ANL-25/59

LLM Generation of Online Courses from a Curated Set of Documents in the Nuclear Safeguards Domain

Strategic Security Sciences Division

ANL-25/59

LLM Generation of Online Courses from a Curated Set of Documents in the Nuclear Safeguards Domain

prepared by
Scott Parent (scottp@anl.gov), Marianne N. Fisher (mnfisher@anl.gov),
Heather Seger (hseger@anl.gov), Benjamin Blakely (bblakely@anl.gov)

Strategic Security Sciences Division, Argonne National Laboratory

September 2025

Executive Summary

A multidisciplinary team at Argonne National Laboratory explores the application of advanced technologies to enhance knowledge transfer and retention within the nuclear safeguards domain. Specifically, it examines the feasibility of leveraging secure *large language models* (LLMs) to streamline the creation of e-learning modules for the U.S. National Nuclear Security Administration (NNSA) Office of International Nuclear Safeguards (NA-241). The initiative addresses the critical need for preserving institutional memory and accelerating skill development amidst the imminent retirement of senior professionals in the field in addition to supporting good knowledge management practices.

The project integrates instructional design theory with cutting-edge AI technologies to transform curated document sets from the *Safeguards Knowledge Repository* (SKR) into modular online courses. By automating the generation of learning objectives and instructional content, the effort aims to reduce manual effort while maintaining high-quality educational outcomes. A limited measure of human supervision, however, ensures accuracy, relevance, and alignment with NNSA's strategic priorities.

Key findings highlight the potential of AI-assisted course generation to support safeguards professionals by creating structured, interactive learning experiences. The report underscores the importance of SME validation to address limitations in AI-generated content, such as terminology errors and gaps in coverage. Recommendations include adopting a structured workflow combining LLM acceleration with expert oversight to ensure accuracy, usability, and alignment with learner needs. This work demonstrates Argonne's commitment to advancing national security and scientific excellence through innovative knowledge management solutions.

Contents

Introduction

A multidisciplinary team at Argonne National Laboratory (ANL), a leading U.S. Department of Energy research institution, explored the application of advanced technologies to enhance knowledge transfer and retention within the nuclear safeguards domain. Specifically, it examined the feasibility of leveraging secure *large language models* (LLMs) to streamline the creation of e-learning modules for the U.S. National Nuclear Security Administration (NNSA) Office of International Nuclear Safeguards (NA-241).

Knowledge transfer in the safeguards domain is critical due to the highly specialized expertise required, the imminent risk of attrition as senior professionals retire, and to support good knowledge management practices. Building on Argonne's prior innovations, this project integrates instructional design theory with cutting-edge AI technologies to create structured, interactive learning experiences. By transforming curated document sets from the *Safeguards Knowledge Repository* (SKR) into modular online courses, the initiative aims to preserve institutional memory, accelerate skill development, and strengthen the safeguards workforce.

This report outlines the methodology, findings, and recommendations from this effort, highlighting the potential of AI-assisted course generation to reduce manual effort while maintaining high-quality educational outcomes. It underscores the importance of human-in-the-loop processes to ensure accuracy, relevance, and alignment with NNSA's strategic priorities. Through this work, Argonne continues to advance its mission of supporting national security and scientific excellence.

Background and Context

Knowledge management, retention, and transfer is a strategic priority for NA-241. This project builds on Argonne's prior innovations to assess the feasibility of scaling and supporting knowledge transfer efforts through the use of secure LLMs.

Argonne National Laboratory has designed and implemented a web-based knowledge management system—originating as a knowledge repository for NA-241—that is now used by multiple NNSA offices. The inaugural system, with the longest managed experience, is the *Safeguards Knowledge Repository* (SKR, pronounced "Seeker"). SKR hosts thousands of safeguards-related documents developed over decades from across the U.S. laboratories and beyond, serving as a critical resource for U.S. safeguards professionals and creating opportunities to learn about various safeguards-related topics. However, providing access to documents alone does not effectively address diverse learning styles. This project integrates instructional design theory to create structured, interactive learning experiences that can assist instructional designers or those who are building online modules to support safeguards training and education efforts.

National Nuclear Security Administration

NNSA is a semi-autonomous body within the *U.S. Department of Energy* (DOE) responsible for the stewardship of the United States' nuclear arsenal among other missions. NNSA is divided into several offices that carry out the various aspects of the agency's mission. More specifically, the NNSA's *Office of Defense Nuclear Nonproliferation* (DNN) focuses on preventing the development and acquisition of nuclear and radiological weapons, materials, equipment, technology and expertise (*Nonproliferation*, n.d.). Within DNN is NNSA's *Office of Nonproliferation and Arms Control* (NPAC) which develops and provides technical and policy solutions related to nuclear nonproliferation, international safeguards, export controls, and arms control. They complete their safeguards mission through NA-241 by developing and implementing new safeguards concepts, approaches, and technologies. The office is especially sensitive to the loss of knowledge and expertise experienced with the retirement of experts in these fields or attrition to other fields (*About NNSA*, n.d.).

Safeguards Knowledge Repository

The U.S. Safeguards Knowledge Repository (SKR) is a web-accessible resource available to experts and practitioners across national laboratories, NNSA, and other U.S. government safeguards professionals—access is controlled and granted to users by NA-241. SKR supports implementation of international nuclear safeguards in line with the agreements under the *International Atomic Energy Agency* (IAEA). SKR hosts thousands of safeguards-related documents—often pertaining to current approaches and associated efforts—drawing from efforts across all national laboratories and the IAEA.

Curated Sets

Within SKR, *Subject Matter Experts* (SMEs) currently package topical content into *curated sets*—where each such set is a list of documents believed by a SME to offer particularly illuminating content on a given safeguards topic, for topics of high relevance to the safeguards workforce. Several curated sets have been developed specifically with educational purposes in mind. SMEs were asked to bring together a reading list of foundational documents relevant to current safeguards topics and research with the idea that SKR users new to that topical area can quickly come up-to-speed.

Learning Objectives

While a simple reading list or a curated collection of documents can be a valuable resource, its effectiveness is often limited. This method primarily serves individuals who are already self-directed and learn best by consuming large blocks of content. However, to create a truly impactful learning experience that reaches a wider variety of learners, we must turn to the principles of *Instructional Design Theory*. This field provides a systematic framework for creating educational experiences that are efficient, effective, and engaging. At the very core of this framework lies the concept of the *Learning Objective* (LO).

Building on the foundational work of educational psychologist Benjamin Bloom, a learning objective—or "Educational Objective" as he originally named it—is a clear and specific statement that describes what a learner should know, be able to do, or value as a result of a learning experience (Bloom, 1956). The introduction of LOs marked a crucial shift in educational philosophy: away from a focus on what the instructor will *teach* and towards a focus on what the learner will demonstrably *achieve*. A well-crafted LO acts as a blueprint for the entire educational process. It is a promise to the learner and a guide for the instructor, providing direction, purpose, and a clear destination for the learning journey.

The power of learning objectives lies in their role in creating alignment. Once LOs are established for a course or module, they become the central pillar connecting content and assessment. This principle, often called "constructive alignment," ensures that every component of the course is purposeful. Content—such as readings, videos, interactive learning activities, and lectures—is no longer just a collection of interesting information; it is strategically selected specifically to help the learner achieve the stated objectives. Similarly, assessments—from quizzes and exams to projects and presentations—are designed not merely to test memory, but to accurately measure the learner's proficiency or mastery with respect to that very same objective. For example, if an objective is to "analyze the causes of the Industrial Revolution," the content must provide the necessary evidence and perspectives, and the assessment must require the learner to perform an actual analysis, not just list dates and names. Furthermore, LOs can be organized hierarchically to create a logical and manageable learning path, a process known as scaffolding. More fundamental knowledge and abilities, such as defining key terms or identifying basic concepts, can be established first. These foundational objectives then serve as the building blocks for more complex objectives that require higher-order thinking, such as application, analysis, or creation. This progression prevents learners from being overwhelmed and ensures they have the prerequisite skills needed to succeed at each new level of complexity.

To help educators classify and structure these objectives, Bloom organized them into a tiered **taxonomy**. Originally published as the cognitive domain, this taxonomy provided a hierarchy of intellectual behaviors, from simple recall of facts to more complex and abstract mental processes. The original levels were Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. This framework was later revised by Anderson (Anderson, 2001), who reframed the levels as verbs to better reflect the active nature of learning: Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. This revised taxonomy provides instructors with a powerful vocabulary to craft precise LOs that target specific cognitive skills, ensuring that a course challenges learners to move beyond simple memorization and engage in deeper, more meaningful thinking. By using learning objectives guided by this taxonomy, instructional designers can transform a simple "reading list" into a dynamic, intentional, and measurable learning experience.

NPAC gAXIS and Moodle

Moodle is an open-source *Learning Management System* (LMS) intended to make, “quality online education accessible for all” (Moodle, n.d.). Moodle is available both for self-hosting and through multiple hosting providers.

Oak Ridge National Laboratory (ORNL) hosts a web-based Moodle instance for the NPAC community packaged as *NPAC gAXIS* (Oak Ridge National Laboratory, n.d.). NPAC’s *Human Capital Development* (HCD) office recognizes NPAC gAXIS as a standard learning platform.

Creators of gAXIS-hosted training programs can apply for a “Teacher” level of administrative access to the Moodle instance which allows the development and maintenance of particular Moodle *courses*. This access allows control over myriad Moodle course features including structural components like *sections, activities,* and *resources*; simple attributes like *full name,* and *description*; directives for behavior, appearance, and availability; and a list of *participants*.

Support for many kinds of activities and resources is supplied with the standard installation package, and additional kinds of activities and resources can be supported through the installation of *plugins*. Moodle distinguishes information, activities, and other notes from a resource in terms of interaction: whereas an *activity* (Figure 1) facilitates learner interaction, a *resource* is in this respect static.

