Policy Considerations for Mitigating Risks Associated with Developing Training Using Generative AI

Dawn Norman, MBA, PMP, CSPO, LBBP, SSCC and Lee W. Lacy, Ph.D., CMSP
Soar Technology, LLC an Accelint Company
Orlando, Florida, USA
Dawn.Norman@SoarTech.com, Lee.Lacy@SoarTech.com

ABSTRACT

New innovative generative AI (GenAI) tools and techniques are revolutionizing training development approaches. Training development organizations are beginning to employ GenAI to support Front-End Analysis (FEA) and develop and to help maintain military training courseware. While these new technologies provide a significant Return on Investment (ROI), they also introduce new challenges that stakeholders should address through risk management policies and procedures.

This paper identifies relevant challenges and recommends risk management policies and procedures for courseware development. Instructional Systems Designers (ISDs) rely on current authoritative data sources to author valid courseware. They must ensure that the data can be used for the intended purposes. As ISDs begin to use GenAI systems that ingest this data, policies are needed to enforce data constraints and to monitor changes in the data that impact developed courseware. GenAI tool developers and users must respect security policies and legal, moral, and ethical issues. Sensitive data such as vendor proprietary information, Controlled Unclassified Information (CUI), distribution restricted data, and classified information must be protected. Policies must define controls and approval processes for using GenAI in courseware development.

Governance frameworks (including model and system cards) should outline which AI models may be used, how outputs are reviewed, and the responsibilities of ISDs in ensuring accuracy and compliance. As GenAI technologies evolve, serious concerns exist regarding the quality of the content they generate. GenAI systems are infamous for mistakes and hallucinations. Courseware development processes must help ensure that ISDs and trainees are provided with verified and validated information. AI-generated content should adhere to a unified tone, style, and terminology to maintain consistency, ensuring "one voice" across all courseware materials. This is critical for preserving instructional coherence and credibility. GenAI systems should produce outputs that comply with style guides and provide trainees with content at appropriate reading and proficiency levels. Potential quality assurance (QA) solutions include automating QA-related processes and workflows to detect inconsistencies, enforce stylistic uniformity, and ensure alignment with established instructional standards. New GenAI-related training development policies and procedures will help fulfill the promise of exciting new cost and time-saving improvements. However, to fully realize these benefits, organizations must establish oversight mechanisms and enforce adherence to established policies. The challenges must be well understood, and innovative policies and procedures must be enacted and followed.

ABOUT THE AUTHORS

Dawn Norman is a Senior Program Manager at Soar Technology, LLC (SoarTech), currently supporting the development and use of AI tools for training. Dawn retired from the U.S. Air Force in 2014. After retirement, Dawn began a career in DoD contracting as an Instructional Designer. Dawn has an undergraduate degree in Workforce Education and Development from Southern Illinois University, a Master of Business Administration degree from the University of North Alabama, and she is currently a doctoral candidate at Liberty University studying Educational Leadership.

Lee Lacy is a Senior Program Manager at SoarTech, currently supporting training development contracts that leverage generative AI. Dr. Lacy began his career in the mid-1980s working on U.S. Air Force research involving automated production of training hypermedia and has since worked on many modeling, simulation, and training contracts for DoD customers, including the U.S. Air Force, U.S. Navy, U.S. Army, DARPA, and ONR. Lee's Ph.D. is in Modeling and Simulation from the University of Central Florida (UCF).

Policy Considerations for Mitigating Risks Associated with Developing Training Using Generative AI

Dawn Norman and Lee W. Lacy, Ph.D., CMSP
Soar Technology, LLC an Accelint Company
Orlando, Florida, USA
Dawn.Norman@SoarTech.com, Lee.Lacy@SoarTech.com

INTRODUCTION

Solution developers are employing artificial intelligence (AI) foundation model technologies to build generative AI (GenAI) applications. Developers train foundation models on an extensive set of data using deep neural networks and self-supervised learning techniques. Large Language Models (LLMs) are foundation models focused on using vast amounts of text data to process text prompts and generate natural language responses. Supported applications (e.g., ChatGPT) interface with LLMs using natural language queries called prompts to perform a wide variety of tasks (Bommasani, 2021). GenAI tools can also generate images and videos.

