The NATO Modelling and Simulation Centre of Excellence team thanks you for participating the 2020 Forum.

This work is copyrighted. All inquiries should be made to: The Editor, NATO Modelling and Simulation Centre of Excellence (NATO M&S CoE), info@mscoe.org.

Disclaimer

The CA²X² Forum 2020 Papers Collection is a product of the NATO M&S CoE. It is delivered to give an outline of the CA²X² Forum 2020 and resulting productions. It does not represents the opinions or policies of NATO and reflects independent analysis, opinion, and the position of his authors.

Release

This document is approved for public release. Portions of the document may be quoted or reproduced without permission provided a standard source credit is included.

Published and distributed by

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

Table of Contents

Introduction……..……7

THEME 1: Modelling and Simulation Standards and Technologies……..8

NATO Federated Mission Networking Standards for CAX……8

The Trend Toward Common Architectures……...20

Multi-Viewpoint Conceptual Modeling *In Support of Simulation Interoperability Readiness Levels (SIRLs)* ……26

Discovering and Leveraging Emerging Technologies for Application in M&S……36

THEME 2: Training and Education……..43

Panel Session Applying and Utilizing New Technologies in Training Challenges and Limitations from a Trainer's Perspective……...……43

Simulation Course – Action Research Approach..……..……45

Virtual Battlespace 4 (VBS4) – Cloud Enabled, High Fidelity and Whole Earth Simulation……..……53

THEME 3: Immersive Technology and Extended Reality (XR) in Support of Operations..……..60

xR in the Operation, Operate in xR - The Tech Intensity of the Future Mission……..……60

Virtual Mission Rehearsal for Special Operations Forces: A Sweden – U.S. Collaborative Effort…..……...…74

Aviator Training Next – Virtual Reality Pilot Program……77

Using Virtual Reality for Collaborative Immersive Operations and Planning……..……..79

THEME 4: Artificial Intelligence..……..……81

A Brief Introduction to Concrete Algorithmic Game Theory Applications..……..……81

Introduction to Artificial Intelligence……...83

Trustable and Ethical Artificial Agents: a Toolkit of Useful Techniques, and Human AI Teaming……86

Modern Intelligent Systems between Explainability and Ethics: Frontiers from NLP Learning Systems...……94

THEME 5: Wargaming...……..108

Introduction to Wargaming: Preparing the Next Generation of U.S. Army Professional Wargamers....……108

Wargaming Cyber: Tactical Hack……..……112

THEME 6: Decision Support and Analysis..……120

Data Farming Services (DFS) for Analysis and Simulation Based Decision Support....…..……120

Collaborative Exercise Planning in an Isolated Environment..……...129

Genomics Enhances the Predictive Power of Machine Learning for Predicting Susceptibility to Heat Illness in Soldiers During Exercise Precise Response 2019....133

Radical Transparency....137

The NATO Modelling & Simulation Specialist Team MSG 189....142

Session 1: Decision Support and Analysis....143

Session 2: Emerging Simulation Technologies for Decision Making Support in the 21st Century....144

Future Command Decision Training Support....145

Talk About CAX Forum....154

Talk About Us....156

The CA²X² Forum 2020

Sponsor Recognition

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.

Michele TURI
Col. ITA Army
NATO M&S CoE DIRECTOR

This book contains the proceedings of NATO M&S CoE's Computer Aided Analysis, Exercise, Experimentation annual conference held, for the first time, online from 22 – 24 September 2020.

The principal theme for the conference was:

'Modelling and Simulation Enabling NATO and Nations'

Through a 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 all of the contributors. Please use these articles as inspiration for further collaboration and contributions to these important themes.

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

CA²X² FORUM 2020

Computer Aided Analysis, Exercise, Experimentation

Modelling and Simulation Enabling NATO and Nations

Opening Address by Colonel Michele Turi
Director, NATO Modelling and Simulation Centre of Excellence.

Dear Generals, Ladies, Gentlemen, Stakeholders and NATO M&S COE friends,

in this collection we present the CA2X2 2020 forum review to publish research, papers and studies presented during our latest international event.

Our pride is to be able to present, even in a moment of pandemic crisis, your and our work to underline the importance of the M&S sector which, as an atypical and transversal doctrine, is able to support many sectors: from training, to education, from decision-making support to research.

Simulation is in an important phase of exponential growth, also driven by the strong technological evolution in communications, information technology and AI.

NATO, which is improving its approach to M&S year by year, sees it as the key technology to implement all other emerging and disruptive technologies to maintain its military and technological advantage over its competitors.

The CA2X2 2020 forum has been an important event because it was carried out only in digital mode, providing an important basis for experimentation also for new distance learning techniques and accelerating the development of skills and adaptation to the new "fluid multidomain environment".

The results of the event are clear: more than 630 attendees, 22 companies, 41 nations represented and more than 60 different papers and research efforts selected and presented in three days of meetings. These important results are an important basis for the next 2021 event which will seek to maintain a high level of challenge and improve it by guaranteeing a digital meeting place for the entire international M&S community of interest.

We will continue to support NATO and nations in their transformation efforts by providing expertise in all aspects of M&S and also to ensure good scientific production and "food for brain" regarding M&S applications in the world of Defense. I wish you good reading and I hope it will inspire you for your work.

Best regards,

P(h)D Col. Michele TURI

NATO M&S CoE Director

THEME 1: Modelling and Simulation Standards and Technologies

NATO Federated Mission Networking Standards for CAX

J. Mark Pullen
GMU C4I & Cyber Center

Claudio Zamponi, Fabio Corona
NATO Modelling and Simulation Centre of Excellence

Abstract

Computer Assisted Exercises (CAX) are a well-accepted capability for achieving collective training of today's national and coalition military in preparation for operations. However, there is a need to harmonize CAX technology with NATO's new approach to operational collaboration, command, and control: Federated Mission Networking (FMN).

Two decades ago in Afghanistan, NATO's International Security Assistance Force was hampered in operations until the Afghan Mission Network (AMN) was assembled to support collaboration and coordination of forces. Today, Allied Command Transformation is preparing for a future where any NATO coalition force has a network far superior to AMN on Day Zero of coalition operations. Toward this end, the FMN project is assembling a framework of NATO and commercial standards with the expectation that the 30 member nations will configure their networking capabilities to interoperate over the FMN standards. FMN standards thus will provide the basis for "train as you fight" communications as well as supporting distributed simulation for that training.

The NATO Modelling and Simulation Group (MSG) Technical Activity 145 and SISO Product Development Group for the C2-Simulation Interoperation (C2SIM) standard have been working together to standardize and operationalize a new capability, which has been described in previous CAX Fora by the author. The team that assembled C2SIM standards now finds a new challenge: assembling and justifying a collection of standards for modeling and simulation (M&S) that suit FMN needs, with C2SIM an obvious cornerstone of that collection. This paper addresses from a CAX viewpoint the technical issues and process whereby standards for networked computer simulation within the FMN are nominated. The paper introduces the FMN concept, followed by a discussion of the role of networking in coalition exercises, and finishes with a review of likely standards for networked military simulation that will be included.

1 Introduction

1.1 Background

Computer Assisted Exercises (CAX) are a well-accepted capability for achieving collective training of today's national and coalition military in preparation for operations. However, there is a need to harmonize CAX technology with NATO's new approach to operational collaboration, command, and control: Federated Mission Networking (FMN). This paper addresses the reasons for NATO development of FMN and how the standards involved impact fielding and execution of CAX. For a more detailed exposition of the concepts and standards behind FMN, see [1].

Before 1995, interoperability in NATO was based on the deployment of Liaison Officers (LNOs) who were attached to a flanking formation on right of the sending organization and one to each subordinate command. They would have radio communications with their parent headquarters but any other form of interoperability posed a challenge. After 1995, which saw NATO deployment of multinational forces, there were issues in balancing the command authority requirements of a force commander versus the reluctance by nations to relinquish national command and control (C2) of forces to a foreign commander. Coupled with this were difficulties that affected the effectiveness of multinational logistics and Communications and Information Systems (CIS) support, where national laws and financial regulations were seen as outweighing the needs of the commander of a multi-national force. As a result, nations sought to embrace new digital communication technologies to securely enhance the decision-making process. Examples of such initiatives were the UK's Network Enabled Capability (NEC) and the USA's Network-Centric Warfare (NCW) [2]. The emergence of these standalone networked CIS systems from 1995 created some added unforeseen problems because of different interpretations that related to security regulations, standards, procurement strategies, industrial self-interest and operating practices. As a result, without major effort and strong leadership the ability to interconnect these networks was challenging. In the C2 domain, this led to deployment of "swivel-chair" interfaces (a situation not dissimilar to that seen even today in simulation interoperability) where an operator used USB sticks and CD-ROMs to bypass the airgaps built into national systems. NATO concluded that a NATO Network-Enabled Capability (NNEC), building on the concepts of NEC and NNEC, was needed [3].

This in turn led to security breaches and other operational problems that sadly exhibited themselves in Afghanistan during the NATO led operations from December 2001 when the International Security Assistance Force (ISAF) was established. Due to the issues faced by forces from NATO and coalition partners in early deployments, the Afghan Mission Network (AMN) was conceived and successfully developed, although it was not without some challenges in its implementation. AMN used some of the basic tenants from NNEC and was designated as the primary Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) Network for the coalition. Thus, it became a "weapon platform" in its own right and most communities of interest moved their C4ISR tools from national networks onto the common information sharing platform. Its topology was not directly related to the structure of the chain of command, enabling efficiencies in CIS resources and sharing of services in different locations.

1.2 Federated Mission Networking

In the aftermath of a number of NATO's operations, and predominantly because of their deployment to Afghanistan, a key lesson identified was the need to have its command and control processes and supporting technology interoperable from the start of a mission, in what is termed "Day Zero Interoperability." In order to achieve this, it was decided by NATO's Military Committee (MC) in 2012 that NATO had to develop Federated Mission Network (FMN) as a common capability among NATO nations [4]. The FMN concept paper was endorsed by nations, an Implementation Plan was drawn up [5], and the North Atlantic Council (NAC) endorsed Version 4 of the FMN Implementation Plan in 2015. In 2016 at the NATO Warsaw Summit, it was stated by NATO leaders that "Interoperability of our armed forces is fundamental to our success and an important added value of our Alliance" [6]. The starting point for FMN was based on the lessons identified from the successful implementation of the AMN. FMN itself cannot be developed in one large acquisition program as was envisaged by concepts like the NNEC, NEC, and NCW. It will evolve over time though "spiral development" with requirements for each spiral established by military needs (see section 3 below). Modeling and Simulation (M&S) was not considered in the early spirals because the priority there was to establish a limited set of functions that could be achieved rapidly. It was however an aspiration from the inception of FMN that M&S, although mainly recognized for its role in supporting training, also would need to be incorporated to support future decision making through Course of Action (COA) analysis, Wargaming and Mission Rehearsal.

