arXiv:2506.19863v2 [physics.comp-ph] 26 Jun 2025

Exploring the Capabilities of the Frontier Large Language Models for Nuclear Energy Research

Ahmed Almeldein, Mohammed Alnaggar, Rick Archibald, Tom Beck, Arpan Biswas, Rike Bostelmann, Wes Brewer, Chris Bryan, Christopher Calle, Cihangir Celik, Rajni Chahal, Jong Youl Choi, Arindam Chowdhury, Mark Cianciosa, Franklin Curtis, Gregory Davidson, Sebastian De Pascuale, Lisa Fassino, Ana Gainaru, Yashika Ghai, Luke Gibson, Qian Gong, Christopher Greulich, Scott Greenwood, Cory Hauck, Ehab Hassan, Rinkle Juneja, Soyoung Kang, Scott Klasky, Atul Kumar, Vineet Kumar, Paul Laiu, Calvin Lear, Yan-Ru Lin, Jono McConnell, Furkan Oz, Rishi Pillai, Anant Raj, Pradeep Ramuhalli, Marie Romedenne, Samantha Sabatino, José Salcedo-Pérez, Nathan D. See, Arpan Sircar, Punam Thankur, Tim Younkin, Xiao-Ying Yu, Prashant Jain, Tom Evans, Prasanna Balaprakash*

Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

{almeldeinam, alnaggarmg, archibaldrk, becktl, biswasa, bostelmannf, brewerwh, bryancd, calleci, celikc, chahalr, choij, chowdhurya1, cianciosamr, curtisfg, davidsongg, depascuales, reedke, gaianrua, ghaiy, gibsonld, gongq, greulichcr, greenwoodms, hauckc, hassanem, junejar, kangs1, klasky, kumara, kumarv, laiump, learcr, liny, mcconnelljh, ozf1, pillairr, raja, ramuhallip, romedennem, sabatinosm, salcedoperjl, seend, sircara, thakurp, younkintr, yuxiaoying, jainpk, evanstm, pbalapra}@ornl.gov

June 30, 2025

Abstract

The AI for Nuclear Energy workshop at Oak Ridge National Laboratory evaluated the potential of Large Language Models (LLMs) to accelerate fusion and fission research. Fourteen interdisciplinary teams explored diverse nuclear science challenges using ChatGPT, Gemini, Claude, and other AI models over a single day. Applications ranged from developing foundation models for fusion reactor control to automating Monte Carlo simulations, predicting material degradation, and designing experimental programs for advanced reactors. Teams employed structured workflows combining prompt engineering, deep research capabilities, and iterative refinement to generate hypotheses, prototype code, and research strategies. Key findings demonstrate that LLMs excel at early-stage exploration, literature synthesis, and workflow design, successfully identifying research gaps and generating plausible experimental frameworks. However, significant limitations emerged, including difficulties with novel materials designs, advanced code generation for modeling and simulation, and domain-specific details requiring expert validation. The successful outcomes resulted from expert-driven prompt engineering and treating AI as a complementary tool rather than a replacement for physics-based methods. The workshop validated AI's potential to accelerate nuclear energy research through rapid iteration and cross-disciplinary synthesis while highlighting the need for curated nuclear-specific datasets, workflow automation, and specialized model development. These results provide a roadmap for integrating AI tools into nuclear science workflows, potentially reducing development cycles for safer, more efficient nuclear energy systems while maintaining rigorous scientific standards.

*Contact Prasanna Balaprakash (pbalapra@ornl.gov) for workshop-related questions; the authors of each section related to the problem are explicitly identified in the report.

1 Introduction

The intersection of artificial intelligence and nuclear energy represents a frontier of immense potential for accelerating scientific discovery and engineering innovation. Both fusion and fission technologies stand as critical pillars for achieving energy abundance and economic prosperity. The promise of virtually limitless, reliable baseload power from nuclear sources could fundamentally transform global energy markets and enable new scales of industrial capability. However, the complexity of nuclear systems—spanning plasma physics, materials science, neutronics, thermal hydraulics, and structural mechanics—presents formidable challenges that traditionally require decades of research and development. The emergence of Large Language Models (LLMs) and advanced AI reasoning systems offers an opportunity to compress these timelines while maintaining the rigorous safety and reliability standards essential to nuclear applications.

The “AI for Nuclear Energy" workshop, convened at Oak Ridge National Laboratory on May 1, 2025, was designed to systematically evaluate how state-of-the-art LLMs could enhance nuclear science research workflows. This one-day workshop brought together experts from diverse disciplines—nuclear engineering, computational science, materials science, and machine learning—to explore practical applications of LLMs in addressing nuclear energy research challenges. The timing of this workshop was particularly significant, as it coincided with rapid advances in AI capabilities, including enhanced reasoning models, deep research functionalities, and improved code generation tools. The workshop's primary objective was to assess whether AI could meaningfully accelerate early-stage research in nuclear energy applications while identifying both the opportunities and limitations of current technology. Unlike traditional computational tools that excel at numerical simulation, LLMs offer unique capabilities in synthesizing vast bodies of literature, generating initial hypotheses, identifying research gaps, and facilitating cross-disciplinary communication—all critical activities in the complex landscape of nuclear research.

Fourteen teams tackled diverse challenges spanning the nuclear energy spectrum: from developing foundation models for real-time fusion reactor control to predicting radiation-induced material degradation, from automating complex Monte Carlo simulations to designing experimental programs for advanced reactor concepts. Each team employed structured workflows that combined prompt engineering, iterative refinement, and careful validation to explore how AI tools could complement traditional physics-based approaches. This report documents the methodologies, findings, and insights from this collaborative effort. We present detailed case studies from each team, highlighting successful applications where AI demonstrated clear value—such as literature synthesis, workflow design, and initial code prototyping—as well as areas where current limitations became apparent, including challenges with novelty, technical accuracy, and the need for expert validation. The synthesis of these experiences provides an advanced understanding of AI's current role in nuclear science: not as a replacement for domain expertise or rigorous physical modeling, but as an accelerant for scientific exploration.

The implications of this work extend beyond the immediate findings. As U.S. race to achieve energy dominance through advanced nuclear technologies, the ability to rapidly explore design spaces, identify promising research directions, and integrate knowledge across disciplines becomes a strategic imperative. The potential for abundant, affordable nuclear energy to power massive data centers, advanced manufacturing, and even space exploration depends on our ability to accelerate the development cycle. This report serves as both a snapshot of current capabilities and a roadmap for future development, outlining how the nuclear community can harness AI tools while maintaining the exceptional standards of safety and scientific rigor that define this field.

Through careful documentation of successes, failures, and lessons learned, we aim to provide the nuclear science community with practical guidance for integrating AI into research workflows, ultimately contributing to the acceleration of energy abundance that will define the next era of human prosperity and technological advancement.

2 Executive summary

The workshop showed LLM's versatility across 14 distinct fusion and fission nuclear energy challenges, from fundamental fusion physics to practical reactor engineering. Teams explored applications spanning materials science, plasma control, neutronics modeling, and data management—representing the full spectrum of nuclear energy research and development.

Success Patterns

Workflow development and integration: Nearly all teams successfully used AI to design complex multiphysics workflows. Notable examples include integrating multiple simulation codes (XGC, NekRS, M3D-C1) for fusion modeling and coupling neutronics–thermal hydraulics–structural mechanics for PWR analysis. AI excelled at identifying data streams, coupling strategies, and process bottlenecks.

Knowledge synthesis and gap analysis: AI models demonstrated exceptional capability in literature review and research gap identification. Teams successfully identified missing experimental data for reactor corrosion, proposed experimental programs with detailed budgets, and ranked candidate materials based on multiple criteria. The medical isotope team correctly identified HFIR production capabilities from dispersed sources.

Hypothesis generation and initial design: Multiple teams generated novel research directions, including alpha particle channeling for fusion optimization, isotopic tailoring for reduced activation, and RF sheath mitigation strategies. The material degradation team achieved high predictive accuracy using AI-generated models.

Visual and structural generation: AI successfully created realistic microscopy images of material defects, atomic void structures, and STEM simulations that aligned with published literature, demonstrating potential for synthetic training data generation.

Limitations

Data and code access barriers: The most significant limitation across projects was the inability to access or process in-house nuclear data. Teams working with SCALE, MCNP, and other restricted codes could not generate functional inputs despite extensive prompting.

Implementation depth: While AI excelled at high-level design and framework development, detailed implementation consistently required human expertise. No team achieved production-ready code, with particular struggles in Python generation and mathematical solver development.

Data access and validation: Projects were hampered by paywalled literature, missing datasets, and the need for extensive expert validation. The fusion data annotation team could not access TRANSP datasets, while the corrosion team missed state-of-the-art studies behind paywalls.

Technical accuracy: Several projects encountered accuracy issues—incorrect parameter values for heat transfer, failed dislocation loop generation, problematic citations, and overly general responses that lacked domain specificity.

Implications

The workshop results reveal LLMs as an accelerator for early-stage nuclear research, particularly in: Rapid literature synthesis and knowledge integration; Cross-disciplinary workflow design; Experimental program planning; Initial hypothesis generation. However, the technology requires careful integration with domain expertise and cannot replace detailed technical implementation or rigorous validation. Success depends on developing secure AI infrastructure for classified data, creating nuclear-specific training datasets, and establishing validation protocols.

The diversity of successful applications, from fusion plasma control to fission material degradation, suggests that LLM integration could provide competitive advantages across the entire nuclear energy sector, potentially accelerating development timelines by effectively harness these capabilities.

Project Title	**Key Challenge**	**Main Outcomes**	**Key Limitations**
		Domain: Fusion	
Foundation Model for Fusion	Real-time fusion reactor control and simulation	Structured workflow for integrating XGC, NekRS, M3D-C1 simulations; identified data pipeline needs	Could not delve into implementation details; requires extensive compute resources

Project Title	Key Challenge	Main Outcomes	Key Limitations
Structural Materials & Molten Salt Compatibility	Material selection for extreme fusion environments	Ranked 5 candidate materials (W, RAFM steels, SiC/SiC); identified MHD interface challenges	No truly novel alloy proposals; limited to known materials
Isotopically Tailored Materials	Reducing radioactive activation in fusion reactors	Identified materials benefiting from isotope engineering; proposed experimental roadmap	Incomplete analysis; AI provided incorrect values for some parameters
Alpha Particle Confinement	Optimizing fusion self-heating from alpha particles	Proposed alpha channeling and divertor optimization strategies; identified experimental gaps	Model struggled with device-specific synthesis; some incorrect citations
Fusion Data Annotation	Semantic enrichment of fusion datasets	Developed RAG pipeline for TRANSP data; created annotation workflow	Could not access needed datasets; incomplete implementation
Multiphysics Modeling for MPEX	Plasma-material interaction modeling	Identified RF sheath erosion as key issue; developed inverse solver for diffusion coefficients	Limited to simplified 1D models; adjoint solver generation failed
Domain: Fission			
Monte Carlo Surrogate Models	Accelerating MCNP simulations	Proposed neural network surrogate architecture; identified data gathering strategies	No existing Shift-based ML surrogates; lacks public training data
Molten Salt Corrosion	Understanding corrosion in nuclear reactor salts	Literature review of MD/AIMD studies; identified need for MLIPs	State-of-the-art studies not captured; limited by paywall access
Material Degradation Prediction	Inverse modeling from acoustic/strain data	Achieved $R^2 \geq 0.90$ for degradation prediction; generated Mathematica ML code	Python code generation problematic; requires extensive validation
Experimental Program Design	AI-directed experiments for advanced reactors	Generated complete experimental program with budget; identified corrosion knowledge gaps	High-level only; lacks technical implementation details
Atomic Structure & S/TEM Imaging	3D reconstruction from electron microscopy	Generated atomic void structures; simulated STEM images	Failed to create accurate dislocation loops; limited to <10nm structures
Pressurized Water Reactor Power Uprate Analysis	Multiphysics framework for reactor uprates	Outlined neutronics-TH-fuel-structural coupling strategy; identified relevant tools	Failed with export-controlled software; could not generate working inputs
Microscopy Defect Imaging	Simulating defect images for SS304	Generated realistic sensitization and crack defect images	No source transparency; unclear if SS304-specific
Nuclear Isotopes & MCNP Automation	Medical isotope ranking and LiDAR-to-MCNP geometry	Correctly identified HFIR isotope capabilities; proposed geometry workflow	Missed latest research; overly general responses

3 Challenge problems explored at the workshop

This section present detailed case studies from 14 interdisciplinary teams that explored diverse applications of AI models across the nuclear energy spectrum. Each team tackled a specific challenge, employing structured workflows that combined prompt engineering, iterative refinement, and domain expertise validation. The projects span both fusion and fission applications, from fundamental physics to practical engineering solutions.

3.1 Towards a Foundation Model for Fusion

Team Members

Furkan Oz, ozf1@ornl.gov
Wes Brewer, brewerwh@ornl.gov
Christopher Calle, calleci@ornl.gov
Scott Klasky, klasky@ornl.gov

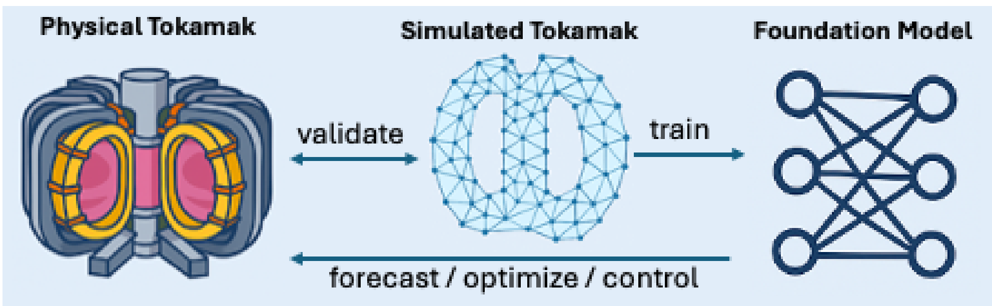


Figure 1: Conceptual overview: A foundation model (FM) trained on high-fidelity fusion simulations provides real-time control and receives validation feedback from a tokamak reactor.

At the start of the workshop, our team held a brainstorming session to identify a fusion-relevant challenge that would meaningfully integrate each member's area of expertise. We aligned on the development of a foundation model for fusion reactor simulation and control as a unifying goal, as illustrated in Figure 1. With backgrounds spanning workflows, magnetohydrodynamics (MHD) simulations, control theory, data management, and high-performance computing (HPC), and AI, each team member assessed their strengths and contributed to the initial problem framing.

To accelerate planning and task division, we provided our brainstorming notes and domain specializations to ChatGPT-4.5, which helped decompose the project into structured tasks and suggested how responsibilities could be distributed. Each team member then worked on their assigned component independently for 2-3 hours, identifying key requirements, developing domain-specific contributions, and collaboratively integrating these elements into a unified workflow. Iterative feedback and refinement—both among teammates and through continued interaction with ChatGPT—helped us converge on a coherent and technically sound concept paper in just a few hours.

The overall resulting workflow—from high-fidelity simulations through data preparation, pre-training, and fine-tuning—is summarized in Figure 2.

3.1.1 Problem Statement

Description: Current fusion simulation workflows are fragmented, relying on isolated high-fidelity codes (e.g., XGC [1], NekRS [2], M3D-C1 [3], OpenFOAM [4]) with incompatible data formats and inefficient communication strategies. This fragmentation leads to significant delays in data integration, limiting real-time prediction capabilities and slowing reactor design iteration cycles.

Motivation: Rapid and reliable integration of multiphysics simulation data is critical to meet DOE's strategic goal of developing fusion pilot plants by the early 2030s. By creating a unified data pipeline, we facilitate fast, iterative design optimization, drastically reducing simulation-to-insight latency and accelerating fusion energy research. The data pipeline will be constructured with a simliar schema which is mostly compliant with the International Thermonuclear Experimental Reactor (ITER) Integrated Modelling & Analysis Suite

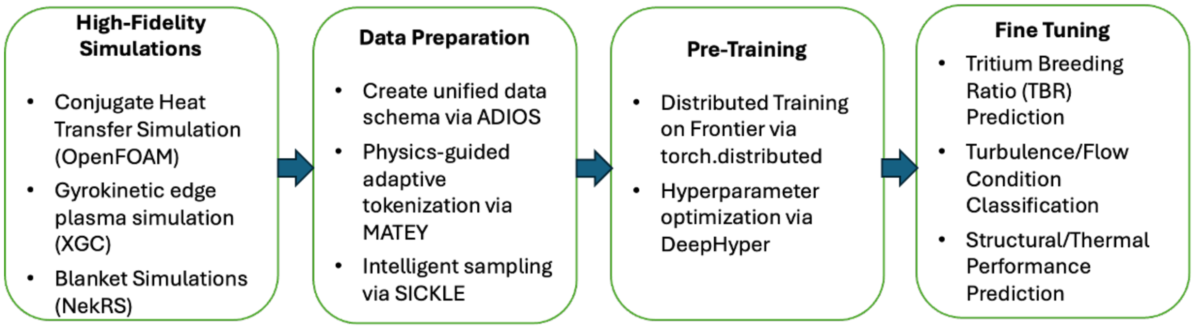


Figure 2: End-to-end workflow from high-fidelity simulations through data preparation and pre-training to fine-tuned foundation model tasks, illustrating how simulation, data, and AI components are integrated into a unified pipeline for scientific modeling.

(IMAS) standard [5] since IMAS is not readily suited for large data. Our extensions will be the inclusion of ADIOS into IMAS which will work for both stellarators and tokamak physics.

3.1.2 Methodology

- **AI Models & Tools:** Our team extensively utilized OpenAI's ChatGPT-4.5, providing initial notes and concepts about our workflow, simulations (XGC, NekRS, OpenFOAM, M3D-C1), and our goals around the IMAS-compliant data schema and HPC integration. ChatGPT assisted in structuring our ideas clearly into a goal statement, objectives, specific tasks, subtasks, and success metrics.
- **Workflow:**
  Data Inventory and Notes Gathering → Problem Formulation → Prompting and Clarification with ChatGPT → Iterative Refinement and Telescoping (in parallel) → Task and Objective Structuring → Success Metric Definition → LaTeX Integration and Formatting → Final Concept Paper Draft

  The team began with Data Inventory and Notes Gathering, compiling individual expertise, relevant simulation tools, and domain-specific challenges. This was followed by Problem Formulation, where we defined the high-level objective of developing a foundation model for fusion reactor simulations.

  We then engaged in Prompting and Clarification with ChatGPT to refine our ideas, decompose the problem into subcomponents, and explore alternative directions. Through Iterative Refinement and Telescoping, we continuously revised our project scope and structure based on AI-assisted feedback and peer discussion.

  Next, Task and Objective Structuring helped us assign responsibilities aligned with each member's domain, enabling parallel development. We then articulated clear Success Metric Definitions, outlining how to evaluate each component's contribution to the overall concept.

  LaTeX Integration and Formatting ensured our concept paper adhered to professional standards and supported seamless collaboration. Finally, we converged on a Final Concept Paper Draft, synthesizing all components into a cohesive document.
- **Team Roles:** - Furkan Oz: Magnetohydrodynamics simulations and Turbulence Modeling - Wes Brewer: AI Model Development and HPC Optimization - Christopher Calle: Plasma Control and Stability - Scott Klasky: Data and Workflow Integration, Project oversight, strategic guidance on exascale data management

3.1.3 Key Findings

- **Results:** Through structured interactions with ChatGPT, our team successfully produced a detailed, coherent concept paper outlining clear project goals, objectives, tasks, and relevant success metrics. ChatGPT facilitated generating concise, accurate language to communicate complex technical concepts clearly, including a detailed LaTeX-formatted workflow figure caption and integration guidance.
- **Insights:** We found that by having a talented and diverse team, along with a structured layout for how each team member would interact with ChatGPT, and a well-defined concept paper template, we could produce a concept paper in just a few hours. ChatGPT helped to synthesize ideas and bring them together in a unified manner.

3.1.4 Limitations & Challenges

- **Technical Constraints:** Due to the complexity of what was being proposed, we were not able to delve into any of the implementation details, but rather focused on the high-level conceptual strategy. Training a foundation model will require running large-scale simulations on Frontier, and then curating training datasets, etc., which involve minimally tens of thousands of node hours.
- **Process Hurdles:** One challenge we encountered was in using ChatGPT-4.5 to generate a visual workflow diagram illustrating the pipeline from simulation to data preparation, pre-training, and fine-tuning Figs. 1 and 2. While the model initially produced a relevant flowchart, the image was cropped and partially cut off. Subsequent attempts resulted in inconsistent diagrams with similar formatting issues. We found it much easier to generate the separate components of the diagrams and then piece them together in PowerPoint. This experience highlighted a clear opportunity for improving OpenAI's diagram generation capabilities—particularly for producing consistent, high-quality visual content that would greatly benefit concept papers like ours.
- **Mitigations Tried:** To address the technical scope limitations, we prioritized defining a clear, modular task structure that could guide future implementation efforts, and focused on aligning high-level objectives with feasible HPC workflows. In response to the diagram generation issues, we created a textual outline of the workflow with ChatGPT's help and used it as a guide to manually construct the final flowchart using external tools. This ensured visual consistency while retaining the semantic structure generated by the model.

3.1.5 Future Directions

- **Next Steps:** Our next goal is to move from the concept paper to a working prototype by developing a small-scale foundation model using existing simulation datasets. This initial effort will lay the groundwork for a future publication related to the topic: "Towards a Foundation Model for Fusion."
- **Potential Extensions:** There are several potential extensions that focus on enhancing our pipeline capabilities. One key area for exploration is the development of real-time visualization and monitoring systems to improve our ability to track and manage data flows more efficiently. Additionally, we see significant promise in AI-driven adaptive workflow optimization, which would enable automatic adjustments to data processing based on real-time HPC performance metrics. We also envision implementing AI-driven topology optimization to better align with our target objective function, alongside integrating neutronics into our framework.

3.2 Compatibility of Structural Materials and Molten Salt Blanket in Fusion Reactors

Team Members

Rinkle Juneja, junejar@ornl.gov
Arpan Sircar, sircara@ornl.gov
Luke Gibson, gibsonld@ornl.gov

Our multidisciplinary team, comprising experts in fusion materials, machine learning, computational modeling, and blanket engineering, brainstormed together to bridge two critical research areas: high-performance fusion structural materials and magnetohydrodynamic (MHD) behavior of molten salt blankets. Through iterative refinements, ChatGPT was able to outline the key drivers for coupling these domains and it mapped out a streamlined research roadmap (as illustrated in Figure 3.)that addresses material–blanket interface challenges.

3.2.1 Problem Statement

Description: It's crucial to simultaneously select a robust structural material [6] and a blanket material [7] suitable for fusion reactor because the isolated selection does not guarantee their mutual compatibility under fusion-relevant conditions. One of the major metrics for compatibility investigated here is electro-magnetic compatibility since large jumps in electrical conductivity in molten salt blankets are difficult to handle numerically [8].

Motivation: Accelerating fusion pilot-plant deployment requires materials that can withstand extreme heat fluxes, high neutron irradiation, particle fluxes, and MHD-induced stresses. By systematically evaluating the interface between candidate structural materials and molten salt blanket concepts, we can de-risk component qualification, integration and shorten the path to commercialization.

