SUNI Review Complete
Template=ADM-013
E-RIDS=ADM-03

ADD: John Lane, Mary Neely
Comment (6)
Publication Date: 4/21/2021
Citation: 86 FR 20744

PUBLIC SUBMISSION

As of: 5/25/21 12:53 PM
Received: May 21, 2021
Status: Pending_Post
Tracking No. koy-lpko-yzfw
Comments Due: May 21, 2021
Submission Type: API

Docket: NRC-2021-0048
Role of Artificial Intelligence Tools in Nuclear Plant Operations

Comment On: NRC-2021-0048-0001
Role of Artificial Intelligence Tools in U.S. Commercial Nuclear Power Operations

Document: NRC-2021-0048-DRAFT-0005
Comment on FR Doc # 2021-08177

Submitter Information

Email: Ryan@forhumanity.center
Organization: ForHumanity

General Comment

The uploaded document contains ForHumanity's response to the Request for Comment on "Role of Artificial Intelligence Tools in U.S. Commercial Nuclear Power Operations."

Attachments

ForHumanity NRC Response May 2021

Comments of

ForHumanity[1]

Ryan Carrier, *Executive Director*
Mark Potkewitz, *General Counsel*

In the Matter of

Role of Artificial Intelligence Tools in U.S. Commercial Nuclear Power Operations

Request for comment

Docket No. NRC-2021-0048

May 21, 2021

[1] ForHumanity (https://forhumanity.center/) is a 501(c)(3) nonprofit organization dedicated to addressing the Ethics, Bias, Privacy, Trust, and Cybersecurity in artificial intelligence and autonomous systems. ForHumanity uses an open and transparent process that draws from a pool of over 350+ international contributors to construct audit criteria, certification schemes, and educational programs for legal and compliance professionals, educators, auditors, developers, and legislators to mitigate bias, enhance ethics, protect privacy, build trust, improve cybersecurity, and drive accountability and transparency in AI and autonomous systems. ForHumanity works to make AI safe for all people and makes itself available to support government agencies and instrumentalities to manage risk associated with AI and autonomous systems.

Introduction and Summary

Artificial Intelligence (AI), including machine learning, statistical and Bayesian approaches, expert systems, reinforcement learning, and autonomous systems, possesses tremendous potential but presents a concomitant level of risk. Current general approaches to the development of AI systems often fail to account for issues related to ethics, bias, trust, privacy, and cybersecurity in their development, deployment, use, and maintenance. Persons looking to use an AI system should ensure that they understand the specific risks associated with that particular system, including the myriad examples of ethical choice embedded in the design and development of systems. The greater the potential impact on humans, human agency, living creatures and the environment, the more exhaustive and exacting the scrutiny and analysis that should be placed on those systems.

The spread of AI tools across various sectors and industries has not increased awareness of the risks associated with these tools. ForHumanity examines the application of AI or autonomous systems when they present a systemic risk to humans, the environment or societal systems. While most systems are believed to be beneficial, industries must demonstrate that this belief is warranted by building trust in those affected or potentially affected by these systems. The introduction of a robust governance system that embeds human agency, governance, oversight, accountability and thorough risk mitigations builds trust. Combined with certain advancements in laws and regulations in the areas of ethics, bias, privacy, trust and cybersecurity, industries can, through transparency, accountability and independent verification, responsibly incorporate AI and autonomous systems into their quotidien systems and practices.

This submission, highlighting the work of 350+ ForHumanity Contributors, explains the risks from these systems and proposes an industry-oriented solution deploying a systemic risk-based approach with transparency and compliance-by-design construction executed throughout the lifecycle of an algorithmic system uniformly across the industry, yet tailored to each individual AI/ML or autonomous system.

Background on Independent Audit

In 1973, the major accounting firms came together and formed The Financial Accounting Standards Board (FASB) which created the Generally Accepted Accounting Principles (GAAP) which still govern financial accounting today. Eventually, Securities and Exchange Commission, and other extranational regulatory agencies, required adherence to the GAAP standard for all publicly listed companies. This clarity and uniformity significantly improved the financial world. An infrastructure of trust has been built over the past 50 years because of critical features such as independence, certified practitioners, and third-party rules that are compliant with the law and best-practices.

ForHumanity has advocated for the adoption of this infrastructure of trust and explained how it can be adapted and adopted for the Governance, Accountability and Oversight of AI and Autonomous Systems.[2] We support the creation and mandate of Independent Audit of AI Systems (IAAIS).

Role on Independent Audit of AI and Autonomous Systems

IAAIS provides a comprehensive solution grounded in the same fundamental principles as Independent Financial Audit.[3] ForHumanity develops and maintains audit and certification criteria designed for a range of industries and jurisdictions.

