










Utilizing a Virtual Sodium-Cooled Fast Reactor Digital Twin to Aid in Diversion Pathway Analysis for International Safeguards Applications

Ryan Stewart , Ashley Shields , Shaw Wen, Frederick Gleicher , Samuel Bays , Mark Schanfein, Jeren Browning , Katherine Jesse  and Christopher Ritter 

Idaho National Laboratory, Idaho Falls, ID, USA

ABSTRACT

Digital twin technology can improve the effectiveness of international safeguards inspectors by providing a tool that can perform an accurate acquisition pathway analysis, identify pathway indicators, develop required sensors to detect indicators, and monitor facilities in real time using critical data streams that benefit from this safeguards-by-design approach. Safeguards inspectors are required to visit facilities and verify the nuclear material to ensure no diversion has taken place and to detect facility misuse; however, this analysis and verification effort is time consuming, and with limited funding, it is imperative that time spent at a nuclear facility is focused on key areas. We developed a virtual digital twin of two general sodium-cooled fast reactors and explored diversion and misuse scenarios to determine how a digital twin could provide inspectors with an understanding of how proliferation may occur and where the most likely areas for proliferation would be. For each of the three reactors, an optimization algorithm was able to find core designs that would be difficult to detect via sensors alone; however, a machine learning adapter provided by the digital twin was able to show general trends in where proliferation is likely to take place.

ARTICLE HISTORY

Received 17 October
2022
Accepted 6 December
2023

Background

International reactor safeguards

The International Atomic Energy Agency (IAEA) is tasked with promoting the safe, secure, and peaceful use of nuclear technologies. This includes providing verification that nuclear facilities are not misused, and that nuclear material is not diverted to develop nuclear weapons. In this regard, the IAEA provides a source of confidence that states are abiding by international commitments while helping minimize the risk of undetected

proliferation. To do this, the IAEA utilizes a combination of techniques, including nuclear material accountancy, containment, surveillance, and design information verification. These techniques provide data to the IAEA that can be used to discern if facilities are being utilized as declared. These techniques use both IAEA personnel visits to facilities to ensure compliance and data acquisition from various sensors, which are transferred remotely and analyzed.

The IAEA is currently experiencing a zero- or minimal-growth budget, which is straining available resources.¹ Given the minimal budget growth and projection of nuclear countries to expand through 2035, additional technologies could allow for a more efficient use of current resources without reducing the quality of work.² To help overcome the minimal-growth budget and projected increase in reactor development, new technologies are currently being examined by various member states. This work examines the use of digital engineering, specifically a virtual digital twin (DT), to aid in diversion pathway analysis (DPA) for a general sodium-cooled fast reactor (SFR).

A virtual DT can be utilized to mirror a physical asset, whereas data obtained from the virtual DT can be used to make predictions about the operations of the physical asset, in this case an SFR. Creating a realistic SFR model requires enough design information to accurately capture the physics occurring, and once the SFR is operational, sensor data will be required to continually update the DT, ensuring an accurate representation of the reactor over time. This work assumes that the IAEA will have access to SFR design information verification documents and will be able to obtain some subset of sensor data at regular intervals to monitor core operations. The IAEA currently has agreements in place to stream data from a facility to its headquarters in Vienna, Austria; the use of DTs seeks to leverage and expand on these types of agreements.³

DTs require significant amounts of data to provide an accurate representation of a physical asset; however, they can be used to perform a preliminary DPA analysis (enhancing safeguards-by-design features) before the reactor is built and can be used to remotely monitor a facility to ensure that it is operating under some umbrella of normal operating conditions. During the design process, the IAEA could utilize a virtual DT to explore DPA and ensure proper instrumentation was deployed. Identifying additional instrumentation requirements early in the design process before a reactor has been fully built could have significant cost implications.⁴ During operations, and depending on the implementation and agreement, a DT could reduce in-person visits. This could allow an operating reactor to run for longer periods of time for power generation or research, thus reducing the burden of safeguards on a state. Along with this, the state would likely have access to the same instruments (especially

if the instruments were dual use) and would provide a plethora of data to ensure the continual operation of the asset. With these boons for a state defined, we explored the use of a virtual DT to perform DPAs to aid in safeguards by design and to provide a framework for the implementation of a DT for an operating reactor once it has been built.⁵

The DT is not meant to replace in-person inspectors but to augment their abilities. In this sense, an inspector can examine the results and conclusions provided by the DT and focus their efforts on specific areas of a reactor that were flagged as off-normal. This could ideally limit the amount of time an inspector was required to be at a facility and provide higher confidence in the inspector's findings.

Digital engineering and digital twin technology

Digital engineering is defined by the U.S. Department of Defense as “an integrated digital approach that uses authoritative sources of system data and models as a continuum across disciplines to support lifecycle activities from concept through disposal.”⁶ Essentially, this approach requires a series of dynamic and integrated models to describe a product, rather than static documents. These models are integrated across various platforms to help produce a design that can be supported across its lifetime.⁷ A DT utilizes digital engineering to comprise a representative model of either a current or future physical product design.⁸ In this sense, we can create a virtual DT to mimic the behavior of a physical product through modeling and simulation.

The existing nuclear fleet has begun to incorporate DTs to improve in areas such as operating, maintenance, and training.⁹ Plant reference simulators act as a limited DT, as they can provide a training tool for operators; however, the reference plant is not actually connected with the power plant. DTs have also begun to be used in condition monitoring, where online performance and condition monitoring can be observed via sensors to provide data that can be utilized by various artificial intelligence and machine learning (ML) algorithms to determine when maintenance should be performed.

Safeguards digital twin

Given the potential benefits of utilizing digital engineering (and DTs specifically), we developed a safeguards digital twin (SG-DT). For this research, the SG-DT represents how a DT can provide the framework for developing a safeguards-by-design approach to perform DPAs. The current work is limited to the “virtual” aspects of a DT, as there is no physical sodium fast reactor to draw data from. Despite this lack of physical data, a wealth of

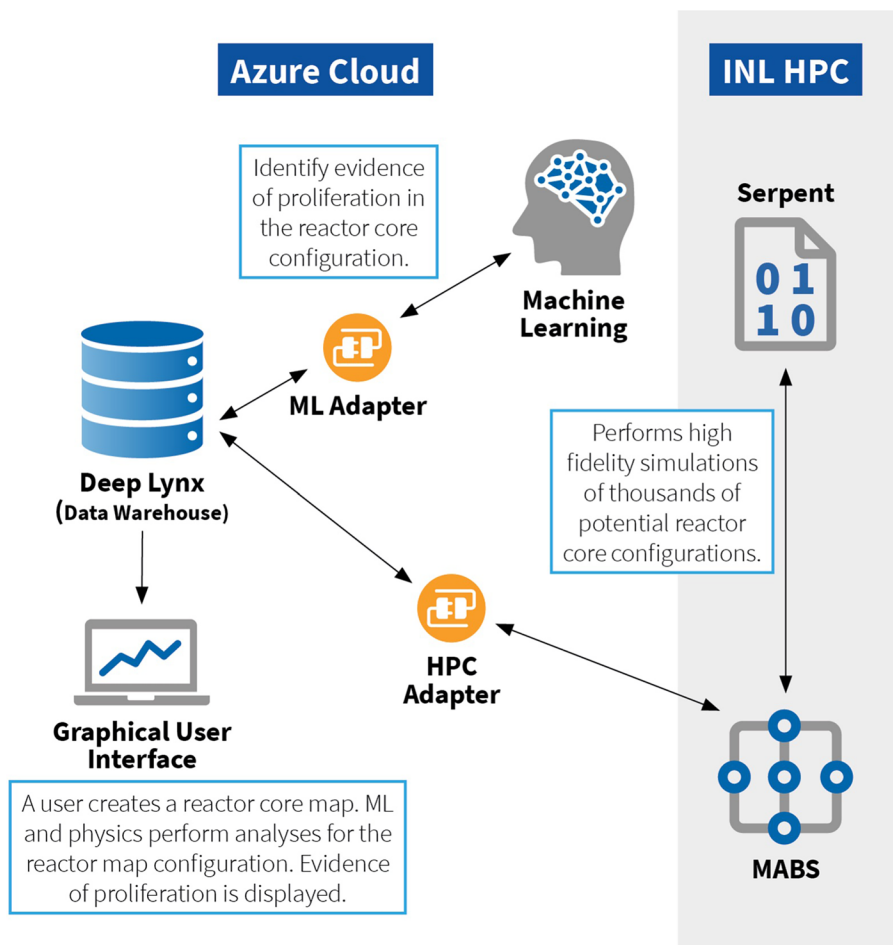


Figure 1. Graphical representation of the SG-DT.

knowledge can be gained from a virtual DT to model, predict, and analyze the behavior of a nuclear reactor to understand safeguards implications.