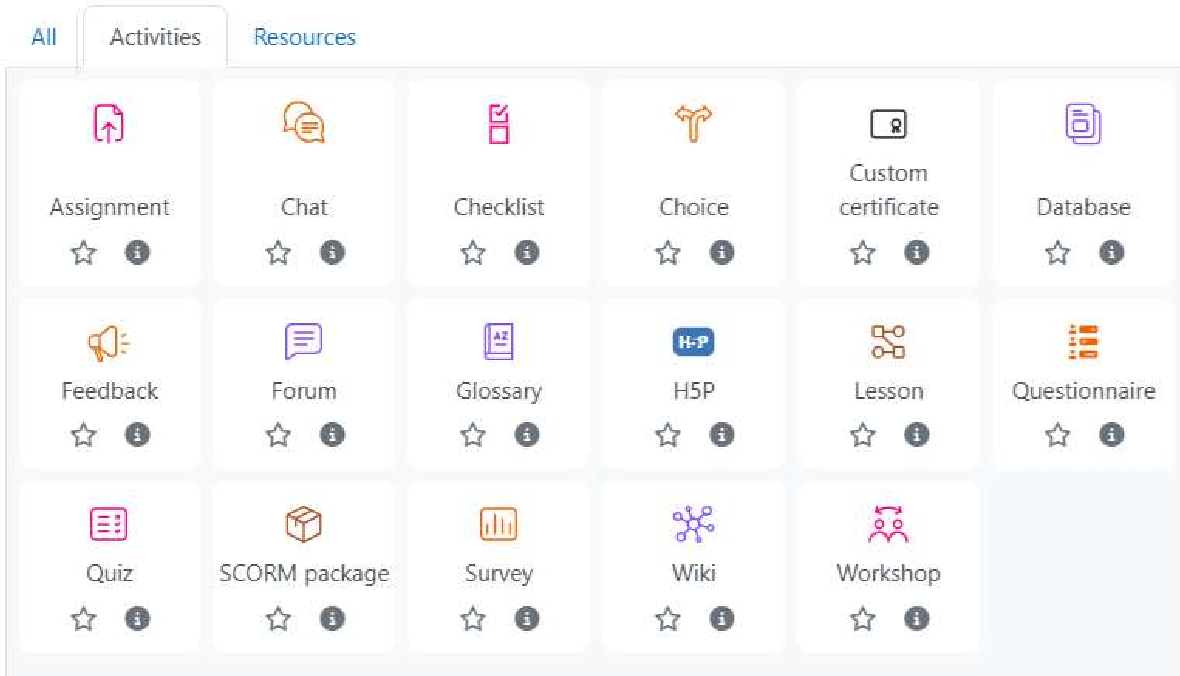

Figure 1 - Activities Available for Moodle Course Construction

Activities options available from Moodle for course construction include different structures for attaching content to learner experience. These options include attachment of content from standard LMS packages. With the basic Moodle installation, two particular LMS standards are supported for this kind of content attachment: *Shareable Content Object Reference Model* (SCORM) and *HTML5 Package* (H5P).

Importance Of Knowledge Transfer in the Safeguards Domain

Knowledge transfer is critical in the nuclear safeguards domain because the expertise required is both highly specialized and predominantly tacit, built over decades of field inspections, complex state-level assessments, data interpretation, and integration of many sources of information to draw conclusions on keeping nuclear technologies and materials in peaceful uses. As many senior specialists approach retirement and as retention of new practitioners is prioritized by NA-241, the risk of knowledge attrition threatens continuity of the readiness of the US safeguards workforce, reduces analytical depth, and lengthens the time needed to qualify new practitioners. This is especially vital as safeguards must adapt to evolving approaches, including novel fuel cycles, safeguards-by-design practices, advanced analytics, integration of AI and *machine learning* (ML) approaches, and increasingly digital, remotely monitored facilities. A robust knowledge transfer strategy preserves institutional memory, accelerates skill development, and strengthens the safeguards culture that underpins credible verification efforts and the integrity of the US safeguards workforce.

Hosting high-quality resources is a force multiplier for this mission, as supported through this scoped course generation effort – particularly leveraging online spaces. Curated learning objectives and modular online content for SKR users and beyond can sustain proficiency, introduce specialized topics, and provide just-in-time refreshers aligned with NA-241 priorities and IAEA competency frameworks. Well-designed e-learning—combining scenario-based exercises, case studies, and more helps practitioners stay current while creating an on-ramp for the next generation of US safeguards practitioners. Integrating contributions from retiring experts, national laboratories, academia, and international partners, and linking content to measurable outcomes will directly support NA-241's mission and strengthen the broader US safeguards community.

Course Creation Effort

e-Learning, course instruction accessed through electronic means such as a computer, tablet or smartphone, is a powerful tool for transfer of knowledge. It creates opportunities that otherwise do not exist. In the case of synchronous e-learning, instructor led courses that students attend in real time, instruction can reach a wider audience as it is not constrained by geography in the way that a traditional classroom setting is. Asynchronous e-learning retains the benefit of reaching a wider audience and also allows students to set their own pace for learning as lessons are accessible by individuals from anywhere at any time.

e-Learning can be an effective form of instruction. It is neither better nor worse than traditional classroom instruction for most topics and individuals. In fact, studies have shown that the medium of instruction has less impact on learning than the design of the course itself (Clark and Mayer, 2016). And in the workplace, training is increasingly done in an e-learning environment with 91% of organizations conducting mandatory training at least partially online and 48% completing all compliance training online in 2024 (Freifeld, 2024). Current statistics show that 100% of large organizations, 94% of midsize organizations, and 82% of small organizations utilize an LMS platform to manage learning programs (Bouchrika, 2025). With 70% of employees stating that online self-paced courses are their preferred method of learning (Employee Training Statistics in 2025), e-learning clearly will continue to be a necessary resource for training and skill development.

In short, developing curriculum is a time-consuming endeavor. The amount of time and effort required varies according to several factors including the level of experience of the instructor, the subject matter of the course, the length of the course, and how detailed the course is. Longer and more advanced courses will take more time to produce. The industry standard is that 10-20 hours of course development should be expected for every hour of instruction (He, 2021). If AI can be successfully leveraged to produce a well-designed course, it can save significant effort on the part of instructors and subject matter experts.

Objectives of the Project and alignment with NNSA's goals

In a word, this project aims to significantly improve the potential return on investment (ROI) for both production and consumption of safeguards-relevant e-learning. Specifically, this work seeks to explore optimizing the balance between human effort and training efficacy with techniques for producing modular e-learning courses from a SKR curated set. These courses, in turn, are intended to reduce the effort required to gain the essential knowledge ascribed to the curated sets from which they were produced.

Aspects of this process explored by this work include:

- Generation of LOs from SKR curated sets to support production of these courses
- Generation of instructional content from these sets to support given LOs
- Deployment of resulting modules to NPAC gAXIS to make them available to safeguards professionals through the NPAC community.

Envisioned Use Cases

Use cases that would benefit from this improvement include assisting instructional designers with course production and assisting learners both by delivering a more efficient learning experience and long term by potentially generating interactive courses on demand (Figure 2).

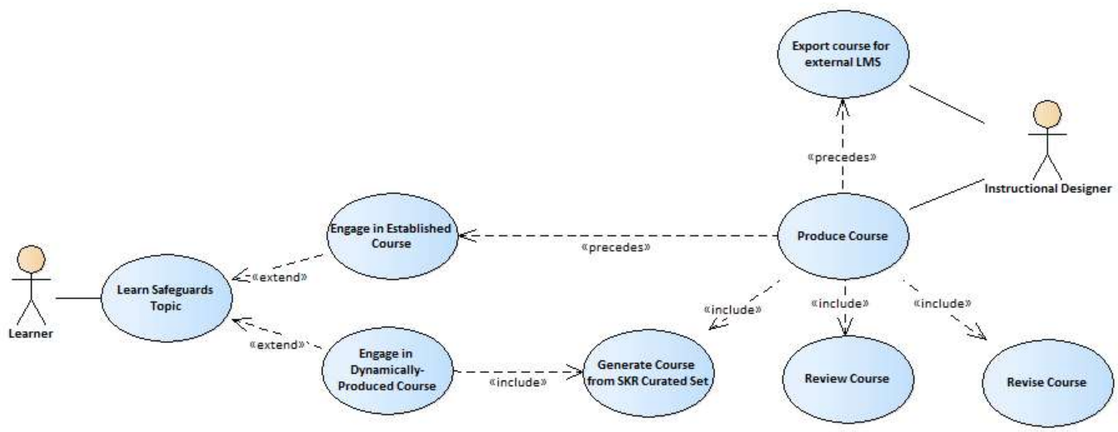

Figure 2 - Use Cases Envisioned to Benefit from Improved ROI

Related Work

This research builds upon a growing body of work exploring the use of LLMs for generating educational and training materials. The primary goal of this project is to create a system that can generate portable safeguards courses from human-curated document sets, a process that involves deriving LOs, summarizing content, and creating assessments. Key areas of prior work relevant to this effort include pedagogical content generation using LLMs, the evaluation of LLM knowledge and reliability, and methodologies for robust LLM assessment.

LLMs for Pedagogical Content Generation

Recent research has demonstrated the potential of LLMs to assist in various instructional design tasks. A broad survey of the field highlights LLM applications in generating diverse content types, including quizzes, teaching plans, and personalized feedback, while also noting the importance of techniques like *Retrieval-Augmented Generation* (RAG) and *Knowledge Graphs* (KGs) to improve factual grounding (Razafinirina, 2024). This validates the general approach of this project, which combines LLMs with a curated knowledge repository. For this project, leveraging knowledge graphs to handle specific safeguards terminology could enhance the accuracy and explanatory power of the generated content.

A foundational step in our course generation pipeline is the creation of learning objectives. Research by Sridhar et al. directly addresses this, investigating the use of GPT-4 to generate LOs for a university course prior to content development. By providing a detailed prompt that included guidelines on LO structure and examples, they found that GPT-4 could produce taxonomically appropriate LOs, with conceptual modules yielding lower-level Bloom's taxonomy verbs (e.g., "describe") and project modules yielding higher-level verbs (e.g., "implement"). This study's implication for our project is significant, as it provides a successful precedent for the automated authoring of LOs, though it also underscores the necessity of human validation to address issues like a lack of specificity or the bundling of multiple verbs in a single objective (Sridhar, 2023).