The proposed system replicates the distributed oversight, accountability and governance needed for AI and autonomous systems in the same manner as financial audit, through audit and pre-audit service providers. These entities will employ certified practitioners to prepare for an eventual independent audit performed by other certified practitioners. The audit criteria are presented transparently to maximize an entity's ability to achieve compliance. Advancements in systems technology allow many of these processes to be automated for entities such as with the Treadway Commissions' Committee of Sponsoring Organization (COSO) framework for internal risk, audit and controls. The result is a fully-integrated, compliance-by-design infrastructure that embeds human agency, transparency, disclosure and compliance from design to decommission.

The audit criteria are applied in two vectors: 1) Top-down accountability, governance and oversight 2) laterally, AI system by AI system. The top-down approach creates accountability systems for ethics, bias, privacy, trust, and cybersecurity for the Board of Directors, Chief Executive Officer and Chief Data Officer. Committee structures are required such as an Algorithmic Risk, Children's Data Oversight, and Ethics to manage the audit/compliance responsibilities. All of these top-down criteria apply to every AI and every autonomous system in the organization. The system-specific audit criteria is designed to ensure legal and best practice compliance tailored to the specific impact of each system on humans. This comprehensive approach ensures consistency across the organization combined with complete risk management coverage of each unique system.

The creation and maintenance of the Independent Audit of AI Systems is an ongoing and dynamic process. It will continue to be fully transparent to all who would choose to participate, provided they join the discussion and participate with decorum. To create each set of audit criteria, ForHumanity engages an international group of experts and seeks points of consensus on its auditable rules. The rules are completely transparent, so when an audit is conducted, compliance is expected.

---

[2] https://forhumanity.center/blog/auditing-ai-and-autonomous-systems-building-an-infrastructureoftrust

[3] For more information about the taxonomy of IAAIS, see Ryan Carrier & Shea Brown: Taxonomy: AI Audit, Assurance Assessment, Feb. 2021. *Available at:* https://static1.squarespace.com/static/5ff3865d3fe4fe33db92ffdc/t/60329e0a4cfbaa172691f7e6/1613929999802/Taxonomy+of+AI+Audit+%282%29.pdf

Independent auditors verify compliance and remain liable for false assurance. Audits must be performed by certified practitioners.

Audit Rules

IAAIS Audit Rules have the following characteristics:

These characteristics prove vital for a variety of reasons. Ambiguous audit criteria only encourage auditors to take a more risk-averse approach and presume noncompliance when faced with non-binary choices. Good audit rules must provide the auditor with binary criteria such that certain elements are either compliant or not compliant. The Auditor remains liable for the final report which will either certify compliance or indicate noncompliance. No entity can be certified by an Auditor as partially compliant.

All of these rules must be implementable. Industry can feed into the creation of the rules to ensure that these rules can be followed. In fact, these rules will likely be built into the systems over time for compliance-by-design.

Risks and Pitfalls

ForHumanity is a mission driven non-profit organization. That mission is *To examine and analyse the downside risks associated with the ubiquitous advance of AI & Automation, to engage in risk mitigation and ensure the optimal outcome… ForHumanity*. As a result of that mission, we are uniquely positioned to aid the Nuclear Regulatory Commission and the industry as a whole to manage these risks. The organizations that design, develop, promote and sell AI/ML tools manage the upside and benefits. Our approach is one of risk control, mitigation and management. Proper

management of downside risks generates better results for everyone. To that end, we have identified five key areas of risk to humans/citizens from applications in the nuclear industry:

1) Ethics
2) Bias
3) Privacy
4) Trust
5) Cybersecurity

We have developed a transparent, crowdsourced service model for governments, regulators and authorities. We craft audit rules and criteria, submitting them to authorities for approval and training individual auditors. We license qualified entities to engage in audits or pre-audit compliance work using authority approved criteria. The NRC is welcome to accept and deploy the same service.

1. What is the status of the commercial nuclear power industry development or use of AI/ML tools to improve aspects of nuclear plant design, operations or maintenance or decommissioning? What tools are being used or developed? When are the tools currently under development expected to be put into use?

ForHumanity's audit criteria requires governance, oversight and accountability throughout the lifecycle of the AI/ML or algorithmic system, regardless of the system as long as it impacts a human. As nuclear power represents an existential risk to neighboring people, all systems fall into this category. AI/ML and autonomous systems must be trustworthy by design and that means mitigation of risk from cyberattack, for which the industry is well aware. What the industry might be less aware of are the risks associated with bias, or more notably data poisoning attacks, broken into two types: 1) Data inputs 2) Training Poisoning. Deepfakes (written, audio or visual) represent a meaningful concern for monitor systems. Each of these forms of attack, often outside the realm of the traditional "cyberattack" represent potential for catastrophic consequences if left unmitigated as they can result in the AI/ML or autonomous system being turned against itself to create either anticipated actions (which can be abused) or outright false steps based upon security protocols which may harm the system or people.