The SG-DT was developed to eventually be deployed as a full DT; as such, the framework for incorporating a virtual DT with a physical asset was kept in mind. The SG-DT consisted of five individual components (see Figure 1): graphical user interface, physics models, optimization algorithm, data warehouse, and ML adapter.

The graphical user interface provides direct access to the core design, where users (i.e., safeguards inspectors) can down select, for example, the removal of a fuel assembly or fuel pin and where they expect fertile targets to be placed or misuse to occur. The interface can be seen in Figure 2, where the determination of appropriate core designs allows domain experts to directly inject knowledge. The interface allows the user to set up a series of general DPAs, where the optimization algorithm can provide

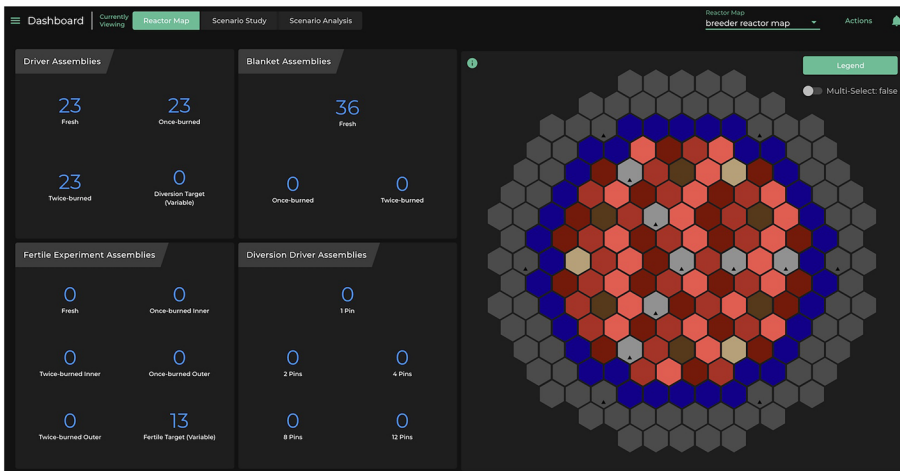


Figure 2. Graphical user interface for SFR core analysis, where users can select fuel bundle positions, likely areas for proliferation, etc.

specific designs that would be the most difficult to detect. The examination of a DPA can also be used from a designer's perspective to help facilitate a safeguard-by-design approach.

The physics models currently consist of high-fidelity Serpent simulations to explicitly model the behavior of each reactor type.¹⁰ While Serpent is the current method, each component within the SG-DT is agnostic to the others, and as such, different physics engines could be replaced with no loss of generality.

A custom optimization algorithm (the multiagent blackboard system [MABS]) for handling large numbers of high-fidelity simulations was developed by the authors to analyze thousands of potential core designs.¹¹ Along with this, the data produced by the optimization algorithm doubles as a training set for the ML adapter and provides a wide sampling of the design space, including an emphasis on optimal designs.

The data warehouse (DeepLynx) is the central hub for storing and transferring data between the reactor physics models and the ML adapter; upon integration with a physical asset (i.e., an operating reactor), DeepLynx can also be able to pull data from any sensors integrated into the system.¹² DeepLynx utilizes a unique ontology for nuclear facilities, which enables a common nomenclature to easily pass information between the physics engine, ML adapter, and future live data streams in a monitoring role.¹³ The data model is encoded using the Data Integration Aggregated Model and Ontology for Nuclear Deployment (DIAMOND), a flexible, open-source model for structuring knowledge of nuclear plant configurations and operations. The DIAMOND ontology is also the basis for automatically parsing and ingesting data into the DeepLynx data warehouse.¹⁴

The ML adapter is trained on data produced by MABS to understand the relationships between the input variables and our quantities of interest. Once trained, the ML adapter examines an individual core to determine if there is a potential for it to be used for diversion or misuse and attempts to narrow in on regions of interest to examine further. The final aspect of the SG-DT is the ability to utilize both cloud computing and high-performance computing (HPC). A majority of the SG-DT is housed on the cloud and communicates with Idaho National Laboratory's (INL's) HPC to perform the optimization algorithm and physics solves. This provides the SG-DT with the computing power associated with HPCs and the versatility and maneuverability of housing the SG-DT in the cloud. As desktop computing power continues to rapidly expand, for local IAEA applications where a data collection and analysis cabinet is installed at the facility, ideally, computations can be done in the cabinet with the analysis remotely transmitted to the IAEA home office.

Implementation and assumptions of the safeguards digital twin

The SG-DT was developed to provide basic safeguards-by-design functionality via DPA. The results generated from the DPA are then fed into an ML algorithm to develop an approach for monitoring a nuclear facility and aiding in the reactor safeguards and nuclear nonproliferation process. Through the SG-DT we will be examining a plausible DPA for a suite of SFR cores. We utilized MABS to determine the most effective core configurations that divert material or misuse the reactor to produce 1 significant quantity (SQ) of special fissionable material (SFM—a material containing plutonium-239 or uranium enriched with uranium-235) with process signals that emulate a nominal operating core.¹⁵ Once a set of reactor cores were identified, we explored how ML can be employed to detect changes in off-normal process signals (i.e., likely indicative of diversion or misuse). Through the SG-DT, our goal is for the ML models to detect an off-normal operating core within a one-year time frame with 95% accuracy.

For this work, we assumed a material balance period (MBP) of one year, providing the time necessary to obtain 1 SQ and remove it from the facility before the next physical inventory is taken. One SQ is the minimum quantity of nuclear material required to develop a weapon; [Table 1](#)

Table 1. SQ values in use by the IAEA¹⁷.

Material	Significant Quantity
Plutonium (<80 wt % plutonium-238)	8 kg
Uranium (<20 wt% uranium-235)	75 kg
Uranium (>20 wt% uranium-235)	25 kg

shows the mass of material required for generating 1 SQ. The detection time is based on the IAEA's goal of timeliness to ensuring no abrupt diversion of 1 SQ has occurred over the MBP.¹⁶

We also assumed that the state is willing to remove or incapacitate specific monitoring and surveillance measures to achieve this goal. This work does not seek to reconcile the monitoring and surveillance measures with the signals obtained from the core. Instead, it seeks to focus purely on the digital signatures obtained from the operating reactor to decide whether the reactor is operating normally. Indeed, future work could leverage the two aspects to further strengthen the safeguards and enhance a security approach.

To evaluate the effectiveness of the SG-DT, we developed a test case by examining an SFR. An SFR was chosen due to its inherent ability to contain large quantities of SFM and generate 1 SQ of plutonium within approximately one year.

Sodium-cooled fast reactor test case

A general sodium-cooled fast reactor (GSFR) was modeled to resemble advanced reactor designs, such as the advanced burner test reactor, advanced burner reactor, and versatile test reactor.¹⁸ While explicit geometries and material compositions are not exact, design preferences draw heavily from these previous designs. The GSFR was designed as a 300 MW plant to be a test and power producing facility. We assumed that its fuel was metallic and comprised of 75 wt% uranium (5 wt% uranium-235/uranium), 15 wt% reactor-grade plutonium (plutonium-239/plutonium ratio of ~ 0.65), and 10 wt% zirconium. The dimensions and general fuel design can be seen in Table 2; more information on the GSFR can be found in Stewart, et al., 2022.¹⁹ The selected fuel matrix resembles current fuel forms for advanced test reactors, such as the versatile test reactor. We assumed that the fuel form will be suitable for generating the high-flux rates necessary for advanced fuels and materials testing in a research reactor without requiring a larger core.²⁰ In addition to this, we selected low-enriched uranium (LEU) fuel to allow the reactor design to be more easily transferred from a weapon to non-weapon state. This also dictated

Table 2. Assembly and fuel design specifications for the GSFR.