Building from LOs to complete lesson plans, other studies show that structuring LLM interactions with pedagogical frameworks improves output quality. Fan et al. developed LessonPlanner, a system that scaffolds GPT-4 outputs using Bloom's Taxonomy for LOs and Gagne's Nine Events for lesson structure. In a user study with novice teachers, this structured approach produced lesson plans of significantly higher quality and reduced the users' cognitive load compared to a free-form ChatGPT baseline. This suggests that our project can accelerate the work of subject matter experts (SMEs) by incorporating established instructional design principles directly into the generation pipeline, although vigilance is required regarding factual accuracy and appropriateness for the target learner level (Fan, 2024). Similarly, Hu et al. found that using a detailed, phase-by-phase prompting framework based on "mathematical problem chains" allowed GPT-4 to generate high school math lesson plans that approached the quality of exemplary human-authored plans, particularly in setting objectives and organizing activities. This result reinforces the strategy that high-quality outputs depend on structured, domain-aware prompting, implying that our project's success will be closely tied to the design of our prompt templates (Hu, 2024).

Evaluating LLM Knowledge and Reliability

While LLMs show promise in content generation, their reliability and factual accuracy remain critical concerns, especially in specialized domains. Work by Acharya et al. is directly pertinent to our focus on nuclear safeguards. They introduced NuclearQA, a benchmark of expert-authored questions in the nuclear domain and found that the performance of four different LLMs was "less than satisfactorily," with frequent hallucinations observed. This finding serves as a critical caution for our project, highlighting the acute need for domain-specific evaluation and rigorous human-in-the-loop oversight from our training SMEs to ensure the factual correctness of generated safeguards content (Acharya, 2023).

A key capability for a system operating on a finite set of curated documents is the ability to recognize when information is not present and to abstain from answering. This concept was formalized in the SQuAD 2.0 dataset, which added unanswerable questions to a reading comprehension task. The dataset was designed to train and test a model's ability to "know what they don't know," a foundational requirement for our system to avoid fabricating content not supported by the source documents in the Safeguards Knowledge Repository (Rajpurkar, 2018). More recently, the RepliQA benchmark was created to specifically test LLM reading comprehension on "unseen" synthetic documents, isolating it from knowledge memorized during training. The study found that all 18 tested models performed significantly worse on this novel content compared to a standard web-based benchmark (TriviaQA) and that their ability to identify unanswerable questions was imperfect. This has direct implications for our project, as our use case mirrors the RepliQA setup of providing proprietary context via RAG. It signals that we must explicitly test our system's performance on extracting information solely from the provided documents and its ability to refuse generation when the source material is insufficient (Monteiro, 2024).

Methodologies for Robust LLM Assessment

The challenges of reliability and domain-specificity necessitate robust and nuanced evaluation methodologies. The HiST-LLM benchmark, built from a structured graduate-level history databank, provides a model for creating assessments from an expert-curated knowledge base. It also revealed that LLM performance was below expert level and varied significantly by historical period and region, cautioning that model knowledge can be uneven. The dataset's distinction between "evidenced" and "inferred" knowledge offers a valuable concept for our own evaluation, potentially allowing us to assess the model's ability to differentiate between explicitly stated facts and reasonable inferences within the safeguards domain (Hauser, 2025).

The efficacy of LLM-generated content is also highly dependent on prompt design. Yin et al. explored this by defining a "knowledge boundary" to distinguish between what an LLM knows regardless of the prompt versus what it only knows with a specific prompt ("Prompt-Sensitive Knowledge"). Their proposed method for finding an optimal prompt for a given query suggests that our project may need to employ sophisticated prompting strategies to reliably extract information. An understanding of a model’s knowledge boundaries in the safeguards domain could inform both model selection and prompt engineering efforts (Yin, 2024).

Finally, the broader field of LLM evaluation offers principles for ensuring our assessments are meaningful. The LiveBench benchmark addresses test set contamination and judge bias by using a frequently updated set of objective, automatically-scored tasks drawn from recent sources. While our use of proprietary data mitigates web contamination, this work's emphasis on objective, ground-truth-based scoring and its inclusion of tasks like summarization and instruction following are directly relevant to our project's evaluation plan (White, 2025). The MixEval framework, which combines real-world queries with ground-truth benchmarks for efficient evaluation, further illustrates the ongoing effort to create scalable and impartial assessment tools (Ni, 2024). Collectively, these works provide a strong methodological foundation for developing a rigorous, multi-faceted evaluation plan for the generated safeguards course content.

Research Plan and Methodology

Methodology

To explore LLM generation of reusable training materials for safeguards professionals, the safeguards training *subject matter experts* (SMEs) on our team began our process by choosing a curated set from which to generate an initial course. SMEs then drafted and analyzed model LOs from which to draw expectations and criteria with which to evaluate generated LOs and content.

Meanwhile, developers on the team constructed a pipeline for generating LOs and content. After receiving the curated set selection from the SMEs, the pipeline generated a baseline set of LOs partitioned into modules. These elements were packaged both as KGs for internal reuse and in human-readable form for SME review. The pipeline was then used to generate baseline instructional content and learning activities for these modules and based on these LOs. Content was packaged as SCORM modules and deployed to Moodle.

SMEs reviewed the generated LOs and content based both on the defined criteria defined earlier, and on a more subjective level—confirming their utility as a starting point but emphasizing the necessity of iterative human involvement to ensure accuracy, relevance, and sensitivity. Early steps in this review included modifying generated LOs with minimal necessary edits to fill observed gaps.

Developers generated subsequent versions of the content based on these edited LOs.

The team analyzed the final version of the generated content and the overall experience to share findings in this report.

Generation of Learning Objectives and Instructional Content

The team constructed key segments of a pipeline that first consumes curated sets from the SKR repository to generate LOs, then retrieves potentially relevant passages from documents within the curated set to generate content for each LOs, and finally packages that content in a SCORM module for deployment to the LMS (Figure 3).

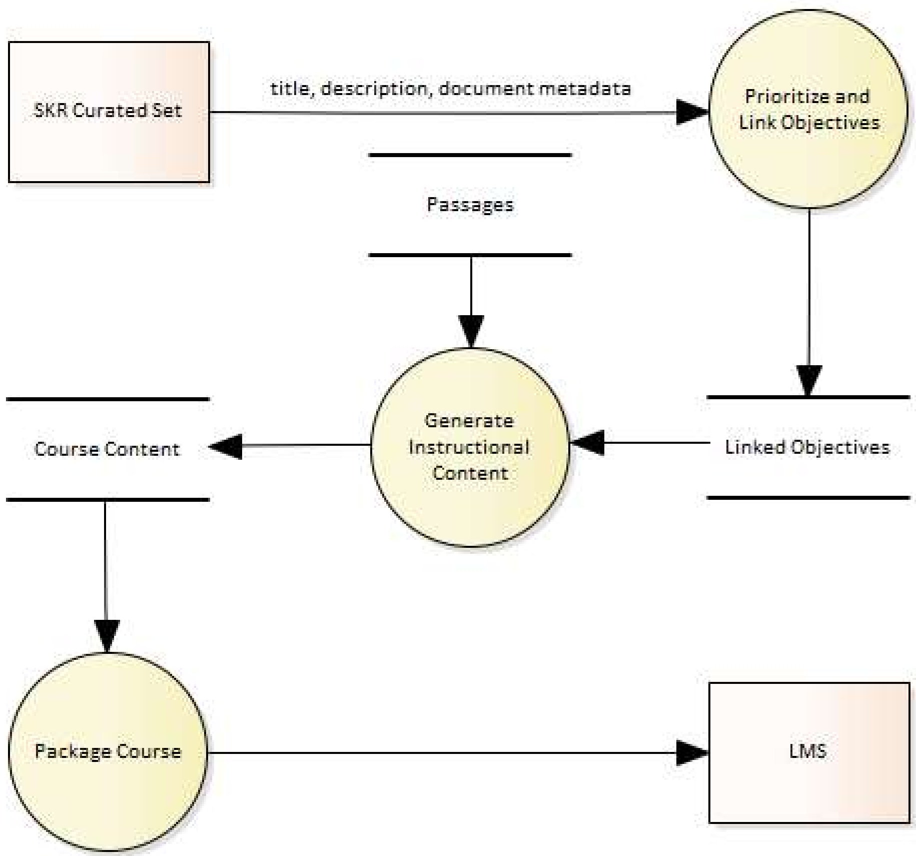

Figure 3 - Pipeline Dataflow

Tools and Technologies Used

SKR is a Java web application using a *Microsoft SQL Server* (MSSQL) database. The pipeline was constructed in Java, reusing SKR components, a copy of the SKR database, and extended to use Argonne's internal LLM façade.

Based on the team members' experience with its speed, reasoning proficiency, and capacity for KG generation, OpenAI o4-mini was selected as the LLM both to generate context for document passages and to generate the baseline set of LOs. Instructional content was generated using OpenAI GPT-5 which, though not as fast as o4-mini, features a larger input capacity and more advanced reasoning capabilities. Prompt templates were created and refined for both operations.

Project Execution

The team reviewed SKR content and selected a curated set from which to generate our initial course, developed criteria with which to evaluate the structure and content of the generated course, generated LOs from the selected curated set, generated instructional content from these LOs, deployed that content to Moodle, and evaluated the generated content. The team also explored a *human-in-the-loop* (HITL) process for improving quality of results.

Selection and Preparation of Curated Sets

The safeguards training SMEs on our team reviewed curated sets available in SKR to identify an appropriate topic for this project that was within their scope of expertise, enabling them to rigorously review and analyze the generated outputs. To maximize utility and minimize complexity for an initial pilot, the team selected an introductory theme, "Intro to Safeguards and IAEA: Structure, Scope, and Core Documents," as the primary curated set (Figure 4). Foundations of safeguards are taught across nearly all safeguards-related curricula and are essential for interns, fellows, and early-career professionals; therefore, this topic offered the broadest application and immediate relevance. The curated set already aggregates documents widely recognized by safeguards professionals as essential references (e.g., foundational treaty texts, IAEA statute and policy materials, core safeguards brochures and glossaries, organizational references, and strategic planning documents). The resulting course would be intended to supplement these documents—not to replace a reading list—by producing LOs, context, sequencing, and formative checks that help learners connect the documents to practical safeguards concepts and workflows.

