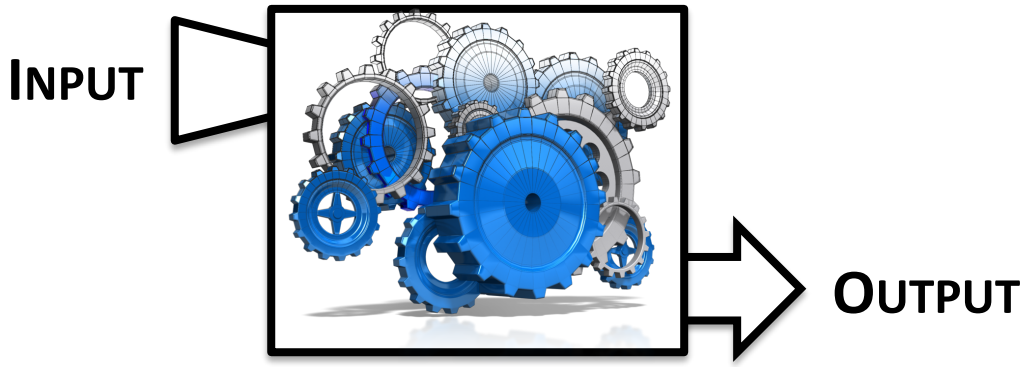
Closed access and fully isolated

Explainability

- To what degree do we understand how the AI is working?
- Establishing a Trustworthy System – Explainability exposes a chain of decision-making for potentially complex logic that is easily interpretable by anyone unfamiliar with the AI system design. This applies to all stakeholders which include reviewers (e.g., regulators) as well as system users.



Black Box AI System

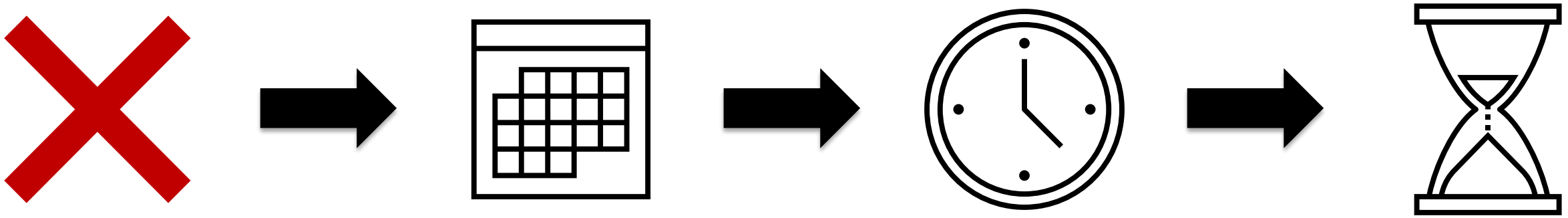


Visibility into the What, How, and Why within an AI System

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Model Lifecycle

- How often the AI is updated and maintained?
- Data Provenance – Based on a graded approach, the modeling data may have a variety of various pedigrees based on the application area (e.g., safety significance).
- Model Updating – Models need to be maintained to avoid performance degradation and kept consistent with the pre-determined change control and notification process for that application.



Frozen or Locked Model

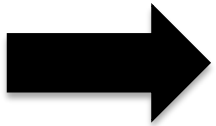
Continuous Updating

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Regulatory Activity

- Is AI being used in a regulated activity?
- Human and Organizational Factors – The context of operation needs to consider the handover to human operation, immediacy for human action, or if placement in a safe stable state is required based on the operational context.

Application Domain
Outside Regulated Activity



AI Supports
Regulated Activity

Regulatory Approval

- What is the level of regulatory approval required?
- Extensive Application Areas - A variety of regulatory requirements apply to various potential AI application areas. Existing requirements may range from evaluation of sufficient functional performance up to specific requirements to ensure AI system safety and security.

Performance Requirements

Prescriptive Requirements for Methods or Approaches

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

AI Maturity

- Is AI commonly used in this way?
- Existing Guidance - Traditional safety, security, software, and systems engineering practices are still applicable as the starting point for good engineering practice.