The mission of FMN is: *Enhanced Operational Readiness & Effectiveness Today and in the Future* and its vision is: *Day Zero Interoperable Forces*. Day Zero capability refers to the minimum capability required to support the needs of the Commander during the pre-deployment and initial deployment phases of an operation, and to support rapid, smooth, and efficient transition from pre-deployment to initial operations. As articulated by NATO Allied Command Transformation (ACT), the FMN vision has two components: 1) *Operate Together: Exploit our Strategic Advantage* and 2) *Adapt Together: Effectively Transform Capabilities to Maintain our Edge* [7]. The first relies on having FMN Ready Forces before the start of a mission. This means that national contributions to a NATO Response Force (NRF) must be declared as FMN compliant, which is achieved through testing and validation activities. The second component is tacit recognition that, in an era of constrained resources and a wide range of potential missions, FMN reflects the need for *federation* as the means to achieve economy of scale and maximum reuse while achieving the full benefit of information sharing. The word "Network" has subsequently been replaced by "Networking" to reflect the fact that FMN is based on an interoperable capability of each nation and is not deployed as a single network under unified management. The FMN capability is composed of a number of elements that collectively comprise the ability to provide mission networking in a federated environment. The primary goal of the FMN capability is to support C2 and decision-making in future operations through improved information-sharing. The approach is distinctive in that it provides the ways and identifies the means to deliver better information sharing. The implementation of this capability is intended to deliver a toolset of processes, organizations, training, technology, and standards provided, in a coordinated approach, by NATO, NATO Nations, and non-NATO nations cooperating together.

FMN Ready Forces are those forces assigned to the NRF, who six months prior to taking on their role within the NRF are interoperable in all elements agreed that will form part of the Spiral Specification. The diagram at Figure 1 illustrates the concept.

Figure 1 – FMN Ready Force Requirements [7]

2 Modeling and Simulation in the FMN

2.1 Characteristics of M&S as Relates to FMN

M&S have been defined in NATOTerm [9] in the following way:

• *Model*: A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process.

• *Simulation*: The execution of a system model over time.

Simulation is categorized as either (1) live, where real people use real systems; (2) virtual, where real people use simulated systems, e.g. a part task driving simulator; or (3) constructive, where the human aspects, behaviors and decision making, are simulated by agent models, scripted logic or human intervention. All three types of simulation have applicability in FMN, particularly the constructive.

A typical simulation has representations of the operational environment, the actors (active objects), and their behaviors. These are supported by an underpinning simulation engine which manages such things as communication among systems, scheduling, and interaction with any operator by means of user interfaces and graphical environments. Simulations execute scenarios, which encapsulate the required operational environment – locations, units, required actions, etc. It is usual to include a logging and replay capability to assist in after action review and analysis processes. Simulation time management is often real-time but there are situations where faster-than-real-time or slower-than-real-time simulation execution is required, for example to quickly assess alternative course of action or to understand quickly evolving situations.

The operational environment can include natural and man-made topography and bathymetry, time-varying weather and oceanographic effects, and electronic environment. Physical models of the units, individuals and equipment, which in turn are represented by sub-models for their components, interact with other physical and environmental models. Behaviors include individual, group, equipment, doctrinal and population and may be 'natural', tasked or requested, background or reactive. Players may be assigned to sides and teams, given allegiances and placed in organizational hierarchies for operational and communication purposes.

Recently there has been a focus on greater composability of simulations and a move towards modeling much more complex environments such as the so-called mega-cities, cyber environments, and a greater number of non-military actors and effects such as social mediainformation (or misinformation) networks. These developments often include technical developments derived from the computer games industry but, when used in support of military ends, are required to comply with approved standards. In line with this, there is also a move towards the use of cloud-based simulation capabilities such as NATO's M&S as a Service (MSaaS) [10].

A number of processes and information exchange standards that have been developed to support the development, integration and execution of M&S systems are introduced later in this paper.

2.2 Potential Application of M&S in FMN

M&S has been used to support a number of military needs such as:

- Individual and collective training;
- Mission rehearsal;
- Operational planning;
- Concept development and experimentation; and
- Acquisition programs, e.g. to support system evaluation.

All these use cases entail M&S interacting with human operators via operational C2 applications that provide means for displaying reports, communicating with other personnel, and preparing plans, orders, tasks and requests. They all can benefit from FMN connectivity and all can support the MDMP and aspects of military operations. It is usual to run a simulation on its own network enclave rather than a shared experimental or operational network. There are sound technical reasons for this, particularly as simulation networks tend to be high volume users and require low latency. The simulation traffic does not need to know about or share information with many of the other applications operating in an FMN environment, e.g. VOIP, operational C2 messaging, email services and shared document repositories. Where it does need to touch FMN is through the command and control applications. For this reason, C2-simulation interoperation is needed as a bridge to provide a compliant means of connecting M&S systems to FMN systems and services [13].

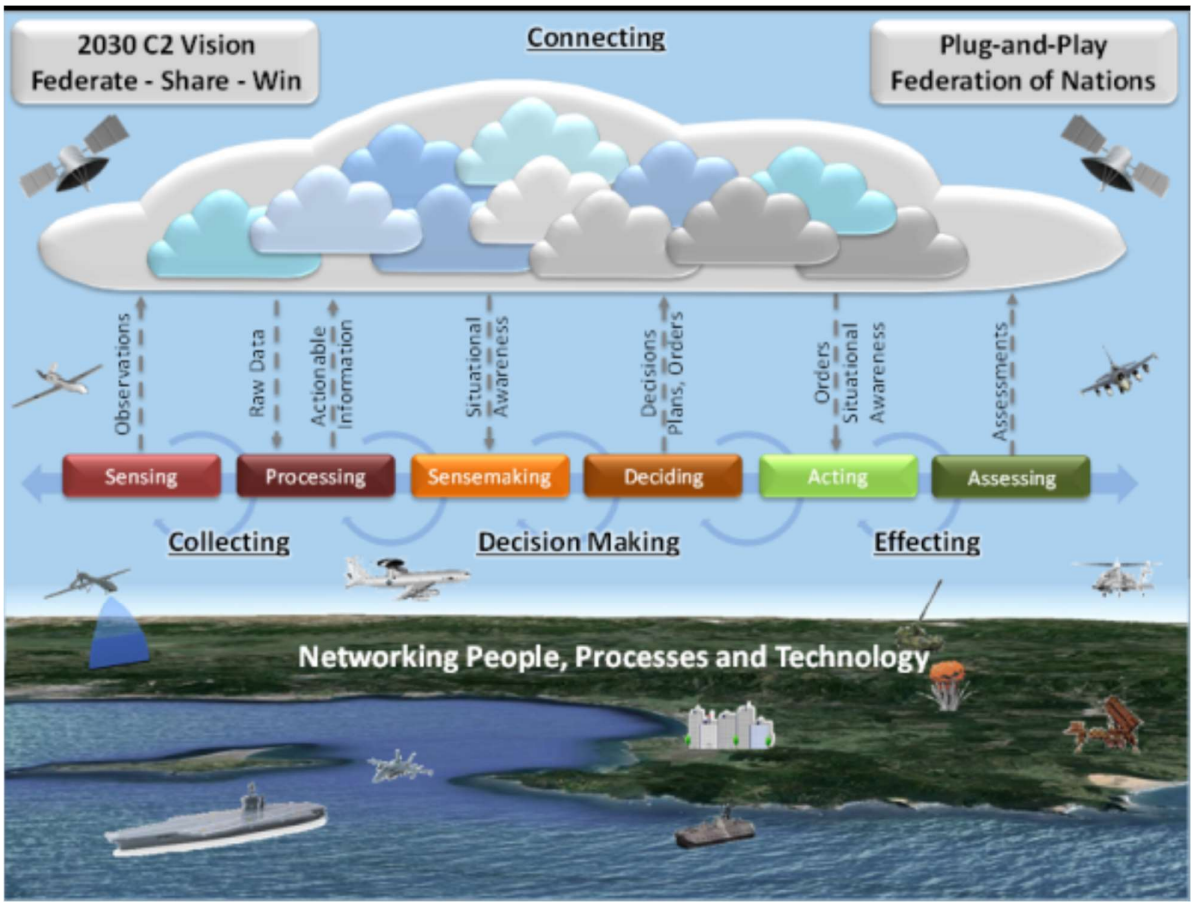


Figure 2 – NATO's 2030 C2 Vision (NATO C2 COE)

In an FMN environment, there are a number of equivalent use cases where M&S could be used to advantage. In a *training environment* the training audience, is presented with an operational situation to work, using their regular operational C2 equipment. A typical set of operational processes, as outlined in NATO's C2 Vision for 2030 [11], covers information collection, decision making and effecting. The information gathering of a live system can be represented using M&S: simulated sensor feeds, blue-force tracker data, reports from battle-space entities and other emulated messages. The information bearers can be represented using a so-called digital range which can represent ideal communications or degradation due to factors such as insufficient bandwidth, jamming or cyber-attack. If the training audience sits at the decision-making part of the process, then their decisions will be based on the information received from the simulated information gathering components. When the training audience has made an assessment and reached a course of action then they can task a simulation to execute the plan thus completing the cycle [13]. As part of a training course, a number of prepared scenarios may be enacted, not only simplifying the roles of the trainers but also helping compare the performance of the different training audiences.

Mission rehearsal is special training given for a specific mission. Here the simulation will replicate as faithfully as possible the proposed mission environment so special care is needed to prepare everything.

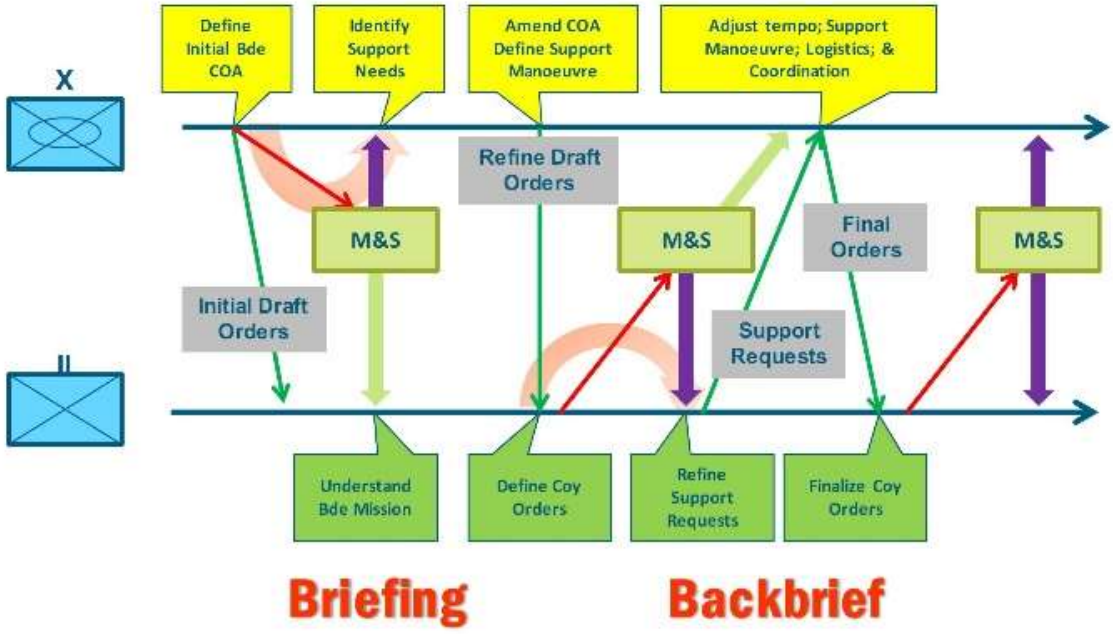


Figure 3 - M&S in Operational Planning

Operational planning can be undertaken in a collaborative way using multiple simulations running in a faster-than-real-time mode (Figure 3, where Coy abbreviates Company and Bde abbreviates Brigade). This allows alternative courses of action to be simulated by groups of planners working at, for example, different echelons or different specialties. Outside the FMN context, this has been demonstrated by NATO Modeling & Simulation Group (NMSG) 085 [13] in an experiment based on the NATO Comprehensive Operational Planning Directive (COPD) [14].