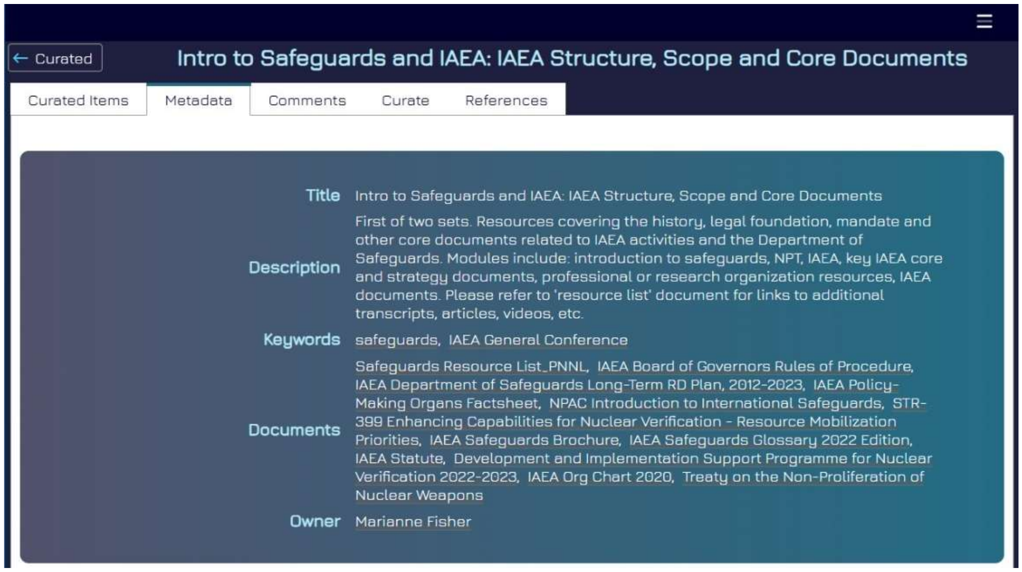

Figure 4 - Intro to Safeguards and IAEA Curated Set in SKR

To prepare the set for LLM-assisted course generation, SMEs and investigators performed light rationalization and enrichment within SKR. Preparation steps included confirming authoritative and current versions of documents, resolving duplications, and ensuring stable references and citation keys. Documents were annotated with SKR taxonomy tags (e.g., legal foundation, organizational structure, core concepts) to support targeted retrieval and alignment to LOs. Where needed, accessible text was ensured to enable accurate passage extraction. Out-of-scope materials were flagged or excluded to maintain appropriateness for introductory audiences. This preparation enabled reliable retrieval-augmented generation, ensured consistent terminology with the Safeguards Glossary, and supported a coherent progression from foundational legal and organizational concepts to application-oriented learning activities.

Development of Criteria for Evaluating Generated LOs and Content

To evaluate and refine the LLM-generated LOs and course content, the team adopted a practitioner-first perspective centered on "what newcomers fundamentally need to know." Two pillars guided the criteria: (1) the legal framework underpinning international safeguards, and (2) the core verification activities used by the IAEA. The evaluation rubric required that LOs and associated content explicitly and correctly introduce the *Treaty on the Non-Proliferation of Nuclear Weapons* (NPT) and IAEA Statute at a high level; the scope and obligations of *Comprehensive Safeguards Agreements* (CSA) *IAEA Informational Circular* (INFCIRC/153) and the Additional Protocol (INFCIRC/540); the role of *Modified Small Quantities Protocol* (SQP/mSQP) in appropriate contexts; key foundational concepts such as significant quantity and Code 3.1 reporting; and the principal verification measures (e.g., nuclear material accountancy, *Design Information Verification* (DIV), containment and surveillance, environmental sampling, information analysis, and complementary access at a conceptual level). The rubric also checked for correct terminology and acronyms consistent with the IAEA Safeguards Glossary, define-before-use sequencing, clear linkage to the curated source documents, and appropriate sensitivity and relevance for an introductory audience. Early iterations showed that some essentials were included while others were omitted or framed incorrectly; the criteria therefore incorporated coverage checks to ensure all major instruments and activities were present and accurately contextualized.

Because the course is intentionally introductory, the criteria enforced guardrails on cognitive load and depth. All LOs were capped at a "beginner" difficulty, emphasizing Bloom's Remembering and Understanding, with occasional low-stakes Applying (e.g., simple matching or identification) when helpful for engagement. Measurability was preserved through concise success criteria (e.g., "name," "describe," "identify"), clear conditions that reference only the curated documents, and small, time-bounded activities suitable for early-career learners. The rubric explicitly discouraged advanced analysis, procedural detail, or sensitive operational content that could overburden or distract newcomers, and it required define-before-use sequencing and one new concept per LO wherever possible. These constraints kept the slope into advanced material manageable, ensured conceptual coherence, and maintained alignment with the foundational purpose of the course, while preserving room for SMEs to iteratively refine and validate accuracy and relevance throughout the generation process.

SME-Proposed Learning Objectives

To explicate, analyze, and exercise evaluation criteria, SMEs proposed a model set of LOs. The first step in developing LOs is to define the purpose and target audience of the course you are creating. In the case of this particular curated set, this was predetermined. The reading list had been developed by Pacific Northwest National Laboratory for NA-241 selected fellows, interns and new hires to better familiarize themselves with the basics of international nuclear safeguards. That reading list had been added to SKR at the request of NA-241, making it available for use in the project. With this in mind, project SMEs developed this model list of objectives: expectations of what the learners would be able to do with the information contained in the course once it had been completed.

After developing an initial list, SMEs evaluated the list against criteria for an effective LO. This led to refinement of the language, ensuring that the LOs were properly structured according to instructional design good practices and utilized the verbs in Blooms taxonomy. The final list of SME-proposed learning objectives can be seen in Figure 5.

- Explain the legal foundation of safeguards between the IAEA and its Member States (e.g. Comprehensive Safeguards Agreement and Additional Protocol)
- Describe the history, organizational structure, and the main activities of the IAEA
- Summarize the mission of the Department of Safeguards and the general objectives of International Nuclear Safeguards
- Describe the organizational structure of the Department of Safeguards
- Summarize the key core IAEA strategy documents

Figure 5 - SME-Proposed Learning Objectives

Criteria for Sufficient Generated Learning Objectives

Following the development of SME-proposed LOs, the SMEs created a more detailed list of the information a learner would be expected to know and understand after reading the materials and taking a course generated from the selected curated set (See Figure 6). For example, while the LOs stated that learners should be able to "Describe the organizational structure of the Department of Safeguards," the various divisions and offices of the Department of Safeguards

- IAEA Objective(s)
- Benefits of Safeguards
- Objectives of the IAEA Department of Safeguards
- Structure of the IAEA Department of Safeguards
  - Division of Concepts and Planning (SGCP)
  - Division of Information Management (SGIM)
  - Division of Technical Support (SGTS)
  - Office of Safeguards Analytical Services (SGAS)
  - Office of Information and Communications Services (SGIS)
  - Operations Divisions
    - There are three Operations Divisions
    - as well as the Office of Verification in Iran
- Responsibilities of the State
- Comprehensive Safeguards Agreement
- Additional Protocol
- Voluntary Offer Agreement
- Significant Quantity
  - SENSITIVE
- Facility-Type Agreement (INFCIRC/66)
  - Sometimes written as INFCIRC-66
- Small Quantities Protocols and modified Small Quantities Protocols
- Nuclear Material Accountancy
- Containment and Surveillance
- United States Support Program
- Integrated Safeguards
- Broader Conclusion
- State-Level Concept/State-Level Approach
- State Systems of Accounting for and Control of Nuclear Materials (SSAC)
- Noncompliance
- Peaceful Uses
- NPT Articles III, IV, VI
- Structure of IAEA
- Structure of the IAEA Department of Safeguards
- Secretariat
- Board of Governors
- Policy-Making Organs
- General Conference
- Nonproliferation Treaty (NPT)

Figure 6 - SME-Defined Checklist for Generated LOs

were spelled out in this list. Generated LOs and content were later benchmarked against this list to ensure that all vital information was covered in the course. After identifying specific topics of interest, safeguards SMEs identified expected associations between documents and LOs. This list can be found in Figure 7 below.

- Draw out the organizational structure of the IAEA (Org Chart);
- Explain the legal foundation and context of the IAEA (NPT);
- Summarize the main activities of the IAEA, the Policy-Making Organs, the Board of Governors, the Secretariat, and the General Conference (Primary: Introduction to International Safeguards, IAEA Factsheet and Statute, Secondary: Provisional Rules of Procedure of the Board of Governors);
- Summarize the IAEA Department of Safeguards (Org Chart and Introduction to International Safeguards);
- Summarize the Strategic Objectives of the Department of Safeguards (Development and Implementation Support Programme for Nuclear Verification and Long-Term R&D Plan);
- Outline the importance of international nuclear safeguards (Introduction to International Safeguards);
- Summarize the core strategy documents of the IAEA Department of Safeguards at a high-level or as an overview (Long-Term R&D Plan and Development and Implementation Support Programme for Nuclear Verification)

Figure 7 - Additional SME Expectations for Generated LOs

Finally, safeguards SMEs were encouraged to consider two questions: “What would make generated objectives as good as SME-produced objectives?” and "What would make generated objectives better than SME-produced objectives?" In answer to the first question, SMEs returned to the criteria for effective LOs that they used to evaluate their own draft learning objectives as well as the expertise that they were able to leverage in developing the objectives found in Figure 8.

- Objectives should be concrete, action-oriented, and measurable.
- Learning objectives should start with an action verb, such as “explain”, “summarize”, “describe”, “develop”, “apply” - avoid basic terms like “understand” or “know”
- Objectives should be meaningful, realistic, and practical to routine work conducted by safeguards professionals.
- Objectives adequately address the many facets and foundations of international nuclear safeguards.
- Objectives could take into account priorities of the Office of International Nuclear Safeguards (NA-241), such as implementation and universalization of the Additional Protocol and the statutory requirement of the US government to assist with international nuclear safeguards training. This also includes focuses on advanced reactors or small modular reactors and incorporating safeguards by design in the development of these novel technologies.
- Objectives should focus on the high-level takeaways of the documents and summarize where possible.
- Objectives should not be too focused on the research priorities for a specific timeframe, the objectives should be more generalist.
- Should the learners understand why the learning objective is important to their role in the safeguards framework – either from a national or international perspective - where possible.

Figure 8 - "What would make generated objectives as good as the SME-produced objectives?"

The answer to the second question was less straightforward. Nevertheless, the SMEs developed a list of qualities that would improve objectives and ultimately the content generated. Many of these track back to good practices in developing training materials such as the ability to distill the information in the documents and identify the biggest take-aways, remaining focused on those takeaways and not including too much information or going too in-depth for an introductory course, and tailoring objectives and learning materials to the needs of the individual (see Figure 9).

