Pacific Northwest NATIONAL LABORATORY

MRP Microreactor Program

PNNL-39277

Draft Feasibility Assessment for Use of AI in Preparing Transportation Safety Analysis Reports

April 2026

Dina E. Carpenter-Graffy
Steven J. Maheras

U.S. DEPARTMENT of ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-39277

Draft Feasibility Assessment for Use of AI in Preparing Transportation Safety Analysis Reports

April 2026

Dina E. Carpenter-Graffy
Steven J. Maheras

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Summary

Preparing transportation safety analysis reports for microreactors is time- and labor-intensive, requiring extensive cross-referencing to Federal regulations, previously approved documents, and expert review comments across structural, thermal, criticality, shielding, and containment. These burdens are magnified by the novelty of microreactor technologies and the evolving regulatory landscape, as well as funding and workforce constraints that may limit available person hours. Generative AI and supporting machine learning tools present an opportunity to accelerate drafting timelines, lift generalized writing burdens, and systematically enforce regulatory adherence through retrieval-augmented generation and other knowledge retrieval and mapping methods.

This draft report presents a preliminary feasibility assessment of the use of AI to expedite the preparation of microreactor transportation safety analysis reports and proposes an initial methodology for doing so.

Acknowledgments

Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract No. DE-AC05-76RL01830. This work was supported by DOE's Microreactor Program. The authors utilized Microsoft Copilot as part of the development of this report.

Acronyms and Abbreviations

AI	artificial intelligence
AiRKS	AI-Enabled Advanced Reactor Knowledge System
BERT	Bidirectional Encoder Representations from Transformers
GPT	Generative Pre-Trained Transformer
LLM	large language model
ML	machine learning
PINN	physics-informed neural network
RAG	retrieval-augmented generation
SAR	safety analysis report
SARP	safety analysis report for packaging
SME	subject matter expert

Contents

1.0 Introduction

Preparing transportation safety analysis reports (SARs) for transportation packages is time- and labor-intensive, requiring extensive cross-referencing to Federal regulations, previously approved documents, and expert review comments across structural, thermal, criticality, shielding, and containment. These burdens are magnified by the novelty of microreactor technologies and the evolving regulatory landscape, as well as funding and workforce constraints that may limit available person hours. Generative AI (GenAI) and supporting machine learning (ML) tools present an opportunity to accelerate drafting timelines, lift generalized writing burdens, and systematically enforce regulatory adherence through retrieval-augmented generation (RAG) and other knowledge retrieval and mapping methods.

Artificial intelligence (AI) is relevant because it can reduce cycle time on standard sections, improve consistency in cross-references, and assist with regulatory knowledge management. GenAI and complementary ML tools can automate boilerplate, retrieve applicable regulations on demand, and maintain explicit links to supporting documents, including official guidance, prior SARs, and resolved review comments. When embedded in a human-in-the-loop workflow with audit trails and approval gates, these capabilities have the potential to improve process timelines without sacrificing safety or quality.

This draft feasibility assessment presents an initial investigation into the use of AI to expedite the preparation of microreactor transportation SARs, specifically focusing on use cases that can leverage Frontier AI models such as those offered by Anthropic, Google, and OpenAI. In addition, to the extent feasible, methodologies proposed in this work would be aligned with other AI-based tools in the nuclear space, such as the AI-Enabled Advanced Reactor Knowledge System (AiRKS) project.

This page was intentionally left blank.

2.0 AI-Enabled SAR Workflow

GenAI systems learn statistical patterns across large corpora of text, code, and structured data to produce coherent, context-sensitive output. In safety and licensing contexts, their utility is maximized when model generation is grounded in domain-specific sources, configured with low-temperature decoding[1] and template-based prompts to minimize variability, and embedded in human-in-the-loop workflows that enforce review and approval gates. Complementary ML methods including surrogate models and physics-informed neural networks (PINNs) can approximate high-fidelity simulations for screening analyses, enabling rapid identification of highest-priority scenarios for computationally intensive, full fidelity modeling. Together, these techniques support a SAR drafting workflow in which AI automates drafting of standard sections, flags gaps and inconsistencies, and continuously maps assumptions and parameters to applicable regulatory criteria.