GenAI technologies support training and education-related applications. Training organizations often seek to address budget and schedule challenges by increasing the productivity of ISDs that need to develop accurate and effective training that satisfies learning objectives. Courseware development tools can leverage foundation models trained on multiple data sources to mimic an understanding of domain subject matter and pedagogical techniques (Bommasani, 2021). These tools generate instructional content, including automatic item generation (AIG) for testing. The generated content can serve generalized audiences or help personalize instruction for specific trainees (Shires, 2024)

Instructional developers are beginning to employ GenAI-empowered tools to develop instructional materials. For example, the U.S. Navy's Ready Relevant Learning (RRL) initiative is producing a Basic Electronics and Electricity (BEE) course using this technology (Lacy, 2025). Examples of these tools include Aptima's NAUTILUS tool and SoarTech's Content Accelerator tool.

Despite expansive claims about the potential of foundation models to transform teaching and learning, historical patterns in educational technology adoption suggest that such optimism should be approached with caution (Blodgett, 2021). The use of GenAI in courseware development raises significant concerns regarding validity, reliability, transparency, fairness, and equity, especially in high-stakes military training contexts (Bulut, 2024).

Large language models (LLMs), when used in training environments, may introduce risks ranging from instructional misalignment and content bias to system vulnerabilities and unintended pedagogical consequences (Blodgett, 2021; Bommasani, 2021). Potential Impacts of these risks materializing include damage to equipment or humans resulting from low-quality or negative training; low scores, information not retained, longer training cycles; and CUI or classified data leakage. Training organizations should implement risk mitigation strategies to reduce these potential impacts of problems arising from the use of GenAI for developing instructional content.

LEVERAGING THE NIST AI RMF

The National Institute of Standards and Technology (NIST) Artificial Intelligence Risk Management Framework (AI RMF 1.0) provides a useful foundation for this approach, emphasizing four key functions: "Map", "Measure", "Manage", and "Govern" (NIST, 2023). These functions provide a lens for evaluating how training development organizations may assess GenAI tools and implement enforceable oversight practices.

Recent policy developments, such as the European Commission's ethical AI guidelines and the Organization for Economic Co-operation and Development (OECD) emphasis on transparency in high-stakes AI use, highlight a global recognition of the need for robust, context-specific governance in educational applications (Bulut, 2024). Noroozi (2024) further emphasizes the importance of embedding ethical guidelines and human oversight to mitigate risks related to bias, accuracy, and privacy.

Training organizations employing GenAI tools for developing courseware can leverage the NIST AI RMF (NIST, 2023) and the companion GenAI Profile (Autio, 2024). This framework structures risk mitigation efforts for AI in a manner like the RMF used to address cybersecurity. Applicable AI RMF tasks include:

1. Establish risk management policies and procedures (initial AI RMF “Governance” tasks),
2. Identify and analyze/assess potential risks (AI RMF “Map” and “Measure” tasks),
3. Prioritize and address risks using mitigation strategies (AI RMF “Manage” tasks), and
4. Manage ongoing risk mitigation efforts (ongoing AI RMF “Governance” tasks).

Training organizations should leverage their existing risk management approaches to document AI-specific policies and procedures. Documented processes should clearly define personnel roles and responsibilities and the specific, actionable tasks that personnel will perform. The organization’s leadership should approve and endorse the policies and procedures and require staff to participate in related training, including AI Literacy (Walter, 2024).

Processes should address the complete lifecycle of activities associated with developing and deploying courseware generated with GenAI-empowered tools. AI tools for educational purposes must be thoughtfully designed, deployed, and monitored to ensure they are safe, robust, transparent, explainable, and responsible (Bulut, 2024). The NIST AI RMF Framework document identifies key stages in the application development lifecycle that include:

- Planning and designing,
- Collecting and processing data,
- Building the AI model,
- Verifying and validating the AI model,
- Deployment and use, and
- Operating and using.

Training organizations should identify the potential risks associated with using GenAI to develop training content and assess the potential impact if those risks materialize. Examples of those risks are described in the “Risk Categories and Concerns” section below.

Training organizations should prioritize identified risks. Risk mitigation priorities will vary based on the organization and its content. For example, misinformation in origami training for elementary school children has a significantly less impact than misinformation in training aircraft mechanics. Once prioritized, training organizations can develop risk mitigation strategies to address high-priority risks. Examples of relevant risk mitigation approaches are described in the “Potential Risk Mitigation Strategies” section below.

Ongoing management of the risk management efforts should include tracking and reporting risk indicator metrics. On a regular basis, the organization should review and update the processes based on lessons learned and changes to the AI technologies.

