AI in Military ETEE ADL: A Toolkit for Local Reflection

This info paper was developed by the PfPC ADL Working Group (WG) to inform next-generation Education, Training, Exercises, and Evaluation among security partners.

This info paper offers military and civilian leaders involved in Military Education, Training, Exercise, and Evaluation (ETEE) a concise and practical tool for reflecting on the role of AI in military ETEE ADL. It introduces a structured approach to integration and evaluation with a shared documentation template, a military-specific AI evaluation rubric, and a repeatable evaluation process helping institutions make informed, mission-aligned decisions about AI integration. Designed to support both strategic dialogue and operational planning, the paper can be read in approximately 20 - 25 minutes.

Defining AI in Military ETEE ADL

Artificial Intelligence (AI) in Military Education, Training, Exercise, and Evaluation (ETEE) refers to algorithmic systems that enhance instructional delivery and learner engagement within Advanced Distributed Learning (ADL) environments. These systems include intelligent tutoring platforms, adaptive learning engines, conversational agents, and scenario generators that dynamically tailor content, provide real-time feedback, and personalize learning pathways based on performance data and operational requirements (Woolf, 2007; Graesser & McDaniel, 2008). AI can automate and support aspects of curriculum development, offer analytics support for instructors, and simulate high-fidelity, mission-relevant training scenarios. It also enhances learner engagement by identifying knowledge gaps, providing timely support, and adapting to individual learning preferences, critical for military personnel operating under pressure or with limited time. The convergence of AI and ADL unlocks new opportunities for scalable, efficient, and mission-aligned training. However, integration across institutions remains difficult and inconsistent, highlighting the need for structured evaluation and responsible implementation (PfPC ADL WG, 2023).

Implementing AI in Military ETEE ADL

One of the primary challenges in unlocking AI's potential in ADL is the lack of standardized methods to identify, assess, validate, and reflect on AI-based support systems within military ETEE contexts. These contexts (from combat simulations to strategic planning exercises) entail diverse instructional and operational requirements, complicating efforts to determine whether AI tools are not only technically functional but also pedagogically sound and operationally aligned (Stanley-Lockman & Christie, 2021). This absence of standardization leads to mismatched solutions, underused tools, and fragmented adoption, which can ultimately undermine training outcomes, mission readiness, resource spending, instructors' bandwidth, and confidence in future AI initiatives.

From Insight to Policy: The Need for Structured Reflection

Understanding the challenges of AI implementation in military learning environments is only a first step. Institutions must move beyond problem identification toward collaborative reflection and structured decision-making. Reflecting on operational needs, ethical standards, and technical readiness helps clarify which AI systems are fit for purpose (Sweeney, Clarke, & Higgs, 2019). Shared evaluation tools are vital to facilitate this process, enabling consistent assessment, fostering open dialogue, and supporting policy development across contexts. To meet this need, the PfPC ADL WG has developed a dedicated toolkit that supports structured reflection (by commanders, instructors, and other stakeholders), enables informed decision-making, and lays a foundation for responsible AI adoption in military ETEE ADL.

Evaluating AI Implementations

Research across education, technology, and defense domains offers critical insights into the integration and evaluation of AI in military ETEE ADL. While no single framework fully addresses the complexity of military ETEE, several established models offer transferable principles and proper starting points. AI Implementation Rubrics (Mackie & Aspenlieder, 2024; OECD, 2022) emphasize reliability, accessibility, interoperability, and cost-effectiveness, ensuring AI tools support learning objectives and align with institutional capabilities. For example the DOMS AI-Ed rubric (Hardman, 2023) introduces dimensions such as learning impact, ethical safeguards, transparency, and scalability, reinforcing the importance of supporting human learning without undermining autonomy or fairness. School-centered frameworks often apply checklist-based approaches focusing on usability, compliance, and infrastructure readiness. Defense-specific literature (Panwar, Li, & T., 2024; Islam, Abrar Jahin, & Mridha, 2024) prioritizes operational robustness, auditability, human oversight, and mission alignment. Whereas the AI Assessment Scale (AIAS) (Perkins, Furze, Roe, & MacVaugh, 2024) includes criteria on academic integrity and institutional readiness.