The main SAR workflow tasks that may benefit from AI integration can be grouped into Writing and Editing, Knowledge Ontology[2], and Physical Surrogate Models. While physical surrogate models are not GenAI, they utilize AI methods and were therefore included. This section evaluates scope and preliminary feasibility for each.

2.1 Writing and Editing

2.1.1 Scope of Potential Work

A practical starting point is to use a SARP[3] (safety analysis report for packaging) checklist (DOE, 2007) or baseline template as a document atlas, within which AI can contribute to section-by-section drafting and validation. Nonspecific or commonly shared text, such as standard definitions, packaging descriptions, or methodological overviews can be autopopulated, with cross-references inserted and kept current as the document iterates. Informational fields can be keyed to microreactor specifics (e.g., transport configurations, shielding assumptions, and dose assessment methods) while maintaining explicit links to source regulations and guidance. In practice, frontier large language models (LLMs) can deliver reliable first-pass text when grounded by outlines, checklists, and controlled vocabularies; adding RAG against authoritative corpora (prior SARs, regulatory guidance, and prior review feedback) improves accuracy even in technical or domain-specific areas.

Criteria-based review agents configured against expert-defined rules could also be employed to score drafted sections, present findings with sources and response confidence levels, and recommend remediation steps.

2.1.2 Feasibility

Frontier LLMs can be effective at producing first-pass prose when they are provided clear structure, such as SAR outlines, controlled glossaries, and checklists; their outputs become materially stronger when the models are trained on and connected to curated document atlases that include domain-specific resources. A human-in-the-loop design remains essential: subject matter experts (SMEs) validate interpretive content, confirm regulatory applicability, and sign off on any AI-assisted text.

Some researchers have found that regulatory writing can confuse LLMs by appearing semantically similar such that the models lose crucial context or correct cross-references (Viswanathan, 2026). Similarly, linguistic heterogeneity in legal documents can make conventional RAG unreliable, which indicates that frontier models may require domain-specific training in addition to data library access (Buster et al., 2024).

Strictly focusing on authorship quality, researchers report that AI-authored technical writing can be overly verbose, insufficiently specific, unclear, or redundant (Eser et al., 2025). It also may struggle with completeness, consistency or standardization, and excess detail in methodology while providing insufficient detail in results.

---

[1] “Temperature” in LLMs refers to the level of statistical confidence it uses to predict text and generate responses—lower temperatures are more deterministic, less creative, and more predictable, whereas high-temperature decoding will demonstrate more randomness. Low temperatures are preferred in settings where source or factual accuracy is a priority.

[2] An ontology is a structured knowledge framework that captures concepts, properties, and relationships between entities. It is a common information structure for AI applications.

[3] A DOE SARP typically follows NRC Regulatory Guide 7.9 for the first eight chapters (General Info, Structural, Thermal, Containment, Shielding, Criticality, Operating Procedures, Acceptance Tests), but adds a 9th chapter specifically for Quality Assurance.

Cumulatively, this indicates that AI tools can provide significant assistance in SAR development, particularly as it relates to first-draft development timelines, reference density, and general editing assistance such as internal consistency-checking. However, research finds that LLMs frequently struggle with highly technical or domain-specific content without additional refinement and training and can require extremely detailed management of prompts and reasoning approaches (Kim et al., 2025). To protect against issues like incorrect information, vague writing, and incomplete source references, model outputs must be constrained via methods like templates, low-temperature decoding, and review gates that ensure clarity, specificity, and completeness before publication.

2.2 Knowledge Ontology

2.2.1 Scope of Potential Work

A queryable nuclear transportation/microreactor ontology can link components, materials, parameters, hazards, and controls to regulatory criteria and precedents, forming the backbone of AI-assisted development. Building this ontology with domain-specific terminology and relationships from sources such as prior review feedback, expert notes, regulatory logic, and previously approved documents can enable automated gap analysis, consistency checks, and an accessible knowledge resource that captures previous work. International frameworks can be included to support cross-jurisdictional considerations.

