Pacific Northwest NATIONAL LABORATORY

MRP Microreactor Program

PNNL-39747

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

August 2026

Dina E. Carpenter-Graffy
Steven J. Maheras
Peter Stefanovic
Veronica M. Wilson

U.S. DEPARTMENT of ENERGY

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

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights**. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

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-39747

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

August 2026

Dina E. Carpenter-Graffy
Steven J. Maheras
Peter Stefanovic
Veronica M. Wilson

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

Pacific Northwest National Laboratory
Richland, Washington 99354

Summary

Preparing the transportation safety analysis report for a new package design is on the critical path to microreactor deployment, and the preparation is often time-consuming. The work cross-references Federal regulations, approved precedents, and expert review comments across six technical disciplines, and the experts who can do it are scarce. Generative AI and supporting machine learning tools present an opportunity to shorten that work without weakening its rigor.

This 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.

This assessment covers four areas of the safety analysis report workflow. Writing and editing, a queryable knowledge ontology, physical surrogate models, and compliance review and gap analysis are each evaluated for scope and preliminary feasibility. Preparing and reviewing safety analysis reports involves information retrieval, cross-referencing, and pattern-matching. This is precisely the domain where AI retrieval-augmented generation excels. Compliance review is the most immediately feasible initial task for AI. AI can systematically identify where a candidate safety analysis report for packaging may be missing information, incomplete relative to acceptance criteria, or internally inconsistent. An AI review can flag those gaps for expert review with citation-level traceability to specific regulatory requirements. The physical surrogate models would join the compliance review as callable checks of the reported values against the physics of the described package.

What would make the approach work is not the model but the systematic encoding of regulatory review expertise into the workflow. Feasibility rests on grounding in a curated corpus, on constrained generation, and on subject matter expert review of every AI-assisted technical claim. The report names the barriers and risks, from domain-specific accuracy to data security, and each shares the same mitigation: no finding and no drafted text reaches the safety analysis report without an expert's decision.

A bounded pilot is proposed to test the approach on one drafted chapter and one package type, with four measures and a go or no-go decision. Either outcome keeps the curated corpus, the requirements traceability matrix, and the encoded review logic, which are platform-agnostic and outlast any particular model.

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 AI assistance as part of the development of this report. The authors reviewed and approved all AI-assisted content.

Acronyms and Abbreviations

AI	artificial intelligence
AiRKS	AI-Enabled Advanced Reactor Knowledge System
BERT	Bidirectional Encoder Representations from Transformers
GenAI	generative artificial intelligence
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

Summary ii
Acknowledgments iii
Acronyms and Abbreviations iv
1.0 Introduction 1
2.0 AI-Enabled SAR Workflow 3
2.1 Writing and Editing 3
2.1.1 Section Drafting and Validation 3
2.1.2 Feasibility of AI-Assisted Drafting 4
2.2 Knowledge Ontology 4
2.2.1 Ontology and Traceability Structures 4
2.2.2 Feasibility of the Ontology and Retrieval 5
2.3 Physical Surrogate Models 6
2.3.1 Technical Discipline (e.g. Thermal, Structural, and Criticality) Consistency Review 6
2.3.2 Feasibility of Surrogate Modeling 7
2.4 Compliance Review and Gap Analysis 7
2.4.1 Acceptance Criteria and Review Depths 7
2.4.2 Feasibility of the Compliance Check 8
3.0 Approach and Implementation 9
3.1 Proposed Workflow 9
3.2 Information Corpus 10
3.3 Limitations of Frontier Models 11
3.4 Data and Retrieval Integrations 11
3.5 Custom Model Development 12
3.6 Architecture and Controls 12
4.0 Conclusions 13
4.1 Benefits and Opportunities 13
4.2 Barriers and Risks 13
4.2.1 Barriers 13
4.2.2 Risks 13
4.3 Proposed Pilot 14
5.0 References 17

Figures

Figure 1. The retrieval foundation. Example published documents are chunked to clause boundaries and tagged at ingest. The tagged corpus and the requirements traceability matrix form the index the compliance check retrieves against. 6

Figure 2. The compliance review loop. A subject matter expert accepts or rejects each finding before the chapter changes. The same chapter reviewed by hand sets the baseline the tool is measured against. 10

Figure 3. The four pilot measures with illustrative thresholds, and the decision the measures feed. The real thresholds are set with the pilot design. Either outcome keeps the corpus and the encoded review logic. 15

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, materials and containment disciplines. 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.