Towards a Toolkit for Reflection

There is a growing need for shared instruments that facilitate structured reflection and knowledge exchange on the use of AI in military ETEE ADL. A dedicated toolkit serves multiple purposes:

1. ***Enable Internal Dialogue*** - The toolkit offers a common language and structured criteria that help instructors, commanders, technologists, and policymakers engage in meaningful conversations about AI. It frames essential questions regarding needs, risks, and system readiness.

2. ***Support Cross-Institutional and Multinational Sharing*** - By standardizing how AI-enabled training scenarios are described and evaluated, the toolkit makes it easier to share lessons and drive interoperability across organizations and nations. For example, an AI-driven language training tool used in one country's officer academy can be adapted when documented using shared templates and rubrics.

3. ***Provide a Consistent Structure for Evaluation and Scaling*** - As institutions test AI in different ETEE ADL areas such as cybersecurity, decision-making, and logistics, the toolkit enables repeatable evaluations. This helps identify solutions that are ready for scaling, those requiring adjustment, and those unsuitable for deployment, transforming isolated pilots into shared learning experiences.

4. ***Incorporate End-User Feedback and Human Oversight*** - The toolkit enables the gathering of structured feedback from instructors, trainees, and other stakeholders. This ensures AI tools remain aligned with operational realities, user needs, and evolving ethical considerations, reinforcing human oversight and trust in AI systems.

Together, these functions position the toolkit as both a reflection aid and a communication enabler, supporting ethical, interoperable, and mission-aligned AI adoption.

The Construction of the Toolkit

Evaluating AI applications in military ETEE ADL presents a unique challenge: the balance between maintaining high-level strategic oversight and capturing nuanced, context-specific details. As AI becomes embedded in both legacy and next-generation systems, its complexity, operational relevance, and pace of change demand adaptable evaluation tools that go beyond generic assessments. To address this, the PfPC ADL WG have developed a modular and scenario-driven evaluation framework composed of three core instruments: a scenario-sharing template, a military-specific AI evaluation rubric, and a structured evaluation process.

1. ***Scenario-Sharing Template*** - This structured template allows stakeholders to document AI-enabled learning scenarios in a consistent, reusable, and evaluable way. It includes fields such as scenario purpose, learner audience, operational context, instructional goals, technology used, ethical considerations, and improvement metrics. By standardizing how scenarios are described, the template facilitates comparison, reuse, and collaborative refinement across institutions and use cases.

2. ***Military-Specific AI Evaluation Rubric*** - To complement the scenario inputs, a tailored evaluation rubric was created by synthesizing established frameworks from NATO, OECD, NIST, and the U.S. Department of Defense. It covers seven categories: Functionality, Adaptability & Personalization, Technical Integration, Pedagogical Impact, Ethics & Security, Cost & Sustainability, and Classified Information & Accreditation. Each category includes specific sub-criteria, evaluated using a three-level concern scale (Works Well, Minor Concerns, and Serious Concerns) to guide nuanced, risk-informed decision-making.

3. ***Structured Evaluation Process*** - Scenarios submitted via the template are analyzed against the rubric. For example, a scenario describing an AI tutoring engine used in pre-deployment training may be assessed for adaptability, cognitive engagement, and Learning Management System (LMS) integration. This process enables both high-level comparability and detailed analysis, translating qualitative scenario data into structured insights. Evaluators can flag operational risks (e.g., lack of multilingual support or

transparency in decision logic), identify scalable innovations, and determine whether interventions are mission-ready, require adjustment, or need redesign. The rubric is intentionally flexible. Evaluators can zoom in on specific technologies (e.g., adaptive feedback or biometric data use) while maintaining a broad overview of strategic alignment, interoperability, and ethical safeguards. The concern-level system helps prioritize further action: "minor concerns" may suggest configuration or compliance tweaks, whereas "serious concerns" could indicate potential blockers to safe deployment in military learning environments.