Compliance benefits most directly from managed retrieval against curated regulatory libraries. A well-designed RAG pipeline should segment source texts by paragraph, sentence, or even clause boundaries, tag authoritative lineage and version, and map ontology fields (component/material/parameter) to specific regulation sections. With this structure, AI can automatically retrieve applicable parts and sections for cited claims, propose citations with confidence scores, and check thermal, structural, or criticality model parameters against regulatory limits. Human reviewers would then confirm applicability and finalize language, maintaining a defensible chain of custody for every regulatory reference.

2.2.2 Feasibility

The SAR workflow can integrate curated regulatory libraries and guidance such as for shipments containing irradiated fuel. A structured interface can be encoded as ontology relationships and checklist validations so that the system flags missing linkages or misapplied criteria. Knowledge-graph tooling and modern vector retrieval make it feasible to assemble a SAR-centric ontology, though it would require ongoing curation and document security controls, particularly if other SARs are used in the knowledge pool.

Out-of-the-box frontier models often lack deep domain understanding in nuclear applications and can struggle on scientific topics without strong grounding and guardrails. Auditability and traceability are essential in regulated environments; raw LLM outputs must be tied back to authoritative sources or acceptance by regulators is unlikely. Additionally, many SAR precedents and expert notes are sensitive, limiting direct ingestion into models and reinforcing the need for secure retrieval rather than wholesale training on controlled documents.

Researchers have found that RAG is at best “partially correct” when applied to National Environmental Policy Act documentation retrieval (Phan et al., 2025). Training can significantly improve domain-specific retrieval accuracy over RAG alone (Buster et al., 2024), as do query approaches that develop a multi-stage methodology, such as fine-tuning a GPT model with domain-specific information and combining RAG with a multiquark approach (Kim et al., 2025). In general, researchers indicate that adapting models to domain-specific knowledge and using several layers of interpretation and retrieval (such as a multiquery approach or domain-specific Bidirectional Encoder Representations from Transformers [BERT]) significantly improve response accuracy in regulatory and/or highly technical topics (Han et al., 2026).

2.3 Physical Surrogate Models

2.3.1 Scope of Potential Work

Surrogate models are compact emulators trained on high-fidelity analysis or test data which can accelerate SAR-relevant screening across thermal, structural, and criticality domains, and help identify scenarios that warrant full-fidelity simulations. While physical models do not incorporate GenAI, the proposed scope does incorporate AI methods, such as ML and PINNs, and was therefore included in the feasibility assessment. Thermal emulators can estimate package/reactor heat-up, cool-down, and transient ambient extremes, and support rapid sweeps across transport configurations and environmental conditions. Structural response surrogates

can approximate shock, vibration, and handling loads across rail, road, and port operations to guide envelope checks and selection of priority cases for full-fidelity analysis. Criticality surrogates that approximate k_{eff} across material compositions, geometries, and moderation/poisoning cases enable prescreening against bounding criteria prior to detailed Monte Carlo studies. Coupled-domain emulators (thermal–structural–criticality) can propagate sensitivities to reveal cross-coupled edge conditions, while digital-twin constructs support continuous scenario exploration. Where data are sparse or extrapolation risk is high, PINNs embed physical laws and relations into the fundamental function of a neural network structure to improve result quality, especially in cases where training data may be limited. Full-fidelity simulations and test data would be utilized for the full evaluations necessary for the SAR.

2.3.2 Feasibility

Surrogate modeling could be feasible within the SAR workflow when confined to clearly defined purposes and governed by robust uncertainty quantification plans. Unique microreactor transport elements, such as unconventional designs, compensatory measures, or strategies for prevention of criticality, benefit from surrogate-enabled triage so SMEs can focus full-fidelity effort where novelty or risk is highest. Using surrogates for rapid thermal/structural/criticality screening could align with iterative SAR drafting cycles and pre-submission refinement. Established uncertainty quantification practices such as error bounds, uncertainty propagation, and coverage metrics, are directly applicable, ensuring that surrogate outputs include defendable limits and confidence measures suitable for licensing-adjacent decision-making.

