

Retrieval-Augmented Generation (RAG)

Transforming Documentation Access in the Nuclear Industry

What is RAG?

Think of RAG as giving an AI assistant a specialized library card system. Instead of relying only on what the AI learned during training, RAG lets the AI search through your actual documents in real-time before answering questions. It's like the difference between asking someone to recall facts from memory versus letting them look up the answer in your company's filing cabinets first.

Why Nuclear Operations Need RAG

The nuclear industry faces a unique documentation challenge:

- Regulatory documents scattered across federal databases (NRC, DOE)
- Operating procedures stored in plant-specific systems
- Maintenance records in CMMS (Computerized Maintenance Management Systems)
- Safety analyses locked in engineering workstations
- Vendor technical manuals sitting in SharePoint or legacy file servers
- Incident reports archived across decades in multiple formats

When an operator, engineer, or compliance officer needs an answer, they're often searching through dozens of systems, multiple formats (PDFs, Word docs, scanned images), and thousands of pages of technical documentation.

How RAG Solves This Problem

The Traditional Problem

Question: "What are the emergency shutdown procedures if coolant pressure drops below 2000 PSI in Loop A during refueling?"

Current Reality: Someone spends 30-60 minutes searching through:

- Emergency Operating Procedures (EOPs)
- Technical Specifications
- Refueling safety analysis
- Previous incident reports
- Vendor pump manuals

With RAG

The same question gets answered in seconds by:

1. Searching across all documents simultaneously - The system converts your question into a mathematical representation and finds every relevant section across all your documentation sources
2. Understanding context - Unlike keyword search, RAG understands that "coolant pressure" relates to "primary loop pressure," "RCS pressure," and "reactor coolant system" - even if those exact words aren't in your question
3. Pulling relevant excerpts - The system retrieves the specific paragraphs, tables, and procedures that answer your question
4. Generating a coherent answer - An AI (like Claude) reads those excerpts and writes a clear response in plain language, citing exactly which documents it used
5. Maintaining traceability - Every answer includes citations back to the source documents, section numbers, and revision dates - critical for regulatory compliance

Complex Query Capabilities for Nuclear Operations

Multi-Document Analysis

Question: "Compare the containment spray actuation setpoints across our three reactor units and identify any discrepancies with NRC Regulatory Guide 1.82."

RAG can:

- Pull specs from all three units simultaneously
- Reference the current NRC guidance
- Identify differences
- Explain which differences are within acceptable variance vs. require attention

Time-Based Queries

Question: "Show me all maintenance procedures that were revised in the last 6 months and explain what changed."

The system tracks document versions and can analyze what actually changed between revisions - invaluable for training and change management.

Cross-Referencing Regulations

Question: "Does our current fire protection system meet the requirements in 10 CFR 50 Appendix R and NFPA 805?"

RAG can simultaneously check your system documentation against multiple regulatory standards and identify gaps or compliance issues.

Incident Investigation

Question: "Have we had similar coolant chemistry issues in the past 10 years, and what corrective actions were effective?"

The system searches historical incident reports, corrective action databases, and chemistry logs to find patterns and solutions that worked previously.

Technical Architecture (Simplified)

Here's what happens behind the scenes:

1. Document Ingestion - Your PDFs, Word docs, databases, and legacy files are processed and converted into searchable text
2. Intelligent Chunking - Documents are broken into logical sections (by procedure step, by regulation paragraph, etc.) that maintain context
3. Mathematical Encoding - Each section is converted into a "vector" (a list of numbers) that represents its meaning mathematically
4. Vector Storage - These vectors are stored in a specialized database optimized for similarity searching
5. Query Processing - When someone asks a question, it's also converted to a vector and compared against all stored document vectors
6. Relevance Ranking - The system finds the most mathematically similar sections (the ones most likely to contain the answer)
7. Answer Generation - An AI reads those sections and crafts a response, always citing sources

Security and Compliance Considerations

For nuclear facilities, RAG implementations include:

Access Control - Users only see answers from documents they have clearance to access. If your question requires Safeguards Information, the system checks your authorization first.

Audit Trails - Every query, every document accessed, and every answer generated is logged with timestamps and user IDs - meeting regulatory record-keeping requirements.

No Training on Your Data - Your proprietary procedures and sensitive information never leave your environment. The AI doesn't "learn" from your documents; it just reads them when needed.

Air-Gapped Options - RAG systems can run entirely on-premises with no internet connection, meeting security requirements for classified facilities.

Version Control - The system knows which revision of which procedure it's citing, preventing outdated information from being used.

Bottom Line

RAG technology transforms scattered, siloed documentation into an intelligent knowledge base that understands natural language questions and provides accurate, cited answers in seconds. For the nuclear industry, where finding the right information quickly can be a matter of safety, compliance, and operational efficiency, RAG represents a significant operational advantage.

The technology is mature, proven, and can be implemented securely within the stringent requirements of nuclear facilities. It doesn't replace human expertise - it amplifies it by making institutional knowledge instantly accessible to everyone who needs it.