ForHumanity argues for third-party, independent audits to verify compliance with procedures designed to track, test, measure, and potentially disclose compliance and satisfaction of both legal and best practice implementation to most effectively manage these risks.

2. What areas of commercial nuclear reactor operation and management will benefit the most, and the least, from the implementation of AI/ML? Possible examples include, but are not limited to, inspection support, incident response, power generation, cybersecurity, predictive maintenance, safety/risk assessment, system and component performance monitoring, operational/ maintenance efficiency and shutdown management.

ForHumanity deals exclusively in downside risk management and mitigation. The systems of accountability, governance and oversight enable a virtuous cycle of feedback, through transparency and expected compliance that raises the floor of governance and ensures regular examination of compliance. All systems should be governed by tailored and system specific audit criteria as well as hierarchical accountability structures that include transparency, disclosure and document annual compliance.

3. What are the potential benefits to commercial nuclear power operations of incorporating AI/ML in terms of (a) design or operational automation, (b) preventive maintenance trending, and (c) improved reactor operations staff productivity?

ForHumanity built Independent Audit of AI Systems to rigorously implement oversight, governance and accountability for all systems. Given the sensitive nature of the nuclear industry, it should require compliance-by-design covering ethical uses, bias mitigation, privacy-by-design, control, safety, transparency and robust cybersecurity requirements which serve as the best means to manage risks associated with all of these systems ranging from poor-design, to negligent governance. Unique questions of control and safety exist for some systems that may operate without continuous human input and only periodic human oversight. Machine learning strategies when not properly administered can lead to amplifications of certain undesirable outcomes at the cost of safety or security to unintended systems or people. In other words, AI can drastically exacerbate negative externalities if not properly managed. Diverse inputs and multi stakeholder feedback risk assessments can help to uncover risk or unknown unknowns. No system is foolproof, but when it comes to the NRC, we know that brakes, seatbelts and airbags are often preferred, overlapping safety protocols. There are brakes, seatbelts and airbags throughout the lifecycle of the algorithm in Independent Audit of AI Systems.

4. What AI/ML methods are either currently being used or will be in the near future in commercial nuclear plant management and operations? Example of possible AI/ML methods include, but are not limited to, artificial neural networks, decision trees, random forests, support vector machines, clustering algorithms, dimensionality reduction algorithms, data mining and content analytics tools, gaussian processes, Bayesian methods, natural language processing, and image digitization.

New systems represent new challenges and require increased thoughtfulness on the management of downside risks. Deploying IAAIS's comprehensive risk management framework can help uncover weaknesses ranging from controls to ethics, data to inputs/outputs. The system is a two-vector approach that is top-down, starting with an organization's management, integrated and enumerated, combined with a horizontal, system-by-system granularity which should systematically identify and quantify severity and likelihood of the majority of AI/ML and autonomous systems risks. The top-down approach provides organizational accountability and promotes strong governance, and the system-by-system approach examines each component that comprises a holistic organizational technology stack. IAAIS intentionally separates risk impact and assessment work from the design and development team to minimize risk from confirmation bias, sunk cost bias and other cognitive biases. Additionally, it eliminates conflicts of interest in the design and development ensuring specialized training for instance of ethical choice and oversight of checks and balances on the system itself. Independent Audit establishes an infrastructure of trust that can be relied upon to maximize oversight and compliance.

5. What are the advantages or disadvantages of a high-level, top-down strategic goal for developing and implementing AI/ML across a wide spectrum of general applications versus an ad-hoc, case-by-case targeted approach?

IAAIS is agnostic to this question as it manages risk from both vectors and is agile and sufficiently thorough to react specifically to a chosen design.

6. With respect to AI/ML, what phase of technology adoption is the commercial nuclear power industry currently experiencing and why? The current technology adoption model characterizes phases into categories such as: the innovator phase, the early adopter phase, the early majority phase, the late majority phase, and the laggard phase.

ForHumanity recognizes that there is both a life cycle for algorithms and a life cycle from testing to decommissioning at the macro-level for complete systems operation and integration. Our forthcoming paper on *Change Management of IAAIS* is designed to explicitly document the differences and similarities between compliance-by-design of existing systems versus new systems with in-built compliance. Notably a triage of risk assessment that identifies the most risky elements of existing systems in order to manage a process of systemic risk mitigation amongst existing systems. New systems should be required to be compliance by design from the outset.

7. What challenges are involved in balancing the costs associated with the development and application of AI/ML tools, against plant operational and engineering benefits when integratingAI/ML into operational decision-making and workflow management?

Good data governance and compliance-by-design can increase the development costs of software. However, the downside risks associated with defective or weak software and software/hardware products that fail and/or underperform in predictable or foreseeable ways will likely result in harms that far outweigh the upfront costs.

