Artificial Intelligence as Part of Army Simulators

Nordefco ADL 2026
From simulator support tools to adaptive adversaries and AI-enabled learning loops
PUBLIC RELEASE

Presentation focus

1. Where AI adds practical value now
2. What a feasible architecture looks like
3. Why governance and Nordic cooperation matter

Core message

AI should strengthen instructor work and training quality - not replace human control.

Dr. Kalle Saastamoinen
Finnish Defence Forces | Simulator Centre of Excellence

Why discuss AI in army simulators now?

Training realism
AI can make entities, scenarios, and feedback more responsive to what trainees actually do, instead of relying only on static scripting.

Decision support
Language models and analytics can help turn mission inputs into structured options faster, while keeping the instructor in control.

Operational adaptability
The same ideas support training in multi-domain, DDIL, and rapidly changing environments.

What has changed?
The difference is not only better models. The important shift is that large language models, generative agents, and real-time analytics can now be combined into one training workflow: from scenario authoring to execution support and AAR. This makes AI relevant as an enabling layer across the training cycle rather than a single point solution.

Current simulation context in the Finnish Army

Live
Saab Live Training enables realistic force-on-force training with rich data capture, control, and AAR.

Virtual
VBS supports tactical training and mission rehearsal, making it a natural target for AI-assisted scenario generation and adaptive behaviour.

Constructive
MASA SWORD supports staff training and multi-domain coordination, where AI can help create and compare plans rapidly.

Specialized
Steel Beasts and other focused systems illustrate the need for simulator-agnostic integration rather than a single platform solution. Saab GC IDT is a focused indoor combat trainer combining replica weapons, a virtual environment, instant feedback, and AAR.

Implication

The strongest opportunity is to connect AI capabilities across this mixed ecosystem. That means common interfaces, shared data models, and outputs that can travel between systems instead of remaining locked inside one tool.

Three immediate AI use cases

1. Scenario generation from natural language

Intent can be entered in structured natural language instead of manual scripting.

Benefits:
- Faster authoring
- Easier iteration
- Better reuse of instructor expertise

2. COA support

AI can propose doctrinally aligned options, compare alternatives, and preserve a human approval loop.

Benefits:
- More options for planning
- Faster refinement
- Better transparency than ad hoc editing

3. Automated AAR

Language-based summarization and analytics can reconstruct what happened and why.

Benefits:
- Less manual workload
- More consistent feedback
- Better learning capture across exercises

Army Academy

ARCHITECTURE

Conceptual architecture: from intent to simulation action

User intent

Commander, instructor, or analyst describes training goals, constraints, and scenario context.

>

AI core

LLM reasoning with retrieved doctrine and scenario corpora.

>

Controlled output

Structured scenario representation plus validation before use.

>

Simulation adapters

C2SIM or other simulator-facing interfaces for VBS, SWORD, live systems, and others.

>

AAR loop

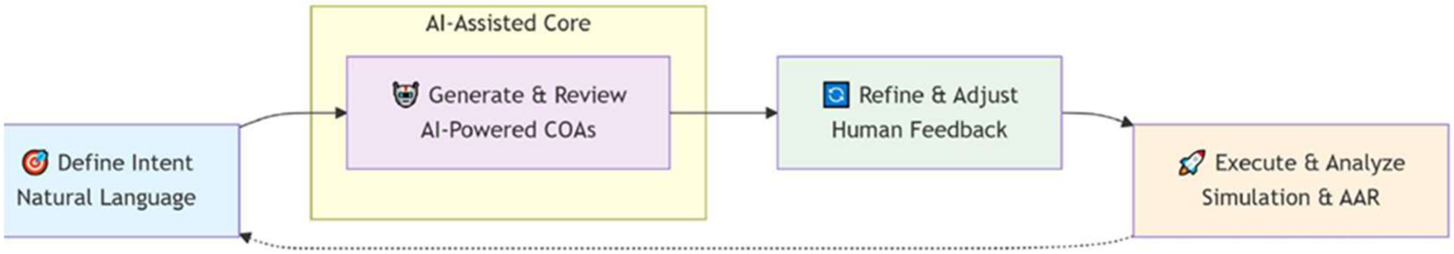
Exercise data feeds summary, coaching insights, and next-scenario refinement.

Design principle

Use AI for speed and variation, but keep doctrine, standards, validation, and human judgement in control. A useful AI capability should not be trapped inside one simulator.

Human-on-the-loop: generate, refine, execute, learn

The system is meant to accelerate instructor work, not remove instructor judgement



Instructor role remains central

- Select among COA variants with different tempo and risk profiles
- Refine using natural-language feedback such as deception, fires, EW, or DDIL injects
- Approve only after plausibility and technical checks are satisfactory

AAR closes the loop by informing the next scenario iteration.

Beyond scripting: adaptive adversaries and generative behaviour

What changes?
Instead of only hand-authored if-then-else logic, opposition forces can combine doctrine, constraints, goals, and learned policies to react more dynamically.

RL and MARL
Reinforcement learning is attractive when we need a tactically adaptive opponent that can change behaviour without following one pre-scripted branch.

Generative support
Language models can help with planning, dialogue, and behaviour shaping - especially when interaction with instructors matters.

Practical implication
A hybrid approach is likely to be strongest: LLMs for intent, structure, and rapid scenario formation; learned agents for dynamic opposing-force behaviour where adaptation matters most.