The development of these instruments was significantly shaped by input collected during the PfPC ADL WG meetings in Tbilisi (Georgia) and Riga (Latvia) and during the NATO NLTIG meeting in Halmstad (Sweden). Through facilitated workshops and scenario contribution sessions, participating nations and institutions provided real-world examples of AI-enabled learning, including tools for adaptive decision support, simulation-based mission rehearsal, and personalized language instruction. These submissions enriched the scenario repository and informed the rubric's structure, ensuring its relevance across operational environments, levels of instruction, and technological maturity. The collaborative discussions also surfaced shared concerns, such as handling classified data, ensuring ethical transparency, and managing human-AI teaming, which were directly integrated into the rubric categories and definitions. This participatory approach ensures the framework reflects the diversity of use cases and doctrinal needs across nations and institutions. It also lays the groundwork for a community of practice or interest around responsible AI integration in military learning.

Together, the scenario template, evaluation rubric, and collaborative validation process form a robust toolkit for assessing AI in military ETEE ADL. They allow institutions to transition from isolated pilot projects to scalable, policy-aligned adoption, ensuring that AI is not just innovative, but also mission-relevant, ethically grounded, and operationally sound.

Recommendations for Use

The toolkit supports institutions at various stages of AI adoption. Its practical, structured components help guide strategic reflection, responsible innovation, and institutional collaboration. Three recommended use cases include:

Internal Dialogue and Strategic Reflection

The toolkit provides a structured foundation for internal dialogue within military institutions, enabling leaders, instructional designers, technology managers, and operational stakeholders to engage in meaningful discussions about the strategic integration of AI. It helps identify:

- Levels of organizational AI readiness and alignment with strategic training goals;
- Instructional challenges that may be addressed through AI augmentation or automation;
- Risks related to ethics, data protection, bias, and system reliability;
- The balance between innovation and operational security;
- The maturity of current digital infrastructure and support systems.

By offering common language and evaluation criteria, the toolkit fosters shared understanding across departments, bridging gaps between technical, pedagogical, and operational perspectives. This internal alignment is essential for ensuring that AI initiatives are scalable, compliant, and mission-relevant from the outset.

Cross-Institutional Knowledge Sharing

In an increasingly interconnected training landscape, the toolkit serves as a mechanism for structured knowledge exchange among nations. The scenario-sharing template and evaluation rubric make it easier to:

- Document and disseminate AI-enabled training practices in a consistent and evaluable way;
- Compare approaches across different operational contexts, learner profiles, and technologies;
- Identify common success factors and recurring implementation challenges;
- Build a repository of validated, real-world use cases that others can learn from or adapt.

This fosters a community of practice, in which insights are exchanged not as isolated anecdotes but as structured, comparable experiences. It also enhances interoperability by aligning terminology, evaluation criteria, and ethical considerations across organizations.

Evaluation Consistency and Informed Decision-Making

As military institutions move from pilot projects to broader implementation, the toolkit supports consistent, transparent evaluation of AI-enabled ADL systems. The military-specific evaluation rubric helps institutions:

- Assess the effectiveness, risks, and limitations of AI tools under realistic conditions;
- Identify whether solutions meet instructional and operational standards before scaling;
- Flag issues requiring redesign, reconfiguration, or further testing;
- Maintain command accountability and integration with existing military evaluation protocols;
- Determine ethical and security implications, including data protection and human oversight.

By shifting from abstract metrics to scenario-based, mission-aligned validation, the rubric ensures that decision-makers are equipped with actionable insights. Appendix 3 brings this to life with a worked example.

Conclusions and a Way Forward

Many institutions are in the early stages of piloting AI in military ETEE ADL, and a logical next step is to establish a structured platform (such as working groups, communities of interest, or regional exchanges) for sharing insights using a common scenario template and evaluation rubric. This structured collaboration would not only enhance transparency and reduce duplication of effort

but also accelerate collective learning across contexts. Over time, such a platform can evolve into a technologically enabled, continuously updated repository of validated, mission-ready AI use cases. This repository would serve as a reference point for instructional designers, procurement authorities, and policy advisors across nations and institutions, ensuring that AI integration efforts are informed by operational relevance, pedagogical value, and ethical soundness.