3.2.2 Methodology

- **AI Models & Tools:** We used ChatGPT-4.5 and o4-mini to iteratively refine our problem statement. ChatGPT was able to list out state-of-the-art structural alloys being explored for fusion reactors, and outline potential coupling strategies with the molten salt blankets. We also used Anthropic's Claude v3.5 Sonnet v2 to explore viable candidates for structural materials in parallel with ChatGPT.
- **Workflow:**
  1. Establish physical conditions that structural materials must withstand in fusion reactor
  2. Identify current state-of-the-art materials and list their advantages and disadvantages
  3. Based on deficiencies in state-of-the-art materials, screen candidate alloys that satisfy the constraints in performance in extreme environments
  4. Generate road map for synthesizing and testing proposed materials
  5. Generate draft of white paper based on proposed road map
- **Team Roles:** Our team members had expertise in fusion materials, machine learning, fusion blanket modeling and MHD modeling. This allowed us to work together on addressing the interface compatibility of fusion structural materials and FLiBe blankets.
  - *Juneja:* Investigated the search for fusion structural materials which can be compatible with the molten salt blankets ChatGPT-4.5 refined the search to design a workflow addressing the material search satistying most of the specified constraints using ChatGPT 04-mini.
  - *Sircar:* Used ChatGPT-4.5 to research how to handle the large change in electrical conductivity of molten salts and structural materials at solid-liquid interfaces in fusion blankets.
  - *Gibson:* Used Claude v3.5 Sonnet v2 from Anthropic to first identify the existing state-of-the-art materials proposed for structural materials and then explored promising candidate materials for comparison with output from ChatGPT

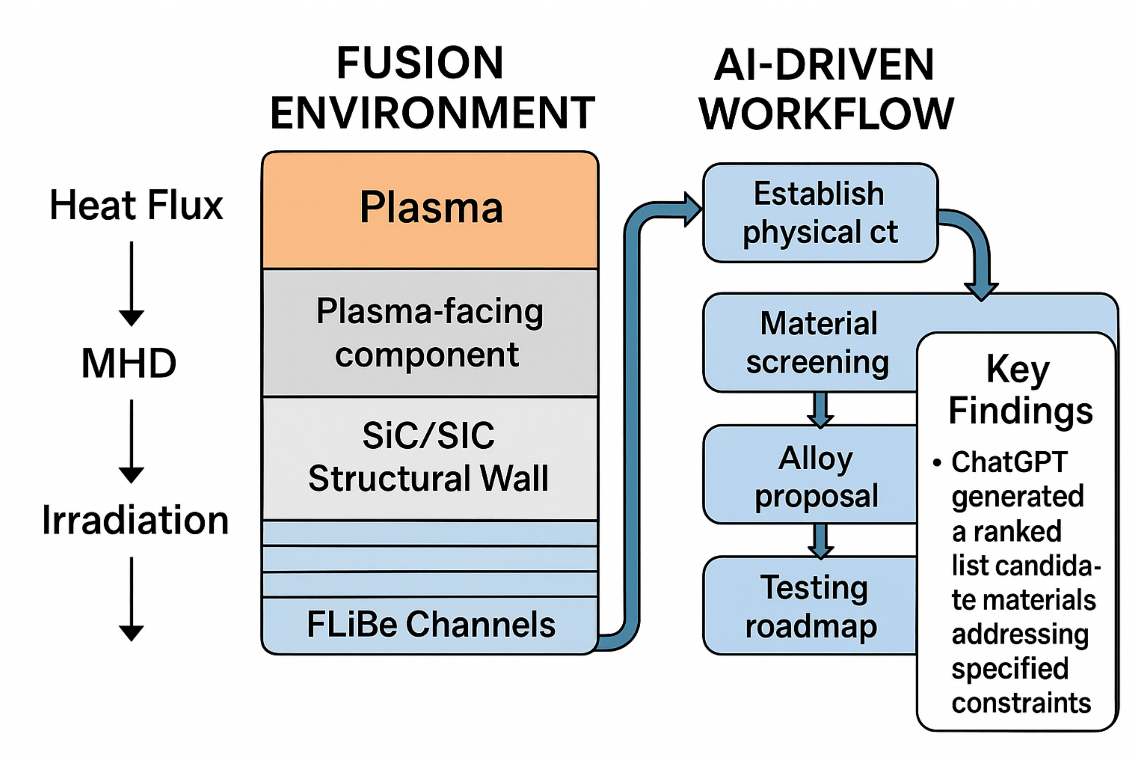


Figure 3: AI-generated modeling workflow highlighting the specified constraints on building a resilient material systems of structural materials and blankets and providing the AI-driven workflow.

3.2.3 Key Findings

- **Results:** ChatGPT generated a ranked list of five candidate structural materials (e.g. Tungsten (pure and Wf/W), RAFM steels, ODS-RAFM, SiC/SiC, and High-Entropy alloy V-4Cr-4Ti), with a summary how each of these can tackle specified four stressors such as heat-flux, neutron irradiation, particle erosion, and MHD disruptions.
- **Coding strategies:** ChatGPT was able to come up with some strategies to tackle the large electrical conductivity jump at interfaces. This included - using a monolithic solver for the solid-liquid coupling, trying out the thin-wall approximation, and fine-tuning the linear solvers and preconditioners for the specific problem [9].
- **Insights:** ChatGPT acknowledged that among the known candidates, no material system yet give performance tackling all the specified sources of damage. However, it ranked these systems based on what should be the best candidates if any of the specified constraints are our top constraints. We also compared the search with Google's Gemini platform. It did not list state-of-the-art materials but ChatGPT was able to do so.

3.2.4 Limitations & Challenges

- **Technical Constraints:** ChatGPT reliably summarized known materials [6, 10, 11, 12] but could not propose truly novel alloys; it did, however, highlight SiC for its extensive validation data [13].
- **Process Hurdles:** ChatGPT was not able to provide high quality summary figures when we asked to generate a visual of high level figure for the proposed research summary. Also, ChatGPT was not able to exactly identify the issue in the simulations for MHD in solid-liquid domains.
- **Mitigation Tried:** We tried different reasoning models to go beyond state-of-the-art materials and it suggested "best-of-all-worlds" systems. We experimented with different prompt strategies, but manual figure curation remains necessary.

3.2.5 Future Directions

- **Next Steps:** ChatGPT suggested systems beyond single materials such as stacking a layer of functionally-graded SiC / Refractory Low-Activation HEA / ODS-RAFM laminate (SiC-HEA-ODS), which might be worth exploring.
- **Potential Extensions:** Discussing the ideas with experimental colleagues.

3.3 Isotopically Tailored Materials for Improvements to Fusion Reactor Design

Team Members

Franklin Curtis, curtisfg@ornl.gov
Tim Younkin, younkintr@ornl.gov
Qian Gong, gongq@ornl.gov
Ahmed Almeldein, almeldeinam@ornl.gov
Ehab Hassan, hassanem@ornl.gov

A multidisciplinary team consisting of experts from fusion, isotope separation, computational modeling, and computer science came together to evaluate the use of AI models to analyze materials within the fusion reactor that would benefit from isotopically tailored materials to improve design and efficiency [14]. The team used the AI tools to explore and generate ideas for materials that are affected by things like neutron bombardment that can benefit from isotopes with different cross-sections to aid with embrittlement, radioactive decay, and other issues related to neutron capture.

3.3.1 Problem Statement

Description: Developing sustainable fusion energy requires addressing critical challenges in the design and operation of plasma-facing components (PFCs) [15], such as divertors, which endure extreme heat and particle fluxes. These components must tolerate heat loads exceeding 10 MW/m^2 while minimizing material erosion and impurity generation that can compromise plasma confinement. In addition to thermal stress, PFCs are exposed to intense 14 MeV neutron flux, leading to the activation of structural materials and the production of radioactive isotopes. To mitigate long-term waste and maintenance challenges, there is growing interest in isotopically tailored materials—engineered to reduce activation through selective nuclear cross-sections—offering potential benefits in safety, sustainability, and lifecycle cost.

Addressing these multiphysics challenges demands the integration of high-fidelity HPC simulation codes with machine learning (ML) models [16, 17, 18]. Surrogate models, trained using tools like NVIDIA Modulus, can replicate complex physics such as turbulence or material response at significantly lower computational cost. Coupling these AI models with fusion codes—either loosely as pre/post-processing tools or tightly within solver kernels—enables scalable, real-time prediction and design-space exploration. This code coupling strategy is essential to unlocking the full potential of exascale computing and accelerating innovation in fusion reactor design through data-driven, multidisciplinary workflows.

Motivation: Achieving practical fusion energy requires addressing critical challenges in the design of plasma-facing components (PFCs), such as divertors, which must withstand extreme heat and neutron fluxes. Traditional workflows for exploring these challenges are time-consuming and fragmented, often requiring deep domain expertise across materials science, computational physics, and reactor engineering. Accelerating scientific discovery in this space demands tools that can support rapid information synthesis, hypothesis generation, and design iteration.

AI-driven language models like ChatGPT offer new opportunities to streamline early-stage research and cross-disciplinary exploration. By enabling rapid comparison of simulation tools, clarification of physical phenomena, and generation of code integration strategies, ChatGPT can assist scientists in evaluating design trade-offs and guiding research directions more efficiently. This is particularly valuable when exploring novel concepts, such as surrogate model integration into HPC workflows or adaptive sampling methods for physics-informed neural networks (PINNs).

A key topic of interest is the use of isotopically tailored materials—elements engineered to exploit differences in nuclear cross-sections to reduce long-term radioactive activation. These materials can lower the cost and complexity of managing activated waste when reactor components are replaced. In fusion environments, tailoring isotopes in structural and shielding materials may provide a means of improving safety and sustainability. Similar approaches have been explored in fission systems, such as the use of gadolinium-157 [19] as a burnable poison due to its favorable neutron absorption properties [20].

3.3.2 Methodology

- **AI Models & Tools:** We used ChatGPT iteratively during technical brainstorming meetings. The interaction involved: (1) Inputting domain-specific prompts extracted from team questions (e.g., thermal limits on tungsten coatings, neutron activation); (2) Requesting comparative assessments of computational platforms (e.g., Vertex vs. COMSOL, Trilinos); (3) Exploring integration approaches for AI surrogates with high-performance simulation codes. The majority of the team did not have much experience using any of the AI tools so utilized the approaches suggested by the workshop organizers in an attempt to extract the best information from the provided platforms.
- **Workflow:** The team began with a general search of what materials would have the largest impact on the fusion reactor for isotopic tailoring. Taking the approach of a naive researcher, little information was provided to the algorithm in an attempt to avoid providing bias to the results. Once this was established, the team began to dive deeper into the results and asked more probing questions with requests for computer codes and figure generation (4).
- **Team Roles:**
  - *Qian Gong*: Used ChatGPT-4o to explore the unique features and performance of VERTEX compared against other state-of-the-art CFD simulation codes. Used ChatGPT-4.5 to brainstorm code-coupling strategies allowing for integrating ML surrogate models with VERTEX code architecture and study examples.
  - *Ehab Hassan*: Used ChatGPT-4o to explore the physical and mechanical properties of materials and designs that extend the sustainability of the divertor under extreme heat and particle fluxes, and develop a scaling-law for erosion rate of several materials and energy confinement time.

3.3.3 Key Findings

- **Results:** ChatGPT significantly enhanced our ability to explore and iterate on complex scientific problems related to fusion reactor design. It facilitated rapid understanding of domain-specific topics such as heat flux limits on divertor tiles, neutron activation of materials, and the viability of

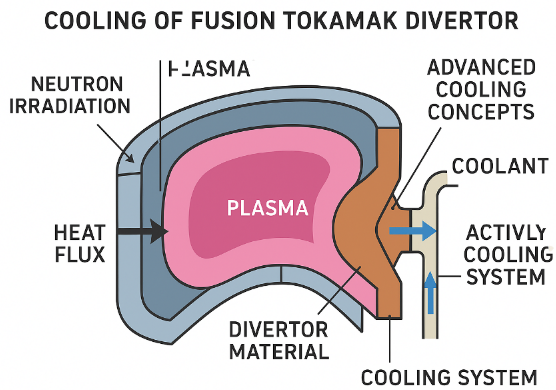


Figure 4: AI generated figure depicting the role of the divertor for a tokamak-style fusion reactor.

isotopically tailored materials. It provided valuable comparions between limits of various materials and geometic designs that can be used in the divertors. Advanced divertor geometries aim to control impurity transport and maximize pumping efficiency:

- **Conventional Divertor**: Minimal flux expansion; high erosion; poor impurity retention.
- **Super-X Divertor (SXD)**: Long leg, extended connection length; excellent heat spreading and impurity screening.
- **Snowflake Divertor (SFD)**: Double X-point improves volumetric radiation and strike-point distribution.
- **X-Point Target (XPT)**: Magnetic mirror geometry reflects impurities and traps neutrals; superior performance in opacity and erosion suppression.

Divertor Type	Impurity Recycling	W Erosion	SOL Control
Conventional	High	High	Poor
SXD	Low	Low	Moderate
SFD	Moderate	Moderate	High
XPT	Lowest	Lowest	Best

Additionally, ChatGPT provides scaling-laws for the rate of erosion of the Tungsten material and their effect of the energy confinement time.
Tungsten erosion rate per unit area (atoms/m^2/s):

$$\dot{E}_W \propto \left(\frac{S}{R^2 A}\right)^{0.5} \left(\frac{I_p}{B_T}\right)^{1.2} \left(\frac{p_{ped}}{\Delta_{ped}}\right)^{0.6} (1-\delta^2)^{0.5} \kappa^{0.3} \left(\frac{1}{\omega_\phi + \epsilon}\right)^{0.4} \left(\frac{1}{t_W}\right)$$

Where:

- S: neutron source rate (n/s)
- R: major radius, A: aspect ratio R/a
- I_p: plasma current, B_T: toroidal magnetic field
- p_{ped}: pedestal pressure, Δ_{ped}: pedestal width
- δ: triangularity, κ: elongation
- ω_ϕ: toroidal rotation, t_W: tungsten thickness

$$\tau_E = \tau_E^{\mathrm{ITER98}} \cdot \left(1 + \alpha \cdot \dot{E}_W^{\mathrm{norm}}\right)^{-1} \cdot \left(1 + \beta \cdot \frac{\lambda_{\mathrm{mfp}}}{\Delta_{\mathrm{SOL}}}\right)^{-1}$$

Where:

- τ_E^{ITER98}: baseline confinement time

  - $\dot{E}_W^{\text{norm}}$: normalized erosion rate
  - α, β: empirical degradation constants

  ChatGPT also provided valuable summaries and comparisons of existing high-performance computing frameworks (e.g., Vertex, MOOSE, WarpX), enabling informed software selection for various simulation tasks. In addition to technical synthesis, ChatGPT helped deconstruct and explore strategies for coupling AI surrogate models—trained using tools like NVIDIA Modulus—with traditional fusion simulation codes. This included guidance on loose and tight integration schemes, performance optimization techniques, and deployment pathways on exascale platforms. As a result, ChatGPT acted as both a knowledge accelerator and a brainstorming assistant, helping us bridge the gap between computational capabilities and scientific objectives.

- **Insights:** We found ChatGPT to be most effective during early-stage exploration, problem scoping, and technical ideation. Its utility was greatest when framing new research directions, breaking down interdisciplinary questions, or synthesizing disparate concepts across physics, materials, and computational science. For instance, concise explanations of reaction-diffusion modeling, surrogate model deployment, and SDF-based adaptive sampling accelerated onboarding and team discussions. However, ChatGPT was not used as a replacement for technical validation or deep-domain authority. Instead, it served as a first-pass assistant to clarify ideas, generate code stubs, identify missing components, or spark alternative hypotheses. Crafting well-scoped prompts and iteratively refining questions yielded the most valuable responses. As such, ChatGPT proved particularly helpful in augmenting the ideation and conceptual design phases of scientific workflows, while downstream modeling, benchmarking, and analysis remained in the hands of domain experts.

3.3.4 Limitations & Challenges

- **Technical Constraints:** Responses require expert validation and sometimes follow-up searches for peer-reviewed confirmation. For example, the team asked what would be required to pull the massive heat out of the plasma and the AI correctly identified the heat transfer equation as:

$$\dot{Q} = hA\,(T - T_f) \tag{1}$$

  where $\dot{Q}$ is the heat flux, h is the heat transfer coefficient of the cooling fluid, A is the area of the surface, T is the surface temperature, and T_f is the fluid temperature. For the heat flux coefficient, the algorithm suggested a cooling fluid velocity of 50,000 m/s, which is unreasonable.

  Although some of the values were unreasonable, the information gathered for relevant materials was used to develop a 'white paper' to begin a first start for funding opportunities. While the algorithm provided this paper, it was not able to provide relevant citations and instead supplied made up citations.

- **Process Hurdles:** Collaboration bottlenecks, tooling gaps, sensitivity issues.

- **Mitigations Tried:** Subject matter expertise was used to weed out irrelevant physics and guide the AI to provide better responses. For the unfamiliar, this could produce results that are misleading.

3.3.5 Future Directions

- **Next Steps:**

  - Develop a fine-tuned version of ChatGPT or RAG (retrieval-augmented generation) system using internal fusion simulation documents and experimental data. The team could use the white paper as a starting point to look at relevant materials for isotopically enriched materials for funding opportunities.
  -

- **Potential Extensions:** New methods, datasets, or interdisciplinary angles to explore.

3.4 Tracing Energetic Particle Pathways in the Lifetime of Alpha Particle Confinement in Fusion Power Plant Reactors

Team Members

Sebastian De Pascuale, depascuales@ornl.gov
Yashika Ghai, ghaiy@ornl.gov
Mark Cianciosa, cianciosamr@ornl.gov
Xiao-Ying Yu, yuxiaoying@ornl.gov

Our team considered the unique challenges that future fusion reactors will face in the burning plasma regime [21] compared to today's experimental devices. Primarily, operating in a fusion scenario will produce from each D-T reaction one neutron of 14.1 MeV energy and one helium ion (alpha particle) of 3.5 MeV energy. This means that approximately 20% of the energy produced by fusion reactions will be lost by magnetic confinement within the plasma or along plasma facing components from diffusive transport and other loss mechanisms. This power fraction from energetic particles would ideally be recirculated within the core plasma. Helium atoms would thereby deposit their energy back into the plasma to reach self-heating (or burning) conditions necessary to promote fusion. However, alpha particles preferentially transfer their energy to electrons via collisions, and these electrons subsequently transfer a proportion of their energy to thermal ions in a very inefficient process of events on the order 30-35%. We investigated the domain reasoning of ChatGPT 4.5 to identify and infer key quantities and constraints that determine the lifetime of an alpha particle, born in the core of a fusion plasma and expiring on the periphery of the vessel through material interactions or pumped exhaust.

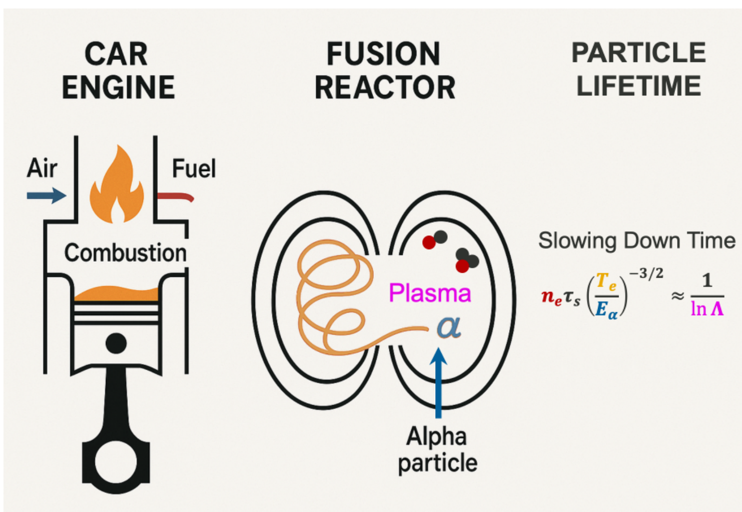


Figure 5: AI generated interpretive diagram showcasing alpha particles (of 3.5 MeV energy, E_α, shown in blue) produced by fusion of D-T fuel (at high densities, n_e, shown in red) as a key ingredient for maintaining sustained fusion reactions through self-heating the plasma [22] (to high temperatures, T_e, shown in yellow) primarily through collisions (indicated by the Coloumb logarithm, $\ln \Lambda$, shown in magenta. The alpha particle slowing down [23] time, τ_s, is an indication of the total confinement of these particles and the triple product, $n_e \tau_e T_e$, is a key metric for obtaining fusion energy producing conditions at high enough density and temperature for long enough time.

The role of energetic particles in fusion reactors is illustrated by an energy cycle, similar to the action of an engine as illustrated in Figure 5, where in the closed system of magnetic confinement the Helium ions act as a principle constituent of the plasma driving energy back into the reactions and in the open system of the plasma interactions with the vessel environment, the Helium ions are removed as "ash" pumped through exhaust pipes or deposited onto the walls of power plants [24, 25]. Since these physical processes encompass multiple physical scales and processes of plasma, particles, materials, and thermodynamics we considered it an excellent challenge problem to test the comprehensive capabilities of ChatGPT from a theoretical level to an experimental level.

3.4.1 Problem Statement

Description: Alpha particles produced by D-T fusion reactions will carry 3.5 MeV of energy and must remain confined within the plasma long enough to transfer energy to elections and ions via collisions, thereby heating the plasma towards self-sustaining steady-state power generation.

Motivation: Helium ions produced by fusion will only remain within the plasma on the order of a typical slowing down time, which can be calculated for a reactor scale device to be of the order of 0.1 to 1 second. This amounts to approximately 200,000 orbits through a toroidal tokamak device before crossing the last closed magnetic flux surface that must carefully balance the degree of energetic particle confinement. Too short of a lifetime within the core plasma and the power plant will suffer from poor heating whereas too long of a lifetime results in fuel dilution of the primary fusion leading to dilution and inefficient energy generation. We considered detailed approaches produced by ChatGPT to optimize alpha particle lifetimes and manage power exhaust for fusion pilot plant reactors.

3.4.2 Methodology

- **AI Models & Tools:** We targeted the OpenAI ChatGPT o4-mini model for deployment solutions as it was the fastest performer in advanced reasoning tasks and, through the new Deep Research feature, this model was able to capture images directly from published reports with accurate and accessible citations.
- **Workflow:** Our team organized its investigation following a hierarchical branching Agile framework, where each member pursued individual research queries according to their own expertise starting from a high level with the goal to produce testable code or calculations. We periodically converged on nodal queries to compare results from different conversational contexts, ensuring that collaboration could synthesize the best results from the reasoning models. Finally, we wrote descriptive assessment of model responses that allowed for changes to our line of inquiry so that more promising directions could be pursued.
- **Team Roles:** Each team member brought important practical knowledge and experience to the use of ChatGPT in this exercise from the plasma theory of energetic particles, research software development, machine learning and artificial intelligence, and material science. Our breadth of expertise allowed us to better consider the cross-disciplinary challenges for power producing fusion reactors in the burning plasma regime.
  - *De Pascuale:* investigated strategies for handling Helium exhaust from alpha particle production, using ChatGPT deep research to assess model extrapolation from present day devices to fusion reactor conditions.
  - *Ghai:* analysed feedback mechanisms to harness improved alpha particle heating, including proposals suggested by ChatGPT for experiments and hardware modifications to test theoretical predictions.
  - *Cianciosa:* generated input/output files from ChatGPT responses for the Integrated Plasma Simulator (IPS) software package and magnetic equilibrium solver to rapidly develop testable modeling cases.
  - *Yu:* explored potential candidate materials and novel compositions to use as plasma facing components in future devices, including expansive key word search with ChatGPT deep research to access literature and database surveys.