Preliminary Evidence: Feasibility from Related Research

Evidence from Multi-Agent Reinforcement Learning (MARL):

- Littoral Warfare Study (Vasankari & Saastamoinen, 2024): MARL algorithms generated COAs that were 5x more likely to win and inflicted 8x the damage.
- Red Force Response Tool (Frazer-Nash, 2025): AI adversary achieved a 91% win rate in wargaming, outperforming all scripted COAs.

Insight: MARL is powerful but computationally expensive for training.

Evidence from LLMs (COA-GPT) and MARL:

- Generates initial, doctrinally-grounded COAs within seconds.
- Outperformed state-of-the-art RL baselines in mission reward.
- Highly adaptable to human feedback, creating an effective human-AI collaboration loop.

The Future is Hybrid: Combining LLMs and MARL

The Best of Both Worlds:

- **LLMs**: Rapid, intuitive planning and scenario generation.
- **MARL**: Adaptive, fine-grained tactical execution and dynamic adversary behavior.

Vision: Use the LLM to generate the high-quality initial plan, then let trained MARL agents execute it or control the adversary for a truly dynamic and challenging training experience.

Automated AAR as the most immediate learning gain

Data in
Logs, movement traces, event timelines, command inputs, outcomes, and relevant exercise metadata.

AI analysis
Reconstruct event narratives
Identify key decision points
Compare actions against intended COA
Summarize for instructor review

Instructor use
Faster debrief preparation
More consistent feedback
Alternative scenario exploration
Better focus on educational value

Next cycle
AAR findings inform the next scenario revision, improving training iteration instead of ending at the debrief.

Why this matters: AAR is where AI can deliver clear value with relatively low operational risk.

CONSTRAINTS

Deployment challenges that matter for defence learning

Edge and DDIL conditions

AI support becomes harder in dispersed or disconnected environments and may require smaller local models or carefully staged workflows.

Governance and doctrine

Output must be doctrinally grounded, validated, and controlled. Retrieval quality matters, but so does corpus governance.

Interoperability

If AI outputs cannot cross simulator boundaries, the educational gain stays fragmented and hard to scale.

Human-on-the-loop remains central

The system should accelerate instructor work, not remove instructor judgement. Human approval is needed for plausibility checks, training value, and operational sensitivity.

Doctrine is not the same as authoritative data

Doctrine / conceptual guidance

- Tactical principles, sequencing, forms of manoeuvre, control logic
- Used to shape plausibility and explain why a COA makes sense
- Best managed as a curated, tiered retrieval corpus

Authoritative structured data

- ORBATs, METs, capability libraries, terrain and operational-environment data
- Must remain authoritative, queryable, and separately validated
- The model should reference these facts — not invent them

RAG helps — but only if the corpus is governed

What RAG adds

- Improves doctrinal plausibility
- Supports traceability back to source passages
- Constrains generation better than free-form prompting alone

Where it starts to break

- Large, heterogeneous corpora reduce retrieval precision
- The model may combine passages that are individually correct but contextually mismatched
- Risk: plausible-sounding but weakly grounded outputs

Why this matters for Nordic cooperation

Alignment points

NATO MSG-229 and MSG-232 activities provide a useful frame for architecture, interoperability, and training innovation.

Shared outputs

Reference scenarios, prompt libraries, validation criteria, and doctrine-grounded data practices could be developed collaboratively.

Educational benefit

Joint experimentation can reduce duplication, improve interoperability, and strengthen the regional defence learning ecosystem.

Nordefco opportunity

Shared innovation in simulation and training technologies can support interoperability while still respecting national doctrine, language, and governance requirements.

A feasible pilot path for 2026-2028

Phase 1
Scenario-authoring assistant

Focus on natural-language input, doctrine grounding, and structured scenario output for one or two training problems.

Phase 2
Automated AAR support

Use exercise logs to generate debrief summaries and identify key decision points for instructor review.

Phase 3
Adaptive opposing force experiments

Evaluate when learned agents improve realism enough to justify added complexity.

Measure success
Compare manual and AI-assisted workflows on speed, usability, doctrinal plausibility, and educational value - not speed alone.

Evaluation: what will be measured?

This is an intended pilot evaluation scenario

Efficiency
Time on task
Iteration count
Baseline vs. assisted workflow

Technical validity
First-pass C2SIM validity
Repair loops
Failure types

Doctrinal plausibility
SME scoring
Traceability to sources
Feasibility / training value

Human performance
Perceived workload
Usability
NASA-TLX

Illustrative pilot setup

Compare manual authoring against the LLM-assisted workflow across multiple scenario types during 2026-2027 staff-training and simulation activities, subject to prototype maturity and exercise schedules.

Why this matters: the evaluation is designed to show not just whether the system is faster, but whether it remains technically usable and militarily plausible.

Takeaway for Nordefco ADL 2026

The central proposition is simple: AI first where it improves learning efficiency and realism without removing human control.

Start with scenario generation, COA support, and automated AAR. Build on governed doctrine, simulator-agnostic interfaces, and a human approval loop. Then extend toward adaptive adversaries where added training value justifies the added complexity.

Discussion prompts

Which simulator layer should be piloted first: authoring, adversary behaviour, or AAR?

How should Nordic countries share validation practices without losing national control?

What is the right balance between central cloud support and edge-capable local AI?

Army Academy

Thank you - discussion

Artificial Intelligence as Part of Army Simulators

Nordefco ADL 2026

Contact

Dr. Kalle Saastamoinen
Finnish Defence Forces
Simulator Centre of Excellence
kalle.saastamoinen@mil.fi
ORCID: 0000-0001-7471-2404

Questions for collaboration are especially welcome around scenario authoring, AAR, validation, and Nordic interoperability.