Vol. n. 8 • 2025 ISBN 979-12-985129-2-4

2025 Annual Review

NATO
Modelling & Simulation
Centre of Excellence

Published by NATO Modelling & Simulation Centre of Excellence, Rome, Italy.

Edition: VIII (February 2026)

ISBN 979-12-985129-2-4

Printed in Italy.

Disclaimer

Release

Published and distributed by

The NATO Modelling and Simulation Centre of Excellence
Piazza Renato Villoresi, 1
00143 Roma
Italy

Table Of Content

Preface

Dear M&S Community of Interest,

As we move forward into another year of progress and transformation, it is my privilege to introduce the 2025 edition of the Annual Review. This publication brings together the collective achievements, lessons, and innovations emerging from both the NATO Modelling & Simulation Centre of Excellence and the wider M&S community.

The theme of 2025 edition, "From Simulation to Action: Wargaming, Exercising, and Experimenting for Multi-Domain Advantage," captures the essence of our shared ambition to translate modelling and simulation insights into tangible operational outcomes. It reflects our growing focus on using M&S not only to understand complexity, but to drive decisions, readiness, and interoperability across all domains.

This Annual Review highlights contributions from across our community and the 2025 CAX Forum, spanning the key thematic tracks of Wargaming, Experimentation and Decision Support in Multi-Domain Operations; Interoperability, Standards, Integration and Development in Synthetic Environments; and Commercial Technology. Together, these perspectives underscore the central role of collaboration and innovation in shaping the future of defense and security.

I would like to express my sincere appreciation to all authors, researchers, and practitioners whose work is featured in these pages. Your dedication and creativity continue to advance the mission of the NATO M&S COE and strengthen our collective capability to act decisively in an increasingly complex environment.

Happy reading!

Best regards,
Col. Felice De Rosa, Director
NATO M&S CoE Director


NATO
OTAN
MODELLING & SIMULATION
HONOS ALIT ARTES
CENTRE OF EXCELLENCE
23 | 24 | 25 SEPTEMBER 2025
ROME, ITALY
20TH EDITION
FROM SIMULATION TO ACTION: WARGAMING, EXERCISING AND EXPERIMENTING FOR MULTI-DOMAIN ADVANTAGE
CA2X2
COMPUTER ASSISTED EXERCISE EXPERIMENTATION
FORUM
MSCOE.ORG

The team of the NATO Modelling and Simulation Centre of Excellence would like to thank you for your participation in the 2025 Forum.

The CA²X² Forum 2025 Sponsors Recognition

Main Sponsor

CALIAN
Confidence. Engineered.

Platinum Sponsors

Bohemia Interactive SIMULATIONS

LEONARDO

MATRIX PRO SIMS

MISSION SYSTEMS

PITCH

BV TECH

Digivents

Gold Sponsors

teorest group

aurizn

Silver Sponsors

bip.

C5I

The NATO Modelling and Simulation Centre of Excellence wishes to thank the sponsors for their contribution to this year's conference and for assisting with making it an incredible achievement.

This book contains the proceedings of NATO M&S CoE's Computer Assisted Analysis, Exercise, Experimentation annual conference held from 23 - 25 September 2025 in Rome, Italy.

The principal theme for the conference was:

'From Simulation to Action: Wargaming, Exercising, and Experimenting for Multi-Domain Advantage'

Through team effort at the M&S COE we have captured the articles from the CA2X2 Forum allowing our readers to reference the great work done by some of the contributors.
Please use these articles as inspiration for further collaboration and contributions to these important topics.

Thank you for the contributions to the forum,
the insightful questions and discussion to advance these topics.
For those that were unable to participate, this collection of articles will help you understand the level of expertise and professionalism that was displayed during the forum.
Enjoy.

If you wish to provide feedback, please send it to us at: info@mscoe.org.

Thank you and good reading!
The NATO Modelling and Simulation Centre of Excellence

CASTLE (CBRN Activities Simulation Total Layer Environment) Project

LTC Piergiorgio Ventura
NATO M&S CoE

www.mscoe.org

Abstract

Modelling & Simulation in support of CBRN (Chemical Biological, Radiological and Nuclear) and Environmental Protection has not been fully exploited to its maximum potential within the M&S areas; namely, Education and Training (Exercises), Support to Operations, Planning (Course of Action Analysis), Execution (Decision Support), Mission Rehearsal, Concept Development & Experimentation (CD&E) and Procurement. Many CBRN tools already exist, such as those providing models to simulate the dispersion of CBRN Agents, or the wearing of IPE during training. However, these tools are limited in their scope. A comprehensive approach to maximize its effectiveness is still missing.

The CASTLE (CBRN Activities Simulation Total Layer Environment) project is an innovative approach, which integrates existing tools and is prepared to integrate future tools not yet developed. It represents a powerful M&S asset to fill the gap in this military problem. This concept was described in detail in the 2023 Report and related I/ITSEC 2023 paper detailing the initial gap analysis. The proposed CBRN layer has been created and tested. Several dispersion simulation tools for the CBRN agent are included in the architecture (ALOHA, HOTSPOT) and the data received from the NATO JCBRN CoE obtained by

HPAC were also included in the testing activities. The architecture includes WISDOM, a Wargaming platform useful for visualization and share information with its HLA plugin, MASA SWORD and VBS4 as CGF (Computer Generated Forces) tools. The scenarios have been built to simulate a synthetic CBRN environment

with contamination and diffusion data. Several tests have been performed to verify the possibility to integrate all components of the architecture using an ad hoc Application Protocol Interface (API) and share with CGF through HLA. The final objective is to test the capability of providing the Commander with a comprehensive synthetic visualization of the CBRN framework on the battlefield.

Keywords: CBRN layer, M&S, Integration, Interoperability.

About The Authors

Author: LTC (ITA – OF4) Piergiorgio Ventura graduated in physics in 1998 with a specialization in Nuclear Physics. He then joined the Italian Army in 1999 with the rank of Lieutenant and began working within experimental firing ranges where missiles, weapon systems and ammunition were tested. After taking a PhD in quantum electronics and plasma physics in 2010, during which a remote sensing detection system to detect and identify chemical compounds, based on optical detection, was developed, he started working in the CBRN field for research, testing and procurement activities. Since January 2022, he has been assigned to the M&S COE as the M&S Concept Development Section Chief, where he is trying to develop new concepts based on his expertise.

1. Introduction

In 2022 and 2023 a Gap analysis had been performed in order to identify what are the gaps to be filled while using M&S for CBRN activities and, on the other way round, how to integrate CBRN activities for wider use of this peculiar type of military activities in commonly used M&S tools. The results was that a CBRN Layer is needed in order to be used in all M&S application areas, namely, Education and Training (Exercise), Support to Operations, Planning (Course of Action Analysis), Execution (Decision Support), Mission Rehearsal, Concept Development & Experimentation (CD&E) and Procurement. The gap analysis had been used also to identify available tools, standards and previous experimental activities. Based on this study, the results showed that many CBRN tools already exist, such as those providing models to simulate the dispersion of CBRN Agents, or the wearing of IPE during training. However, a comprehensive approach to maximize its effectiveness is still missing. These analysis had been published in I/ITESEC proceedings 2023. For these reasons, a new integrated architecture to develop a CBRN Layer is under development at the NATO M&S Centre of Excellence, named CASTLE (CBRN Activities Simulation Total Layer Environment) project. It is an innovative approach, which integrates existing tools and provides those not yet developed. It represents a powerful M&S asset to fill the gap of this military

problem. ALOHA and HOTSPOT dispersion simulation tools for the CBRN agent are part of the CASTLE architecture and the NATO JCBRN CoE HPAC data were included in the testing activities. CASTLE consist also of MASA SWORD and VBS4 as CGF (Computer Generated Forces) tools. The scenarios is built to simulate a synthetic CBRN environment with contamination and diffusion data. Several tests have been performed to verify the possibility to integrate all components of the architecture using an ad hoc Application Protocol Interface (API). The final objective is to test the CASTLE capability of providing the Commander with a comprehensive synthetic visualization of the CBRN framework on the battlefield.

2. *Castle Proof Of Concept Description*

The goal of the project is to integrate, in the same architecture, all available CBRN "expert systems" in order to share CBRN events data, such as the propagation of a chemical plume over a specific terrain with Computer Generated Forces (CGF) M&S tools. The architecture will also include an expert engineering system for specific calculations (modelling of sensors, protective equipment, etc.).

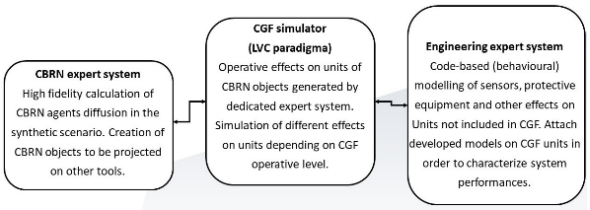


***Figure 1**. CASTLE project general concept.*

Starting from this objective, a proof of concept has been developed. In details, the architecture's structure includes several tools to be used as expert systems and is designed with the ability to incorporate additional expert systems. For this prototype, tools that are available for free and provide outputs that are reasonable, useful and easy to handle have been included as "Expert Systems", as identified during the gap analysis described in ref. 1. Furthermore, the possibility to use such free software allows every organization to start using the proposed architecture without costs, a cooperation agreement with NATO M&S CoE is the only preliminary constrain. The tools chosen among the one with free download are ALOHA for chemical events and HOTSPOT for the radiological and nuclear events. Additional tools, such as CBRN Analysis and HPAC are available to the center thanks to cooperation from other organizations and have been considered for inclusion in the project. While both useful for any kind of CBRN event. CBRN analysis provides an oversimplified prediction tool that is not useful to share technical data and has not been included in the prototype. HPAC provides good quality technical prediction and therefore has been included thanks to the cooperation with Joint CBRN Centre of Excellence.

The main characteristics of the tools are described below:

The HotSpot Health Physics Codes, or HotSpot program, developed and regularly updated by the USA National Atmospheric Advisory Center (NARAC), provides a first-order approximation of the radiation effects associated with the atmospheric release of radioactive materials. It was created to equip emergency response personnel and planners with a fast, field-portable set of software tools for evaluating incidents involving radioactive material. The software is also used for safety-analysis of facilities handling radioactive material. This program is designed for short-range (less than 10 km), and short-term (less than a few hours) predictions.

ALOHA®, developed and maintained by the USA Environmental Protection Agency (EPA), is a hazard modelling program which is used widely to plan and respond to chemical emergencies. It allows you to enter real or potential chemicals, release detailed data and then generate threat zone estimates for various types of hazards. It also models toxic gas clouds, flammable gas clouds, BLEVEs (Boiling Liquid Expanding Vapor Explosions), jet fires, pool fires, and vapour cloud explosions. The threat zone estimates are shown on a grid, and they can also be plotted on maps in MARPLOT® (Mapping Application for Response, Planning, and Local Operational Tasks), Esri's ArcMap, Google Earth, and Google Maps. The threat zone depicted in red represents the worst hazard level, and the orange and yellow threat zones represent areas of decreasing hazard. The results are a static image after one hour.

HPAC, owned by Defense Threat Reduction Agency (DTRA), models and predicts human collateral damage for events involving intentional or unintentional release of chemical, biological, or nuclear materials into the atmosphere or enclosed space. It provides a suite of models for simulating the release of CBRNE materials into the atmosphere and their associated dispersion using detailed meteorological information. These

predictions are used to estimate the effects of these CBRNE agents on the physical environment and, to a lesser extent, the resulting impact of that release on an exposed population. It can describe the transport/dispersion of hazardous materials through the atmosphere due to attacks or accidents resulting in radiological, chemical, or biological releases. It uses information on the material source, the amount released into the atmosphere, high-resolution weather forecasts, and particulate transport to model the hazard areas produced by such events.

Computer Generated Forces (CGF) tools are the second pillar of the project. These are used to model and simulate the behavior of military units in the battlefield. There are several tools used for this purpose among NATO countries, many of which are able to share information about their objects and entities using a standardized protocol named High Level Architecture (HLA). Therefore, some specific tools have been used for our testing activities, but potentially any tool able to properly use HLA protocol, including the CBRN specific protocol, developed by NATO as a specific Federation Object Model (FOM), could be potentially be included in the federation. The tools have been consequently chosen for their availability and flexibility for the testing activities. For this reasons, SWORD, developed by MASA Company has been chosen for the Operational level testing, whereas VBS4, developed by Bohemia Interactive Simulation has been chosen for Tactical Level testing.

The main characteristics of the tools are described below:

SWORD, developed by MASA Company, is a suite dedicated to staff education and training, which includes a scenario building application, an aggregate simulator and an analysis tool. Its main purpose is within the training environment but it also support planning and education. SWORD allows aggregate-level simulations for tactical training, which is oriented towards operational field decisions. It is normally used at Operational level, it has an HLA plugin and is compatible with several languages including C#. Lastly, it has several embedded CBRN features and is able to share them with HLA, making it a very versatile platform, easily adaptable to receive and share information with others systems.

VBS4, developed by Bohemia Interactive Simulations, is a whole-earth virtual desktop trainer and simulation host that allows you to create and run any imaginable military training scenario at tactical level. The VBS4 workflow steps the user through Prepare - Execute - Assess phases, facilitating fast and effective skills enhancement. In VBS4, users create "Battlespaces" that are a collection of terrain edits, mission plans, scenario files and after-action reviews. With VBS4, you can plan your mission, build your terrain with easy-to-use interfaces, and focus on the learning points of the exercise rather than the technical aspects of setup. It has several embedded CBRN features and it is able to share them with HLA. It is therefore a good platform to be adapted to receive information and share with others.

The main pillar of the architecture is WISDOM (Wargaming Interactive System Digital Overlay Model), a software developed over the years by NATO M&S CoE, useful for the configuration of geographical scenarios. This platform is particularly suitable to carry out Wargaming activities and is aimed at those who need to perform education & training activities, experimentation, AAR or preparation for real missions, both in the military and civilian context. The platform is designed to reuse and tailor available scenarios or to develop new scenarios in the tactical, operational or strategic level. WSIDOM uses a data model from PostGIS DataBase (spatial extension of PostgreSQL Database) based on Linux Ubuntu. Data are elaborated/created from QGIS which is a software application able to manage spatial data (visualize, analyze editing). With QGIS, one can edit GIS data and create a further new database for PostGIS; this database will be afterward loaded and utilized on WISDOM that is in turn based on QGIS Server. WISDOM therefore is a GIS server platform based on WebGIS Architecture.

Due to its state of the art architecture, WISDOM allows wargamers around the globe to access the platform from their location as it is built to support distributed execution. The MSCOE is presently upgrading WISDOM flexibility and integration capability with M&S tools and is used in this project as bridge between ESs and CGFs. For this purpose, a specific HLA plugin has been implemented to share entities with M&S platforms even without a time management process.

In order to use HLA, a Run Time Infrastructure (RTI) is also required. Among the few available on the market, the product most frequently used by our possible partners has been chosen, in order to limit as much as possible mistakes not determined by our architecture but instead related to different customizations with third parties. The product Pitch pRTI™ is an implementation

of the IEEE 1516 Interface Specification. It lets you integrate simulations in an HLA compliant way. You can mix different operating systems and programming languages. It gives you the ability to integrate your existing C/C++ simulators with platform independent Java systems. Pitch pRTI™ provides APIs for both C++ and Java, so you can use federates written in any of those languages together in the same federation. It provides advanced debugging capabilities. It has an extensive GUI that allows you to inspect the state of your federation during runtime as well as a powerful set of debugging tools.

The links used from ESs to WISDOM are Application Protocol Interfaces (API), based on JAVA and Python. That allows to handle the original data format from expert systems to be modified and included in the Quantum GIS database where WISDOM is based to run its own database. The basic idea is to modify the data structure while importing it in the virtual machine that host WISDOM.

Some of the Expert systems and CGF tools have also been used as Engineering expert systems, considering that elaborated data, such as the people affected, were already calculated within specific branches of their algorithms.

One important characteristic of the project is that it utilizes an open architecture. It relies on the flexibility and adaptability of WISDOM to be ready to include results from other expert systems that provide outputs, requiring only development of the specific Application Protocol Interface, if needed. The other part of the architecture, going towards CGF tools, is even simpler because it is based on a civil protocol, IEEE 1516 "Standard for modeling and simulation (M&S) high level architecture (HLA) - framework and rules", that become also a military standard, STANAG 4603 "Modelling and simulation architecture standards for technical interoperability: high level architecture (HLA)". So it is really easy to include these tools; they just need to be complaint with the standard and that, among the available protocols, the specific CBRN NATO Education and Training Network Federation Object Model is included. If they are, they just need to join the federation, as if they were in a standard exercise. WISDOM could be updated and modified with the only costs of the developing phase (if not performed by Ministry of Defense personnel) because it is not based on costly third party licenses. Being used as test benchmark, it is easily attracting potential partners which would like to be included in the architecture, representing a win-win solution.

Thanks to the most recent developments, WISDOM is now able to visualize and update rapidly evolving CBRN objects, continuously sharing them in real time using its HLA plugin. These results have been appreciated by MASA company, which helped in the experimental activities with SWORD and stated these kinds of updates have never been done in the past. Moreover, WISDOM is now able to handle time so as to be easily managed to get synchronization with any other tool in the federation.

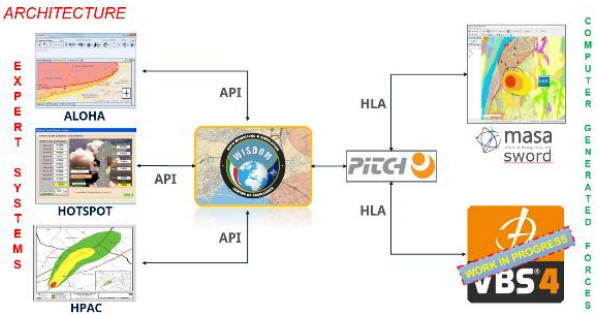


***Figure 2**. CASTLE project architecture.*

3. *Proof Of Concept Testing Activities*

The CASTLE project has completed its concept development phase and a proof of concept has been delivered. The system is able to share the information related to CBRN events calculated by Expert systems with other tools, mainly CGF using WISDOM as main database, visualization tools and bridge. Many different kinds of events have been calculated and then imported in the WISDOM database. As first example, a simulation performed by HOTSPOT and shared within the architecture is shown in figure 3. On the left side the original calculation related to a plutonium contamination is shown, then the visualization on WISDOM is visible in the center. Finally, using it as an HLA bridge, the contaminated area is shared on SWORD, as shown on the right side.

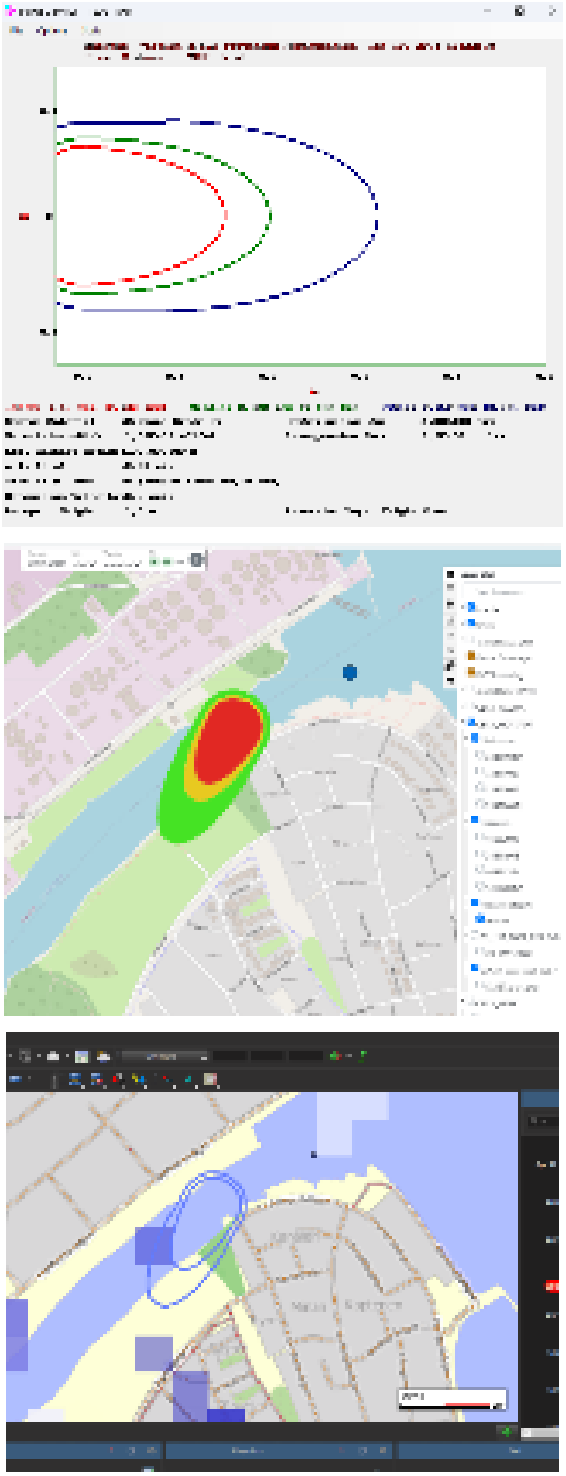


***Figure 3**. Plutonium plume shared from HOTSPOT (up) through WISDOM (center) to SWORD (down)*

Also several chemical events have been included in the WISDOM database, using ALOHA as "expert systems". An example of the events used for the test is a SARIN contamination shown in figure 4. Also in this case the event is imported from ALOHA to WISDOM and then shared using HLA with SWORD.

