INL/CON-24-79831-Revision-0

Human-centered and explainable artificial intelligence in nuclear operations

August 2024

Changing the World's Energy Future

Anna Hall, Patrick Murray, Ronald Laurids Boring PhD, Vivek Agarwal

Idaho National Laboratory

INL/CON-24-79831-Revision-0

Human-centered and explainable artificial intelligence in nuclear operations

Anna Hall, Patrick Murray, Ronald Laurids Boring PhD, Vivek Agarwal

August 2024

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Human-centered and explainable artificial intelligence in nuclear operations

Anna Hall[1], Patrick Murray[2], Ronald Boring[1], & Vivek Agarwal[1]

[1]Idaho National Laboratory

[2]EQRPI

The operating fleet of light water nuclear reactors in the United States is undergoing modernization from analog to partially digitalized environments, and the advent of new plant designs will feature fully digital control rooms. Retrofitting legacy instrumentation and controls (I&C) and new digital concepts of operations require significant increases in the level of automation, incorporating artificial intelligence (AI) and machine learning (ML). Human-centered AI (HCAI) is a combination of human-centered design (human factors, human-in-the-loop, etc.) combined with AI/ML to design an efficient, reliable system with full consideration for human engagement. The transition from digital to AI-automation does not have an industry-ready template, and significant research is needed to bridge the gaps and maximize the application of AI. This paper provides a comprehensive and novel discussion of HCAI considerations in nuclear power, highlighting unique applications for the existing fleet and new advanced reactor designs. We end with a real-life use case of AI applications to work management in nuclear power plants (NPPs) and lessons learned.

INTRODUCTION

The operating fleet of light water nuclear reactors in the United States relies heavily on analog I&C systems. A recent wave of modernization has brought upgrades in the form of partially digitized control rooms, while new plant designs feature fully digital control rooms. With the advent of smaller, advanced reactors like microreactors and small modular reactors, there is a need to reduce human staffing levels to ensure operating efficiencies and cost competitiveness. Such concepts of operations will require significant increases in the level of automation deployed at future NPPs, which would involve incorporating AI and ML. The transition from digital to automation does not have an industry-ready template, and significant research is needed to bridge the gaps and maximize the application of AI/ML, especially in control systems.

Human-centered AI (HCAI) is a combination of human-centered design (human factors, human-in-the-loop, etc.) combined with AI to design an efficient, reliable system with full consideration for human engagement (Shneiderman, 2020). This approach supports AI development, evaluation and use with humanistic design and control, ensuring human-in-the-loop with respect to sustained interaction and ongoing collaboration between humans and the technology throughout its lifecycle. This includes controlling how to arrive at an endpoint and moves away from historical viewpoints of AI as pure science and mathematical advancements, towards one that is interactive with humans and used to empower human capabilities.

HCAI is an evolution of human-centric automation which has been around for several decades. First principles have been developed in the aviation space positing that automation technologies be designed as tools with *human* use as the primary *focus* of attention (Billings, 1996). This is because as with nuclear power operators, aircraft operators bear responsibility for safety, and so must possess ultimate authority. Human-centered automation requires that operators be actively involved and appropriately informed, be able to understand and predict the automation, and benefit from automation that offers checks and balances to human actions when necessary.

While the application of human factors engineering in nuclear power operations is reviewed and regulated by the U.S. Nuclear Regulatory Commission (NRC), the human factors implications for AI in nuclear are in development. An HCAI-nuclear framework provides for function analysis that supports humans in their new roles at the plants alongside the AI technology, but to do so automation is needed that teams well with the operator, including interventions for safety reasons.

We firstly introduce AI/ML applications in nuclear that are both currently underway and in the works. These span concepts that are both retrofitted to the large traditional reactors that comprise the existing fleet, as well a new advanced reactors such as small modular and microreactors. We then describe the human factors and human-centered design considerations that can help support efficient, reliable systems with full consideration for human engagement. Although AI/ML technologies face different challenges across existing and new reactors, HCAI considerations are similar. We end with examples of real-world AI deployment within the industry that has produced mixed outcomes and that can inform an HCAI framework moving forward.