Assembly pitch (cm)	14.5
Pins per assembly	217
Pin diameter (cm)	0.808
Clad thickness (cm)	0.0559
Fuel height (cm)	101.6
Fuel diameter (cm)	0.651
Mass per pin (uranium-235 g, plutonium g)	81.1, 324.7
Mass per assembly (uranium-235 kg, plutonium kg)	17.6, 70.5

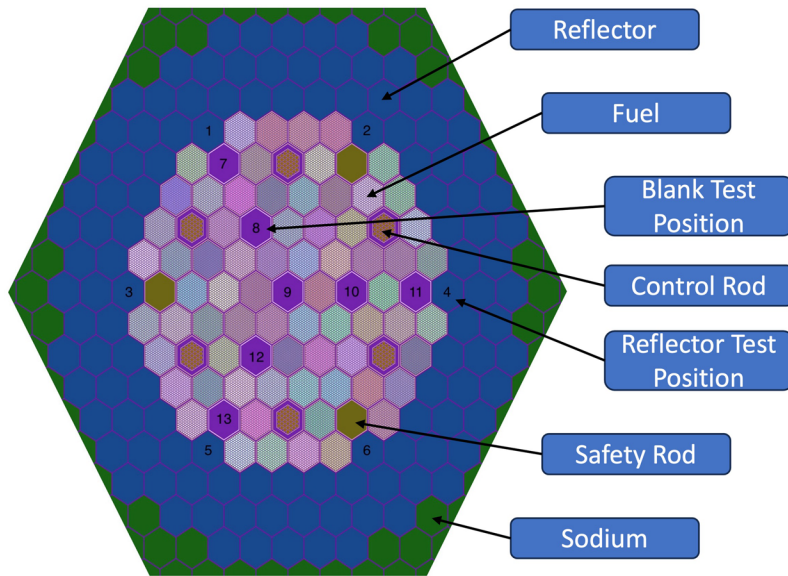


Figure 3. Reference 300MW GSFR core layout.

the use of reactor-grade plutonium. While this work focuses on LEU and reactor-grade plutonium, this framework could easily be extended to other designs.

To capture the fuel cycle of an SFR, approximately one-third of the core was replaced at the end of each cycle, so that one-third of the core was fresh fuel, one-third was once-burned fuel, and one-third was twice-burned fuel. A full description of the core design can be found Stewart, et al., 2022.²¹ The reference GSFR is shown in Figure 3; the numbering in Figures 3–4 refers to the misuse scenario and is described in further detail in the subsequent paragraphs. For the reference GSFR, the central core region consisted of fuel assemblies, safety control rods, control rods, and blank test positions. Surrounding the core region were three rows of reflector assemblies and sodium. The reference core design was created to simulate an SFR test facility that did not intend to breed declared plutonium. The reference GSFR was then augmented create a 300MW test SFR with breeding capabilities (Figure 4).

The 300MW breeder design maintained the same blank test positions and control rod positions; however, a row of stainless-steel reflectors was replaced with blanket assemblies that would be used to breed plutonium to utilize a closed fuel cycle. These blanket assemblies are assumed to be “declared” and subject to the same safeguards inspections as the fuel assemblies. The two SFR configurations provide a realistic test case for the DT and a virtual reactor to test the methodology for determining material diversion and reactor misuse.

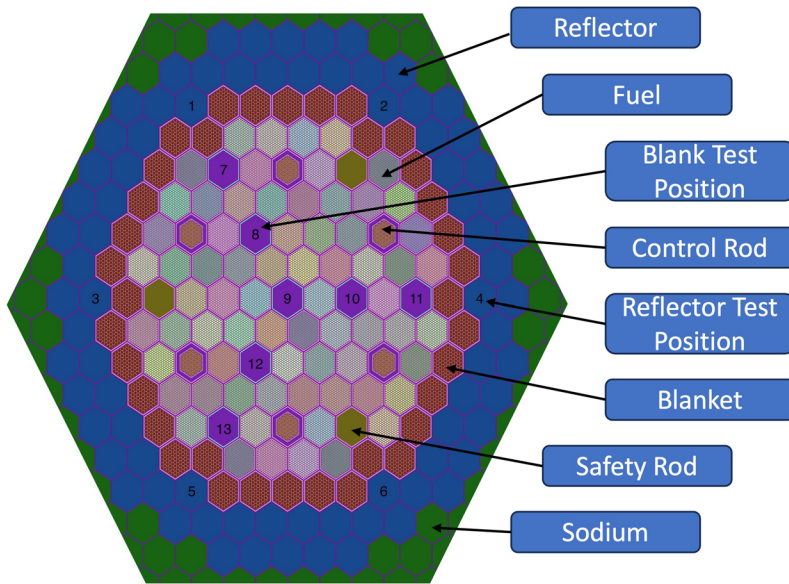


Figure 4. 300MW GSFR with breeding capabilities.

This study examines two potential DPAs: material diversion (i.e., declared SFM is removed from an assembly before being loaded into the core) and reactor misuse (i.e., undeclared fertile material was placed in the core to breed plutonium). The burner test reactor examines both material diversion and reactor misuse, while the breeder test reactor only examines reactor misuse. For the breeder test reactor, the inherent potential for breeding plutonium was determined to be the most likely scenario.

Material diversion was examined by allowing one, two, four, eight, or 12 fuel pins from a fresh fuel assembly to be replaced by 90 wt% natural uranium and 10 wt% zirconium. The diversion scenario created an optimization problem with approximately 30 design variables (one for each fresh assembly in the core). For material diversion, the fuel material is unattractive from a weapons perspective; however, it is still treated as SFM and, as such, is examined. While this analysis looks at a single fuel type, this framework could be extended without a loss of generality, where the results may change but the analysis techniques would remain constant.

Reactor misuse was characterized by the placement of undeclared fertile experimental assemblies (FEAs) into the core. FEAs were characterized by burnup variations ranging from fresh to twice burned, where previously irradiated assemblies were further broken down by their original placement in the core. The use of previously irradiated FEAs was envisioned to encompass previously declared experiments that were placed in the core under the guise of materials testing and were being placed back in the core to help generate 1 SQ. For reactor misuse,

each of the test cases allowed for a variable number of FEAs to be placed in the core.

The focus of this work was to devise a method for detecting material diversion and reactor misuse; as such, quantitative measures were required to differentiate normal reactor operations from off-normal conditions. Two major quantities of interest were the control rod insertion depth and assembly powers measured throughout the core. The assembly power (P_n) is defined as the power of the n^{th} assembly at a specific snapshot in time; as such, each core will have N assembly powers, where N is the total number of monitored assemblies. This assumes that the monitoring agency is only receiving a handful of data in specified intervals for their analysis. [Table 3](#) outlines the normal operating conditions for both the burner and breeder test reactors; the quantities of interest are generated from the nominal reactor physics model. Both numerical indications of the quantities of interest and three-sigma uncertainty for each quantity of interest in a reactor design without diversion or misuse occurring are shown in the table.

We assumed a cycle length of 400 days, where there is significant excess reactivity to allow for the placement of declared experiments in the core. For the assembly power measurements, the current assumption is that each of the fuel assemblies is monitored at the same interval (0, 200, and 400 effective full power days [EFPD]) as the control rod (CR) insertion depth. For this work, it is assumed that the power of every assembly is known. This could be characterized by flow monitors placed over each assembly or via some assembly power reconstruction based on a subset of measurements.²² The assumed uncertainty in each measurement is derived from either the physical uncertainty in detection limits or reasonable uncertainties that may arise during normal operations.²³ A detailed discussion of the uncertainties can be found in Stewart, et al., 2022.²⁴ Uncertainty provides an envelope for the state to perform off-normal operations under the guise of nominal operations.

To reiterate, obtaining appropriate sensors data will require a discussion between the state, IAEA, and reactor designer. This discussion will be imperative to ensure enough data is present for the SG-DT. Utilizing the virtual SG-DT will assist the IAEA in performing their safeguards analysis

Table 3. Quantities of interest for detecting reactor diversion and misuse for each configuration that was explored during this work, including CR insertion depths at 0, 200, 400, and 600 EFPD, assembly power, and cycle length.