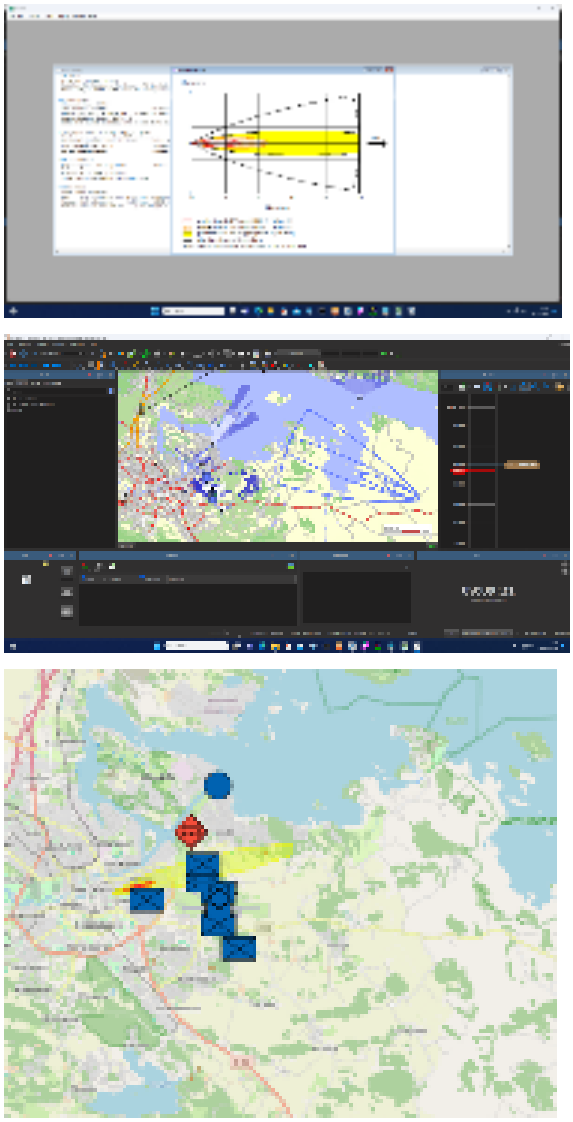


***Figure 4**. Sarin plume sharing from ALOHA (up) through WISDOM (center) to SWORD (down)*

The use of HPAC is not directly included in the proof of concept architecture. However, the HPAC files integration capability has been verified using data directly provided by the NATO JCBRN CoE and therefore the tool can be used as "expert systems" for any kind of CBRN events too. The testing was conducted with a simulation of a SARIN contamination imported on the WISDOM database and, in this case, evolving over time. The results of the experiment confirmed that the architecture was able to share the files with SWORD using HLA. The three steps of the evolution over time are shown in WISDOM (figure 5 up) and the equivalent in SWORD (figure 5 down).

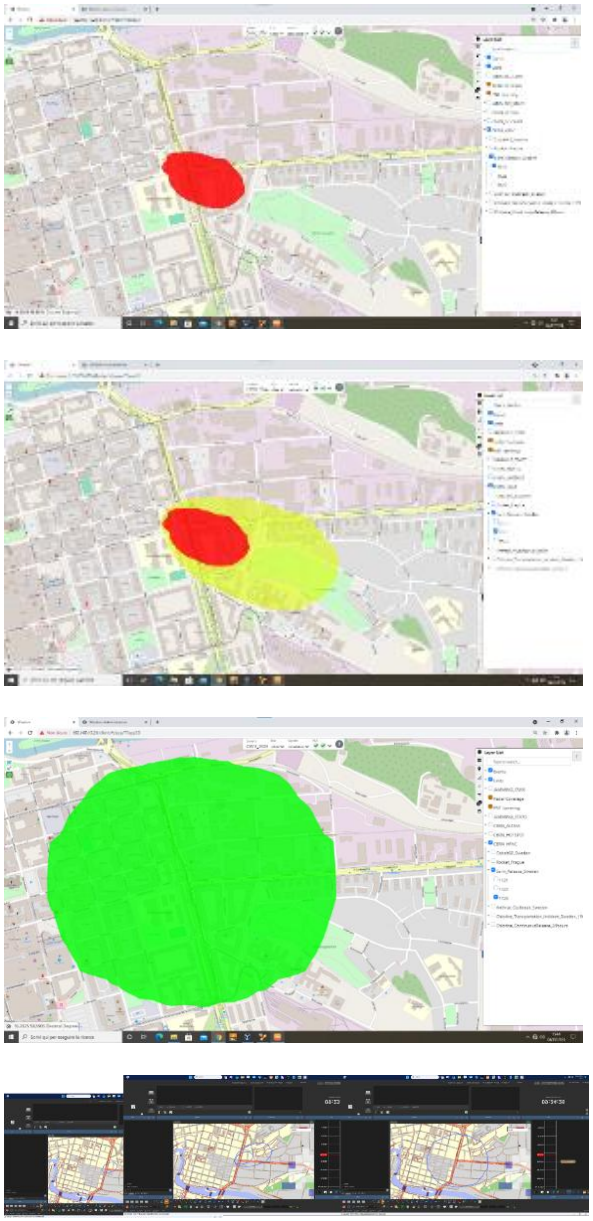


***Figure 5**. Sarin plume sharing from HPAC evolving over time in WISDOM (up) and shared in SWORD (down)*

To make a database able to represent the full spectrum of CBRN events, also Biological examples have been included. The following figure shows Anthrax diffusion generated by HPAC and shared by WISDOM to SWORD through HLA, as in the previous cases.

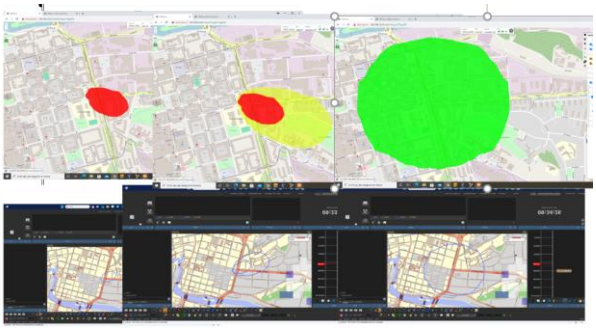


***Figure 6**. Antrax plume sharing from HPAC evolving over time in WISDOM (up) and shared in SWORD (down)*

4. Future Activities and Conclusion

The CBRN and M&S integration project (CASTLE), conducted using the available tools at the MSCOE and performing the testing activities described, confirms that the initial Concept is valid and the developed architecture is effective. The project has completed its first phase of experimentation as planned in the MS COE program of work (POW). The next phase will be conducted with additional interoperability tests and architecture implementation as follows:

Interoperability tests with additional Computer Generated Forces, starting from VBS4 in order to represent CBRN data on a tactical simulator (FOM adaptation might be required). Many experiments are planned while participating to Coalition Interoperability Warrior eXercise (CWIX) during 2025.

Assess the "CBRN effects" realism processed and delivered by the Expert System (HPAC, HOTSPOT) or internally by Computer Generated Forces tool.

Assess the NETN FOM capability to take into consideration the CBRN effects and propose modification and/or integration to the NMSG standardization community, if required.

Include an Engineer Expert System in the architecture to cover specific CBRN effects (e.g. filter duration).

Include C2 Systems in the architecture to increase interoperability capabilities of the project.

The MS COE is exploring the possibility of conducting parallel activities linked to the CBRN project involving additional stakeholders from the M&S/CBRN community of interest (COI) as follows:

foster cooperation with other institutions to integrate their activities in one broader project, bringing all the

stakeholders around the same table, sharing the existing resources and analyzing additional capabilities with the objective to extend the software included in the architecture and the potential users, so being capable of providing effective decision making support to the Alliance.

Test individual capability and the whole architecture for the verification and validation process.

Make the deliverable available to the entire COI and to NATO in order to obtain the maximum benefit for the Alliance and improve the tool itself through users and developers contributions.

The architecture developed proved to be effective and able to share CBRN activities, as calculated by expert systems, with WISDOM and, through it, with potentially any possible simulation systems that rely on HLA to share information. Using the system will push it towards users' needs and, thanks to their contribution, it will continue to evolve extending its capabilities, importing new tools to become a "system of systems" architecture.

The challenge for the future seems to be sharing this objects within the C2 architectures, where these seem not to be usually visualized. A different approach could be to share the effects, e.g. export the information that a unit or platform is contaminated because it enter in a contaminated area. In the latter case, the evaluation on each contaminated platform is made within CASTLE and then exported towards the C2 system using the applicable Tactical Data Link Protocol to modify the status of each units or platform entering the contaminated areas.

References

[1] LTC Piergiorgio Ventura (ITA Army) and CPT Salvatore De Mattia (ITA Army), Modelling & Simulation in support of a comprehensive CBRN Layer development, in 2023 Interservice/Industry Training, Simulation and Education Conference, Orlando, FL, Nov 2023.

[2] Charles McLean, Y. Tina Lee, Dr. Sanjay Jain, Dr. Charles Hutchings, Modeling and Simulation of Hazardous Material Releases for Homeland Security Applications. – National Institute of Standards and Technology - NISTIR 7786, 2011

[3] NATO STANAG 2499 "The effect of wearing CBRN individual protection equipment on individual and unit performance during military operations" – ATP-65(B)

[4] NATO STANAG 4625 "Assessment of effect levels of classical chemical warfare agents applied to the skin to be used in the design of protective equipment" – AEP 52

[5] Lt Col Walter David et al., Crisis Decision-Making with M&S Support in Complex Urban Environments, I/ITSEC 2018, from https://www.researchgate.net/publication/331044317_Crisis_Decision-Making_with_MS_Support_in_Complex_Urban_Environments

[6] Jon Lloyd, Nathan Newton and Richard Perkins, A Chemical, Biological and Radiological Modelling Capability to Support Acquisition Advice and Re-use as a Common Cross-Domain Capability., DSTL – UK, - STO-MP-MSG-126, from https://www.sto.nato.int/publications/STO%20Meeting%20Proceedings/STO-MP-MSG-126/MP-MSG-126-09.pdf

[7] Jon Lloyd, Nathan Newton, Jose Ruiz, David Desert, Antony Hubervic, Lennart Olsson and Russell Mills, A Common Chemical, Biological & Radiological modelling capability: UK and NATO HLA-Evolved experimentation. from https://www.sisostds.org/DesktopModules/Bring2mind/DMX/API/Entries/Download?Command=Core_Download&EntryId=42367&PortalId=0&TabId=105

[8] Orlin NIKOLOV, M&S support for Crisis and Disaster Management Processes and Climate Change implications, https://drmkc.jrc.ec.europa.eu/Portals/0/Partnerships/Seminars/3_Scientific_Seminar_DRMKC/Presentations/session_4b/pdf/3_orlin_nikolov-m_and_s_support_for_crisis_and_disaster_management_processes_and_climate_change_implications.pdf

[9] Ms. Gail Cayce-Adams and Mr. Michael Kierzewski, JPEO CBRND develops a new concept to better manage its portfolio, from https://asc.army.mil/web/news-alt-jas18-analytical-framework/

[10] Hazard Prediction and Assessment Capability (HPAC), by Defense Threat Reduction Agency

[11] VBS4, from https://bisimulations.com/wp-content/uploads/2025/03/bisim_product_flyers_2024_vbssubscription.pdf

[12] MASA SWORD, from https://www.masasim.com/en/sword

[20] ARCHARIA, from https://www.mscoe.org/nato-ms-coe-archaria/

[21] STANAG 4603 "Modelling And Simulation Architecture Standards For Technical Interoperability: High Level Architecture (HLA)"

[22] IEEE 1516 Standard For Modeling And Simulation (M&S) High Level Architecture (HLA) - Framework And Rules

[23] SDR 4603.1 National Systems Conformant To Hla Standards Formats

[24] IEEE 1516.1 Standard For Modeling And Simulation (M&S) High Level Architecture (HLA) - Federate Interface Specification

[25] IEEE 1516.2 Standard For Modeling And Simulation (M&S) High Level Architecture (HLA) - Object Model Template (OMT) Specification

Mathematical Modelling of Cyber Warfare: A Quantitative Framework for Pre-Conflict Analysis

LTC Bernd Weissenberger
NATO M&S CoE

www.mscoe.org

Abstract

The evolution of warfare in the digital age has transformed cyberspace into a central domain of strategic competition and military operations. Nation-state actors increasingly employ cyber tools to achieve geopolitical objectives without resorting to kinetic force, exploiting vulnerabilities in interconnected critical infrastructures. This paper presents a mathematical framework to quantify and simulate the pre-conflict phase of cyber warfare, integrating offensive and defensive cyber capabilities (OCC and DCC) within the PMESII analytical model. The proposed approach provides a structured, quantitative foundation for assessing the strategic effects of cyber operations and supports decision-making under uncertainty.

Keywords: cyber warfare, mathematical modelling, PMESII, simulation, hybrid warfare, strategic analysis

1. Introduction

1.1 Background

In the 21st century, cyberspace has emerged as a crucial theater of conflict and influence. As military, political, and economic systems become increasingly interconnected, cyber operations offer states non-kinetic means to project power, influence outcomes, and destabilize adversaries. These operations - ranging from espionage and disruption to full-scale strategic attacks - are now integral components of national defense strategies.

1.2 Purpose of the Paper

This paper proposes a mathematical modelling framework to quantify the strategic effectiveness of cyber operations in the pre-conflict phase of hybrid warfare. It seeks to bridge the gap between qualitative cyber strategy research and quantitative simulation by integrating probability theory, differential equations, and systems modelling into a cohesive analytical structure.

1.3 Paper Structure

Following the Introduction, Section 2 (Motivation) outlines the limitations of current approaches to cyber modelling. Section 3 (Framework) presents the conceptual basis linking cyber warfare to PMESII domains. Section 4 (Methodology) details the mathematical formulation. Section 5 (Example) shows a first prototype, and Section 6 (Conclusion) summarizes findings and future directions.

2. Motivation

2.1 Problem Statement

While cyber operations are widely analyzed from political, ethical, and strategic perspectives, few studies offer quantitative tools to model their systemic effects. Traditional wargames and simulations often lack dynamic, time-dependent mechanisms to capture cascading consequences across political, military, economic, social, information, and infrastructure domains.

2.2 Goal

The goal of this study is to create a scalable and data-driven model for simulating cyber conflicts, capable of quantifying both the probability of successful attacks and their cross-domain impacts using PMESII indicators. This enables scenario planners and analysts to assess cyber power as a measurable strategic factor.

3. Cyber Warfare Framework

3.1 Conceptual Foundations

Cyber power, as defined by Nye [1], combines hard and soft power in the digital domain. Libicki [2] emphasizes the signaling and shaping potential of cyber operations, while Rid [3] highlights their ambiguous nature between espionage, sabotage, and warfare. Building on this foundation, the proposed model aligns cyber dynamics with PMESII dimensions to represent systemic impacts across interdependent domains.

3.2 Integration with PMESII

The model extends the PMESII framework by mapping cyber effects across Political, Military, Economic, Social, Information, and Infrastructure systems. For example, an

attack on information infrastructure may indirectly degrade economic and social stability, captured mathematically through cross-domain coefficients and feedback loops [4].

4. Methodology

4.1 Model description

The model employs equations to represent time-dependent relationships between offensive (OCC) and defensive (DCC) cyber capabilities. Probabilistic functions estimate success rates of attacks, while resilience metrics adjust dynamically based on cumulative system degradation and recovery (Figure 1: System dependencies).

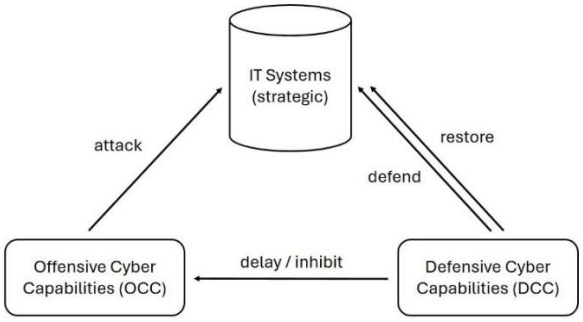


Figure 1: System dependencies

The model comprises three principal components. First, the *IT-system layer* (see Figure 2) does not denote a single service; rather, it represents the aggregate of all IT assets and services within an organisation. Each modelled system is assigned a category - critical, high, medium, or low (C/H/M/L)—as defined in consultation with subject-matter experts from the German Federal Office for Information Security (BSI) [5]. In addition, the level of protection of each system is assumed to improve over time, reflecting rising organisational awareness, enhanced personnel proficiency, and the maturation of software security features. For integration with the strategic simulator, every IT system is mapped to the PMESII schema, enabling cross-domain analysis of cyber effects

$$PMESII_s = \sum_i w_{is}$$

where w_{is} are mapping weights.

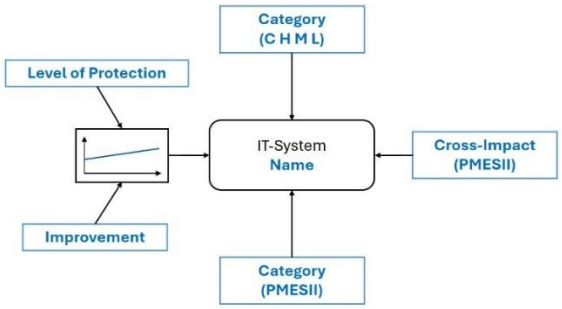


Figure 2: IT-System

Second, the Defensive cyber capabilities (DCC) (Figure 3). During a workshop with representatives from the BSI, the DCCs were described not merely as human resources, but as encompassing the collective skills, structures, and capabilities of the entire cyber defence organisation. According to BSI experts, these capabilities are typically organised in response teams, which can be formed on demand depending on the incident type and severity. Each team possesses specialised, problem-specific competencies. The exact number of deployable teams remains classified. In the model and subsequent simulation, two parameters are therefore considered: 1. the skill level of each tea and 2. the number of teams available for deployment.

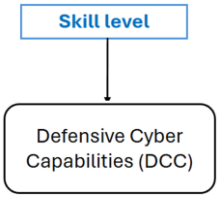


Figure 3: Defensive Cyber Capabilities (DCC)

Third, the Offensive cyber capabilities (OCC) (Figure 4). Similar to the defensive domain, the third and final component of the model represents the Offensive cyber capabilities. Based on discussions with experts at both national and international levels, as well as insights gained from professional conferences such as DEF CON and Black Hat, the concept of OCC is understood to encompass far more than mere personnel strength. It includes the skills, tools, and operational vectors that constitute the offensive cyber domain - notably the utilisation of advanced capabilities such as zero-day exploits and other specialised attack methods. Accordingly, the model incorporates two primary parameters: (1) the skill level of offensive teams and (2) the number of available teams, each with its own distinct expertise and attack vectors (the details of which are classified).

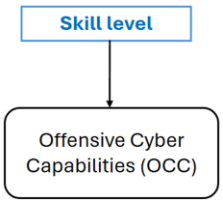


Figure 4: Offensive Cyber Capabilities (OCC)

4.2 Data Sources

Empirical data from exercises such as NATO CCDCOE's Locked Shields provide baseline parameters for cyber attack success rates and defensive efficiency. Simulation outputs are calibrated using open-source datasets and expert assessments to ensure both realism and adaptability to diverse conflict scenarios.

4.3 Mathematical Model Design

As described in the preceding sections, the overall model consists of three core components, each with its own specific parameters. In the following subsections, these components are treated separately and their behaviour during a cyber attack is translated into mathematical form.

4.3.1 Systems

According to the German Federal Office for Information Security, every IT system possesses a basic level of protection at any given time. This protection improves gradually due to continuous updates, patches, investments in security, and the increasing training level of administrators (Figure 5).

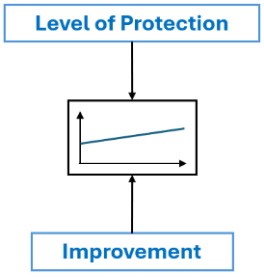


Figure 5: Protection Level of IT-Systems

Over the time horizon of the simulation, we therefore model the protection level of an individual system ias a linearly increasing function:

$$S_i(t) = m_i + c_i$$

where:

- $\boldsymbol{S_i(t)}$ denotes the effective protection level of system iat time t,
- $\boldsymbol{c_i}$ is the initial baseline protection at the start of the simulation,
- $\boldsymbol{m_i}$ is the rate at which protection improves over time (e.g. due to hardening measures and organisational learning).

Systems are categorised as critical, high, medium, or low (C/H/M/L) importance based on expert input, and each system is mapped to at least one PMESII dimension to enable cross domain analysis of cyber effects (a purely linear trend is a simplification and may later be replaced by saturating growth).

4.3.2 Defensive cyber capabilities (DCC)

The mathematical representation of the DCC is deliberately simple. Following BSI subject matter experts, the overall defensive performance of a response team can be approximated by a percentage value between 0% and 100%, reflecting skills, training, processes, and available tools. For a given defensive team j, we model its effectiveness as a constant:

$$D_j(t) = d_j, 0 \leq d_j \leq 0$$

where d_j captures the aggregated training level and organisational maturity. In a more advanced version, d_j could itself evolve over time as a function of experience and resource allocation, but for the present study it is assumed to be time invariant.

4.3.3 Offensive cyber capabilities (OCC)

Among the three components, the Offensive cyber capabilities (OCC) are the most complex to model. The corresponding function should represent the temporal progress of an attack from initial reconnaissance to a successful compromise. Combined with the protection level of the target system and the additional protection provided by the DCC, this function yields the overall probability of a successful intrusion.

Both personal experience and expert interviews with penetration testers in governmental and private settings indicate that this process is not linear in time. Instead, the success probability is initially low, then increases as

attackers discover viable vectors, and eventually saturates: if no compromise has occurred after a certain time, additional time does not substantially increase the chance of success.

To support this intuition with data, empirical evidence was obtained from the Locked Shields cyber defence exercise organised annually by the NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE). Locked Shields [6] is the world's largest and most complex live fire cyber defence exercise, involving multinational teams defending critical infrastructure systems under realistic attack.

For the 2025 iteration, approximately 8,000 systems were deployed and defended by about 4,000 blue force personnel, while a three digit number of red team attackers conducted operations. From the captured network traffic (around 2 TB of data), 9,775 distinct cyber attacks were identified. Using packet timestamps, the duration from attack start to successful compromise was determined. These attacks were grouped into time intervals of roughly 30 time units (Figure 6); for example, in the interval 8.5–37.9 three attacks were successful, while no attack succeeded after time 361.7.