8. What is the general level of AI/ML expertise in the commercial nuclear power industry (e.g. expert, well-versed/ skilled, or beginner)?

ForHumanity creates tools to evaluate overarching compliance-by-design in ethics, bias, privacy, trust and cybersecurity is at best a beginner. Governance, accountability and oversight remain nascent and present catastrophic risk in the nuclear industry. Any regulatory guidance that does not marry AI/ML and autonomous system promotion and adoption with a third-party, independent system of checks and balances and embedded governance with regular re-examinations will allow for inherent weakness and likely failure. The consequences may be minimal, but the critical and existential nature of failure in the nuclear industry demands the highest oversight.

9. How will AI/ML effect the commercial nuclear power industry in terms of efficiency, costs, and competitive positioning in comparison to other power generation sources?

The potential harm to the nuclear industry's top-of-stack position in power generation may be at greater risk from AI/ML/Autonomous system adoption than the other generators in the power stack. With a substantial natural generational efficiency advantage, adding AI/ML/Autonomous systems will not meaningfully increase that advantage. Instead, as has always been the case, risk in the nuclear industry is the one variable that can hold back or even curtail the industry's role in the stack.

AI/ML/autonomous systems here may be a double-edged sword. With systems being designed to increase security, improve safety and enhance autonomy, it is natural to think that these tools will secure the nuclear industry's role at the top. However, the added complexity, and increased attack vector risk from Data Entry Point Attacks, unmitigated bias, genuine control and safety issues embedded in operationalization and uncertainty associated with instances of uncovered ethical choice related to AI/ML/autonomous systems may introduce sufficient concern to merit excess scrutiny and result in increased offline operations. The wrong incident (e.g. a Data Poisoning attack

or Deepfake) could easily introduce sufficient fear for a comprehensive deep dive across the entire industry as to severely disrupt production and/or instill doubt in the system.

Therefore, ForHumanity strongly recommends oversight, accountability and governance by design for all implementations of AI/ML/Autonomous systems so that the transparency, disclosure, documentation are constantly reviewed and current. This would represent a robust risk management and mitigation process.

10. Does AI/ML have the potential to improve the efficiency and/or effectiveness of nuclear regulatory oversight or otherwise affect regulatory costs associated with safety oversight? If so, in what ways?

At the outset, regulatory governance and oversight, if accomplished by Independent Audit of AI Systems would result in decreased efficiency and increased cost, but with tremendous gains in oversight, governance, accountability, transparency and trust. This is likely a temporary state. New developments and systems should and will be created compliance-by-design which, over time, result in synergies and cost savings. Once established, maybe over a 3-5 year period, then through systemic compliance, and transparency, documentation and disclosure requirements — compliance costs will stabilize, if not decline.

The value for the NRC is anticipated to be enormous, as the AI/ML/Autonomous system regulatory framework could be levelled and normalized. Compliance-in-a-box solutions could create a systemic funnel of normalized and automated compliance resulting in tremendous leverage for the NRC.

11. AI/ML typically necessitates the creation, transfer and evaluation of very large amounts of data. What concerns, if any, exist regarding data security in relation to proprietary nuclear plant operating experience and design information that may be stored in remote, offsite networks?

The size and turnover of data is a new security vector for the operators and the NRC to consider and secure. Data labelling attacks, model inversion, membership inference and other Data Entry Point attacks can render models useless or in a worst-case scenario adversarial to the safe function of a nuclear facility. Large sums of data or source code present tremendous cover for malicious entry, such as the SolarWinds hack, which the entire Federal Government was largely susceptible towards. This highlights a protocol concern about segmentation and separation of AI/ML/Autonomous systems. National Institute of Standards and Technology and NRC frameworks crafted into auditable rules examined by third-party independent auditors will represent the highest possible level of governance, accountability and oversight.

Conclusion

Because of the elevated level of risk associated with nuclear power, the checks, safeguards, and consideration must match the equivalent level of risk. Among the various methods and strategies for identifying and analyzing downside risk, Independent Audit of Autonomous Systems presents the most thorough and holistic approach to risk identification and mitigation throughout product development, deployment and use. Any AI system remains vulnerable in areas of ethics, bias, privacy, trust, and cybersecurity, but proper management of the downside risks associated with each area can help to mitigate risk and will result in better, safer, and more sustainable AI and autonomous systems implementation under the scrutiny of the NRC. These criteria, submitted by ForHumanity's crowd of experts, are reviewable by the NRC. While ForHumanity will continue to encourage and promote these principles in industry, we recognize that controlling authorities remain in the hands of legislators and administrative agencies. Legislative and regulatory bodies have the authority and position to drive industry standards and practices, and ForHumanity remains eager to support and serve those charged with making AI and autonomous systems safe for humanity.