	300 MW Burner Test SFR	300 MW Breeder Test SFR
0 EFPD CR Insertion Depth (cm)	42.6 ± 7.5	46.7 ± 7.5
200 EFPD CR Insertion Depth (cm)	28.7 ± 7.5	35.3 ± 7.5
400 EFPD CR Insertion Depth (cm)	11.8 ± 7.5	23.3 ± 7.5
Assembly Power	$P_n \pm 15.0\%$	$P_n \pm 15.0\%$

and allow for the insertion of detectors before the plant is built, highlighting the SG-DT's use in a safeguards-by-design analysis.

The quantities of interest were used to construct an optimization problem for the MABS system. For each test case, an optimal core would attempt to produce CR insertion depths and assembly powers within the uncertainty of the CR's insertion depth and assembly powers.

The reference GSFR burner test core allowed a total of 13 experimental positions, as shown in [Figure 3](#). Seven of the 13 allowable positions are represented by dedicated experimental facilities (i.e., located internally within the core). The remaining six encompass the positions in the periphery of the core and would replace the stainless-steel reflectors. The GSFR breeder test core also allows for 13 experimental positions (as seen in [Figure 4](#)); however, the positions in the periphery of the core are pushed back one row to allow for the inclusion of the blanket region.

Machine learning for proliferation detection

The SG-DT provides a unified approach for analyzing nuclear reactors to help detect if proliferation is occurring. The optimization algorithm acts as a rogue state and attempts to find core configurations that can produce 1 SQ of material within the MBP and within the uncertainty provided by [Table 3](#). This would constitute a core configuration that could evade detection using only the sensors from the core to detect off-normal operations. Once data is obtained from MABS, 75% of the data was used for the training the ML models and 25% was used for validation.

Utilizing this data, the ML adapter can provide three levels of information for misuse scenarios to the user through the implementation of a scikit-learn (Pedregosa et al, 2011) multilayer perception (MLP) algorithm:²⁵ core-level plutonium detection (MLP Regressor: kilograms), assembly-level plutonium detection (MLP Regressor: kilograms), assembly classification (MLP Classifier: fertile experiments, test positions, reflectors, etc.). Additionally, the ML adapter can provide assessments for diversion detection using similar MLP implementations. This includes identifying the quantity of pins diverted (MLP Classifier: one, two, four, eight, or 12 pins) and identifying if any diversion has occurred (MLP Classifier: diversion, no diversion). Along with each method, the accuracy (R^2) and explainability, through the implementation of Local Interpretable Model-Agnostic Explanations (LIME) are presented here.²⁶

Results

Results are presented in the next section for each of the test reactor designs for both diversion and misuse. These results combine the

determination of reactor configurations that can generate 1 SQ, as provided by the MABS, and a detailed analysis utilizing ML to understand how and where proliferation can occur, along with the confidence in the resulting detection.

300 MW burner test SFR

Case 1: Diversion

For the diversion scenario, each of the fresh assemblies was allowed to have one, two, four, eight, or 12 fuel pins replaced with a dummy pin. Figure 5 shows the Pareto front for the 300 MW burner test SFR material diversion scenario. The Pareto front is a set of optimal solutions obtained using a multiobjective optimization algorithm. These designs can concurrently optimize each objective without penalizing the other objectives.

For the burner test SFR diversion case, the MABS was able to find a very narrow Pareto front that only consists of seven entries. There were over 1,000 viable designs; however, these seven entries were present on the Pareto front, meaning they were more optimal than the other viable designs. These seven designs maintained a CR insertion depth roughly 2–3 cm below the nominal value. Given the approximate uncertainty of 7.5 cm, these designs would likely appear as nominal. The consistent 2–3 cm insertion depths below the nominal values are due to the reduction of fuel in the core, which causes the fuel that is present to accrue more

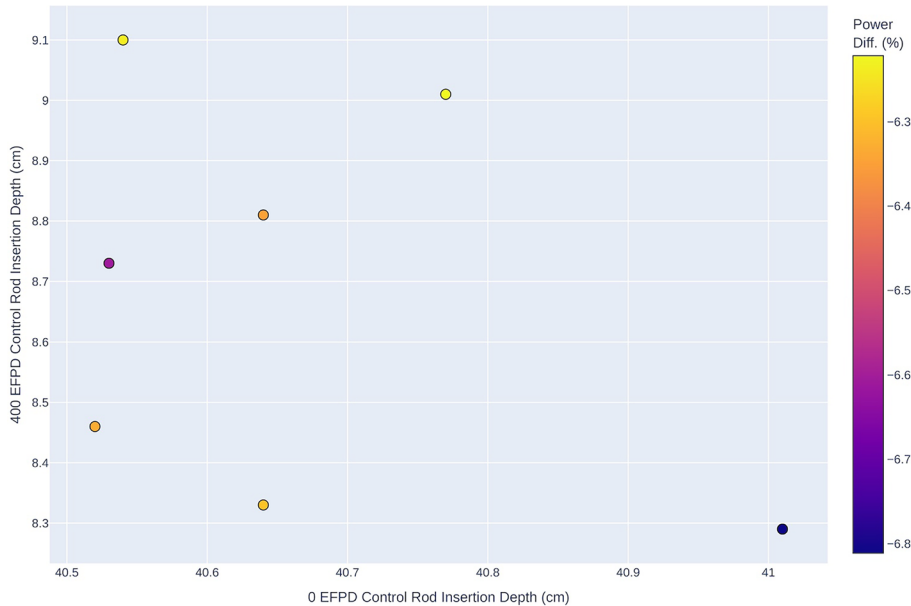


Figure 5. Pareto front for reactor diversion using the burner test SFR, comparing CR insertion depths and assembly power differences.

burnup in a given time span. The increased burnup then burns more fuel, which requires the CRs to be withdrawn further from the core.

For the power difference, the diversion cases hover between -7.0% and 4.5% ; only the largest negative differences are shown. Physically, the assemblies with negative power differences contain the dummy pins, and as such produce less power; to make up for the loss in power from assemblies with dummy pins, the remaining assemblies produce slightly more power, resulting in the positive power differences. For the diversion case, the designs on the Pareto front required between 16 and 19 fuel assemblies to divert fuel pins. This indicates that 83–89 fuel pins were diverted to generate 1 SQ of plutonium.

Figure 6 shows the accuracy of the ML adapter for predicting if diversion was occurring in each of the fresh fuel assembly positions. Assemblies highlighted by the accuracy were fresh fuel assemblies that could have fuel diverted. Blank assemblies constituted all other types of assemblies not examined for diversion. Figure 7 shows the accuracy of the ML adapter for predicting the number of pins diverted from the fresh assemblies. The ML adapter has a high accuracy ($>90\%$) rating for both predicting diversion and the number of pins. This accuracy is due to monitoring each of the fresh assembly powers, which allowed the ML adapter to accurately capture power differences in these diverted positions.

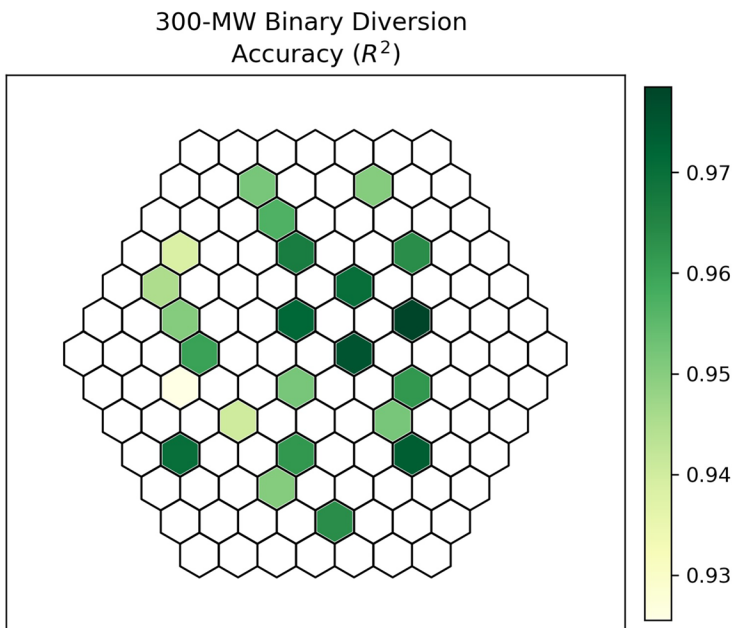


Figure 6. Accuracy of the ML adapters prediction if diversion occurred in each available position for the 300MW burner diversion case.