RISK CATEGORIES AND CONCERNS

Risk identification and analysis involving the use of GenAI for the generation of courseware and associated artifacts is likely to result in concerns that fall into the following categories:

- Inaccurate generated content,
- Ineffective training, and
- Improperly Handled Information.

Risk of Inaccurately Generated Content

LLMs may produce text that appears fluent and coherent yet contains factual inaccuracies or fabricated information, a phenomenon commonly known as “hallucinations" (Kasneci, 2023). These errors occur because LLMs generate content based on statistical patterns, rather than verified knowledge or a comprehensive understanding of the subject matter (Walter, 2024). In the context of military training, hallucinated content poses a significant risk, as it may introduce misleading procedures, misrepresent technical specifications, or provide dangerously incorrect information

(Eastgate Software, 2023; Hall, 2025). Even when GenAI tools use quality-controlled data, they may mistakenly combine unrelated facts or generate plausible but inaccurate assertions. These inaccuracies may propagate into training materials, leading to learner confusion or negative training that results in operational errors.

GenAI systems can potentially introduce biased perspectives into their outputs (Schatz, 2024). GenAI models can reflect and amplify the biases embedded in their training data (Center for Teaching and Learning, n.d.). These biases may manifest in subtle ways, such as stereotypical characterizations, gendered or culturally biased language, or unbalanced representation of concepts or practices. In the military training context, bias poses a dual threat: it may erode learners' trust and credibility in the training content, and more critically, affect operational readiness by skewing the representation of tactics, procedures, or mission-critical concepts. If GenAI outputs favor one doctrinal perspective over another, the resulting courseware may offer an incomplete or distorted instructional view.

ISD and trainee trust of the results of GenAI systems may be unwarranted (Blodgett, 2021). Inaccurate GenAI outputs can lead to unwarranted trust on the part of the ISDs and trainees. There is a tendency to mistake GenAI's fluent and confident output with actual subject matter expertise. This misplaced trust may lead users to accept content at face value without an appropriate level of validation. In training environments where safety, compliance, and technical accuracy are critical, blind reliance introduces unacceptable risks (McLean, 2024; Schatz, 2024). ISDs may assume that GenAI-generated text is inherently trustworthy because it sounds correct, uses correct grammar, or mimics the structure of validated instructional materials. However, although GenAI models may have prior domain knowledge, they may lack the grounding/validation to generate responses based on probability rather than understanding. As a result, they may inaccurately represent complex concepts, oversimplify nuanced procedures, or omit critical context that affects instructional fidelity. This overconfidence may be compounded when organizations institutionalize GenAI tools without also strengthening review processes. For example, if GenAI is used to accelerate development timelines but Subject Matter Expert (SME) involvement is reduced to meet efficiency goals, the risk of undetected errors or misrepresentations increases significantly.

Risk of Ineffective Training

GenAI-generated content may lack effective instructional design leading to ineffective training. Outputs may lack instructional scaffolding and misalign with objectives if it is not based on appropriate taxonomies or military instructional frameworks (Squalli Houssaini, 2024; Shires, 2024). Teaching an AI tool "how to teach" requires explicit alignment with service-specific guidance and the Military Handbook for Instructional Design. Prompting alone is insufficient without domain-specific model training and QA. The absence of human oversight may lead to a decline in educational rigor and consistency.

Traditional ISD methods include Planning, Analysis, Design, Development, Implementation, Evaluation, and Maintenance (PADDIE+M) and the Successive Approximation Model (SAM). Unlike these traditional ISD methods, which are grounded in structured pedagogical models, GenAI applications (without the assistance of agentic AI) may lack the embedded logic required to ensure content aligns with learning objectives, instructional strategies, and evaluation criteria. Without direct human intervention, courseware generated by GenAI may inadvertently skip prerequisite knowledge, fail to scaffold learning appropriately, or introduce material in an illogical sequence. These instructional breakdowns may confuse learners and reduce the effectiveness of training.

In addition, basic GenAI systems may fail to differentiate between instructional formats, such as distinguishing between demonstration, guided practice, and independent application phases. As a result, training content may become overly generic, miss critical instructional cues, and omit learner engagement techniques that are essential to successful knowledge transfer in military environments.

Instructional drift may also occur at scale. When large volumes of courseware are generated without a rigorous and centralized QA framework, inconsistencies in tone, terminology, and instructional flow may emerge across training modules. This undermines the "one voice" standard and introduces the risk of fragmented or contradictory learning experiences.