3.4.3 Key Findings

- **Results:** ChatGPT o4-mini correctly characterized the 3.5 MeV alpha particle as an important agent in the fusion process. It specified that once the energetic ion is thermalized with the plasma, it effectively becomes Helium ash at ambient ion temperatures and can no longer sufficiently heat the plasma. In this manner, Helium doesn't participate in fusion reactors and can reduce the overall reaction rate leading to decreased fusion power output. Additionally, a population of thermal Helium can contribute to bremsstrahlung and line radiation which increases the net energy loss from the plasma.
- **Insights:** Initial conversations with ChatGPT led to two proposed mechanisms to manage energetic particle lifetimes in a fusion reactor. One concept is the process of **alpha particle channeling**, where a radio-frequency antenna generates plasma waves that can extract energy from fusion-born

alpha particles *before* they thermalize with electrons. This allows for a more efficient pathway of energy transfer directly to the bulk ions which sustain the plasma temperature through preferential diffusion. It was suggested that Ion Cyclotron Range of Frequencies, Lower Hybrid Waves, or mode conversion from Alfvén to Ion Bernstein waves could facilitate in this energy transfer. The other mechanism identified was **helium ash removal** through divertor design. Carefully shaping the magnetic configuration allows or increased pumping pressures and management of heatflux, necessary to mitigate excess power loss conditions along the scrapeoff layer. Turbulent diffusion and convection tend to carry thermalized helium ions away from the core towards the plasma edge, a phenomenon termed impurity screening. Additionally, in high performance plasmas transient magnetohydrodynamic events called edge-localized modes can periodically purge particles from the plasma edge down towards the divertor. ChatGPT produced a comparison of different optimized geometries that could better handle Helium exhaust, including enhanced selective pumping mechanisms and active exhaust control.

3.4.4 Limitations & Challenges

- **Technical Constraints:** As there are a variety of experimental tokamak research facilities, the model had a difficult time synthesizing the results across devices into one proposed fusion pilot plant design. We decided to use the international ITER experiment to base our findings on and familiarity with the literature to assess model performance.
- **Process Hurdles:** We noticed that different conversation iterations could lead to several dead-ends, where the comment history adversely affected ChatGPT responses. In these cases the model was incorrectly linking different factual pieces of information together in order to produce a reasonable seaming answer, such as the name of a researcher who published a particular citation or the name of a code and its particular use cases.
- **Mitigations Tried:** We decided to start new conversations from scratch in order to proceed with our investigation or to test the same query in a different reasoning model in order to cross-validate the responses.

3.4.5 Future Directions

- **Next Steps:** Interestingly, the ChatGPT o4-mini model noted that burning plasma regimes aren't accessible in today's magnetic confinement fusion devices and instead proposed a series of experiments in order to test energetic particle physics. Some of these responses were based on previously published literature, such as the use of neutral beams to preferentially heat the plasma, however other cases indicated experimental upgrades to facilities in order to access new radio-frequency resonances or improved ion diagnostics in order to better track the particle trajectories of fast ions.
- **Potential Extensions:** The Deep Research feature of ChatGPT o4-mini applied to our investigation indicates that with proper synthesis of not only the published literature from experiments, but also with unused diagnostic or operational data from comparable research facilities a fine-tuned reasoning model could propose optimization of device design in order to address important physics challenges in fusion energy science.

3.5 AI-Assisted Semantic Annotation of Fusion Data via Retrieval-Augmented Generation

Team Members

Ana Gainaru, gaianrua@ornl.gov
Chris Bryan, bryancd@ornl.gov

3.5.1 Problem Statement

Description: Fusion research generates high-frequency, multi-modal measurements and simulation datasets [26]. These datasets often lack consistent metadata or semantic annotations, limiting their reuse, interpretability, and integration across studies [27]. One solution of integration expert knowledge in data-driven model extractions can mitigate these challenges [28], but it introduces a high overhead. We envisioned a trained AI system based on a Retrieval-Augmented Generation (RAG) framework to automate the semantic enrichment of measurement and simulation data that can provide a schema and transformation recommendations for datasets generated with different experiments and simulation codes or versions. Such a capability would allow for fast and reliable evaluation of operation and simulation data for determination of causal relationships and trend evaluation for nuclear power and research reactors across unrelated datasets.

Our **main goal** can be summarized as: Given a raw dataset (e.g., NetCDF, CSV, or binary log) containing fusion measurements or simulation data, we aim to:

- Extract and annotate variables with semantic metadata (units, geometry, definitions, relevance).
- Ground annotations in credible scientific sources (e.g., dataset readme files, publications).
- Provide structured outputs suitable for machine learning, visualization, or scientific interpretation.

Motivation: Most nuclear power and research reactors have generated significant data over their lifetimes. Some of this data is typically used during operation by reactor control systems to adjust and control the reactor operating parameters or indicate to human reactor operators the status of specific systems [29]. Some data is captured but not used in operations; however this data can be mined later for better understanding of the reactor operating conditions at the time. While most of this data generated in the past 30 years is digital, some reactors started operations with analog pen plotters, so a significant amount of reactor data exists in different formats. Similar issues can be seen with simulation data [30]. Despite the vast number of datasets collected over decades from both simulations and experiments, fusion scientists often cite a scarcity of usable data—largely due to inconsistent metadata, undocumented variables, and the difficulty of integrating or interpreting legacy and heterogeneous sources.

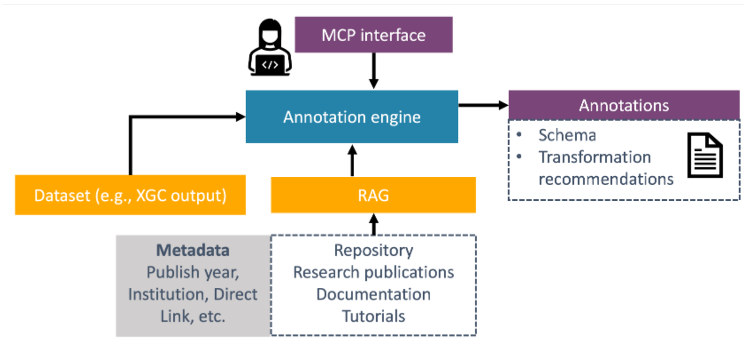
We investigated a system for AI-assisted ingestion and semantic annotation of raw fusion data using Retrieval-Augmented Generation (RAG). Our approach combines data chunking, scientific document indexing, and large language models (LLMs) to enrich raw data with contextually relevant metadata derived from associated publications, technical documentation, repositories and analysis codes, and domain-specific ontologies. We investigated how RAG can improve the accessibility, standardization, and utility of fusion datasets.

3.5.2 Methodology

- **AI Models & Tools:** ChatGPT o3 and Gemini were used as the main reasoning models and ChatGPT o4-mini-high for the initial code generation. We used the models to recommend the steps needed to build our framework, an exemplar application that we can test our methodology on and an evaluation plan.
- **Workflow:** Our pipeline includes the following components:
  - *Document Corpus Construction.* For this we used the reasoning models to find ways of collecting and preprocessing scientific documents, including: i) Dataset README files; ii) XML/JSON metadata; iii) Open-access scientific papers (e.g., IPCC, NOAA, PANGAEA); iv) Repositories with analysis codes
  - *Metadata cleaning.* The reasoning models recommanded ways of cleaning each document, chunked by semantic unit (paragraph or section), and enriched with metadata (source, section, tags). E.g. year information for papers can give context to datasets generated around that time.
  - *Embedding and Indexing.* The cleaned chunks are encoded using domain-aware language models (e.g., SciBERT, OpenAI embeddings). We investigated ways of fast semantic retrieval during annotation.
  - *RAG-Based Annotation* For each data record or variable, we use a predefined API to form a query, top-k relevant chunks are retrieved, an LLM is prompted with the data and retrieved context and the model generates structured annotations, including field definitions, units, and scientific relevance.
  - *Output Normalization* The generated annotations are converted into a unified schema (e.g., JSON-LD or CF-convention-compliant metadata) for integration with downstream tools.
- **Team Roles** A.Gainaru was in charge of designing the entire workflow using Gemini and applying it to a commonly used dataset using ChatGPT and C. Bryan investigated ways of creating a metadata dataset with relevant publications and documentation using ChatGPT. K. Dayman has investigated the use of RAGs for the purpose of dataset annotation. All together conceived the research problem and design the workflow for adding semantics to datasets.

3.5.3 Key Findings

- **Results:** The workflow used is presented in Figure 6. We applied the suggested methodology for Ion temperature profile in the TRANSP simulation outputs [31] and the and we plan to generalize the




```json
{
  "command": "annotate_dataset",
  "inputs": {
    "dataset_uri": "file:///mnt/data/transp_output_2020.h5",
    "format": "HDF5",
    "context_mode": "RAG",
    "retrieval_config": {
      "corpus": "fusion_docs_index_v2",
      "top_k": 5
    },
    "schema_guidance": {
      "allow_inference": true,
      "expected_units": ["keV", "m^-3", "W"],
      "domain": "tokamak_core_profiles"
    }
  },
```


Figure 6: Methodology for users to request semantics about a fusion dataset

methodology in the future. Our plan was to be able to read the Ion temperature in the same format for datasets generated with different versions of TRANSP without any manual intervention. Most of the time was spent developing and putting the RAG pipeline was in place. While progress has been made, we were unable to achieve our goal since we could not gather the datasets needed to test different parts of the pipeline. To overcome this, we used the MITIM framework that integrates various fusion modeling tools, including TRANSP [32]. While it doesn't provide the TRANSP code, it offers scripts and documentation for setting up and running simulations using TRANSP within the MITIM environment. The ChatGPT Reasoning model seemed to generate a plausible coding outline for a minimal working RAG pipeline to go through the MITIM documentation and generated datasets.

3.5.4 Limitations & Challenges

- **Technical Constraints:** Finding publicly available TRANSP simulation datasets spanning the past decade can be challenging due to access restrictions and the proprietary nature of some data. We were able to play with generated data in different formats, but we were unable to test the RAG pipeline since there was no documentation for the generated data.

- **Process Hurdles:** Setting up the whole workflow is a bottleneck since not all the team members are able to access the data necessary for testing the hypothesis. In addition, ChatGPT and Gemini tend to give a lot of information that is not necessary useful for the given task, a lot of time was spent in extracting the needed information and double checking what was generated (this is especially true for the code generation for the RAG pipeline. The reasoning models were unable to help us track down datasets that can be used to test our pipeline.

- **Mitigations Tried:** For the coding part, corrections and changes were requested until the code was usable. For the dataset discovery, many interactions were required to find datasets that are generated by codes hosted on GitHub. The TRANSP-IMAS Translator was suggested to understand how to generate TRANSP data without having access to the code.

3.5.5 Future Directions

- **Next Steps:** We plan to try the methodology proposed to other datasets for which we have access. We plan to make experiments using XGC or GENE that are well-documented and provide comprehensive resources to understand the format of data.

- **Potential Extensions:** We plan to investigate MCP (Model Control Protocol)-style interfaces to allow users (or data management and query systems) issue structured requests to models or agents to perform tasks like annotating a dataset.

3.6 AI-assisted Multiphysics Modeling for MPEX

Team Members

Atul Kumar, kumara@ornl.gov
Paul Laiu, laiump@ornl.gov
Jong Youl Choi, choij@ornl.gov

3.6.1 Problem Statement

Description: The Material Plasma Exposure eXperiment (MPEX) [33, 34, 35, 36, 37] is a high-power linear plasma device designed to study plasma–material interactions under reactor-relevant conditions. A key scientific and engineering challenge is modeling the tightly coupled physical processes involved—RF wave propagation, impurity erosion and redeposition, plasma transport, and sheath dynamics—across disparate time and spatial scales.

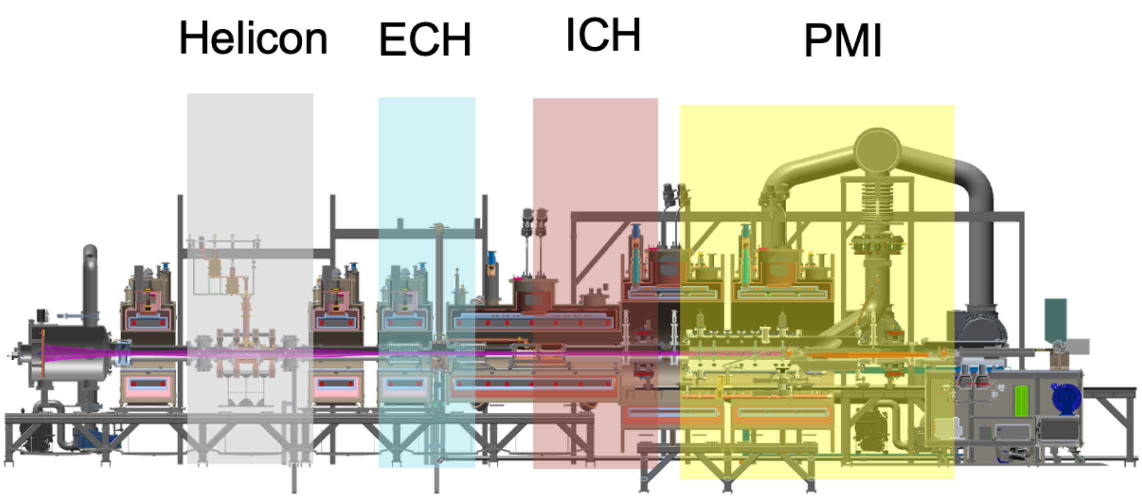


Figure 7: Geometry of the Material Plasma Exposure eXperiment (MPEX) device, highlighting the major operational regions along the axial direction. The plasma is generated and sustained through sequential heating stages: the Helicon region (gray) initiates the plasma using RF helicon waves; the ECH region (blue) applies electron cyclotron heating to increase electron temperature; the ICH region (red) uses ion cyclotron heating to further energize ions; and the PMI region (yellow) is dedicated to plasma-material interaction studies, where high-flux, high-temperature plasma is directed onto material targets for erosion, deposition, and transport analysis.

At the same time, achieving predictive modeling across these regimes requires not only forward modeling but also solving complex **inverse problems** to infer difficult-to-measure quantities (e.g., cross-field diffusion coefficients) from experimental or high-fidelity simulation data. This is particularly important when bridging different physical models (kinetic vs. fluid) and codes (e.g., PICOS++ [36, 35] vs. SOLPS[37]).

Motivation: The multiscale, multiphysics complexity of fusion devices like MPEX demands a robust and integrated modeling approach. Different simulation tools—such as COMSOL [38] for RF wave propagation, SOLPS [37] for fluid plasma transport, and GITR [34, 39, 40, 41] for impurity dynamics—are typically developed in isolation, each tailored to specific physics regimes and dimensional constraints. This siloed structure creates inconsistencies when interpreting experimental data or designing mitigation strategies, especially for coupled phenomena like RF sheath-induced sputtering, impurity transport, and edge plasma behavior.

To address this, we explored the use of AI-assisted workflows to identify missing physics, propose model extensions, and streamline coupling strategies across simulation codes. Large language models (LLMs) like ChatGPT and Gemini were used not only to synthesize domain knowledge but also to help construct proof-of-concept workflows—including forward modeling, parameter scanning, and inverse problem-solving techniques. One such effort focused on using kinetic simulations (e.g., PICOS++) to inform radial diffusion parameters in SOLPS, with AI/ML used as a bridge between high-fidelity physics and reduced models.

Overall, the motivation behind this work is to accelerate the development of predictive, experimentally grounded, and physically consistent across scales modeling pipelines for MPEX and future fusion platforms—leveraging AI where it enhances physical reasoning, solver construction, and system integration.

3.6.2 Methodology

- **AI Models & Tools:**
  - **ChatGPT (GPT-4o)** was used to:
    1. Diagnose shortcomings in existing RF/plasma/impurity transport models.
    2. Generate conceptual and quantitative mitigation strategies (e.g., Faraday shielding, ECH screening, impurity traps).
    3. Propose multi-code modeling workflows (e.g., COMSOL + AORSA [42] + SOLPS).
    4. Draft physics-informed inverse problem frameworks involving kinetic-to-fluid translation of transport coefficients.
  - **Gemini 2.0 and ChatGPT 4-mini** were used to:
    1. Formulate Python-based inverse solvers.
    2. Implement synthetic tests for inferring Bohm-like diffusion coefficients from simplified 1D plasma profiles.
- **Workflow:**
  - Defined a multiphysics problem structure based on observed discrepancies in Proto-MPEX (e.g., hollow T_e profiles [43], target impurity deposition [39], sheath erosion [38]).
  - Used ChatGPT to analyze literature, synthesize diagnostic observations, and propose model extensions.
  - Constructed an AI-assisted **inverse solver prototype** to derive perpendicular diffusion coefficients from 1D kinetic-like simulations (intended to mimic results from PICOS++).
  - Broke the solver into modular Python components for easier debugging and refinement.
  - Tested LLM-generated code using synthetic datasets to compare inferred vs. target profiles.
- **Team Roles:**
  - *Kumar:* investigated solutions to the overall physics and operational challenges associated with the MPEX experiment, used chatGPT 4o with deep research.
  - *Laiu:* developed an AI-assisted inverse solver prototype to determine accurate diffusion coefficients for radial (cross-field) transport, used Gemini 2.0 and ChatGPT 4-mini.
  - *Choi:* explored AI-assisted pseudo turbulent flow simulator for generating synthetic turbulent data to aid the development of detector for the transition from transient to steady state.

3.6.3 Key Findings

- **Results:**
  - By employing ChatGPT-4o with deep research during the AI for Nuclear Workshop at ORNL, it identified RF sheath-induced sputtering from the AlN helicon window as the dominant impurity source in Proto-MPEX. ChatGPT-4o explored relevant literatures [38, 39] on Proto-MPEX and correlated it with experimental deposition data to guide this interpretation.
  - It further suggested that COMSOL cold-plasma models fail to reproduce the experimentally observed hollow T_e profiles, primarily due to missing physics such as kinetic damping (e.g., Trivelpiece–Gould absorption) and RF sheath losses. ChatGPT-4o contributed by interpreting discrepancies between simulation and experiment, identifying key mechanisms omitted in the models, and referencing relevant theoretical and experimental studies.
  - It proposed use of **AORSA** to resolve off-axis power deposition more accurately, with coupling to SOLPS for transport response.
  - Recommended the following hardware and operational mitigations:
    1. Slotted tungsten **Faraday shield** to reduce sheath-induced PMI while maintaining RF coupling.
    2. **ECH-based impurity screening** to suppress impurity flux to the target by increasing T_e.

3. Magnetic geometry modifications and impurity traps near the source to intercept impurities early.

– With the help of Gemini 2.0 and ChatGPT 4-mini, developed a simplified inverse solver using LLMs to infer Bohm-like diffusion coefficients from 1D plasma profiles (density, temperature, magnetic field). The solver used gradient-free optimization and was validated on synthetic test data:

```
import numpy as np
import scipy.sparse as sparse
import scipy.sparse.linalg
import matplotlib.pyplot as plt
from scipy.optimize import minimize

#===================== 1. Diffusion Coefficient Model =====================
def parameterized_diffusion(r, T, n, B, params):…

#===================== 2. 1D Transport Solver =====================
def solve_bohm_diffusion_1d(r, D, S, bc_left, bc_right, dt, num_time_steps):…

#===================== 3. Objective Function =====================
def objective_function(params_values, r, T, B, S, bc_left, bc_right, dt, num_time_steps, n_target,  params_keys):…
```

Figure 8: LLM-generated code skeleton for computing parameterized diffusion coefficients from plasma profiles.

```
# 1. Define the Grid
num_points = 100
r = np.linspace(0.1, 1, num_points) #Radial grid

# 2. "Experimental" Profiles (replace with your data!) - SMOOTHER PROFILES
# Example 1: Parabolic profiles
T_peak = 800  # eV (peak temperature)
n_peak = 1e18 # m^-3 (peak density)
B_0 = 3       # Tesla (magnetic field at r=0)

T = T_peak * (1 - (r/r[-1])**2)  #Parabola
n_target = n_peak * (1 - (r/r[-1])**2) + 1.0e17
B = B_0 / (1 + r) #Decreasing profile

#Ensure data quality to not mess up the code.
T = np.where(T < 0, 0, T)
n_target = np.where(n_target < 0, 0, n_target)

# 3. Define Source
S = np.ones(num_points) * 1e17

#Boundary Conditions
bc_left = n_target[0]   # Density at r[0]
bc_right = n_target[-1]  # Density at r[-1]

#Simulation Parameters
dt = 1e-6
num_time_steps = 1000

# 4. Define Initial Guesses and Bounds for Parameters
initial_params = {…

bounds = {…

# 5. Optimization
params_keys = list(initial_params.keys()) #Make keys to a list
initial_values = [initial_params[key] for key in params_keys]  # Starting point

#Organize the bounds based on the list for minimize function
bounds_list = [bounds[key] for key in params_keys]

result = minimize(objective_function, initial_values,
                  args=(r, T, B, S, bc_left, bc_right, dt, num_time_steps, n_target, params_keys),
                  method='Nelder-Mead', # No bounds needed for Nelder-Mead
                  options={'maxiter': 1000, 'disp': True})  # More iterations, show progress
```

Figure 9: LLM-generated proof-of-concept script for testing the solver on synthetic plasma data.

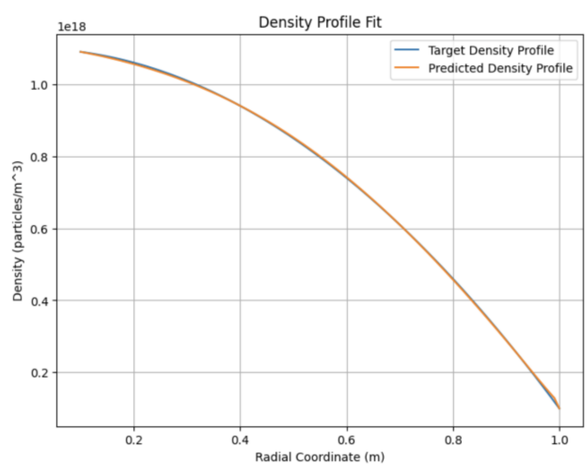


Figure 10: Comparison between target and simulated density profiles using LLM-inferred diffusion coefficients.

– The toy model served as a stepping stone to a more realistic scenario: using PICOS++ to model parallel transport and using AI/ML to infer perpendicular diffusion coefficients for SOLPS. This would create a physics-informed pathway for radial transport calibration in MPEX-like regimes.

- **Insights:**

  – AI-assisted exploration confirmed the hypothesis that spatial RF sheath voltages (modeled via COMSOL) correlate with erosion zones on the helicon window—validating sheath-induced PMI as a dominant erosion mechanism.

– Comparative reasoning with PISCES-RF showed that impurity redeposition can be enhanced via magnetic or flow control, guiding MPEX design strategies.
– Higher azimuthal mode numbers ($m > 1$) must be included in RF simulations to explain observed off-axis power deposition and hollow heating.
– PICOS++ was identified as a viable kinetic tool for modeling flux-surface-resolved parallel transport. When combined with AI/ML-based inverse optimization, it provides a practical approach to infer radial diffusion coefficients in SOLPS.
– Faraday shielding and wall conditioning strategies emerged as essential tools to reduce source impurity loads. AI-synthesized design trade-offs between AlN, BeO, and pBN windows highlighted the importance of evaluating thermal, RF, and PMI resistance properties in tandem.
– In building the inverse solver, we learned that LLM-generated solutions work best when decomposed into small, modular functions, and that derivative-free optimization (though inefficient) was most reliable given the current limits of LLMs in generating adjoint-based solvers.

3.6.4 Limitations & Challenges

- **Technical Constraints:** High-fidelity kinetic solvers like AORSA and PICOS++ are resource-intensive. LLMs struggled to generate robust adjoint-based solvers due to limited public code examples and gaps in memory/context.
- **Process Hurdles:** Cross-code parameter mapping remains a bottleneck. Code generation from LLMs is sensitive to prompt structure and benefits from step-by-step modularization. Optimization techniques used in inverse modeling were simple but not scalable.
- **Mitigations Tried:** Used surrogate sheath models in COMSOL. Reverted to gradient-free optimization methods for inverse solver prototyping. Introduced unit testing and modular code generation strategies to improve LLM usability.