AI/ML APPLICATIONS IN NUCLEAR POWER

A small number of AI applications are already in use, albeit in a limited capacity. These mainly involve automating routine processes including work management, document retrieval, and using AI to help with regulatory affairs (Nuclear Energy institute, 2024). Notwithstanding, there is great interest to expand AI use and in coming years the deployment of AI technologies within the energies industries is anticipated to be prolific. To this end, the International Electrotechnical

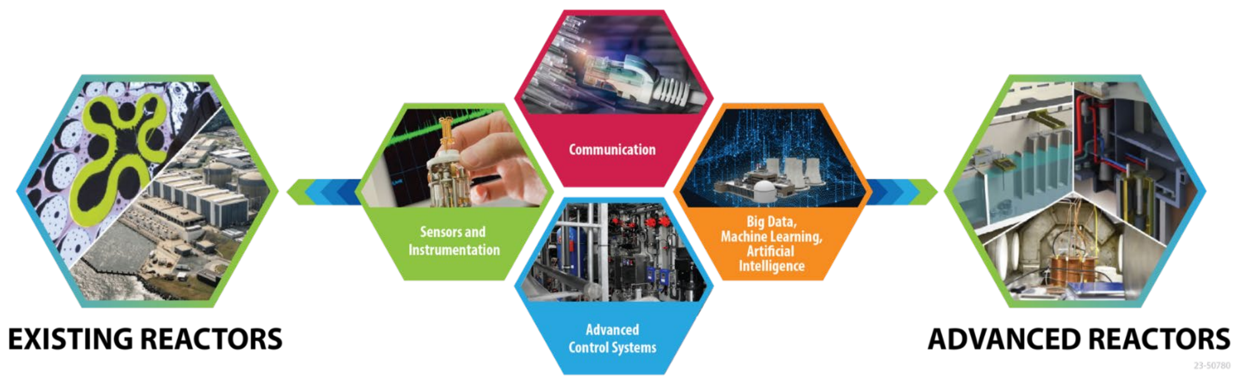

Figure 1. Applications of AI/ML in nuclear power

Commission (IEC) have begun issuing standards for AI development and process (IEC, 2024), data life cycle models (IEC, 2023) and assessment of ML classification performance (IEC, 2022). In the advanced reactor space, AI applications are poised to span the lifecycle including reactor design, operations and maintenance (O&M), materials, grid, storage, and through decommissioning (Vilim and Ibarra, 2022).

Figure 1 depicts four key research areas in which AI/ML will be used to convert data into information generating new actionable insights. Each of these new technical areas can lower O&M costs, increase reliability and with new advanced designs, AI may be used to support remote, and near-autonomous operations (Agarwal, 2024).

Sensors and Instrumentation

A significant cost to existing NPPs is preventative maintenance programs in which plant equipment, components and systems undergo cyclic manual inspections to ensure functionality and avoid failure. These labor-intensive rounds are carried out routinely whether the parts and systems appear degraded or not.

One of the modernization strategies in recent years has been to install data capture technologies such as sensor hardware with corresponding bandwidth and computing capabilities (Al Rashdan and St Germain, 2018). Digitizing infrastructure in this way has allowed for vast quantities of component data that can feed ML algorithms to shift maintenance activities from preventive to predictive measures, resulting in significant cost-savings.

A recent application of this nature has been to identify the likelihood of waterbox fouling (Agarwal et al., 2021, Agarwal et al., 2022; Walker et al., 2023), a common maintenance activity for pressurized water reactors. This water pump blockage occurs when debris such as grass and leaves from the water source used to cool the reactor (e.g., river) builds up over time. Thus, instead of periodic, paper-based and manual inspections, ML algorithms can make use of plant indicator data already available such as motor vibration, current, temperature, system condenser, and turbine and vacuum values to form condition-based monitoring that can generate waterbox fault predictions.