Concept development uses M&S to represent evolving processes and behaviors in a safe environment. This may well include implementing lessons identified in an operational environment.

Acquisition programs use M&S to simulate equipment before it has been built or put into production to help identify potential problems or to test operational capabilities. M&S can be used to help choose between rival suppliers' solutions by comparing them with a common set of simulated conditions, including ones which would be expensive, difficult or unsafe to test in a real-world situation.

3 FMN Spiral process

3.1 Why Spirals?

The concept of FMN development laid out in [8] follows in a general way the spiral development approach that has become very popular in the commercial sector, where a short sequence of phases is continually repeated, with active user involvement, coming closer in each cycle to the goals of the development. The concept of developing systems in such "spirals" as described in [15] and [16] has been shown to be more effective at developing systems that meet user needs better, more rapidly, and at lower cost. Proponents contend that these benefits result from users coming to understand their needs better as they help steer the development toward better results, combined with developers achieving better technical results by frequent evaluation and, if necessary, revision of their work. The approach has been described as repeated plan-a-little/build-a-little/test-with-users/rethink-results [15]. Commercial development spirals can be as short as one month in duration.

The Spirals in FMN are much longer than those in commercial development; each one has planned duration of about two years. The phases of each Spiral are described in [5] as Operational and Security Requirements, Proposed Specifications, Final Specifications, Emerging Operational Use, and afterward Preferred Operational Use. The process is standards-based; products of each spiral are characterized as Requirements, Interoperability Architecture, Standards Profile, and Instructions. The approach has been driven by the need to involve 30 NATO nations that are at various levels of technical sophistication and the fact the process necessarily involves government bureaucracies. Like its commercial counterpart, each spiral builds on previous ones incrementally; but the Spirals are overlapped to shorten overall development time. This approach is seen by many as a great improvement over traditional military system development that takes many years and, as a result, often produces results that are outdated by the time the systems are produced. (Please note that here we are addressing here systems that are primarily software based; developing major military hardware platforms necessarily has a different set of characteristics).

3.2 How will Spirals work?

In addition to the Enabling Framework described above, per [8] the FMN will have "Common and Permanent Management." [17] defines a management structure consisting of overall management and support, plus a collection of *working groups* that meet separately and then come together in the FMN track of ACT's "TIDE Sprint" assembly twice yearly [17]:

- Overall management group
- Supporting secretariat staffed by Allied Command Transformation
- Operational coordination working group linking to NATO commands
- Multinational security management working group
- Capability planning working group and syndicates
- Change and implementation working group
- Coalition interoperability assurance and validation working group

Among the above, the colorfully-named *syndicate* is an interesting innovation. According to [19] "syndicates are informal working bodies - often already existing as collaborative undertakings for a specific subject, product or community of interest - focused at providing expert advice and tangible input for one or more FMN working groups." While this concept is not unheard-of in government, when combined with the Spiral concept it provides an interesting extension to the more typically bureaucratic structure of FMN management, allowing for participation of technical laboratory staff, industry experts, and academics. This is consistent with the FMN goal "adapting existing capabilities for quick start." As an example, FMN architectural planning is driven by the concept of a *mission thread*: an operationally driven, technically supported description of the end-to-end set of activities required to execute a mission or mission task.

Whatever way the Spirals are planned, their effectiveness depends on the various national affiliates collaborating to reach and test implementable specifications based on existing NATO and commercial standards. To that end, ACT carries out the annual Coalition Warrior Interoperability Exploration, Experimentation, Examination and Exercise (CWIX) interoperability testing, involving all stages of each ongoing Spiral tested either in person or via secure CFBLNet or semi-secure Internet VPN.

3.3 Current status of FMN spiral development

It should come as no surprise that an effort of such scope occasionally fails to meet its goals. Indeed, in CWIX there

is a viewpoint that it is much better to have some early failures than to establish a Spiral specification that can't work or doesn't meet user needs. Such an early failure is seen as a success of a different sort and is consistent with the nature of the other major assembly sponsored by ACT, the twice-yearly *Think-Tank for Information, Decision and Execution Superiority (TIDE) Sprint*. (The term Sprint is taken from the Agile methodology [16] where each sprint makes rapid, short progress of one to four weeks toward a system goal.)

TIDE Sprint is scheduled for a one-week period twice a year, with location alternating between Europe and North America (but after the pandemic experience in 2020, likely also offering Internet participation). Its purpose is stated as "survey requirements, identify issues and make recommendations" [18] which then make their way into Spirals and from there to CWIX and deployment in NATO nations for Day Zero use.

Development of Spirals in FMN is ongoing; of the four stages Draft, Candidate, Proposed and Final, Spirals 1 through 4 have reached Final stage. Spirals 5 and 6 are Proposed and still the topic of ongoing work, while Spirals 7 and 8 are just beginning. Modeling and Simulation was scheduled to be addressed in Spiral 6 but currently lacks a Syndicate to support that. A proposal is expected in the NATO Science and Technology Organization (STO) Modeling and Simulation Group (MSG) to address this shortcoming by establishing a Specialist Team of MSG experts as an M&S Syndicate. This paper is a first step by its authors to prepare for participation in a possible Spiral 6 M&S Syndicate.

4 M&S in CAX Support

A well-known M&S application field is individual and collective training. NATO defines how a synthetic environment can support an exercise in automating the processes, preventing duplication of work, enhancing the exercise environment and ensuring that the exercise process flows towards the objectives [21]. Computer Assisted Exercise (CAX) is a particular Synthetic Exercise (SYNEX) where a Command Post Exercise (CPX) is executed with the support of computers simulating the operational environment and providing event resolution, in a distributed or not-distributed form or a combination of both. CAX support replaces or helps response cell, High Level Command (HICON) and Low Level Command (LOCON) by providing stochastically computed results to decisions and requests of the training audience (TA).

For this goal, M&S in support of CAX should fulfill a set of requirements. M&S tools have to compute the possible outcomes of commands given to the simulated units and entities with automated representation of friendly and opposing force actions, reducing the requirement for exercise control staff and response cells. The picture of the exercise should be complete temporarily and spatially, so M&S has to simulate the entities and conditions not controlled by the TA or Exercise Control (EXCON). M&S tools must be interoperable with Command and Control (C2) systems to stimulate them and to receive orders saving resources for LOCON and EXCON. M&S tools should operate seamlessly with existing and planned NATO operational CIS through interoperability standards and making simulation transparent to users. For these reasons, FMN is relevant for M&S applications and M&S has to be included in the FMN specifications.

Even if the first impression could be that M&S in CAX support means mainly running military constructive simulation systems, actually CAX support tools are involved in all stages of the exercise process, so they can categorized into four classes: planning and management tools; constructive simulation systems and ancillary tools; interfaces to C2 systems and functional area services; experimentation and analysis tools. Each of these is essential to CAX but also has broader applicability to training, COA analysis and missions rehearsal in the networked operational environment of FMN.

4.1 Planning and management tools

CAX support tools used during the Exercise Planning and Product Development Stage should allow a collaborative development of: all scenario modules with respect to georeferenced data; information and documentation fully in compliance with NATO policy, doctrine, forces' standards, mission essential tasks and interoperability requirements of Functional Services; pre-scripted events, injections and information flows to support achievement of the exercise aim and objectives. Thereafter, M&S tools should assist in the preparation of the scenario related products, sustainment and deployment planning as well as for war-gaming purposes. CAX support tools used during the first phases of an Exercise should support: Force Activation; Deployment; Reception, Staging and Onward Movement (RSOM) and Integration. They should present to the training audience the data and information in the expected formats and levels of granularity that they would expect to see if the situation were real.

4.2 Constructive simulation systems and ancillary tools

The constructive simulators are those M&S tools which play the role of the LOCON providing simulation of friendly and opposing force, their actions, effects of events and conditions not controlled by TA or EXCON. They compute the possible outcomes of commands given to the simulated units and entities and must be interoperable with C2 systems to receive orders and to send back reports. Among these, the Joint Theatre Level Simulation (JTLS) and the Joint Conflict And Tactical Simulation (JCATS) are more used by NATO simulation centres.

4.3 Interfaces to C2 systems and functional area services

CAX support tools must replicate C4I environments during CAXs. So, constructive simulation systems and all the other related software must be transparent to the TA, which should operate as in an operation and command their subordinates by using C4I systems normally available to them. This transparency can be achieved by the mediation tools between the simulations and C4I systems or by a combination of mediation tools and standards developed ad-hoc. The new C2SIM standard described below offers way to do this that will provide plug-compatibility to C2IS.

4.3 Experimentation and analysis tools

M&S tools can support in the observation collection and analysis tasks for comprehensive post-exercise analyses and production of reports on achievement of exercise aim, objectives and requirements. They can assist in reconstructing events and derive lessons for users in real-world operations.

In the following section, a set of M&S related standards for networked military simulation are reported. They can provide such a support in both CAX and real operations for all the phases of planning, execution and analysis.

5. M&S standards for networked military simulation

The Simulation Interoperability Standards Organization (SISO) has been at the forefront in developing simulation standards and processes. These have all been developed by multinational domain specialist teams drawn from across industry, government and academia. Many have been adopted by IEEE and NATO and are in widespread use.

5.1 Command and Control – Simulation Interoperation (C2SIM)

The SISO C2SIM standard [22, 23] has been developed to provide a means of exchanging information between C2 systems and modelling and simulation (M&S) systems, particularly constructive and virtual simulations. C2SIM can also be used to exchange information between different C2 systems and between C2 systems and autonomous systems. In Figure 4, C2SIM is represented by the arrows joining the different types of system.