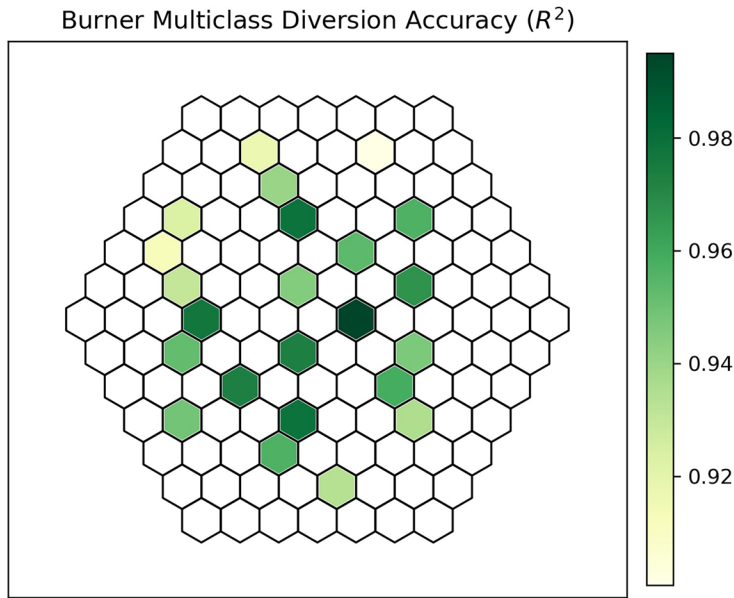


Figure 7. Accuracy of the ML adapters prediction of the number of pins diverted in each available position for the 300MW burner diversion case.

Case 2: Misuse

For the misuse scenario using the 300MW burner test core, FEAs could be placed in the six periphery or seven inner core region locations. The Pareto front acquired from the MABS can be seen in [Figure 8](#).

For the designs on the Pareto front, the CR insertion depths at 0 EFPD are within 1 cm of the nominal core. This deviation grows as a function of time where, at 400 EFPD, the CR insertion is on there is a 1–2 cm higher than the nominal case. This indicates that there is more fissile material in the core at the end of 400 EFPD due to plutonium being bred into the FEAs.

For the assembly power difference, the majority of designs have power differences within approximately 5% of the expected assembly power, where a few designs show larger deviations. The designs with a larger power deviation have FEAs placed in the last row of the core rather than in the periphery locations. This causes a larger perturbation to the flux, which in turn causes larger differences in assembly power.

For reactor misuse, between three and six assemblies are required to generate 1 SQ of plutonium. This resulted in between 8 and 20 kg of plutonium being generated in a 400 day period.

[Figure 9](#) shows the accuracy of the ML adapter to predict misuse for each assembly position. Assemblies highlighted by the accuracy were experimental positions or reflectors where an FEA could be placed. Blank

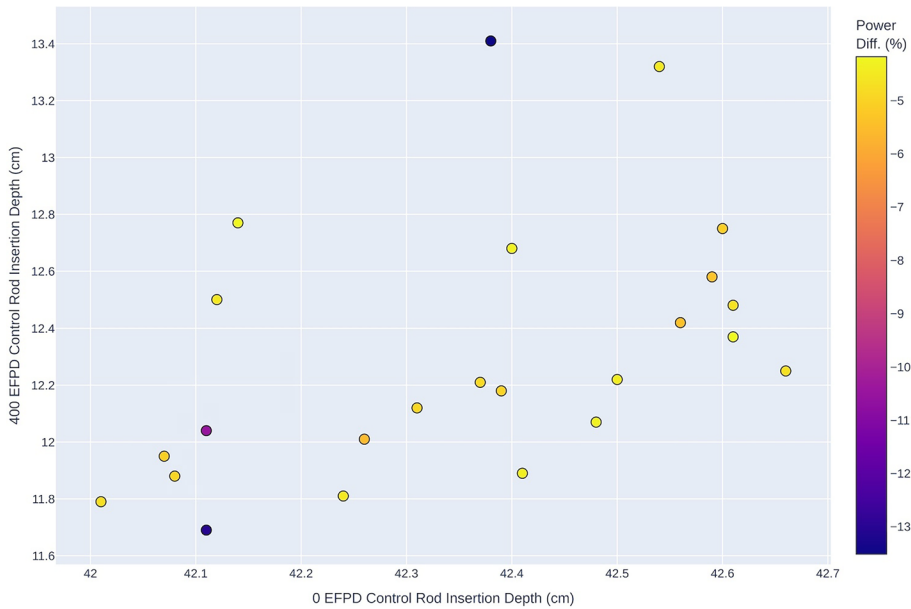


Figure 8. Pareto front for reactor misuse using the burner test SFR, comparing CR insertion depths and power difference.

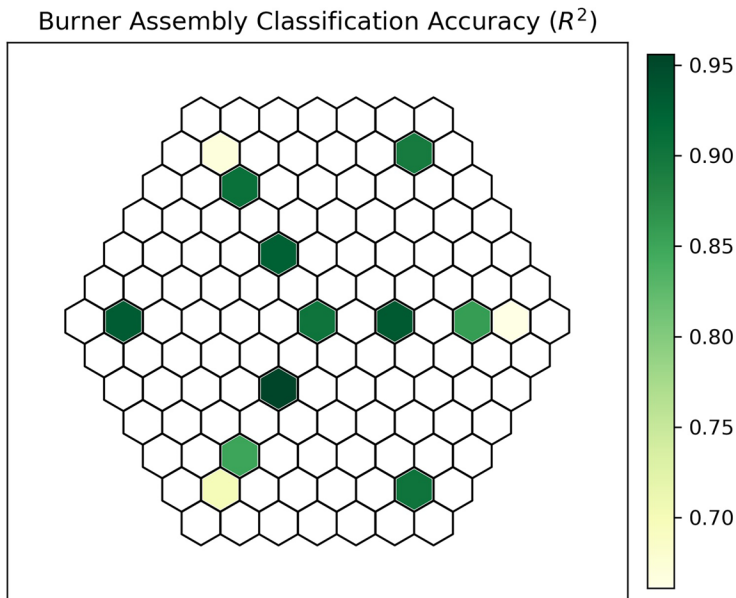


Figure 9. Accuracy of the ML adapters prediction of assembly type in each available position for the 300MW burner misuse case.

assemblies constituted all other types of assemblies not examined for misuse. [Figure 10](#) shows the accuracy in predicting the amount of plutonium generated in each assembly position. The center of the core has a

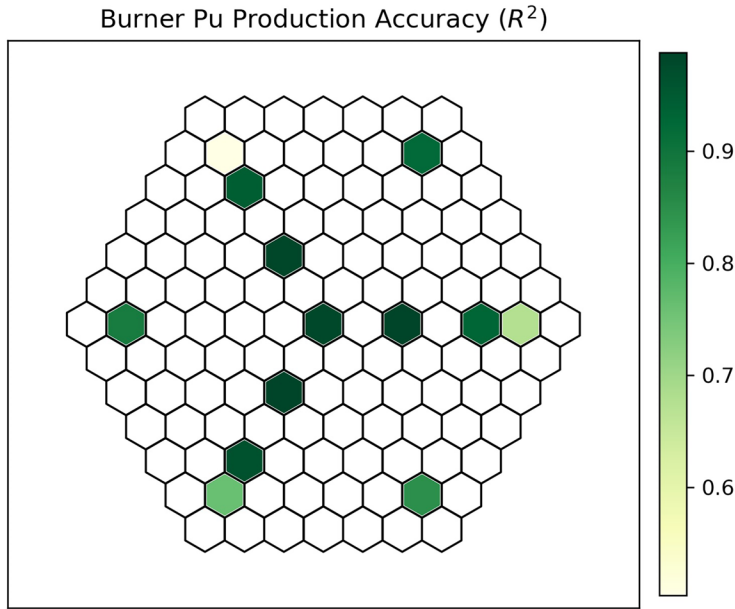


Figure 10. Accuracy of the ML adapters prediction of plutonium generation in each available position for the 300MW burner misuse case.

much higher accuracy than the periphery positions. Of note is the fact that the most difficult assemblies to detect lie in the reflector region and are adjacent to test positions. These positions have less data to draw conclusions from because they are surrounded by fewer sensors and therefore are more difficult to accurately predict.