8.492	3
37.9243917	60
67.3567833	489
96.789175	1311
126.221567	1584
155.653958	2085
185.08635	2066
214.518742	1348
243.951133	605
273.383525	181
302.815917	40
332.248308	2
361.6807	0

Figure 6: Table of successful attacks

When plotted, the distribution of successful attacks over time resembles a normal (Gaussian) distribution (Figure 7).

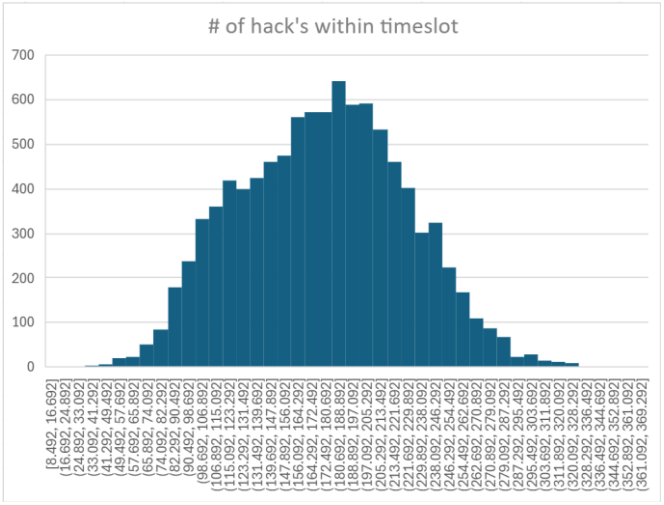


Figure 7: Distribution Over Time

An analysis of the empirical distribution yields a skewness of 0.057 (no significant left/right skew) and a kurtosis of –0.429 (slightly flatter than a perfect Gaussian) (Figure 8).

Metric	Value
Mean	175.1241089
Standard Deviation	49.87257723
Skewness	0.057023164
Kurtosis	-0.428520213

Figure 8: Parameters of the normal distribution

A Q–Q plot confirms the goodness of fit (Figure 9):

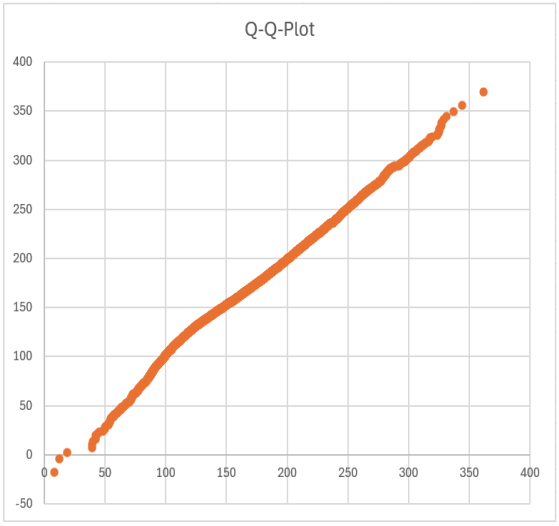


Figure 9: Q-Q-Plot

Based on these findings, the cumulative distribution function (CDF) of a normal distribution is a reasonable

model for the time dependent success probability of an offensive operation (Figure 10),

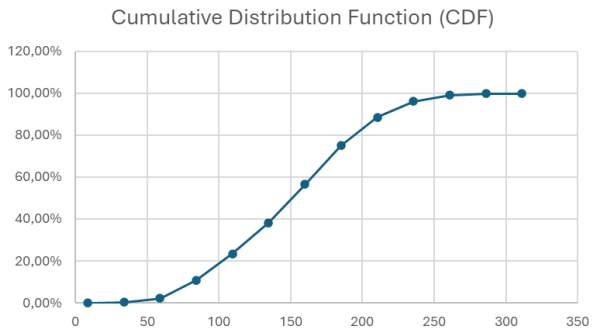


Figure 10: Cumulative Distribution Function (CDF)

$$\Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

where **μ** and **σ** are fitted to the Locked Shields data.

However, the exact CDF involves an integral over the normal density, which is computationally expensive in large scale Monte Carlo simulations. To reduce runtime, the CDF is approximated by a scaled sigmoid function:

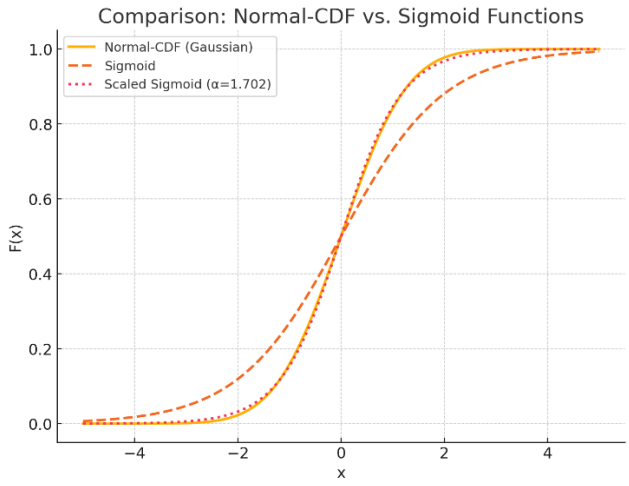


Figure 11: Normal Distribution vs. Sigmoid Function

Figure 11 compares the curves of a complex CDF (orange line) with those of a sigmoid function (red dash line) and a scaled sigmoid function (red dot line). It is evident that the CDF is approximately identical to the scaled sigmoid function (scale factor = 1.702). Therefore, the following function can be used instead of the complex CDF:

$$\Phi(x) \approx \sigma_{1.702}(x) = \frac{1}{1 + e^{-1.702x}}$$

With an appropriate scaling factor (here approximately 1.702), the maximum absolute error is around 1% in the central region, and typically between 0.5% and 1.0% across the relevant domain. Given the high speed up (approximately a factor of three in computation time), this approximation is acceptable for Monte Carlo simulation purposes.

4.3.4 Recovery Process

The final part of the model concerns the recovery of compromised systems. BSI experts describe recovery as a phased process: *"System recovery is carried out in phases (Figure 12). First, the core functions—representing roughly two-thirds of total system capability—are restored. Subsequently, the remaining, less critical functions are gradually brought back online until full system functionality is achieved"*.

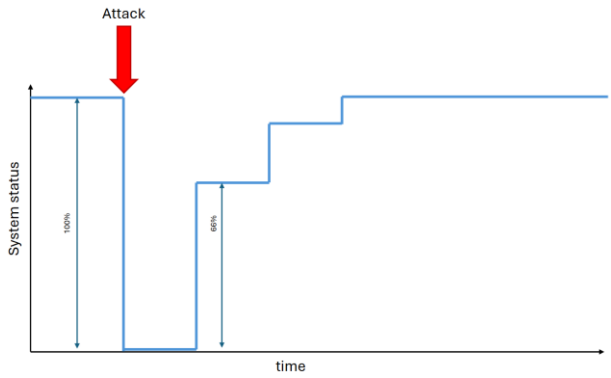


Figure 12: Recovery Process I

This behavior can be approximated by the charging curve of a capacitor in electrical engineering (Figure 13 green line):

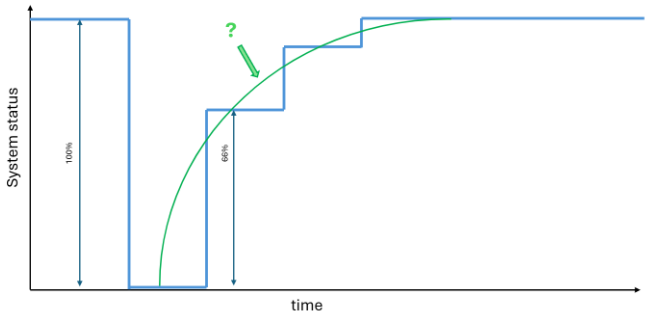


Figure 13: Recovery Process II

$$R(t) = R_{\max}(1 - e^{-t/\tau})$$

where:

- $\boldsymbol{R(t)}$ is the recovery level at time t,
- $\boldsymbol{R_{max}}$ is the fully restored capability (normalized to 1),
- $\boldsymbol{\tau}$ is the time constant capturing both DCC effectiveness and system complexity.

After approximately 5τ, the system is considered fully restored for practical purposes. While this continuous function smooths over the stepwise nature of real recovery, it captures the empirically observed pattern of rapid restoration of core functions followed by slower restoration of non-critical services.

5. Example and Prototype

Using the proposed model, it is possible to simulate, on an evidence based footing, the probability of successful cyber attacks against specific systems under defined conditions. A Monte Carlo simulator is particularly well suited for this purpose, as it allows many stochastic realisations of OCC, DCC, and system parameters to be explored. The aggregated results of this pre simulation are then passed to a higher level heuristic or strategic simulator as input parameters.

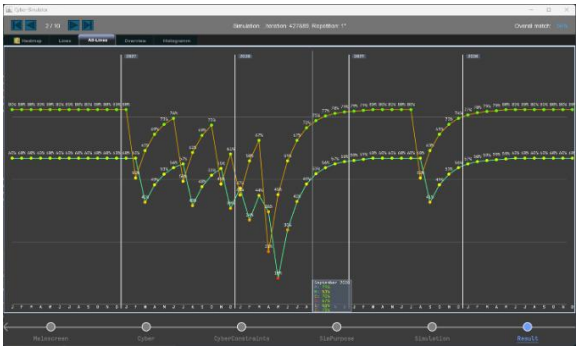


Figure 14: Prototype

Combined with additional elements such as a PMESII cross impact matrix and country specific characteristic, this enables the exploration of potential cross domain consequences of cyber campaigns (Figure 14). For example, the prototype demonstrates how cyber-attacks against infrastructure systems can, with a delay, degrade military capability and how long full restoration of those systems would take.

6. Conclusion

6.1 Summary

This study provides a quantitative foundation for cyber warfare modelling, addressing a key gap in the integration of cyber power within strategic analysis. By formalizing OCC and DCC interactions and aligning them with PMESII dimensions, the model offers a reproducible tool for pre-conflict simulation and policy evaluation.

6.2 Outlook

Future research should focus on enhancing empirical calibration, incorporating AI-driven adaptive models, and integrating human decision-making variables to capture the complex dynamics of cyber deterrence and escalation.

References

[1] J. S. Jr. Nye, "Cyber Power." Harvard Kennedy School, May 2010. Accessed: Nov. 29, 2025. [Online]. Available: https://www.belfercenter.org/sites/default/files/pantheon_files/files/publication/cyber-power.pdf

[2] M. L. M. C. L. is the Maryellen, R. L. K. D. V. P. in C. S. S. at the U. S. N. Academy, adjunct senior management scientist at the R. C. H. work involves the national security implications of information technology, N. as I. I. Cybersecurity, cyberwar H. lives in Kensington, and M. M. S. F. T. A. V. Biography, "Cyberspace in Peace and War, Second Edition," U.S. Naval Institute. Accessed: Nov. 29, 2025. [Online]. Available: https://www.usni.org/press/books/cyberspace-peace-and-war-second-edition

[3] T. Rid, Cyber war will not take place. Oxford New York: Oxford University Press, 2013.

[4] B. Weissenberger, "Seeding Success: Generating Valid and Realistic PMESII Start Values for Serious Wargames and Simulators," in NATO Modelling & Simulation Centre of Excellence, Rome, Sept. 2024. Accessed: Nov. 29, 2025. [Online]. Available: https://www.mscoe.org/document/seeding-success-generating-valid-and-realistic-pmesii-start-values-for-serious-wargames-and-simulators/

[5] "Bundesamt für Sicherheit in der Informationstechnik," Bundesamt für Sicherheit in der

Informationstechnik. Accessed: Nov. 29, 2025. [Online]. Available: https://www.bsi.bund.de/DE/Home/home_node.html

[6] "Locked Shields." Accessed: Nov. 29, 2025. [Online]. Available: https://ccdcoe.org/exercises/locked-shields/

CACTUS: Enhancing Strategic Decision-Making for Urban CBRNe and TIC/TIM Events through Simulation and AI

Marina Massei, Antonio Giovannetti, Marco Gotelli, Filippo Ghisi, Luca Cirillo, Xhulia Sina, Massimo Pedemonte

SIM4Future, Simulation Team

www.simulationteam.com

Abstract

The increasing complexity of urban and industrial environments and the growing threats associated with CBRNe (Chemical, Biological, Radiological, Nuclear, and Explosive) scenarios, particularly those involving Toxic Industrial Chemicals and Materials (TIC/TIM), require innovative solutions for training, mitigation, and decision support. This paper introduces CACTUS (City Analysis of Contamination by Tic/Tim Urban Area Simulation), an innovative simulation platform designed to dynamically model the dispersion, impact, and mitigation of multiple hazardous events. CACTUS integrates open geospatial data, decision-making support, stochastic threat modeling, and weather dynamics to simulate agent behaviors, population exposure, urban and industrial flow effects, and countermeasure deployment. The simulator supports urban mapping, building permeability, toxic persistence, domino effects, and evacuation scenarios. CACTUS is used also for the protection and risk assessment of critical infrastructures, including pipelines, industrial plants, and high-risk urban assets. Through the use of Strategic Engineering, CACTUS provides decision-makers with actionable insights to evaluate mitigation strategies, prioritize resources, and reduce population and infrastructure vulnerability during CBRNe and TIC/TIM events.

1. Introduction

CBRNe threats in urban, industrial areas and critical infrastructures represent a critical risk with potentially catastrophic consequences. From toxic industrial accidents to deliberate releases of chemical or radiological agents, these events challenge traditional response mechanisms due to their complexity, rapid evolution, and the urban environments' geometrical constraints. In particular, TIC and TIM incidents have become more probable given the high concentration of hazardous materials in industrial zones adjacent to populated areas.

In this context, decision-makers face growing difficulty in understanding the scale of an incident, predicting its evolution, and choosing the most effective countermeasures under uncertainty. The complexity of such scenario requires simulation platforms capable of integrating environmental, infrastructural, behavioral, and operational data in a coherent, flexible, and dynamic framework.

To address these challenges, we introduce CACTUS, a CBRNe simulator developed by Simulation Team, which reproduces the dynamics of multi-agent and multi-source hazardous events in complex urban and industrial scenarios, including critical infrastructures such as pipelines and industrial plants. Unlike traditional simulators focused on physical dispersion only, CACTUS integrates human behavior, evacuation flows, building permeability, meteorological effects, and countermeasure evaluation.

The simulator provides several functionalities, including:

- modeling of chemical, biological, radiological, nuclear, and explosive threats.
- simulation of TIC/TIM releases via both accidental and malicious events.
- integration of open-source geospatial data to recreate realistic urban maps.
- support for 3D terrain analysis and detailed meteorological boundary conditions.
- implementation of Intelligent Agents to simulate population behavior and emergency responders.
- strategic/tactical dashboards for decision-making support in real time.

This paper presents CACTUS's architecture, describes its core modules and simulation capabilities, and outlines several applications including urban evacuation planning, hazard mapping, and crisis management training.

2. State of the Art

The modeling and simulation of CBRNe (Chemical, Biological, Radiological, Nuclear, and Explosive) scenarios in urban environments has evolved significantly over recent decades. Initial models predominantly focused on the physical aspects of contaminant dispersion, relying on simplified geometries and static boundary conditions. However, as urban and industrial infrastructures have become more complex and threats more multifaceted, especially with the proliferation of Toxic Industrial Chemicals (TIC) and Materials (TIM), there is a growing need for integrated, dynamic simulation frameworks that incorporate both environmental and human factors. Traditional dispersion models such as Gaussian plumes or CFD (Computational Fluid Dynamics) approaches offer insight into the physical behavior of hazardous substances. Yet, these models often assume idealized terrains or steady-state conditions, making them poorly suited for the heterogeneous and dynamic nature of modern cities. To improve realism, advanced approaches consider complex geometries derived from Digital Elevation Models (DEM) and 3D urban data, as highlighted by Fellini et al. (2021) and Soulhac et al. (2017). These models address how urban canyons, building permeability, and microclimatic effects influence the flow and stagnation of toxic agents.

The catastrophic events occurred in Seveso and Bhopal remain historic examples underscoring the need for predictive modeling and regulation (Fabiano et al., 2017; Yang et al., 2015) and modern simulators must be capable of reproducing these scenarios where cascading failures, such as simultaneous gas release and fire, can occur. TIC and TIM events, whether accidental or deliberate, introduce additional complexity due to the diversity of substances involved, their interactions with infrastructure, and the potential for domino effects. This challenge necessitates both temporal and spatial resolution in simulations, as well as the ability to assess multiple release modes and vectors.

Beyond physical modeling, one of the most crucial aspects of CBRNe simulation is the incorporation of population dynamics. The presence, density, and behavior of individuals under threat conditions significantly impact evacuation efficiency, casualty estimation, and resource allocation. Approaches based on Intelligent Agents (IA), as developed in previous Simulation Team models (e.g., PONTUS, IDRASS), enable the representation of heterogeneous behaviors including panic, compliance, and communication effects (Bruzzone et al., 2022; Massei & Tremori, 2014). These models are essential for reproducing emergent phenomena such as bottlenecks or spontaneous regrouping, which affect both threat evolution and response effectiveness.

Strategic Engineering, as proposed by Bruzzone et al. (2021), blends M&S (Modeling & Simulation), data analytics, and IA to support real-time and offline decision-making. These frameworks are particularly suited to urban crises, where information overload, time constraints, and conflicting objectives dominate. Recent advancements have been focused on integrating

Artificial Intelligence (AI) to support scenario generation, outcome prediction, and decision-making under uncertainty, while AI is also used to dynamically simulate Courses of Action (COAs), assess mitigation strategies, and prioritize interventions based on stochastic impact evaluations (Wilner & Babb, 2021; Regal et al., 2022). An additional evolution in the state of the art is the emphasis on system interoperability. Tools such as JISR (Joint Intelligence, Surveillance, and Reconnaissance) networks, GIS systems, and weather forecasting engines are being integrated into simulation platforms to support live-data-driven simulation. The ability to fuse sensor data — from UAVs, fixed stations, and satellites — enables CACTUS-like systems to operate in "fast time" with real-time updates, ensuring relevance throughout the crisis life cycle. The use of frameworks like HLA (High Level Architecture) further supports integration with legacy and emerging command-and-control (C2) systems.

Despite the advances described, most existing commercial and academic platforms focus on singular aspects — for instance, wind dispersion or evacuation planning — and lack the holistic, interoperable architecture needed to address full-spectrum CBRNe crises in real cities. Additionally, many systems are not scalable, suffer from limited accessibility to source code for modification, and fail to simulate compound threats or cross-domain effects (cyber, psychological, kinetic). This motivates the development of integrated systems like CACTUS, designed to support complex, dynamic, and realistic multi-agent crisis scenarios.

3. General Architecture of the System

CACTUS solution is designed to simulate the dispersion of hazardous substances in urban and industrial areas by integrating geospatial data, environmental physics, building permeability, and human exposure dynamics. Its modular architecture ensures that the simulator can flexibly adapt to various urban contexts and threat scenarios. CACTUS operates by ingesting real-world data, configuring the scenario according to a user-defined configuration file, and simulating the event through time-stepped physics and behavioral models. The simulator retrieves data from open sources such as OpenStreetMap, extracting the street network, building footprints, and land use categories. In parallel, it incorporates Digital Elevation Models (DEM) or Digital Terrain Models (DTM) to capture the local orography, elevation gradients, and slope characteristics generating a 3D representation of the urban environment from the real-world geographic and cartographic data.

The resulting environment is then discretized into a structured mesh composed of volumetric elements (voxels), which distinguishes between terrain surfaces and building volumes. Each voxel is associated with physical properties including elevation, ground type, building permeability, and occupancy attributes. This spatial discretization enables realistic simulation of physical processes

such as gas flow, surface interaction, and infiltration dynamics. CACTUS supports full scenario configurability through a dedicated configuration file read at simulation startup. This file defines the structural and dynamic attributes of the simulation, including the demographic and behavioral distribution of the population. Occupancy levels in buildings are estimated based on the typology of each structure, residential, commercial, educational, or industrial, and are modulated by temporal parameters such as hour of the day or day of the week.

Environmental conditions are also configured at this stage. Users select either static or time-varying values for wind speed and direction, temperature, and humidity. These meteorological conditions are subsequently used to drive the dispersion dynamics. The user also specifies the hazardous agent to be simulated, selecting from a predefined library of substances with associated toxicological, physical, and chemical properties. The release event is characterized by its spatial location, its modality (e.g., explosive, slow leak, or rapid release), and its initial mass or concentration. Once the 3D model and scenario configuration are complete, the simulator performs an analysis of terrain gradients to determine preferential flow directions for the dispersion of gases or aerosols. By calculating slope vectors and coupling them with user-defined wind fields, CACTUS constructs a dynamic advection field that governs the movement of contaminants.

This stage is critical for capturing the influence of topography and urban morphology on dispersion behavior. Wind-shadow zones, channeling effects along streets, and orographic barriers are all inherently represented in the gradient and flow field computation. These spatial dynamics directly impact the speed and direction of toxic plume propagation.

To ensure integration with external simulators, sensor systems, and command platforms, CACTUS adopts the IEEE 1516 High Level Architecture (HLA) standard.

The HLA-based federation enables CACTUS to:

- Operate as part of a larger simulation federation (e.g., NATO CAX, civil-military training environments).
- Share time-synchronized data on physical variables (e.g., contamination levels, casualty counts), agent states, and environmental dynamics.
- Interoperate with legacy defense systems, city digital twins, emergency response simulations, or incident command systems.
- Synchronize federated simulations with Run-Time Infrastructure (RTI) to support both real-time and fast-time execution modes.

This modularity ensures CACTUS can evolve with new modules or integrate domain-specific simulators such as HPAC or custom CFD engines.

CACTUS models the spread of hazardous agents (chemical, biological, radiological, etc.) in a three-dimensional urban grid. Each cell in the mesh represents a volumetric element with terrain, building, or atmospheric properties.