3.6.5 Future Directions

- **Next Steps:**
  1. Integrate COMSOL, AORSA, SOLPS, and GITR in a unified framework with physics-informed feedback loops.
  2. Scale the inverse modeling approach by incorporating realistic PICOS++ outputs for full-core MPEX scenarios.
  3. Test and deploy Faraday shields and impurity traps in Proto-MPEX to evaluate operational viability.
  4. Enhance LLM usage by combining with symbolic computing and physics-informed neural networks (PINNs) for faster inversion.
- **Potential Extensions:**
  – Develop surrogate ML models for expensive solver stages (e.g., sheath modeling, impurity redepositions).
  – Deploy LLMs as real-time assistants during experiments for scenario adjustment and model-informed decision-making.
  – Expand AI-generated solver frameworks to support transport coefficient inference in other edge/scrape-off layer codes (e.g., UEDGE, BOUT++).

3.7 Surrogate Models for Monte Carlo N-Particle Transport

Team Members

Rick Archibald, archibaldrk@ornl.gov
Cory Hauck, hauckc@ornl.gov
Jono McConnell, mcconnelljh@ornl.gov
Anant Raj, raja@ornl.gov

3.7.1 Problem Statement

Description: MCNP (Monte Carlo N-Particle) simulations are used to solve complex physical problems involving the transport of neutrons, photons, electrons, and other particles through matter. These simulations aim to model and predict nuclear processes and radiation interactions with materials in a wide range of applications. Our problem statement was building real-time surrogate models to replicate the accuracy of MCNP simulations (particularly for advanced or transient reactor analysis) using machine learning, with an emphasis on:

- Coupling to thermal-hydraulic (TH) models
- Leveraging real or benchmarked reactor physics data
- Training models that can approximate quantities like keff and pin-level power.

Here, keff (k-effective) is a dimensionless parameter that indicates the neutron multiplication factor in a nuclear system. The system is critical for keff = 1, subcritical when keff < 1.0, and supercritical for keff $>$ 1.0. Pin-level refers to the granular spatial resolution in a nuclear reactor core model, specifically at the level of individual fuel pins (fuel rods) within an assembly. Figure 11 is an image generated by ChatGTP to summarize building real-time surrogate models to replicate the accuracy of MCNP simulations (particularly for advanced or transient reactor analysis) using machine learning.

Motivation: Building surrogates for slow but accurate MCNP simulations enhances computational efficiency, reduces simulation times, and allows for real-time analysis in fusion and fission science. These models enable efficient exploration of design spaces through parameter studies and sensitivity analyses, facilitating optimized configurations. They support multi-physics simulations and integration with other codes, leading to comprehensive system evaluations. Surrogates also aid in validation and benchmarking against experimental data, enhancing confidence in results. Additionally, they serve as educational tools for new researchers, simplifying complex concepts. Overall, surrogates are crucial for advancing nuclear technologies by improving simulation efficiency and supporting safer reactor designs

Team Roles: The team work interactively on this problem. Our expertise was split, with have half having a machine learning background and the other half having a nuclear background. We used each others expertise to judge quality of information being generated by AI.

3.7.2 Methodology

We found that questions that were descriptive of what we wanted to solve, including the constraints to problems, were the most effective in getting compelling answers from ChatGPT -03. Using general questions provided broad general answers, missing the challenges and limitations, and making it hard to later focus on specific aspects of the problem. In general, we found asking a long specific question provided better results over telescoping. For example two questions that gave a lot of traction were:

1. 'Provide an overview of the current challenges in getting data for training models for advanced reactor systems. Focus specifically with respect to (i) neutronics, (ii) thermal hydraulics, (iii) fuel response. On the third one, look at different scales from atomistic response, mesoscale microstructure and grain response and engineering scale thermo mechanical response and fission gas release. Connect it to applications like reactor design optimization, materials design etc. Also, explore development of foundation models in these domains. How can we utilize the data and models from existing reactor systems?'
2. 'We would like to build real-time capable surrogate models for the Monte Carlo code Shift for advanced reactors and for transient simulations. Our specific question is how to gather data for training. Have any surrogate AI models been designed using Shift, even in the steady-state case and/or for standard reactors, like pressurized water reactors (PWR) and boiling water reactors (BWR)? If so, how do such surrogates couple to thermal hydraulics models? Is there experimental data available for validation?'

Shift targets neutron and photon transport physics, with a strong emphasis on reactor core analysis, criticality safety, and high-resolution simulations for advanced nuclear systems [44]. The MOOSE (Multiphysics Object-Oriented Simulation Environment) framework is an open-source, high-performance computational platform developed by Idaho National Laboratory (INL) [45]. It is designed to solve complex, coupled multiphysics problems, making it especially powerful for nuclear reactor simulations. MOOSE is used to simulate the behavior of nuclear reactor systems by coupling various physics modules (e.g., heat transfer, neutronics, fuel performance, structural mechanics). It enables the development of advanced, multiphysics simulation tools tailored to reactor design, analysis, and safety.

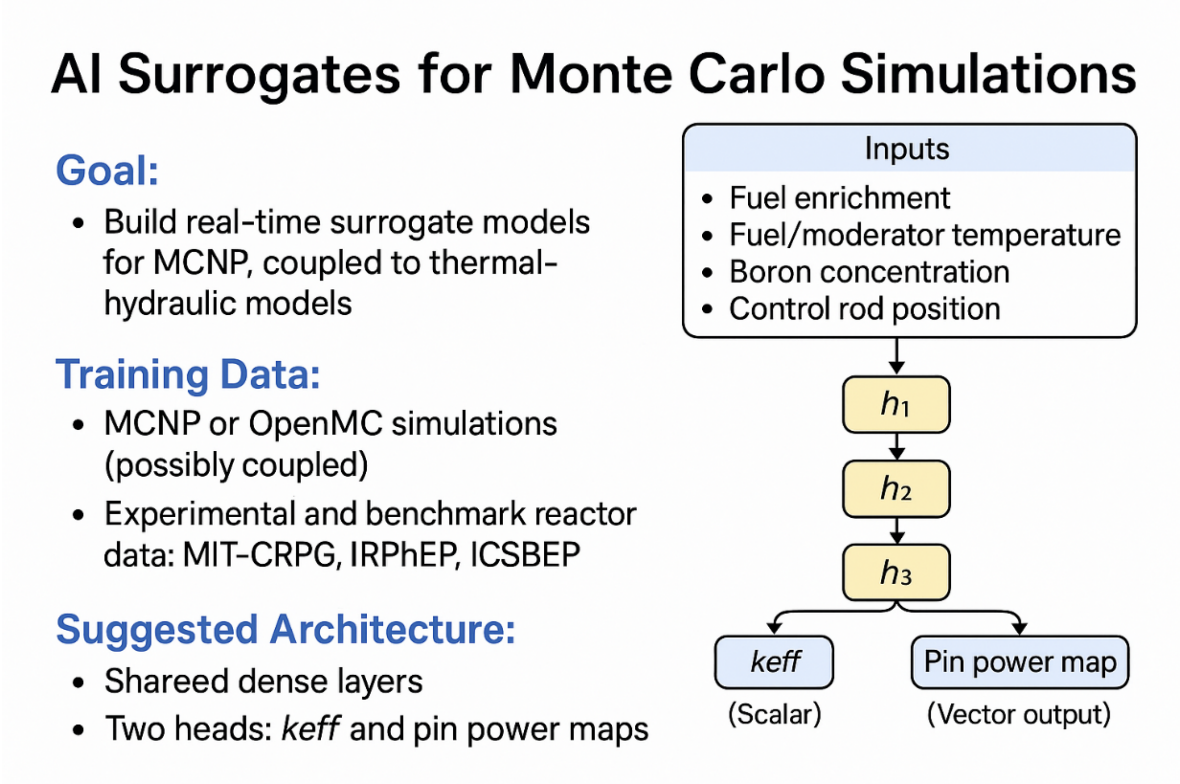


Figure 11: Image generated by ChatGTP to summarize building real-time surrogate models to replicate the accuracy of MCNP simulations (particularly for advanced or transient reactor analysis) using machine learning

For the rest of this section we will focus on the teams experience with the second question. The proposed workflow that ChatGPT-03 offered for the second question was a detailed and reasoned approach and summarized as:

Summary Statement for Surrogate Modeling of Shift

- `No published Shift-based ML surrogate exists yet; you'll likely be breaking new ground.`
- `Data gathering will hinge on a well-designed DoE (possibly multifidelity) using Shift (for neutronics) + Griffin (for low-fidelity coupling) runs.`
- `Coupling uses the same MOOSE MultiApp pattern already exploited by Griffin/Pronghorn: simply substitute your ML surrogate for Griffin in the loop.`
- `Validation should leverage established experimental benchmarks (MSRE, LOFT, BEAVRS/VERA, NEA transients).`

`If you'd like pointers to detailed methodologies for DoE sampling, active learning, or embedding in MOOSE, let me know!`

ChatGPT-03

3.7.3 Key Findings

Overall, using ChatGPT -03 to develop a research plan to create a surrogate model for the ORNL based exascale code Shift [46] was well structured. It provided useful workflow and suggestions on how to build the entire framework. It operated at a general knowledge level for the required domains needed to solve this problem. As described in the limitations and challenges section, there are minimal to no known public databases to train a surrogate model for MCNP. ChatGPT -03 identified this limitation, however was still able to describe the process and write reasonable code that noted if a suitable data-set is generated, how to proceed with all the other steps.

3.7.4 Limitations & Challenges

Developing accurate surrogate models typically requires a substantial amount of high-quality data from full MCNP simulations. Obtaining this data can be resource-intensive and time-consuming. There are currently no sources available and data generation for these types of challenge could be a value add for the DOE community. In the nuclear field, meeting regulatory standards for safety and reliability can add additional layers of scrutiny. While the application of AI in generating surrogates for MCNP code presents exciting possibilities for improving simulation efficiency, it also brings forth significant challenges that must be addressed. Ensuring high-quality data, model accuracy, interpretability, and successful integration into existing workflows are critical for leveraging the full potential of AI in this domain.

There was some unique datasets that were discovered, the MIT Reactor Physics Benchmark Models Repository [47], and alternative codes suggested than the more costly models. However, the main process hurdle was high quality curated data sets ready for AI/ML are hard to come by for nuclear reactors. results. Large Language Models (LLMs) can significantly assist in tagging metadata by leveraging their ability to understand context, semantics, and patterns in natural language and structured data. LLMs can take previously available data—such as documents, logs, images with descriptions, simulation results, or sensor outputs—and automatically generate or enhance metadata (descriptive labels or annotations). This is valuable for improving dataset quality, discoverability, and readiness for ML tasks.

3.7.5 Future Directions

With the strengths of Nuclear science at ORNL and the resources available, creating a high quality surrogate of MCNP is possible. However, the main challenges remain for surrogate modeling of complex physical process. Specifically, the development of accurate surrogate models necessitates a substantial amount of high-quality data, which can be resource-intensive to obtain. Additionally, the complexity of the physical processes involved can make it difficult for AI to accurately capture all nuances, and the resulting models may suffer from issues with generalization and overfitting. Moreover, the training of complex models may demand significant computational resources, potentially undermining efficiency gains. Integration with existing workflows can also be challenging due to compatibility issues and user adoption hurdles. Additionally, the validation and verification of these surrogate models against traditional MCNP results are crucial yet complex, particularly in ensuring regulatory compliance. Finally, quantifying uncertainties within the surrogate predictions and the need for ongoing updates as new materials and trends emerge further complicate their implementation in the dynamic fields of fusion and fission science.

3.8 Investigating Corrosion in Molten Salts

Team Members

Rajni Chahal, rchahal@tntech.edu; chahalr@ornl.gov
Cihangir Celik, celikc@ornl.gov
Flavio Dal Forno Chuahy, dalfornochuf@ornl.gov
Tyrone Harris, harristc@ornl.gov

3.8.1 Problem description

Due to high temperatures and radioactive environment in the nuclear reactors, the effect of foreign species, e.g. salt impurities and metallic particles from salt-alloy/metal interface, on the structure and properties of molten salts is difficult to investigate solely using experimental techniques due to difficulties in precise measurements in extreme environments as well as their underlying costs [48]. Thus, there are gaps where modeling can help: systematically investigating structure-property relations [49], predicting solvation structures of ions/complexes that are hard to isolate [50], predicting solubility limits of corrosion species within reactor temperature ranges [51], computing redox potentials [52] [53] and activity coefficients for multi-component salts [54], and simulating electrode processes (nucleation, growth) that experiments cannot easily resolve.

Motivation: Molten salts are relevant for both fusion (e.g., molten salt blankets) and fission (e.g., coolant and fuel salts) nuclear reactors. In such applications, due to high temperatures and radioactive environment, corrosion in molten salts is accelerated, which in turn can affect the structure and properties of the molten salts. This problem is critical to address for the successful reactor operations and their licensing [55].

3.8.2 Methodology

- **AI Models & Tools:** OpenAI's Deep Research was deployed as the reasoning model to investigate: 1) what has been done before to study the problem of corrosion in molten salts, 2) what modeling and simulations efforts can be helpful, 3) what existing and new data will be needed for model development, and 4) what experimental data can be used for validation of models. Initial prompts for this model were generated using OpenAI's GPT4.5.
- **Workflow:** Problem breakdown into small segments, e.g., existing literature on simulations, existing literature on experiments, the gaps in simulations and experiments to further address the problem of corrosion in molten salts
  - generate prompts for each problem subsection in OpenAI's GPT4.5,
  - feed the prompts to Deep Research,
  - thorough review of each response by domain expert in the team,
  - cross-checking the statements made with the reported references,
  - checking the validity and trustworthiness of reference sources,
  - obtain short summary of each response based on verified content,
  - LaTeX Integration and Formatting.
- **Team:** Rajni Chahal (expertise in molten salts' modeling and simulations, MLIPs), Flavio Dal Forno Chuahy (expertise in computational fluid dynamics), Tyrone Harris (expertise in radiation detection), Cihangir Celik (expertise in modeling and simulation for nuclear criticality safety and radiation shielding)

3.8.3 Key Findings

- **Results:** We obtained a brief review of existing literature reporting previous efforts done to study the structure of molten salts with corrosion solute species using classical molecular dynamics (MD), *ab initio* MD (AIMD), and MD simulations driven by machine learning interatomic potentials (MLIPs). The limitations of each method was determined correctly but state-of-the-art computational studies using each method were not reported. Open gaps and opportunities were identified correctly, which include (1) developing accurate MLIPs potentials for actinide/transition metals, (2) bridging atomistic MD with continuum models, (3) understanding cluster formation (e.g. metallic nanoparticle solubility) in salt. In short, while MD (classical or AIMD) has detailed local structure and transport, it often struggles to fully capture redox chemistry and rare events; improved potentials and multiscale methods are needed. A slightly better review of experimental works was reported listing studies on spectroscopy and scattering, redox potentials, solvation structures of the species, and measured transport properties of molten salts. In this case as well, several important studies were unlisted. No novel hypothesis was reported by Deep Research model.
- **Insights:** No unexpected observations or new domain-specific learnings were reported in Deep Research model's responses.

3.8.4 Limitations & Challenges

- **Technical Constraints:** The model was not able to report the state-of-the-art computational studies in this area. Also, we traced the references and found that the sources were not necessarily reliable. This could be due to the fact that reliable or journal articles in this area are not open-access. The fact that some statements in model's response were pulled out of the conference abstracts (generally available openly and free to cost) also supported our hypothesis that the model may not be trained on *closed-access* journal articles. A better performance of model was observed in the case of reviewing experimental literature for corrosion in molten salts. This could also be due to the problem of data imbalance. For example, less number of computational studies exist in this area compared to the number of experimental studies.
- **Process Hurdles:** A comparison of Deep Research's responses with other available models was not performed. However, response time of the Deep Research was considerably slower than a regular inquiry response time which might be a limiting factor for context-based follow up inquires that researchers tend to come up frequently.

- **Mitigations Tried:** A workaround for using Deep Research effectively would be preparing tailored prompts using regular AI interfaces to ask Deep Research.

3.8.5 Future Directions

- **Next Steps:** 1) New AIMD data needs to be generated to study the corrosion solute species in molten salts for the cases where experimental data is available for model's validation. OLCF computing facility can be utilized for this purpose.
  2) Machine learning interatomic potentials should be developed on AIMD data using uncertainty quantification approach for efficiently sampling the potential energy surface in order to obtain a trustworthy and robust MLIP model. OLCF computing facility can be utilized for their development.
  3) Using MLIP models, the quantities of interest, such as solubility limits, redox potentials, activity coefficients of species in molten salts need to be calculated for the temperature range for reactor operations. The effect of solute species and their subsequent aggregation on molten salt's transport properties needs to be evaluated.
  4) Existing experimental data [56] [57] can be used to validate MLIP model's findings. New experiments need to be proposed through collaboration for further validation of the simulations' predictions.
  5) Investments need to be made in both AIMD datasets (for model's training) and experimental datasets (for model validation).
- **Potential Extensions:** Multiscale modeling approaches on "how to link MD with higher length-scale models or continuum models?" can be investigated to gain a better understanding of corrosion of molten salts when interfaced with alloy/metallic structural components in nuclear reactors. This could include feeding-forward the MD-determined diffusion coefficients and other transport properties to higher-scale models allowing investigation of effect of corrosion in molten salts during reactor's life cycle.

3.9 Predictive Modeling of Nuclear Material Degradation

Team Members

Samantha Sabatino, sabatinosm@ornl.gov
Pradeep Ramuhalli, ramuhallip@ornl.gov
Tom Beck, becktl@ornl.gov

3.9.1 Problem Statement

Description: Characterizing degradation in nuclear structural materials, including reinforced concrete and irradiated metals, poses a complex inverse problem. Internal damage (e.g., microcracking from alkali–silica reaction (ASR) or cracking in weldments) is difficult to detect directly during operation, requiring nondestructive examination and inference from sparse measurement data [58]. Two variations on the science problem were studied: (1) the problem of estimating the modulus loss and crack density in nuclear concrete at 60 years using acoustic and strain measurements; (2) reconstruction of the shape of internal discontinuities (such as internal voids or cracks) from a limited set of acoustic measurements [59].

For the first case, the expected input data included synthetic acoustic waveform features and strain measurements at mid-life (30 years), with targets being modulus loss and crack density at 60 years. For the second case, expected input data was from a linear acoustic scan of the material under test.

In each case, the objective was to use the AI model to accelerate all aspects of solving the problem, from finding or synthesizing data to generating code that may be used to solve the inverse problem. Overall, the inverse problem is often ill-posed and computationally intensive: small changes in observations can correspond to large differences in internal state, and simulations needed to predict material behavior are demanding [60].

Motivation: This challenge is critical in the context of fission and future fusion reactors, where structural integrity over 80+ years must be ensured under extreme conditions. Traditional inspections often miss early-stage degradation, increasing the need for AI-driven methods that can infer internal damage states from indirect acoustic or strain measurements. Reliable inverse modeling enables better aging management and safety assurance in nuclear infrastructure [61].

3.9.2 Methodology

- **AI Models & Tools:** We employed OpenAI's ChatGPT (o4-mini or its variants) to identify methods that have been proposed in the literature, with a focus on machine-learning methods. The LLM was further utilized to determine a reasonable workflow for this problem. The scope of the machine learning methods identified for use within the workflow included linear regression, random forests, gradient-boosted trees, Gaussian process regression, and neural networks (MLP and CNNs).
- **Workflow:**
  - **ChatGPT**
    1. Problem scoping: discussion with ChatGPT to frame the inverse problem and explore solution strategies.
    2. Code generation: ChatGPT produced `Python` and `Mathematica` code to simulate, optimize, and visualize data.
    3. Debugging: errors encountered in execution were discussed with ChatGPT, which provided corrections and suggestions.
    4. Documentation: ChatGPT drafted explanatory content, summaries, LaTeX output, and figure descriptions.
  - **Inverse Problem**

    An overview of the following inverse modeling architecture is shown in Figure 12.
    1. Synthetic data generation: generate data with probabilistic variations in material, environmental, and geometric parameters [62].
    2. Preprocessing: waveform normalization, noise injection, feature extraction.
    3. Hypothesis generation: selecting strain and acoustic signal features as proxies for damage.
    4. Code development: modular `Mathematica` or `Python` framework for training and evaluating AI models.
    5. Evaluation: RMSE, MAE, R^2, and depth-dependent prediction plots.

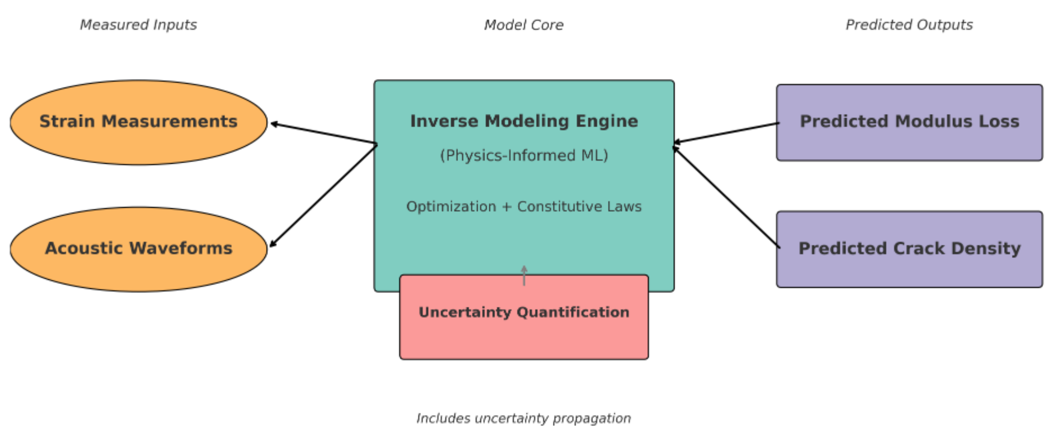


Figure 12: Conceptual diagram of a physics-informed inverse modeling pipeline that maps measured strain and acoustic data to predicted material degradation with uncertainty quantification.

- **Team Roles:** Samantha Sabatino provided workflow integration, interfaced with ChatGPT to develop appropriate code to generate synthetic ASR damage data, and debugged `Mathematica` code that facilitated machine learning. Pradeep Ramuhalli discovered and tested the limitations of ChatGPT, developed and debugged `Python` code, and interfaced with ChatGPT to provide summary text about the domain problem. Tom Beck provided strategic project oversight, as well as data and workflow management.

3.9.3 Key Findings

Key findings include both findings relative to the use of ChatGPT as well as from the technical problem selected for this workshop (inverse analysis).

ChatGPT

- **Results:** ChatGPT provided a reasonable workflow for the problem of interest and was able to provide compilation of information relevant to the problem. Code was generated (`Python` as well as `Mathematica`) for solving several components of the workflows. The `Mathematica` code generated was able to be successfully executed (see **Inverse Problem: Results** below for a summary of the results from this activity. The `Python` routines generated focused on the data generation aspect for the generation of acoustic scan data, and were limited in their ability to execute successfully (missing functions, apparent code generation for incorrect domain areas, incorrect results of code execution).
- **Insights:** ChatGPT appears to be very well suited for information compilation, though the provided information did not seem specific enough but was sufficient for a beginner to understand the problem space and initiate research. While its code generation capabilities seem to be good, greater specificity in queries appears to be needed. It should be noted that the accuracy of data or domain problem results generated using ChatGPT generated code was not evaluated, nor was the accuracy of the provided workflows compared against other, known, solutions that have been proven to work. As a result, at this stage in the evaluation, ChatGPT appears to be a good way to initiate and support research where the researcher needs to conduct due-diligence. Note that other LLM models were initially considered but were not used in this work for a variety of reasons (lack of access, initial inaccurate information generated).