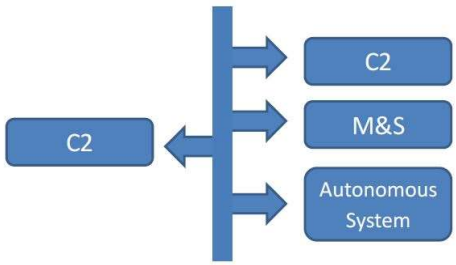


Figure 4: C2SIM Overview

C2SIM was developed by SISO, who are currently ratifying it, in collaboration with NATO STO and will be proposed as a NATO STANAG in 2020 allowing it to be specified in procurement proposals, etc. C2SIM uses a common data model which permits unambiguous data to be exchanged between systems to convey: initialization information (e.g. force structures and dispositions), plans, orders, tasks, requests and reports. C2SIM unifies and replaces two earlier standards: Military Scenario Definition Language (MSDL) and Coalition Battle Management Language (C-BML).

C2SIM is highly pertinent to FMN in that it is aimed at plug-and-play compatibility between command and control information systems (called C2IS, C2 systems or Mission Command systems) and military simulations. Principal uses for this capability are coalition operational training, course of action analysis, and mission rehearsal [22, 23]. The vision of C2SIM is articulated as:

We are working toward a day when the members of a coalition interconnect their networks, command and control (C2) systems, and simulations simply by turning them on and authenticating, in a standards-based network environment.

This vision is for a system of systems where each national component uses its own, familiar C2 system and is represented in the simulated Coalition by a national simulation that accurately depicts its staffing, equipment, and doctrine. C2SIM was developed by the Simulation interoperability Standards Organization (SISO) working in cooperation with international teams from NATO STO. It is expected to be approved as a SISO Standard in June 2020, after which plans are already underway to propose it as a NATO Standardization Agreement (STANAG). It is SISO's second generation standard in C2-simulation interoperation and was subjected to extensive validation in CWIX 2018 and 2019. Implemented under FMN, C2SIM can bring the power of accurate simulation to Mission Command of a multinational coalition. The initial implementation is based around a server, but a multicast-based implementation is possible. C2SIM is completely compatible with all of the standards that are described below.

Mission Threads is a NATO process [24] which has been developed to help develop operational scenarios in a uniform way in accordance with the NATO FMN Implementation Plan (NFIP). For FMN a mission thread is described as:

An operationally driven, technically supported description of the end-to-end interrelated activities required to accomplish the execution of a mission or mission task

and is:

Comprised of the step-by-step description of a mission or activity, the information exchange requirements of the mission or activity and the identification of systems and services that are needed to accomplish it.

Mission threads, as developed for general use for FMN, provide a way forward to developing scenarios for C2SIM M&S applications.

5.2 Modeling and Simulation as a Service (MSaaS)

MSaaS [25] is a NATO approach to provide a means of delivering reusable, composable simulation to the user using a service-based architecture, typically cloud-based, as shown in Figure 5.

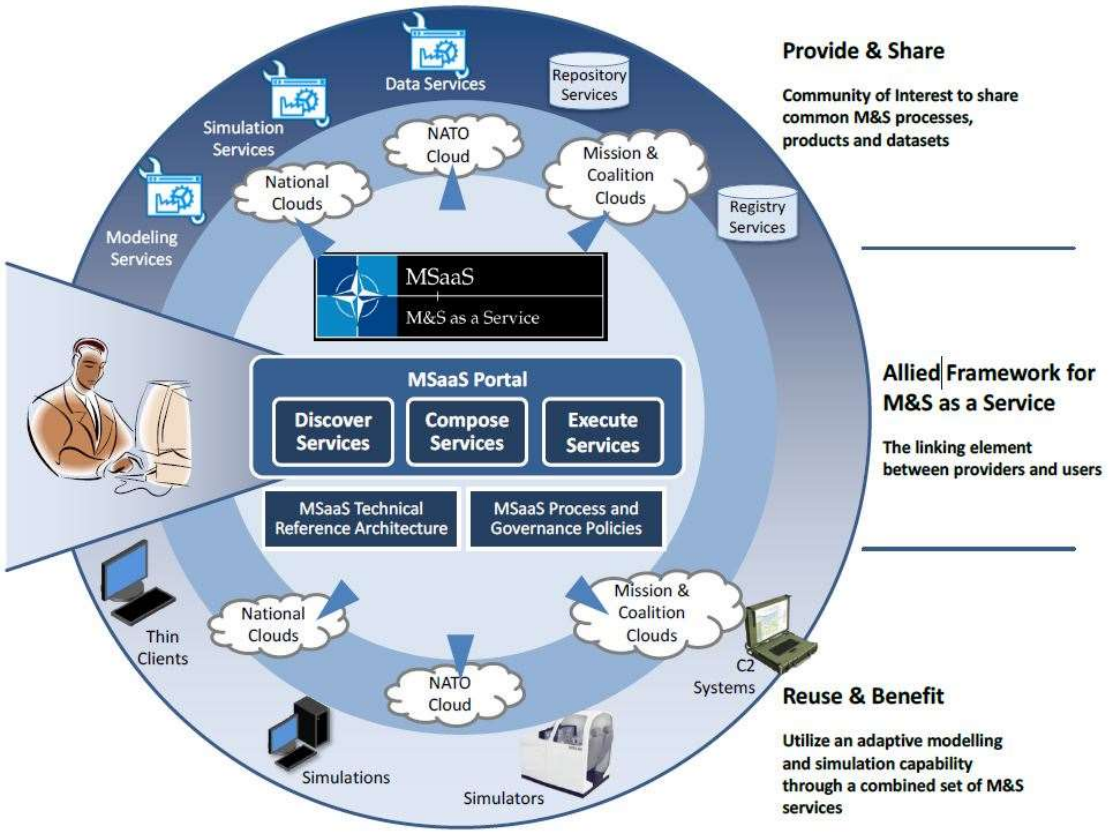


Figure 5: MSaaS Phase 2 [23]

MSaaS helps simulation designers provide better scalable and fault-tolerant simulations. It follows a three-stage process of discovery, composition and deployment/execution. The discovery phase uses searchable simulation repositories to find simulations appropriate for the simulation task in hand. Composition is the building and configuration of the simulation from discovered components. The composability approach has the advantage that 'best-of-breed' or new models may be used for particular aspects of the whole simulation. Deployment/execution is the final phase where the configured simulation is ready to be used. MSaaS also represents a potential path into the FMN for the amalgamation of Live, Virtual, and Constructive training capabilities described in section 2, which presents significant technical challenges. Ideally, the combination of these disparate approaches will yield an experience that is experienced as seamless by the trainees. SISO been developing the WebLVC standard to combines the latest World-Wide-Web distributed system communication technologies with simulation as shown in Figure 6. WebLVC is in balloting as a SISO standard [26] and could be employed productively in the FMN to support coalition training.

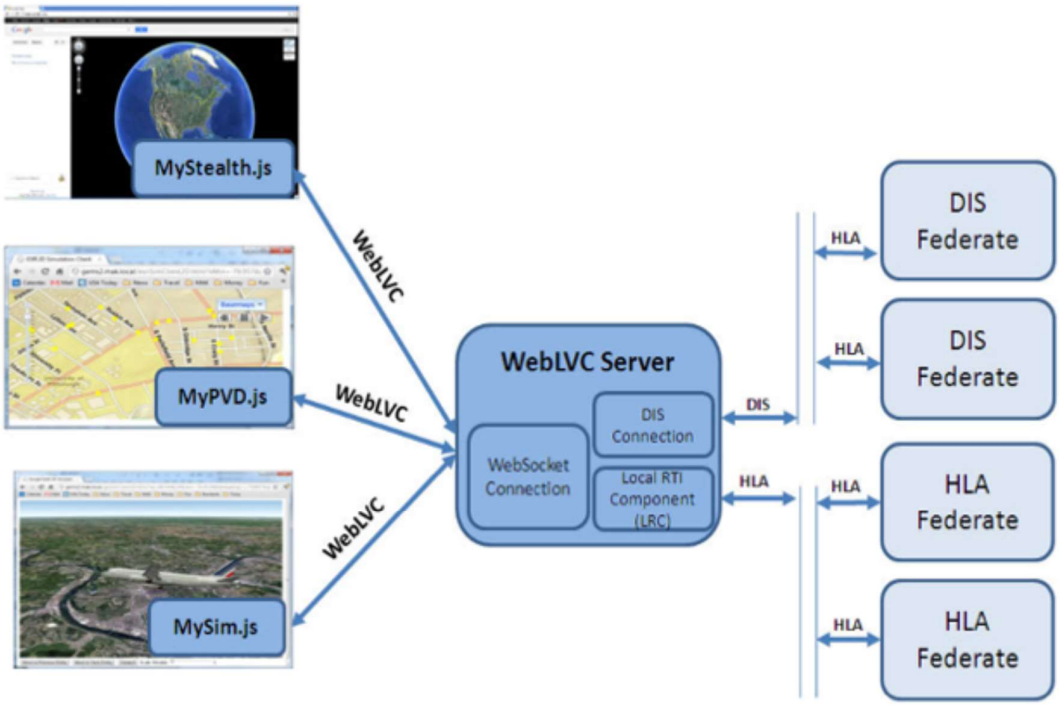


Figure 6: SISO WebLVC

5.3 High Level Architecture (HLA) for Modeling and Simulation

HLA [27] is an IEEE simulation interoperability standard developed by SISO that has been adopted as NATO STANAG 5603. HLA uses an object model approach to define the information that may be exchanged between simulations. The most important are objects (persistent items such as physical entities) and interactions (usually transient events such as weapon detonations). HLA is supported by its own management services for things such as object management and time management. The interfaces and underlying services are provided by supporting software known as the Run-time Infrastructure (RTI). The objects, interactions and associated ancillary information are defined in a Federation Object Model (FOM). HLA terminology gives the names federate to any HLA-compliant application and federation to a group of federates operating together using the same FOM and RTI.

5.4 Distributed Interactive Simulation (DIS)

DIS is the original SISO standard for networked military simulation [28]. Using it, entity-based simulations interoperate by exchanging state several times per second over a broadcast or multicast network. Typically, the simulations are physics-based and reflect the performance of a platform such as a tank or helicopter, although dismounted soldiers also are possible. The simulator displays show the out-the-window/viewfinder battlefield at the platform level along with vehicle dashboard status. This can support exercises in teamwork critical to military organization performance. DIS has the maturity of a 30-year-old technology and many implementations are available. It is notably simpler to implement and administer than HLA but is limited in scope to interactions of at most a few hundred battlefield objects (most often it is used for under one hundred battlefield objects, a reasonable match for recent NATO deployed missions). DIS is generally considered to be simpler to implement than HLA due to its simple, real-time object model, However, it lacks the broader scope of HLA, which is able to federate a variety of advanced, complex composite and distributed systems, and has not been adopted as a NATO STANAG.

5.5 Verification, Validation and Accreditation (VV&A) of military simulations

The IEEE standard Recommended Practice for VV&A [29] was developed by SISO to help guide simulation developers through an accredited verification and validation process. These guidelines have been refined further by the US DoD MSCO, who have taken the process further, refined it as US Mil-Std 3022 and published templates for V&V plans and reports and accreditation plans and reports. Providing these VV&A plans and reports fits in well with the FMN systems engineering acceptance process.

5.6 Distributed Simulation Engineering and Execution Process (DSEEP)

The IEEE standard DSEEP [32] gives a well understood way to manage a complete simulation process from an operational concept through to final execution and analysis. There are sevenstages in DSEEP as shown in Figure 7.