- If the objectives are able to grab the biggest take-aways from nearly all included documents (potentially excluding the Enhancing Capabilities for Nuclear Verification and Org Chart documents). Moreover, considering that there might be information 'overload' and to account for the high-level summaries stemming from the Safeguards Glossary, Long-Term Strategic Plan, Provisional Rules of Procedure of the Board of Governors documents.
- There may be the issue of 'too much information' and striking the balance between what is important and not for a summary – in the case of an 'introductory course' – is paramount.
- Focus on knowledge development, do not emphasize that learners would be an 'expert' following an introductory presentation for example.
- Not too focused on the many nuances of the legal documents, research directions, and other instruments, instead focusing on the overarching safeguards regime.
- Drawing conclusions that are interpreted and not necessarily literal – for example, the fact that the NPT is based on three pillars: nonproliferation, disarmament, and peaceful uses of nuclear energy – as the pillars of the nonproliferation regime.
- Basic: assume objectives will be for adult learners with limited knowledge of safeguards who need to have a basic understanding of the frameworks of international nuclear safeguards.
- Advanced: tailor objectives to specific roles that folks will be expected to conduct in the safeguards realm. Will the person taking the course try to become an inspector or an open-source analyst? Then the overview and its learning objectives could be modified slightly.

Figure 9 - "What would make generated objectives generated better?"

Generation of Learning Objectives

The first prompt template produced by the team channeled metadata—both for a given curated set and for the documents contained by that set—into a prompt for generating LOs and grouping those LOs into modules. The course, modules, and LOs were structured as a knowledge graph and generated using the XML dialect of W3C *Resource Description Framework / Extensible Markup Language* (RDF/XML). Once the initial curated set was selected, the test harness obtained the necessary metadata for that set to instantiate the prompt shown in

Appendix B: Prompt Template to Generate Baseline L.

The team selected OpenAI o4-mini as the LLM to generate the LOs based on its relative speed and the proficiency it has demonstrated with other Argonne projects for completing RDF/XML KGs. The pipeline channeled the prompt to this LLM, cleaned the output to isolate the completed RDF/XML document, and parsed that document with *Apache Jena* (https://jena.apache.org/) both to validate the document and to create the knowledge graph. A human-readable presentation of this KG is shown in Appendix A: Generated Baseline Learning Objectives.

Generation of Baseline Instructional Content

Feeding all of the documents from the curated set into an LLM at once would unnecessarily overload any of the LLMs to which the team had access. To prepare for generation of course content, each of these documents was *chunked* into coherent passages of a manageable size. Coherent passages were obtained using the *Grobid* parser (https://github.com/kermitt2/grobid), a machine-learning base parser trained to recognize structural elements of a discourse-oriented PDF. Passages were then processed sequentially for storage in an MSSQL *passages* table. For each passage, both the passage content and extracted context were stored in separate columns of a single row in this table. The table itself was indexed using MSSQL's full text indexing capability.

To generate course content, then, the pipeline traversed the KG defining the course's modules and LOs. A local filesystem directory was created for each module, and a separate html file was produced for each LO within that module.

The first step in creating instructional content for an LO was to select relevant document content from the passages table using a technique akin to *retrieval augmented generation*. In short, attributes for the LO, the module, and the course, are obtained from the KG and transformed into a query: a collection of terms paired with weights. Each term is a word or a phrase extracted from the obtained attributes. The weight associated with each term is used by MSSQL to prioritize rows matching the term. Rows matching terms with greater weight received proportionately higher priority.

SKR is driven by an *ontology*: a KG that defines and relates core structural and behavioral concepts. SKR integrates this with a SME-managed *taxonomy* which defines and relates hundreds of essential safeguards domain concepts. The pipeline for this work harnessed both the ontology and the taxonomy to identify phrases, acronyms, and synonyms to improve both the precision and recall of the search for passages relevant to a given LO.

Once the pipeline is in possession of passages for an LO, another prompt template channels these passages along with acquired attributes for the LO, module, and course to instantiate a prompt for generation of instructional content. The team selected OpenAI GPT-5 to generate this content based on its superior reasoning capabilities and large input capacity. In particular, the prompt instructs the LLM to produce instructional content that is engaging and professional looking. The resulting content includes information presentation and a variety of interactive learning activities—examples of which can be seen below in the Results and Findings section.

Instructional content for an LO is generated as HTML for use in a SCORM module, and this HTML is stored in the local file associated with the LO. This content contains a variety of learning activities

The pipeline employs a final prompt template to summarize the generated content—in accordance with module and course attributes—in two ways: as an introduction to the module and as a module summary. That is, the same prompt is used to generate both of these components, but the template varies the instructions to let the LLM know which of these components it is generating.

Once each of the modules was generated, it was compressed into a zip file for deployment to the LMS.

Deployment to Moodle

The initial course was created through the Moodle web application. Zip files for each of the four modules were then uploaded—also through the Moodle web application.

Results and Findings

Evaluation of Generated Learning Objectives

After generating LOs, the team conducted a practitioner-led evaluation to determine accuracy, relevance, and appropriateness for an introductory audience. Using the beginner-focused rubric described earlier, safeguards SMEs reviewed the material line-by-line for legal precision, terminology, scope, sequencing, and sensitivity. Overall, the LLM-produced LOs provided a useful scaffold. The generated LOs were found to have met all eight of the predefined conditions that would make generated LOs as good as SME-proposed LO—and six out of seven (86%) of the predefined conditions that would make them better.

Evaluation of the generated content, however, revealed multiple areas in which LOs should be adjusted and rebalanced to improve content generation. In particular, portions of Module 2 (policy-making procedures and organizational detail) and Module 3 (R&D emphasis and topic sensitivity) were "in the weeds" for newcomers and unlikely to affect the day-to-day work of most safeguards practitioners. Conversely, the modules that foreground the full legal framework and the breadth of verification activities—and clarify roles and responsibilities of the IAEA and the State—were missing key information and needed to be prioritized and expanded.

SMEs also noted that assigning any LO a difficulty level above "beginner" quickly pushed content into advanced territory that was not appropriate for an introductory course and did not meet the needs of the target audience. All targets needed to be capped at the beginner level to keep cognitive load appropriate for an introductory or foundational course.

It was also observed that differentiating between *Terminal Learning Objectives* (TLOs) – what the learner is meant to accomplish through the course - and *Enabling Learning Objectives* (ELOs) – how the learner was to accomplish the goal – would be a useful tool in further refining the content and ensuring that it remained within the correct scope.

Generated Instructional Content

After uploading each module's zip file to Moodle, course content became available to team SMEs for review (Figures 10-13).

When uploading the SCORM modules, the containing title for each module was simply set to "Module *x*" where x is the module number. Clicking on a module navigates the learner to the instructional content with that module.