Transportation packages are certified against 10 CFR Part 71, which sets the packaging and transportation requirements for radioactive material (NRC, 2024). The shipments themselves fall under the U.S. Department of Transportation hazardous materials regulations at 49 CFR Parts 100 through 185 (DOT, 2024). NRC Regulatory Guide 7.9 sets the standard format and content of a Part 71 application (NRC, 2005). NUREG-2216 consolidates the acceptance criteria NRC staff apply when reviewing a Part 71 application (NRC, 2020). These four documents define the structure a SAR follows and the criteria its completeness is judged against. Any retrieval or citation-checking method proposed in this report would work over that same body of information. The use of “SAR” and “SARP” can be used interchangeably in this context as SAR names the safety analysis report in general, and SARP names the safety analysis report for packaging that supports a Part 71 application. SARP is the terminology more commonly used in DOE packages, while SAR is more commonly used in commercial packages.

A microreactor cannot reach its destination without an approved, e.g. certified, transportation package. Transportation package certification is directly on the critical path to reactor deployment. Licensing challenges in reactor module transport are foreseen, and new assessment approaches will be needed (INL, 2019). A new package design will need its own SARP, so the preparation work this report examines is on the same critical path.

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 sections of the report, 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 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], Physical Surrogate Models, and Compliance Review and Gap Analysis. While physical surrogate models are not GenAI, they utilize AI methods and were therefore included. This section evaluates scope and preliminary feasibility for each. The order is cumulative. Compliance review comes last because the review draws on the drafting structure, the corpus, and the surrogate models that the earlier sections describe. The assessment finds compliance review the most immediately feasible of the four.

2.1 Writing and Editing

2.1.1 Section Drafting and Validation

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[4], within which AI can contribute to section-by-section drafting and validation. The DOE checklist already pairs every requirement with the SARP section that must satisfy it, which is the paper form of the traceability structure Section 2.2.1 proposes. 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

[1] “Temperature” in LLMs is a setting that controls how randomly the model chooses each next word. 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.

[4] A document atlas organizes a set of documents by their structure, so drafting and retrieval can be pointed at a specific section rather than at the whole set. The DOE checklist is the atlas for the SARP being drafted, and the curated corpus in Section 3.1 is the atlas for retrieval.

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 of AI-Assisted Drafting

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.

Regulatory writing also resists automated quality measurement (Viswanathan, 2026). Two passages can differ in wording and both be correct. A pair that reads alike may still diverge from the source data. Judging an AI-assisted draft therefore needs measures anchored in the source documents and the regulatory template rather than surface similarity. Linguistic heterogeneity in legal documents can make conventional RAG unreliable (Buster et al., 2024), which indicates that frontier models may require domain-specific training in addition to data library access (Kim et al., 2025).

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.

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 Ontology and Traceability Structures

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.

Two structures turn that mapping into something a tool can use. A requirements traceability matrix pairs each acceptance criterion with the SARP section that has to satisfy it, which gives drafting and review a single place to look. A metadata scheme tags every ingested passage by regulatory source, applicable SARP chapter, technical discipline, package type, and condition type. Tagging at ingest is what lets retrieval be narrowed before the model sees anything, rather than filtered after the fact.

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 technical model parameters, such as thermal, structural, or criticality, 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 of the Ontology and Retrieval

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.

A first corpus can be built from published material alone, which avoids much of that constraint. The regulations and the review guidance cited in this report are public documents. Controlled documents can be added later under access controls, once the value of adding them justifies the handling burden. Figure 1 shows the resulting foundation, from the public corpus to the traceability matrix in Section 2.2.1.

Researchers testing frontier models on National Environmental Policy Act documents found that retrieval-augmented models answer more accurately than models given the whole document as long context, and that the result held for every frontier model tested (Phan et al., 2024). Applying domain adaptation to a general-purpose model can improve retrieval accuracy in a specialized regulatory corpus (Han et al., 2025), as do query approaches that develop a multi-stage methodology, such as fine-tuning a GPT model with domain-specific information and retrieving on both the user question and a hypothetical answer generated by that fine-tuned model (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 domain-specific Bidirectional Encoder Representations from Transformers [BERT]) significantly improve response accuracy in regulatory and/or highly technical topics (Han et al., 2025).