Communications

Advances in wireless communications are being applied to industrial automation (Manjunatha & Agarwal 2022). In NPPs, communications advances that provide continuous access to plant data are necessary to help reduce industrial risk and enhance operational efficiency and effectiveness. Given the increased cybersecurity risk that NPPs face, the nuclear industry has traditionally been cautious with regard wireless communications, largely opting to keep data private and guarded, in closed loop systems within the plant (Hall & Agarwal, 2024).

The full business potential afforded by a wireless infrastructure stems from plant data being monitored and analyzed with subsequent diagnosis and prognosis. To realize this benefit, data quality and integrity must be intact (i.e., garbage-in-garbage-out (Awati, 2023)). Research efforts are underway to apply AI to predict data anomalies at the sensor level (Agarwal, 2024). This smart sensor application filters sensor data through an AI model validation process that can detect malfunction, operating limits and baselines values before ongoing data transmission.

Big data, ML, AI

Most U.S. reactors were commissioned in the 1970s and 1980s and are among the oldest NPPs worldwide with a mean age M=41.6 years (U.S. Energy Information Administration, 2023). Initially licensed for 40 years, many plants have either been approved, or are in the approval process for 20 or 40 years licensure extensions. To maintain safety, structural degradation due to age and wear and tear is carefully monitored. However, this process currently lacks for advanced technology solutions that can make use of big data and AI/ML.

The structural integrity of the plants can be compromised by Alkali-Silica Reaction (ASR), an intrinsic chemical reaction that forms a gel in concrete pores, that expands causing stress and cracking. In addition, ASR can be correlated with corrosion of steel embedded in the concrete. AI can be applied to detect, localize, and estimate ASR damage using a physics-informed ML approach. This technique creates diagnostic and prognostic models that offer data-based decision-making on structural health monitoring to replace the current offline and manual approach.

Advanced Control Systems

Although not yet commercially operational in the U.S., intense research and development is underway for advanced reactors. These new concepts of operations (conops) promise improvements such as passive safety features, better fuel and material performance, reduced waste, and many O&M benefits. Unlike large, traditional NPPs that generate 700+ MWe per unit, and that can power small cities, most advanced reactors are smaller in size such as small modular reactors (~50-300MWe) and microreactors (<50MWe). These transportable conops can be broadly distributed and are envisioned for industrial applications and to power remote and isolated areas such as mining sites (International Atomic Energy Agency, 2023). Other advanced conops include fusion or radioisotope technologies.

Research efforts are underway that strive to realize AI-enabled remote and autonomous operations. One such project is an anticipatory control strategy for microreactors, that uses neural-network-based models to self-regulate under varying operational conditions (Lin et al., 2024). These conditions include steady state and transient operations, load following, and failure or degraded operations. In comparison to physics-based models, data-driven anticipatory control approaches are stated to produce computationally efficient, accurate, and adaptive semi-autonomous operations (Agarwal, 2024).

HCAI CONSIDERATIONS FOR EXISTING AND ADVANCED REACTORS

Figure 2 describes a human-centered approach to AI deployment in nuclear such that humans be involved in all phases of the technology's lifecycle from development, through evaluation and use. For many of the advanced AI applications described, from predictive maintenance to structural health monitoring, their success depends on data quality and integrity. While AI tools may help detect sensor data anomalies, human verification will be necessary to ensure accurate, reliable and contextualized data is feeding the algorithms in the first place. This may require new personnel roles be created such as Data Scientist or Analyst. In terms of AI evaluation, there are training implications because the ability to interpret and verify ML recommendations generated from multiple data sources is a new analytical skill and new mode of O&M not currently conducted at plants.

In addition, to satisfy defense in depth safety requirements, the technology must be designed to not only support employees with deep understanding of the automated systems, but also each intelligent component must possess an understanding of the function and intent of all other intelligent automated systems (Billings, 1996).