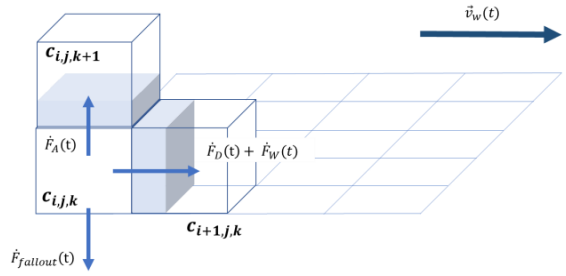


Figure 1: Voxel representation

The Dispersion and Interaction Engine simulates the time evolution of the hazardous agent within the 3D environment using a time-stepped mass-balance approach. The substance mass in each voxel evolves according to a system of differential equations that capture multiple interacting physical phenomena:

$$dMi/dt = \nabla \cdot (uMi) + D\nabla 2Mi + B(Mi) - R(Mi)$$

Where:

Mi is the mass of the agent in voxel i,

$\vec{u}$ is the local advection vector (derived from wind and terrain slope),

D is the diffusion coefficient,

B represents buoyancy-driven vertical transport,

R captures reaction and removal processes, such as chemical decay, deposition, or precipitation.

The simulation also accounts for the persistence of the agent on surfaces and its infiltration into buildings, which is modeled based on material permeability and structural characteristics. Each time-step updates the local concentrations in the air and computes the fraction absorbed by terrain or internalized into structures.

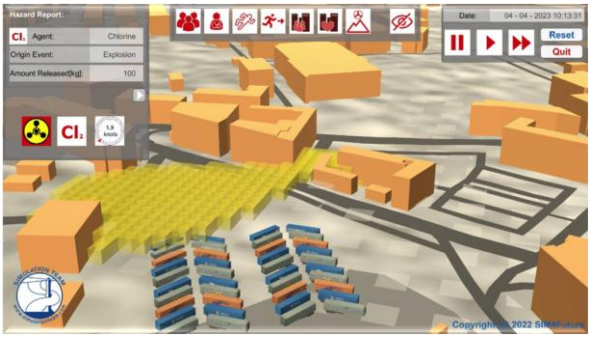


Figure 2: CACTUS Simulation Interface

CACTUS supports multiple threat sources:

- Point releases (e.g., IED explosion, pipe rupture)
- Area sources (e.g., storage tank leak)
- The impact of the hazardous agent on the population is evaluated using an agent-based human behavior model. Individuals are instantiated within buildings and outdoor areas based on the initial configuration and are assigned attributes such as age, health status, mobility, and risk perception.
- As the concentration of the toxic substance reaches a threshold in a given voxel, the corresponding agents react according to a Perceive-Decide-Act loop. The agents assess the perceived risk in their surroundings, make decisions about evacuation, sheltering, or seeking help, and act by moving through the street network using shortest-path algorithms affected by crowd dynamics and environmental conditions.
- Health outcomes are computed based on dose-response curves, taking into account the duration of exposure and substance-specific toxicity levels. The simulator tracks metrics such as number of exposed individuals, severity of intoxication, fatalities, and successful evacuations.

CACTUS provides both 2D and 3D visualization tools to support situational awareness and decision-making. Users can observe the evolution of the simulation in real-time or fast-time mode, visualizing:

- Contaminant concentration maps.
- Infiltration heatmaps.
- Population density and dynamic number of casualties and hospitalizations.
- The simulator can be used in fast-time mode for scenario exploration or in real-time mode for training and support in live operations. Outputs are also available in tabular format for downstream analysis and reporting.

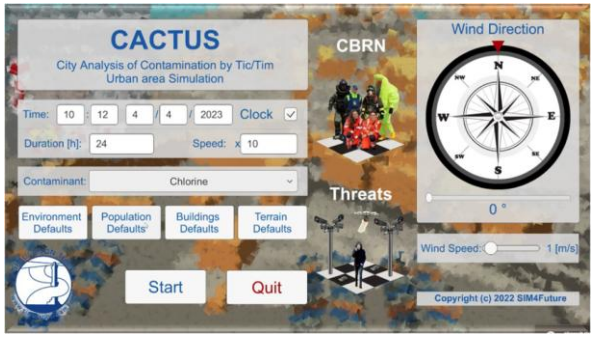


Figure 2: CACTUS Configuration UI

4. HLA Integration

The integration of CACTUS within an HLA federation employs the AMSP-04/NETN-CBRN Federation Object Model (FOM) as a framework to ensure consistent and reliable data exchange. CACTUS, with its ability to simulate the release and atmospheric dispersion of hazardous gases, naturally aligns with the core intent of the FOM: representing contamination and its dynamic evolution within a shared synthetic environment. Within this integration effort, CACTUS's foundational principle of discretizing the simulated atmosphere into cuboid volumes offers a practical method for representing contamination as it evolves in space and time. Each cuboid could be thought of as a microcosm of the real-world environment, containing detailed information about gas concentration, contaminant type, and the potential for exposure to occupants in adjacent structures. As gas releases unfold, whether as continuous emissions or sudden, catastrophic discharges, CACTUS continuously calculates the concentration of contaminants in each cuboid, adjusts the volume of the affected airspace, and considers environmental factors such as building shielding and infiltration. The AMSP-04/NETN-CBRN FOM provides a structured vocabulary to publish these detailed calculations as standardized objects. In this context, CACTUS's calculated contamination volumes are mapped

directly to the Contaminated Zone object class of the FOM, allowing other simulation components to comprehend and respond to these evolving hazards. Attributes such as contaminant type, concentration, and the geometry of the affected zone are published as they emerge and change in the simulation, reflecting the model's precise grasp of gas behavior and environmental interactions. As the simulation progresses, CACTUS's updates are fed directly into the federation, capturing every significant alteration in the contaminated landscape, from the expansion of a toxic cloud to its eventual dilution or containment. Notably, CACTUS's attention to buildings and their occupants is crucial in achieving realistic representations of exposure and risk. The model calculates the mass of contaminants that infiltrate indoor spaces and adjusts the projected number of affected individuals based on estimated building occupancy and air exchange dynamics. This nuanced assessment of indoor exposure is essential in modern hazard modeling, as real-world consequences hinge not merely on outdoor contamination but on the subtler, often deadlier infiltration of gases into occupied spaces. By taking advantage of the object structure of the FOM, these detailed aspects of CACTUS's outputs, such as changing concentrations in occupied spaces, or the dynamic shapes of contamination clouds, are made available to other federated systems. Command and control tools, evacuation planners, and medical response models could draw directly from this data, ensuring a shared, consistent understanding of the evolving threat.

5. Conclusions

This paper introduces CACTUS (City Analysis of Contamination by Tic/Tim Urban Area Simulation), a novel simulation platform designed to address the growing need for accurate modeling and simulation to support decision-making in the face of urban and industrial CBRNe and TIC/TIM crises. By integrating open geospatial data, high-resolution terrain and building models and agent-based human behavior, CACTUS enables a dynamic representation of complex threat scenarios. The system's architecture supports dispersion of hazardous substances through detailed gradient-based flow modeling, considering wind, terrain, building permeability, and environmental conditions. CACTUS also simulates population exposure, evacuation dynamics, and health outcomes based on time-dependent contamination levels. Moreover, its modular and interoperable design, compliant with the IEEE 1516 High Level Architecture (HLA), ensures compatibility with broader simulation ecosystems and operational command frameworks. In conclusion, CACTUS ability to model and assess the vulnerability of critical infrastructures, including industrial plants, and urban utility networks, providing a strategic layer of analysis to support protection and resilience planning.

References

[1] Bruzzone, A. G., Gotelli, M., Giovannetti, A., De Paoli, A., Ferrari, R., Pedemonte, M., Martella, A., Reverberi, A., Faccio, F., Cirillo, L., Ghisi, F.,Bucchianica L. & Frosolini, M. (2023). Strategic Engineering for Decision Making during Urban Crises, DHSS, I3M Athens 2023

[2] Bruzzone, A. G., Vairo, T., Cepolina, E. M., Massei, M., De Paoli, A., Ferrari, R., Giovannetti A. & Pedemonte, M. (2022, October). Cooperative use of autonomous systems to monitor toxic industrial materials and face

accidents & contamination crises. In International Conference on Modelling and Simulation for Autonomous Systems (pp. 231-242). Cham: Springer International Publishing

[3] Bruzzone, A. G., Massei, M., Sinelshchikov, K., Giovannetti, A., & Gadupuri, B. K. (2021, July). Strategic engineering applied to complex systems within marine environment. In 2021 Annual Modeling and Simulation Conference (ANNSIM) (pp. 1-10). IEEE

[4] Fabiano, B., Vianello, C., Reverberi, A. P., Lunghi, E., & Maschio, G. (2017). A perspective on Seveso accident based on cause-consequences analysis by three different methods. Journal of Loss Prevention in the Process Industries, 49, 18-35

[5] Fellini, S., Salizzoni, P., & Ridolfi, L. (2021). Vulnerability of cities to toxic airborne releases is written in their topology. Scientific Reports, 11(1), 23029

[6] Massei, M., & Tremori, A. (2014). Simulation of an urban environment by using intelligent agents within asymmetric scenarios for assessing alternative command and control network-centric maturity models. The Journal of Defense Modeling and Simulation, 11(2), 137-153

[7] Regal, G., Murtinger, M., & Schrom-Feiertag, H. (2022, May). Augmented CBRNE responder-directions for future research. In 13th Augmented Human International Conference (pp. 1-4)

[8] Soulhac, L., Nguyen, C. V., Volta, P., & Salizzoni, P. (2017). The model SIRANE for atmospheric urban pollutant dispersion. PART III: Validation against NO2 yearly concentration measurements in a large urban agglomeration. Atmospheric environment, 167, 377-388

[9] Wilner, A., & Babb, C. (2021). New technologies and deterrence: Artificial intelligence and adversarial behaviour. NL ARMS Netherlands Annual Review of Military Studies 2020: Deterrence in the 21st Century—Insights from Theory and Practice, 401-417

[10] Yang, M., Khan, F., & Amyotte, P. (2015). Operational risk assessment: A case of the Bhopal disaster. Process Safety and Environmental Protection, 97, 70-79

Integrating Cognitive Warfare into Multi-Domain Wargaming: The CW-BRAINWARE Approach based on Strategic Engineering

Marina Massei, Antonio Giovannetti, Marco Gotelli, Filippo Ghisi, Luca Cirillo, Xhulia Sina

SIM4Future, Simulation Team

www.simulationteam.com

Abstract

This paper presents CW-BRAINWARE, an innovative AI-powered wargame that integrates Modeling & Simulation (M&S) with Large Language Models (LLMs) to enable complex scenario analysis, decision support, and strategic experimentation across physical, virtual, and cognitive dimensions. Developed by the Simulation Team, CW-BRAINWARE introduces the Strategic Engineering approach for audacious wargaming (AuW) capability development, that combines a multi-domain chessboard, stochastic simulation, and cognitive behavior modeling to evaluate the cascading impacts of kinetic, cyber, and cognitive attacks on critical infrastructures and populations. The platform enables players to plan and execute operations across seabad, sea, land, air, space, cyberspace, and cognitive domains, generating and countering influence campaigns, disinformation, and public sentiment manipulation. LLMs are used for the generation of context-sensitive messages, sentiment analysis, and strategic counter-narrative development, providing real-time cognitive response planning capabilities. The integration of agent-based models, demographic stratification, and AI-generated feedback empowers commanders to assess operational outcomes while considering emotional resilience, morale, and social cohesion. This paper explores CW-BRAINWARE's architecture, simulation methodology, and a representative scenario based on a destabilization campaign against strategic energy and communication infrastructures.

1. Introduction

Modern warfare has entered an era of unprecedented complexity, where strategic decision-making must account for operations that simultaneously span across the physical, cyber, and increasingly, the cognitive domains. The rise of multi-domain operations (MDO)—encompassing seabad, sea, land, air, sea, space, cyber, and information spaces—demands new simulation-based tools capable of capturing the interdependencies between these layers of conflict and influence. In particular, the cognitive domain—targeting human perception, morale, decision-making, and public sentiment—has become a decisive factor in shaping both the course and outcome of conflicts. Within NATO and allied defense ecosystems, there is an urgent need to provide strategic foresight, training, and experimentation platforms that not only simulate kinetic and cyber engagements, but also model the emotional, social, and informational dynamics of the operational environment. Traditional Computer-Assisted Exercises (CAX) and simulations, while effective in rehearsing physical scenarios, often fall short when attempting to integrate disinformation, propaganda, psychological operations, and their effects on populations and decision-makers.

To address this gap, we present CW-BRAINWARE—a strategic wargame and simulation environment that uses Artificial Intelligence (AI), Large Language Models (LLMs), and agent-based simulation to model the impact of multi-domain operations on critical infrastructures and the human cognitive landscape. CW-BRAINWARE is an audacious wargaming (AuW) solution designed to immerse decision-makers in dynamic, multidimensional scenarios that include kinetic attacks, cyber intrusions, and cognitive operations via Broadcast, Print, and Social Media (BPSM). Through the use of intelligent agents, demographic modeling, and AI-driven messaging systems, CW-BRAINWARE supports experimentation and operational planning by simulating the behavior of civilian populations, the resilience of infrastructure, and the cognitive effects of disinformation campaigns. LLMs are employed both to generate and counter influence operations, enabling the simulation of narrative warfare, emotional disruption, and counter-propaganda strategies. This paper provides an overview of CW-BRAINWARE's architecture, its role in supporting NATO-aligned initiatives, and its application in a near-future operational scenario involving the protection of energy and communication infrastructures in a contested region. By bridging simulation, AI, and cognitive warfare, CW-BRAINWARE represents a new frontier in strategic wargaming, offering a critical tool for

enhancing situational awareness, resilience, and decision superiority in multi-domain environments.

2. State of the Art

The modern battlefield has evolved into a highly interconnected and information-rich environment, where multi-domain operations (MDO) and cognitive warfare are reshaping doctrines, capabilities, and strategic planning. In response, the scientific and defense communities have increasingly explored the intersection of AI-enabled wargaming, simulation-based experimentation, and strategic decision support.

MDO requires synchronized operations across land, air, sea, cyber, and space, with increasing emphasis on cross-domain synergies and time-sensitive decision-making. Traditional wargames have primarily focused on the physical dimensions, relying on deterministic models and pre-scripted scenarios. However, these approaches struggle to reflect the adaptive, nonlinear, and hybrid nature of modern conflicts. NATO publications emphasize the necessity of integrated M&S environments that can address cross-domain threats and support joint force experimentation and training (NATO STO, 2021).

The emergence of computer-assisted wargames for strategic planning has been widely documented (Cayirci et al., 2022). Systems such as the CAPIAS Wargame have laid the groundwork for agent-based simulations that incorporate cognitive variables, but often with limited support for dynamic narrative warfare or LLM-driven message generation (Bruzzone et al., 2023).

Cognitive warfare (CW) targets human perception, behavior, and decision-making, leveraging information manipulation, psychological pressure, and emotional disruption (Claverie & Du Cluzel, 2022). In this context, modeling public sentiment, morale, belief systems, and susceptibility to disinformation becomes central.

Early efforts in Human Behavior Modeling (HBM) and agent-based simulation focused on stress responses, cultural variables, and social behaviors (Bruzzone et al., 2014; Bruzzone et al., 2011). These were later extended with intelligent agents in training environments for asymmetric warfare and civil-military operations. Recent simulation systems have explored the use of factional variables, moral indices, and belief systems to simulate perception shifts and community reactions to hybrid operations (Mazal & Bruzzone, 2019).

The integration of Artificial Intelligence (AI) and Large Language Models (LLMs) into wargaming represents a transformative leap, offering sophisticated simulation of military scenarios and augmenting strategic decision-making with autonomous technologies. LLMs can simulate human-like decision-making processes, providing valuable perspectives in crisis escalation scenarios (Chen, 2024, Lampath et al., 2024)

The advent of Large Language Models (LLMs) has opened new frontiers in simulation-based decision-making. LLMs, such as GPT and LLaMA, are capable of generating strategic

narratives, simulating adversary behavior, and modulating cognitive attacks through context-sensitive messaging. Their zero-shot and few-shot capabilities allow real-time adaptation without domain-specific retraining, making them suitable for unpredictable and evolving scenarios (Hogan & Brenner, 2024; Weller et al., 2024).

Simulation frameworks that incorporate Retrieval-Augmented Generation (RAG), sentiment analysis, and emotional scoring can simulate and counter disinformation with unprecedented fidelity. These capabilities are not yet fully integrated in NATO-standard CA2X2 tools, but platforms like CW-BRAINWARE mark a step forward by combining AI-driven reasoning with interactive wargaming, enabling strategic experimentation in contested narrative environments.

3. CW-BRAINWARE Architecture

CW-BRAINWARE is an innovative simulation-based strategic wargame that integrates artificial intelligence, Human Behavior Modeling, and multi-domain dynamics within a unified architecture. It is designed to support experimentation, training, and strategic planning in complex operational environments where physical, virtual, and psychological dimensions intersect. Its architecture reflects the increasing demand for simulation tools capable of incorporating not only kinetic and cyber actions, but also the far-reaching implications of cognitive warfare on populations, decision-makers, and critical infrastructures. At the core of CW-BRAINWARE there is a stochastic discrete-event simulation engine that governs the evolution of operational scenarios involving both

traditional and hybrid threats. This engine simulates the behavior of physical systems—such as military units, critical infrastructure components, and urban populations—across seabad, sea, land, air, space, and cyber domains. Interdependencies among these domains are modeled in detail, including the cascading effects of failures in interconnected systems like energy grids, telecommunications networks, and supply chains.

A distinctive feature of CW-BRAINWARE is its integration of cognitive warfare into the operational landscape. The system models the intentional manipulation of information through Broadcasting, Print, and Social Media (BPSM), enabling the simulation of disinformation campaigns, psychological pressure, and influence operations. These cognitive actions impact not only public sentiment but also troop morale, trust in authorities, and the coherence of strategic narratives. The effects of such operations are calculated using agent-based models informed by demographic, psychographic, and emotional variables, such as belief systems, resilience levels, and moral integrity.

Artificial intelligence is embedded within the architecture through the incorporation of Large Language Models (LLMs), which serve multiple functions within the decision cycle. These models generate cognitively tailored messages, forecast emotional and behavioral responses, and simulate adversarial narrative strategies. LLMs also assist in counter-messaging by analyzing the intent and target of incoming cognitive attacks and proposing context-specific reactions. Their capabilities are further enhanced through the use of retrieval-augmented generation (RAG), allowing CW-BRAINWARE to access dynamic knowledge bases and adapt messaging to real-time developments in the simulated environment. The representation of human actors is achieved through a detailed behavioral engine that models individuals and population clusters as agents governed by demographic and cognitive parameters. Agents are differentiated by characteristics such as age, education, health status, religious affiliation, political orientation, and social group membership. These variables inform each agent's susceptibility to cognitive actions and contribute to emergent behaviors at the societal level. Emotional states, morale, trust, and motivation evolve dynamically in response to simulated events and external stimuli, including kinetic strikes, infrastructure disruptions, or information flows.

Interaction with the simulation is facilitated through an advanced user interface, which includes a multidomain chessboard and a strategic dashboard. Players interact with the scenario by assigning high-level tasks, initiating operations, deploying cognitive or cyber assets, and managing responses to adversarial actions. The system visualizes the evolution of sentiment, infrastructure status, unit effectiveness, and decision impact in real time, allowing users to iteratively adapt their strategy.

Figure 1: Land and Air Layer of the Multi-domain Chessboard

Figure 2: On the left one of the Monitoring Panel for Critical Infrastructures KPIs. On the right the 2-D control panel

Figure 3: Message Manager and Control Board for Cognitive and Messaging management

CW-BRAINWARE simulates kinetic and cyber operations by extending classical attrition-based models to incorporate dynamic variables that reflect modern hybrid warfare. These models not only account for traditional combat factors such as unit strength, equipment capabilities, and operational tempo, but also integrate cyber vulnerabilities, psychological stressors, and infrastructure interdependencies. The outcomes of kinetic engagements are influenced by factors such as troop morale, leadership coherence, and battlefield information asymmetry, while cyber-attacks consider the structure and resilience of digital systems, the sophistication of intrusion vectors, and the

preparedness of defensive protocols. Infrastructure degradation is not treated as a binary event but rather as a gradual process, evaluated through a composite metric that considers the operational condition of each component subsystem and its functional relevance within broader interlinked networks. The representation of human actors within CW-BRAINWARE is grounded in cognitive modeling, where both population clusters and key decision-makers are treated as agents with evolving internal states. These agents are characterized by attributes such as morale, emotional resilience, belief stability, and susceptibility to external influence. Their behavior adapts over time in response to environmental stimuli—ranging from kinetic attacks and service outages to cognitive operations and public messaging campaigns. The agents operate within simulated social networks where proximity, communication, and group identity play central roles in shaping individual and collective responses to ongoing events. LLMs components are employed to generate and disseminate strategic messaging intended to either degrade or strengthen cognitive coherence within targeted audiences. Offensive operations may include the simulation of disinformation, panic induction, or rumor propagation aimed at destabilizing population trust or weakening command structures. Defensive cognitive strategies, in contrast, involve the crafting of reassurance messages, the reinforcement of institutional legitimacy, and the generation of counter-narratives capable of neutralizing adversarial influence campaigns.

The generation of such messages follows a structured process that mirrors the operational planning of psychological operations in real-world settings. Initially, the system profiles the intended targets based on demographic, socio-political, and emotional characteristics. This enables the tailoring of communication strategies to exploit cognitive biases, cultural contexts, and media preferences. Once the target profile is established, the platform constructs prompts that encode the strategic intent of the operations such as inciting fear, undermining trust, or promoting resilience. Large language models, such as LLaMA 3.1, are then employed to generate linguistically coherent and contextually relevant content using either zero-shot capabilities or through retrieval-augmented generation techniques.

The emotional tone and likely psychological impact of the generated message are evaluated using natural language processing tools, which estimate sentiment polarity, emotional charge, and intensity across multiple dimensions. This analysis supports the selection or refinement of messages based on their predicted effectiveness and risk of unintended consequences. Finally, the propagation of the message is simulated using the Simulation Team's Information Propagation Model, which accounts for network topology, agent connectivity, and dynamic sentiment transfer mechanisms. This model enables the evaluation of reach, influence, and feedback loops, allowing decision-makers to explore the consequences of cognitive actions across time and space.