Figure 1. The retrieval foundation. Example published documents are chunked to clause boundaries and tagged at ingest. The tagged corpus and the requirements traceability matrix form the index the compliance check retrieves against.

2.3 Physical Surrogate Models

2.3.1 Technical Discipline (e.g. Thermal, Structural, and Criticality) Consistency Review

Surrogate models are compact emulators trained on high-fidelity analysis or test data which can accelerate SAR-relevant screening across technical disciplines such as thermal, structural, and criticality, and help identify scenarios that warrant full-fidelity simulations. 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 the effective neutron multiplication factor 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[1] 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.

[1] A digital twin is a model of a specific physical asset that is kept current with data from that asset, so the model can be run against the conditions the asset is expected to meet.

2.3.2 Feasibility of Surrogate Modeling

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.

AI trained surrogate models have not currently been approved for use within radioactive and fissile material transportation SARP analyses approved by the NRC. Use of surrogate models at this stage would likely be confined to preparation work used to help analysts focus efforts to identify limiting conditions or to perform reviews of calculated values as discussed in Section 2.4. All licensing calculations within the SAR will ultimately be performed with codes and models developed using software and methodologies based on physics principles and found acceptable by the NRC.

Other agencies and programs are increasingly deploying custom, task-specific AI (Karanicolas, 2024), but these tools are frequently developed in-house and may not be able to leverage frontier AI tools. 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.

Surrogate models would enter this workflow as tools the compliance review in Section 2.4 can call, checking reported values against the physics of the described package. The models themselves are in-house builds, and Section 3.4 sets out that development path.

2.4 Compliance Review and Gap Analysis

2.4.1 Acceptance Criteria and Review Depths

A compliance check grounded in the curated regulatory corpus can systematically identify where a candidate SARP may be missing information, incomplete relative to acceptance criteria, or internally inconsistent. Those gaps would be flagged for expert review with citation-level traceability to specific regulatory requirements.

The acceptance criteria for the check would come from NUREG-2216 (NRC, 2020). Evaluating a drafted SARP against those criteria would also flag areas likely to trigger requests for additional information before the application is submitted.

A compliance review also has more than one depth. NUREG-2216 directs the reviewer to compare reported results against the acceptance criteria and then to grade confirmatory analysis by the size of the margin (NRC, 2020). The graded analysis runs up to independent calculations of source terms, component temperatures, and reactivity. A reported value can meet its regulatory limit and still be inconsistent with the package the SARP describes.

That second depth is a physical consistency check, and the surrogate models of Section 2.3 are the tool for it. Where a validated surrogate exists, the review would check the reported values against the physics of the package the SARP itself describes. Reported dose rates would be checked not only against the regulatory limits but against the stated source term and shielding design. That comparison ties the chapters of the application together, which is how a reviewer

reads a SARP. Sections 3.1 and 3.5 describe how the review would call a surrogate and how each call would be controlled and logged.

2.4.2 Feasibility of the Compliance Check

A compliance review is fundamentally an information retrieval, cross-referencing, and pattern-matching task, performed across the same structured regulatory corpus that supports drafting. This task is precisely the domain where retrieval-augmented generation excels. The model is asked what the corpus already says instead of composing new text, so the exposure to fabricated content is lower than it is for the drafting use case in Section 2.1. Grounding comes from the same curated corpus and the same clause-level retrieval that Section 2.2 describes, which means this use case adds no new data requirement of its own.

The output is easier to govern than drafted prose. A finding list is accepted or rejected one item at a time, and every item names the passage it came from. Rejecting one finding costs nothing else. No finding reaches the SARP unless a subject matter expert decides it should.

Calling a surrogate would be a bounded use of frontier model capability, with the physics staying in the surrogate. A wrong extraction or a call outside the surrogate's validated range would be caught at the expert gate rather than in the SARP, through the controls in Section 3.5. Each surrogate is an in-house build with its own validation and uncertainty quantification.