For inspection tasks that are replaced by sensors, situation awareness may decrease as elimination of manual inspection also eliminates operators noticing other fault scenarios on their

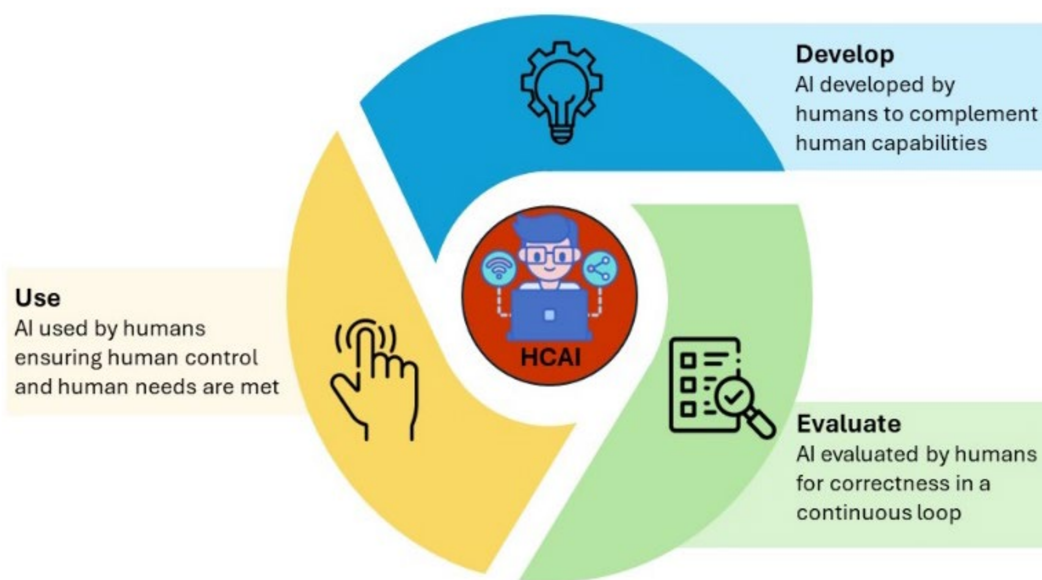

Figure 2. Human-centered AI in the technology's lifecycle

rounds (e.g., leaking pipes unrelated to the system the AI is monitoring). Further, design of any ML interface should comport with HCAI principles such that the information is presented in a digestible and explainable manner that the humans can understand.

For control tasking to meet autonomous and near-autonomous operations, all existing conops invariably require humans to monitor the automation, and certainly act as failsafes or "backups" when the automation fails (Hall et al., 2024). Several decades of human factors research highlight automation 'trade-offs', including turning once-experts into novices when put in situations where passive monitoring is required and there are no longer any active tasks to perform (i.e., skills degradation; McLeod et al., 2022). Further, vigilance decrement occurs in humans after about 10 minutes of monitoring, which results in reduced situation awareness, pointing to humans being ill-suited for automation monitoring in the human-AI teaming relationship (Sheridan, 2002). The precise role of humans in AI deployment for control functions will have to be carefully articulated, tested and verified to ensure safe and efficient operations.

Trust and explainability are central components to HCAI. Trust has been shown to be one of the leading factors in whether users will rely on automated processes. Trust in automation has a rich literature in human factors research and can be considered either as an information-based set of beliefs about the automation (cognitive), or the corresponding reliance and use of the automation (behavioral). EXplainable AI (XAI) is synonymous with transparency in automation and has grown in importance as increasingly larger numbers of people outside computing science or intelligent systems interact with AI. XAI serves to ensure human-in-the-loop with respect to transparency in reaching an endpoint and increases user trust (Ali et al., 2023).

Advanced reactors can benefit from HCAI being built into the design, while legacy plants must contend with AI retrofitting, and the corresponding changes to staff roles and responsibilities. However, without advanced reactors possessing established operational safety and reliability, adding layers of functional complexity through AI may prove challenging for the regulator. Further, and the lack of historical operational experience data needed to support AI applications will further exacerbate these concerns.

Advanced reactor designs with in-built AI applications will be seeking regulatory assurances and may be viewed in light of demonstrated success of AI-applications to the existing fleet. New AI-assisted functions will realize new economic and efficiency benefits, but they will also bring about new costs and challenges. Greater departures from current O&M activities will attract greater scrutiny, especially for AI technologies involved in control and safety functions. The role and authority of the human operators will be of chief concern (Billings, 1996).