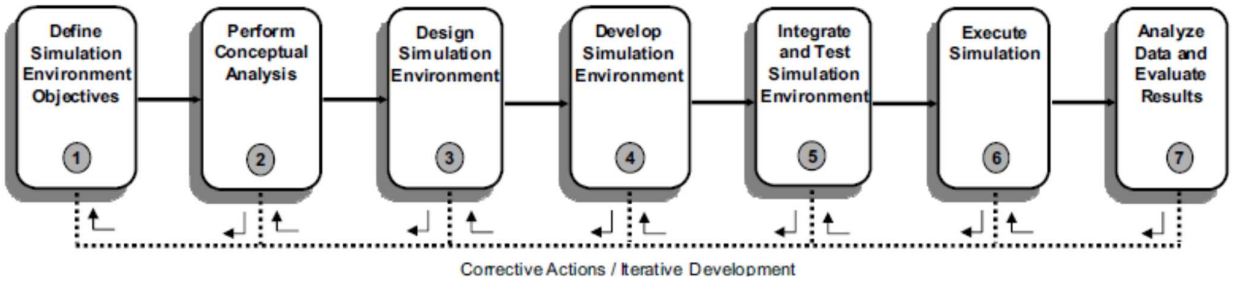


Figure 7 - DSEEP Phases

DSEEP does not explicitly address the design of scenarios, there are other processes for this such as the SISO GSD and NATO Mission Threads approach, but it does address the systems engineering aspects of designing a system to execute a scenario. SISO has defined a DSEEP overlay process which can be adapted for related activities, in the current case to introduce C2SIM capability into an FMN environment.

The Guideline on Scenario Development (GSD) [33] is a document published by a SISO product development group, which aims to help a scenario developer work through stages 1 to 5 of DSEEP. It defines three phases of scenario development from the operational, usually provided in response to a requirement, possibly that of a military user, through a formalized conceptual specification, expressed, for example using C2SIM to a final executable scenario customized for the specific simulations it is to be run on. GSD is pertinent to development of simulation scenarios for FMN.

6. Conclusion

We have described the need, development methodology, and plans for Federated Mission Networking, a major step forward in preparing the NATO Coalition for multinational deployments. The "Day Zero Interoperability" concept of FMN is well suited to incorporation of a variety of modeling and simulation standards as described above. It is our intention to participate in the FMN Spiral process to achieve this, in order that NATO will have capabilities necessary to

continuing its role of sustaining international peace.

Ultimately, M&S in the FMN can extend the force-multiplier effect of the FMN beyond the initial coalition training M&S focus. Operational use of M&S for COA analysis and ultimately for mission rehearsal are capabilities that should extend the capabilities of NATO coalition forces, resulting in greater effectiveness and including the ability of smaller, more nimble forces to achieve NATO missions. All of these capabilities need to be implemented in CAX for a "train as you fight" military capability.

7. References

[1] Pullen, J., K. Galvin and R. Brook, "Simulation in NATO Federated Mission Networking," International Command and Control Research and Technology Symposium, 2020, to appear

[2] Lanbert, D. and J. Scholz, "A Dialectic for Network Centric Warfare," Command and Control Division Defence Science & Technology Organisation Defence Science & Technology Organisation, Australia, 10th International Command and Control Research and Technology Symposium, 2005

[3] Palaganas, R. Implementing NATO Network Enabled Capability: Implications for NATO Response, US Army War College, https://csl.armywarcollege.edu/usacsl/publications/Section%20Three%20-%20Palaganas.pdf, last visited 22 June 2020

[4] Allied Command Transformation, Federated Mission Networking, https://dnbl.ncia.nato.int/FMNPublic/SitePages/Home.aspx, last visited 12 June 2020

[5] Allied Command Transformation, FMN Specifications, https://tide.act.nato.int/tidepedia/index.php/FMN_Spiral_Specifications, last visited 12 June 2020

[6] 2016 Warsaw Summit Communiqué: https://www.nato.int/cps/en/natohq/official_texts_133169.htm

[7] Allied Command Transformation, FMN Vision, https://tide.act.nato.int/tidepedia/index.php/FMN_Vision, last visited 12 June 2020

[8] Pezzato, J., "Federated Mission Networking," briefing to NATO TIDE Sprint, March 30, 2020

[9] North Atlantic Treaty Organization, "NATOTerm: The Official NATO Terminology Database," https://nso.nato.int/natoterm/content/nato/pages/home.html?lg=en, last visited 24 June 2020

[10] NATO Collaboration Support Office (MSG-136) "Modelling and Simulation as a Service: Technical Reference Architecture," 2017

[11] NATO C2 COE, "C2 Vision 2030," https://c2coe.org/c2-focus-area/

[12] NATO Science & Technology Organization Modeling & Simulation Group Technical Activity 145, "Final Report - Operationalization of Standardized C2-Simulation Interoperability," 2020

[13] NATO Science & Technology Organization Modeling & Simulation Group Technical Activity 085, "Standardisation for C2-Simulation Interoperation," 2014

[14] NATO Allied Command Transformation, Comprehensive Operational Planning Directive, 2013

[15] Hopkins, R. and K. Jenkins, Eating the IT Elephant, IBM Press, 2008

[16] Scrum, J., Agile Project Management Guide, Amazon Digital Services, 2020

[17] Allied Command Transformation, FMN Management Structure, https://tide.act.nato.int/tidepedia/index.php/FMN_Management_Structure, last visited 12 June 2020

[18] Allied Command Transformation, TIDE Sprint, https://www.act.nato.int/tide-sprint, last visited 12 June 2020

[19] Allied Command Transformation, FMN Syndicates, https://tide.act.nato.int/tidepedia/index.php/FMN_Syndicates, last visited 12 June 2020

[20] Pullen, J. and K. Galvin, "New Directions for C2-Simulation Interoperability Standards," International Command and Control Research and Technology Symposium 2016, London, UK

[21] NATO, BI-SC Collective Training and Exercise

Directive (CT&ED) 075-003, update Dec 2014, 2014

[22] SISO-STD-019-2020 Standard for Command and Control Systems - Simulation Systems Interoperation, 2020

[23] SISO-STD-020-2020 Standard for Land Operations Extension (LOX) to Command and Control Systems - Simulation Systems Interoperation, 2020

[24] North Atlantic Treaty Organization, NATO Mission Thread Development Guide. 2017

[25] Siegfried, R., "NATO Modelling & Simulation Group Contributions to Implementing NATO S&T Strategy," briefing to NATO TIDE Sprint , April 1, 2020

[26] SISO Digital Library, WebLVC overview, last visited 19 June 2020

[27] IEEE Standard 1516-2010, Standard for Modelling and Simulation (M&S) High Level Architecture (HLA) – Framework and Rules, 2010

[28] IEEE Standard 1278, Distributed Interactive Simulation, 2012

[29] IEEE Standard 1516.4 Recommended Practice for Verification, Validation, and Accreditation of a Federation - An Overlay to the HLA Federation Development and Execution Process, 2007

[30] SISO-GUIDE-006-2018 Guideline on Scenario Development, 2018

[31] US DoD MSCO Verification, Validation and Accreditation, http: vva.msco.mil, April 2020, last visited 23 June 2020

[32] IEEE Standard 1730-2010: Recommended Practice for Distributed Simulation Engineering

and Execution Process (DSEEP), 2010 [33] SISO GSD PDG. SISO-GUIDE-006-2018 Guideline on Scenario Development for Simulation Environments https://www.sisostds.org/DesktopModules/Bring2mind/DMX/API/Entries/Download?Command=Core_Download&EntryId=47237&PortalId=0&TabId=105

The Trend Toward Common Architectures

Peter Swan, Director International Sales,
pswan@mak.com

Len Granowetter, Chief Technology Officer,
lengrano@mak.com

Bob Holcomb, Chief Architect, Scalability,
bholcomb@mak.com

MAK Technologies, Cambridge MA, USA

Abstract

Defense forces around the world are starting to realize the benefits of a common distributed simulation architecture for collective training, whether in the land or air domain. Leaders are frustrated by having to pay for terrains and models multiple times, spending time and effort attempting to federate incompatible training systems, and being unable to guarantee a fair fight between these systems. This paper will describe two different architectural approaches being taken to resolve these issues and to deliver collective training.

1 Introduction

The benefits of distributed simulation have been well known, even since the early days, epitomized by the Defense Advanced Research Projects Agency (DARPA), SIMNET (SIMulationNETworking) Program [1]. Those early systems struggled with high latencies and a lack of bandwidth. It was not feasible to transmit large quantities of data around – just the bare minimum needed for distributed simulations to interoperate. However, the advent of fiber networks providing 200+ Mbs into our homes, and the availability of pervasive cloud environments, is making simulation-based training available at the point of need, as well as enabling distributed, collective training.

This paper describes two specific architectural approaches – a Scalable Holistic Cloud Architecture and a Common Components Architecture - and how they are leveraging modern networking and cloud environments to deliver training.

2 Scalable Holistic Cloud Architecture

2.1 STE Overview

The US Army initiated the Synthetic Training Environment effort in 2016 with the mission of providing a cognitive, collective, multi-echelon training and mission rehearsal capability for the Operational, Institutional and Self-Development training domains; of converging the virtual, constructive and gaming training environments into a single Synthetic Training Environment (STE) for Active and Reserve Components as well as civilians; and to provide training services to ground, dismounted and aerial platforms and command post (CP) points-of-need (PoN) [2]. MAK was selected to deliver initial prototypes for the Training Simulation Software (TSS) and Training Management Tool (TMT) components of the STE Common Synthetic Environment (CSE). As part of this effort, MAK developed a hybrid cloud-based scalability architecture for handling very large entity counts.

2.2 STE CSE System Architecture

The STE vision is to develop and deliver a Common Synthetic Environment underpinning the virtual simulators, semi-automated forces and rendering engines across all future US Army collective trainers. The architecture has the following attributes:

- An open systems architecture with a suite of **documented APIs, SDKs, and editors** to allow users to customize, extend, and adapt any elements of these systems, and to directly integrate new and existing models, simulations and client applications into the Common Synthetic Environment

- A truly **modular architecture that supports scalability** by combining multiple instances of the simulation engine to share the simulation load and communicate with each other through a common, network protocol independent interface

2.3 Hybrid Cloud Model

The system architecture is based on a hybrid cloud model, with the flexibility to enable training wherever it is needed, whether at the point of need, at homestation, in a training center, or distributed across many sites. The architecture must support the processing where it is needed for optimal performance. For example:

- Rendering is done locally on edge computers for Mixed Reality applications using headsets, to avoid delays and jitter caused by the network, but can be done remotely in a cloud for desktop training applications.

- Simulation is conducted locally for small local exercises when connectivity is missing, but can run entirely in the cloud for large constructive simulations.