Figure 10 - Four Modules Shown Within the Moodle Course

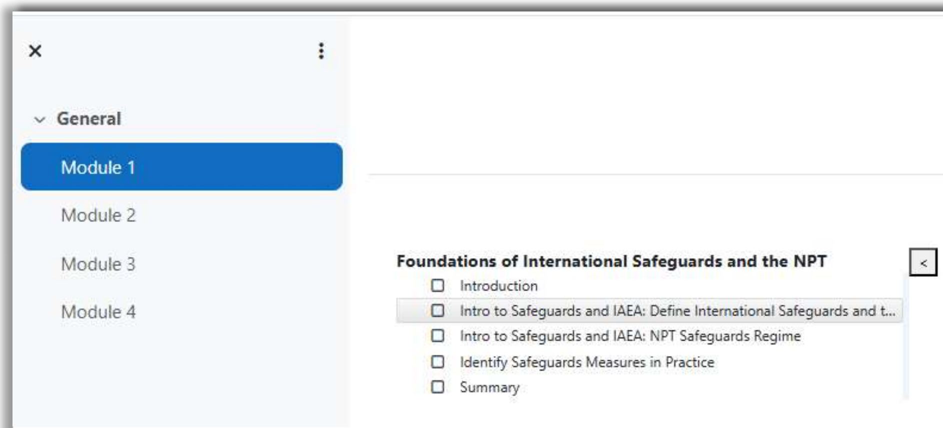

Figure 11 - Navigation Among and Within Modules

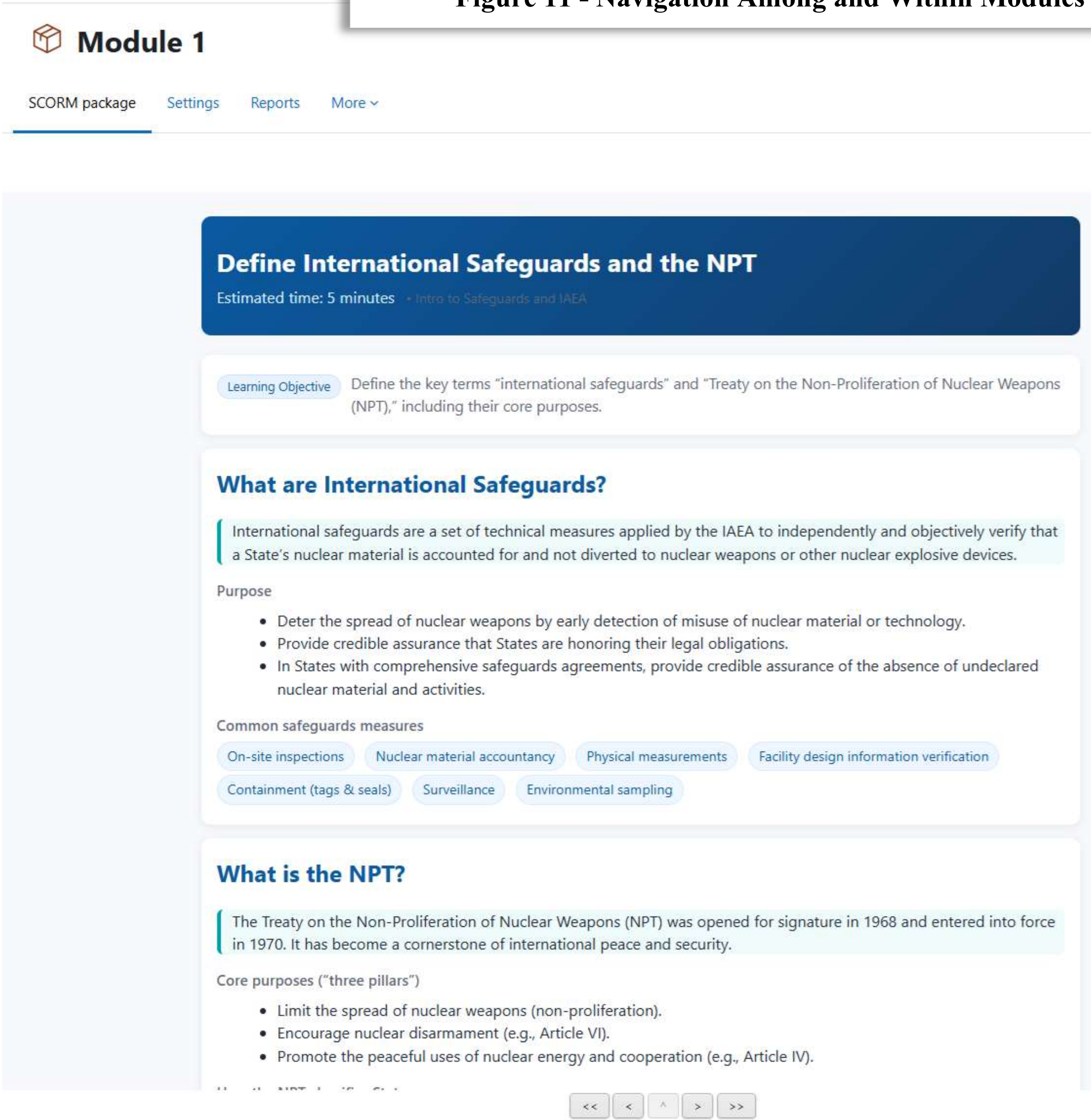

Figure 12 - First Lesson in First Module

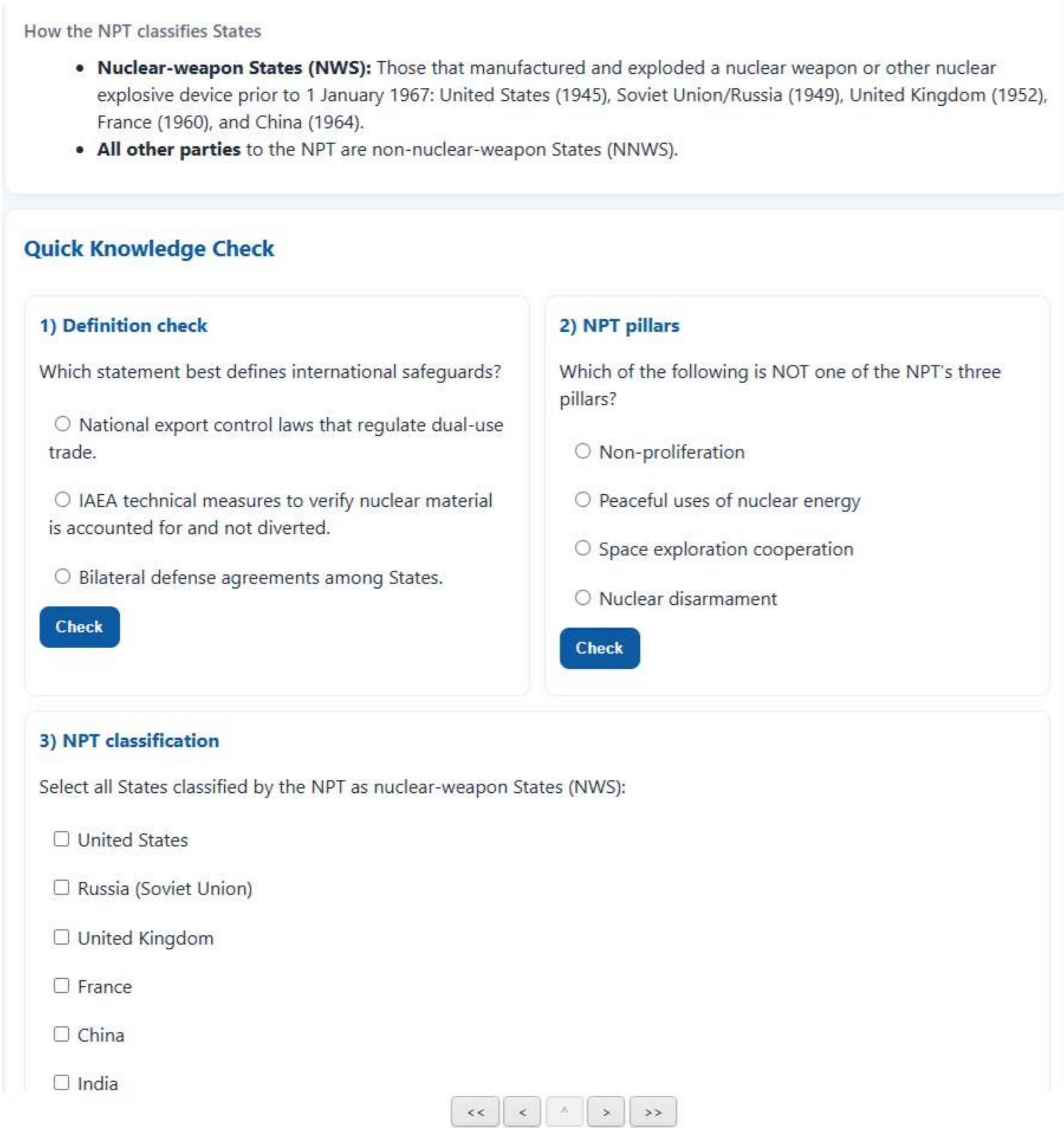

Figure 13 - A Variety of Learning Activities in the Generated Instructional Content

Evaluation

Evaluation of the course content focused on aligning content to the essentials that new safeguards practitioners "must know" and correcting subtle but consequential inaccuracies. Key improvements included: reframing the introduction to "international nuclear safeguards concepts" (plural); removing limitations that implied safeguards' purpose applies only to CSA States; correcting terminology (e.g., "DIV," not "facility DIV"); maintaining NPT in present tense and avoiding overreach into the "grand bargain" narrative; ensuring "peaceful uses" phrased as "applications"; and introducing define-before-use sequencing (e.g., presenting significant quantity before referencing it). The evaluation team incorporated missing instruments and identifiers—including explicitly citing the Model Additional Protocol as INFCIRC/540 and adding the mSQP—and removed mischaracterizations (e.g., VOA is not a complementary agreement). Practice activities and quizzes were simplified to match the beginner level and to ensure unambiguous alignment between scenarios and answer choices. Accessibility issues (e.g., low-contrast text) were also flagged and corrected.

SME-Augmented Learning Objectives

After evaluating generated LOs and content, SMEs iteratively refined the generated LOs to emphasize beginner-appropriate levels, precise terminology, and define-before-use sequencing. The augmented set preserved areas where the model performed well—foundational concepts, the high-level legal instruments (NPT; CSAs/INFCIRC/153 (Corr.); AP/INFCIRC/540; mSQP), and the basic structure of the IAEA—while correcting subtle but consequential issues. In particular, SMEs clarified or added missing definitions (e.g., complementary access), standardized terms (e.g., replacing "voluntary offer" with the correct "*Voluntary Offer Agreement"* (VOA)), and rebalanced objectives to avoid drifting into intermediate or advanced territory. As a result, definitions are clearer, omissions are reduced, and the objectives better reflect what entry-level safeguards practitioners should learn about the field and work.

SMEs also tightened flow and removed duplications that could confuse learners, especially where the draft overlapped coverage of agreements across adjacent sections. The SMEs also in later iterations caught extraneous examples (e.g., an "exports of scrap" case that was out of scope for an introductory course) and flagged ambiguous statements (e.g., "Immediate feedback reinforces correct reasoning") that did not map to outlined criteria nor further learning objectives. The augmented and enhanced objectives – after two thorough evaluations – now more directly align to practitioner priorities, sequencing legal foundations and core verification activities before introducing organizational and resource topics. In short, the tandem model—LLM drafting with SME review and augmentation—demonstrates clear promise.

During the process of augmenting the generated LOs, it was difficult to keep the TLO at the appropriately high level. To correct some of the deficiencies that they saw in the content, the SMEs tended toward adding wording that would be more appropriate for an ELO. Ultimately, the need for ELOs was identified as needing to be addressed in a future iteration of this work.

Evaluation of Resulting Instructional Content

SMEs reviewed the first round revised SCORM-packaged modules for accuracy, level, and alignment to the augmented objectives. The model's narrative overview of the NPT safeguards regime was notably strong and provided a clear, beginner-friendly through-line. However, several content elements and practice items drifted beyond the intended level—e.g., too detailed procedural comparisons of the Board of Governors and General Conference, fine-grained numbering from resource prioritization documents, and scenario quizzes invoking terms not yet defined. SMEs recommended simplifying comparisons to "at-a-glance" essentials, leading Module 3 with a high-level introduction to core safeguards documents and ensuring glossary-driven activities follow explicit introductions to basic safeguards concepts, not advanced ones. They also advised using department acronyms and full names consistently (e.g., SGIM, SGTS) and standardizing legal terms (e.g., consistently "Voluntary Offer Agreement").

The evaluation further identified places where the content assumed practitioner knowledge, missed define-before-use sequencing, or introduced advanced or sensitive material prematurely. Corrections included clarifying complementary access at a conceptual level, removing out-of-scope examples, and tailoring terminology to entry-level roles with suggestions in future adaptations to 'adapt by track' in specific areas within sub-modules (e.g., tailoring towards technical vs. policy learners). Several items should also be streamlined or deferred (e.g., mentorship planning) to preserve focus on foundational knowledge and beginner-appropriate practice.

A pipeline issue that had allowed content order to drift from LO order was corrected, and a second round of revised SCORM-packaged modules was generated with the augmented LOs, and with these adjustments—and modest flow fixes to reduce duplication—the resulting instructional content is materially stronger than the first draft: clearer, more coherent, and better aligned to the essentials newcomers must master.