Novel AI Application with Minimal Experience

Commonplace AI Application with Extensive Usage

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Summary Considerations (1/2)

- Existing Guidance – Traditional safety, security, software, and systems engineering practices are still applicable as the starting point for good engineering practice.
- Establishing a Trustworthy System – Explainability exposes a chain of decision-making for potentially complex logic that is easily interpretable by anyone unfamiliar with the AI system design. This applies to all stakeholders which include reviewers (e.g., regulators) as well as system users.
- Safety Principles using Risk or Determinism – In the absence of the ability to quantify risk, there are good engineering principles (e.g., defense-in-depth) that can be used to guard against unintended consequences.
- Open-Source Tools – Use of open-source tools are not precluded, but using non-specialized software solutions means that there are steps taken to rigorously confirm the safety and security of the implemented solution.

Summary Considerations (2/2)

- Failure and Consequence Identification – A first step as part of AI systems engineering, a formalized process to quantify the hazards and modes of operation can be considered to ensure adequate system design.
- Data Provenance – Based on a graded approach, the modeling data may have a variety of various pedigrees based on the application area (e.g., safety significance).
- Model Updating – Models need to be maintained to avoid performance degradation and kept consistent with the pre-determined change control and notification process for that application.
- Human and Organizational Factors – The context of operation needs to consider the handover to human operation, immediacy for human action, or if placement in a safe stable state is required based on the operational context.
- Extensive Application Areas – A variety of regulatory requirements apply to various potential AI application areas. Existing requirements may range from evaluation of sufficient functional performance up to specific requirements to ensure AI system safety and security.

NRC AI Considerations

Current

Traceable and Auditable Evaluation Methodologies

Understanding Licensee and Applicant AI Usage

Future

Regulatory Guidance and Decision-Making Development

Differentiating AI Usage for Reactor Design Versus Autonomous Control

Explainable AI and Trustworthy AI – Reliability and Assurance

Internal AI Budget Predicated on Emergent Industry Applications

Moving Forward and Stakeholder Engagement

- Continued safety and security in the nuclear industry is paramount
- Embrace new and innovative ways to meet NRC's mission
- Maintain strong partnerships with domestic and international counterparts
- Engage with the NRC early and often on plans and operating experience

Future Activities

- Advisory Committee on Reactor Safeguards subcommittee meeting on AI (November 15, 2023)
- Regulatory framework applicability assessment of artificial intelligence in nuclear applications (Summer 2023-Spring 2024)

Contact Information

- **Matt Dennis**
  Data Scientist
  Office of Nuclear Regulatory Research
  matthew.dennis@nrc.gov

- **Luis Betancourt, P.E.**
  Chief, Accident Analysis Branch
  Division of Systems Analysis
  Office of Nuclear Regulatory Research
  luis.betancourt@nrc.gov

- **Victor Hall**
  Deputy Division Director
  Division of Systems Analysis
  Office of Nuclear Regulatory Research
  victor.hall@nrc.gov

BACKUP SLIDES

Acronyms

- AI – Artificial Intelligence
- AICoP – Artificial Intelligence Community of Practice
- AISC – Artificial Intelligence Steering Committee
- CNSC – Canadian Nuclear Safety Commission
- DOE – U.S. Department of Energy
- EO – Executive Order
- EPRI – Electric Power Research Institute
- FDA – U.S. Food and Drug Administration
- FRN – Federal Register Notice
- FY – Fiscal Year
- GAO – U.S. Government Accountability Office
- GSA – U.S. General Services Administration
- IAEA – International Atomic Energy Agency
- IEC – International Electrotechnical Commission
- ML – Machine Learning
- MOU – Memorandum of Understanding
- NLP – Natural Language Processing
- NRC – U.S. Nuclear Regulatory Commission
- OMB – U.S. Office of Management and Budget
- ONR – U.K. Office for Nuclear Regulation
- NEI – Nuclear Energy Institute
- NIST – National Institute of Standards and Technology
- XAI – Explainable Artificial Intelligence

Other Regulatory and Risk Management Approaches

- United Kingdom AI Regulation: A Pro-Innovation Approach
- European Union AI Act
- U.S. Food and Drug Administration AI Regulatory Framework for Medical Devices
- U.S. Department of Health and Human Services Trustworthy AI Playbook
- U.S. National Institute of Standards and Technology AI Risk Management Framework
- U.S. Department of Energy AI Risk Management Playbook

Additional AI References

- United Kingdom AI Standards Hub
- United Kingdom Centre for Data Ethics and Innovation (CDEI) AI Assurance Techniques
- OECD AI Policy Observatory
- Partnership on AI
- AI Incident Database