2.4 Scalability

The STE Architecture is required to support scalability in terms of:

Capacity for terrain data. MAK's VR-TheWorld Server and the MAK Earth terrain engine are built to handle whole-earth terrain at unlimited resolution. Unlimited layers of source imagery, elevation, features, and land use data can be stored on one or more copies of VR-TheWorld Server, which can run in a cloud environment to access growing storage resources. MAK permanently hosts an instance of VR-TheWorld Server on the Amazon Web Services network to demonstrate this capability. Source layers are tiled into quad-trees so that client applications can have immediate access to the specific tiles they need – without being overwhelmed by data outside their areas of interest. The MAK Earth terrain engine can be configured at each client to dictate how aggressively to page in each layer of terrain data, and the level of styling detail to apply. As compute and GPU resources continue to increase, clients can choose to use higher-resolution textures, more detailed building models, or more detailed deterministic fractal noise, so that visual fidelity keeps up with future expectations.

Number of players. To ensure player entities can be modelled at a high fidelity and high frame rates, even in large simulation exercises, responsibility for simulating ownship entities can be assigned to a dedicated instance of the simulation engine. Resources on the player's station can be reserved for high-performance 3D rendering or other tasks by offloading simulation responsibility onto remote machines. For cloud-based deployments with lightweight client devices, we can offload both simulation and visualization onto server-side (cloud) resources so that the player's system simply forwards user input to the simulation server and displays a video stream generated by graphics resources in the cloud. The player's UI/control/display can be fully decoupled from the responsibility to simulate the physics and behavior of the entity. Exercise control can be distributed across multiple

stations at multiple sites for large exercises or compressed into a single station for small exercises.

Entity count. The STE architecture must support Brigade level exercises, with the capacity to handle millions of entities representing a city's population. VR-Forces has a built-in ability to divide simulation responsibility across many instances of the VR-Forces simulation engine. Entities can be assigned to simulation engines by the scenario author, or the system will evenly distribute entities across engines automatically. Each user of the VR-Forces GUI can access the full set of entities in a scenario without having to worry about which specific copy of the simulation engine owns each entity.

In Addition, the scalability solution takes advantage of:

- Spatial Organization – responsibility for simulating entities is shared across multiple simulation engines, each covering a specific geographic area.
- Ownership transfer - ownership is transferred from one simulation engine to another as the entity moves from one region to another.
- Load balancing - the size of the region covered by a simulation engine is automatically adjusted based on the density of entities in the region.
- Interest Management – the simulation clients register interest in entities based on certain criteria, typically only those that they might interact with.

Within a real-time High Level Architecture (HLA) federation, the simulation engine will support on the order of 50,000 complete and intelligent semi-automated entities, constrained largely by the fact that HLA transmits data separately for each entity.

For larger entity counts, MAK is developing a scalability architecture we are calling Legion. A centralized Entity Server maintains a mirror of each Sim Engine's Data Store. Data stays in the same contiguous layout all the way from the Sim Engine's Data Store, to the network, to the Entity Server's Data Store.

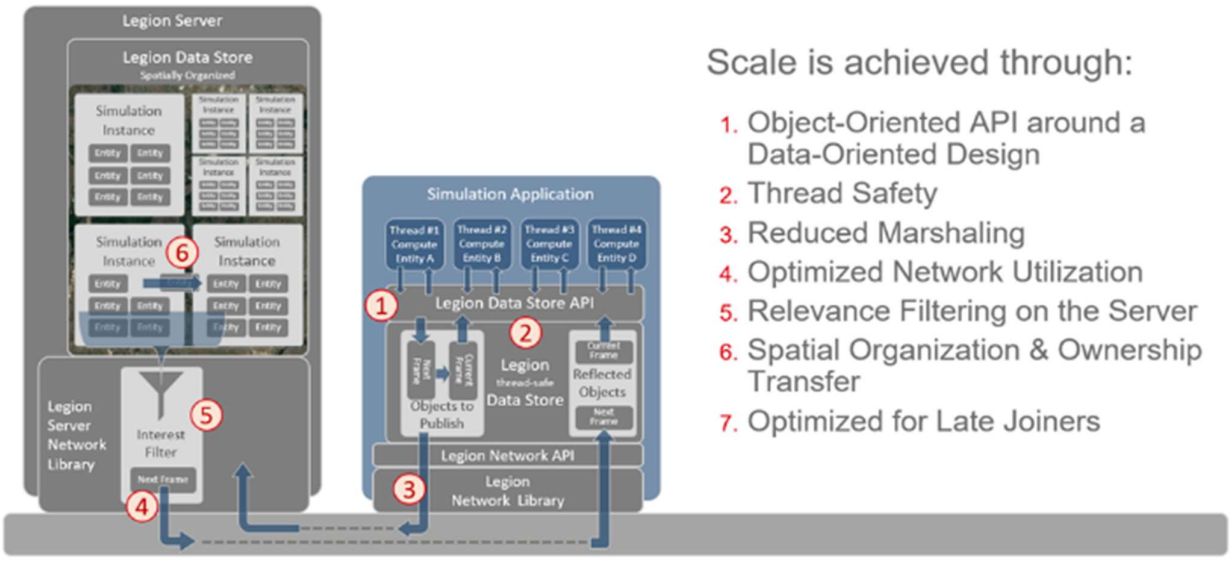


Figure 1 - Legion Scalability and Interoperability Architecture

This eliminates expensive marshalling and copying. Network API abstraction allows various network protocol choices. The default TCP implementation ensures reliability (even over WANs) and allows the efficiency of sending large packets that include many entities' state.

When there are too many Sim Engine instances for one Entity Server to handle, the load is shared across multiple Entity Servers. The Entity Server filters the data against interest criteria, builds a large message of just the data required by a particular client, and sends the message directly to the network.

2.5 Interoperability

Legion provides an Object-Oriented API that incorporating the familiar simulation concepts that simulation developers have been using for decades. This API is backed by an efficient data-oriented, thread-safe data storage library, allowing developers to take advantage of the full processing power of the target hardware. Users of the Legion API may write multithreaded code without needing to worry about locking or ensuring data consistency.

The architecture also provides a networking library that is designed to take advantage of the internal data layout in the data storage library to facilitate optimized replication of the simulation data with other clients. The networking library communicates with the Legion Server to ensure that only data that is required for the simulation application is transmitted.

The default object model that is provided with the Legion Architecture is derived from the DIS and HLA RPR FOM object models. It utilizes the SISO Enumerations document to define common entity types, munition effects, and other data shared between clients. By leveraging similar concepts to well established simulation protocols, building a heterogeneous simulation exercise that uses both new and existing systems is much easier.

The VR-Exchange product bridges DIS, HLA, DDS and Legion protocols, sharing content between all simulation systems.

Given that the Legion data store and the Legion network library are independent of one another, MAK is planning to implement Legion connections that utilize both DIS and HLA for data replication and transport. This would allow simulation application developers to build their application against the Legion API and then choose at run-time the networking protocol that will support their exercise best.

MAK is committed to continuing support of DIS and HLA as open standards and the libraries that we provide to implement them. Legion is another architecture that is useful in some use cases and we recognize that it is not a one size fits all solution. It is our intent to propose the Legion API to SISO for standardization and to be made an open standard.

3 Common Components in an HLA Federation

3.1 DOTC (Air)

In May 2019, the UK RAF awarded Boeing Defence UK the contract to deliver the Defence Operational Capability (Air) Core System and Services (DCS&S) contract. DCS&S will create a capability, known as Gladiator, that will support multiple complex training scenarios, simultaneously and independently of each other. The system will allow personnel at different sites to train together in their own high-fidelity simulators, linked by a secure network to a new hub at RAF Waddington. The system will securely manage training events across locations and classification levels, allowing RAF crews to experience the same battle environment and threats, including in joint training with their counterparts in the United States.

3.2 DOTC (Air) System Architecture

Rather than a clean sheet, new architecture as envisioned for STE, DCS&S will link together existing and new training centers developed by different suppliers such as BAE Systems, Boeing, CAE and Thales. It will need to resolve the correlation and 'fair fight' issues created when different simulations are linked together.

The DOTC (Air) network conceived by the RAF will be implemented using the High Level Architecture (HLA) standards specified in the NATO Education and Training Network (NETN) federation agreement document. It will provide centralized services for instructor control, threat modelling, pattern of life, terrain, weather, communications and visualization.

DCS&S will provide the common Computer Generated Forces application, Image Generator, and Role Player Stations for UK RAF distributed training exercises based on MAK's COTS products, extended to support Tactical Data Links and high fidelity flight and electronic warfare models.

Boeing will deliver Generic Exercise Management Systems (GEMS) at the central hub and at each main operating base. The GEMS will be reconfigurable for scenario generation, After Action Review and as an Instructor Operator or Role Player Station.

3.3 DOTC (Air) Common Components

The DCS&S contract was let to Boeing Defence UK to provide centralized services to create a networked training environment for the Royal Air Force and to deliver a suite of common components with the following characteristics:

- **Commercial Off The Shelf (COTS) software**.
- **Modularity**. The MAK ONE common components are made up of a set of modules (see Figure 2) that provide support for terrain, interoperability, rendering and simulation.

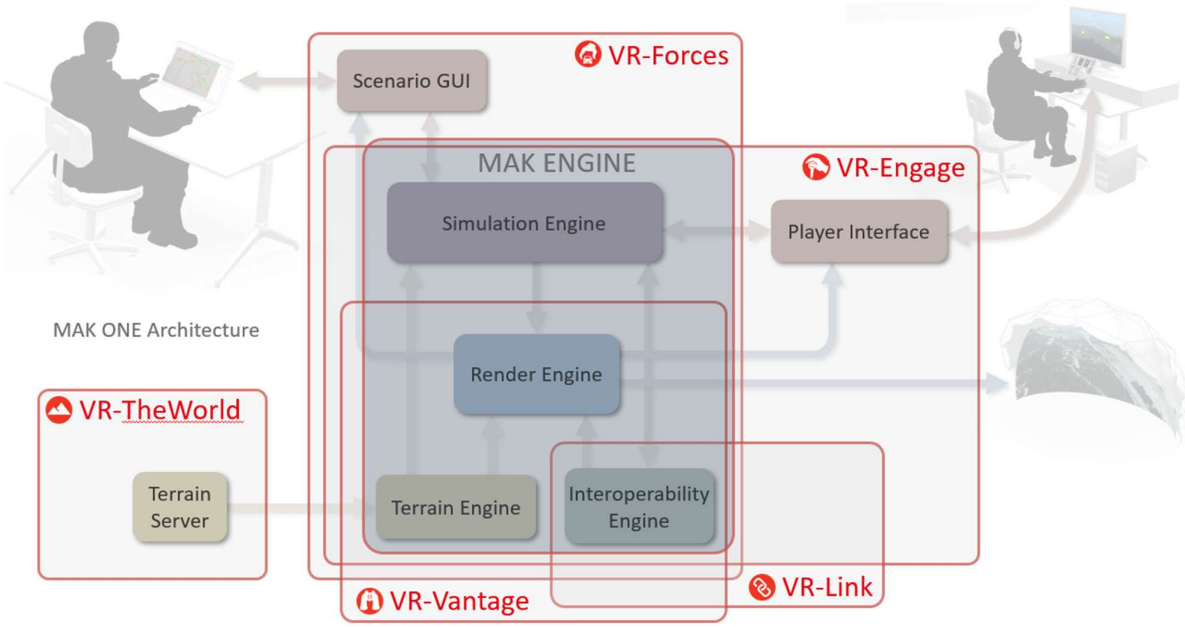


Figure 2 - DOTC(Air) Common Components

- **Toolkits with open APIs**. MAK ONE includes a modular architecture and complete SDK to support customization of existing and future components. Modular design has been a hallmark of the MAK ONE platform from the very beginning. Each module shown in Figure 2 has its own API. The SDK allows developers to replace and adapt components of the software to meet specific program needs. Rather than providing a limited plug-in API, the SDK provides header files, libraries, and documentation for all of the implementation classes. This allows developers to override almost any aspect of MAK ONE functionality (large or small) through subclassing and "factory" patterns. For example, the VR-Forces simulated entities are comprised of:

  - Sensors providing perceived-truth information about other entities, the terrain, and the environment.
  - Actuators modelling the physics or dynamics of the entity, compute its state, and update the state repository for the entity, which is shared with other entity models and other instances of the simulation engine, and with the GUIs and visual systems.
  - Controllers implementing decision logic – executing tasks, plans, or behavior scripts. They use input from users, sensors, and task/script parameters, and knowledge about the terrain to make decisions, and generate appropriate inputs to the actuators.