Inverse Problem

- **Results:** All models achieved strong predictive performance ($R^2 \geq 0.90$) for both modulus loss and crack density. Gradient-boosted trees and linear regression consistently performed best on this dataset. To better understand how predictive accuracy varies with sensor depth, we evaluated RMSE and R^2 trends across the concrete biological shield (CBS) thickness for five machine learning models trained on synthetic data generated in `Mathematica`. As shown in Figure 13, prediction error (RMSE) increases with depth, while model fit (R^2) decreases, reflecting higher uncertainty and signal degradation at deeper regions. These results reinforce the importance of data quality and spatial distribution when designing inverse models for degradation prediction. CNN-based classifiers on waveform data outperformed feature-engineered approaches in ASR detection scenarios [63]. The final model framework supports inverse prediction of degradation state using sparse mid-life sensor inputs.
- **Insights:** Waveform-based features (e.g., normalized velocity) were more predictive than summary statistics alone. Even simple models (linear, RF) captured degradation trends well due to strong signal in features. Combining acoustic and strain data improved accuracy. However, model generalization to real reactor conditions requires further validation with field or high-fidelity experimental data [64].

3.9.4 Limitations & Challenges

Again, we include challenges for both the LLMs as well as the domain problem.

ChatGPT

- **Technical Constraints:**
  - Limited memory window resulted in context loss.
  - No internal code execution; all results had to be externally validated.
  - Domain-specific assumptions sometimes incorrect.
  - Limited access to real-world labeled data to provide additional context for ChatGPT.
- **Process Hurdles:**
  - No built-in version control; iterations had to be tracked manually.
  - Prompt history had to be actively managed.

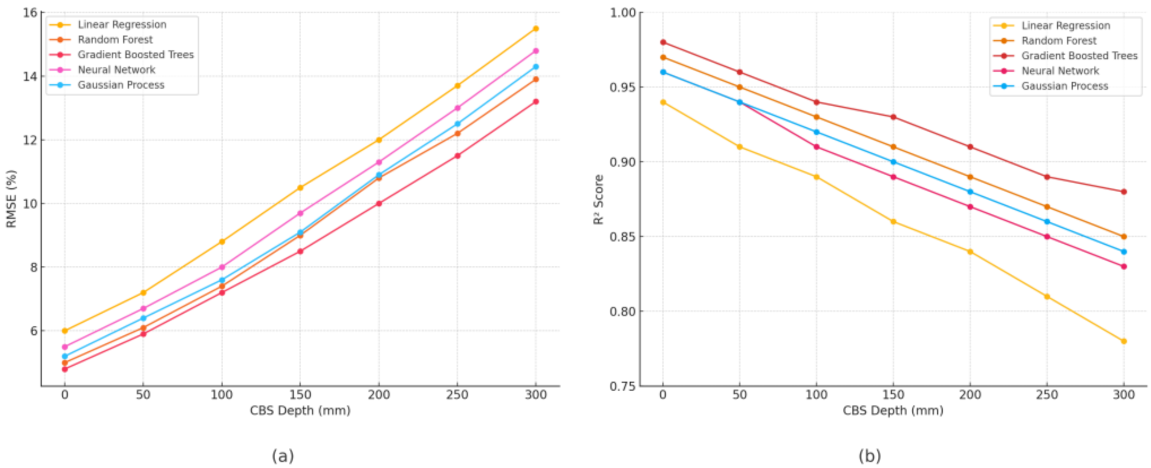


Figure 13: Model performance as a function of CBS depth, showing (a) prediction error (RMSE) and (b) model fit (R^2) across five AI models using synthetic data.

  - Multi-turn prompts required careful planning.

- **Mitigations Tried:**
  - Code was modularized to reduce context load.
  - Periodic summaries and external notes maintained continuity.
  - Prompting strategies were refined iteratively.
- **Mathematical/Computational Challenges:** Occasional delays in responses from ChatGPT (upwards of an hour or two) were encountered though it is unclear if this is related to technical issues with the servers or computational challenges with the model and the size of the queries/generated results.

Inverse Problem

- **Technical Constraints:** There was limited access to real-world labeled data and a heavy reliance on synthetic scenarios. Memory and compute constraints in cloud-based environments limited model complexity and dataset size (∼2,000 samples).
- **Process Hurdles:** Tooling gaps between `Mathematica` and `Python`-based ML frameworks slowed integration. Acoustic signal variability due to environmental noise required careful feature design [65].
- **Mitigations Tried:** We reduced the sample size for tractable training and trained one model at a time using modular loops. Furthermore, simplified feature sets and targeted visualization enabled efficient debugging and validation.
- **Mathematical/computational challenges:** Basic research is needed into how best to incorporate physical constraints in solving inverse problems that utilize limited data to update and improve computational models of stress-strain behavior, fracture, etc [66]. These improved models can in turn be used for prediction of critical failures in the materials within uncertainty bounds.

3.9.5 Future Directions

From the standpoint of the domain problem, next steps and potential extensions are described below. The capabilities and potential limitations of the LLM (ChatGPT) (as encountered during the workshop) are also enabling a review of other potential extensions.

- **Next Steps:** Ideally, we would validate the current models on real ASR-affected structures and irradiated metal data from surveillance programs [67]. Additionally, we would expand the waveform library using laboratory experiments or physics-based signal simulation.

- **Potential Extensions:**
  - Integrate physics-informed neural networks (PINNs) to constrain model predictions.
  - Extend to 3D geometry and time-series forecasting for progressive damage.
  - Apply to fusion materials (e.g., Tungsten, Reduced Activation Ferritic/Martensitic steels) under extreme neutron and thermal conditions.
  - Incorporate uncertainty quantification for risk-informed decision support.

3.10 AI-directed Experimental Program for Sodium Fast Reactors and Molten Salt Reactors

Team Members

Marie Romedenne, romedennem@ornl.gov
Gregory Davidson, davidsongg@ornl.gov
Nathan D. See, seend@ornl.gov

Our team considered the development of an AI framework that would integrate simulation, uncertainty quantification and experimental design. The framework consists of a physics-based simulator for coolant materials, physical and thermodynamic properties, surrogate and physics-informed neural network (PINN) data pipeline, OpenMC loop modeling, AI-directed experiment and gap detection, high-fidelity corrosion simulation and 40-year lifetime transient extension.

3.10.1 Problem Statement

Advanced nuclear reactors, including molten salt and liquid metal-cooled systems, present distinct corrosion challenges owing to their high operational temperatures and the chemically aggressive nature of their coolants (i,e., salts and liquid metals) [68, 69, 70]. Current worldwide modeling tools and data integration methodologies are limited in their experimental coverage and responsiveness [71]. This research plan aims to address these limitations by integrating surrogate models with physics-informed neural networks (PINNS) that account for uncertainties and incorporate an active learning loop. Unlike traditional frameworks that rely on one-way validation and model calibration before development, this approach is designed to be more adaptive and data-efficient, enhancing the understanding and management of corrosion phenomena in advanced nuclear reactor environments.

Description: Our investigation focused on creating a design for an experimental program aimed at generating validation data for a physics-informed neural network (PINN)-based surrogate model, which is trained using a coupled multiphysics simulation capability calibrated through a targeted experimental campaign. The development of this multiphysics simulation capability was guided by an AI algorithm. Additionally, the experimental program was crafted by employing AI to assess and identify existing knowledge gaps in the corrosion data related to molten salts and liquid metals within the current accessible scientific literature.

Motivation: Advanced nuclear fission reactors, including molten salt reactors (MSRs) and sodium fast reactors (SFRs), lack the extensive operational data that traditional light water reactors possess, which limits validation opportunities for newm high-fidelity simulation frameworks. To develop a robust simulation framwork that supports the design, licensing, and operation of advanced MSRs and SFRs, it is essential to identify the gaps in existing experimental data. A comprehensive research program must be established to address these gaps and reduce uncertainties in the simulaton models. The present work focused on experimental data related to corrosion and compatibility issues.

3.10.2 Methodology

- **AI Models & Tools:** Team 11 dedicated its efforts to utilizing various models of ChatGPT, with a preference for ChatGPT 41o due to its advanced reasoning capabilities. In instances where code snippets were required, ChatGPT 4o-mini-high was employed. Importantly, no data was uploaded to ChatGPT other than the prompt queries used throughout the investigation.ChatGPT was used to enhance the writing of this section as well.
- **Workflow:** Our workflow involved leveraging ChatGPT to develop an experimental program through a sequence of prompts. Initially, we requested ChatGPT to outline the steps necessary for designing an experimental program to support an AI-generated MSR surrogate simulator. In response, ChatGPT provided a five-step plan that included the creation of a surrogate model using physics-based methods,

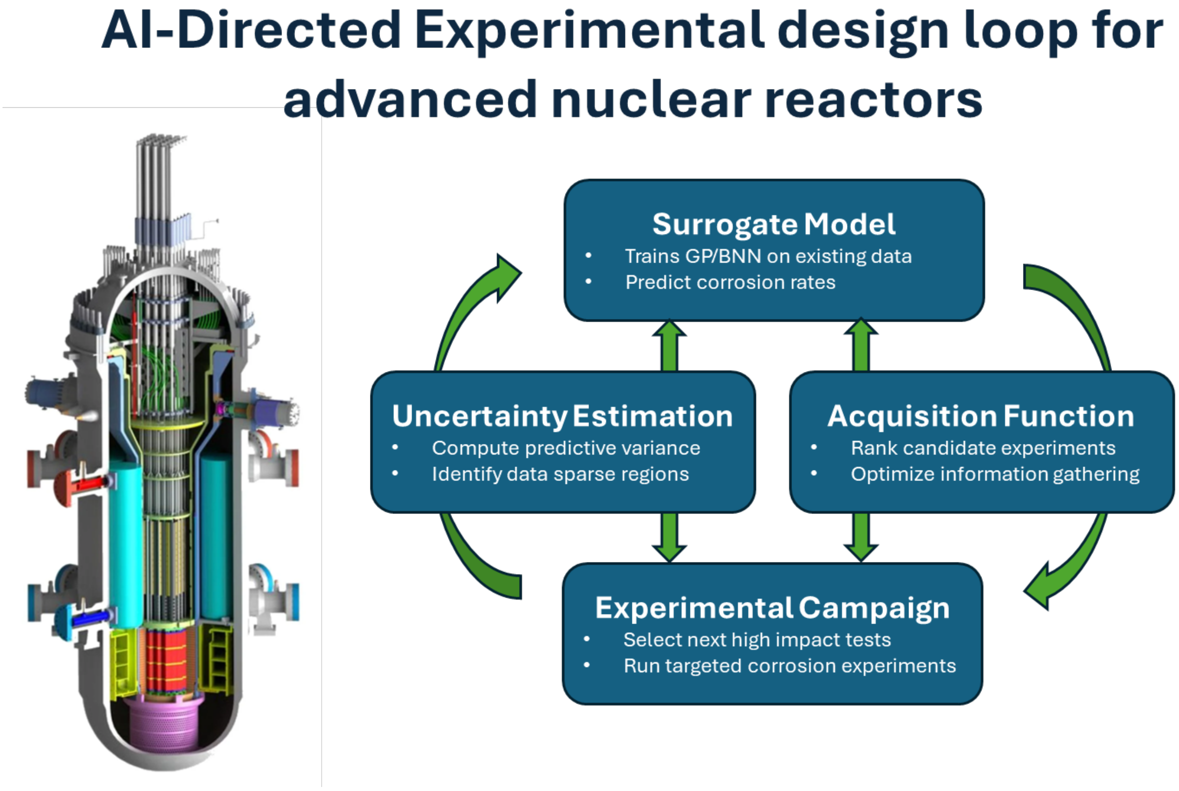


Figure 14: Physics-informed neural network (PINN)-based surrogate model trained using a coupled multi-physics simulation capability calibrated through a targeted experimental campaign

followed by uncertainty quantification, which would inform the experimental design. Next, we prompted ChatGPT to propose an initial coupled multiphysics simulation framework for generating synthetic data to inform the AI-driven surrogate model. ChatGPT provided a plausible Python-based multiphysics simulator that utilized pre-existing solvers for the relevant physics. It accurately identified key components, including neutron transport, thermohydraulics, and corrosion chemistry, and furhter promts led to the inclusion of nonlinear feedback mechanisms among these components using a Gauss-Seidel iteration scheme. To explore neutron transport, we specifically instructed ChatGPT to use OpenMC as an example and requested a Python script for generating OpenMC inputs tailored for an MSR model. ChatGPT successfully responded with a plausible initial script for this purpose. Subsequently, we inquired about chemistry codes for simulating corrosion in an MSR. ChatGPT provided a comprehensive table listing various codes along with their licensing information (commercial, open-source, etc.) key features, and a discussion of the pros and cons associated with each software package. Upon indicating our preference for Thermochimica as the coolant chemistry simulation code, we asked ChatGPT to integrate it to the previously proposed multiphysics framework. This prompted the generation of additional Python code to facilitate the integration. Finally, we requested a Python code from ChatGPT for traning a PINN using the established multiphysics framework, which resulted in futher code inputs. After completing the above steps, we expanded our analysis to include sodium-cooled fast reactors (SFRs) to assess ChatGPT's adaptability to novel problems. It successfully identified the relevant corrosion mechanisms for SFRs and how they differ from those in MSRs.

In conclusion, ChatGPT effectively summarized the entire process into a cohesive research program. With additional prompts, it identified gaps in the current corrosion knowledge base, proposed experiments to address these gaps, and located existing experimental facilities capable of conducting the necessary research.

- **Team Roles:** Gregory acted as query architect and development subject matter expert (SME) while Marie and Nathan acted as scientific SMEs to review and suggest new pathways.

3.10.3 Key Findings

- **Results:**
  ChatGPT was able to propose plausible experiments for filling current experimental gaps in both SFR and MSR corrosion chemistry knowledge. In Figure 15 we show a screenshow ChatGPT produced for an experimental program to fill knowledge gaps in SFR corrosion.

#	Alloy	T (°C)	O_2 (wppm)	Flow (m/s)	Duration	Mode	Primary Uncertainty Addressed
1	316 SS	550	2	5	1 000 h	Steady	Baseline at low-end intermediate O_2
2	316 SS	550	10	5	1 000 h	Steady	Mid-band O_2 kinetics transition
3	316 SS	550	20	5	1 000 h	Steady	Upper-band O_2 kinetics transition
4	316 SS	550	30	5	1 000 h	Steady	High-end "challenge" intermediate band
5	316 SS	350	1	5	1 000 h	Steady	Low-T protective-film behavior below 450 °C
6	316 SS	800	10	5	500 h	Steady	High-T dissolution/spallation near 725–900 °C
7	316 SS	600	5	15	500 h	Steady	Mass-transfer-limited regime at high flow (>12 m/s)
8	316 SS	350⇄600 (cyclic)	5	5	100 cycles	Thermal cycling	Startup/shutdown transient effects
9	316 SS	550	5	5	10 000 h	Steady	Ultra-long exposure & scale coarsening
10	ODS Steel	550	5	5	1 000 h	Steady	Novel alloy response under intermediate O_2

Figure 15: Screenshot from ChatGPT showing a proposed SFR experimental program.

ChatGPT was also able to provide a proposed budget for the entire research program. This is shown in Figure 16.

Component	Description	# Runs / Scope	Estimated Cost (2025 USD)
MSR Corrosion Experiments	Targeted studies varying salt chemistry, T, flow, impurities, radiation, alloys, cycling, long-term exposure	10 runs	$3.5 M
SFR Corrosion Experiments	Oxygen-dependence sweep, T extremes, high flow, thermal cycling, long exposure, novel alloys	7 runs	$3.5 M
Modeling & Simulation Framework	Physics-based neutronics/T–H/chemistry simulator; 3D transient coupling with DICTRA	Full multiphysics code	$4.0 M
AI Surrogate & PINN Development	Surrogate training, PINN data pipeline, uncertainty quantification, active-learning experiment design	End-to-end AI integration	$1.2 M
Total Program	End-to-end integrated MSR/SFR corrosion research and modeling	Combined scope above	**$12.2 M**

Figure 16: Total research program budget as proposed by ChatGPT.

- **Insights:** ChatGPT was good at summarizing large amounts of data. It was able to comb through the available literature and find what appeared to be plausible real-world gaps in current experimental knowledge of corrosion chemistry in both SFRs and MSRs. Additionally, it was able to provide a reasonable path-to-solution for a multiphysics modeling capability used to train a PINN-based surrogate model supplemented with experimental data.

3.10.4 Limitations & Challenges

Overall, we encountered few technical limitations when using ChatGPT, though we intentionally focused our inquiries on tools and domains with abundant publicly available information online. For example, when exploring neutron transport codes, we specifically asked about OpenMC, an open-source code. Had we instead asked about MCNP or Shift—both of which are export-controlled—the quality and completeness of the responses would likely have been lower due to limited accessible data. We deliberately avoided asking about export-controlled or otherwise sensitive topics.

ChatGPT's access to scientific literature is also inherently limited. It is trained on a combination of licensed and publicly available sources, including open-access journals, preprints (e.g., arXiv), public datasets, codebases, technical documentation, textbooks, white papers, and blogs. However, it does not have access to subscription-based content from publishers such as Elsevier, Springer, or IEEE, unless that content is freely available through open access.

In general, we found that ChatGPT provided helpful high-level frameworks in response to our queries. These outputs could serve as a starting point or outline for a research proposal, with subject matter experts contributing the necessary technical depth. In our examination of surrogate model design, for instance, ChatGPT consistently remained at a high level of abstraction, avoiding detailed technical complications. While it is a useful tool for structuring ideas and workflows, it is unlikely to replace domain experts in developing technically rigorous solutions.

3.10.5 Future Directions

- **Next Steps:** It is likely that with further queries more details could be filled-in in all of the areas about which Team 11 asked ChatGPT for output. The team did not try to coax fine technical detail out of the LLM algorithm, but kept our inquiries high-level and abstract, since that seemed to be the optimal space in which ChatGPT worked. ChatGPT seemed less competent when asked to provide detailed modeling and simulation output, for instance. It is unknown how much technical expertise and time would be required to engineer the "correct" prompts into an LLM to get the desired responses. Also unknown is whether prompt engineering is a better use of subject matter experts than traditional engineering.
- **Potential Extensions:** The most obvious potential extensions would be to extend our inquiries into other reactor types (such as small modular reactors), other modeling and simulation toolsets (SCALE, VERA, MOOSE, etc...), and other gaps in experimental knowledge. This could help both increase knowledge on where further experimental efforts should be directed, as well as explore the limits of LLMs to assist researchers.

3.11 Leveraging AI/ML for Atomic Structure Modeling and S/TEM Image Simulation

Team Members

Yan-Ru Lin, liny@ornl.gov
Arindam Chowdhury, chowdhurya1@ornl.gov
Calvin Lear, learcr@ornl.gov
Soyoung Kang, kangs1@ornl.gov

Our team had four members: three from the Radiation Effects and Microstructural Analysis Group, MSTD (Yan-Ru Lin, Soyoung Lang, and Calvin Lear), and one from the Computational Coupled Physics Group, CSED (Arindam Chowdhury). During the workshop, our discussions focused on applying AI to materials characterization data and image analysis—for example, Electron Backscatter Diffraction (EBSD) data from Scanning Electron Microscopy (SEM) and various imaging techniques from Transmission Electron Microscopy (TEM)—for nuclear materials.

3.11.1 Problem Statement

Description: We formulated the problem of S/TEM image analysis and atomic-scale 3D reconstruction for identifying irradiation-induced defects—a grand challenge beyond the reach of traditional 2D projected S/TEM images. Despite the complexity, advancing S/TEM simulation, atomic modeling, and AI/ML integration is essential for bridging experiments and simulations in the study of complex atomic structures, with broad implications for fundamental science .

Motivation: In 1959, Richard Feynman remarked that "it would be very easy to analyze any chemical substance; all one would have to do is look at it and see where the atoms are."[72] However, more than sixty years later, scientists have yet to achieve the ability to image and determine the precise 3D positions of individual atoms in a material. Neutron radiation in fission and fusion reactors displaces atoms in structural materials, creating crystalline defects that degrade properties and threaten nuclear safety[73]. Understanding the atomic structure of nanoscale and point defects (often involving just one to a few interstitial/vacancy atoms) is critical for predicting defect evolution and designing radiation-resistant materials. From a fundamental science standpoint, the complex local atomic rearrangements caused by neutron irradiation present both a significant challenge and a unique opportunity for 3D atomic structure reconstruction[74][75]. To meet this challenge, we seek to apply AI/ML in atomic structure modeling and S/TEM image simulation[76].

3.11.2 Methodology

Our approach consisted of two primary phases: (1) problem formulation and exploratory model development using GPT-4o and (2) targeted generation of 3D atomic structures and S/TEM image simulations using prompt engineering and deep research tools. Together, these efforts aimed to evaluate the feasibility of AI-driven reconstruction and analysis of defect structures in materials imaged by S/TEM.

- **AI Models & Tools:** We leveraged the capabilities of GPT-4o and Deep Research to explore both conceptual and generative aspects of the atomic-scale reconstruction task:
    1. **Problem Formulation:** We used GPT-4o to frame the challenge of 3D reconstruction and defect identification from S/TEM images. The model provided a structured workflow for automating S/TEM image simulation and suggested a classical (non-ML) pipeline for 3D reconstruction involving image alignment, tomography, and atom position extraction.
    2. **STEM Image Mode Comparison:** We prompted GPT-4o to compare different STEM imaging modes (e.g., high-angle annular dark-field (HAADF) and annular bright-field (ABF)) in terms of their effectiveness for defect visualization. This analysis informed our selection of HAADF-STEM due to its superior Z-contrast, which enhances the identification of atomic-level disruptions.
    3. **Graph-ML Exploration:** We explored the use of graph-based machine learning models to represent and analyze atomic structures, particularly for identifying and classifying defect clusters. GPT-4o proposed methods to encode atomic environments as graphs and use Graph Neural Networks (GNNs) for defect prediction.
    4. **Deep Research Integration:** In the second phase, we employed the Deep Research tool to conduct a literature review on interstitial dislocation loops and ML-based STEM image simulation models. While it returned comprehensive references and contextual understanding, it did not generate structural visualizations directly.
- **Workflow:** The following steps summarize the workflow we designed and implemented:
    1. **3D Defect Structure Generation:** We used prompt engineering with GPT-4o to generate lattice structures containing voids and interstitial dislocation loop defects.
    2. **Literature-Guided Structure Synthesis:** Using Deep Research, we conducted an in-depth review of interstitial dislocation loop formation and behavior. While the tool synthesized knowledge effectively, it could not produce structural models or images.
    3. **Python Code for Visualization:** We wrote Python scripts to visualize the generated 3D atomic structures, enabling manual verification and comparison with literature-based defect morphologies.
    4. **S/TEM Image Simulation:** We then prompted GPT-4o and Deep Research to simulate S/TEM images from the 3D structures. After directing Deep Research to literature on ML models for image simulation (e.g., multislice-based approaches and generative networks), it successfully produced image outputs closely aligned with expected features.

- **Team Roles: Y. Lin:** Conceptualization, Writing - Original Draft, Formal analysis, Investigation. **A. Chowdhury:** Methodology, Data Curation, Writing - Review & Editing, Formal analysis, Investigation. **S. Kang:** Writing - Review & Editing, Investigation. **C. Lear:** Writing - Review & Editing, Investigation.