Other agencies and programs are increasingly deploying custom, task-specific AI, but these tools are frequently developed in-house and may not be able to leverage frontier AI tools (Karanicolas, 2024). For this application, where the basis of the AI tool is highly mathematical or extremely dependent on subject-matter-specific training data, in-house development may be the only feasible option.

This page was intentionally left blank.

3.0 Approach and Implementation

3.1 Potential Workflow

A feasible proposed workflow for SAR development would include the following:

- query templating,
- LLM encoding/decoding,
- vector search structures
- managed RAG via a knowledge-base service
- a curated document atlas (prior SARs, regulatory guidance, and review comments), and
- section drafting and consistency checks.

To support this pattern, the tool would establish data cleanup and structuring methods that condense and organize the corpus of regulations, existing SARs, and related documents; if the model is being fine-tuned, training data must be labeled to support reliable retrieval and evaluation. A complementary set of query rules would guide how users formulate requests and how the system constrains responses. Domain-specific fine-tuning of a publicly available, general-purpose model would reduce training overhead relative to a custom-developed model by focusing on nuclear transportation specificity. A domain-adapted BERT retriever or similar approach could interface between user queries and the corpus, while semantic analysis resolves the correct regulation and clause in response to a query. A frontier LLM would then synthesize grounded answers, with low-temperature, template-bound decoding to ensure auditability and minimize the risk of hallucination.

3.2 Limitations of Frontier Models

While frontier AI models are well-suited for several steps in this process, they also come with potentially significant drawbacks rooted in their generalization. Out-of-the-box frontier models demonstrate generally poor domain-specific understanding of nuclear science (Acharya et al., 2023). Even non-frontier models trained on a knowledge base of science texts can struggle significantly with accuracy on scientific topics (Munikoti et al., 2024). As previously discussed, limitations on data security (e.g., proprietary information) may severely constrain the training corpus and document atlas, which would lead to limited domain specificity and issues with technical accuracy. The total size of a document library may also be a limiting factor, since many frontier models have token limits that preclude the use of large files in training or retrieval structures (Buster et al., 2024). Finally, many LLMs are considered to have “black-box” decision-making processes, meaning that the reasoning behind responses is not available; without sufficient auditability requirements for responses, this could lead to a lack of proper citations, non-auditable methodologies, and unclear analysis.

3.3 Necessary Integrations

Reliable AI-enabled SAR drafting retrieval depends foremost on a curated, authoritative document atlas with provenance tracking and version control. To make content AI-ready, the program will need to implement source tagging and lineage, appropriately-sectioned clause-level chunking aligned to regulatory boundaries, ontology fields that map components and parameters to regulations, and workflows for parsing and reducing documents to manage the overall volume of input data. Robust output instructions and controls, like low-temperature generation, would help ensure that retrieved content is accurate to original sources and minimizes the risk of hallucinated or otherwise incorrect results. Document templates and similar structural tools are also key for setting retrieval targets and limiting opportunities for AI tools to draft sections without sufficient knowledge resources. Human oversight would be crucial regardless of tasks or methodology wherein SME authority would have the final say on citation applicability, content accuracy, any analytical conclusions, and overall SAR correctness.

3.4 Custom Approaches

Beyond frontier models, task-specific non-generative AI could be built in-house to address particular analytical needs. Thermal, structural, and criticality screening could begin with approximations generated by ML models that were trained on real-world and validated full-fidelity simulation data. In data-sparse regimes or where extrapolation risk is high, PINN approaches embed governing equations and constitutive relations into model reasoning. These early surrogate findings can guide prioritization of full-fidelity simulations. However, these

models would not leverage Frontier LLM capability, and so were not identified as a priority area in this analysis.

3.5 Necessary Architecture

All approaches would require human-in-the-loop review so that every AI contribution is examined and signed off by SMEs. A secure, up-to-date document atlas and ontology are necessary for accurate retrieval and grounding. Model outputs will be generated with low temperatures and constrained through templates and deterministic decoding to minimize hallucinations and stylistic drift.