The retrieval limits in Section 2.2.2 apply here, and one failure mode is specific to this use case. Retrieval that misses an applicable criterion produces a silent gap rather than a visible error, and a completeness check is exactly where that failure mode does the most damage. Measuring recall against an expert baseline is therefore the first thing a pilot would need to establish.

3.0 Approach and Implementation

3.1 Proposed 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 fissile and radioactive material 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.

The compliance review use case in Section 2.4 would reuse this workflow with the direction reversed. The system would take a drafted SARP chapter as input and retrieve the acceptance criteria that apply to that chapter. Instead of drafted text, the output would be a list of findings. Each finding would name the SARP section, the acceptance criterion, and the source passage behind the finding. For a discipline with a validated surrogate, the system would also extract the design parameters the chapter states. Those parameters would go to the surrogate through a tool call[1] with preset instructions, and the surrogate's estimate would be compared with the reported value. A consistency finding would name both numbers, the tolerance, and the surrogate's uncertainty bounds. A subject matter expert would accept or reject every finding before the SARP changes. Figure 2 shows this loop, with the expert baseline beside it.

[1] A tool call is a step where the language model invokes an external program through a defined interface. The model supplies the inputs and receives the result, and the computation itself runs in the program.

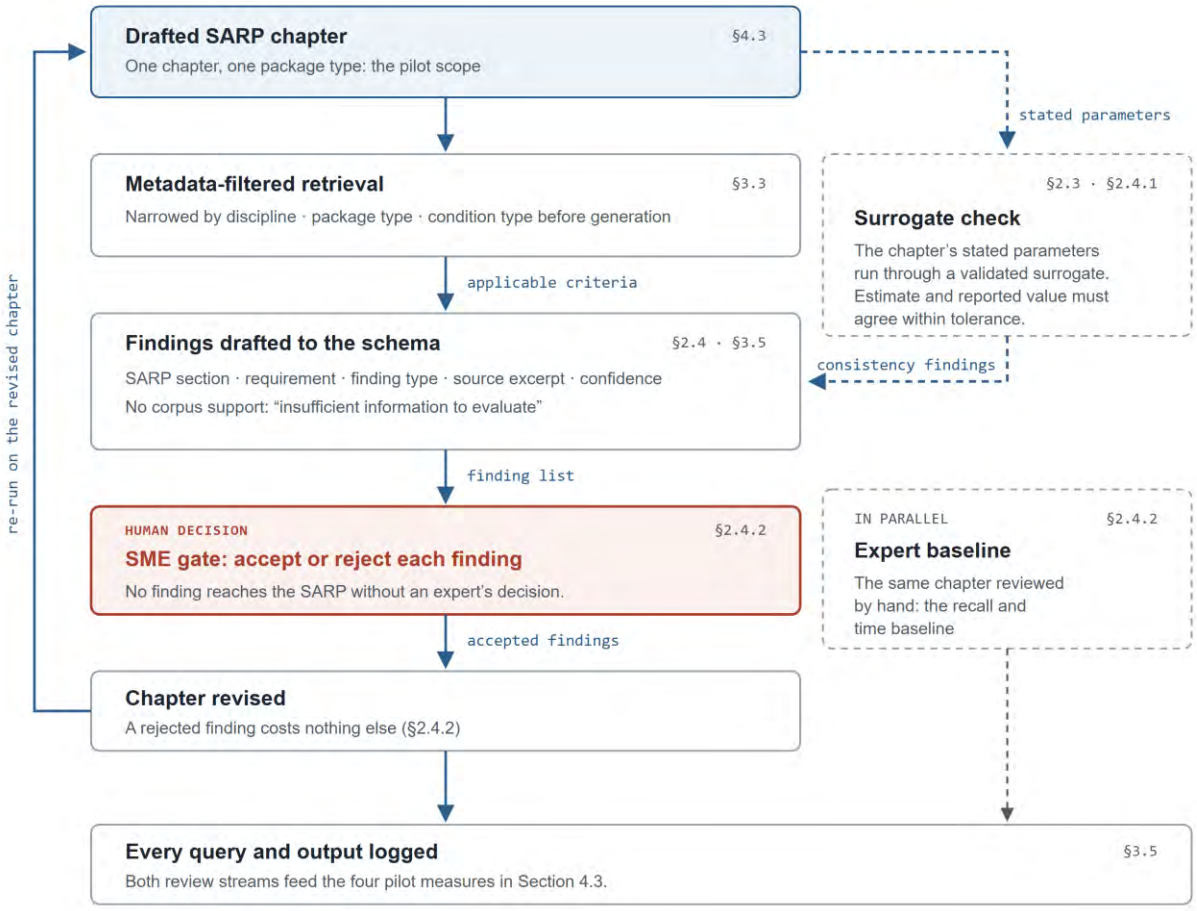


Figure 2. The compliance review loop. A subject matter expert accepts or rejects each finding before the chapter changes. The same chapter reviewed by hand sets the baseline the tool is measured against.

3.2 Information Corpus

The corpus of information used for the AI models would be developed to be technical discipline specific (e.g. shielding, criticality, structural, etc.) and based on NRC regulations, guides and information used to in licensing of fissile and radioactive material transportation packages such as:

- 10 CFR Part 71 Regulations
- DOT Hazardous Material Regulations
- The NRC standard review plan for transportation of radioactive and fissile packages (NUREG-2216)
- Applicable NUREGs and Regulatory Guides (e.g. RegGuide 7.9)
- Applicable standards (ASME Code Standards, ANS/ANSI Standards)

- Publicly available SARPs and NRC staff safety evaluation reports (SERs) and Requests for Additional Information (RAIs)

NRC documents are available on the ADAMS database and can be programmatically aggregated via the ADAMS public search API (https://adams-api-developer.nrc.gov/).

For some microreactors that may not meet the traditional deterministic regulations, NRC has proposed a modification to 10 CFR Part 71 (NRC, 2026) that allows the use of a previously approved or similar risk informed approach. This is a newer approach for approval of radioactive and fissile material transportation packages regulated by the NRC. Available information (such as Coles, et al. 2024) will be used to train AI models to assist with SARP development that uses risk informed approaches.

