PUBLIC SUBMISSION

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

ADD: Matthew Dennis, Tray Hathaway, Mary Neely
Comment (3)
Publication Date: 7/5/2022
Citation: 87 FR 39874

As of: 8/18/22, 7:47 AM
Received: August 17, 2022
Status: Pending_Post
Tracking No. l6x-zo9u-86m3
Comments Due: August 19, 2022
Submission Type: Web

Docket: NRC-2022-0095
NRC's Fiscal Years 2023-2027 Artificial Intelligence Strategic Plan

Comment On: NRC-2022-0095-0001
NRC's Fiscal Years 2023–2027 Artificial Intelligence Strategic Plan

Document: NRC-2022-0095-DRAFT-0004
Comment on FR Doc # 2022-14239

Submitter Information

Name: Tyler Cody
Address: United States,
Email: tcody@vt.edu

General Comment

Members of the NRC,

In the submitted document I respond to the posed questions for comment in order. I hope the comments are useful. I am available for follow-up discussions.

Very respectfully,
Tyler Cody, Ph.D.
Research Assistant Professor
Intelligent Systems Division
Virginia Tech National Security Institute
tcody@vt.edu

Attachments

NRC Comments

NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

August 17, 2022

Comments Regarding NRC's Fiscal Years 2023–2027 Artificial Intelligence Strategic Plan

Members of the NRC,

In the following I respond to the posed questions for comment in order.

1. Are there any specific recommendations or improvements to consider in the development of the AI Strategic Plan?

On line 34 of page 4-2 in section 4.1, "Strategic Goal 1":

The document states:

"The NRC will undertake research to develop an AI framework to determine the approach to assess areas such as, but not limited to, explainability, trustworthiness, bias, robustness, ethics, security, risks, and technical readiness of AI."

While these topics are certainly top-of-mind for the computer science community, from a systems engineering perspective, they are akin to "properties" one would like AI solutions to have. There are other concerns which are fundamental to engineering generally but take on different forms for AI, and need to be addressed in whatever AI framework is developed. I will mention two interrelated basic issues. The first is test and evaluation (T&E), which is needed to see if these properties are actually possessed by the AI. The second, related issue is life cycle.

As stated in the abstract, "AI are machine-based systems that go beyond defined results and scenarios". Accordingly, there are two issues with T&E: the first is the prevailing focus on "held-out-but-identically-distributed" test sets of data. Identically distributed testing checks that the algorithm is working–not that the function approximations it produces will work as intended in the variety of scenarios it will face during operation. The second is that AI solutions are often an input-output component, but AI solutions are influenced by the systems they operate within, they create systems-level effects and and systems-level outcomes. Component-level testing lacks the scope to properly identify the operating envelopes of AI solutions in terms of the context of their system and environment. Therefore, T&E for AI has dual challenges of (1) insufficient component-level tests and (2) component-level tests only being a piece of the picture.

AI solutions have life cycles, but they are understudied. Whatever framework that is proposed should involve plans for T&E over the life cycle, to monitor the "health" of AI solutions, as well as address system maintenance (retraining, recalibration, etc. of AI solutions) and retirement. In dynamic settings, which I am not sure exist widely in nuclear applications, continuous T&E and continuous re-engineering are important concepts. So, in short, the specific recommendation is to consider an emphasis on T&E and on life cycle management in addition to those "properties" that are currently listed on line 34 of page 4-2. While not all applications of AI may merit all "properties" currently listed, all applications will merit a T&E and life cycle management plan.

2. What goals, objectives, or strategies within the NRC's current strategic plan should be added, enhanced, or modified in the AI Strategic Plan?

The response to Question 1 applies to Question 2. Additionally, I think there should be a strategy which raises assurance in AI by focusing on engineering *processes* for training, deploying, maintaining, and retiring models. The emphasis on processes is meant to distinguish assurance approaches which focus on individual models or solution methods. There is a tendency in AI to focus on validating and accrediting trained models, as opposed to the processes that train, deploy, maintain, and retire them.

3. What are potential near-term, or far-term, AI activities that the NRC should be aware of when finalizing and prioritizing the AI Strategic Plan, or associated supporting research?

The systems engineering research community has a growing, active interest in systems engineering methods and best practices for AI and AI-enabled systems. These lines of research are directly concerned with the engineering process, from need analysis, through decomposition into components, recomposition into a system, deployment, maintenance, and retirement.

In the near term, this work addresses a gap in the literature and research community. There is heavy research into AI algorithms, and into secondary concerns like those properties listed under Strategic Goal 1, but there is not as much research into engineering processes, which, if followed, will produce AI solutions that meet stakeholder needs and requirements.

In the far-term, the use of digital models as part of digital engineering and model-based systems engineering activities will expand to AI-enabled systems. This means digital processes for verification, validation, and accreditation of AI solutions using digital twins and model-based systems engineering to evaluate test performance and to test against requirements in (virtually) varied conditions. The Department of Defense has a growing portfolio of this research, and so, especially for Strategic Goals 2 and 3, it is suggested that the NRC connect with the systems engineering research community.

4. What are potential challenges the NRC should be aware of when preparing to review potential use of AI in nuclear applications?

Regarding operations, the no free lunch theorems of statistical learning theory suggest that no single model can be optimized for all conditions at once. And so, the concept of T&E as a pre-deployment activity is in conflict with the first principles behind ML solutions. That is, as conditions change, e.g., between operations, or as platforms degrade, or with changes in use, if there is a material difference in the data that flows through the model, then the performance of the model is expected to change. This suggests that domain adaptation is the rule, not the exception. Conversely, so-called universal models, e.g., general purpose vision models, are the exception, not the rule. Thus, operators looking for accreditation ought to have in place procedures for mitigating risks due to changes in the distribution of data. Important to such procedures is an acknowledgement of catalysts; what aspects of the system wherein the AI model

will operate can change and can those changes create a material difference in the data that flows through the AI model? Maintenance of physical parts and machinery, changes in use, and variation in environmental factors are examples of catalysts that exist *outside the scope* of the AI model itself.

I hope the above comments are useful. I am available for follow-up discussions.

Very respectfully,

Tyler Cody, Ph.D.

Research Assistant Professor
Intelligent Systems Division
Virginia Tech National Security Institute
tcody@vt.edu