Bulut (2024) emphasizes the importance of aligning item generation with assessment objectives and integrating established learning and measurement theories into AI-driven content development. Without this instructional content, the content may lack a strong educational foundation. These potential foundational gaps underscore the importance of grounding GenAI adoption in sound instructional design principles and domain expertise (Noroozi, 2024).

Military training materials should adhere to established tone, terminology, and formatting conventions. Without standardization, training content becomes fragmented and inconsistent, undermining learner confidence and instructional coherence (Noroozi, 2024). GenAI models, unless guided by clearly designed prompts or templates, may produce inconsistent outputs that vary in tone, vocabulary, or complexity. One output might be overly casual; another may introduce jargon inconsistently or diverge from service-specific terminology.

Risk of Improperly Handled Information

GenAI systems must properly handle source materials and outputs and avoid issues associated with cybersecurity, IP infringement, and data distribution restrictions. Some systems may be required to adhere to data protection regulations and standards such as the European Union's General Data Protection Regulation and the Health Insurance Portability and Accountability Act (HIPAA) (Shires, 2024) GenAI systems may require proprietary, sensitive, controlled, or classified data. If improperly managed, their use may violate ethical standards or legal regulations. GenAI must comply with data protection laws and DoD-specific policies such as Department of Defense Instruction (DoDI) 8510.01. For example, ingesting Controlled Unclassified Information (CUI) or third-party licensed material into an unapproved model, especially one hosted on public infrastructure, may result in data leakage, intellectual property infringement, or non-compliance with DoD cybersecurity protocols. Legal risks may arise from a lack of transparency in how GenAI systems are trained and how their outputs are generated. If a GenAI model's behavior cannot be traced or explained, it becomes difficult to demonstrate that courseware complies with security classification or government procurement requirements (Ammanath, 2022; Lawton, 2025).

POTENTIAL RISK MITIGATION STRATEGIES

Risk mitigation begins with establishing documented risk management processes and procedures. Training organizations can define processes and procedures based on the tasks identified in the NIST AI RMF. A variety of strategies can help mitigate the GenAI risks such as the ones identified above. These strategies include:

- Establishing policy and performing governance,
- Ensuring Accurate Generated Content,
- Ensuring Effective Instruction, and
- Protecting information.

Establishing Policy and Governance

While identifying risks is a critical first step, the growing deployment of GenAI in defense training environments necessitates proactive governance measures and structured policy interventions. Bulut (2024) emphasizes that many GenAI-generated assessment items lack theoretical foundations and advocates for alignment with learning and measurement principles. Blodgett (2021) discusses that the history of educational technology is marked by overpromising tools that fail because they lack proper checks and balances. For GenAI to contribute meaningfully to instructional outcomes, deployment must be framed by deliberate institutional safeguards that control access and ensure accountability.

Training organizations should adopt governance frameworks that validate model provenance and monitor system behavior to ensure accountability and transparency. Model usage should be documented in accordance with the acquisition and Information Technology (IT) policy. Personnel from compliance and legal teams should participate in the development of GenAI policy to ensure its lawful deployment and operation. Before any GenAI tool is integrated into a training development environment, it should be evaluated and approved through a formal review process. This process should assess the tool's capabilities, security controls, limitations, provenance, and compliance with relevant standards such as DoDI 8510.01 (RMF for DoD IT). A centralized governance body should maintain an approved list of tools and periodically review usage across programs. Organizations should also define clear boundaries for how GenAI may be used within the instructional systems design process. For example, certain tasks like generating initial text drafts or content templates might be permitted, while direct generation of test items or critical safety-related content may be prohibited unless validated by SMEs. Policies must also clarify who has the authority to use GenAI, under what conditions, and within what environments. These constraints should be embedded in acceptable use agreements, training system authority-to-operate (ATO) documentation, and workforce onboarding materials. Regular audits should ensure compliance and capture lessons learned to inform ongoing tool governance (Lacy, 2025).

Ensuring Accurate Generated Content

Training organizations can help ensure that instructional content developed by GenAI tools is accurate by using domain-specific information from authoritative data sources and validating the resulting outputs. ISDs should focus on guaranteeing the veracity of information and assessing the dependability of sources (Conklin, 2024). Training organizations should work with SMEs to document a list of Authoritative Data Sources (ADSs).

GenAI systems generating instructional content can use technologies such as Retrieval Augmented Generation (RAG) to supplement general information provided by LLMs with domain-specific content (Lacy, 2025). RAG architectures use extensive background information from retrieved external authoritative documents to form precise instructions (Lewis, 2020). The U.S. Navy is using RAG architectures to infuse its GenAI training content development tools with domain knowledge and pedagogical guidance (Lacy, 2025).