4.0 Preliminary Conclusions

4.1 Benefits and Opportunities

The use of AI to assist in the preparation transportation SARs has the potential to yield shorter drafting timelines by auto-generating standard sections, inserting validated citations, and continuously checking parameter-to-regulation consistency. By alleviating boilerplate work, SMEs can focus on technical judgments and other knowledge-intensive tasks. High-capacity semantic analysis may also identify precedents when microreactor technology lacks fully settled regulation, and automated gap analysis could maintain coherence across assumptions, parameters, and regulations as the document evolves.

4.2 Risks and Barriers

4.2.1 Barriers

Frontier models are limited by corpus size and domain specificity; without robust retrieval and grounding, technical accuracy can suffer. Auditability demands may exceed “black-box” capabilities unless reasoning explanation and information sourcing are integrated from the outset. Information security constraints and total document volume may also limit knowledge-base completeness. Finally, a robust verification and validation process would need to be developed.

4.2.2 Risks

Primary risks include hallucinations of data, simulation results, or regulatory text if grounding is weak, and potential data-security exposures if controlled documents are used. There is also a potential expertise drain if routine knowledge-gathering tasks are over-automated without balancing SME engagement, and a risk of perceived administrative legitimacy loss or elevated safety risks if AI contributions dominate the process (Karanicolas, 2024). Accuracy risks are most pronounced for specialized technical topics; therefore, every model-generated technical claim must undergo SME review.

This page was intentionally left blank.

5.0 References

Acharya, A, S. Munikoti, A. Hellinger, S. Smith, S. Wagle, and S. Horawalavithana. 2023. "Nuclear QA: A Human-Made Benchmark for Language Models for the Nuclear Domain." *arXiv*. https://doi.org/10.48550/arXiv.2310.10920.

Buster, G., P. Pinchuk, J. Barrons, R. McKeever, A. Levine, and A. Lopez. 2024. "Supporting Energy Policy Research with Large Language Models." arXiv. https://doi.org/10.48550/arXiv.2403.12924.

DOE (U.S. Department of Energy). 2007. *SARP Completeness Checklist for EM-60*. Revision 1. U.S. Department of Energy, Washington, D.C.

Eser, U, Y. Gozin, L. J. Stallons, A. Caroline, M. Preusse, B. Rice, S. Wright, and A. Robertson. 2025. "Human-AI Collaboration Increases Efficiency in Regulatory Writing." *arXiv*. https://doi.org/10.48550/arXiv.2509.09738.

Han, I, M. Roh, and M. Kong. 2026. "A RAG-based Q&A System for Ship Regulations Applying Domain Adaptation." *International Journal of Naval Architecture and Ocean Engineering* 18: 100735. https://doi.org/10.1016/j.ijnaoe.2025.100735.

Karanicolas, M. 2024. *Artificial Intelligence and Regulatory Enforcement.* Report to the Administrative Conference of the United States, Washington, D.C.

Kim, J, M. Hur, and M. Min. 2025. "From RAG to QA-RAG: Integrating Generative AI for Pharmaceutical Regulatory Compliance Process" paper presented at Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing (SAC'25). Catania, Italy, March 31–April 4, 2025. https://doi.org/10.1145/3672608.3707749.

Munikoti, S., A. Acharya, S. Wagle, and S. Horawalavithana. 2024. "Evaluating the Effectiveness of Retrieval-Augmented Large Language Models in Scientific Document Reasoning." *arXiv*. https://doi.org/10.48550/arXiv.2311.04348.

Phan, H., A. Acharya, S. Chaturvedi, S. Sharma, M. Parker, D. Nally, A. Jannesari, K. Pazdernik, M. Halappanavar, S. Munikoti, and S. Horawalavithana. 2025. "RAG vs. Long Context: Examining Frontier Large Language Models for Environmental Review Document Comprehension." *arXiv*. https://doi.org/10.48550/arXiv.2407.0732.

Viswanathan, A. 2006. "Measuring Quality in the Age of AI: Why Regulatory Writing Needs a True Standard." Clinical Researcher. Accessed April 24, 2026. https://acrpnet.org/2026/04/14/measuring-quality-in-the-age-of-ai-why-regulatory-writing-needs-a-true-standard.