Comparison of LOs: Generated, SME-produced, and SME-Augmented

The generated LOs provided broad coverage of fundamentals and a fair outline of the IAEA's structure and core legal instruments; yet they sometimes missed practitioner priorities, introduced undefined terms (e.g., complementary access), and wandered above the beginner level in examples and assessments. The SME-produced LOs, by contrast were more practitioner-centered and tightly aligned—but slower to develop and limited by the time available for authoring. The SME-augmented set marries the speed of generation with the precision of practice: objectives retain the model's structure while incorporating SMEs' refinements to terminology, scope, and sequencing, including corrections and defining key concepts before use.

Across the developed sets, differences are most evident in measurability, level, and relevance. Generated objectives tended to bundle verbs and assume context, which complicated assessment design and occasionally pushed cognitive demands past "beginner." SME-produced objectives were more clear and measurable but required significant authoring effort. SME-augmentation closed these gaps—tightening criteria and conditions, removing duplications, and reordering content to emphasize the legal framework and core verification activities first. The result is a harmonized, beginner-appropriate path that minimizes omissions, improves definitional clarity, and better reflects "what is really important" to practitioners, while preserving the acceleration benefits of LLMs. This comparison underscores a durable conclusion for safeguards coursework: LLMs can effectively bootstrap structure and coverage, but SMEs remain essential to ensure accuracy, appropriateness, and usable learning experiences.

Evaluation

After revision, the LOs and content present a clearer, beginner-appropriate path through the legal framework, the spectrum of verification activities, and the practical roles of the IAEA and States. The resulting sequence reduces cognitive load, avoids sensitive or unnecessarily detailed material, and better maps to the needs of interns, fellows, and early-career practitioners, while preserving room for future advanced modules that can build on this foundation.

However, the revised materials also further supported the need for SME review, both of the initial draft and second draft of the content materials. There is no question that AI-generated content is useful for reducing the time needed to create learning-level appropriate materials, but it is not able to fully eliminate the need for validation by a safeguards expert.

Strengths and Limitations of this Approach to Course Generation

This approach using LLMs can rapidly bootstrap a coherent course structure, propose draft learning objectives, and populate modules with plausible definitions, activities, and summaries. That speed and breadth are valuable for leveraging curated sets of safeguards relevant documents, especially when time and staff are constrained. However, the drafts reliably reveal limitations that matter in safeguards education: subtle but consequential terminology errors (e.g., misnaming terms, implying limited purposes of specific legal instruments), gaps and omissions in the legal framework (e.g., missing key documents; not citing important acronyms or terms), uneven treatment of sensitive topics, pedagogical sequencing issues (e.g., invoking the use of a term before defining it), and occasional overreach into organizational detail not useful for most practitioners. Accessibility and usability details (e.g., contrast, readability) and unsubstantiated rationales (e.g., "cost efficiency" for export-state measures) further underscore the use of expert curation somewhere in the review process before deploying for full use. In short, LLMs are effective accelerators, but without practitioner review they can entrench inaccuracies, dilute priorities (e.g., CSA/AP emphasis), and create training that is either too shallow in the right places or too deep in the wrong ones.

Recommendations and Next Steps

Recommendations include adopting a structured, expert-curated workflow, to include having the LLM produce an outline mapped to explicit learning objectives (both terminal and enabling) and competency statements first, then conduct a practitioner scoping pass to prioritize key documents, identify missing topics, and setting sensitivity. This human-in-the-loop approach preserves the speed benefits of LLMs while ensuring the accuracy, relevance, and usability needed by safeguards practitioners.

Conclusion

Automation of course generation should proceed only after a vetted curated set and a concise coverage rubric are in place to ensure complete treatment of the legal framework (NPT; CSAs/INFCIRC/153 (Corr.); AP/INFCIRC/540; mSQP) and the core verification activities (nuclear material accountancy, design information verification, containment and surveillance, environmental sampling, information analysis, complementary access). All objectives and content should be constrained according to the goal of the course being generated. For example, if a course is introductory, then objectives should be capped at beginner level with sensitivity and scope controls applied to exclude highly sensitive topics and avoid unnecessary procedural detail.

Indeed, the use of a safeguards subject matter expert throughout the workflow for safeguards course generation is recommended. While the use of an LLM can rapidly draft outlines and modules mapped to clear learning objectives, a safeguards SME is useful to validate and refine the content. The practitioner ensures precise terminology, legal completeness, and appropriate handling of sensitive topics, sound pedagogical sequencing, and usability for users. This tandem model preserves the speed and breadth of LLM drafting while safeguarding accuracy, relevance, and alignment with priorities. In practice, this approach has the potential to drastically reduce the time a practitioner spends compared to authoring new course content from scratch, while improving quality and consistency across modules.

To sustain quality and alignment, learning objectives should follow a standardized schema (clear Bloom's taxonomy verbs, concise success criteria, simple conditions), with define-before-use sequencing and terminology aligned to the Safeguards Glossary. SME engagement should remain integral at three checkpoints—LO validation, module outline and sequencing, and iterative expert assessments. Source-grounded citations tied to the curated documents, retrieval quality assurance, and explicit abstention when evidence is insufficient are required, followed by a pilot evaluation with SMEs using an objective rubric.

Finally, packaging and usability checks for SCORM/Moodle, version control and governance for updates, and documented acceptance criteria should be established so SMEs transition from primary authors to final approvers. With these controls and ongoing SME tuning automation can deliver speed and consistency without compromising accuracy, relevance, or learner appropriateness, and will effectively support instructional designers and trainers developing safeguards coursework.

Acknowledgements

This material is based upon work supported by *Laboratory Directed Research and Development* (LDRD) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357

The authors would also like to thank Laura Christiansen for software development, research, and editing assistance, Matthew Swindall for research and editing assistance, Douglas Johnson for data management and editing assistance, Joe Petersen for systems administration assistance, and both Julie Muzzarelli and Ike Therios for review and editing assistance.

Appendix A: Generated Baseline Learning Objectives

Course : Intro to Safeguards and IAEA: IAEA Structure, Scope and Core Documents

This course introduces interns and new hires to international safeguards, focusing on the IAEA's structure, scope, and core documents. Over four 30-minute modules, learners will explore the history, legal foundation, governing bodies, strategic plans, and essential resources of the IAEA and its Department of Safeguards.

Module 1: Foundations of International Safeguards and the NPT

Introduce the concept of international nuclear safeguards and the framework of the Treaty on the Non-Proliferation of Nuclear Weapons (NPT).

Knowledge Learning Objective 1.1: Define international safeguards and the NPT	
Learn the definitions and purposes of nuclear safeguards and the Treaty on the Non-Proliferation of Nuclear Weapons.	
Estimated Minutes	5
Difficulty Level	beginner
Criteria	Correctly state the purpose of safeguards and list at least two obligations of NPT signatories
Condition	Using the NPT treaty text and NPAC Introduction to International Safeguards document
Behavior	Define key terms related to international safeguards and the NPT

Comprehension Learning Objective 1.2: Explain the safeguards regime established by the NPT	
Interpret how the NPT framework supports IAEA safeguards activities.	
Estimated Minutes	8
Difficulty Level	beginner
Criteria	Provide a coherent summary of at least three NPT provisions that enable safeguards
Condition	Given excerpts from the NPT treaty and NPAC Introduction document
Behavior	Explain the relationship between the NPT and the IAEA safeguards system
Prerequisite	1.1

Application Learning Objective 1.3: Identify safeguards measures in practice	
Apply knowledge of safeguards to real-world inspection scenarios	
Estimated Minutes	7
Difficulty Level	beginner
Criteria	Correctly pair at least three examples with the appropriate safeguards measure
Condition	Using case examples from the NPAC Introduction and IAEA Safeguards Brochure
Behavior	Define key terms related to international safeguards and the NPT
Prerequisite	1.2

Module 2: IAEA Governance and Organizational Structure

Examine the IAEA's policy-making bodies and organizational chart to understand decision-making processes.

Knowledge Learning Objective 2.1: Recall roles of IAEA policy-making organs	
Identify the functions of the General Conference, Board of Governors, and Secretariat.	
Estimated Minutes	6
Difficulty Level	beginner
Criteria	Accurately name at least three organs and describe one key function of each
Condition	Using the IAEA Policy-Making Organs Factsheet
Behavior	List the main IAEA policy-making organs and their primary roles

Analysis Learning Objective 2.2: Compare the Board of Governors and General Conference	
Analyze decision-making differences between the two bodies.	
Estimated Minutes	10
Difficulty Level	intermediate
Criteria	Identify at least two procedural or membership differences
Condition	Given the Board of Governors Rules of Procedure and Org Chart
Behavior	Contrast the roles and voting procedures of the Board of Governors and the General Conference
Prerequisite	2.1

Application Learning Objective 2.3: Interpret the IAEA organizational chart	
Map the reporting lines within the Department of Safeguards	
Estimated Minutes	8
Difficulty Level	intermediate
Criteria	Correctly list at least three units and their reporting relationships
Condition	Using the IAEA Org Chart 2020
Behavior	Identify which divisions report to the Department of Safeguards

Module 3: Core IAEA Safeguards Documents and Strategies

Review key documents guiding IAEA safeguards policy, research, and resource mobilization.

Knowledge Learning Objective 3.1: List IAEA core safeguards documents	
Recognize the range of foundational IAEA publications relevant to safeguards	
Estimated Minutes	6
Difficulty Level	beginner
Criteria	Correctly name at least five documents and categorize each by type (statute, strategy, glossary, etc.)
Condition	Given access to IAEA Statute, Safeguards Brochure, Glossary, *Development and Implementation Support Programme* (D&ISP) 2022-2023, STR-399, and Long-Term R&D Plan
Behavior	List major IAEA safeguards documents

Comprehension Learning Objective 3.2: Summarize the Long_Term R&D Plan objectives	
Explain the primary goals and milestones of the Department of Safeguards R&D plan.	
Estimated Minutes	8
Difficulty Level	intermediate
Criteria	Identify at least three strategic objectives and corresponding milestones
Condition	Using the IAEA Department of Safeguards Long-Term R&D Plan 2012-2023
Behavior	Summarize key objectives and milestones of the Long-Term R&D Plan
Prerequisite	3.1

Evaluation Learning Objective 3.3: Evaluate alignment of STR-399 with the R&D Plan	
Assess how the Resource Mobilization Priorities report supports the Long-Term R&D objectives.	
Estimated Minutes	10
Difficulty Level	advanced
Criteria	Produce a brief comparison noting at least two areas of alignment or gaps
Condition	Given STR-399 and the Long-Term R&D Plan documents
Behavior	Evaluate consistency between STR-399 priorities and the Long-Term R&D Plan
Prerequisite	3.2

Module 4: Professional Resources and Glossary Application

Utilize professional guides and the glossary to deepen understanding and plan further study.