300 MW breeder test SFR

Case 1: Misuse

For the misuse scenario, FEAs could be placed in the six periphery or seven inner core region locations. The Pareto front acquired from the MABS is in [Figure 11](#).

For the breeder case, a multitude of solutions appear within one standard deviation from the expected 0 and 400 EFPD CR insertion depths of 46.69 and 23.34 cm, respectively. The same holds true for the CR insertion depth at 200 EFPD. For the maximum power differences throughout the cycle, many solutions are on the cusp of one standard deviation, with a maximum difference of ~6.5%. Optimal designs placed FEAs in the periphery of the core, where between three and six assemblies were required for obtaining 1 SQ of plutonium. For the breeder core, this meant placing FEAs outside the blanket region.

[Figures 12 and 13](#) shows the accuracy in predicting the assembly type and the quantity of plutonium for each position. Two distinct regions,

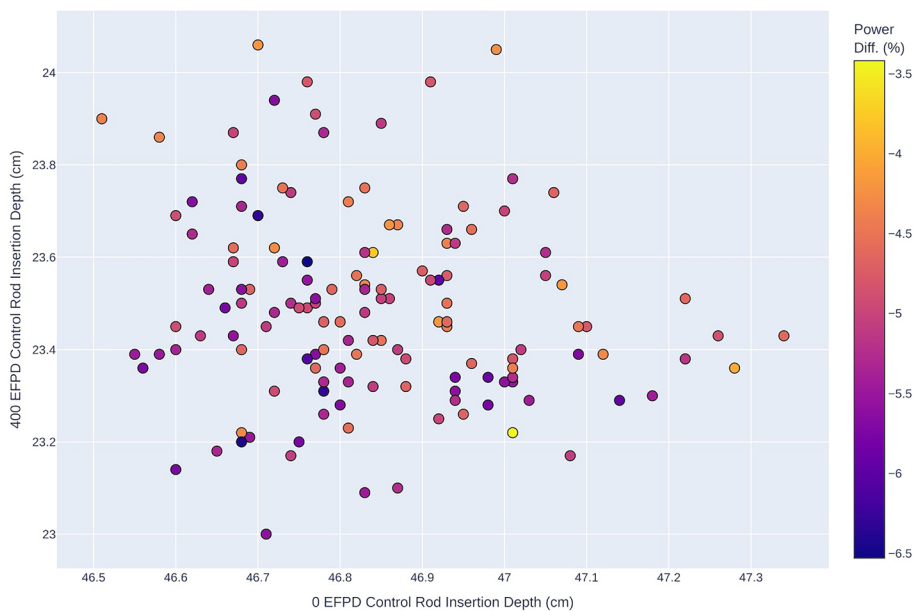


Figure 11. Pareto front for reactor misuse using the breeder test SFR, comparing CR insertion depths and power differences.

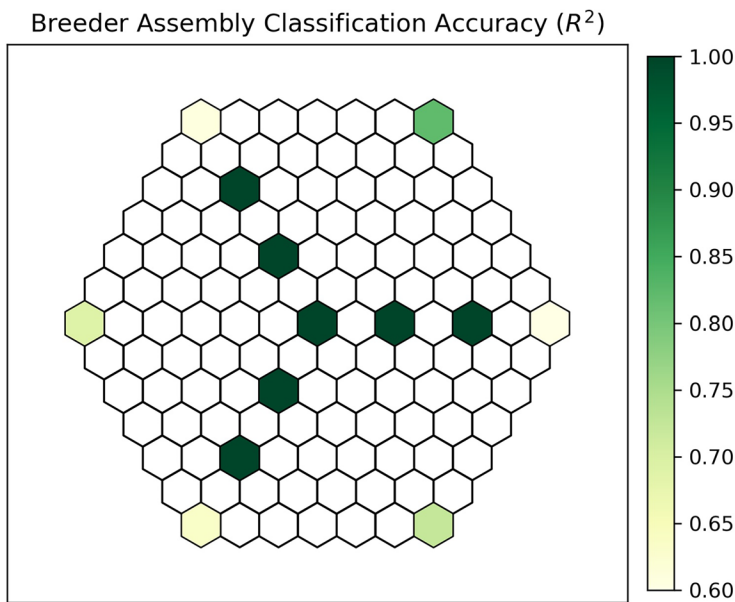


Figure 12. Accuracy of the ML adapters prediction of assembly type in each available position for the 300mw breeder misuse case.

specifically the inner and outer cores, were recognized as plutonium prediction zones. In the center of the core, the ML adapter can accurately predict the amount of plutonium generated for each position. As we move

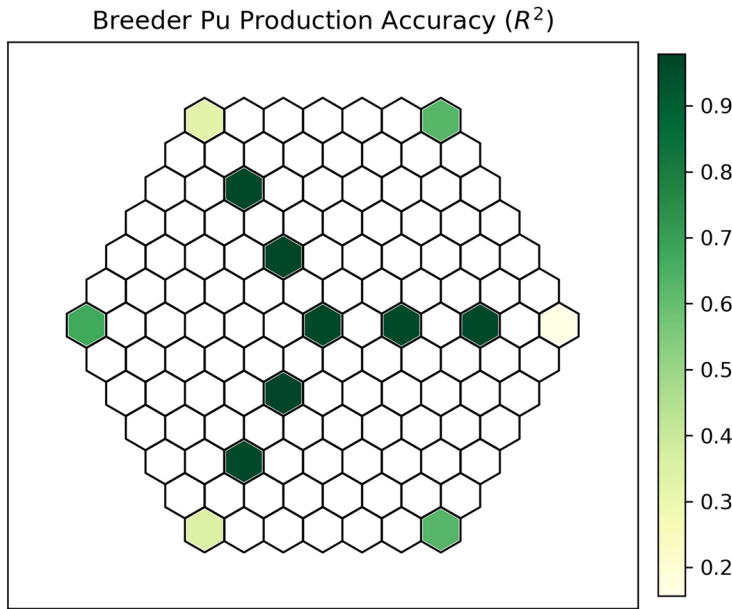


Figure 13. Accuracy of the ML adapters prediction of plutonium generation in each available position.

to the periphery of the core, where FEAs would be placed in the same row as the blanket locations, the ML model fails to accurately predict what is occurring. To explain this, we must rely on what is physically occurring in the system. These assemblies are far away from the core, and inserting an FEA into the periphery likely has a very small perturbation on the CR insertion depth, as seen in [Figure 11](#). While power detectors were placed in the blanket assemblies, the blanket assemblies are producing an order of magnitude less power than the fuel assemblies in the core; due to this, they are less sensitive to perturbations in local power differences. Finally, the FEAs in the corner assemblies are absorbing fewer neutrons to begin with since they are now separated by blanket assemblies and the core (in the plutonium production rate), and thus FEAs are also of lower importance and are more difficult to distinguish total plutonium production rates.

Detector selection and explainable machine learning

The ability for ML models to “see” into an SFR core is advantageous on its own, but beyond that benefit, the results can be leveraged for safeguards by design. As part of this work, the LIME explainability packages were used to develop an approach for performing instrument configuration assessments. LIME was performed on 440 reactor designs and identified the top 10 most significant variables (i.e., assembly power locations) for

Pu Quantity Model: Variable Importance
LIME top 10 normalized frequency (σ)

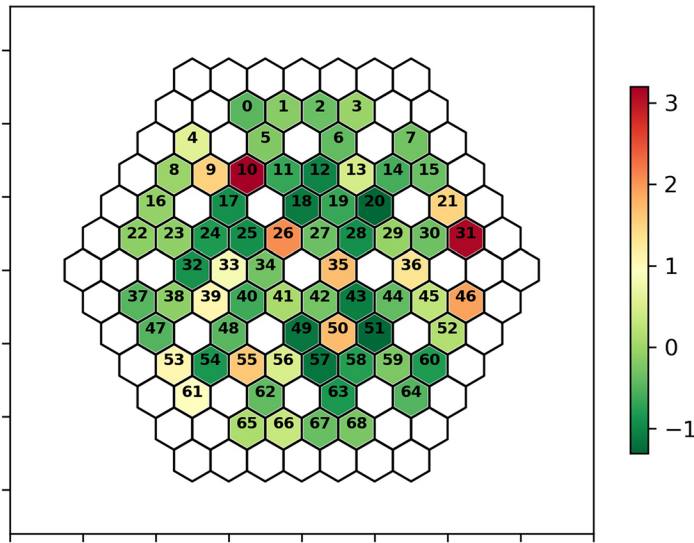


Figure 14. Map showing the variable importance for determining the amount of plutonium produced at each assembly location in an SFR core, where red locations indicate the most important locations and green the least important locations.

each design to determine the overall importance of each sensor location. These results were summarized and sorted based on how often each location appeared within the top 10 variables. Models were iteratively generated, eliminating the least important sensor location at each iteration, and R^2 values were generated. Though more sophisticated methods for variable selection exist, this approach has the benefit of being intuitive and easy to explain both conceptually and visually and therefore carries a level of trustworthiness for end users that is difficult to achieve using ML methods.