AI/ML USE CASES IN NUCLEAR POWER

The application of AI in nuclear plants is already underway. Currently, it is largely confined to automating routine processes that require a great deal of paperwork and with checks and balances across different departments. Here we evaluate real world instances in which AI has been applied, and discuss lessons learned for HCAI moving forward.

NPP performance improvement processes have evolved over time adding more rigor and capability, but also adding more administration to manage the improvements. One of the most important performance improvement processes is the Corrective Action Program (CAP) which consists of a series of actions from reporting through screening, evaluating, responding to, and trending of issues that arise at the plant. The administration of all these functions is labor intensive and time consuming.

In 2016, the Nuclear Energy Institute (NEI), the policy organization for the nuclear utilities, in collaboration with the NRC issued a series of documents under the 'Delivering the Nuclear Promise' (DNP) initiative (NEI, 2016, 2017), which was a multi-year strategy aimed at transforming the industry and finding significant performance and economic improvements while maintaining safety. Over the course of the next several years, the DNP initiative rolled out a series of guidance documents on how to achieve this vision, including two concerning greater efficiency in CAP.

As a result, many utilities made changes to their CAP by reducing perceived low value processes within the program that added significant "administrative burden". For some, this meant implementing ML-driven AI programs that performed screening functions of the condition reports, the first step in the CAP process, which was previously performed by a collegial group of individuals from various departments. These changes resulted in immediate O&M cost savings for the utility, and Top Industry Practice awards were issued to some utilities that implemented AI applications in the CAP process. This acknowledgement drove other U.S. utilities to strive for the same type of AI programs to help with their CAP administration.

However, although implementation of the AI reduced the administrative burden of managing CAP, one of the consequences of the automation was a reduction in cognitive information processing and trending of condition reports, that were previously performed by humans. Together these provided a defense to more significant events, or failures. Consequently, some of the utilities that have implemented AI enhancements to CAP have not only realized more failures, but also without the ability to fully understand the underlying causes due to inadequate human engagement in processing at the lower-level condition reports.

In some cases, the utility saw an overall decline in plant performance. Although it is difficult to attribute this performance decline directly to the implementation of the AI, a lack of information to feed human decision-making could be a contributor. The old CAP process historically offered an effective real-time communication pipeline between departments and levels of the organization to raise issues unfiltered. Disruption of this process via AI-tools may have led to lowered collective situation / system awareness of the issues within the plant, and thus lowered the possibility to identify and troubleshoot problematic trends over time that highlighted underlying organizational and / or programmatic causes. In at least one plant, there was an instance of a downstream reduction in the plant's senior leadership team to be able to make informed decisions concerning the condition of the plant's equipment, and the organization as a whole.

These examples underscore the importance of human engagement in the development, evaluation and use of AI technologies in nuclear (Figure 2) for work management processes, as well as the unforeseen consequences that can be brought about absent human factors considerations. Beyond failures of the automation in its primary screening function for CAP, an HCAI framework can help mitigate problems downstream by ensuring the correct level of human engagement in the ongoing evaluation of issues raised at the plant.

CONCLUSIONS

NPPs contain complex sociotechnical integration systems, and while automation and AI technologies can and do improve system performance, full consideration of HCAI is yet to be understood. Human factors research is necessary to examine human-system performance indicators with AI-enhanced tools, especially with advanced controls. The human-intelligent agent teaming relationship and training requirements for AI-augmented job functions will also have to be explicated.

REFERENCES

Agarwal, V. (February, 2024). Application of artificial intelligence and machine learning in nuclear: Benefits, challenges, and opportunities. 5th International Conference on Reliability, Safety, and Hazard. DAE Convention Center, Anushaktinagar, Mumbai, India.