The first iteration of this toolkit was developed to support this evolution by stimulating structured reflection and enabling policy alignment at both national and institutional levels. It provides a practical and accessible entry point for institutions exploring AI in military ETEE ADL. The toolkit outlines key principles, ethical considerations, implementation risks, and functional requirements for using AI in military learning environments, while recognizing the diversity of operational settings, learning audiences, and digital maturity levels. By combining scenario-based documentation, robust multi-criteria evaluation, and collaborative validation practices, the toolkit empowers stakeholders to move beyond isolated, experimental pilots toward a coherent, ethical, scalable, and mission-driven adoption strategy.

In doing so, it helps bridge the gap between innovation and standardization, ensuring that local experimentation contributes to shared learning and collective readiness. The annexes include the full set of instruments (such as the scenario template and evaluation rubric) as well as a worked example. These materials are designed to be easily adaptable for local use, while also supporting alignment with broader national or coalition-wide objectives.

Annexes:

1. Scenario-Sharing Template
2. Military-Specific AI Evaluation Rubric
3. Worked Example and Structured Evaluation

Acknowledgements:

This paper was drafted by the Partnership for Peace Consortium Advanced Distributed Learning Working Group in conjunction with its meetings in Tbilisi (Georgia) and Riga (Latvia). The meetings were co-chaired by COL Stephen Banks and Dr. Stijn Van Laer (NATO Allied Command Transformation), Dr. Aaron Presnall (Jefferson Institute, USA), and LTC Nickolaus Michael (Bundeswehr, DEU); hosted by BG Mamia Balakhadze, Commandant of the Military Academy of Georgia; and by COL Māris Utināns Commandant of the Latvian National Defense Academy, and the PfP Consortium Secretariat at the George C. Marshall Center.

The core author group was led by Stijn Van Laer (NATO ACT).

Iskanderov Nicat (AZE); Nuraddinov Anar (AZE); Harutyunyan Tigran (ARM); Margaryan Astghik (ARM); Radoslavov Dobril (BGR); Keremidchieva Greta (BGR); Kucsmik-Horváth Éva (HUN); Dionisie Ciubotaru (MDA); Babara Irina (MDA); Bogatinov Dimitar (NMK); Banks Stephen (NATO); Roman Gigi (NSO); Zoran Karavidić (SRB); Radu Catalin (ROU); Presnall Aaron (USA); Schatz Sae (USA); Jinnestrand Ulf (USA); Nesic Aleksandra (USA); Oper Marika (EST); Jenna Vekkaila (FIN); Savoranta Ville (FIN); Pankotsch Philipp (DEU); Kerigan-Kyrou Dinos (IRL); Mikalainis Povilas (LTU); Gunter Heather (NSO) Pietraszkiewicz Danuta (NATO DEEP); Kowalczys Justyna (NATO DEEP); Geiss Tanja (NATO DEEP); Tomasik Jerzy (NATO DEEP); Hommes Hillery (NLD); Arnuga Edvard (SVN); Melanie Nacimento (FRA); Brown Ash (GBR); Melmoth David (GBR); Roe David (GBR); Hartley David (GBR); Fernandexz Kendra (USA); Tyrrell Gina (USA); Hartley David (USA); Walsh Wendy (USA); Russell Nancy (USA); Kwon Paul (USA); Goldberg Ben (USA); Erin Czerwinski (USA); Margaret Korosec (GBR); Nyland Rob (USA); and Skirtladze Giorgi (GEO).

References:

Graesser, A. C., & McDaniel, B. (2008). Conversational Agents Can Provide Formative Assessment, Constructive Learning, and Adaptive Instruction. In C. A. Dwyer, *The Future of Assessment* (p. 28). New York: Routledge.

Hardman, P. (2023). *The DOMS AI-Ed Tools Evaluation Rubric: A Practical Checklist for Evaluating the Efficacy & Safety of AI Education Tools.* London: DOMS.