To highlight the methodology, we examined the amount of plutonium generated at each assembly location within the GSFR burner core. Figure 14 shows the variable importance for 69 locations as the standard deviation of location frequencies, with the red locations being the most important and the green locations being the least important. Sixty-nine models were eliminated by iteratively eliminating the least important assembly power location. Figure 15 shows these results, where the number of assembly power locations is shown on the x-axis and the resultant R^2 is shown on the y-axis. In this example, the target result of an R^2 value of 0.75 can be achieved using 18 assembly power locations, but similar results can be obtained with as few as 15 assembly power locations. This figure also demonstrates that, in this example, any additional monitored assembly power locations beyond 36 have a low impact on accuracy improvements.

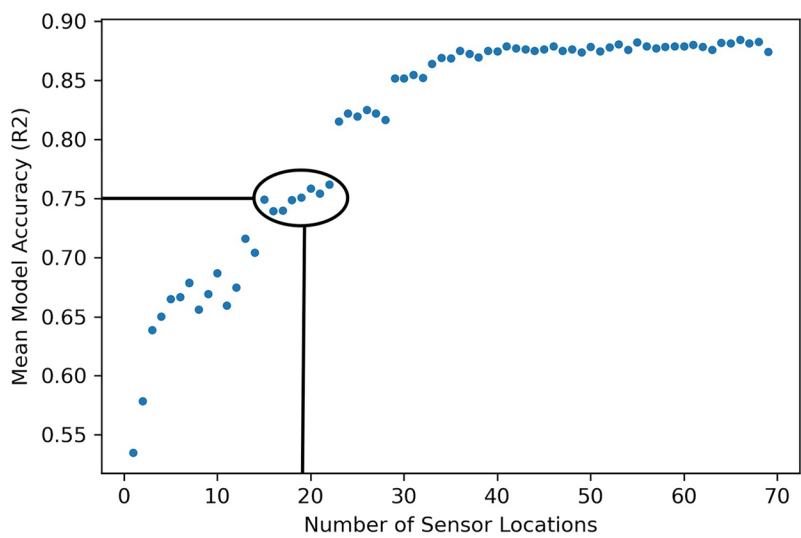


Figure 15. Graph showing the R^2 value vs. the number of sensor locations included in model development, where the grey box indicates the number of locations required to achieve an R^2 of 0.75 and the dark circle shows the range of sensor locations that provide similar results.

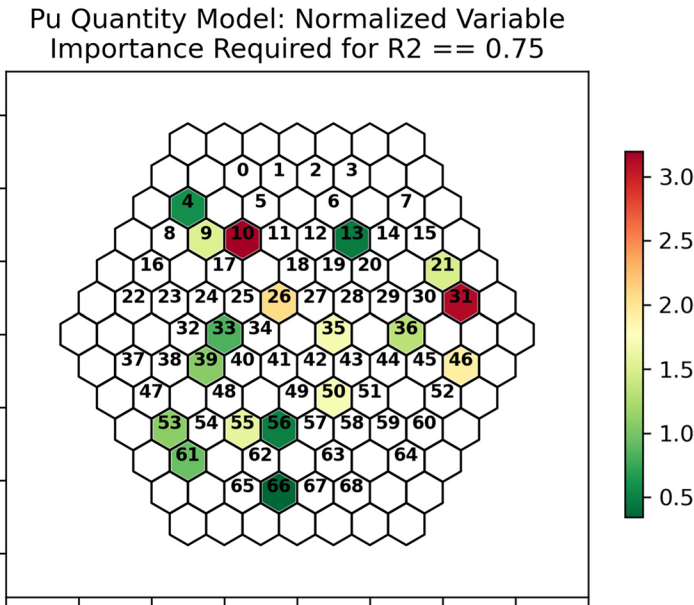


Figure 16. Map showing the recommended sensor configuration for achieving a $0.75 R^2$ value, with the most important locations shown in red and the least important in green.

The final monitored assembly power configuration recommended to achieve a $0.75 R^2$ value is shown in [Figure 16](#), where the most important locations are shown in red and the least important locations in green. It's worth noting that it is unlikely in a human-driven approach (rather than

a data-driven approach) that this configuration would be identified, making our approach advantageous for determining the optimal scenario for monitoring an SFR core with confidence, reducing the frequency of inspection-related interruptions to operations, and providing an informative tool for inspectors to leverage in their verification strategies.

Discussion

The results highlighted in the previous section provided two important pieces of information. First, detecting FEAs in the central region of the core would likely be manageable using the SG-DT. Second, predicting plutonium generation in the periphery of the core was significantly more difficult for each of the scenarios. While these results highlight limitations on the model's ability to accurately predict plutonium generation in the periphery of the core, vital information can still be gained for the center of the core and the results in the previous section can help inspectors know where to scrupulously inspect during visits. For example, the inspector can be confident of what is observed in the interior of the core, allowing them to target locations that require additional verification attention but with the understanding that the exterior portion will by default require additional attention regardless of our results.

The SG-DT is not meant to replace inspectors who need to physically inspect a nuclear facility. Instead, it is meant to provide a focus for inspectors to reduce the time required to inspect and verify in a targeted manner. Given the zero growth for inspection funding, and potential for new reactors coming online, it is imperative that inspectors can spend their limited time at a facility efficiently.

From a monitoring perspective, the SG-DT provides a significant amount of data, which would provide an inherent safeguards-by-design reactor if the diversion pathway results are incorporated in the design while in the conceptual design phase. Incorporating various sensors (such as assembly power) and creating a data feed to monitor the CR insertion depths would prevent unnecessary additions during the life cycle of the reactor, which would likely cost significantly more. Along with this, incorporating multiple sensors during the design process would allow plant operators to create a DT-like device for their own needs. This could include aiding with automated controls, reducing operating and maintenance costs, and scheduling preventative maintenance during outages.

The SG-DT provides a significant amount of data to help perform a DPA for the design and operations of a nuclear reactor. One major question to explore will be the potential misuse of an SG-DT. Given the sensitivity of this type of information, it may be prudent to shield the inner works of the DT from a state. Similarly, we imagine that a state

with access to the same data could potentially develop a similar DT to thwart the IAEA. In either case, if the state has access to the same data stream, it would be prudent for the IAEA to have the same if not more enhanced capabilities to ensure the state does not develop a viable diversion pathway. To this end, the DT would be one tool in addition to the others available to the IAEA to help detect undeclared diversion; ideally, the amalgamation of the data from the DT and other tools will provide a clear understand of how the facility is being operated.

Concluding remarks

We developed an SG-DT to examine three proliferation scenarios for a set of GSFRs, such that an inspector would be able to model a design provided by a monitored state. This design could then be examined in detail to determine how or where proliferation could occur within a GSFR. Expert knowledge could be applied to the design to help reduce proliferation scenarios.

Material diversion and reactor misuse were examined for burner and breeder test SFRs. For each case, we used the MABS to probe the design space and determine reactor configurations that would be difficult to detect due to uncertainties placed on the various sensors. For each of the scenarios using each of the GSFR types, the MABS was able to find a set of optimal solutions that would likely be difficult for an inspector to discern by hand. To help further understand the proliferation scenarios, the ML adapter used the data produced by the MABS to create a predictive model. The ML model provided three levels of understanding: if production or diversion of unaccounted material was occurring in the core, where that production or diversion was likely occurring, and the accuracy of those predictions. [Table 4](#) shows an overview of the ML model’s ability to detect the various levels of understanding for the SFR burner and breeder cases. While the ML model was not able to predict proliferation with near 100% accuracy, we did find that certain regions of the core were more attractive for proliferation from an adversary standpoint. For the reactor misuse case, this region was the periphery of the core; whereas in the center of the core, it was typically easy to discern if undeclared materials were placed in the core.