4. Conclusions

The increasing complexity of modern conflicts, driven by the convergence of kinetic, cyber, and cognitive threats, require new paradigms in wargaming and decision support. CW-BRAINWARE represents an innovative response to this challenge—providing a strategic simulation environment that integrates artificial intelligence, large language models, and behavioral modeling to support multi-domain operations in contested, ambiguous, and information-saturated environments.

Through its integrated architecture, CW-BRAINWARE enables decision-makers, planners, and analysts to explore the interplay between physical operations and cognitive effects, simulating not only attacks on critical infrastructures but also the cascading influence of disinformation, psychological operations, and social destabilization. The platform's AI-driven capabilities allow for the generation and analysis of realistic cognitive operations, enhancing situational awareness and enabling proactive responses to adversarial influence campaigns. The incorporation of agent-based population dynamics and emotional propagation models further allows the system to assess the second- and third-order effects of strategic actions on civilian morale, trust, and behavior.

References

[1] Bruzzone, A. G., Massei, M., Gotelli, M., Giovannetti, A., & Martella, A. (2023). Sustainability, Environmental Impacts and Resilience of Strategic Infrastructures. In Proceedings of the International Workshop on Simulation for Energy, Sustainable Development and Environment, SESDE., Rome 2023

[2] Bruzzone, A., Massei, M., Longo, F., Poggi, S., Agresta, M., Bartolucci, C., & Nicoletti, L. (2014, April). Human behavior simulation for complex scenarios based on intelligent agents. In Proceedings of the 2014 Annual Simulation Symposium (pp. 1-10)

[3] Bruzzone, A. G., Tremori, A., Tarone, F., & Madeo, F. (2011). Intelligent agents driving computers generated forces for simulating human behaviour in urban riots. International Journal of Simulation and Process Modelling, 6(4), 308-316

[4] Cayirci, E., AlNaimi, R., AlNabet, S. (2022). Computer assisted military experimentations. 2022 Winter Simulation Conference (WSC), IEEE

[5] Chen, Y., & Chu, S. (2024). Large language models in wargaming: Methodology application and robustness. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2894-2903)

[6] Hogan, D. P., & Brennen, A. (2024). Open-ended wargames with large language models. arXiv preprint arXiv:2404.11446

[7] Lamparth, M., Corso, A., Ganz, J., Mastro, O. S., Schneider, J., & Trinkunas, H. (2024, October). Human vs. machine: Behavioral differences between expert humans and language models in wargame simulations. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (Vol. 7, No. 1, pp. 807-817)

[8] Mazal, J., & Bruzzone, A. G. (2019). NATO needs of Future Strategic Engineers. In Workshop on Applied Modelling & Simulation (Vol. 35)

[9] NATO Allied Command Transformation. (2024). Multi-Domain Operations and Digital Transformation

[10] NATO Allied Command Transformation. (2024). Cognitive Warfare.

[11] Weller, D., Meltschack, M., & Schwindling, D. Leveraging Large Language Models for Enhanced Wargaming in Multi-Domain Operations. 2024, NAT

Intelligent Agents in Wargaming Simulations

Giacomo Del Rio, Jonas Fontana, Matthias Sommer

Armasuisse, IDSIA, Supsi

www.armasuisse.ch
www.ar.admin.ch
www.supsi.ch

Abstract

Every entity participates in a wargaming simulation is required to have its behavior defined. This typically happens either in real-time by operators controlling the entities or by using condition-based scripts. Additionally, with emerging capabilities of AI, it has become feasible to add autonomous intelligent agents to the simulation. In doing so, not only is the variance in the simulation increased, but also the human effort required is reduced and thus large-scale simulations enabled. In this work, we report on a research project with exactly that goal. Algorithms from Reinforcement Learning receive the operational picture of the simulation and send orders through a semantic standard to enable doctrinal behaviour and interoperability. By proceeding in that way, the AI becomes independent of the simulation software used and can be easily connected to any other simulation engine supporting the standard. Potential applications are training, decision support and concept development.

1. Overview

The central concept of this paper is the development of "intelligent agents" that can control entities (Individual systems, aggregates, or even whole sides/coalitions) in wargaming simulations. The motivation for this is threefold:

- increasing the variance since AI-based agents explore tactics unfamiliar to human players.
- supporting wargaming operators by delegating the control of entities to AI.
- enabling large-scale experimentation and Monte-Carlo simulations by providing behaviors for every single entity in the scenario.

A common challenge in using advanced simulation and action spaces. In particular, if the chosen action space is too large (in the sense that too many actions are accessible to the agent at a given time) the DRL algorithm may not find a meaningful action to achieve the task at hand. A natural choice for the representation is thus one that is compact and based on standards, e.g., HLA or C2SIM. The latter in particular, through its semantic nature, provides a promising approach for interfacing DRL agents with wargaming simulations, and is the solution applied in this paper. The C2SIM standard [SIS20, NAT24] developed to unify communication between Command and Control (C2) systems and simulations, provides a means of adopting doctrinal behavior and — as a side effect — creates simulation platform independence.

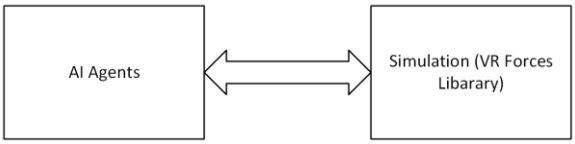


Figure 1: Training architecture. The DRL agent communicates directly via a dedicated message protocol with the VR-Forces engine. The engine is encapsulated into a custom executable via the provided C++ library.

Our architectural approach to build an AI agent that can control an entity within a simulation (being an airplane, a tank, or a platoon of soldiers) is to clearly separate the phase of training, Figure 1, from the phase of simulation, also called the inference phase, Figure 2. During the training phase, the simulation platform (in our case VRForces [MAK25]), is used as an engine that allows us to run the studied scenario in a controlled and repeatable way: the DRL agent is exposed to a large number of situations sampled from the scenario, with the aim of learning to act proficiently in each of them. During the simulation phase, the trained agent is connected via standard protocols (in our case C2SIM) to the simulation engine, allowing other participants, being human or artificial agents themselves, to interact with it. In this phase, the agent does its best according to what it has seen in the training phase, but it can't learn any new behavior: everything it knows is acquired during the training phase. It is worth noting that, in both phases, the action space and the observation space are kept the same, assuring a smooth transition between them.

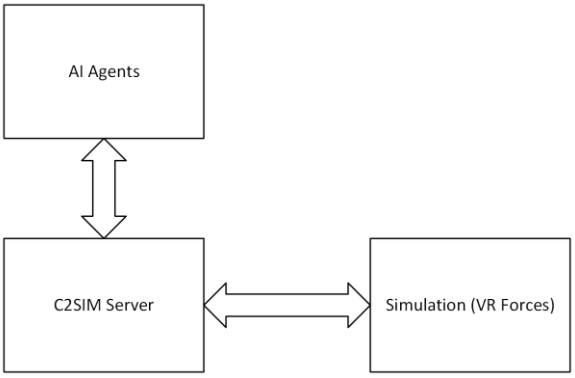


Figure 2: Inference architecture. The trained agent is connected to the simulation engine via reference server implementation for the C2SIM standard.

2. Training RL Agents

In this section we focus on the training phase and on the algorithms that we run to teach the agent to act appropriately in a given scenario. In DRL [MKS+15], an agent interacts with an environment in discrete steps: at each step the agent observes the state of the environment and executes an action choosen from a predefined set, then the environment reacts to this action by updating its internal state and providing to the agent a reward, i.e., a number that tells the agent how good or bad was the action just executed. This agent-environment interaction proceeds in episodes, starting with an initial configuration until the objectives of the mission are completed or a maximum number of steps is reached. The goal of the agent is to maximize the return, defined as the sum of the rewards obtained during an episode. In our setup, the environment is represented by a VR-Forces scenario instance, and the agent is the entity that must be trained. During the training, all the entities that are not directly controlled by the trained agent follow a predefined deterministic behaviour that remains fixed across episodes. We have selected three RL algorithms from the many available: Deep Q-Network (DNQ) [MKS+15], Proximal Policy Optimization (PPO) [SWD+17] and AlphaZero [SHS+18]. DQN is the neural network extension of the popular Q-Learning algorithm [Wat89]: it was the first algorithm able to reach a human-level play in the Atari game suite. PPO is a newer algorithm, very popular and effective in control task, i.e. driving a car or piloting an aircraft. AlphaZero was the first RL algorithm to reach master level in chess and Go with no human demonstration, learning completely from scratch and self-play: it is known to be very effective in strategy games. All these algorithms are widely used and tested, and cover most of the use case that we can encounter in a wargame scenario, like manoeuvring an entity and deciding for a strategy. At the time of writing, the state-of-the-art versions of the three selected algorithms are written in the Python language and they expect the environment to satisfy the Gymnasium requirements: right now, VR-Forces does not yet provide a native way to expose a scenario in the form of a Python compatible Gymnasium [TTK+23] environment, ready to be used for training. As a consequence, to concretely connect the VR-Forces simulation engine with an instance of one of the three mentioned algorithms, we developed a custom training infrastructure. Such infrastructure allows us to launch many parallel instances of the same VRForces scenario, either in the same or in different machines, and to manage the communication required to connect them to the RL algorithm. The exact format of the messages exchanged follows a custom designed binary protocol, tailored to be fast and efficient, sent and received over a standard TCP/IP socket. The power of this solution comes with a limitation too: for any new scenario that we want to simulate, some messages of the protocol need to be adapted to support the specific characteristics of the scenario at hand. Given a scenario, the total training time required by an agent to solve it could vary greatly: we start from a few hours for simple scenarios, to several days for complex ones. This variability can be explained not only with the complexity of the scenario, but also with the chosen algorithm and the performance of the simulation engine. A wargame simulation like VR-Forces has been designed to provide realistic simulation behaviours, focusing on real time interaction and gameplay: this is at the same time an advantage and a limitation. An advantage because the results are accurate and reliable, a limitation if we consider that the speed at which it can interact with the agent is in the order of ten interactions per second: a couple of order of magnitude slower than a common Gymnasium environment usually employed to develop and test RL algorithms. To overcome this slowness, a massive parallelization is required: many copies of the environment are run in parallel so that the agent can accumulate experience faster. With our hardware, a 64-cores CPU and 256GB of RAM, we could run up to 10-20 parallel environment, depending on the complexity of the scenario.

Regarding the three algorithms tested, two of them, PPO and DQN, are model-free, which means they can learn only from the interaction with the environment, as described above. The AlphaZero algorithm instead needs an additional ingredient, called the model. The model is a copy of the environment, that can be used by the algorithm to plan future moves in advance, before acting in the main environment. Planning is a powerful tool that enables stronger behavior, especially in strategy games. By leveraging the ability of VR-Forces to save and restore the state of a game at will, we enable the use of model-based algorithms in our infrastructure. The limitation in this case is a tiny but noticeable drop in performance and the need of powerful hardware, especially in terms of memory, since states saved by the simulation engines must be kept in RAM.

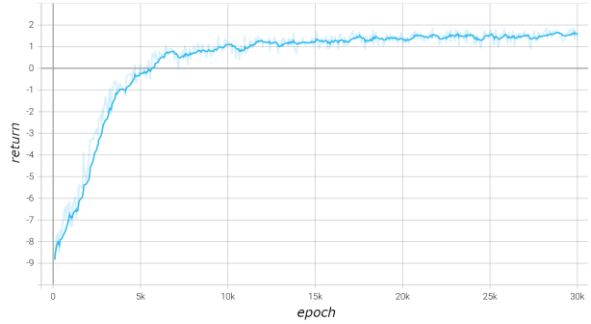


Figure 3: Typical learning curve for a DRL agent.

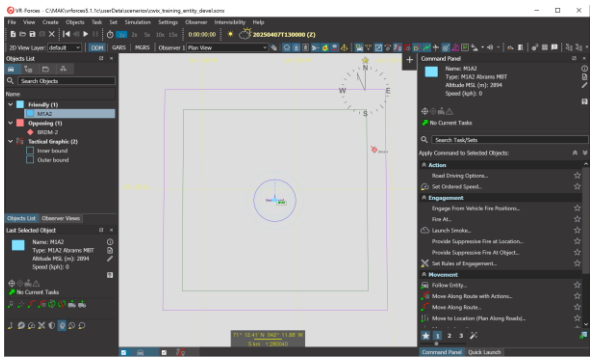


Figure 4: The tank scenario on VR-Forces

The scenario that we designed to test the training and deployment of a DRL agent is the following: there is a tank that is waiting for some threat entering in his assigned patrolling area. If an hostile ground entity is discovered, it tries to seek it and neutralize it. The apparent simplicity of the scenario allows also a great flexibility: the trained agent is not tied to a particular scenario but it is general enough to be run in many different contexts. In Figure 3, we can see how the episode return (Y axis) increases as the agent is faced with many kinds of enemy intrusion (X axis). As the learning progresses, the agent is increasingly able to seek and neutralize hostile entity. In Figure 4, we can see an example of an episode in which the enemy is approaching from the upright quadrant of the patrolling tank.

3. *Standards integration and Interoperability*

In this section we talk about the simulation phase, where we make use of an already trained agent to control entities in a scenario.

As stated in the introduction, we adopted the C2SIM standard to bridge the gap between DRL agents and the virtual environments used for training and experimentation. In military contexts, C2 systems issue orders and receive reports, while simulations provide a safe and flexible environment to test responses. C2SIM formalizes this exchange, ensuring that the flow of information between the two domains is consistent and unambiguous.

In our context, an AI agent acting as a decisionmaker can be viewed as a functional equivalent of a human-operated C2 system. The standard allows us to plug the agent into the simulation just as we would connect a human-operated C2 tool, without requiring modifications to the simulation itself. This makes the integration clean and scalable.

Another reason for adopting C2SIM is interoperability. The military simulation landscape is highly diverse, with many different systems developed for specific use cases, and it is unlikely that a single simulator will ever satisfy all operational needs [arm24, GKH+20]. Standards provide the necessary glue to connect these systems. By working through C2SIM, we gain the ability to use the same AI agent across different simulators, as long as they support the standard. This multiplies the value of training investments and facilitates coalition and joint operations, where different partners rely on different tools but need to work together.

At the same time, working with C2SIM introduces constraints. The standard is still evolving and currently supports only a limited set of orders (e.g., attack, defend, scout, and move) and reporting mechanisms. This means that the action space is narrower than what an AI agent might ideally exploit. Agents usually benefit from richer and more frequent information [HS15], for example continuous updates on the location and state

of all entities in the scenario. To deal with this, we faced an implementation choice: either consult the agent only when a C2SIM report arrives (using the information contained in the report itself), or maintain our own internal model of the scenario, updated whenever reports are received, and query the agent at regular intervals. We chose the second option, which means we always keep track of the last known position and status of all entities. As expected, this model is always slightly outdated, but it allows us to stabilize the interaction and control the frequency at which the agent is consulted. All this functionality has been encapsulated in a software layer interface that stays between the trained DRL agent and the C2SIM server.

4. Conclusions

Besides preliminary experiments, we tested the proposed approach in a combat scenario where the DRL agent needs to control a tank that operates in a battlefield. The results were highly promising: our trained agent was successfully integrated into a real-world exercise, demonstrating its practical applicability and effectiveness in realistic scenarios.

The next step was to validate the same setup in a broader interoperability context during the NATO Coalition Warrior Interoperability eXploration, eXperimentation, eXamination eXercise (CWIX)[Tra25]. In this case, we connected our framework not only to VR-Forces but also to simulations provided by partner nations. Specifically, we replaced VR-Forces with KORA [Gmb25] (Figure 5) and later with OneSAF [AD09], starting from a new scenario previously unknown to both the agent and the framework. In each case, the agent successfully issued attack orders to the controlled unit and acted as desired.

The most significant achievement is that switching between simulators requires no change in our AI framework. Instead of connecting VR-Forces to the C2SIM server, we simply connected to another simulator. From the perspective of both the agent and the framework, the underlying simulation system was transparent. This illustrates the true power of adopting a standard: without C2SIM, every new simulator would have required its own dedicated middleware or adapter. With C2SIM, integration was achieved within minutes, enabling seamless interoperability with systems that neither we nor the AI agent had encountered before.

References

[1] United States Army and Defense Technical Information Center (DTIC). Onesaf technical documentation. Technical Report ADA501150, DTIC, 2009. Accessed: 2025-09-15.

[2] armasuisse Science and Technology (S+T). Simlab: A demonstrator of a standardised simulation landscape for the armed forces. https://www.ar.admin.ch/en/simlab-en, 2024. Accessed: 2025-09-16.

[3] Mario Golling, Robert Koch, Peter Hillmann, Volker Eiseler, Lars Stiemert, and Andres Rekker. On the evaluation of military simulations: Towards a taxonomy of assessment criteria. arXiv preprint arXiv:2004.09340, 2020. Accessed: 2025-09-16.

[4] IABG GmbH. Software and system solutions — defence — iabg. https://www.iabg.de/en/industries/defence/ software-and-system-solutions, 2025. Accessed: 2025-09-15.

[5] Matthew Hausknecht and Peter Stone. Deep recurrent q-learning for partially observable mdps. arXiv preprint arXiv:1507.06527, 2015. Accessed: 2025-09-16.

[6] MAK Technologies, Inc. Vr-forces / mak one applications. https://www. mak.com/mak-one/apps/vr-forces, 2025. Accessed: 2025-09-15.

[7] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. Nature, 518(7540):529–533, Feb 2015.

[8] NATO Science and Technology Organization (STO). Nato modelling and simulation group activity msg-211: Modelling and simulation standards in nato federated mission networking [9] NATO Allied Command Transformation. (2024). Multi-Domain Operations and Digital Transformation.

[10] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. A general reinforcement learning algorithm that masters

Chess, Shogi, and Go through self-play. Science, 362(6419):1140–1144, 2018.

[11] Simulation Interoperability Standards Organization SISO. Standard for command-and-control systems– simulation systems interoperation (c2sim). https://cdn.ymaws.com/ www.sisostandards.org/resource/ resmgr/standards_products/ siso-std-019-2020_c2sim.pdf, 2020. Accessed: 2025-09-15.

[12] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. ArXiv, 2017

[13] NATO Allied Command Transformation. Cwix25 starts now! nato's largest digital interoperability exercise kicks off. https://www.act.nato.int/ article/cwix25-starts-jftc/, 2025. Published June 4, 2025; Accessed: 2025-09-15.

[14] Mark Towers, Jordan K. Terry, Ariel Kwiatkowski, John U. Balis, Gianluca de Cola, Tristan Deleu, Manuel Goul˜ao, Andreas Kallinteris, Arjun KG, Markus Krimmel, Rodrigo Perez- Vicente, Andrea Pierr´e, Sander Schulhoff, Jun Jet Tai, Andrew Tan Jin Shen, and Omar G. Younis. Gymnasium, March 2023. (accessed July 4, 2025).

[15] C. J. C. H. Watkins. Learning from Delayed Rewards. PhD thesis, King's College, Oxford, 1989.

Enhancing the Deployment of Multidomain Defense Systems through AI-Driven Automation and Integrated Engineering Processes to gain Strategic Advantage

Diego Tornese, Maria Giuseppina Motta, Gianluca Toscano, Antonio Zagaria, Antonio Tedone

Teoresi

www.teoresigroup.com

Abstract

The automotive sector's rigorous safety standards, such as ISO 26262, provide a valuable reference for improving safety, traceability, and robustness—especially as defence technologies increasingly rely on electronics, software, and simulation. These methodologies, matured through high-volume production and rapid iteration, can be aligned with System Life Cycle Management (SLCM) and Functional Safety (FS) principles in defence applications. Modern defence architecture now supports Multi-Domain Operations (MDO), integrating land, air, sea, space, and cyber domains. In this context - especially complex C4ISR systems - SLCM frameworks ensure end-to-end traceability, quality assurance, and regulatory compliance across all phases, from requirements engineering to operational support. Functional Safety ensures that systems maintain safe, predictable behavior even in the presence of faults, a critical requirement for mission assurance. This paper introduces AITestX, an AI-enhanced Test-as-a-Service (TaaS) framework developed by Teoresi. It integrates seamlessly with lifecycle management tools and supports the full SLCM process—from design, through development and qualification, to deployment and in-field validation. Aligned with NATO's CA2X2 objectives, AITestX enables Simulation-to-Action transitions, accelerating system certification, mission readiness, and operational decision-making.

1. Introduction

As defense systems grow in complexity and autonomy, safety and assurance frameworks must evolve accordingly. Modern platforms must operate seamlessly across land, air, Maritime, space, and cyber domains, a paradigm now formalized under the term **Multi-Domain Operations (MDO)**. Defense technologies are increasingly dependent on **software-defined systems**, **sensor fusion**, and **simulation-driven design**, creating a compelling need for robust System Life Cycle Management (SLCM) and **Functional Safety (FS)** methodologies.

This paper presents **AITestX**, an AI-enhanced Test-as-a-Service framework developed by Teoresi Group, designed to bridge the gap between system development, validation artifact generation, wargaming environments, and operational decision-making. An efficient process produces high-quality outcomes with minimal resource consumption while maintaining robustness, adaptability, and sustainability throughout the entire lifecycle.

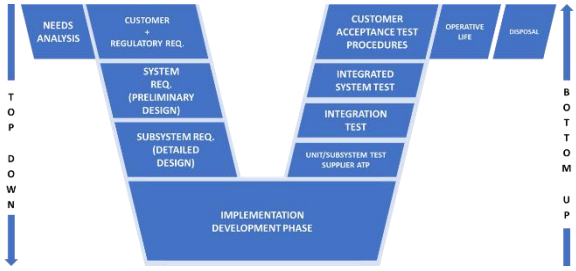


Figure 1: ***V Cycle***

The transition from fragmented to integrated approaches enables consistency from requirements to operations, directly supporting complex scenarios such as MDO and wargaming.