Islam, A., Abrar Jahin, M. F., & Mridha, N. D. (2024). A Unified Framework for Evaluating the Effectiveness and Enhancing the Transparency of Explainable AI Methods in Real-World Applications. *arXiv*.

Mackie, K., & Aspenlieder, E. (2024). *Rubric for Evaluating AI Tools: Fundamental Criteria.* Hamilton: McMaster University. Retrieved from https://ecampusontario.pressbooks.pub/app/uploads/sites/3696/2024/02/Rubric-for-AI-Tool-Evaluation-Fundamental.pdf

OECD. (2022). *OECD Framework for the Classification of AI systems.* OECD Digital Economy Papers. Paris: OECD Publishing.

Panwar, R. S., Li, Q., & T., S. j. (2024). *Military Artificial Intelligence Test and Evaluation Model Practices.* Geneva: INHR.

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment. *Journal of University Teaching & Learning Practice, 21*(6), 1-18.

PfPC ADL WG. (2023, November). *Info Papers.* Retrieved from Partnership for Peace Consortium: https://www.pfp-consortium.org/products/info-papers

Stanley-Lockman, Z., & Christie, E. H. (2021, October 25). *An Artificial Intelligence Strategy for NATO*. Retrieved from NATO Review: https://www.nato.int/docu/review/articles/2021/10/25/an-artificial-intelligence-strategy-for-nato/index.html

Sweeney, A., Clarke, N., & Higgs, M. (2019). Shared Leadership in Commercial Organizations: A Systematic Review of Definitions, Theoretical Frameworks and Organizational Outcomes. *International Journal of Management Reviews, 21*(1), 115-136.

Woolf, B. P. (2007). *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning.* Amherst: Morgan Kaufmann.

Annex 1: Scenario-Sharing Template

By completing this template, users will generate a standardized scenario profile that supports reflection reuse, comparison, and evaluation across contexts. Each item is accompanied by a guiding question.

	Field	Guidance
Scenario and Contributor Information	**Scenario Short Name / Title**	Title describing the learning challenge, context, learner impact, or some combination.
	Contributor Institution	Which institution submitted the scenario?
	Contributor Nation	Which nation submitted the scenario?
	Contributor Name	Who submitted the scenario? (POC)
	Contact email	Provide a contact email. (POC)
AI System Overview	**Scenario Description**	What is the AI-enabled tool or solution? How is AI incorporated?
	Purpose	General purpose of the AI system in this context?
	Operational Context	Where would this scenario be deployed?
	Application Area	Where in the instructional lifecycle does this apply?
	Technology Used	AI or related tools/languages used.
Stakeholders and Users	**Primary Learner Audience**	What is ETEE audience that will use the AI system?
	Primary User Role	Who uses the AI system?
	Intended Audience	Broader intended stakeholder group.

Learning Objectives and Challenges	**Intended Goals Category**	What category of instructional or operational goals does the scenario target?
	Intended Goal Description	What specific instructional or operational goals does the scenario target?
	Central Challenge	What problem or pain point was the scenario addressing?
	Outcomes Supported	Skills, competencies, or objectives the scenario supports.
Evaluation and Impact	**Improvement Metrics and Indicators**	What data were measured/captured to evaluate effectiveness? (e.g., engagement %, test scores, time on task, retention, decision speed, etc.)
	Impact Description / Level	What was the measured level of impact from evaluations?
Strategic Alignment	**Mission-Relevance**	Level of connection to operational or strategic objectives.
	Mission-Relevance Explained	Specific connection to operational or strategic objectives.
Technical Implementation	**Technical Implementation**	Describe in general terms the tool development process from analysis to evaluation.
	Procurement	How were resources obtained?
	Integrations and/or Standards	What technical standards or systems does it connect to?
	Security Level	What security level does the tool operate at?
	Security Implementation	Describe in general terms how the tool is tuned / implemented for the security level.
Compliance and Ethics	**Ethical Considerations**	Any open ethical issues?
	Sources and/or References	Papers, documentation, standards, or studies cited.
Sharing and Documentation	**Is the Scenario Complete?**	Is all the required information present for internal reuse?
	Willingness to Share Details	Are you willing to share details on the implementation of the scenario?