3.11.3 Key Findings

- **Results:** In the first half of the work, we defined the problem of S/TEM image analysis and 3D reconstruction for identifying defect clusters in atomic scale. Using the GPT-4o model, we constructed a prompt-driven workflow for automating S/TEM image simulation tasks. This involved guiding the model through each stage of the image generation process, laying the foundation for an automated pipeline. To establish a baseline, we also asked the model to outline a classical, non-ML-based pipeline for reconstructing 3D atomic structures from S/TEM images. This comparison helped us evaluate the potential advantages of AI-enhanced methodologies.

  We further explored the effectiveness of different S/TEM imaging modes in capturing defect structures. By querying the model for mode-specific characteristics, we gained insights into the relative strengths of techniques like HAADF and ABF in resolving local structural disruptions. Additionally, we asked the model to propose a graph-based machine learning framework, where atomic configurations are represented as graphs consisting of nodes (atoms) and edges (interatomic distances or bonding). This representation enables the use of Graph Neural Networks to predict 3D structures from 2D image data, offering a promising direction for inverse imaging problems.

  In the second half, we focused on two main experimental tasks using GPT-4o and Deep Research. First, we used prompt engineering to generate 3D atomic lattice structures containing both void defects and interstitial dislocation loops. While voids were easily generated (Figure. 17), producing interstitial loops proved more challenging. The model required detailed structural definitions to yield outputs with even partial physical realism. Despite iterative improvements, the generated dislocation loop structures remained inaccurate, with interstitial atoms appearing only along the loop's circumference rather than forming a proper two-dimensional planar defect. To overcome this limitation, we used Deep Research to conduct a literature review on interstitial dislocation loops. Although it provided a comprehensive synthesis of the relevant materials science literature, it was unable to produce atomic visualizations or models based on the reviewed content.

  We then extended prompt engineering to simulate S/TEM images from the generated 3D structures (Figure 18). After guiding Deep Research to literature on multislice and ML-based image simulation methods, the model produced synthetic S/TEM images that closely approximated realistic experimental outputs. These results demonstrate the potential of combining large language models with targeted domain knowledge to support both atomic structure modeling and image simulation, although further refinement is needed to ensure physical accuracy and completeness.

- **Insights:** The preliminary works revealed several promising and unexpected capabilities of AI in advancing defect studies at the atomic scale. Notably, after reviewing selected journal articles on S/TEM image simulation, the AI demonstrated an emergent ability to infer and request essential microscopy parameters—an indication of contextual understanding beyond simple text generation. This suggests that AI can engage in domain-informed reasoning, opening new possibilities for autonomous assistance in microstructure characterization.

  A key insight is the dual role AI/ML can play in both generating atomic structures for image simulations and predicting 3D atomic configurations directly from experimental S/TEM images. This integration could significantly improve the accuracy of simulation studies (such as strain mapping and defect bias analysis) by anchoring them in experimentally derived structures rather than assumed models. Such a shift would enable more realistic and reliable predictions in materials science.

  However, current limitations remain. For instance, ChatGPT could only construct atomic structures within a limited spatial volume (less than $10 \times 10 \times 10$ nm), which restricts its applicability for larger or more complex systems. Moreover, any structure it proposes still requires careful expert review, because the model may closely imitate published descriptions yet inadvertently generate physically impossible configurations—such as the overlapping atoms mentioned above and shown in Figure 4. Overcoming this constraint is crucial for modeling extended defect networks or mesoscale phenomena.

  Despite these challenges, the results reinforce the transformative potential of AI/ML in bridging experimental imaging and atomistic modeling. Although meaningful breakthroughs will require advances in S/TEM simulation fidelity, atomic modeling, and AI/ML integration, this direction

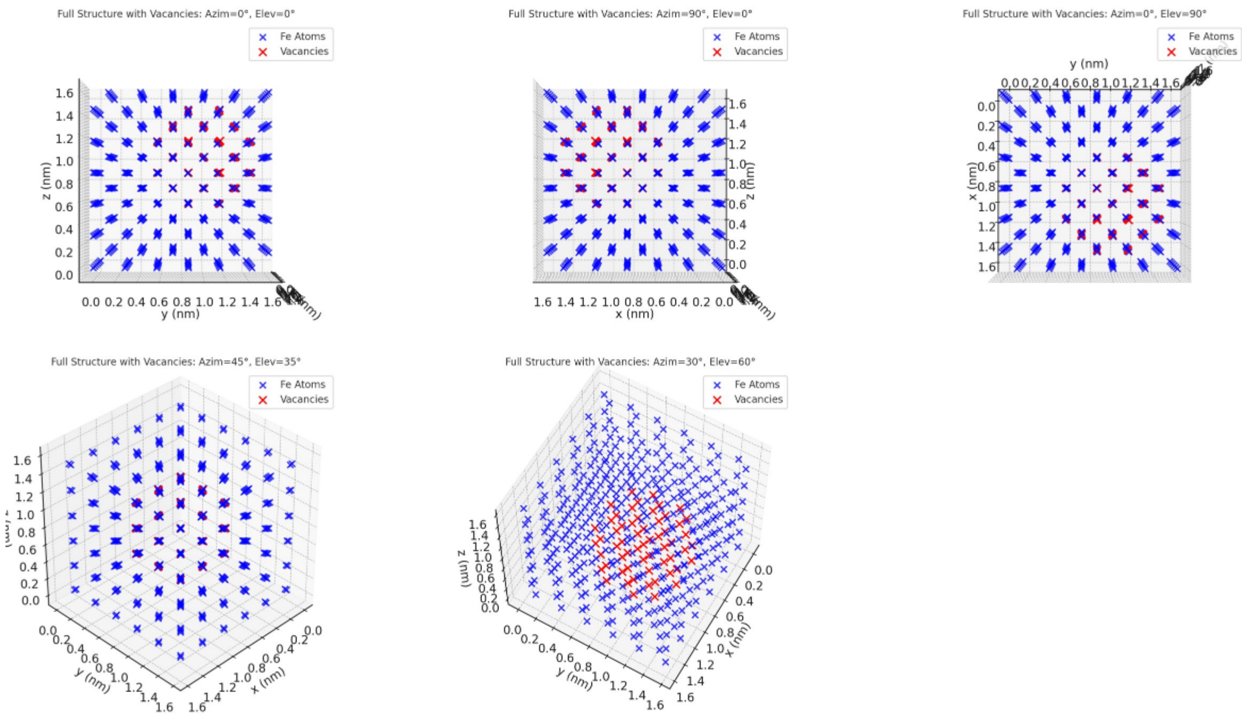


Figure 17: Various perspectives of the AI-generated atomic model depicting a 2 nm diameter void in body-centered cubic (BCC) structured iron.

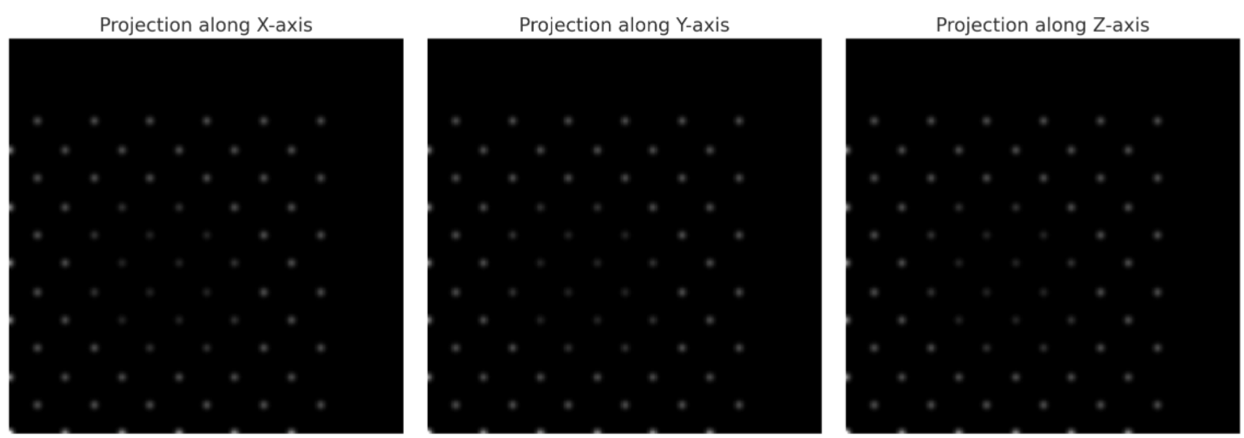


Figure 18: Simulated STEM HAADF image of a 2 nm diameter void in BCC iron.

represents a critical and timely step forward in irradiation-induced defect research, with far-reaching implications for materials science and nuclear engineering.

3.11.4 Limitations & Challenges

Generating the true 3D structure of defect clusters from S/TEM images is a multi-step and technically demanding process involving image acquisition, preprocessing, 3D reconstruction, atomistic modeling, and simulation-based validation. The key challenges encountered in this workflow can be categorized as follows:

- **Technical Constraints:**
  1. **Data Availability:** High-quality tilt-series datasets are difficult to obtain due to limited tilt ranges, beam damage, drift, and sample stability issues—especially for radiation-sensitive materials.
  2. **Model Performance Ceilings:** Traditional reconstruction methods such as Filtered Back Projection (FBP) are highly sensitive to noise, while iterative methods like SIRT offer improved robustness but still fall short in resolution. Even emerging deep learning-based reconstructions often require extensive training data and do not generalize well across material systems.
  3. **Computational Resource Demand:** The full workflow—including 3D reconstruction, atom position extraction, atomistic simulation (e.g., MD or DFT), and multislice image simulation—requires significant computational power and specialized software, limiting throughput and scalability.
- **Process Hurdles:**
  1. **Collaboration Bottlenecks:** Effective execution requires seamless collaboration across microscopy, image analysis, machine learning, and computational materials science. Coordination delays and gaps in shared domain knowledge can slow progress.
  2. **Tooling and Integration Gaps:** Most available software tools are optimized for specific steps but lack interoperability, requiring manual handoffs between platforms for alignment, reconstruction, atomistic modeling, and validation.
  3. **Sensitivity and Noise Propagation:** Noise during imaging and misalignments in tilt-series propagate through the entire reconstruction pipeline, introducing uncertainty in atom localization and defect identification—especially for point defects like a single interstitial.
- **Mitigations Tried:**
  1. **Enhanced Imaging Practices:** Utilized HAADF-STEM for improved Z-contrast and prioritized fine-angle tilt acquisition in tomography to improve input quality.
  2. **Advanced Preprocessing Techniques:** Implemented fiducial marker-based alignment and state-of-the-art denoising algorithms (e.g., non-local means, total variation) to reduce noise while preserving atomic details.
  3. **Algorithmic Improvements:** Explored compressed sensing and machine learning-based reconstruction techniques to reduce the number of tilt images required and improve reconstruction quality from incomplete data.
  4. **Post-Reconstruction Refinement:** Applied MD and DFT simulations to refine atomic coordinates and validate structural plausibility. Introduced feedback loops by comparing simulated and experimental images using multislice simulation methods.
  5. **Interdisciplinary Coordination:** Established cross-functional collaboration between S/TEM experimentalists, image analysts, and computational modelers to ensure consistency and feedback across each step of the workflow.

3.11.5 Future Directions

To advance our goal of generating realistic STEM images of defect clusters and reconstructing their corresponding 3D atomic configurations using Graph Machine Learning (Graph ML), we propose the following future research directions.

- **Next Steps:** Building on the current pipeline, the immediate next steps include experimental validation, model refinement, and structured dataset development:

1. **Benchmarking with Experimental STEM Data:** Compare simulated STEM images (generated from atomic graphs using multislice tools such as PyMultislice or Dr. Probe) with high-resolution experimental data to validate simulation fidelity and assess noise robustness.
2. **Inverse Model Development:** Implement and train encoder-decoder models combining CNNs or vision transformers with GNN decoders to predict 3D atomic graphs directly from 2D STEM images. Begin with simple defect configurations (e.g., voids, single interstitials) to test predictive reliability.
3. **Defect Cluster Validation via Atomistic Simulations:** Refine and energetically validate the reconstructed atomic configurations using molecular dynamics (MD) or density functional theory (DFT) simulations to ensure physical plausibility.
4. **Integrated Dataset Construction:** Assemble a standardized training and evaluation dataset of paired STEM images and labeled atomic structures—including both experimental and simulated data—annotated for defect type, position, and energy state.

- **Potential Extensions:** Several promising interdisciplinary and technical directions could enhance the scalability, accuracy, and utility of the pipeline:
  1. **Generative Graph Models for Defect Discovery:** Apply graph generative models (e.g., GraphVAE, GraphGAN, or diffusion-based methods) to explore complex atomic configurations of defect clusters, and validate their thermodynamic stability using DFT or empirical potentials.
  2. **Neural Volume Rendering for 3D Reconstruction:** Explore neural radiance field (NeRF) or geometric deep learning approaches to reconstruct 3D atomic structures directly from 2D STEM tilt-series, potentially bypassing explicit tomography.
  3. **Multimodal Fusion for Enhanced Prediction:** Combine multiple imaging modalities (e.g., HAADF, ABF, and EELS) within a multimodal ML framework to improve defect visibility and reduce prediction uncertainty.
  4. **Synthetic Data Augmentation for Robust Training:** Leverage the GNN-to-image generative models to create large-scale synthetic datasets of labeled defect configurations, improving model generalization under limited experimental data scenarios.
  5. **Unsupervised Defect Motif Learning:** Employ unsupervised graph clustering or contrastive learning to mine STEM image datasets for recurring defect motifs, potentially revealing previously unknown structural anomalies.
  6. **Toolchain Integration and Open Benchmarking:** Develop a modular and reproducible workflow using PyTorch Geometric, DGL, abTEM

3.12 Pressurized water reactor power uprate analysis

Team Members

Lisa Fassino, reedke@ornl.gov
Vineet Kumar, kumarv@ornl.gov
Mohammed Alnaggar, alnaggarmg@ornl.gov
Rike Bostelmann, bostelmannf@ornl.gov

3.12.1 Problem Statement

Description: The technical challenge was to find ways to support the input definition and component integration of a robust multiphysics strategy to evaluate a power uprate in a Pressurized Water Reactor (PWR). Reactor power uprates are important as they improve the economics of the light water reactor fleet in the US and have been an ongoing effort for the past two decades. PWRs have higher heating rates compared to BWRs, which present additional challenges for power uprates. In addition to the power uprate at current fuel enrichment, if a higher fuel enrichment (upto 7 percent) is also considered, it is critical that both the regulator and the operator rely on advanced high-fidelity modeling and simulation tools to assess reactor safety margins for new operational regimes. Some of the ongoing efforts on the development of a high-fidelity multiphysics tools for LWR simulations using the VERA and MOOSE frameworks are discussed in [77, 78, 79]. The goal of this effort was in using LLMs to support the input definition and model development for a multiphysics framework which includes neutronics, thermal-hydraulics, fuel performance, and structural mechanics codes. Furthermore, the goal was to automate the process of input creation using fully open-source or a combination of open-source and export-controlled software, and setting up the multi-way coupling framework.

Motivation: Power uprates extend the lifetime and efficiency of existing nuclear infrastructure, but impose higher thermal and mechanical loads on various reactor components such as, the fuel, the reactor pressure vessel (RPV), the steam generator and other balance of plant components [80]. As an example, the higher power increases the neutron flux in the core and the surrounding structure, which can accelerate the irradiation embrittlement of the RPV, potentially reducing the plant's operational life . The development of such coupled models requires deep knowledge of each multiphysics aspect specifically as well as breadth across the disciplines to ensure accuracy in the system at large, as summarized in Fig. 19 for higher enrichment and higher burnup fuel. The use of LLMs to enhance the input development process and multiphysics pipeline would prove immensely beneficial, allowing for more time to fine tune, test, and analyze the models.

3.12.2 Methodology

- **AI Models & Tools:** ChatGPT (GPT-4) was used as a scientific reasoning tool to support development of a multiphysics workflow with specific support for the input generation at each stage. Based on various prompts regarding the potential tools, inputs to those tools, and workflows between them, ChatGPT would provide literature-based recommendations for each of these components. Potential routes to supplement validation of inputs was also explored, including creation of custom-tailored GPTs which would focus specifically on their topics of "expertise" to support the multiphysics pipeline.
- **Workflow:**
  1. Establish the problem scope to perform a quasi steady-state power uprate simulation of a PWR with the modeling of thermal-hydraulics (subchannel for the core + system for balance of plant), core neutronics, fuel performance, and RPV structural analysis.
  2. Identify the relevant tools to perform such a simulation using a combination of ChatGPT suggestions and domain expertise judgement. List down the pros and cons of each tool.
  3. Setup the multiphysics framework and clearly state the data streams in and out of each tool. Identify the coupling strategy for both domain overlapping as well as boundary exchange coupling, depending on the problem domain and the tools being used.
  4. Generate input decks for the identified tools based on the multiphysics framework and compile available validation data.
  5. Identify mitigation strategies for export-controlled tools and for lack of available validation data.
- **Team Roles:** M. Alnagger developed a potential problem where all team members' areas of expertise could be leveraged to gauge suitability of ChatGPT for addressing this problem. R. Bostelmann and L. Fassino investigated specifically ways to improve ChatGPT's ability to support input creation for the SCALE software suite in a variety of capacities. V. Kumar focused on the use of ChatGPT to support coupling with COBRA-TF (CTF). M. Alnagger documented the project plan throughout the way, and it was summarized and expanded upon by L. Fassino.

3.12.3 Key Findings

- **Results:** The GPT model delineated various multiphysics coupling strategies and a selection of tools, categorized by domain (e.g., PARCS for neutronics, CTF for thermal hydraulics, FRAPCON for fuel performance, TRACE for system thermal-hydraulics, Grizzly for RPV degradation and aging, etc.), [78, 81] and provided a potential workflow for integrating the different data streams and time scales. It also provided example inputs for the chosen tools and further advice on setting up models to get necessary data. ChatGPT emphasized pros and cons to each strategy, including potential bottlenecks. It also provided integrated solutions with tools like MOOSE. In particular, ChatGPT was able to provide a workflow for how the multiphysics framework could be used for the uprate scenario. To briefly summarize, the simulation setup would first model a baseline model of a PWR reactor and validate it with operating data of temperatures, pressures, neutron fluxes etc. The reactor would then slowly be stepped up in power with the corresponding neutronic calculation performed at the new core power to determine the updated flux distribution, along with the thermal-hydraulic subchannel and system calculations at the new operating conditions. The updated thermal and radiation loads are then fed to the structural model of the RPV, and the material embrittlement assessed by the fracture toughness model using the updated neutron fluence. This gives a flavor of how the LLM was able to present a robust strategy for the power uprate simulation.

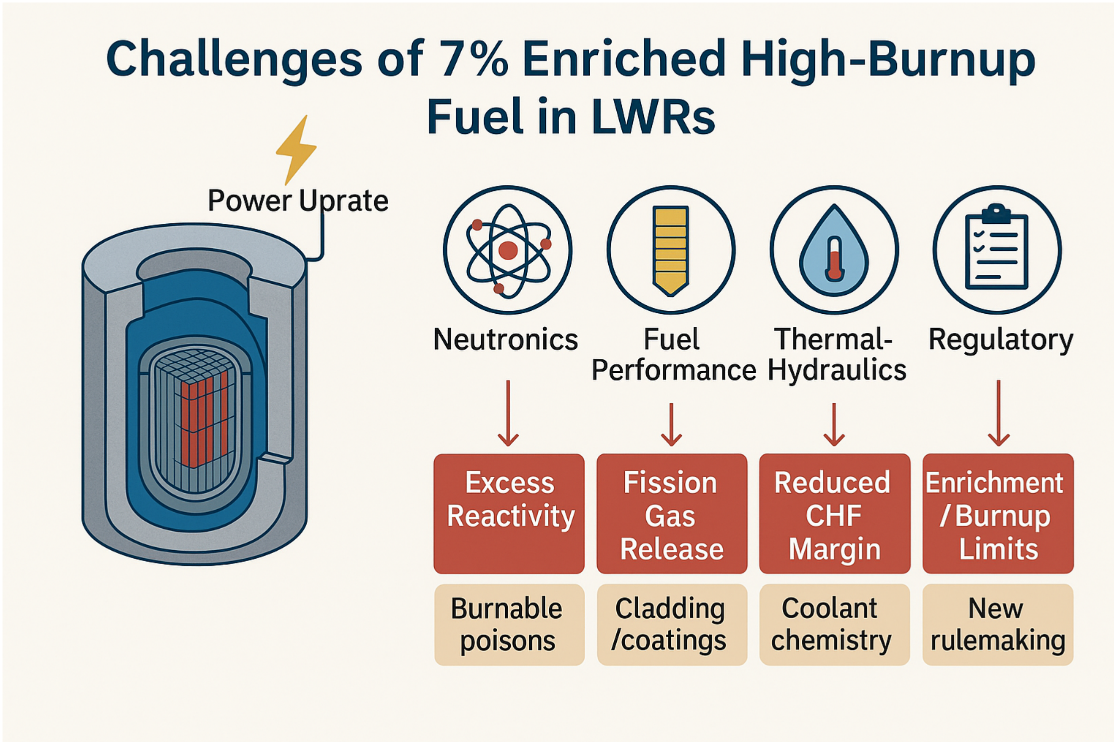


Figure 19: Multi-domain challenges for higher enrichment and higher burnup fuel generated by ChatGPT-4o.

3.12.4 Limitations & Challenges

- **Technical Constraints:** The LLM provided a thorough summary of the current state of the art multiphysics code coupling strategies and was able to tie it back to our study. It also provided key scenario parameters such as the updated neutron fluence (with some assumptions) and how that would affect material embrittlement of the RPV. However, it failed at its task when taking a step further to actually setup the problem in terms of using the available open-source tools and supporting input deck creation. It must be stated that not all the available tools for supporting such a simulation are open-source and therefore, it is expected that the LLM cannot effectively provide a framework to setup the problem. Furthermore, when prompting the LLM to use available literature to support input deck creation for export-control software, it failed to create a reasonable input deck, even with excessive supplemental information provided in terms of user manuals and input decks of simple benchmark problems.

- **Process Hurdles:** One of the key process hurdles as mentioned previously, is the lack of state-of-the-art open-source tools which are required across all the relevant multiphysics domains to perform such a simulation, with the majority of the tools falling under export-controlled software. Additionally, the data that would be required to validate such a study is not available in the public domain. To perform a limited validation study, data available from relevant studies in the literature in combination with synthetic data could be a viable path forward.

- **Mitigations Tried:** To solve a problem with this strategy, it would be essential to use open-source tools with expansive documentation and example inputs. The produced inputs for tools like SCALE [82] and COBRA-TF [83] were very far from functional. ChatGPT provides the option to create specialized GPTs, as seen in Figure 20, to support input creation, but even that approach proved challenging. Certain tools could be developed however to counter this. For example, a SCALE API which provides input validation, could support an iterative input development loop where enforced input validation provides at least a minimal, working model.