  The SDK allows developers to modify or replace these individual low-level components of entities. An example would be to replace the existing entity controllers with an external AI engine. See Figure 3.

  The MAK ONE architecture isolates the GUI elements from the underlying simulation engine. The SDK provides a Remote Control API that enables alternate GUIs and

other remote applications to control any aspect of the simulation. Other APIs (i.e., abstraction layers) are provided to enable easy replacement of interface devices, operational software, and 3rd Party software libraries as well as connectivity to legacy and future systems.

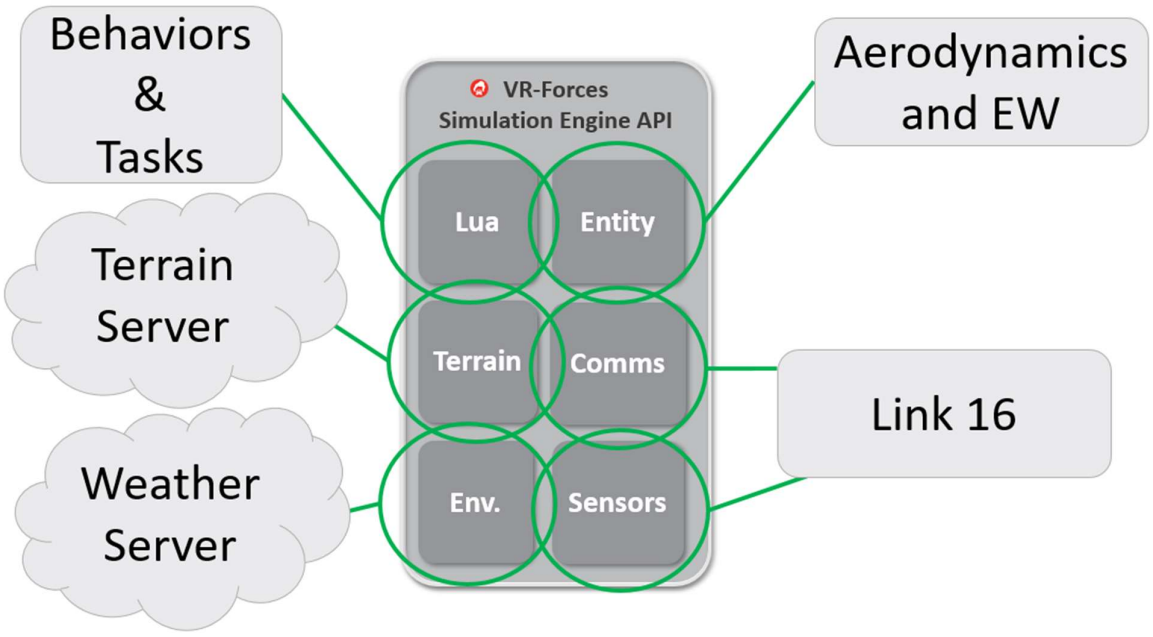


Figure 3 - VR-Forces Simulation Engine API

- **Based on open standards**. The interoperability engine used in all MAK ONE products, supports all flavors of DIS and HLA out-of-the-box, including the NATO Education and Training Network (NETN) FOM. The networking API allows for relatively easy integration with other networking protocols such as SpatialOS, DDS, TENA and MAK's own Legion.

MAK ONE supports other standards such as:

- CIGI – the Common Image Generator Interface, a SISO standard that specifies communication between an Image Generator and its host simulation, is supported by VR-Vantage.
- MSDL - The Military Scenario Definition Language is a SISO standard for scenario laydown. VR-Forces supports import and export of most MSDL information.
- Open Geospatial Consortium streaming terrain standards in both our VR-TheWorld streaming terrain server and simulation and visualization clients. MAK ONJE supports streaming of elevation, imagery and feature data to build and procedurally enhance terrain on the fly.
- WebLVC – initiated by MAK, this soon-to-be SISO standard is a protocol for enabling web and mobile applications (typically JavaScript applications running in a web browser) to play in traditional M&S federations

(which may be using Distributed Interactive Simulation (DIS), High Level Architecture (HLA), Test and Training Enabling Architecture (TENA), or related protocols and architectures). MAK's WebLVC Server can be used to extend the simulation out to thin client web applications.

- **Supports a central terrain server**. Imagery, elevation, and features including roads, buildings and trees - will be stored centrally in standard GIS formats and streamed to the local federates using Open Geospatial Consortium (OGC) protocols TMS and WMS-C. Whole-Earth terrains will be procedurally generated from GIS data streamed from the terrain server, with the addition of cut-in high-resolution site models.

- **Supports a central weather server**. A weather server will provide centralized environment services to the federation.

4. Conclusion

Two major simulation-based training programs have chosen significantly different approaches to meeting their distributed simulation requirements:

- Scalable Holistic Cloud Architecture: The US Army is developing an entirely new architecture and common simulation software for the Synthetic Training Environment; and
- Common Components Architecture: The UK Royal Air Force has developed the concept of an HLA based federation linking together and providing centralized services for existing and new training centers for the DOTC(Air) Program.

Future programs can perhaps learn from these programs how best to implement their own architectures based on their specific requirements.

References

[1] D.C. Miller, SIMNET and Beyond: A History of the Development of Distributed Simulation, I/ITSEC (2015)

[2] PEO STRI Website: https://www.peostri.army.mil/synthetic-training-environment-ste-

Multi-Viewpoint Conceptual Modeling *In Support of Simulation Interoperability Readiness Levels (SIRLs)*

Dr. Katherine L. Morse
JHU / APL
katherine.morse@jhuapl.edu

Mr. David L. Drake
JHU / APL
david.drake@jhuapl.edu

Abstract

Several attempts have been made over the last two decades to produce a conceptual modeling format and process that are understandable by all stakeholders, both operational and engineering. Because a conceptual model is the bridge between these stakeholders through which they create a shared understanding of what will be represented in the simulation, it's critical that the format be unquestionably lucid, while providing sufficient detail from which to engineer the solution. The Johns Hopkins Applied Physics Laboratory (JHU/APL) has developed such a format on behalf of the OUSD(R&E) Modeling & Simulation Enterprise (MSE). Multi-Viewpoint Conceptual Modeling (MVCM) provides multiple views of the conceptual model: intended use, narrative, behavior, causal chains, entity subset conceptual models, characteristics, references, and assumptions. Through a facilitated process, all stakeholders create a conceptual model using successive decomposition from the intended use and narrative down to the detailed views of causal chains and characteristics. The level of detail is managed within each view to ensure that sufficiently precise information is collected to support engineering, while avoiding the pitfall of getting down to design. Opportunities for tool support of the process are highlighted.

1 Problem Framing

Given any 2 simulations, how do we rapidly and accurately determine if they will interoperate?

The Multi-Viewpoint Conceptual Modeling (MVCM) approach was motivated by an effort within Simulation Interoperability Standards Organization's (SISO) Federation Engineering Agreements Template (FEAT) Product Support Group (PSG) to define Simulation Interoperability Readiness Levels (SIRLs) [1]. Rapidly and

accurately determining whether a set of simulations can be integrated into a working, federated simulation producing valid results remains one of the key challenges for distributed simulation. The SISO SIRL effort was started to address this challenge. Table 1 shows the levels that make up SIRL. Note that the top level, Conceptual, only identifies one item of engineering evidence, the conceptual model, a long-standing, known simulation interoperability gap.

Interoperability Readiness Level	Evidence Assessment Effort	Engineering Evidence
Conceptual Level: "The description of what the model or simulation will represent, the assumptions limiting those representations, and other capabilities needed to satisfy the user's requirements." - DoD M&S Glossary. This is entities, behaviors, and their interactions; who does what to whom.	Interoperability assessment requires systematic analysis by all stakeholders	• Conceptual Model
Modeling Level: "A physical, mathematical, or otherwise logical representation of a system, entity, phenomenon, or process." - DoD M&S Glossary. These are choices about how to model real world phenomena.	This level needs a human to determine interoperability. Interoperability assessment is constrained to individual aspects.	• Terrain representation and correlation • Environmental effects assumed or modeled • Simulated time representation • Kinetic effects adjudication • Aggregation and disaggregation algorithms • Communications (representation) • Non-kinetic effects adjudication • Inter-visibility • Movement (physics) • Decision-making behavior • Human performance • Platform / equipment performance (not movement, e.g., reliability) • Entity resolution • Dead reckoning • Human behavior representation (pattern of life)
Simulation Control Level: These are design decisions about controlling the execution of the simulation.	Interoperability may not be determined in a semi-automated fashion depending on simulation interoperability solution.	• Update rates • Execution states: initialization, start, stop, save, restore, join, resign • Synchronization points • Interest management • Control transfer • Time advancement • Simulation control mechanisms (HITL, test harnesses, federation manager) • Fault tolerance mechanisms (failure modes)
Data Level: These are the formats and models of data to be used for initialization, run time data exchange, and logging.	Interoperability can be determined in a semi-automated fashion. Human analysis can be carried out on a case-by-case basis.	• Initial laydown of entities and entity starting states • Terrain playbox • Terrain format • Terrain resolution • Environment format/representation • Authoritative data sources • Data storage formats • Data exchange models • Coordinate systems
Technical / Syntactic / Communication Level: These are decisions about the simulation architecture including hardware, software, and protocols.	Easily assessed in an automated fashion.	• Middleware / protocol • Secondary communication channels • Hardware • OS • Simulation support tools

Table 1. SIRL Hierarchy

At its most basic, a conceptual model describes entities and behaviors modeled within simulations and what triggers / interactions occur between simulations. It answers the question, "who does what to whom?"