Annex 2: Evaluation Rubric

Category	Sub-Category	Works Well	Minor Concerns	Serious Concerns
Functionality	**Scalability**	AI supports large-scale deployments with no performance degradation.	Occasional slowdowns during high loads.	Consistently struggles with large numbers of users.
	Ease of Use	Intuitive interface with a minimal learning curve.	Some users require assistance.	Complex and non-intuitive, hindering effective use.
	Tech Support	Dedicated support with military-specific guidance.	Generalized AI support, lacking military focus.	Limited or no support, impeding usability.
	Responsiveness	Provides real-time, accurate responses.	Some delays or inaccuracies.	Slow response times and frequent inaccuracies.
Adaptability & Personalization	**Adaptability**	Adjusts to different military training needs and user levels.	Limited adaptation options.	One-size-fits-all; No adaptation possible.
	Accessibility	Meets military accessibility standards (e.g., WCAG, DoD requirements).	Partially compliant with accessibility guidelines.	Lacks accessibility features.
	Localization	Supports multiple languages, including military lexicon.	Some localization issues.	Lacks multilingual support.
Technical Integration	**LMS Integration**	Fully integrates with military LMS.	Some features unavailable within the LMS.	Requires separate access, disrupting workflow.
	Device Compatibility	Works across all military-approved devices.	Some compatibility limitations.	Requires specialized or high-end hardware.
	Offline Functionality	Offers key features offline for deployed environments.	Limited offline capabilities.	Fully dependent on internet connectivity.

	Interoperability	Fully integrates with existing military systems, software, and platforms.	Some interoperability challenges requiring additional configuration.	Does not support integration with military infrastructure.
Pedagogical Impact	**Cognitive Engagement**	Enhances critical thinking and decision-making training.	Somewhat supports cognitive engagement.	Does not promote meaningful learning.
	Scenario-Based Learning	Effectively supports war-gaming, simulations, and real-world scenarios.	Supports scenarios adequately but with some customization or integration limitations	Poor scenario support or integration that significantly limits training effectiveness.
	Learning Analytics	Provides actionable insights for instructors and trainees as well as detailed and interoperable data feeds for custom analysis.	Provides rudimentary dashboards or some data in a raw format.	No learning analytics available.
	AI-Powered Feedback	Offers personalized, constructive feedback.	Feedback is generic or inconsistent.	No feedback available.
Ethics, Bias, & Security	**Bias Mitigation**	Actively prevents bias in decision-making and training.	Some bias concerns present.	AI exhibits unchecked bias, leading to misinformation.
	Data Privacy & Security	Fully compliant with military data protection standards.	Some security risks identified.	Poses significant security risks to classified data.
	Transparency	AI decision-making processes are well-documented.	Some transparency issues.	Operates as a black box, with no insight into decision-making.
Cost & Sustainability	**Cost Efficiency**	Provides value within military budgets.	Some hidden costs.	Prohibitively expensive.
	Energy Efficiency	Optimized for low resource consumption.	Moderate energy use.	High energy demands, impacting operational feasibility.

	Long-Term Viability	Sustainable development with ongoing updates.	Updates are infrequent.	AI tool is becoming obsolete.
Classified Information & Accreditation	**Classified Data Handling**	Meets military standards for classified data storage and transmission (e.g., DoD, NATO).	Some compliance gaps, requiring additional security measures.	Not suitable for handling classified information.
	Data Storage & Sovereignty	Ensures data is stored within military-approved facilities with proper access controls.	Some data may be stored in non-military jurisdictions.	Data stored in unsecured or non-compliant locations.
	Accreditation & Certification	Holds necessary certifications (e.g., FedRAMP, STANAG, ISO/IEC 27001).	Certification pending or partially compliant.	Lacks necessary accreditation, making it unsuitable for military use.