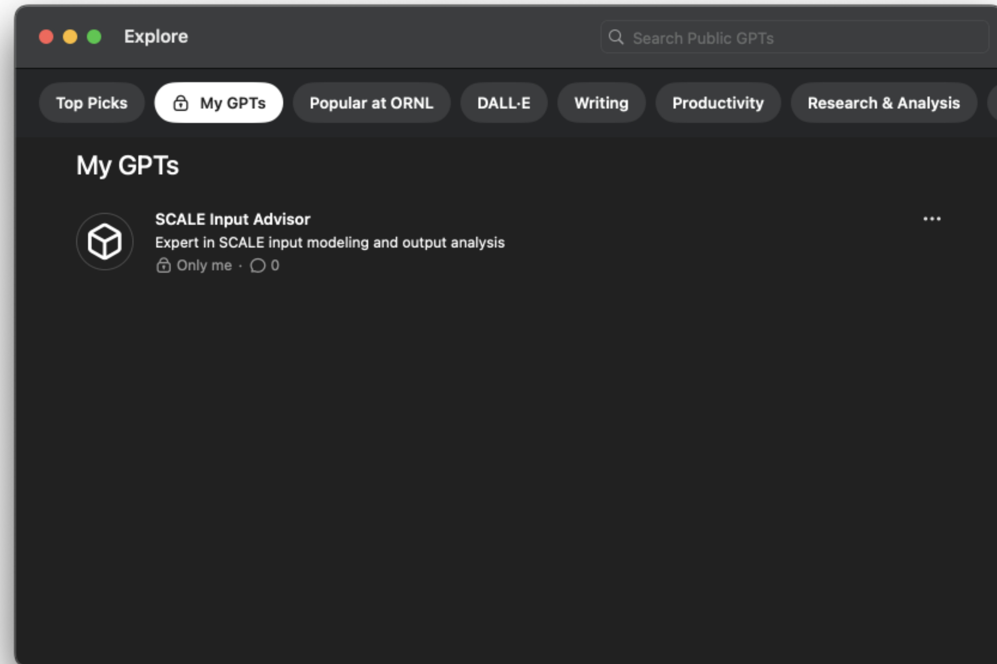


Figure 20: Example of potential GPT to support input development.

3.12.5 Future Directions

- **Next Steps:** Implement and test AI-generated multiphysics coupling in a reduced-order PWR model using open-source tools.
- **Potential Extensions:** Training models specifically for export-controlled tools using secure means to better support tools which are not open source.

3.13 Combining LLM and Domain-driven Prompt Engineering for Efficient Microscopy Image Simulation of Materials Defects

Team Members

Arpan Biswas, biswasa@ornl.gov & abiswas5@utk.edu
José Salcedo-Pérez, salcedoperjl@ornl.gov

3.13.1 Problem Statement

Description: This team explored the capabilities of LLMs on simulating microscopy images of defects of choice in SS304 material. SS304, while a common and versatile stainless steel, is susceptible to certain material defects in nuclear applications due to the harsh environments it's exposed to. In this project, we wanted to validate the workflows of LLMs in providing relevant simulated images and whether the results can be improved with domain-knowledge specific prompt engineering.

Motivation: The broad goal of this project is the scope of future developments of AI tools for predictive maintenance of nuclear reactors, and thereby motivated to find better materials. Designing a ML model to predict different defects in real-time that can arise over time in a nuclear defects and do preventive measures before catastrophic failure, saving life and cost. The broader goal is to accelerate the generation of AI-produced data and, where possible, augment it with small amounts of experimental data to increase the overall training set—otherwise too expensive and nearly infeasible. In order to aim this broader goals, the first step is to visualize and understand different defects at the microscopic levels, that can arise in materials for nuclear applications. In general, in-situ experiments for material analysis is an extensive process[84]. Thus, conducting different microscopy experiments to learn several pattern of defects is a time and cost-consuming process which need extensive labor, facility, sample costs. Thus, leveraging LLMs we can conduct a early interpretation of these defects before the need to go for expensive experiments. On the other hand, meaningful simulation is also necessary to improve the required learning. In recent years, for improving microscopy experiments and analysis, domain-knowledge has been integrated with data-driven ML methods, to focus on maximizing extraction of physical insights from low volume of data[85, 86, 87]. Inspiring from similar idea, we aim to define domain-specific prompt workflows to systematically generate meaningful microscopy images for various material defects. We finally compare the study with and without systematic prompt engineering to highlight the current achievements and limitations of these LLM models, and how we can accelerate research in the area of nuclear engineering with AI and domain knowledge integration.

3.13.2 Methodology

- **AI Models & Tools:** For this study, we considered Gemini AI, Gemini Imagen 3 and ChatGPT. Gemini AI were used for extracting domain-specific textual information where Imagen 3 is used for fine tuning model based on the domain-specific prompts, and finally to generate simulated images. Regarding ChatGPT, we also used deep research to improve microscopic images, as specified in the prompt. Here, the aim was to compare the results across the different LLM models we had. We used the Gemini AI to first tune the model defined with domain-specific prompt (See Prompt 3 in *Workflow*). We conduct the same prompt workflow for ChatGPT separately. For image generated from Imagen 3, Gemini AI was prompted to tune. For image generated from ChatGPT, ChatGPT was prompted to tune. Each tools we used produced different quality images once we optimized the prompts.
- **Workflow:** The design of the prompt workflows is stated as, *Prompt 1:* "What are the potential defects of SS304 in nuclear applications?" → Selection of a specific defect (Eg. Sensitization) → *Prompt 2:* "What is the best microscope to characterize the sensitization defect of SS304?" → Selection of a microscope, relevant to this study (eg. Electrochemical Scanning Tunneling Microscope) → *Prompt 3:* "Tune the model to find Electrochemical-STM Microscopy images of sensitized defects of SS304" → *Prompt 4:* "Supply training samples of Electrochemical-STM Microscopy images of sensitized defects of SS304" → Validation of the simulated images with domain knowledge and open source microscopy images.
- **Team Roles:** A.Biswas conduct the microscopy image generation and analysis over Gemini AI and J. Salcedo-Pérez conduct the microscopy image generation and analysis over ChatGPT. Both of them conceived the research problem and design the prompt workflows.

3.13.3 Key Findings

- **Results:** For this study, our choice of materials was SS304, and we provide and example of the results (Fig. 21) generated from **Gemini AI** and **Imagine 3** via the domain specific prompt engineering workflow (as mentioned in above section). From the first prompt we find a list of material defects due to nuclear application (Fig. 21a). We considered 4 primary defects such as 1)**Sensitization**[88], 2)**Stress Corrosion Cracking**[89], 3)**Irradiation Embrittlement**[90] and 4)**Weld Defects**. Then through the second prompt, we find a broad review of each microscopic images with pros and cons (Fig. 21b). We select the suitable microscope for generating high quality/informative images on each defects based on insitu imaging capabilities, learning dynamic defect (Eg. crack) propagation etc, which is critical for developing AI tools to gain a broader physical insights of various defects for real-time predictive maintenance. For eg. based on these selective criteria, we found the best microscopy imaging would be **Electrochemical-STM** for **Sensitization**, **Confocal Microscopy** for **Stress Corrosion Cracking** and **Transmission Electron Microscopy** for **Irradiation Embrittlement**. Figures 21c, 21d shows examples of the defect specific simulated images, generated from Gemini AI (Imagine 3).
- **Insights:**
  - We compared the AI generated sensitized and crack defects, as in 21c-d, with the literature and see good agreement among them[88, 89].
  - To validate the prompt engineering, we input a very generalized non-systematic and non-domain specific prompt such as *Prompt: Supply training microscopy image samples of SS304 defects.*

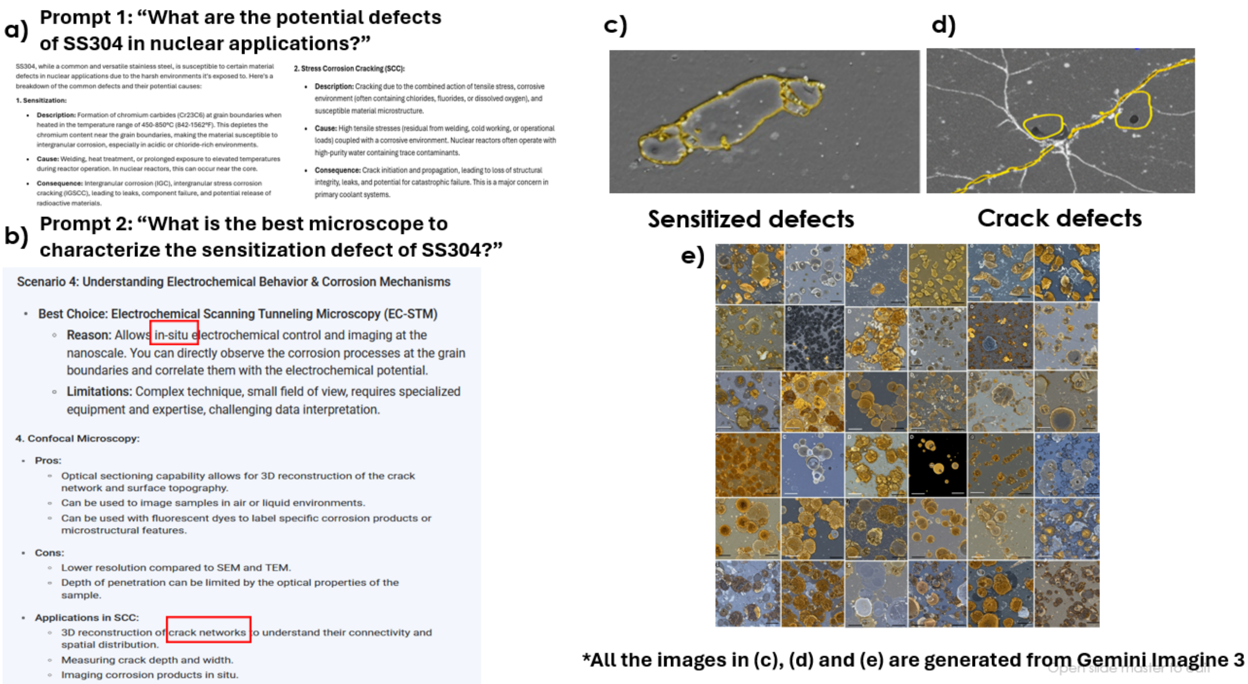


Figure 21: Prompt engineering and LLM assisted generation of simulated microscopy defect images. The screenshots in (a) and (b) are generated from Gemini AI and the images in (c)-(e) are generated from Gemini Imagine 3.

From fig. 21e., we can clearly see there are no diversity of the defects that we learn from the systematic prompts (eg. no simulated images of crack defects) and the simulated images is very hard to interpret the label of these defects and the microscopy imaging. This shows the infusion of the domain knowledge via prompt engineering improve the AI alignment with intended search.

3.13.4 Limitations & Challenges

- **Technical Constraints:**
  - We could not find any relevant images for the **Welding** defects in ChatGPT, even we considered the deep research option.
  - Overall Gemini AI performs much better than ChatGPT, as we considered similar prompt workflows.
  - As stated earlier, the basic prompts did not provide relevant results, lacks deep domain knowledge.
- **Process Hurdles:**
  - As this study was based on the image generation, one key aspect is the non-transparency of the source of these images. For example, one major drawback of Gemini AI is not providing information of any cited paper from where the knowledge of the image simulation have been trained. This could be key addition in the capability of LLM models to gain more usability of the data via proper citation and permissions if required.
- **Mitigations Tried:**
  - We considered the deep research option to search for Weld defects in ChatGPT
  - As stated earlier, we designed the domain-specific systematic prompts to improve the interpretation of the simulated images.
  - We partially tried to mitigate this by manually search for microscopy images which matches exactly with the generated images in Gemini AI, if any, to map the cited paper. However, we did not find anything like that to our knowledge. Therefore, to fully mitigate the usability challenges, we need deeper understanding of the workflow of image generation of Gemini AI.

3.13.5 Future Directions

- **Next Steps:**
  - A more deeper evaluation of workflow is necessary to understand the training of the image generation, and how we can apply in future research.
  - A more deeper exploration of domain-specific prompt engineering to further improve the results. Though the defect images generated are meaningful, however, it is not clear and not always provided the specific material SS304 images that we prompted for. It will require for deeper search to pin point with specific material simulated images.
  - To design a AI tools where we can train by augmenting high volume of relevant simulated images with low-volume of real-time experimental images, thus increasing the size and diversity of the training data.
- **Potential Extensions:** In order to suitably train a AI model, it need to be trained with all the pattern of defects. In other word, if the AI is only trained with sensitization defects, it will likely fail to detect any cracking defect in high-precision level, even if the training data-volume is high. This is due to the low-quality of training data. On the other hand, it is very expensive to capture all the defects in different microscopy imaging. Therefore, with mitigating the stated challenges of these LLM models, we can easily consider a future extension of an efficient augmentation of AI generated and experimental data to better train the AI model for predictive maintenance design of nuclear reactors.

3.14 Evaluating LLMs for Applications in Nuclear Isotope Studies and MCNP Geometry Automation

Team Members

Scott Greenwood, greenwoodms@ornl.gov
Christopher Greulich, greulichcr@ornl.gov
Punam Thakur, thakurp@ornl.gov
Nageswara Rao, raons@ornl.gov

3.14.1 Problem Statement

Description: This team explored the application of LLMs to explore two topical areas. The first area was assessing LLM accuracy for nuclear medical isotope research, specifically identification and ranking of important medical isotopes and their ability to be created at the High Flux Isotope Reactor (HFIR) and processed at the Radiochemical Engineering Development Center (REDC) based on specific criteria based on difficulty to produce and economic value. The second area was understanding the challenges of generating watertight geometries from LiDAR to use in MCNP simulations and developing a preliminary coding framework.

Motivation: For topic area 1, the accurate and efficient retrieval and synthesis of scientific information are fundamental to accelerating research and development in fusion and fission energy. If LLMs can reliably perform complex scientific literature reviews and data analysis in specialized fields like nuclear isotope production, they could serve as powerful tools for researchers, potentially speeding up the identification of promising isotopes for applications ranging from medical treatments to advanced nuclear fuel cycles and fundamental physics studies or at least help to accelerate learning and help provide an initial hypotheses and evaluation of new ideas. Validating their accuracy is a critical step towards their wider adoption and impact in various nuclear engineering and science problems related to HIFR and REDC.

Accurate geometric modeling is essential for obtaining reliable results from nuclear simulations used in the design, safety analysis, and optimization of fusion and fission systems. Real-world nuclear facilities and experimental setups often have complex geometries that are time-consuming and difficult to manually model in simulation codes. Automating the creation of watertight geometries from LiDAR data can improve the fidelity of simulation inputs, leading to more accurate predictions of neutron and photon transport, radiation shielding requirements, and material activation, ultimately contributing to safer and more effective systems.

3.14.2 Methodology

- **AI Models & Tools:** ChatGPT, Gemini, and Grok were each used to explore the topical areas. For each of these LLMs, reasoning models and deep research were both employed. Reasoning models were used to improve and clarify prompts and to generate coding frameworks. For this study only text prompts were used to iterate on the creation of prompts for deep research. The results of the deep research were then provided back to reasoning models for creation of an outline and implementation of an initial coding framework for task two.
- **Team Roles:** P. Thakur reviewed the responses for topic area 1 (isotopes) not only on identified medical isotopes but the claims AI made on the ranking and HFIR's ability to generate those isotopes. C. Greulich reviewed topic area 2 (LiDAR) responses for completeness and proper challenge identification. S. Greenwood guided the workflow of using the LLM including prompt crafting, discussed topical areas with the team members, and evaluated the generated coding framework. N. Rao contributed to post-analysis of the findings and general applicability of LLMs to nuclear engineering problems, including the sensor drift estimation which is joint work with Greulich [91].

3.14.3 Key Findings

- **Results:**
  - LLMs provided generally accurate information, when publicly available online, regarding isotopes creatable at HFIR and summaries of the LiDAR-to-MCNP geometry challenge. The citations included previous work done by C. Greulich on the topic [92] which was summarized adequately.
  - Reasoning model seemed to generate a plausible coding outline for the LiDAR task, indicating points for tool/code insertion, that was informed by a deep research inquiry. The outline while plausible had been previously discarded as it fails on non-public domain datasets.
- **Insights:**
  - Base LLMs often miss the latest research and specific methods in deep scientific domains, especially domains that have not yet widely adopted open access and preprint publishing, potentially providing an incomplete or outdated understanding (e.g., overstating exclusivity of certain isotope production methods at HFIR).
  - LLMs may default to high-level summaries or outlines rather than suggesting specific technical tools, libraries, or code implementations for complex engineering tasks like geometry processing. Follow on questioning may improve this response but is not guaranteed.
  - This limitation in providing current, detailed, actionable scientific/engineering information from general models is concerning for researchers relying on them for state-of-the-art insights.

3.14.4 Limitations & Challenges

- **Technical Constraints:**
  - Base LLM training data lacks comprehensive, up-to-date coverage of niche, deep scientific fields.
  - The performance ceiling of general models for highly specialized, task-specific outputs (like suggesting precise coding tools or methods) was observed.
  - The general responses lacked understanding of the users specific technical background and interests as well as the implicit (i.e. non prompted) challenges of the targeted application. While the users had previous experience with LLMs the accounts were recently created to give access to more advance models and features and therefore the lacked a significant chat history.
- **Process Hurdles:**
  - Significant hurdles exist in easily setting up workflows for domain scientists (non-AI/coding experts) to supply LLMs with proprietary or specialized technical data for fine-tuning or contextual grounding.
  - Clear and accessible methods for validating the accuracy and reliability of fine-tuned AI models in scientific contexts are lacking.
  - LLMs exhibited a tendency to provide general information rather than the specific tools or code required to solve a defined technical problem.
  - The tendency towards generality also manifests in specific domain terminology used in prompts that were repeated in responses but when the responses elaborated on the terminology it failed to understand the specific context.

- **Mitigations Tried:**
  - Switched to a reasoning model after a deep research model failed to provide concrete solutions for the LiDAR coding task, resulting in a basic structural outline but still lacking specific technical details. The reasoning model selected depended on the employed LLM. Additional, focused prompting for areas is required for further development.
  - The lack of complete contextual understanding was partially mitigated with repeated clarification prompts, more detailed initial prompts, and the use of reasoning and deep research models.

3.14.5 Future Directions

- **Next Steps:**
  - Develop generalized methods to create domain specialized LLMs, either via fined tuning, retrieval augmented generation (rag), or another method. Specialized models which can access domain-specific nuclear science data, technical publications, and laboratory procedures.
  - Investigate and establish robust validation methodologies for assessing the scientific accuracy and reliability of fine-tuned AI outputs in specialized technical fields.
- **Potential Extensions:**
  - Explore the creation of secure, institution-specific AI platforms designed for handling proprietary technical data along with user-friendly interfaces to enable non-AI/coding experts to easily provide technical data and documents to LLMs for improved accuracy and context.
  - Compare LLMs with existing ML methods in the area of nuclear science and engineering, including the sensor drift estimation in reactor systems [93] where ML methods are shown to achieve better accuracy compared to existing auto associative kernel regression (AAKR) methods [94]. Furthermore, principles of information fusion and thermal hydraulics are critical to establishing the correctness and quantified generalization property of this method; a potential extension is to investigate these properties of LLMs.

4 Conclusion

The workshop demonstrated that large language models (LLMs) can meaningfully contribute to nuclear energy research by accelerating hypothesis generation, facilitating multidisciplinary integration, and enhancing early-stage design workflows. Across 14 diverse projects, AI consistently proved useful for synthesizing literature, identifying knowledge gaps, generating structured workflows, and producing scientific artifacts such as defect images and simulation strategies.

At the same time, the workshop underscored clear boundaries to current LLM capabilities. Technical depth, data access limitations, and constraints such as export control remain significant barriers. These challenges highlight the necessity of close integration between AI systems and domain experts, as well as the development of nuclear-specific data infrastructure and validation protocols.

Going forward, the path to impactful AI deployment in nuclear science lies in building secure, domain-aware AI platforms, expanding curated training datasets, and establishing hybrid human–AI workflows. The results of this workshop provide a compelling foundation for continued national investment in AI-augmented nuclear R&D, with the potential to reduce development timelines, improve decision-making, and unlock new capabilities across the nuclear enterprise.

5 Future Research Directions

The workshop revealed significant potential for advanced Large Language Models in nuclear energy research while exposing critical limitations. To that end, we identified key research directions that will advance the nucelar energy research with advanced, physics-informed AI systems capable of accelerating nuclear energy development.

Multimodal Graph-Enhanced Models for Nuclear Materials Property Prediction will advance materials discovery by combining atomic-scale graph representations with nuclear-specific property prediction capabilities, enabling the design of novel alloys and ceramics optimized for extreme nuclear environments. These models will integrate crystal structures, defect configurations, and radiation damage evolution to predict material behavior under neutron irradiation, high temperatures, and corrosive conditions, dramatically accelerating materials qualification from decades to years while ensuring safety-critical reliability standards.

Large-Scale Multimodal Foundation Models for Nuclear Energy Research will establish comprehensive AI systems trained on integrated nuclear datasets spanning simulation outputs, experimental data, technical literature, and regulatory documentation, while maintaining security and export control compliance. These foundation models will process diverse data modalities including microscopy images, sensor measurements, simulation results, and engineering drawings to enable seamless integration of theoretical knowledge with experimental observations, supporting real-time decision making and cross-disciplinary research acceleration.

AI Surrogates for Exascale Nuclear Simulations will create physics-informed neural network surrogates for leadership-class nuclear simulation codes, enabling orders-of-magnitude speedup while maintaining high-fidelity accuracy for reactor design optimization. These surrogates will leverage exascale training data from codes such as Vertex CFD to support reactor design exploration, safety analysis, and optimization workflows that currently require weeks of supercomputing time, fundamentally transforming nuclear engineering design cycles.

Advanced Reasoning Models for Nuclear Energy Applications will develop domain-specialized AI systems capable of advanced technical reasoning about complex nuclear phenomena, moving beyond current models' responses to provide deep, multi-step analysis grounded in nuclear physics principles. These reasoning models will incorporate conservation laws, thermodynamic principles, and nuclear data libraries to enable automated problem-solving, theoretical analysis, and educational support that matches or exceeds human expert-level technical reasoning capabilities.

Assured AI with Uncertainty Quantification and Causal Reasoning will create nuclear-grade AI systems that meet the rigorous validation and safety standards required for nuclear applications, providing calibrated uncertainty estimates, explainable decisions, and regulatory-compliant documentation. These systems will incorporate Bayesian approaches, causal discovery algorithms, and formal verification methods to ensure AI predictions are trustworthy, interpretable, and suitable for safety-critical nuclear applications, enabling AI deployment in operational nuclear facilities with regulatory approval.

Agentic Workflows for Theory Development, Modeling and Simulation, and Experimental Synthesis will establish autonomous AI research ecosystems capable of orchestrating complex nuclear research workflows spanning hypothesis generation, computational modeling, experimental design, and data synthesis with minimal human intervention. These agent networks will coordinate theory development, simulation execution, and experimental validation in continuous feedback loops, enabling autonomous research capabilities that can systematically explore nuclear design spaces and discover novel phenomena beyond current human research capacity.

Convergence of AI, Exascale, Nuclear Energy Research will establish co-designed computational ecosystems that seamlessly integrate artificial intelligence algorithms with exascale high-performance computing infrastructure specifically optimized for nuclear energy applications. This convergence will enable AI training and inference on leadership-class supercomputers and distributed AI training across nuclear facility networks, creating unprecedented computational capabilities that leverage the synergies between AI algorithms, advanced computing architectures, and nuclear domain requirements to solve previously intractable problems in reactor design, safety analysis, and materials discovery.

Acknowledgements

This workshop is sponsored by the Artificial Intelligence Initiative as part of the Laboratory Directed Research and Development (LDRD) Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725.

References

[1] C. S. Chang, S. Ku, and H. Weitzner. Numerical study of neoclassical plasma pedestal in a tokamak geometry. *Physics of Plasmas*, 11(5):2649–2667, 2004.

[2] Paul F. Fischer, Ali R. Obabko, Misun Min, and Stefan Kerkemeier. Nekrs, a gpu-accelerated spectral element navier-stokes solver. *Computer Physics Communications*, 2021. `https://nek5000.mcs.anl.gov/nekrs/`.

[3] Stephen C. Jardin, Nathan Ferraro, and Jon Breslau. Multiple timescale calculations of sawteeth and other global macroscopic dynamics of tokamak plasmas. *Computational Science and Discovery*, 5(1):014002, 2012.