Agarwal, V., Manjunatha, K., Smith, J.A., Gribok, A.V., Yadav, V., Palas, H., Yarlett, M., Goss, N., Yurkovich, S., & Diggans, B. (2021). Machine learning and economic models to enable risk-informed condition-based maintenance of a nuclear plant asset.

Agarwal, V., Walker, C.M., Manjunatha, K., Mortenson, T.J., Lybeck, N.J., Hall, A., Hill, R., & Gribok, A.V. (2022). Technical basis for advanced artificial intelligence and machine learning adoption in nuclear power plants. Idaho National Laboratory. INL/RPT-22-68942.

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M.,... & Herrera, F. (2023). Explainable artificial intelligence (XAI): What we know and what is left to attain trustworthy artificial intelligence. *Information Fusion, 99,* 101805.

Al Rashdan, A.Y., St Germain, S.W., 2018. Automation of data collection methods for online monitoring of nuclear power plants. Idaho National Laboratory. INL-EXT-18-51456.

Awati, R., 2023. Garbage in, garbage out. TechTarget. Retrieved 1 JUNE from: https://www.techtarget.com/searchsoftwarequality/definition/garbage-in-garbage-out.

Billings, C. E. (1996). *Human-centered aviation automation: Principles and guidelines* (No. NASA-TM-110381).

Hall, A., & Agarwal, V. (2024). Barriers to adopting artificial intelligence and machine learning technologies in nuclear power. *Progress in Nuclear Energy, 175,* 105295.

Hall, A., Le Blanc, K., Miyake, T., Boring, R., & Kovesdi, C. (2024). Draft landscape analysis of anticipated use of autonomous operations. Idaho National Laboratory. Research Information Letter. Office of Nuclear Regulatory Research.

International Atomic Energy Agency. (2023). What are Small Modular Reactors (SMRs)? Retrieved from https://www.iaea.org/newscenter/news/what-are-small-modular-reactors-smrs

International Electrotechnical Commission, 2022. Information Technology — Artificial Intelligence — Assessment of Machine Learning Classification Performance (ISO/IEC TS 4213:2022, Issue.

International Electrotechnical Commission, 2023. Information Technology — Artificial Intelligence — Data Life Cycle Framework. ISO/IEC 8183:2023, Issue.

International Electrotechnical Commission, 2024. Information Technology — Artificial Intelligence — Guidance for AI Applications. ISO/IEC 5339:2024, Issue.

Lin, L., Oncken, J., & Agarwal, V. (2024). Autonomous control for Heat-Pipe microreactor using Data-Driven model predictive control. *Annals of Nuclear Energy*, *200*, 110399.

Manjunatha, K. A., & Agarwal, V. (2022). Multi-band heterogeneous wireless network architecture for industrial automation: A techno-economic analysis. *Wireless Personal Communications*, *123*(4), 3555-3573.

McLeod, R. (2022). Human factors in highly automated systems [White Paper]. Chartered Institute of Ergonomics and Human Factors.

Nuclear Energy Institute (2016). Reduce cumulative impact from the corrective action program. Efficiency Bulletin: 16-10.

Nuclear Energy Institute (2017). Improving the effectiveness of issue resolution to enhance safety and efficiency. NEI 16-07 (Revision A).

Nuclear Energy Institute (2024). Survey on Uses of Artificial Intelligence. [Draft]. Personal communication.

Sheridan, T. (2002). Humans and Automation: System Design and Research Issues. New York: Wiley & Sons.

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495-504.

US Energy Information Administration, 2023. Nuclear explained: U.S. nuclear industry. Retrieved 10 DECEMBER from. https://www.eia.gov/energyexplained/nuclear/us-nuclear-industry.php.

Vilim, R., & Ibarra, L. (2022). Artificial Intelligence/machine Learning Technologies for Advanced Reactors: Workshop Summary Report.

Walker, C.M., Agarwal, V., Lin, L., Hall, A.C., Hill, R.A., Boring PhD, R.L., Mortenson, T. J., Lybeck, N.J., 2023. Explainable artificial intelligence technology for predictive maintenance. INL/RPT-23-74159. https://doi.org/10.2172/1998555.