Annex 3: Worked Example Scenario and Structured Evaluation

EXAMPLE: AI-Enhanced Adaptive Tutoring for Pre-Deployment Decision-Making

Scenario Short Name / Title	AI-Enhanced Adaptive Tutoring for Pre-Deployment Decision-Making
Contributor Institution	XXX
Contributor Nation	XXX
Contributor Name	XXX
Contact email	XXX
Scenario Description	An AI-powered Intelligent Tutoring System (ITS) embedded in a pre-deployment course for junior officers. It adapts decision-making scenarios in real time using learner analytics, prior training data, and NLP-based feedback. The system personalizes the learning path, simulates joint ops environments, and works offline in field conditions.
Purpose	To accelerate and personalize decision-making training for personnel before deployment, under operationally realistic, multilingual, and rules-based simulated conditions.
Operational Context	In national pre-deployment centers, joint multinational exercises, and field-device-enabled remote training setups.
Application Area	Applied during training delivery and formative assessment; also supports remedial instruction and adaptive learning paths.
Technology Used	Natural Language Processing (NLP); Reinforcement Learning for scenario branching; Rule-based decision trees; Dashboard analytics with AI outputs; SCORM/xAPI-compatible learning architecture
Primary Learner Audience	Personnel preparing for deployment, including officers and non-commissioned personnel involved in command, control, and field decision-making.
Primary User Role	Learners, instructors, and training evaluators involved in nation-led or partner-affiliated pre-deployment training programs.
Intended Audience	Training centers, operational commanders, policy makers, and instructional designers within the nation and partner defense education institutions.
Intended Goals Category	Cognitive readiness, situational judgment, multilingual decision-making, and rules of engagement (ROE) scenario training.
Intended Goal Description	Enhance judgment under uncertainty; Improve decision-making under operational constraints; Adapt responses to dynamic ROE and cultural considerations; Reinforce pre-deployment mission preparedness
Central Challenge	Inconsistent training scalability across national units, lack of personalization in traditional instruction, insufficient feedback loops, and the need for high-fidelity scenario exposure before deployment.
Outcomes Supported	Tactical and operational decision-making; Rules of engagement comprehension; Cultural awareness; Multilingual scenario navigation; Adaptive thinking and flexibility under stress
Improvement Metrics and Indicators	Learner engagement and completion rates; Scenario decision accuracy and speed; Instructor override frequency; Feedback timeliness; System stability across device types; Learner satisfaction (qualitative surveys)
Impact Description / Level	Pilot tested across multiple cohorts with scalable implications for institutional-wide adoption; supports mission-specific training and broader coalition interoperability.

Mission-Relevance	High, directly aligned with operational readiness and interoperability goals.
Mission-Relevance Explained	Supports the national mandate to prepare the force for future operations; reinforces the Military Training and Education Policy (MTEP) focus on adaptive learning and mission relevance.
Technical Implementation	Needs Analysis (pre-deployment scenario gaps identified); Design & Prototyping (AI engine and branching scenarios); Iterative Development (with military SMEs and field instructors); Pilot Testing (multinational cohorts); Evaluation & Optimization (based on learner data and instructor feedback)
Procurement	Through a vendor-university partnership funded by a national innovation grant.
Integrations and/or Standards	SCORM, xAPI, LMS, WCAG 2.1 (Accessibility Compliance), SCORM / xAPI documentation, MTEP Guidance, ISO/IEC 27001 & FedRAMP Medium (pending compliance), secured laptops and rugged field devices, Pending real-time dashboard integration
Security Level	Operates at NATO unclassified level; not authorized for use with classified materials.
Security Implementation	Runs on accredited cloud infrastructure; content reviewed for security restrictions; data stored in compliant European jurisdictions; logs from pilot sessions temporarily used non-sovereign storage, now corrected.
Ethical Considerations	Data Privacy, Algorithmic Bias, Human Oversight, Consent / Autonomy; Some cultural bias detected in scenario phrasing and mission language; Instructor training needed for full interpretability of AI feedback visualizations; Voice accessibility not yet available for visually impaired users
Sources and/or References	Requirements, Policy, Support Documents.
Is the Scenario Complete?	Yes
Willingness to Share Details	After contact