Effective and safe use of GenAI tools in courseware development requires deliberate prompt engineering strategies that guide the model to produce accurate, relevant, and appropriately scoped outputs. Prompt engineering is not merely a technical skill; it is a critical quality control function that must be formalized within training development workflows. In-depth prompt templates aligned with learning objectives and model capabilities are crucial for avoiding hallucinations and inconsistencies (Fang, 2024). Although RAG may help by conditioning the model, it is not guaranteed alone to eliminate hallucinations. Other important techniques, such as post-hoc fact-checking, play a complementary role by ensuring outputs are validated against authoritative sources before being integrated into course materials.

Policies should require ISDs to use pre-approved prompt templates that align with best practices in instructional design. Prompts should be structured to include contextual information, intended learning outcomes, and constraints that shape the tone, scope, and format of the content.

Just as important as the prompt itself is the ability to trace it. All prompts and corresponding GenAI outputs should be logged and version controlled. These records support transparency, facilitate quality assurance, and provide critical evidence in the event of disputes or compliance reviews. Version control and traceability tracking support transparency and compliance (Wang, 2023).

Documentation protocols should include metadata such as the date and time of generation, the model version used, and whether a prompt was modified during an iterative review process. This level of rigor not only supports reproducibility and accountability but also builds institutional knowledge of what prompt formats are most effective in different training scenarios.

Training organizations should implement structured content validation workflows that include factual review checkpoints and standardized content acceptance criteria. Policies should counter trust-related risks by emphasizing the limitations of GenAI and formally requiring SME validation for all AI-generated instructional content. Human-in-the-loop validation integrates human expertise with AI-based decision-making (Bulut, 2024).

ISDs ensure instructional integrity while SMEs remain the authoritative validators of domain-specific content. QA policies should formalize SME engagement in the review process through structured workflows that define when, how, and by whom SME approval is required (McLean, 2024). SME reviews should be more than informal feedback; these checks should be documented steps in the QA chain, with clear responsibilities, sign-off protocols, and escalation paths in place if disagreements arise. Collaboration platforms or review dashboards may be useful for tracking SME input, validating edits, and verifying that feedback has been implemented.

In high-velocity production cycles, SME availability may become a constraint. In such cases, tiered review models may be used, prioritizing SME time for critical content while routing routine verifications through experienced ISDs or quality analysts with SME oversight. To manage GenAI outputs at scale, organizations need QA tools that go beyond manual review. Natural language processing (NLP) utilities, taxonomy-matching software, style checkers, and metadata tagging systems may enhance consistency and reduce human burden. Dashboards should report metrics such as hallucination frequency, SME rejection rates, prompt-response accuracy, and time-to-approval. This data helps identify systemic weaknesses in prompts, pedagogical agents, tool configurations, or review practices. They also support continuous improvement by identifying training gaps or content categories prone to GenAI error (Blodgett, 2021; Bommasani, 2021).

A balance of automation and human QA ensures that efficiency does not come at the cost of instructional quality or learner safety. As GenAI tools evolve, so too must the QA infrastructure that supports them. Organizations should consider implementing disclaimers, prompts, or embedded markers within GenAI outputs to flag content for mandatory review. Promoting a culture of scrutiny, where outputs are consistently evaluated rather than assumed to be correct, will help reduce the likelihood of instructional errors reaching learners. Content review should include proactive detection and correction of biased patterns. Policies should require the use of bias auditing tools, SME reviews with diverse perspectives, and prompt design strategies that minimize bias exposure (Humble, 2024). Incorporating fairness metrics and inclusive design practices into courseware development helps safeguard content integrity and learner trust.

The integration of a Human-in-the-Loop (HITL) model is critical to maintaining the instructional integrity, factual accuracy, and domain validity of GenAI-generated training content. GenAI should never be the final authority in any content generation workflow. Instead, every output must be reviewed, edited, and validated by qualified instructional systems designers (ISDs) and subject matter experts (SMEs). The HITL model should be embedded as a requirement at multiple checkpoints in the content lifecycle. This includes initial review of AI outputs, validation of alignment with learning objectives, and final quality assurance prior to delivery or publication. In high-risk content areas such as safety procedures, weapons systems, or tactical doctrine, multiple SMEs may be required to reach consensus before content can proceed.