3.3 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 a single context window or retrieval structure (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.

3.4 Data and Retrieval 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.

The compliance review logic described in Section 2.4 would execute against a metadata-filtered index that constrains retrieval by discipline, package type, and condition type. Narrowing before generation is cheaper than checking afterwards, and it removes a class of error rather than detecting it.

3.5 Custom Model Development

Beyond frontier models, task-specific non-generative AI could be built in-house to address particular analytical needs. Section 2.3 describes what the surrogate models would do. This section describes what building one takes. The training set comes from validated full-fidelity simulation runs that span the parameter range the package analyses need, supplemented by test data where test data exists. In data-sparse regimes or where extrapolation risk is high, PINN approaches embed governing equations and constitutive relations into model reasoning. Each model would pass the uncertainty quantification practices in Section 2.3.2 before the model is treated as validated, and each would be delivered as a callable tool with a fixed interface and a stated valid input range. These models are not generative, but the compliance review in Section 2.4 would call them as tools, with the frontier model extracting the inputs from the SARP and drafting the finding from the comparison.

3.6 Architecture and Controls

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.

Constrained decoding is necessary but not sufficient. A trustworthiness control layer would govern all outputs. Every finding would conform to a structured schema. The schema would specify the SARP section evaluated, the regulatory requirement, the finding type, supporting source excerpts, and a confidence indicator. The system would report “insufficient information to evaluate” rather than generating unsupported findings, and all queries and outputs would be logged for audit and reproducibility. A surrogate call would run under the same layer. The log would record the preset instructions, the extracted inputs, the surrogate version, and the returned estimate beside the finding. A physical consistency finding would be one more finding type in the same schema. A call outside the surrogate's validated range would be refused and reported, under the same rule as retrieval without corpus support. Each consistency finding would quote the extracted inputs with their source passages, so a wrong extraction is visible at the expert gate.

4.0 Conclusions

4.1 Benefits and Opportunities

The use of AI to assist in the preparation of 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. Package certification is directly on the critical path to reactor deployment, so time saved preparing the SARP is time saved on deployment itself (Section 1.0). The curated corpus and ontology would also capture expertise that is scarce: what a senior reviewer knows about precedent and applicability becomes a queryable resource rather than personal memory (Section 2.2.1).