REPORT: AI-Enhanced Adaptive Tutoring for Pre-Deployment Decision-Making

Category	Concern Level
Functionality	Works Well
Adaptability & Personalization	Minor Concerns
Technical Integration	Minor Concerns
Pedagogical Impact	Works Well
Ethics, Bias, & Security	Minor Concerns
Cost & Sustainability	Minor Concerns
Classified Info & Accreditation	Serious Concerns

The scenario demonstrates strong performance in both functionality and pedagogical impact, showcasing a mature tutoring design and effective learner adaptability. Minor concerns were identified in areas such as adaptability, technical integration, ethics, and sustainability, largely due to limitations in localization, partial LMS synchronization, moderate energy demands, and the need for continuous oversight and refinement. However, a serious concern was flagged under Classified Information & Accreditation, as the system is not yet certified for handling classified data and lacks the necessary accreditation for deployment in secure operational environments.

Categories	Sub-Category	Concern Level	Justification
Functionality	**Scalability**	Works Well	Successfully deployed across multiple cohorts without system degradation, with positive feedback on consistency in learner experience.
	Ease of Use	Works Well	Learners required minimal onboarding; UI is intuitive and adapted to both desktop and field-device formats.
	Tech Support	Minor Concern	While support is responsive, it is currently vendor-managed and lacks round-the-clock availability with military-specific SMEs.
	Responsiveness	Works Well	AI provides instant feedback and adjusts content fluidly during scenario progression, including in offline simulations.
Adaptability & Personalization	**Adaptability**	Works Well	System adapts to multiple user levels and tracks individual learning curves effectively.
	Accessibility	Minor Concern	Largely compliant with WCAG 2.1, but voice control features for visually impaired users are not yet available.
	Localization	Minor Concern	Multilingual support exists (EN, FR, DE), but field-specific military jargon is not equally well translated across all interfaces.

Technical Integration	**LMS Integration**	Minor Concern	Integrates with LMS, but real-time sync with performance dashboards is still under refinement.
	Device Compatibility	Works Well	Fully functional across standard secured laptops, rugged tablets, and offline-compatible field devices.
	Offline Functionality	Works Well	Core features (scenario playback, basic feedback) work offline; syncing resumes when connection is restored.
	Interoperability	Minor Concern	Compliant with SCORM and xAPI, but lacks integration with some simulation command-and-control systems.
Pedagogical Impact	**Cognitive Engagement**	Works Well	AI challenges learners through branching scenarios requiring judgment, tactical flexibility, and reflection.
	Scenario-Based Learning	Works Well	Excellent fit for pre-deployment training; integrates decision points, varying rules of engagement, and multilingual dilemmas.
	Learning Analytics	Minor Concern	Delivers rich data, but instructors report difficulty interpreting some visualizations without training.
	AI-Powered Feedback	Works Well	Feedback is tailored, timely, and delivered in natural language. Instructors can review and override AI suggestions.
Ethics, Bias, & Security	**Bias Mitigation**	Minor Concern	Datasets reviewed for demographic and cognitive diversity, but edge cases (e.g., cultural biases in mission language) still appear occasionally.
	Data Privacy & Security	Minor Concern	Uses accredited cloud but pending final audit for classified data handling.
	Transparency	Works Well	Decision pathways and scoring logic are available via instructor dashboard; AI outputs are interpretable.
Cost & Sustainability	**Cost Efficiency**	Minor Concern	Developed under a grant, but replication cost for other units may be high without subsidization.
	Energy Efficiency	Minor Concern	Efficient under standard loads, but requires GPU acceleration during scenario branching or NLP generation.
	Long-Term Viability	Works Well	Backed by a vendor-university partnership with clear update schedule and roadmap for continued support.
Classified Information & Accreditation	**Classified Data Handling**	Serious Concern	Scenario is not authorized for use with classified documents or data and lacks classified handling certification.

	Data Storage & Sovereignty	Minor Concern	Data stored in accredited European servers, but some logs from pilot sessions were temporarily stored in non-sovereign jurisdictions.
	Accreditation & Certification	Minor Concern	Pending ISO/IEC 27001 and FedRAMP Medium certifications; partial compliance with accreditation norms.