Policies should clearly define roles and responsibilities within the HITL process and ensure that accountability is documented (Bozkurt, 2023). Organizations may consider implementing checklists, annotation tools, or content review logs to enforce review discipline. Furthermore, HITL validations should include a focus on more than just factual accuracy; they should also examine pedagogical alignment, readability, and learner appropriateness. By institutionalizing HITL processes and equipping personnel with both authority and guidance, training organizations may preserve the benefits of GenAI while safeguarding against its most dangerous vulnerabilities. HITL validation is the foundation of any quality assurance framework for GenAI-generated courseware. AI systems should be viewed as support tools, rather than final authorities. Every output intended for learner consumption must be reviewed by trained instructional systems designers (ISDs) and validated by qualified subject matter experts (Walter, 2024).

The HITL approach should include multiple layers of review. ISDs should assess instructional flow and format adherence, followed by SMEs, who should validate technical and operational accuracy. For content involving critical safety or mission-essential tasks, a second SME or cross-functional review may be required to ensure content accuracy. Review logs and digital approval signatures may reinforce accountability and ensure traceability. Incorporating HITL review into workflows mitigates the most pressing risks associated with hallucination, bias, and misalignment with training objectives. It also reinforces the integration of human judgment in areas where interpretation and domain expertise are essential.

Ensuring Effective Instruction

Training organizations can mitigate the risk of ineffective training by encoding pedagogy and verifying instructional alignment. Ensuring effective instruction is being produced is still a highly manual process. Human expertise is still crucial in areas such as: understanding the specific needs of learners, designing engaging and motivational content, and ensuring ethical considerations are addressed (Shires, 2024).

Domain information for instruction will vary based on learning objectives. However, GenAI tools developing instructional content should consistently follow the same approaches for providing content based on domain-specific information that is pedagogically sound.

To mitigate pedagogical drift, organizations should enforce strict adherence to style guides, mandate instructional design validation, and utilize workflow tools that embed instructional checkpoints throughout the content development process. GenAI tools should include a robust system that monitors the creation and management of user roles and permissions for approving content. Policies should require ISDs to trace GenAI outputs back to instructional objectives and ensure that all courseware reflects validated learning outcomes and domain standards.

As organizations move beyond identifying risks and establishing policy guardrails, attention must shift to quality assurance (QA) as a vital pillar in ensuring the effective development of AI-enabled training. Without rigorous QA practices, even the most well-governed GenAI systems may introduce factual inaccuracies or pedagogical misalignments into military courseware.

Addressing these quality challenges requires organizations to implement systematic review mechanisms, define output standards, and apply a Human-in-the-loop (HITL) model across the entire GenAI-assisted development pipeline. As Conklin, Dorgan, and Barreto (2024) note, GenAI is not inherently pedagogically aware; therefore, QA must compensate for this gap by aligning outputs with validated instructional frameworks and SME feedback. Well-developed GenAI tools should move to embed a pedagogical agent.

Effective courseware should align with defined learning objectives and cognitive expectations that are appropriate to the target audience. GenAI outputs, however, may diverge from these standards unless explicitly instructed otherwise. Instructional alignment verification should entail mapping AI-generated content to task analysis results and approved learning objectives. This type of information is often an output of the analysis phase of the PADDIE+M model (also referred to as front-end analysis). This alignment should include verifying that prerequisite knowledge is acknowledged, instructional scaffolding is in place, and the cognitive level aligns with the intended outcomes (e.g., recall vs. analysis). Outputs must be mapped to learning objectives and verified using outcome taxonomies, such as Bloom's (Kasneci, 2023).

Automated validation tools may assist in tagging TLOs and ELOs or comparing generated content to outcome taxonomies such as Bloom’s, but human validation is a vital step to ensure instructional fidelity. This step ensures that content not only covers the right topics but also does so in a way that builds learner competence. To maintain a “one voice” standard, organizations should develop and enforce the use of centralized style guides specific to GenAI use. Outputs should be evaluated against these style guides as part of routine QA. Where possible, model fine-tuning or foundational prompt templates should be aligned with these linguistic expectations.

Protecting Information

Training organizations should ensure that GenAI tools do not compromise the confidentiality or integrity of sensitive instructional content. When processing Controlled Unclassified Information (CUI) or classified data, GenAI systems may introduce risks related to data leakage and unauthorized access. If not properly managed, these risks may lead to mission degradation or violations of regulatory and legal requirements (Lawton, 2025).