The review side has benefits of its own. A compliance check run before submission would flag the gaps that would otherwise come back as requests for additional information (Section 2.4.1). Findings would be accepted or rejected one at a time with citation-level traceability, which makes each expert decision auditable (Section 2.4.2). Where a validated surrogate exists, the reported values would also be checked against the physics of the described package, which is a check a regulatory limit alone cannot provide (Section 2.3). Surrogate triage would focus full-fidelity simulation effort where novelty or risk is highest (Section 2.3.2). Because every query and output would be logged and every finding would follow one schema, a review would be repeatable (Section 3.5).

4.2 Barriers and Risks

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.

4.3 Proposed Pilot

This report has evaluated the feasibility of AI assistance across the SAR workflow from the published record and the research literature. The practical next step to complete the feasibility case is a bounded pilot. A single drafted SARP chapter and a single radioactive material or fissile package type, checked against the regulations already cited, would give the pilot enough scope to exercise the workflow end to end. As no microreactors have yet been approved for transportation and there are currently no publicly available SARPs, a package such as a fissile (Type AF) or a spent nuclear fuel (Type BF) would be used for this pilot. Compliance review is the natural first target, because it needs no capability that drafting does not already need and its output is easier to check.

The pilot would begin with the foundation. The public regulations and review guidance would be indexed at clause level and tagged by discipline, package type, and condition type. A requirements traceability matrix would pair each acceptance criterion with the SARP section that must satisfy it (Section 2.2). The review loop would then run against that index. Retrieval would return the criteria that apply to the drafted chapter, findings would be drafted to a fixed schema, and a subject matter expert would accept or reject each finding before the chapter changes (Sections 2.4 and 3.1). The same chapter would be reviewed by hand in parallel. That independent review would set the recall and time baseline the tool is measured against, and every query and output would be logged. Figure 1 shows the foundation and Figure 2 shows the loop.

Four measures would tell the program whether the approach works:

- whether a retrieved citation points at the criterion it claims to,
- how much a subject matter expert edits an AI-assisted draft before it is usable,
- whether repeated runs on the same input return the same findings, and
- how elapsed time compares against the same work done by hand.

Thresholds for each belong with the pilot design rather than with this assessment. Setting them here would commit the program to numbers this report has not derived. An illustrative set shows the form the thresholds would take:

- citation accuracy of at least 80 percent at the paragraph level,
- at least 60 percent of findings accepted by the expert without edits,
- at least 90 percent identical findings across three repeated runs on the same input, and
- review time an order of magnitude below the same review done by hand.

The numbers are examples of form rather than commitments. The pilot design would set the real numbers and would tighten the thresholds as the corpus and the iteration cycles grow. Figure 3 shows the measures with the illustrative thresholds and the decision the measures feed.