2. Safety-Critical Development and the Automotive Analogy

The automotive sector, guided by standards such as ISO 26262 [1], has established advanced methodologies for safety, traceability, and high-assurance verification within accelerated development cycles. These practices—rooted in architectural discipline, systematic fault analysis, and continuous verification—offer valuable insights for defence applications, particularly in the context of Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) systems.

The success of the automotive industry in delivering safe and traceable systems at scale—while maintaining rapid development—stems from a foundational insight: when engineering processes are sound and enforced through integrated tooling and automation, safety and speed are not mutually exclusive. ISO 26262 codifies this approach, emphasizing that functional safety is achieved not through post-development manual verification, but through continuous, embedded validation throughout the lifecycle.

Frameworks such as Automotive SPICE formalize the organizational and managerial structures that support scalable verification. Techniques like Hardware-in-the-Loop (HIL) testing enable realistic validation of embedded software without requiring full physical prototypes. These methodologies are directly applicable to defence systems, where mission-critical capabilities demand both accelerated certification and robust assurance. Their integration into C4ISR platforms enhances safety, traceability, and operational readiness, even in contested and adversarial environments.

Defence systems, particularly those operating under **mission assurance** constraints, require similar rigor. Unlike consumer vehicles, military platforms must be certifiable across **heterogeneous, contested, and adversarial environments**, which adds additional layers of uncertainty and functional dependency. Still, the **systemic frameworks**—from structured requirementtracing to **fault-tolerant behavior assurance**—are readily translatable.

3. *System Life Cycle Management (SLCM) and Functional Safety (FS) in MDO*

The Integrated System Life Cycle Management (SLCM) framework establishes a structured and traceable foundation for designing, validating, and sustaining complex defence systems, particularly C4ISR system. The approach emphasizes a seamless connection between requirement definition, functional validation, verification, and in-field operation through a unified engineering platform. By leveraging automation, AI-driven testing, and standardized data management, SLCM enhances robustness, consistency, and adaptability across multiple domains.

Modern systems—especially C4ISR—rely on modular open system approaches (MOSA), necessitating robust traceability from requirements to fielded capabilities.

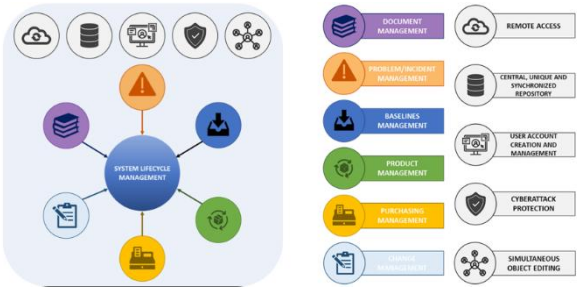


Figure 2: Integrated System Lifecycle Management

Functional Safety, when integrated into this SLCM framework, adds an additional layer of rigor focused specifically on ensuring that systems behave in safe, predictable ways even when faults or failures occur in compliance with standards such as ISO 26262, DO-178C [3], and IEC 61508 [2]. Rather than achieving safety through perfection—the attempt to design systems that never fail—functional safety acknowledges that all systems will eventually experience faults due to manufacturing imperfections, wear and tear, component degradation, or environmental stresses. Instead of attempting to prevent all failures, functional safety design ensures that even when faults occur, the system either detects the fault and takes corrective action, or the system fails in a way that is non-harmful or minimizes potential harm.

In MDO contexts, where interconnected systems operate across multiple domains and latency-sensitive data chains, FS becomes a mission-critical property. Traceability, documentation, and **correct-by-construction validation pipelines** are essential to prevent cascading failures or information asymmetry during real-time operations.

4. *Test as a service*

Test as a Service (TaaS) is an integrated, automated and adaptive software testing delivery model, enabling automated or manual testing on applications, embedded systems or simulated environments, leveraging technologies such as containers, orchestrators, APIs and virtualized environments. The service can include functional, non-functional (performance, security, compatibility), hardware-in-the-Loop (HIL) testing, integration testing, and regression testing, with support for CI/CD and DevOps.

Test Management is the second pillar of the integrated lifecycle framework. The Teoresi Test-as-a-Service (TaaS) model follows seven structured phases: preliminary analysis, environment setup, planning and strategy, execution, defect management, reporting, and continuous improvement. This framework ensures that testing is not a standalone activity but an integral part of the engineering lifecycle, providing transparency, consistency, and continuous feedback loops.

4.1 Teoresi TaaS Model

Test Management is the second pillar of the SLCM framework. The Test-as-a-Service (TaaS) model includes seven phases, covering planning, execution, reporting, and continuous improvement. It represents a complete and standards-aligned testing ecosystem integrated with lifecycle management tools.

The Teoresi TaaS model has seven stages:

1. Preliminary Analysis
2. Test Environment Setup
3. Test Planning & Strategy
4. Test Execution
5. Defect Management
6. Test Monitoring & Reporting
7. Continuous Improvement & Maintenance

Each phase corresponds to a control point in the validation process and helps ensure scalability, traceability, and reusability.

- Phase 1-2 (Preparation and Setup): define the functional scope, quality metrics and test environment (hardware, software and simulation).
- Phase 3-5 (Execution and defect management): represent the operational core of TaaS, where the test strategy is implemented in an automated manner and managed through CI/CD pipelines and issue tracking tools.
- Phase 6-7 (Monitoring and Improvement): introduce the data-driven dimension, with continuous feedback and performance and coverage measurements critical to ensure improvement and regulatory compliance.

4.2 Added Value

Standardization: the process is replicable and compliant with safety standards (IEC 61508, ISO 26262, DO-178C), enabling end-to-end auditing and traceability [3].

Automation: integrated with tools such as Robot Framework, Jenkins, Polarion, TaaS enables automated, manual and AI-driven testing to be merged into one coherent flow.

Multi-domain scalability: suitable for automotive, aerospace and defense systems due to modularity and centralized management of test environments.

5. AITestX Framework Overview

AITestX represents an integrated approach to automating the verification and validation aspects of SLCM while maintaining the rigor demanded by functional safety principles. The framework operates by orchestrating a seamless flow of information and processing across multiple integrated components, each of which contributes specialized capabilities to the overall system. Test automation transforms testing into a continuous pipeline. Instead of sporadic manual checks, the pipeline continuously pulls the latest requirements, executes related test cases, and returns results promptly. This enables earlier feedback, higher coverage, and fewer late-stage defects. Automation ensures repeatability, reduces variability and human error, and centralizes results for easier traceability and auditing. Productivity is enhanced as engineers spend less time on setup and manual review, while parallel execution reduces verification time and overall cost.

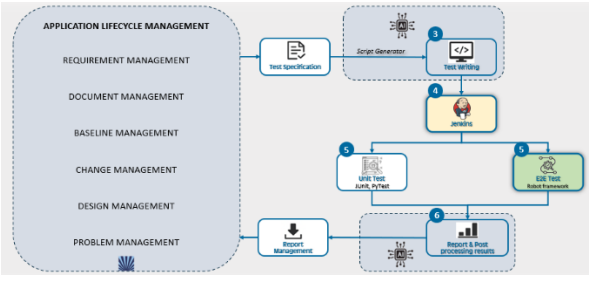


Figure 3 - Test Automation Workflow

To address these challenges, Teoresi introduces **AITestX**: an **AI-enhanced Test-as-a-Service** platform that integrates with the entire system development lifecycle. Built on principles of

automation, continuous integration, and traceability, AITestX provides:

- Integration between ALM and automation: complete traceability from requirements to testing.
- AI to support test writing: error and time-to-market reduction.
- CI/CD orchestration: continuous and automated execution.
- Multilevel coverage (unit + E2E)
- Feedback loop with reporting: continuous improvement.

This entire flow operates continuously, triggered either by schedule or by code changes in the source repository. When an engineer commits changes to the codebase, the CI/CD pipeline automatically exports the current requirements from Polarion [5], regenerates test specifications using the LLM generator, creates or updates the test scripts, executes all tests across all environments, collects results, and updates the verification status in Polarion. This continuous execution model means that verification status is never more than a few hours out of date, enabling rapid identification of regressions or newly introduced issues.

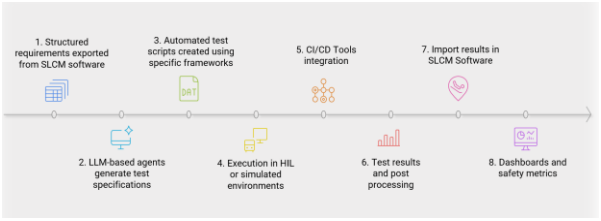


Figure 4 - AITestX: Key Features

5.1 Requirements Engineering via RESTful APIs

The starting point for AITestX is structured requirements exported from the SLCM platform—specifically, the Polarion Application Lifecycle Management system. Unlike traditional exports of requirements as unstructured documents or spreadsheets, the AITestX approach leverages Polarion's RESTful APIs to export requirements in carefully structured formats (XML, JSON, or CSV) that preserve all metadata associated with each requirement. This metadata includes the requirement identifier, the functional description, associated attributes such as priority level and safety integrity level assignment, trace relationships to parent and child requirements, and links to associated design documentation. This structured export is critical because it enables the downstream AI components to interpret requirements not merely as natural language text, but as formally attributed elements within the system architecture.

5.2 Natural Language Test Generation via LLMs

AITestX incorporates **Large Language Models (LLMs)** to interpret natural language requirements and generate formal test cases.

From these structured requirements, the second component of AITestX - the LLM-based test specification generator - interprets each requirement and generates a comprehensive test specification that defines how the requirement will be verified. Using **prompt engineering** aligned with FS constraints, these agents ensure tests are relevant, bounded, and auditable The prompt structure guides the LLM to consider multiple verification approaches, including nominal case testing (the happy path where the system operates normally), boundary value testing (testing behavior at the limits of specified operating ranges), and fault injection testing (deliberately introducing faults to ensure the system detects and responds appropriately). The LLM generates these test specifications in a structured format that preserves traceability to the originating requirement and includes acceptance criteria that precisely define what constitutes a successful test outcome.

5.3 Test Automation with Robot Framework

Generated test specifications are converted into executable scripts using **Robot Framework** [4] or other standard libraries. For systems with web-based or API-based interfaces, the Robot Framework provides a natural language like approach to test specification that produces highly readable, maintainable test code. For embedded systems and software-in-the-loop testing, frameworks like PyTest and JUnit provide the flexibility and performance needed for comprehensive unit-level testing. The critical difference between these automated scripts and manually written tests is that the automated scripts maintain explicit traceability back to the originating requirement, enabling automatic verification that all requirements have received appropriate test coverage.

5.4 CI/CD Pipeline and Execution Environments

Test execution is integrated into **Continuous Integration/Continuous Deployment (CI/CD)** pipelines using tools like **Jenkins** or **GitLab CI**. Tests run in both **simulated** (Model-in-the-Loop, Software-in-the-Loop) and **HIL** environments, enabling early fault detection and system behavior verification. This multi-level execution strategy enables comprehensive fault detection—errors in algorithm logic are caught early in MIL testing; interactions with hardware are validated in SIL testing; and realistic timing and computational load effects are verified in HIL testing. By executing all tests across all environments in parallel through the CI/CD pipeline, total test execution time remains manageable despite the comprehensive coverage.

5.5 Bidirectional Traceability and Results Mapping

Test outcomes are automatically mapped back to the originating requirement in **Polarion**, ensuring full **bidirectional traceability**. This is crucial for satisfying certification requirements, generating safety cases, and maintaining **configuration integrity** across iterations. As each test completes, the result—pass or fail—is automatically recorded in Polarion, associated with the requirement that was being tested. If a test fails, a defect is automatically created and linked to the requirement, enabling developers to understand exactly what requirement is not yet satisfied. As defects are fixed and tests are rerun, the requirement status in Polarion is automatically updated. This continuous feedback loop enables project managers to maintain real-time visibility into verification progress.

5.6 Reporting and Safety Metrics

AITestX provides comprehensive reporting and safety metrics through integrated dashboards and analysis tools. Allure and ReportPortal provide visualization of test execution results, including pass/fail rates, execution time trends, and regression identification. These tools integrate with Polarion to provide unified visibility across the entire verification process. Safety-specific metrics—such as test coverage broken down by safety integrity level, verification method distribution (how many requirements are verified through black-box testing, white-box code coverage, or formal analysis), and fault injection coverage—are automatically calculated and displayed. These metrics provide evidence essential for certification authorities and safety reviews.

6. Benefits and Impact

The integration of SLCM and AI-driven automation significantly reduces manual effort and testing time while improving consistency, safety, and traceability. By aligning with NATO's CA2X2 Simulation-to-Action objectives, the framework supports continuous verification and mission readiness, enabling adaptive decision-making in dynamic operational contexts.

The AITestX framework provides several tangible benefits:

- **50% Reduction in Test Design Time**: AI-generated test specifications drastically reduce manual effort in verification planning.
- **DevSecOps Alignment**: Modular pipelines enable **security-integrated verification** at every development stage.
- **Accelerated Certification**: Traceability and structured metrics support quicker compliance with defence safety standards.
- **Operational Agility**: Continuous updates to test suites enable rapid adaptation to evolving mission requirements.
- **Resilience and Robustness**: FS validation ensures safe behavior under fault conditions, reinforcing mission assurance.

7. NATO Alignment and Simulation-to-Action Readiness

AITestX is built with direct alignment to NATO's **CA2X2** objectives—supporting seamless transition from simulation to operational action. This is particularly crucial in **training**, **decision support**, and **doctrine evolution**, where test artifacts and safety models can directly inform mission rehearsal and live execution environments.

The framework supports **real-time synchronization between simulation outcomes and operational data**, enabling **evidence-based decision-making** in high-tempo environments. By combining test-driven development with M&S (Modeling and Simulation), AITestX transforms **validation artifacts** into **tactical assets**.

8. Conclusion

As the defense sector embraces multidomain complexity, new paradigms are required for **system validation, safety assurance, and lifecycle governance**. By importing proven methodologies from the automotive sector and leveraging AI-driven automation, AITestX provides a powerful platform for accelerating development, ensuring safety, and supporting operational decision superiority.

By bridging the gap between structured engineering and intelligent automation, AITestX exemplifies how **AI and SLCM** can coexist to meet the rigorous demands of modern defence systems—today and in the evolving future battlespace.

References

[1] International Organization for Standardization (ISO). "ISO 26262: Road vehicles — Functional safety," Parts 1–12. ISO, Geneva, 2018.
https://www.iso.org/standard/68383.html

[2] IEC. "IEC 61508: Functional safety of electrical/electronic/programmable electronic safety-related systems," Parts 1–7. International Electrotechnical Commission, Geneva, 2010.
https://webstore.iec.ch/en/publication/5515

[3] Discover DO-178C Guidance DO-178C: Software Verification

[4] Robot Framework Foundation. Robot Framework User guide
https://robotframework.org/robotframework/latest/RobotFrameworkUserGuide.html

[5] Siemens Digital Industries Software Polarion ALM documentation
https://polarion.plm.automation.siemens.com/techsupport/documentation

Simulating the Narrative Battlespace: Integrating Synthetic Social Media, AI Scenario Generation, and Autonomous Adjudication into Wargaming

Etienne van de Bijl, Daphne van Dulst, Huib Smit, Ragger Jonkers, Chihab Amghane

Royal Netherlands Aerospace Centre – NLR

www.nlr.nl

Abstract

This paper introduces a novel wargaming framework that describes how simulated narratives and information flows in social media can enhance wargaming. The wargaming framework is built around three integrated AI-powered components. First, an AI-powered scenario generator enables the wargame sponsor to define key parameters and strategic themes, generating tailored multi-domain scenarios. The scenario is used as input for a military training simulator such as Virtual Battlespace (VBS), as well as an AI–powered social media simulator. The social media simulator produces dynamic, persona-driven content that simulates public discourse and narrative development. It models audience sentiment affected by real-time input, including VBS, the facilitator and players. Finally, a conceptual AI-powered white cell is introduced to support the facilitator by modelling the consequences of player decisions, updating the game state on the fly and simulating realistic adversarial responses.

This integrated wargaming framework supports experimentation and training by:

- Enabling rapid generation of complex MDO scenarios
- Incorporating the IE into development and analysis of possible Courses of Action (CoA)
- Supporting the facilitator by rapidly simulating the effects of various CoA on the current scenario

To conclude, this research contributes to creating a more realistic, immersive wargaming experience adding educational value and improved planning exercises.

1. Introduction

Recent global conflicts have highlighted the significant impact of social media on narrative shaping, information dissemination, and morale influence (Amghane & de Marez Oyens, 2024); Centre for Countering Disinformation, 2025). To get a grip on the societal consequences, it is crucial to understand how information affects public perception and resulting events during warfare. Social media intelligence aids defence personnel in refining targeting strategies that intersect with multiple strategic domains. Military operations in the (public) physical environment might be severely hindered by negative online sentiment among civilians towards military operations and-actions or even the military as a whole. For example, this could lead to protests. The evaluation of conflict consequences via social media analysis is restricted to real-time or retrospective assessments, lacking predictive capabilities that are crucial for training, planning, and guideline formulation (Amghane & de Marez Oyens, 2024). However, legal constraints prevent military organizations from collecting and analyzing public social media data effectively (Ducheine et al., 2024).

To overcome the aforementioned privacy regulations of working with public social media data, we are currently investigating how, with help of Large Language Models (LLM), this data gap can be closed. Therefore, the aim of the article is to provide a safe and realistic environment that can be used to raise awareness, create knowledge and build skills in analyzing and using social media during conflicts.

Simulated social media environments can be integrated into wargaming. Wargames are designed to immerse participants in a variety of scenarios, encouraging them to confront complex challenges and gain insights into operational decision making (Montessori, 2024). The combination of social media simulation and wargaming offers a novel approach to preparing military personnel for the information challenges of modern warfare.

We present a conceptual framework that allows for rapid development of comprehensive MDO-enabled wargaming scenarios by introducing Artificial Intelligence (AI) to the wargaming method's components. Our framework emphasizes the

importance of understanding the IE and becoming aware of the risks, as well as the possibilities and benefits that it presents. Furthermore, it is aimed at enhancing *cognitive resilience* of defense personnel. Cognitive resilience is the capacity to anticipate, prepare, withstand and recover quickly from cognitive attacks through the effective preparation of groups and individuals (De Reus et al., 2025). Cognitive resilience is a combination of skill and attitude and thus can be trained – something for which wargaming with a social media simulator lends itself as a training method. We illustrate how existing AI algorithms can be effectively integrated into the construction of wargaming scenarios, particularly during the planning phase.

This paper begins by introducing characteristics of wargames and which characteristics will be used in our framework. This is followed by an explanation of the components of this new wargaming method: the social media simulator, the scenario generator and the conceptual AI-powered white cell.

2. *Wargaming*

Wargaming is a structured method for examining conflict or competition within an immersive safe-to-fail environment at a relatively low cost. At its core, it is a scenario-based model in which events, human decisions and resulting outcomes mutually influence one another (NATO Handbook; Bundeswehr). Participants are presented with a situation (either real-world or hypothetical) which they analyse and in which they make decisions. These decisions shape the following course of events, which in turn influences the subsequent choices participants must make. This cyclical interaction creates an environment of competitive challenge and creativity, overseen by an adjudicator to maintain structure and consistency (DEFD UK).

As illustrated in *Figure 1*, interrelated variables together determine the overall scope, structure and analytical value of a wargame. Wargames can be categorized along three dimensions: characteristics, purpose, and adjudication style. The level of analysis as a characteristic defines the scale or hierarchical layer at which decisions are made: strategic, operational, tactical or technical. This level shapes the game's objectives, participant composition, data requirements and the type of insights it can produce. The mode of play, another characteristic, can be physical, such as a board game, or digital, with online players. Lastly, wargames can be classified as either open or closed, depending on the level of information transparency among players. In an open wargame, all information is available to all players, resulting in a transparent and predictable game. In contrast, a closed wargame restricts information availability, increasing realism but also introducing uncertainty among players regarding their actions.

If we look at the purpose of a wargame, there two types: learning wargames for training and education, and analytic wargames for broader analytical objectives. Learning wargames focus on training and education, allowing players to apply new knowledge and receive feedback on their decisions. Their analysis is simpler and primarily assesses whether participants have gained a better understanding of key concepts. Analytic wargames, on the other hand, are designed to contribute to a broader analytical effort, guided by specific research questions. These games require a more rigorous analytical process to generate insights and evaluate plans, concepts, or strategies.

The adjudicator is responsible for determining the outcomes of player decisions and interactions within the scenario. This involves applying rules, data or professional judgment in a consistent and transparent manner. The credibility of a wargame often depends on participants' confidence in the impartiality and accuracy of these determinations. The facilitator, in contrast, manages the flow of play, ensuring that participants understand the sequence of activities, that tempo and focus are maintained, and that communication remains clear and productive. Together, facilitators and adjudicators uphold both the procedural integrity and the analytical credibility of the exercise.

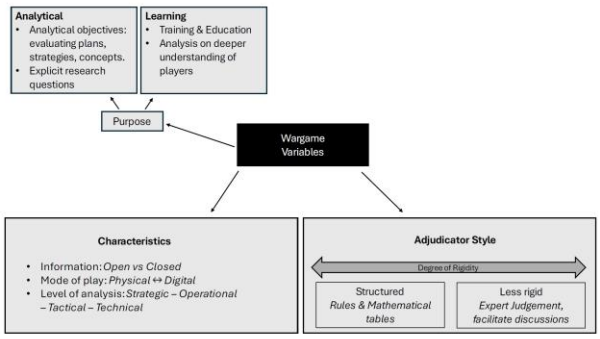


Figure 1: Main characteristics of wargames

Referring to the wargame variables in *Figure 1*, our framework is focused on *learning* objectives by exposing players to an interactive IE. The adjudicator is structured, since an AI-based adjudicator uses underlying mathematical models to make decisions. In terms of characteristics, the framework allows for *o en* information components, such as the generated content in social media to be analyzed in real-time. This analysis is mostly on the operational level. Additional to physical mode of play, autonomous adjudication supports digital mode of play due to fast and structured decisions.