To reduce exposure, GenAI tools should operate within secure, government-accredited environments that comply with cybersecurity guidance, such as DoDI 8510.01. GenAI tools should be designed without the need to transmit content over public networks or access external data sources. In high-risk scenarios, deploying models within DoD-controlled infrastructure may be necessary to ensure full control over both input and output data (Williams, 2025). Information protection measures should include access control, encrypted storage, network segmentation, and audit logging. These safeguards support visibility into how data is processed and help detect anomalies or unauthorized use. Periodic reviews should confirm that inputs and outputs remain within the system’s approved classification and authorization boundaries.

Establishing and enforcing policies that cover data sensitivity, input source restrictions, user role definitions, and system usage limitations should be a priority for organizations that use GenAI tools. Integrating a secure infrastructure surrounded by policy controls preserves the reliability and trustworthiness of GenAI tools used in courseware development. The deployment and operation of GenAI tools for military training must take place within secure and controlled environments that comply with DoD cybersecurity standards. These environments should be designed to prevent unauthorized access, data leakage, and any unintended exposure of training content, including Controlled Unclassified Information (CUI), proprietary data, or other sensitive instructional material (Williams, 2025). GenAI tools must not be allowed to access or transmit data over the public internet. Instead, they should be hosted within accredited networks, isolated from open systems, and protected with strict identity and access management protocols. Depending on the sensitivity of the content, local model deployment (i.e., running GenAI models within DoD-controlled infrastructure) may be necessary to ensure full control over input and output data and model behavior.

Organizations should also establish policies for configuration control, software patching, and log monitoring specific to GenAI-enabled environments. These technical safeguards should be supplemented with user access controls, audit trails, and usage monitoring systems to detect anomalies, policy violations, or potential compromise. Ensuring a secure and controlled environment helps protect against threats such as prompt injection, model inversion, and adversarial attacks. Over the past two years, the OWASP GenAI Security Project has advanced research in this area; however, these risk categories remain underexplored and may have significant implications for defense training systems.

SUMMARY

The use of GenAI to develop military courseware offers meaningful opportunities to enhance instructional system design and streamline content production. However, organizations must identify, evaluate, and address these potential risks, especially within high-stakes defense training environments.

GenAI systems may generate content that lacks accuracy and fails to align with learning objectives. The presence of biased information, security vulnerabilities, and unverified outputs further underscores the need for a comprehensive policy. To manage these risks, training organizations should adopt a structured risk management approach informed by the NIST AI Risk Management Framework that includes establishing clear policies and procedures, identifying and analyzing risks, prioritizing those with the highest potential impact, and managing mitigation efforts over time.

Human-in-the-loop validation is another important aspect that may be used to preserve instructional integrity. Qualified instructional designers and subject matter experts should verify the accuracy, pedagogical alignment, and operational relevance of all GenAI content. Institutional practices should include standardized prompt engineering and traceable output logging, supported by the integration of authoritative data sources and outcome-based validation protocols.

GenAI technologies support courseware development when implemented within secure, well-governed environments that enforce instructional rigor and accountability to manage risks. The value of these tools depends on integration into quality assurance workflows that maintain human oversight and uphold the standards required for effective and mission-relevant training. Recognize the risks associated with using GenAI for developing courseware and associated artifacts

REFERENCES

Ammanath, B. (2022). Trustworthy AI: A business guide for navigating trust and ethics in AI. John Wiley & Sons.

Autio, C. , Schwartz, R. , Dunietz, J. , Jain, S. , Stanley, M. , Tabassi, E. , Hall, P. and Roberts, K. (2024), Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD, [online], https://doi.org/10.6028/NIST.AI.600-1, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=958388

Blodgett, S. L., & Madaio, M. (2021). Risks of AI foundation models in education. arXiv. https://arxiv.org/abs/2110.10024

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., .& Liang, P. (2021). On the opportunities and risks of foundation models. Stanford Center for Research on Foundation Models (CRFM). https://arxiv.org/abs/2108.07258

Bozkurt, A., & Sharma, R. C. (2023). Generative AI and prompt engineering: The art of whispering to let the genie out of the algorithmic world. Asian Journal of Distance Education, 18(2), i–vii. Retrieved from https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/749

Bulut, O., Beiting-Parrish, M., Casabianca, J. M., Slater, S. C., Jiao, H., Song, D., Ormerod, C. M., Fabiyi, D. G., Ivan, R., Walsh, C., Rios, O., Wilson, J., Yildirim-Erbasli, S. N., Wongvorachan, T., Liu, J. X., Tan, B., & Morilova, P. (2024). The rise of artificial intelligence in educational measurement: Opportunities and ethical challenges. Chinese/English Journal of Educational Measurement and Evaluation, 5(3), Article 3. https://www.ce-jeme.org/journal/vol5/iss3/3/