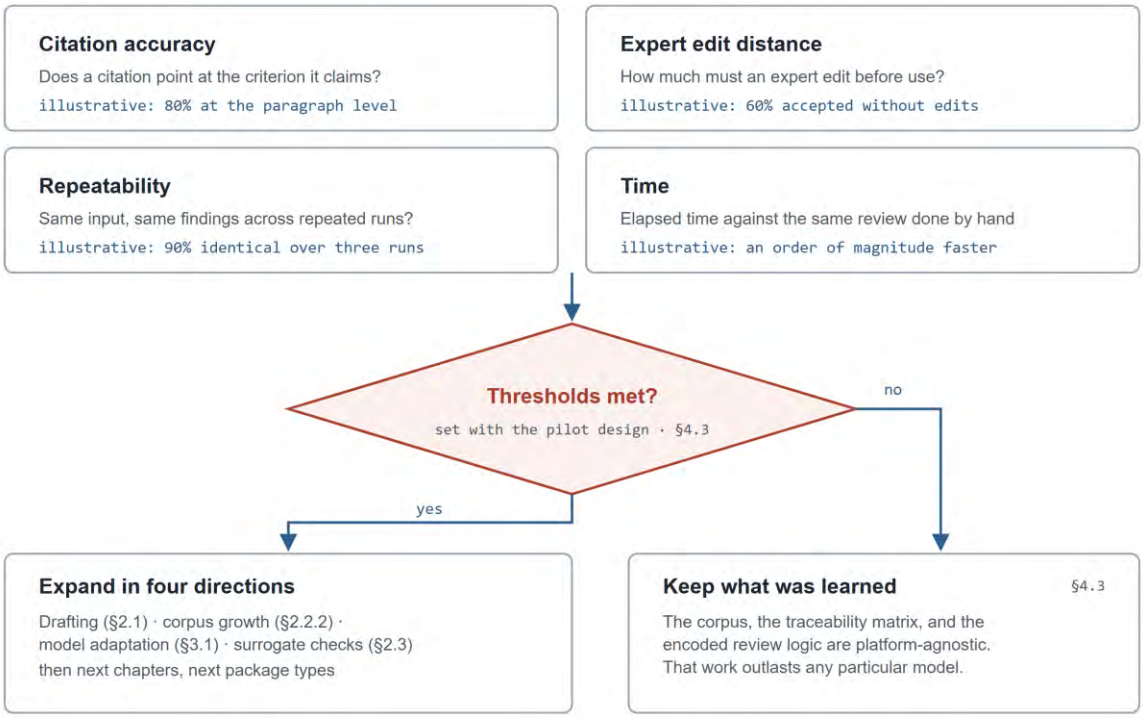


Figure 3. The four pilot measures with illustrative thresholds, and the decision the measures feed. The real thresholds are set with the pilot design. Either outcome keeps the corpus and the encoded review logic.

The overall finding is that AI is a promising approach for this challenge. What would make it work is not the model but the systematic encoding of regulatory review expertise into an AI-capable framework. The curated corpus, the requirements traceability matrix, and the discipline-specific review logic that encoding produces are platform-agnostic by design, and they outlast any particular model.

This page was intentionally left blank.

5.0 References

Acharya, A., S. Munikoti, A. Hellinger, S. Smith, S. Wagle, and S. Horawalavithana. 2023. "NuclearQA: 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.

Coles G.A., T.A. Ikenberry, S.M. Short, M.S. Taylor, H.E. Adkins, P.P. Lowry, and C.A. Condon, et al. 2024. "Development and Demonstration of a Risk Assessment Approach for Approval of a Transportation Package of a Transportable Nuclear Power Plant for Domestic Highway Shipment." Richland, WA: Pacific Northwest National Laboratory.PNNL-36380 Revision 1.

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

DOT (U.S. Department of Transportation). 2024. *Hazardous Materials Regulations*. Title 49, Code of Federal Regulations, Parts 171–180. Office of the Federal Register, 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. 2025. "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.

INL (Idaho National Laboratory). 2019. *Key Regulatory Issues in Nuclear Microreactor Transport and Siting*. INL/EXT-19-55257. Idaho National Laboratory, Idaho Falls, Idaho.

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.

NRC (U.S. Nuclear Regulatory Commission). 2005. *Standard Format and Content of Part 71 Applications for Approval of Packages for Radioactive Material*. Regulatory Guide 7.9, Revision 2. U.S. Nuclear Regulatory Commission, Washington, D.C.

NRC (U.S. Nuclear Regulatory Commission). 2020. *Standard Review Plan for Transportation Packages for Spent Fuel and Radioactive Material*. NUREG-2216. U.S. Nuclear Regulatory Commission, Washington, D.C.

NRC (U.S. Nuclear Regulatory Commission). 2024. *Packaging and Transportation of Radioactive Material*. Title 10, Code of Federal Regulations, Part 71. Office of the Federal Register, Washington, D.C.

NRC (U.S. Nuclear Regulatory Commission). 2026. "Modernizing Package Certification Requirements." *Federal Register* 91 (July 27, 2026): 46869–81.

Phan, H., A. Acharya, S. Chaturvedi, S. Sharma, M. Parker, D. Nally, A. Jannesari, K. Pazdernik, M. Halappanavar, S. Munikoti, and S. Horawalavithana. 2024. “RAG vs. Long Context: Examining Frontier Large Language Models for Environmental Review Document Comprehension.” *arXiv*. https://doi.org/10.48550/arXiv.2407.07321.

Viswanathan, A. 2026. “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.