Knowledge Learning Objective 4.1: Locate definitions and further readings	
Navigate the Safeguards Glossary and Resource List to find key terms and resources.	
Estimated Minutes	5
Difficulty Level	beginner
Criteria	Find definitions for three technical terms and list two recommended resources
Condition	Using the IAEA Safeguards Glossary 2022 Edition and Safeguards Resource List_PNNL
Behavior	Retrieve definitions and associated resources

Application Learning Objective 4.2: Apply glossary terms in a scenario	
Use glossary definitions to clarify technical language in a hypothetical safeguard event.	
Estimated Minutes	7
Difficulty Level	intermediate
Criteria	Correctly define and use three glossary terms in context
Condition	Given a description of an undeclared nuclear activity
Behavior	Explain technical terms within a given scenario
Prerequisite	4.1

Synthesis Learning Objective 4.3: Design a mentorship resource plan

Combine professional resources to outline a traning plan for new safeguards staff.

Estimated Minutes	10
Difficulty Level	advanced
Criteria	Include at least four resources aligned with two specific training objectives
Condition	Using the Development and Implementation Support Programme for Nuclear Verification 20 2023 and the Safeguards Resource List
Behavior	Create an outline mapping resources to learning goals
Prerequisite	4.2

Appendix B: Prompt Template to Generate Baseline LOs

You are designing an online course for interns and new hires who aspire to become nuclear safeguards professionals.
The course should be scoped so that the learner can learn content and demonstrate proficiency in learning objectives over a small number of 30-minute modules.
The course is titled, "${title}".
The course is based on a curated set of documents which is described as follows: ###
${description}
###
The course will only draw content from these documents, and these documents are listed below.

${content-documents}

Please complete the RDF/XML knowledge graph below by providing a list of learning objectives (as defined by Bloom's Taxonomy) in a learnable order for this course.
For each module,
1. add a skr:Module instance with dc:title and dc:description properties.
2. the skr:Module instance should also include a skr:index property containing a module number: this should be 1 for the first module and incremented by 1 for each successive module.
3. add to the course instance a skr:module property referencing the module instance.
4. for each of the module's learning objectives,
   a. add an instance of the appropriate learning module class. Classes of learning objectives are,
      bloom:KnowledgeLearningObjective,
      bloom:ComprehensionLearningObjective,
      bloom:ApplicationLearningObjective,
      bloom:AnalysisLearningObjective,
      bloom:SynthesisLearningObjective, and
      bloom:EvaluationLearningObjective.
   b. The instance should include dc:title and dc:description elements.
   c. The instance should also include a skr:index property for the learning objective index. This should be a hierarchical dotted identifier starting with the module number, so the first learning objective will have an index of "1.1".
   d. The instance should also include a bloom:behavior property describing what this learning objective should enable the user to do.
   e. The instance should also include a bloom:condition property for each condition, constraint, or contextual assumption under which the above behavior will occur.
   f. The instance should also include a bloom:criteria property for each measure the learner must meet in to demonstrate competence in the learning objective.
   g. The instance should also include a bloom:difficultyLevel property with one of the following literal values:
      "beginner", "intermediate", or "advanced".
   h. The instance should also include a bloom:estimatedMinutes property providing the estimated time required to attain competence.
   i. If the learning objective is part of a larger learning objective, add a skos:broader property referencing the larger objective.
   j. If the learning objective requires the learner attain competence in a prerequisite learning objective, add a bloom:prerequisite property referencing the prerequisite objective.
   k. add to the module instance a skr:objective property referring the learning objective instance.

Finally, always be sure to properly escape ampersands and other XML delimiters with the appropriate entity.
###
```
<rdf:RDF xmlns="https://skr.anl.gov/generated/"
         xmlns:bloom="http://bloom.nsis.anl.gov#"
   xmlns:skr="http://skr.nsis.anl.gov/vocabulary/1.0#"
   xmlns:dc="http://purl.org/dc/elements/1.1/"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:prov="http://www.w3.org/ns/prov#"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
   xmlns:schema="http://schema.org/"
   xmlns:sem="http://semanticweb.cs.vu.nl/2009/11/sem/"
   xmlns:time="http://www.w3.org/2006/time#"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
   xmlns:skos="http://www.w3.org/2004/02/skos/core#"
   xmlns:skr-system="https://skr.nsis.anl.gov/"
   xmlns:skr-contents="https://skr.nsis.anl.gov/contents/"
   >
         <skr:Course rdf:about="https://skr.anl.gov/generated/course-1">
                  <dc:title>${title}</dc:title>
                           <dc:description><!-- derive a course description from the curated document set description--
         ></dc:description>
                  <!-- add skr:module property for each module here -->
         </skr:Course>

         <!-- add modules and learning objectives here -->
 </rdf:RDF>
```

References

About NNSA. (n.d.). Energy.Gov. Retrieved September 16, 2025, from https://www.energy.gov/nnsa/about-nnsa

Acharya, Anurag, Sai Munikoti, Aaron Hellinger, Sara Smith, Sridevi Wagle, and Sameera Horawalavithana. 2023. "NuclearQA: A Human-Made Benchmark for Language Models for the Nuclear Domain." arXiv:2310.10920. Preprint, arXiv, October 17. https://doi.org/10.48550/arXiv.2310.10920.

Anderson, L. W. (Ed.). (2009). A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives (Abridged ed., [Nachdr.]). Longman.

Bloom, B. S. (1956). Taxonomy of Educational Objectives - The Classification of Educational Goals Handbook 1: Cognitive Domain". Longmans, Green and Co.

Bloom's Taxonomy. (n.d.). Retrieved April 2, 2025, from https://fctl.ucf.edu/teaching-resources/course-design/blooms-taxonomy.

Fan, Haoxiang, Guanzheng Chen, Xingbo Wang, and Zhenhui Peng. 2024. "LessonPlanner: Assisting Novice Teachers to Prepare Pedagogy-Driven Lesson Plans with Large Language Models." Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology (New York, NY, USA), UIST '24, October 11, 1–20. https://doi.org/10.1145/3654777.3676390.

H5P. (n.d.). https://docs.moodle.org/500/en/H5P

Hauser, Jakob Elias. 2025. "HiST-LLM." Zenodo, January 16. https://doi.org/10.5281/ZENODO.14671247.

Hu, Bihao, Longwei Zheng, Jiayi Zhu, Lishan Ding, Yilei Wang, and Xiaoqing Gu. 2024. "Teaching Plan Generation and Evaluation With GPT-4: Unleashing the Potential of LLM in Instructional Design." IEEE Transactions on Learning Technologies 17: 1445–59. https://doi.org/10.1109/TLT.2024.3384765.

Monteiro, João, Pierre-André Noël, Étienne Marcotte, et al. 2024. REPLIQA: A Question-Answering Dataset for Benchmarking LLMs on Unseen Reference Content.

Moodle. (n.d.-a). About Us: Making quality online education accessible for all. Moodle. Retrieved September 16, 2025, from https://moodle.com/about/

Moodle. (n.d.-b). Moodle Key Terms. https://docs.moodle.org/500/en/Moodle_key_terms

Ni, Jinjie, Fuzhao Xue, Xiang Yue, et al. 2024. "MixEval: Deriving Wisdom of the Crowd from LLM Benchmark Mixtures." arXiv:2406.06565. Preprint, arXiv, October 12. https://doi.org/10.48550/arXiv.2406.06565.

Nonproliferation. (n.d.). Energy.Gov. Retrieved September 28, 2025, from https://www.energy.gov/nnsa/nonproliferation

Oak Ridge National Laboratory. (n.d.). NPAC gAXIS. NPAC G. Retrieved September 16, 2025, from https://npac.gaxis.ornl.gov/login/index.php

Rajpurkar, Pranav, Robin Jia, and Percy Liang. 2018. "Know What You Don't Know: Unanswerable Questions for SQuAD." arXiv:1806.03822. Preprint, arXiv, June 11. https://doi.org/10.48550/arXiv.1806.03822.

Razafinirina, Mahefa Abel, William Germain Dimbisoa, and Thomas Mahatody. 2024. "Pedagogical Alignment of Large Language Models (LLM) for Personalized Learning: A Survey, Trends and Challenges." Journal of Intelligent Learning Systems and Applications 16 (4): 4. https://doi.org/10.4236/jilsa.2024.164023.

Sridhar, Pragnya, Aidan Doyle, Arav Agarwal, Christopher Bogart, Jaromir Savelka, and Majd Sakr. 2023. "Harnessing LLMs in Curricular Design: Using GPT-4 to Support Authoring of Learning Objectives." arXiv:2306.17459. Preprint, arXiv, June 30. https://doi.org/10.48550/arXiv.2306.17459.

White, Colin, Samuel Dooley, Manley Roberts, et al. 2025. "LiveBench: A Challenging, Contamination-Limited LLM Benchmark." arXiv:2406.19314. Preprint, arXiv, April 18. https://doi.org/10.48550/arXiv.2406.19314.

Yin, Xunjian, Xu Zhang, Jie Ruan, and Xiaojun Wan. 2024. "Benchmarking Knowledge Boundary for Large Language Models: A Different Perspective on Model Evaluation." arXiv:2402.11493. Preprint, arXiv, May 29. https://doi.org/10.48550/arXiv.2402.11493.

Clark, R. C., & Mayer, R. E. (2016). e-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning (4th ed.). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119239086

Freifeld, L. (2024, November 20). 2024 Training Industry Report. Training. https://trainingmag.com/2024-training-industry-report/

90+ Employee Training Statistics in 2025: Investing in Employee Development. (2025). Shortlister. https://www.myshortlister.com/insights/employee-training-statistics

He, Symon. (2021). *How much time do I need to create an online course?* Six Figure Instructor. https://sixfigureinstructor.com/learning-center/how-much-time-needed-to-make-online-course/

Strategic Security Sciences
Argonne National Laboratory
9700 South Cass Avenue
Lemont, IL 60439

www.anl.gov

Argonne National Laboratory is a U.S. Department of Energy
laboratory managed by UChicago Argonne, LLC