Center for Teaching and Learning. (n.d.). Generative AI in teaching and learning: Biases and risks. The University of Texas at Austin. https://ctl.utexas.edu/generative-ai-teaching-and-learning-biases-and-risks

Conklin, S., Dorgan, T., & Barreto, D. (2024). Is AI the new course creator? Discover Education, 3(285). https://doi.org/10.1007/s44217-024-00386-2

Eastgate Software. (2023, January 10). 10 potential negative effects of AI in education. Eastgate Software. https://eastgate-software.com/10-potential-negative-effects-of-ai-in-education/

Fang, B., & Broussard, K. (2024). Augmented course design: Using AI to boost efficiency and expand capacity. Educational Technology & Society, 27(1), 45-62. https://er.educause.edu/articles/2024/8/augmented-course-design-using-ai-to-boost-efficiency-and-expand-capacity

Hall, Rachel, The Guardian. (2025, May 4). 'Dangerous nonsense': AI-authored books about ADHD for sale on Amazon. https://www.theguardian.com/technology/2025/may/04/dangerous-nonsense-ai-authored-books-about-adhd-for-sale-on-amazon

Humble, N. (2024). Risk management strategy for generative AI in computing education: How to handle the strengths, weaknesses, opportunities, and threats? International Journal of Educational Technology in Higher Education, 21(1), Article 61. https://doi.org/10.1186/s41239-024-00494-x

Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F,. ChatGPT for good? On opportunities and challenges of large language models for education. Learn Individual Diff. 2023;103: 102274. https://doi.org/10.1016/j.lindif.2023.102274.

Lacy, L.W & Jones, K.. (2025, April). "Developing Aircraft Maintenance Training Using Generative AI". Presentation at the WATS 2025 Maintenance Training Conference, Orlando, FL.

Lawton, G. (2025, March 3). Generative AI ethics: 11 biggest concerns and risks. TechTarget. https://www.techtarget.com/searchenterpriseai/tip/Generative-AI-ethics-8-biggest-concerns

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474. https://doi.org/10.48550/arXiv.2005.11401

McLean, M. (2024). AI Tools to Enhance Teaching and Reduce Workloads. Agora, 59(2), 59-61. https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/scholarly-journals/ai-tools-enhance-teaching-reduce-workloads/docview/3085714896/se-2

Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

National Institute of Standards and Technology. (2023). NIST AI RMF Playbook. U.S. Department of Commerce. https://www.nist.gov/itl/ai-risk-management-framework/nist-ai-rmf-playbook

Noroozi, O., Soleimani, S., Farrokhnia, M., & Banihashem, S. K. (2024). Generative AI in education: Pedagogical, theoretical, and methodological perspectives. International Journal of Technology in Education, 7(3), 373-385. https://doi.org/10.46328/ijte.845

Schatz, S., Stodd, J., & Stead, G. (2024, December 2). Navigating the generative AI revolution [Conference tutorial]. Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL. https://www.iitsec.org/get-involved/authors/best-papers-and-tutorials-archive

Shires, R., & McCormack, R. (2024, December 2). Unleashing the potential: Harnessing large language models and generative AI in military and industry applications [Tutorial]. Interservice/Industry Training, Simulation and Education Conference (I/ITSEC), Orlando, FL.

Squalli Houssaini, M., Aboutajeddine, A., Toughrai, I., & Ibrahimi, A. (2024). Development of a design course for medical curriculum: Using design thinking as an instructional design method empowered by constructive alignment and generative AI. Thinking Skills and Creativity., 52. https://doi.org/10.1016/j.tsc.2024.101491

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. International Journal of Educational Technology in Higher Education, 21(15). https://doi.org/10.1186/s41239-024-00448-3

Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L., & Feng, M. (2023). When adaptive learning is effective learning: Comparison of an adaptive learning system to teacher-led instruction. Interactive Learning Environments, 31(2), 793-803. https://doi.org/10.1080/10494820.2020.1808794

Williams, L. C. (2025, April 26). The Army has rolled out a generative AI workspace to improve daily operations. Defense One. https://www.defenseone.com/defense-systems/2025/04/army-preparing-generative-ai-workspace-improve-daily-operations/404876/