3. Framework to simulate the narrative battlespace

In this section, we conceptually describe the components of our framework and afterwards give a formal definition of how these components are connected. We have developed a cohesive and immersive training tool that brings together several complementary components designed to enhance realism and interactivity in wargaming. In the following sections, we introduce each of these components and explain how they collectively contribute to the overall training experience. These elements and their dependencies are illustrated in Figure 2 and briefly introduced at the beginning of this section.

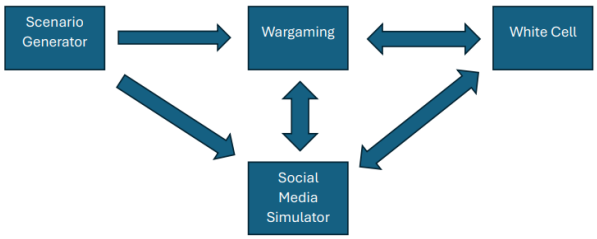


Figure 2: Concept for wargaming with integration of a scenario generator and social media simulator and white cell

The scenario generator delivers a tailored operational context, defining mission objectives, environmental conditions and potential threats that define the initial state of the world for both the wargame and the social media simulator.

The wargame Go4It (Hulst et al., 2012), has served as an example and inspiration for our proposed concept. Go4It is a wargame designed to raise awareness and support decision making through structured play and collaboration, illustrating how interactive game mechanics can foster learning and reflection in complex environments. Go4It provides the overarching narrative, role distribution and decision-making framework that pursues a multitude of objectives.

One of the wargaming objectives in our framework is understanding changes in public discourse, narrative shifts and sentiment changes modelled by the social media simulator. The social media simulator dynamically generates persona-driven content that reflects this public discourse, in response to events in the scenario. The social media domain as effector on the world state is positioned as the central training focus, placing participants in a dual decision-making space: the physical-operational environment and the IE. By observing how information spreads, communities polarize and perceptions being influenced, participants gain a deeper understanding of the potential risks and opportunities that social media presents in contemporary conflict.

The AI-powered white cell ensures that scenario progression remains coherent and credible. It adjudicates the outcomes of player actions, updates the scenario and simulates adversarial behaviour (acting as the opposing force in both the operational and IEs). This automation reduces the cognitive load on facilitators, allowing them to focus on guiding player interaction and conducting meaningful debriefs.

4. Formalization

In the previous section, we introduced our framework in a conceptual manner. Now, we will formulate the framework in a formal way and make the connections between the different components explicitly by casting it in the Reinforcement Learning (RL) paradigm. This is a concept in the realm of AI. In Figure 3, the elements of the proposed framework are illustrated in the RL concepts, which we will now describe.

In an RL problem, each agent, or participant in the wargame, selects an action At at timestep t. at timestep $t$in order to achieve a goal (multiple/collectively shared). A timestep here can be seen as the turn in a wargame and $t \in \{1,2,3, \dots, T\}$ where T is the final turn. At each turn, each agent selects an action to be evaluated by the environment.

Agents take actions based on the state of the environment, which we denote by St . Here, St can be observed as a multi-dimensional variable describing the

state of the environment at turn t. For example, in the context of wargaming, one dimension in the multi-dimensional state space can describe the number of blue agents in the field. One can consider more dimensions, all that are relevant to describe the state of the environment. The state $S1$ is the initial situation and the state $St+1$ is the state of the next turn, which is determined in the environment. Agents observe this snap-shot and formulate actions.

The goal of wargaming is not necessarily to win, but to gain insights. An important aspect here is the reward function. The primary goal of an agent is to maximize the cumulative reward over time. The environment provides a reward signal, which is a single number at each time step, indicating the desirability of the agent's actions. This reward signal defines the good and bad events for the agent, guiding its decision-making process to achieve long-term objectives (Sutton & Barto, 2014). In the context of wargaming, an example is that one can think of saving hostages or winning a battle. However, the overall goal might be to win the war (even though this is not primarily the aim of wargaming). In the context of the social media simulation, we can express the reward by indicating the social unrest for example (stances/narratives).

As described above, wargaming can be tackled with RL methods. In Figure 3, RL is depicted including the elements of the proposed framework. In the wargame (augmented by the social media simulator) there are players interacting with the environment trying to achieve their own goal (or a collective goal). The arrows and corresponding Action/State/Reward shows what each of the component has as input and outputs and how the variables are determined in the next turn.

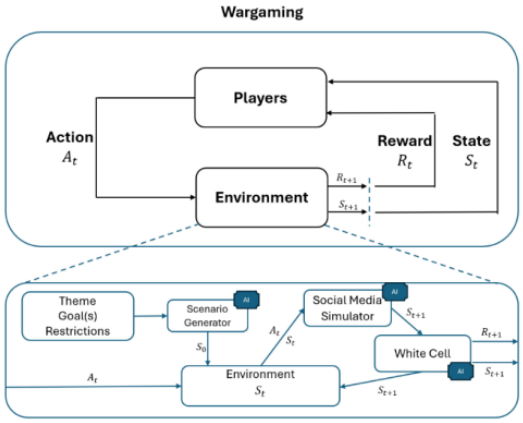


Figure 3: Reinforcement learning applied to the concept framework

5. AI- driven Scenario Generator

In the context of wargaming, scenario generation is both a preparatory and operationally critical function. Scenarios establish the initial conditions, constraints and narrative framework in which player decisions and interactions unfold. Their quality, coherence, and adaptability directly determine the realism, analytical thoroughness and training value of the wargame. Traditionally, creating such scenarios has been a time-consuming and resource-intensive task, requiring significant expertise to achieve the necessary fidelity.

To explore how this process could be accelerated and enhanced, a proof-of-concept system was developed that enables the collaborative creation of wargame scenarios through interaction between a Subject Matter Expert (SME) and an AI-based scenario generator (Van Oijen et al., 2025). The process begins with the SME formulating an operational question or training requirement, for example: "want to further evelo the conce t of increasing awareness of the consequences of social me ia using a wargame." The SME then provides structured input in the form of natural language prompts specifying the platform, mission objectives, environmental conditions, threats or other relevant constraints. An LLM extracts key topics from this input and uses Retrieval-Augmented Generation (RAG) to search an external database for relevant domainspecific information. Here, data is stored as vectors that represent unstructured data such as images, text or audio. The retrieved knowledge, combined with the SME's query and predefined system instructions, forms the complete prompt for the LLM. Based on this input, the scenario generator produces a tailored scenario whose specificity reflects the detail provided by the SME. This may range from a broad conceptual framework suitable for further refinement to a fully detailed specification containing geographic settings, force compositions, environmental conditions and mission phases, ready for integration into a simulation or wargaming environment (see Figure 4). The scenario generator could also be used to produce scenarios for wargames that do not include any modelling of the information environment

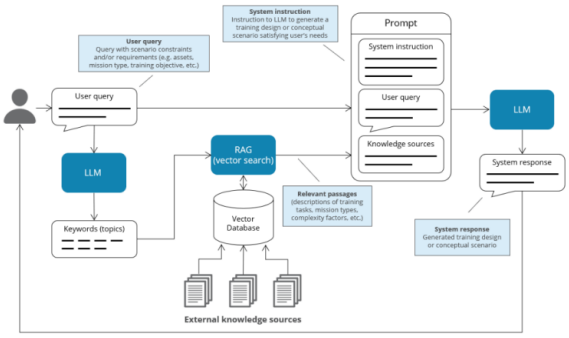


Figure 4: Interaction design between user and LLM

6. AI- driven Social Media Simulator

Conceptually, our scenario generator is linked to the social media simulator. In this configuration, the social media simulator would receive the initial world design from the scenario generator. The simulator produces dynamic, persona-driven content that reflects how those events might be perceived, discussed and amplified publicly. This integration would allow the wargame to incorporate both the physical-operational and informational dimensions of conflict.

By incorporating social media in a simulated manner, analysing the IE can be trained, as well as the 'live' use of dynamic social media information to support the mission objectives. In earlier work by (Park et al., 2022) a social design of communities was formulated specifying a community as a set of goals, rules and personas. The personas are the people that are part of a group and participate in discussions related to the goal whilst adhering to the specified set of rules. The behaviour of the persona is modelled by a set of parameters describing the persona, such as age, gender and stance on a variety of topics. The number and level of detail of these parameters can vary and dictates the complexity and realism of the behaviour observed in the simulated discourse. Essentially, a social media simulator should mimic how people behave and interact in groups, e.g. communities, and participate in discussions on online platforms.

7. Simulating social media with large language Models

One point of interest in the research on social media network simulation is how information is propagated, known as influence iffusion. (Kempe et al., 2015) describe an approach to determine the ke la ers, formulated as finding the nodes that maximize information spread. As an adaption on this approach, a study leverages the power of LLMs to simulate social media networks (Zhang et al., 2025). Qualitative analysis of user behaviour in the network is provided by visualizing LLM-generated responses over time. It must be noted that the content of these LLM generated responses do not affect whether a user becomes 'influenced' or not.

A more agentic approach is adopted in the research of (Ferraro et al., 2025) in which a social media network is simulated by modelling each user as an LLM-powered agent capable of making decisions that a real user could also make. These actions include writing a post, responding to someone's post, resharing a post, and not taking any actions. The action the agent performs is based on their persona and the content they are presented with. A vector-based recommendation system is used to expose the agent to certain content, behaving similarly to contemporary social media platforms (Edelson et al., 2025).

The proposed social media network simulator in this paper is a privacy respecting solution for analysing users and groups. Even though real world social media data could theoretically be used as a starting point (Gao et al., 2025), in this research we will not use any social media content of real life users, but rely on creating enough topic context from public articles. This decision was made to ensure full compliance with applicable privacy and data protection regulations governing the use of user-generated online content. Similar to aforementioned work, we represent each user in the social media network as an LLM agent with their own personality and one or multiple stances on topics.

The simulated social media network can be analysed to gain insights into its users. One such insight is to identify the target audience for practicing effective strategic communication. Another example is gaining insight on the spread of (mis)information and fake news; how audiences can be deceived by presented content on the platform. These capabilities can be deployed to make people (e.g. operators of the MoD, also society) aware of the benefits and risks of social media, as well as providing it as a training application.

In a real social media network, the person behind the post is often unknown, yet the content of the post can signal that persons ideological beliefs or political stance

on a topic. In our simulated social media network the stance of each user can change over time based on the dynamics of the network and the content the user is exposed to. Therefore, the average stance on a topic is on a certain point in time is measurable after content and interaction has been simulated. More generally, the topic sentiment of a post can be observed ("What emotion does the message convey? Is the user angry?") and predicted.

Figure 5 illustrates the visualization ofthe development of the stances of users over turns of social media users in wargaming. Two groups, each consisting of 25 individuals, are formed with stances ranging from -1 to 1. The extremes (-1 and 1) represent the strongest opinions, while 0 denotes a more neutral stance. As social media users interact with each other by posting/messaging each other, their groups become increasingly polarized. If the stances within a group become too dissimilar, the group may even become disconnected.

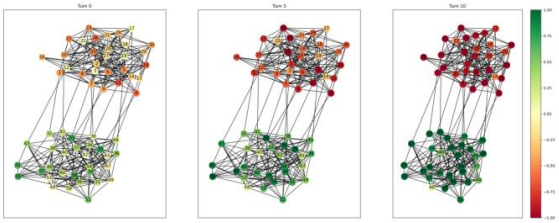


Figure 5: Illustration of two groups polarization during wargaming with social media simulator

8. AI- driven White Cell

Historically, human adjudicators have relied on a variety of tools and techniques to support their work. Operational analysis can provide probabilistic ranges of possible outcomes. Computational models can simulate engagements with high fidelity. Moderation can adjust outcomes toward training or analytical objectives and role play can inject human unpredictability into the scenario. Each of these methods, however, requires active human interpretation and integration, under limited time. This creates persistent challenges in terms of speed, consistency and scalability, particularly in wargames with dynamic and high-volume data flows.

In our proposed framework, these functions remain clearly delineated. The adjudicator role is performed by an AI-enabled system capable of processing large volumes of operational and scenario data in real time, applying consistent adjudication logic, dynamically modelling adversary behaviour and updating scenario progression. This automation is intended to reduce cognitive load, improve decision speed, and increase the consistency of outcomes. The facilitator role, however, remains in human hands. By retaining a human facilitator, the wargame benefits from the ability to guide interactions in real time, interpret player dynamics and conduct in-depth post-game debriefs that probe the rationale behind player choices and stimulate discussion.

As outlined earlier in this paper, the AI White Cell is presented as a conceptual construct intended to illustrate how AI could augment facilitation and adjudication in wargaming. It serves as a conceptual model to illustrate how AI could support or partially assume the functions of facilitation and adjudication in future wargames. While the operational implementation of this concept lies beyond the scope of the present study, its underlying principles were previously explored in an experimental wargaming environment described by Muurmans et al. (2024). In that earlier work, AI-driven agents and reasoning components were used to simulate decision-making dynamics and partial adjudication tasks within a synthetic command-and-control context. The observations from those experiments provided initial evidence of how LLMs can mediate between simulation data, human intent and scenario logic. These are insights that directly informed the conceptual framework presented here. A full-scale technical implementation could Figure 5: Illustration of two grou s olarisation uring wargaming with the social me ia simulator. Links in icate connections between in ivi uals an colour in icates stance towar s a certain subject. As time rogresses, the grou s become more iverte in their stances. be developed and validated in future research phases, once the necessary technical, methodological and ethical considerations have been addressed.

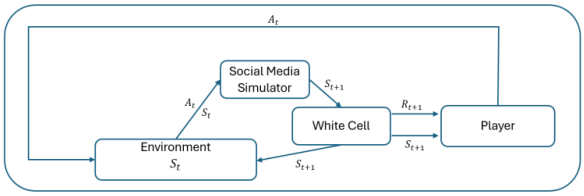


Figure 6: Schematic overview of the proposed AI-White Cell framework

The environment represents the overall simulation world at state St . Based on this state, the social media simulator processes both the current environment (St) and player actions (At), producing an updated

intermediate state (St+1). This new state is then passed to the white cell, which functions as the intelligent adjudicator. The white cell updates the global environment with the new state, while also generating the reward signal (Rt+1) and communicating both Rt+1and St+1 back to the player. The player interprets this feedback and selects a new action (At), which is sent to the environment, initiating the next decision cycle. This structure enables the AI-Driven white cell to mediate between player intent, simulated effects (e.g., social or informational dynamics) and the evolving operational environment, thereby maintaining a coherent and adaptive wargame flow.

9. Discussion

In this study, we propose a conceptual framework that integrates the IE into wargaming using AI. The two primary objectives of this integration are enhancing the cognitive resilience of defence personnel within IE in modern conflicts and to skill them in exploiting the capabilities of social media to obtain strategic goals. Additionally, our proposed framework aims to ease the complex process of constructing a wargame.

This paper presents the results for establishing a framework that integrates a conceptual AIdriven scenario generator and adjudicator with a social media simulator. Due to this exploratory nature, the conclusions are primarily theoretical and have not yet been supported by empirical testing in real-world or training environments. Two future work prospects are identified that can be addressed in future research.

Future work is to validate the proposed framework with empirical results showing that indeed defence personnel become more resilient in the IE and can exploit the existing benefits of social media for obtaining strategic goals. The state of social network at each turn is crucial here. Endusers should be able to extract relevant information of the social media domain to be able to understand the state of the network. A possibility for additional research lay in the post-game analysis of social media content generated during the wargame. This would be particularly valuable for personnel working in strategic communications, who could be tasked with reconstructing the course of events based solely on the simulated social media output. Such an exercise would provide insights into the extent to which narrative and sentiment patterns can Figure 6: chematic overview of the ro ose AI-White ell framework reveal operational developments, thereby further enhancing skills in detecting, interpreting and responding to information environment dynamics.

In addition, the possible actions of players should be known to them; even though outcomes might be an expected result of actions, they should be (somewhat) explainable. Furthermore, the simulated social media environment's fidelity can be considered as limited due to potential biases and unrealistic discourse patterns resulting from the LLMs' training data and system prompts, as well as the absence of live social media data.

10. Conclusion

This study presents a conceptual, novel wargaming framework that integrates an AI-powered scenario generator, a dynamic social media simulator and an AI-enabled white cell into a single cohesive training environment. The combination of these components enables faster scenario development, consistent adjudication and the creation of realistic responses. The framework expands the scope of wargaming beyond the physical-operational environment to include the information environment as a central element of play. By doing so, it allows participants to experience first-hand the interplay between operational decisions and narrative dynamics and to observe how social media can influence perceptions, shape public discourse and impact mission outcomes.

References

[1] Amghane, C., & de Marez Oyens, P. M. W. (2024). Social Simulator Madness: Simulating Social Behavior in Dynamic Environments. 2024 Interservice/In ustr raining, imulation, an E ucation onference, 24232, 1–11. https://s3.amazonaws.com/amz.xcdsystem.com/44ECEE4F-033C-295CBAE73278B7F9CA1D_abstract_File18333/FinalPaperStage3_24232_0825052944.pdf

[2] De Reus, A., Doherty, G., & Van Dulst, D. (2025). Evaluation Criteria and Use Cases for Information Operation/Social Media Simulators. NA cience an echnolog rganization. https://www.sto.nato.int/document/evaluation-criteria-and-use-cases-for-informationoperation-social-media-simulators/

[3] Ducheine, P. A. L., Pijpers, B. M. J., & Zwangenburg, M. C. (2024). an en voor e Leeuw: Een voor haar oel en o haar taak bereken e krijgsmacht in e informatie-omgeving (pp. 1–54). Amsterdam Center for International Law (ACIL). https://www.tweedekamer.nl/downloads/document?id=2024D51201

[4]Edelson, L., Haugen, F., & McCoy, D. (2025). A Comparative Survey Of Algorithmic Feed Recommendation System Designs. A M ransactions on Recommen er stems, 3757327. https://doi.org/10.1145/3757327

[5] Ferraro, A., Galli, A., La Gatta, V., Postiglione, M., Orlando, G. M., Russo, D., Riccio, G., Romano, A., & Moscato, V. (2025). Agent-Based Modelling Meets Generative AI in Social Network Simulations. In L. M. Aiello, T. Chakraborty, & S. Gaito (Eds), ocial Networks Anal sis an Mining (Vol. 15211, pp. 155–170). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78541-2_10

[6] Gao, C., Lan, X., Lu, Z., Mao, J., Piao, J., Wang, H., Jin, D., & Li, Y. (2025). ^{3}: ocial-network imulation stem with Large Language Mo el-Em owere Agents (No. arXiv:2307.14984). arXiv. https://doi.org/10.48550/arXiv.2307.14984

[7] Hulst, A., Christoph, N., Maij, A., & Aldershof. (2012, January 1). GCAM: A modeling approach to the comprehensive approach. In rocee ings of the Nato Mo elling imulation an Gaming Worksho , tockholm. https://www.researchgate.net/publication/269332641_GCAM_a_modeling_approach_to_the_comprehensive_approach

[8] Kempe, D., Kleinberg, J., & Tardos, E. (2015). Maximizing the Spread of Influence through a Social Network. heor of om uting, 11(1), 105–147. https://doi.org/10.4086/toc.2015.v011a004

Strategic Foresight under Uncertainty: Simulating the Collapse and Regeneration of Strategic Postures

Asmaa Taib

Strategos MSc, University of Genoa

strategos.simulationteam.com

Abstract

Conventional strategic planning often assumes linear execution from objective to outcome, an assumption that collapses under real-world complexity. This study introduces a simulation-based foresight methodology that intentionally induces strategic breakdowns to examine recovery and adaptation. The framework integrates hybrid Agent-Based and System-Dynamics modeling with Reinforcement Learning, transforming simulation from a validation tool into a diagnostic instrument for strategic survivability. Results from ultra-stochastic multi-agent experiments show that cooperative agents develop adaptive behaviors, extend mission duration, and sustain performance despite persistent instability. The findings demonstrate that resilience and learning emerge not from the avoidance of collapse but from structured engagement with it, supporting NATO's call for anticipatory and adaptive planning under deep uncertainty.

1. *Introduction*

Strategic planning has long rested on an assumption of linearity: that clearly defined objectives, rational actors, and sufficient resources will yield controllable and predictable outcomes. In contemporary security environments, this premise is routinely invalidated by dynamic complexity and irreducible uncertainty. Political, cognitive, technological, and informational disruptions introduce volatility that can render ostensibly robust strategies obsolete within days or hours. In multi-domain operations, assumptions of continuity and predictability are quickly overturned; even well-crafted plans prove brittle when conflict dynamics shift abruptly.

This paper advances a simulation-centered approach to strategic foresight that treats collapse as a diagnostic condition and strategic survivability as the capacity to break, learn, and recover in time to remain relevant. A Collapse and Recovery Simulation intentionally induces cognitive, technological, and narrative breakpoints within a reinforcement learning decision environment, making adaptation observable through the resulting regenerative dynamics of reconfiguration, redirection, or replacement. The framework moves simulation beyond scenario validation toward structured disruption and recovery rehearsal, in line with allied calls for anticipatory thinking and flexible doctrine. Regeneration levers and clearly defined Measures of Effectiveness are used to quantify recovery, while robustness under deep uncertainty is assessed by examining strategy performance across stressed conditions. The result is a practical pathway for training, planning, and foresight that privileges survivability over nominal optimality.

This study addresses three guiding questions:

1. How can simulation be used to assess the adaptive capacity of strategic postures under deep uncertainty?
2. What collapse–recovery mechanisms emerge when interacting agents learn through reinforcement within volatile, multi-domain environments?
3. Which indicators most effectively measure strategic survivability as a system-level property?

In the following sections, we elaborate on the conceptual foundations for strategic survivability (Section 2), the simulation methodology (Section 3), the observed results (Section 4), and conclusions (Section 5).