Table 4. Overview of the average probability to detect diversion and misuse in an assembly using the SG-DT.

	Binary Diversion (R^2)	Multiclass Diversion (R^2)	Misuse at Core (R^2)	Misuse at Assembly (R^2)
SFR Test Burner	0.96	0.95	0.99	0.89
SFR Test Breeder	N/A	N/A	0.93	0.73

The SG-DT was created to provide an investigation for an additional toolkit for safeguards inspectors to both perform an acquisition pathway analysis and then to use for monitoring purposes on reactor data streams with ML analysis and remote data transmission to the IAEA for a timely response capability. The SG-DT has the potential to help produce reactor cores, which incorporates the safeguards-by-design approach and would benefit both the monitoring agency and the operating state. By incorporating the necessary sensors and detectors for the SG-DT, the state would also gain valuable operating data for the generation of their own DT, providing a further incentive for the use of an SG-DT. We hope that DTs would be utilized in conjunction with inspections to help provide targeted regions for scrupulous investigations. The SG-DT is only the first step for developing a framework that would be adopted by the IAEA; additional research and discussion are imperative to make DT technologies viable for international safeguards.

Notes

1. IAEA, "The Agency's Budget Update for 2023," International Atomic Energy Agency, Vienna, 2022.
2. IAEA, "Energy, Electricity and Nuclear Power Estimates for the Period up to 2050," International Atomic Energy Agency, Vienna, 2021; IAEA, "The Agency's Budget Update for 2023."
3. L. Cheng, "Gen-IV PR&PP," Brookhaven National Laboratory, 2022.
4. L. Cheng, "Gen-IV PR&PP"; R. Stewart et al., "Utilizing digital twin for nuclear safeguards and security," in *63rd Annual INMM Meeting*, Vienna, 2023.
5. IAEA, "Safety of Nuclear Power Plants: Design," International Atomic Energy Agency, Vienna, 2016; IAEA, "Technical Challenges in the Application and Licensing of Digital Instrumentation and Control Systems in Nuclear Power Plants," International Atomic Energy Agency, Vienna, 2015; IAEA, "Options to Enhance Proliferation Resistance of Innovative Small and Medium Sized Reactors," International Atomic Energy Agency, Vienna, 2014.
6. "Digital Engineering Strategy," Department of Defense, Washington, D.C., 2018.
7. V. Yadav et al., "The State of Technology of Application of Digital Twins," U.S. Nuclear Regulatory Commission, Washington, D.C., 2021; V. Yadav et al., "Technical Challenges and Gaps in Digital-Twin-Enabling Technologies for Nuclear Reactor Applications," U.S. Nuclear Regulatory Commission, Washington, D.C., 2021.
8. Ibid
9. P. Apte, "Digital Twins of Nuclear Power Plants," The American Society of Mechanical Engineers, 2021. [Online]. Available: <https://www.asme.org/topics-resources/content/digital-twins-of-nuclear-power-plants>; L. Thompson, "Argonne to explore how digital twins may transform nuclear energy with \$8 million from ARTP-E's GEMINA program," Argonne National Laboratory, 2020. [Online]. Available: <https://www.anl.gov/article/argonne-to-explore-how-digital-twins-may-transform-nuclear-energy-with-8-million-from-arpaes-gemina>; B. Kochunas and X. Huan, "Digital Twin Concepts with Uncertainty for Nuclear Power Applications," *Energies* 14 (2021): 4235.

10. J. Leppanen et al. "The Serpent Monte Carlo code: Status, development and applications in 2013," *Annals of Nuclear Energy* 82 (2015): 142–150.
11. J. Leppanen et al., "The Serpent Monte Carlo code: Status, development and applications in 2013," *Annals of Nuclear Energy* 82 (2015): 142–150.
12. R. Stewart et al, "A digital twin of the AGN-201 reactor to simulate nuclear proliferation," in *63rd Annual INMM Meeting*, Vienna, 2023.
13. J. Darrington, J. Browning and C. Ritter, "Deep Lynx: Digital Engineering Integration Hub," U.S. DOE Office of Energy Efficiency and Renewable Energy, 2020.
14. J. Browning et al., "Data Integration Aggregated Model for Nuclear Deployment," U.S. DOE Office of Nuclear Energy, 2020.
15. IAEA, "IAEA Safeguards Glossary," IAEA, Vienna, 2022.
16. Ibid.
17. Ibid.
18. E. Hoffman, W. Wang and R. Hill, "Preliminary Core Design Studies for the Advanced Burner Reactor over a wide range of conversion ratios," Argonne National Laboratory, ANL-AFCI-177, Argonne, 2006; Y. Chang, P. Finck and C. Grandy, "Advanced Burner Test Reactor Preconceptual Design Report," Argonne National Laboratory, Argonne, 2006; N. Martin, S. Stewart and S. Bays, "A multiphysics model of the versatile test reactor based on the MOOSE framework," *Annals of Nuclear Energy* 172 (2022): 109066.
19. R. H. Stewart et al., "Examination of Diversion and Misuse Detection for a Generalized Sodium-Cooled Fast Test Reactor," *Journal of Nuclear Material Management* 2 (2022).
20. N. Martin, S. Stewart and S. Bays, "A multiphysics model of the versatile test reactor based on the MOOSE framework."
21. R. H. Stewart et al., "Examination of Diversion and Misuse Detection for a Generalized Sodium-Cooled Fast Test Reactor."
22. S. Nomoto, H. Yamamoto, Y. Sekiguchi and Tamura, "Measurement of subassembly outlet coolant temperature in the JOYO experimental fast reactor," *Nuclear Engineering and Design* 62 (1980): 233–239; C. Day, "FFTF core and primary sodium circuit instrumentation," Hanford Engineering Development Laboratory, HEDL-SA-1082, Richland, WA, 1975; A. Chenu, R. Adams, K. Mikityuk and R. Chawla, "Analysis of selected Phenix EOL tests with the FAST code system – Part I: Control-rod-shift experiments," *Annals of Nuclear Energy* 49 (2012): 182–190.
23. K. Korsah et al, "Assessment of Sensor Technology for Advanced Reactors," Oak Ridge National Laboratory, ORNL/TM-2016/337, Oak Ridge, 2016; Nuclear Regulatory Commission, "Westinghouse Technology Systems Manual (R304P), Section 3.1.4, Control Rod Drive Mechanisms," Nuclear Regulatory Commission, Technical Training Center; C. Day, "FFTF core and primary sodium circuit instrumentation," Hanford Engineering Development Laboratory, Richland, 1975.
24. R. H. Stewart et al., "Examination of Diversion and Misuse Detection for a Generalized Sodium-Cooled Fast Test Reactor."
25. F. Pedregosa et al, "Scikin-learn: Machine learning in python," *The Journal of Machine Learning Research* 12 (2011): 2825–2830.
26. M. Ribeiro, S. Singh and C. Guestrin, "“Why should i trust you?” Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2016.

Acknowledgments

This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government

retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes. This research made use of the resources of the High-Performance Computing Center at Idaho National Laboratory, which is supported by the Office of Nuclear Energy of the U.S. Department of Energy and the Nuclear Science User Facilities under contract no. DE-AC07-05ID14517. The authors would like to thank the editors and reviewers for their insights and comments. Their efforts significantly added to the quality of this article.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ryan Stewart  <http://orcid.org/0000-0003-4867-6555>
Ashley Shields  <http://orcid.org/0000-0002-9039-4706>
Frederick Gleicher  <http://orcid.org/0000-0001-6483-3931>
Samuel Bays  <http://orcid.org/0000-0003-4086-4594>
Jeren Browning  <http://orcid.org/0000-0002-0482-4395>
Katherine Jesse  <http://orcid.org/0000-0002-5862-4962>
Christopher Ritter  <http://orcid.org/0000-0002-5662-1506>