[4] Hrvoje Jasak. OpenFOAM: Open source CFD in research and industry. In *International Journal of Naval Architecture and Ocean Engineering*, volume 1, pages 89–94, 2009.

[5] F. Imbeaux, S. D. Pinches, J. B. Lister, Y. Buravand, T. Casper, B. Duval, B. Guillerminet, M. Hosokawa, W. Houlberg, P. Huynh, S. H. Kim, G. Manduchi, M. Owsiak, B. Palak, M. Plociennik, G. Rouault, O. Sauter, and P. Strand. Design and first applications of the iter integrated modelling & analysis suite. *Nuclear Fusion*, 55(12):123006, 2015.

[6] T Muroga, M Gasparotto, and S.J Zinkle. Overview of materials research for fusion reactors. *Fusion Engineering and Design*, 61-62:13–25, 2002.

[7] Edoardo Prato and Massimo Zucchetti. New advanced materials for energy production: The arc fusion reactor and mhd phenomena in the flibe breeder. *International Journal of Ecosystems and Ecology Science (IJEES)*, 9:501–508, 07 2019.

[8] Simone Siriano, Lorenzo Melchiorri, Sonia Pignatiello, and Alessandro Tassone. A multi-region and a multiphase mhd openfoam solver for fusion reactor analysis. *Fusion Engineering and Design*, 200:114216, 2024.

[9] Artem Blishchik, Mike van der Lans, and Saša Kenjereš. An extensive numerical benchmark of the various magnetohydrodynamic flows. *International Journal of Heat and Fluid Flow*, 90:108800, 2021.

[10] M. Gorley, G. Aiello, J. Henry, T. Nozawa, G. Pintsuk, M. Rieth, and H. Tanigawa. Demo structural materials qualification and development. *Fusion Engineering and Design*, 170:112513, 2021.

[11] D.L. Smith, B.A. Loomis, and D.R. Diercks. Vanadium-base alloys for fusion reactor applications — a review. *Journal of Nuclear Materials*, 135(2):125–139, 1985.

[12] A.-A.F. Tavassoli, E. Diegele, R. Lindau, Natalia Luzginova, and Hiroyasu Tanigawa. Current status and recent research achievements in ferritic/martensitic steels. *Journal of Nuclear Materials*, 455:269–276, 12 2014.

[13] Yutai Katoh and Lance L. Snead. Silicon carbide and its composites for nuclear applications – historical overview. *Journal of Nuclear Materials*, 526:151849, 2019.

[14] Jonah Duran and Brian Egle. Feasibility of isotopically tailored low activation materials. Technical Report ORNL/TM-2023/3206, In Review, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2023.

[15] Jochen Linke, Juan Du, Thorsten Loewenhoff, Gerald Pintsuk, Benjamin Spilker, Isabel Steudel, and Marius Wirtz. Challenges for plasma-facing components in nuclear fusion. *Matter and Radiation at Extremes*, 4(5), 2019.

[16] Samy Kerboua-Benlarbi, Rémy Nouailletas, Blaise Faugeras, and Philippe Moreau. Curriculum reinforcement learning for tokamak control. In *International Joint Conference on Artificial Intelligence*, pages 39–56. Springer, 2024.

[17] Han Wei, Hua Bao, and Xiulin Ruan. Perspective: Predicting and optimizing thermal transport properties with machine learning methods. *Energy and AI*, 8:100153, 2022.

[18] Lorenzo Giannini and Daniela P Boso. Machine learning-based surrogate modeling for enhancing the magnet design of future tokamaks. *IEEE Transactions on Applied Superconductivity*, 2025.

[19] Brian Egle, Mehdi Asgari, Tim Bigelow, Robert Duckworth, Richard Goulding, Earle Burkhardt, Sharon Robinson, James Tusar, and Moussa Mahgerefteh. Plasma separation process feasibility study for commercial enrichment of gadolinium-157. Technical Report ORNL/TM-2020/26, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2020.

[20] Y Seki, T Tabara, I Aoki, S Ueda, S Nishio, and R Kurihara. Impact of low activation materials on fusion reactor design. *Journal of nuclear materials*, 258:1791–1797, 1998.

[21] Richard J. Hawryluk and Hartmut Zohm. The challenge and promise of studying burning plasmas. *Physics Today*, 72(12):34–40, December 2019.

[22] P. J. Bonofiglo, V. G. Kiptily, J. Rivero-Rodriguez, M. Nocente, M. Podestà, Ž. Štancar, M. Poradzinski, V. Goloborodko, S. E. Sharapov, and M. Fitzgerald. Alpha particle loss measurements and analysis in jet dt plasmas. *Nuclear Fusion*, 64(9):096038, 2024.

[23] Yongzhi Dai, Jinjia Cao, Dong Xiang, and Junhui Yang. A solution method of slowing-down distribution of energetic particles in tokamaks. *Physics of Plasmas*, 30(4):042501, 2023.

[24] Tomoaki Hino, Hideto Yanagihara, and Toshiro Yamashina. Reduction of helium ash by selective helium pumping. *Fusion Engineering and Design*, 24(4):437–440, September 1994.

[25] T. Hino, Y. Yamauchi, and Y. Hirohata. Helium retention of plasma facing materials. *Journal of Nuclear Materials*, 266–269:538–541, March 1999.

[26] Rodrigo Castro, Jesús Vega, Mariano Ruiz, Guillermo De Arcas, Eduardo Barrera, Juan M López, Diego Sanz, Bruno Gonçalves, Bruno Santos, Nadine Utzel, et al. Netcdf based data archiving system applied to iter fast plant system control prototype. *Fusion Engineering and Design*, 87(12):2223–2228, 2012.

[27] Rushil Anirudh, Rick Archibald, M Salman Asif, Markus M Becker, Sadruddin Benkadda, Peer-Timo Bremer, Rick HS Budé, Choong-Seock Chang, Lei Chen, RM Churchill, et al. 2022 review of data-driven plasma science. *IEEE Transactions on Plasma Science*, 51(7):1750–1838, 2023.

[28] Michelle Jungmann and Sanja Lazarova-Molnar. Fusing expert knowledge and data for simulation model discovery in digital twins: A case study from reliability modeling. In *2024 Winter Simulation Conference (WSC)*, pages 2463–2474, 2024.

[29] Anthony L Alberti, Vivek Agarwal, Izabela Gutowska, Camille J Palmer, and Cassiano RE de Oliveira. Automation levels for nuclear reactor operations: A revised perspective. *Progress in Nuclear Energy*, 157:104559, 2023.

[30] R. Michael Churchill, Jong Choi, Ralph Kube, C. S. Chang, and Scott Klasky. Machine learning for the complex, multi-scale datasets in fusion energy. In Jeffrey Nichols, Becky Verastegui, Arthur Barney Maccabe, Oscar Hernandez, Suzanne Parete-Koon, and Theresa Ahearn, editors, *Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI*, pages 269–284. Springer International Publishing, 2020.

[31] Jeongyun Kim, Myung-Hwan Jeon, Sangwoo Jung, Wooseong Yang, Minwoo Jung, Jaeho Shin, and Ayoung Kim. Transpose: Large-scale multispectral dataset for transparent object. *The International Journal of Robotics Research*, 43(6):731–738, 2024.

[32] P. Rodriguez-Fernandez, N.T. Howard, and J. Candy. Nonlinear gyrokinetic predictions of sparc burning plasma profiles enabled by surrogate modeling. *Nuclear Fusion*, 62(7):076036, may 2022.

[33] Juergen Rapp, T. M. Biewer, T. S. Bigelow, J. B. O. Caughman, R. C. Duckworth, R. J. Ellis, D. R. Giuliano, R. H. Goulding, D. L. Hillis, R. H. Howard, T. L. Lessard, J. D. Lore, A. Lumsdaine, E. J. Martin, W. D. McGinnis, S. J. Meitner, L. W. Owen, H. B. Ray, G. C. Shaw, and V. K. Varma. The development of the material plasma exposure experiment. *IEEE Transactions on Plasma Science*, 44(12):3456–3464, 2016.

[34] Juergen Rapp, Matthew J. Baldwin, Clyde J. Beers, Timothy S. Bigelow, John B. O. Caughman, Richard H. Goulding, Atul Kumar, and Cornwall Lau. Research and development to reduce impurity production and transport of the impurities to the target in linear plasma devices using helicon plasma sources. *IEEE Transactions on Plasma Science*, 52(9):3885–3891, 2024.

[35] A. Kumar, J. F. Caneses-Marin, J. Rapp, C. Lau, and R. Goulding. Density drop at the divertor target in the prototype material plasma exposure experiment (proto-mpex). *Physics of Plasmas*, 31(12), 2024.

[36] A. Kumar, J. F. Caneses-Marin, C. Lau, and R. Goulding. Parallel transport modeling of linear divertor simulators with fundamental ion cyclotron heating. *Nuclear Fusion*, 63(3), 2023.

[37] M. S. Islam, J. D. Lore, A. Kumar, G. Dhamale, M. J. Baldwin, D. Nishijima, G. R. Tynan, and J. Rapp. Simulation of plasma and neutral transport in pisces-rf using solps-iter*. *Plasma Physics and Controlled Fusion*, 67(2), 2024.

[38] C. J. Beers, D. L. Green, C. Lau, J. R. Myra, J. Rapp, T. R. Younkin, and S. J. Zinkle. Rf sheath induced sputtering on proto-mpex. i. sheath equivalent dielectric layer for modeling the rf sheath. *Physics of Plasmas*, 28(9), 2021.

[39] C. J. Beers, C. Lau, J. Rapp, T. R. Younkin, T. M. Biewer, T. Bigelow, J. F. Caneses, J. B. O. Caughman, D. L. Green, H. Meyer, J. R. Myra, and S. J. Zinkle. Rf sheath induced sputtering on proto-mpex part 2: Impurity transport modeling and experimental comparison. *Physics of Plasmas*, 28(10), 2021.

[40] T. R. Younkin, D. L. Green, A. B. Simpson, and B. D. Wirth. Gitr: An accelerated global scale particle tracking code for wall material erosion and redistribution in fusion relevant plasma–material interactions. *Computer Physics Communications*, 264, 2021.

[41] G. Dhamale, M. J. Baldwin, M. S. Islam, A. Kumar, H. M. Meyer, D. Nishijima, L. Nuckols, M. I. Patino, W. Tierens, G. R. Tynan, and J. Rapp. Impurity transport in pisces-rf*. *Plasma Physics and Controlled Fusion*, 66(9), 2024.

[42] C. Lau, E. F. Jaeger, N. Bertelli, L. A. Berry, D. L. Green, M. Murakami, J. M. Park, R. I. Pinsker, and R. Prater. Aorsa full wave calculations of helicon waves in diii-d and iter. *Nuclear Fusion*, 58(6), 2018.

[43] N. Kafle, L. W. Owen, J. F. Caneses, T. M. Biewer, J. B. O. Caughman, D. C. Donovan, R. H. Goulding, and J. Rapp. Plasma flow measurements in the prototype-material plasma exposure experiment (proto-mpex) and comparison with b2.5-eirene modeling. *Physics of Plasmas*, 25(5), 2018.

[44] Tara M Pandya, Seth R Johnson, Gregory G Davidson, Thomas M Evans, and Steven P Hamilton. Shift: a massively parallel monte carlo radiation transport package. In *Mathematics and Computations, Supercomputing in Nuclear Applications and Monte Carlo International Conference, M and C+ SNA+ MC 2015*, pages 1472–1485. American Nuclear Society, 2015.

[45] Casey Icenhour, Shane Keniley, Corey DeChant, Cody Permann, Alex Lindsay, Richard Martineau, Davide Curreli, and Steven Shannon. Multi-physics object oriented simulation environment (moose). Technical report, Idaho National Lab.(INL), Idaho Falls, ID (United States), 2018.

[46] Tara M. Pandya, Seth R. Johnson, Thomas M. Evans, Gregory G. Davidson, Steven P. Hamilton, and Andrew T. Godfrey. Implementation, capabilities, and benchmarking of Shift, a massively parallel Monte Carlo radiation transport code. *Journal of Computational Physics*, 308:239–272, March 2016.

[47] Reactor physics benchmark models. `https://github.com/mit-crpg/benchmarks`. Accessed: 2025-05-06.

[48] K. Sridharan and T.R. Allen. 12 - corrosion in molten salts. In Frédéric Lantelme and Henri Groult, editors, *Molten Salts Chemistry*, pages 241–267. Elsevier, Oxford, 2013.

[49] Rajni Chahal, Santanu Roy, Martin Brehm, Shubhojit Banerjee, Vyacheslav Bryantsev, and Stephen T. Lam. Transferable deep learning potential reveals intermediate-range ordering effects in lif–naf–zrf4 molten salt. *JACS Au*, 2(12):2693–2702, 2022.

[50] Luke D. Gibson, Santanu Roy, Rabi Khanal, Rajni Chahal, Ada Sedova, and Vyacheslav S. Bryantsev. Tracing mechanistic pathways and reaction kinetics toward equilibrium in reactive molten salts. *Chem. Sci.*, 15:3116–3129, 2024.

[51] Kasey Hanson, Krishna Moorthi Sankar, Philippe F. Weck, Jacob K. Startt, Remi Dingreville, Chaitanya S. Deo, Joshua D. Sugar, and Preet M. Singh. Effect of excess mg to control corrosion in molten mgcl2 and kcl eutectic salt mixture. *Corrosion Science*, 194:109914, 2022.

[52] Hyo On Nam and Dane Morgan. Redox condition in molten salts and solute behavior: A first-principles molecular dynamics study. *Journal of Nuclear Materials*, 465:224–235, 2015.

[53] Luke D. Gibson, Rajni Chahal, and Vyacheslav S. Bryantsev. Computing chemical potentials with machine-learning-accelerated simulations to accurately predict thermodynamic properties of molten salts. *Chem. Sci.*, 16:3078–3091, 2025.

[54] Mathieu Salanne, Christian Simon, Pierre Turq, and Paul A. Madden. Calculation of activities of ions in molten salts with potential application to the pyroprocessing of nuclear waste. *The Journal of Physical Chemistry B*, 112(4):1177–1183, 2008. PMID: 18177029.

[55] W B Cottrell, H E Hungerford, J K Leslie, and J L Meem. Operation of the aircraft reactor experiment. Technical report, Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States), 09 1955.

[56] M. A. Bredig. Mixtures of metals with molten salts. Technical report, Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States), 08 1963.

[57] E. A. Ukshe and N. G. Bukun. The Dissolution of Metals in Fused Halides. *Russian Chemical Reviews*, 30(2):90–107, February 1961.

[58] L. Ai, V. Soltangharaei, and P. Ziehl. Evaluation of asr in concrete using acoustic emission and deep learning. *Nuclear Engineering and Design*, 380:111328, 2021.

[59] G. K. Kocur, E. H. Saenger, C. U. Grosse, and T. Vogel. Time reverse modeling of acoustic emissions in a reinforced concrete beam. *Ultrasonics*, 65:96–104, 2016.

[60] K. G. Field, I. Remec, and Y. Le Pape. Radiation effects in concrete for nuclear power plants—part i: Quantification of radiation exposure and radiation effects. *Nuclear Engineering and Design*, 282:126–143, 2015.

[61] Harleen Kaur Sandhu, Saran S. Bodda, and Abhinav Gupta. A future with machine learning: Review of condition assessment of structures and mechanical systems in nuclear facilities. *Energies*, 16(6):2628, 2023.

[62] G. G. Aliyeva, Y. Le Pape, A. Gupta, and S. Samaddar. Modeling concrete expansion due to alkali-silica reaction. In *Proc. 26th Int. Conf. on Structural Mechanics in Reactor Technology (SMiRT-26)*, Berlin/Potsdam, Germany, 2022.

[63] V. Soltangharaei, R. Anay, L. Ai, E. R. Giannini, J. Zhu, and P. Ziehl. Temporal evaluation of asr cracking in concrete specimens using acoustic emission. *Journal of Materials in Civil Engineering*, 32(10):04020285, 2020.

[64] W. Zhang, H. Liu, Y. Zhou, K. Liao, and Y. Huang. Effects of neutron irradiation on densities and elastic properties of aggregate-forming minerals in concrete. *Nuclear Engineering and Technology*, 55(6):2147–2157, 2023.

[65] Á. Carrasco, F. Méndez, F. Leaman, and C. Molina Vicuña. Short review of the use of acoustic emissions for detection and monitoring of cracks. *Acoustics Australia*, 49:273–280, 2021.

[66] Y. Le Pape, K. G. Field, and I. Remec. Radiation effects in concrete for nuclear power plants—part ii: Perspective from micromechanical modeling. *Nuclear Engineering and Design*, 282:144–156, 2015.

[67] Bozhou Zhuang, Anna Arcaro, Bora Gencturk, Ryan M. Meyer, Assad Oberai, Anton Sinkov, and Morris Good. Non-destructive evaluation and machine learning methods for inspection of spent nuclear fuel canisters: A state-of-the-art review. *Progress in Nuclear Energy*, 185:105697, 2025.

[68] R. Pillai, S.S. Raiman, and B.A. Pint. First steps toward predicting corrosion behavior of structural materials in molten salts. *Journal of Nuclear Materials*, 546:152755, 2021.

[69] M. Romedenne, R. Pillai, B. Harris, and B.A. Pint. Compatibility of fe-cr-al and fe-cr-al-mo oxide dispersion strengthened steels with static liquid sodium at 700 °c. *Journal of Nuclear Materials*, 569:153919, 2022.

[70] M. Romedenne and B. A. Pint. Corrosion in sodium fast reactors. Technical report, Nuclear Regulatory Commission, 2021.

[71] R. Pillai and W. Reed. Preliminary assessment of models for generating predictions of long-term corrosion in molten salts. Technical report, Nuclear Regulatory Commission, 2024.

[72] Richard P. Feynman. There's plenty of room at the bottom. 1959.

[73] S.J. Zinkle and G.S. Was. Materials challenges in nuclear energy. *Acta Materialia*, 61(3):735–758, 2013. The Diamond Jubilee Issue.

[74] Y. R. Lin, L. G. Chen, C. Y. Hsieh, M. T. Chang, K. Y. Fung, A. Hu, S. C. Lo, F. R. Chen, and J. J. Kai. Atomic configuration of point defect clusters in ion-irradiated silicon carbide. *Scientific Reports*, 7(1):14635, Nov 2017.

[75] Yan-Ru Lin, Yajie Zhao, Michael J Zachman, Jose' D Arregui-Mena, M Grace Burke, and Steven J Zinkle. Understanding formation of irradiation-induced defects through 4d-stem, electron tomography, and wbdf-stem. *Microscopy and Microanalysis*, 30(Supplement_1):ozae044.122, 07 2024.

[76] Zhiheng Cheng, Chaolun Wang, Xing Wu, and Junhao Chu. Review in situ transmission electron microscope with machine learning. *Journal of Semiconductors*, 43(8):081001, aug 2022.

[77] Kyle A. Gamble, Richard L. Williamson, Christopher Newman, Yaqi Wang, Javier Ortensi, and Shane Stimpson. Vera–bison coupling for multi-physics fuel performance analysis. In *Proceedings of the International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (M&C 2019)*, Portland, Oregon, USA, August 2019.

[78] Maria Avramova, Agustin Abarca, Jason Hou, and Kostadin Ivanov. Innovations in multi-physics methods development, validation, and uncertainty quantification. *Journal of Nuclear Engineering*, 2(1):44–56, 2021.

[79] Yiqi Yu, April Novak, Dillon Shaver, and Elia Merzari. Coupled simulation of reactor pressure vessel (rpv) subjected to pressurized thermal shock (pts) using cardinal. In *Proceedings of Advances in Thermal Hydraulics, ATH 2022 - Embedded with the 2022 ANS Annual Meeting*, pages 758–770. American Nuclear Society, 2022.

[80] Alejandro Nuñez-Carrera, Ana Lidia Carreño-Padilla, Erick Gilberto Espinosa-Martínez, and Raúl Camargo-Camargo. Analysis of power uprate in nuclear power plants. *Energy Research Journal*, 8(1):1–10, 2017.

[81] Ben Spencer, Jeremey Busby, Richard Martineau, and Brian Wirth. A proof of concept: Grizzly, the lwrs program materials aging and degradation pathway main simulation tool. Technical Report INL/EXT-12-27559, Idaho National Laboratory, October 2012.

[82] William A. Wieselquist and Robert Alexander Lefebvre. Scale 6.3.2 user manual. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 02 2024.

[83] Robert Salko Jr., Ana Gomez, Caleb Jernigan, et al. Ctf user's manual (v.4.2). Technical Report ORNL/SPR-2020/1234, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2020. This document describes how a user should go about using the CTF pre-processor tool to create an input deck for modeling rod-bundle geometry in CTF.

[84] S B. Harris, A. Biswas, S J Yun, K M. Roccapriore, C M. Rouleau, A A. Puretzky, R K. Vasudevan, D B. Geohegan, and K. Xiao. Autonomous synthesis of thin film materials with pulsed laser deposition enabled by in situ spectroscopy and automation. *Small Methods*, 8(9):2301763, 2024.

[85] A. Biswas, M. Ziatdinov, and S V. Kalinin. Combining variational autoencoders and physical bias for improved microscopy data analysis. *Machine Learning: Science and Technology*, 4(4):045004, 2023.

[86] A. Biswas, Y. Liu, N. Creange, Y C. Liu, S. Jesse, J C. Yang, S V. Kalinin, M. Ziatdinov, and R K. Vasudevan. A dynamic bayesian optimized active recommender system for curiosity-driven partially human-in-the-loop automated experiments. *npj Comput Mater*, 10(29), 2024.

[87] Sergei V Kalinin, Yongtao Liu, Arpan Biswas, Gerd Duscher, Utkarsh Pratiush, Kevin Roccapriore, Maxim A Ziatdinov, and Rama Vasudevan. Human-in-the-loop: The future of machine learning in automated electron microscopy. *Microscopy Today*, 32(1):35–41, January 2024.

[88] Carlos Alberto Gonzalez Rodriguez Sr. and et. al. Evaluation of the degree of sensitization of a stainless steel 304 using local electrochemical impedance spectroscopy. *ECS Transactions*, 47(123), 2013.

[89] W. Guo, N. Ding, L. Liu, N. Xu, N. Li, F. Zhang, and L. Chen. Stress corrosion cracking of a 304 stainless steel elbow. *Journal of Failure Analysis and Prevention*, 20:483–493, 2020.

[90] P. Kumar, R K. Soni, and B S. Dehiya. Gamma radiation effects on pre- and post-welding ss304 and ss316 stainless steels' ductility and hardness. In *Govindan, K., Kumar, H., Yadav, S. (eds) Advances in Mechanical and Materials Technology . EMSME 2020. Lecture Notes in Mechanical Engineering. Springer, Singapore.*, 2022.

[91] N. S. V. Rao, P. Ramuhalli, C. Greulich, and S. M. Cetiner. Sensor error estimation for reactor coolant system with generalization error equations. *Transactions of the American Nuclear Society*, 121(1), 11 2019.

[92] Christopher Greulich, Mathew Swinney, Will Ray, Daniel Archer, Mark S Bandstra, M Salathe, and Brian Quiter. A method for semi-automated point cloud to watertight mesh conversion for rapid mcnp input file generation of indoor scenes. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2022.

[93] N. S. V. Rao, C. Greulich, P.amuhalli, A. Gurgen, F. Zhang, and S. M. Cetiner. Estimation of sensor measurement errors in rector coolant systems using multi-sensor fusion. *Nuclear Engineering Design*, 2020. in press.

[94] Jianping Ma and Jin Jiang. Applications of fault detection and diagnosis methods in nuclear power plants: A review. *Progress in Nuclear Energy*, 53(3):255–266, 2011.