2. *Conceptual Foundations for Strategic Survivability*

Effective strategy operates within ambiguity, complexity, and uncertainty, where the assumption of linearity routinely fails. Modern conflicts present wicked problems in which attribution of cause and effect is uncertain and continuous adaptation is required. While foresight tools such as scenarios and red teaming seek anticipatory awareness, most simulations avoid provoking or observing systemic collapse, leaving a gap in understanding regeneration. This section reframes collapse as a deliberate diagnostic tool, shifting analysis from optimizing expected performance to testing structural adaptability and long-term strategic survivability.

2.1 The strategic Environment: Dynamic Complexity and Irreducible Uncertainty

The contemporary strategic environment is characterized by dynamic complexity, where effects are rarely proportional to causes and linear analysis loses validity. Complex systems consist of interdependent parts whose interactions generate emergent and often unpredictable behavior. Strategists must therefore operate in conditions defined by Volatility, Uncertainty, Complexity, and Ambiguity (VUCA), where unpredictability is intrinsic. Military strategy faces wicked problems that cannot be solved through technical expertise or management alone, as their causes and effects are intertwined and evolving. In such contexts, point prediction is futile; instead, disciplined interpretation of feedback becomes essential to avoid cognitive bias and misjudgment. Strategy must function in a world where chance, uncertainty, and ambiguity dominate.

2.2 The Evolution of Strategic Foresight Methodologies

Strategic foresight represents a shift from deterministic prediction toward envisioning multiple possible futures to strengthen present decision-making. It is defined as the systematic debate of complex futures, serving as a precursor to planning conditions policies for robustness and flexibility.

Core foresight methodologies include:

• Horizon Scanning: A systematic examination of weak signals that may indicate emerging developments, threats, or opportunities. It supports the intelligence-gathering phase of strategy by anticipating potential future conditions.

• Scenario Analysis: The structured generation of alternative futures, distinguishing between exploratory scenarios (what might happen) and normative scenarios (desired futures and the pathways, back casting, needed to reach them).

• Adversarial Challenge (Red Teaming): The deliberate testing of assumptions, plans, and strategies from alternative or adversarial perspectives to ensure analytical rigor.

Despite their utility, traditional foresight applications often remain anchored in preserving stability or optimizing existing capabilities, seldom engaging with the dynamics of systemic failure and recovery.

2.3 From Resilience to Regeneration

Resilience is commonly understood as the ability to withstand shocks or maintain operations under stress, but in complex, multi-domain environments this definition is insufficient. True strategic resilience demands structural adaptation when foundational assumptions collapse. Most existing military simulations focus on optimizing performance under expected conditions or addressing localized disruptions, leaving systemic breakdown and institutional regeneration largely unexplored. Understanding how organizations recover, reorganize, and transform after collapse constitutes a necessary extension of current foresight methodologies.

2.4 Strategic Collapse as a Diagnostic Condition

In complex systems, the breakdown of a strategic posture should not be viewed solely as failure but as a diagnostic condition that reveals underlying structural limits. By intentionally modeling collapse as a controlled stressor, simulation becomes a means to identify the most critical variables to failure, to locate potential recovery levers, and to estimate the timelines required for re-stabilization. Complexity theory holds that genuine adaptive capacity emerges not in equilibrium but at the edge of chaos, where coherence is tested. Simulation therefore functions as a laboratory for adaptive intelligence, enabling observation of how strategy evolves, reorganizes, and regains functionality under extreme conditions.

2.5 Doctrinal Context: Toward Adaptive Doctrine

Contemporary doctrine, particularly within NATO's Allied Command Transformation (ACT), emphasizes anticipatory planning and the integration of simulation into training. Yet, doctrinal practice often treats simulation as an instrumental validation tool rather than a mechanism for structured disruption. Advancing simulation to function as a strategic foresight engine requires its use in questioning assumptions, revealing fragility, and rehearsing adaptation. This perspective aligns with the understanding that strategy is a continuous process of evolution rather than a fixed endpoint. The durability of a strategic posture depends

less on its initial design and more on its capacity to transform when destabilized. Consequently, strategic agility, the ability to adapt, learn, and recover, must be regarded as doctrine itself. The following section operationalizes these conceptual and doctrinal insights through the development of a simulation framework designed to test strategic survivability under induced collapse conditions.

3. Methodology: The Collapse and Recovery Simulation Framework

This study operationalizes strategic survivability through a simulation that reproduces collapse and recovery dynamics under deep uncertainty. The framework moves the use of simulation beyond traditional planning and validation toward structured disruption and recovery rehearsal. The methodological intent is to provide a strategic foresight engine capable of testing assumptions and adaptive capacity under stress. Simulation serves here not only as an analytical instrument but as an epistemological bridge between theory and experimentation. In complex adaptive systems, understanding emerges through iterative modeling that tests structural assumptions against controlled disruption. Following Sterman's and Epstein's view of simulation as a form of "virtual experimentation", this framework treats collapse as an experimental condition through which latent feedback and adaptive mechanisms become observable. Reinforcement learning functions as the computational analogue of adaptive behavior in complex systems, while the simulation environment itself embodies the foresight principle of learning through exploration of multiple plausible futures.

3.1 Simulation Objectives and Design

The objective is to observe how collapse propagates through interdependent components and how adaptive behavior emerges as agents learn from failure. The simulated setting represents key elements of a strategic posture, including decision centers, logistical nodes, and informational actors that interact across cognitive, technological, and narrative dimensions. Reinforcement learning agents adjust policies over time to favor survivability and the reestablishment of coherence rather than short term efficiency. Each run spans a finite horizon with stochastic stress events applied at variable intensity and timing to induce failure and reveal recovery pathways. The framework is instantiated on fictitious, semi fictitious, or realistic baselines, allowing systematic comparison of behaviors under different contextual assumptions. Outputs include trajectories for reward, days survived, and coherence measures that allow evaluation of recovery time, adaptation rate, operational continuity, and strategic coherence.

3.2 Simulation Architecture

The environment is structured as a hybrid system in order to analyze complex situations with high uncertainty where information cannot be fully quantified. System dynamics captures macroscopic and nonlinear behavior through stocks, flows, and feedback loops that represent accumulated quantities such as material resources, legitimacy, and situational awareness, as well as delays and interactions among subsystems. Agent based modeling represents heterogeneous actors operating with bounded rationality. Each agent perceives state variables, often with delay or noise, and applies decision rules based on information known to be available rather than the true state of the system. The combination of qualitative and quantitative elements yields a model that is consistent with the conceptual foundations and robust enough to expose structural effects. The prototype is implemented in Python using the Mesa framework for agent-based modeling, integrated with Stable-Baselines3 for reinforcement-learning components. System-dynamics feedback loops are represented through parameterized equations following Sterman's (2000) formalism. This hybrid implementation enables the simultaneous representation of structural feedback and adaptive decision-making.

3.3 Induced Collapse and Recovery Mechanisms

Collapse is intentionally provoked through stressors introduced with variable intensity and timing. When stress scenarios are needed to diversify conditions, they can be generated algorithmically to ensure coverage of both common and rare events. The simulation employs three families of disruptions that function as strategic breakpoints. Cognitive shocks degrade coordination and decision quality by simulating internal failures such as misaligned mental models, ambiguous orders, communication breakdowns, and time pressured bottlenecks. Technological disruptions remove or neutralize key capabilities, for example the denial of intelligence, surveillance, and reconnaissance or command and

control infrastructure, or the exploitation of vulnerabilities in autonomy. Narrative erosion reduces perceived legitimacy and weakens shared intent through misinformation that reframes strategic aims and undermines public or allied trust. Once a disruption produces structural failure, agents update their policies through an adaptive learning cycle based on reinforcement learning. The state, action, reward, and update sequence enables agents to discover new courses of action that reestablish function and pursue mission coherence despite obstacles.

3.4 Evaluation and Experimentation

System behavior is evaluated with measures aligned to strategic survivability and tied to observable outputs. Recovery time is the number of steps required for the system to return from a reward/stability trough to its preshock rolling band. Learning progress is assessed from post-shock improvements in cumulative rewards and the shortening of re-stabilization intervals. Operational continuity is the proportion of the episode in which stability remains above a set threshold (or, equivalently, the minimum reward maintained during disturbance), supporting sensitivity analysis of components that fail first and most often. Strategic coherence is gauged by the persistence of coordination among cooperative agents, operationalized as rolling correlations and gaps in their rewards; deviations serve as early indicators of systemic strain. Experiments use multiple runs under ultra-stochastic conditions to identify recurring collapse–recovery patterns and to separate signal from noise.

3.5 Validation and Limitations

Validation at this stage focuses on internal coherence and face validity. The model is assessed for logical consistency of feedback, plausible collapse–recovery sequences, and stable behavior under repeated runs with identical settings. External or empirical calibration has not yet been undertaken. The prototype remains a conceptual demonstrator. Parameters are illustrative rather than data-derived, state variables are parsimonious, and several mechanisms are represented at aggregate level. These constraints limit quantitative inference but are suitable for exploring structural effects and adaptive patterns. With the framework specified, Section 4 reports the observed collapse–recovery dynamics and discusses their implications for strategic survivability.

4. Results and Discussion: Observing Collapse–Recovery Dynamics

The multi-agent reinforcement learning (MARL) simulation was executed under ultra-stochastic conditions to observe how collapse and recovery unfold in a complex, adaptive system exposed to persistent uncertainty. Each episode represented an operational campaign of 180 simulated days, during which agents interacted, learned, and adapted in response to continuous cognitive, technological, and narrative stressors. The resulting data reveal distinct phases of collapse, stabilization, and regeneration that collectively illustrate the system's emerging strategic survivability.

4.1 Overview of Simulation Runs

Across more than two thousand episodes, the model generated time-series data for cumulative reward, mission duration, and systemic stability variables. In early training, episodes ended prematurely, often within 60 to 100 days, demonstrating the fragility of uncoordinated policies under compounded stress. As learning progressed, the agents gradually discovered cooperative behaviors that extended survival and reduced the amplitude of systemic oscillations. The probability distribution of "days survived" shifted steadily toward the upper boundary of 180 days, marking the transition from reactive endurance to adaptive recovery. These trends indicate that, over time, the simulated actors internalized strategies capable of maintaining functionality despite repeated perturbations.

4.2 Collapse Dynamics

The initial training phase was characterized by frequent and abrupt collapses driven by overlapping disruptions in coordination, communication, and perception. Cognitive stressors—manifested as decision noise, delayed perception, and misaligned objectives among agents—produced the earliest breakdowns, rapidly cascading across the network. These failures disrupted shared situational awareness, leading to reactive decision cycles and loss of coherence in joint actions. Functional degradation followed as system stability declined below operational thresholds, reflecting the inability of the Blue Government and Humanitarian agents to sustain effective cooperation under compounding volatility. Simultaneously, narrative distortion, simulated through stochastic interference

in the Cognitive Media agent's information state, accelerated systemic distrust and further reduced coordination. Feedback analysis revealed that these collapse sequences were non-linear: minor misperceptions or reward conflicts amplified through feedback loops, producing disproportionate systemic degradation. The cumulative effect demonstrates that in complex strategic environments, breakdown originates not from external shocks alone but from endogenous interaction failures within the adaptive architecture itself.

4.3 Recovery and Adaptation Patterns

As reinforcement learning progressed, the agents gradually improved their capacity to anticipate and absorb environmental volatility. Recovery occurred cyclically, with each collapse followed by partial stabilization and a gradual re-emergence of coordinated behavior. Reinforcement-learning updates allowed agents to refine their policies through exploration, reducing overreaction and improving alignment among decision nodes. Emergent adaptation was visible in the form of improved information handling by the Cognitive Media agent, smoother coordination between Blue Government and Humanitarian actors, and faster re-stabilization after cognitive or narrative disruptions. The steady increase in episode duration confirmed that resilience arose not from avoiding collapse but from learning to reconstruct systemic coherence more efficiently after each breakdown.

4.4 Reward Evolution and Systemic Performance

The evolution of cumulative reward provides quantitative evidence of adaptive learning. During the first 500 episodes, reward trajectories fluctuated around zero, reflecting exploratory behaviour and unstable coordination. As training progressed, the cooperative agents—Blue Government, Humanitarian, and Cognitive Media—gradually achieved higher and more stable returns, indicating convergence toward policies that preserved systemic stability. The Red Adversary, designed with its own independent objective rather than a mirrored loss function, maintained oscillating and often lower rewards, reflecting its destabilizing influence within the environment. This divergence highlights the emergence of adaptive balance: while Red introduced persistent disruption, the Blue-side agents collectively learned to counter volatility and sustain performance. The overall upward trajectory of the rolling mean reward demonstrates that the system not only tolerated uncertainty but benefited from it, reflecting antifragile adaptation in Taleb's sense, where exposure to variability strengthens long-term effectiveness.

4.5 Behavioral Insights from the Final Run

The dynamics of the final trained run illustrate how adaptive behavior unfolds across multiple timescales. During stress peaks, global stability and cooperative rewards declined but recovered through iterative policy updates and improved coordination. These oscillations indicate that the system did not settle into equilibrium; rather, it maintained bounded instability consistent with complex adaptive behavior at the "edge of chaos." The extension of episode duration to near-maximum values demonstrates that the agents collectively evolved policies emphasizing structural flexibility and mutual reinforcement. Final-run reward distributions show differentiated specialization among agents. The Blue Government acted as the integrator of systemic stability and efficiency, the Humanitarian agent sustained welfare and continuity, and the Cognitive Media actor improved information balance and coherence. The Red Adversary continued to introduce perturbations but no longer destabilized the system entirely, revealing a learned equilibrium between disruption and adaptation. The emergent interplay among agents produced a composite resilience that could not have arisen from any actor operating in isolation.

4.6 Interpretation and Implications

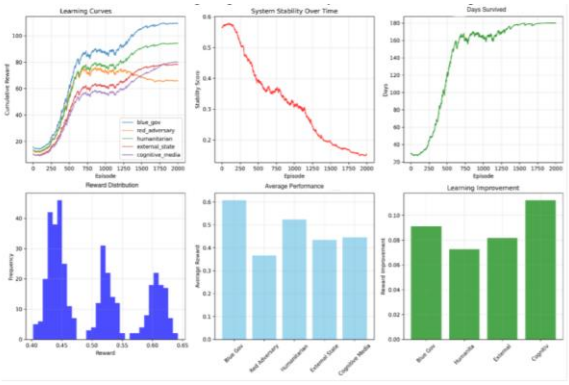
These results substantiate the methodological proposition advanced in Section 3: collapse–recovery simulation serves as a foresight instrument for assessing adaptive capacity. The learning trajectories and survivability patterns indicate that robustness arises not from resistance to disturbance but from the ability to reorganize and learn under pressure. In this sense, the approach shifts simulation from plan validation to structured discovery, exposing hidden dependencies and revealing how policy architectures fracture and reconstitute under uncertainty.

The observed oscillatory cycles of breakdown and renewal are consistent with complex-systems theory. They echo Holling's adaptive cycle, in which systems progress through disturbance and reorganization, and

align with Taleb's notion of antifragile adaptation, where exposure to variability enhances long-term performance. Reinforcement learning provides the computational mechanism for this principle, converting environmental shocks into opportunities for policy revision and improved coordination. The framework thus operationalizes strategic survivability as a measurable property: not the absence of collapse, but the capacity to recover faster and more coherently after each disruption.

4.7 Visual Synthesis of Adaptive Behavior

Figure 4.7.1 summarizes the core behavioral patterns observed throughout the training process. The learning curves in the upper-left panel demonstrate the gradual improvement in cumulative rewards across cooperative agents, reflecting the emergence of adaptive coordination and mutual reinforcement. The central panel shows the evolution of the global stability index, which declines gradually over time as the environment remains ultra-stochastic. This trend does not signify failure but indicates that the agents continue to operate effectively within increasing volatility, maintaining coherence despite the erosion of nominal stability. The right panel captures the steady extension of episode duration, with survival time converging near the maximum horizon, confirming that resilience is achieved through repeated recovery rather than static equilibrium



Simulation outcomes illustrating adaptive learning and systemic regeneration through repeated collapse–recovery Cycles

The lower panels illustrate additional properties of the learning dynamics. The reward distribution shows alternating phases of high and low performance, typical of systems operating at the edge of chaos. Average performance comparisons reveal differentiated roles among agents, with Blue Government and Humanitarian actors ensuring structural stability and Cognitive Media driving information resilience. The relative improvement measure emphasizes that the most significant learning gains occurred in the cognitive domain, validating the central hypothesis that narrative and informational adaptation are decisive levers of systemic regeneration. Together, these indicators visually reinforce the central argument of this study: that exposure to volatility enhances long-term survivability by stimulating adaptive learning across interacting agents.

5. Conclusions and Future Work

This study presented a simulation-based approach to strategic foresight that treats collapse as a diagnostic condition and recovery as an adaptive process. The hybrid multi-agent reinforcement-learning model reproduced breakdown–regeneration cycles under deep uncertainty and showed that learning emerges through exposure to volatility: survival horizons extended, cumulative rewards improved, and systemic coherence was repeatedly re-established after disruption. Findings indicate that strategic survivability derives less from rigid planning than from continuous recalibration of structure and intent. Collapse operates as structured discovery, revealing hidden dependencies and fragilities while enabling policy revision. The observed oscillations are consistent with complex-systems perspectives in which systems strengthen near the edge of instability, where feedback and learning are most active.

The approach reframes simulation from plan validation to structured discovery. Rather than confirming a preferred course of action, the framework tests adaptability, surfaces vulnerabilities in advance of crisis, and supports the cultivation of agility, decision diversity, and institutional self-awareness in volatile environments. Future work will extend the architecture with richer inter-agent communication and cross-domain couplings, develop a staged validation pathway using expert elicitation and comparative exercises, and explore integration with wargaming and decision-support settings to employ collapse–recovery rehearsal in practice. An open invitation is extended for collaboration to refine shock modeling, co-develop validation protocols, and embed the framework in operational analysis; joint development will accelerate calibration and broaden practical applicability.

In sum, the evidence supports a shift in strategic foresight from predicting futures to rehearsing adaptation, positioning survivability as the capacity to recover faster and more coherently after each disruption.

References

[1] P. Cilliers, Complexity and Postmodernism: Understanding Complex Systems. London, U.K.: Routledge, 1998.

[2] N. N. Taleb, Antifragile: Things That Gain from Disorder. New York, NY, USA: Random House, 2012.

[3] N. N. Taleb, The Black Swan: The Impact of the Highly Improbable. New York, NY, USA: Random House, 2007.

[4] C. S. Holling, "Resilience and stability of ecological systems," Annual Review of Ecology and Systematics, vol. 4, pp. 1–23, 1973.

[5] W. Walker, R. Lempert, and J. Kwakkel, "Deep Uncertainty," in Encyclopedia of Operations Research and Management Science, S. I. Gass and M. C. Fu, Eds. Boston, MA, USA: Springer, 2013.

[6] J. D. Sterman, Business Dynamics: Systems Thinking and Modeling for a Complex World. New York, NY, USA: McGraw-Hill, 2000.

[7] P. W. F. van Notten, J. Rotmans, M. B. A. van Asselt, and D. S. Rothman, "An updated scenario typology," Futures, vol. 35, no. 5, pp. 423–443, 2003.

[8] K. E. Cuhls, "Horizon Scanning in Foresight – Why Horizon Scanning is only a part of the game," Futures & Foresight Science, vol. 2, no. 1, 2020.

[9] Z. Lou, Z. Chen, M. Sim, J. Xie, and P. Xiong, "A Unified Framework for Robust Decision Models Integrating Robust Optimization and Robust Satisficing Paradigms," presented at 25IC Conf., 2023.

[10] W. Suo and L. Wang, "A Novel Framework Integrating Generative Artificial Intelligence (GenAI) with System Dynamics for Critical Infrastructure Resilience," presented at 25IC Conf., 2023.

[11] N. Oreskes, K. Shrader-Frechette, and K. Belitz, "Verification, validation, and confirmation of numerical models in the earth sciences," Science, vol. 263, no. 5147, pp. 641–646, Feb. 1994.

[12] The Royal College of Defence Studies (RCDS), Getting Strategy Right (Enough). London, U.K.: UK Ministry of Defence, 2017.

[13] NATO, Allied Joint Publication-5 (AJP-5), Edition A Version 2: Allied Joint Doctrine for the Planning of Operations. Brussels, Belgium: NATO Standardization Office, May 2019.

[14] NATO, Comprehensive Operations Planning Directive (COPD), Interim Version 2.0. Mons, Belgium: Allied Command Operations, Oct. 2013.

[15] T. Deibel, Foreign Affairs Strategy: Logic for American Statecraft. Cambridge, U.K.: Cambridge University Press, 2010

Talk About Us

Italian Ministry of Defense

www.difesa.it

NATO Allied Command Transformation

www.act.nato.int

NATO Science and Technology Organization

www.sto.nato.int

ST Engineering Antycip

www.steantycip.com

Calian Group

www.calian.com

CYBERNAUA

www.cybernaua.it

MathWorks

it.mathworks.com

Report Difesa

www.reportdifesa.it

Teroresi SpA

www.teoresigroup.com

NATO M&S CoE Annual Review 2025

- CASTLE (CBRN Activities Simulation Total Layer Environment) Project
- Mathematical Modelling of Cyber Warfare: A Quantitative Framework for Pre-Conflict Analysis
- CACTUS: Enhancing Strategic Decision-Making for Urban CBRNe and TIC/TIM Events through Simulation and AI
- Integrating Cognitive Warfare into Multi-Domain Wargaming: The CW-BRAINWARE Approach based on Strategic Engineering
- Intelligent Agents in Wargaming Simulations
- Enhancing the Deployment of Multidomain Defense Systems through AI-Driven Automation and Integrated Engineering Processes to gain Strategic Advantage
- Simulating the Narrative Battlespace: Integrating Synthetic Social Media, AI Scenario Generation, and Autonomous Adjudication into Wargaming
- Strategic Foresight under Uncertainty: Simulating the Collapse and Regeneration of Strategic Postures

ISBN 979-12-985129-2-4

NATO OTAN

MODELLING & SIMULATION
HONOS ALIT ARTES
CENTRE OF EXCELLENCE

www.mscoe.org