2 Previous Approaches

Several previous approaches to conceptual modeling have been tried but failed to gain significant traction [2]. In this section, we identify some of these approaches and discuss their shortcomings.

Natural language is tempting because we all speak it, but it's too vague to derive the engineering detail necessary to evolve the conceptual model to a simulation. For example, "Represent Joint Close Air Support (JCAS) using UH-1Y Super Hueys." This approach lacks specificity such as the intended use, which will drive resolution of simulations including what specific systems and environment need to be represented.

Some engineering views can be understandable by all stakeholders, e.g., sequence diagrams as in Figure 1, and do a good job of representing causality, triggers, and message specifications. But doing them for anything beyond a simple vignette is intractable.

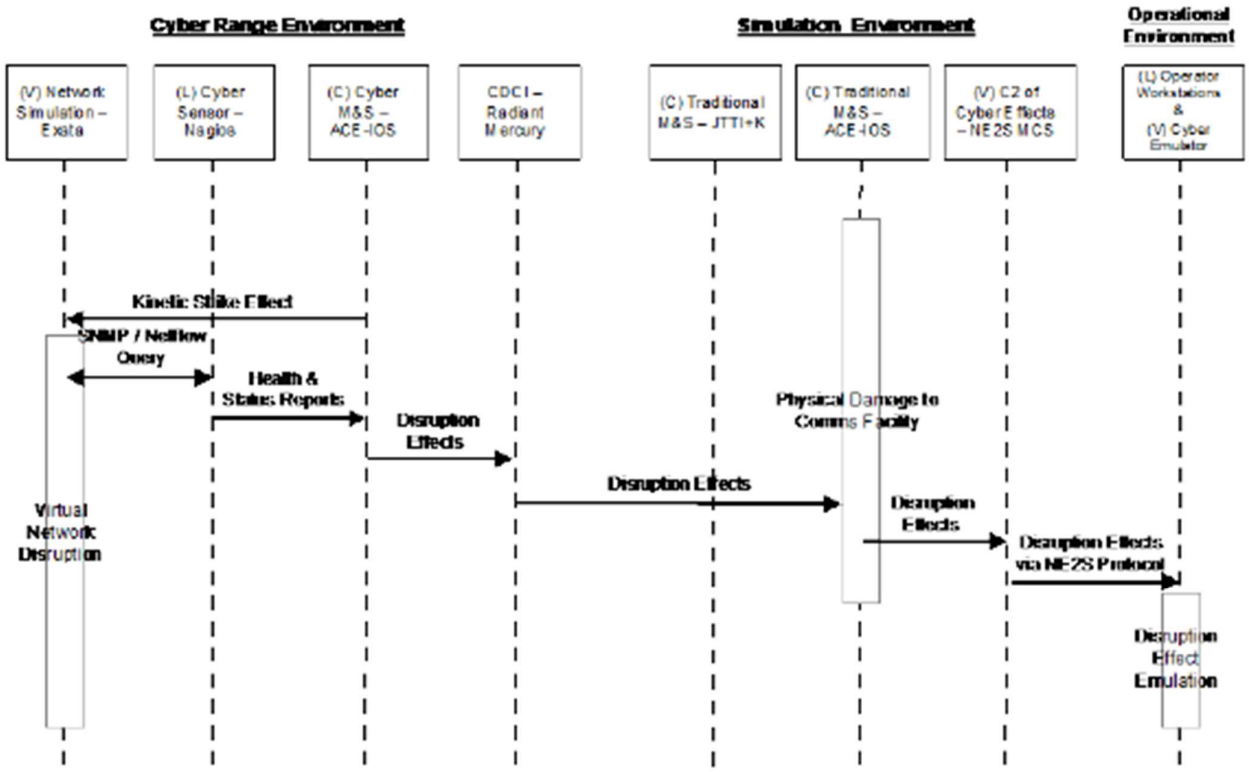


Figure 1 – An Example Sequence Diagram [3]

Solutions such as technically complete UML / SysML diagrams such as the one in Figure 2 have historically been too technical for non-engineers. What do the different arrowheads mean? What's the difference between a solid and dashed line? Even when these questions are answered in the course of defining the conceptual model, they're a barrier to easy understanding.

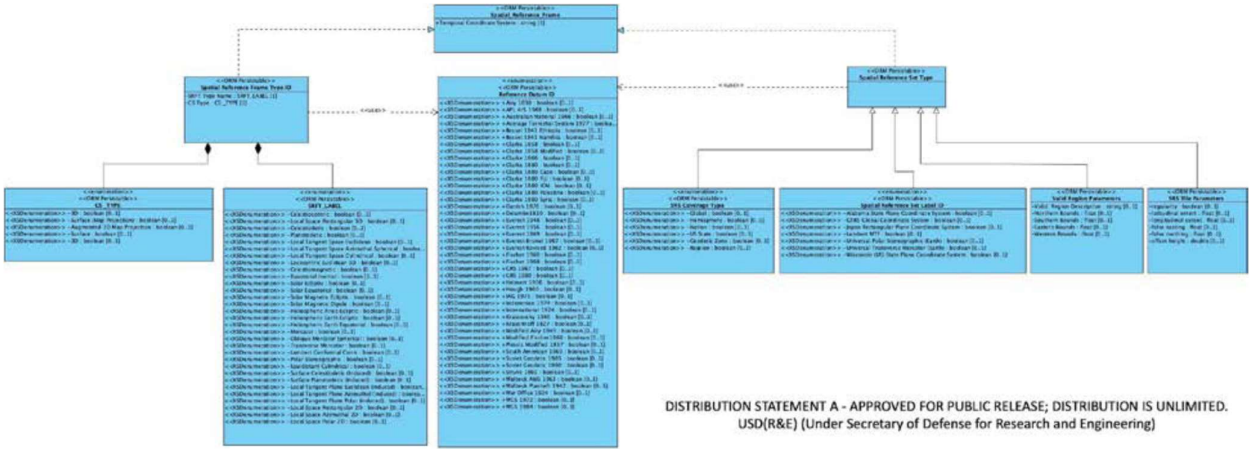


Figure 2 – Example Class Diagram for Metadata Definition

"Human readable" is a critical requirement. The MVCM approach is a fresh look at conceptual modeling from first principles.

3 Bridging the Gap with MVCM

MVCM is designed to bridge different stakeholder perspectives by providing multiple viewpoints into all aspects of the conceptual model. It is also the bridge to all the requirements and supporting engineering evidence for the other levels of SIRL.

A key to bridging the gap between stakeholders is to translate between natural language and engineering detail. In this translation, things = nouns = entities with adjectives ≈ characteristics. Verbs = behaviors and subject vs direct object = triggers. All of these relationships become structures, i.e., the conceptual model views described in the next section.

The intent of MVCM is to address this conundrum:

We can't decide if a simulation does what we need until we know what we need the simulation to do.

4 MVCM Views

The audience for these views includes both stakeholders and engineers, so managing the level of detail is critical. For the sake of keeping this paper to a reasonable length, the details of the examples in the following views are not all complete, e.g., authoritative references, icons, and characteristics. We hope to present more complete examples in future work.

In addition to managing the level of detail, MVCM strives to manage the complexity of developing and reviewing a conceptual model. Trying to view and understand the *entirety* of even a reasonably complex conceptual model is intractable. So, the views are designed to focus on *individual perspectives* as well as enabling limiting the percentage of the conceptual model that must be viewed at any time, reducing cognitive burden.

We are deeply indebted to MSE and the Marine Corp M&S Office for the joint personnel recovery use case example applied in the following subsections.

4.1 Intended Use View

The intended use view, Figure 3, captures domain, mission, goals, objectives, and users. This is the first step in down selecting simulations that could potentially be used. In this view, simulation domain and Service are constrained, while mission and goals / objectives are not

	Simulation Domain	Service	Mission
	Training	Joint	Joint personnel recovery
ID	Goals / Objectives		
G1	Training coordination of assets for joint personnel recovery		
O1	C2		
O2	Locate and positively identify IP		
O3	Navigate and ingress to pick up zone		
O4	Clear and isolate zone		
O5	Secure zone		
O6	Near positive ID		
O7	Secure IP		
O8	Egress		
	Training	Air Force	
	Testing	Army	
	Analysis	Coast Guard	
	Experimentation	Marines	
		Navy	
		Joint	
		Special Forces	

Figure 3 – Intended Use View

4.2 Narrative View

The narrative view, Figure 4, is the first step in translating the language to structure. The natural language narrative of the conceptual model is broken down into discrete elements in the narrative column. Then each element is analyzed to identify the entity (source) that initiates the action (behavior) and the entity (sink) that observes the action. When preparing this view, it's critical to focus on behaviors *between* simulations rather than *within* them because these represent elements of the runtime data exchange model and behaviors that are probably subject to federation agreements [4].

Subsequent analysis is critical to verifying:

- All objectives from the intended use view are addressed in the narrative
- All entities and behaviors are represented in the entity and behavior views
- All entities and behaviors represented in the entity and behavior views are required

See the discussion of tool support to this analysis in section 5.2.

Narrative	Source	Behavior	Sink	Notes / Questions
Then, at 2:08 a.m. Thursday, Bosnia time, an F-16 from O'Grady's squadron patrolling in the area of the shootdown heard Basher 52 come up on his radio.	IP Radio	Communicate	FAC A Radio	
It took about 12 minutes for procedures that positively identified O'Grady, and from there things moved rapidly.	FAC A Radio	Communicate	IP Radio	
	IP Radio	Communicate	FAC A Radio	
	FAC-A	Identify	IP	
	FAC-A Radio	Communicate	IP Radio	Requesting location and condition
	IP Radio	Communicate	FAC-A Radio	Reporting location and condition
	FAC-A Radio	Communicate	Rescue Team Radio	Reporting location and condition
	FAC-A Radio	Communicate	RESCORT Radio	Reporting location and condition
	FAC-A Radio	Communicate	Recovery Vehicle Radio	Reporting location and condition
At 5 a.m., six minutes before dawn, the rescue team began lifting off the Kearsarge.	Recovery Vehicle	Maneuver		
	RESCORT	Maneuver		
At 5:50, the rescue aircraft made their run: two CH-53 Sea Stallion helicopters with two dozen hand-picked Marines under 1st Lt. Martin Wetterauer; escort Cobra AH1W helicopters under Maj. Scott Mayleby, and above them a pair of Harrier jump jets led by Maj. Michael Ogden.	Recovery Vehicle	Maneuver		
	RESCORT	Maneuver		
	RESCORT Radio	Communicate	Rescue Team Radio	
	RESCORT Radio	Communicate	Recovery Vehicle Radio	

Figure 4 – Narrative View

4.3 Entity View

The entity view, Figure 5, specifies all the entities to be represented in the simulation as initially described in the narrative view. It includes:

• Their designation within the conceptual model, which will be used to cross reference to behaviors and characteristics

• Identifier / nomenclatures, which can be used to identify authoritative references

• Icon for representation in causal chains, entity subset conceptual models, and tools

• Potential variances where simulations may not exist and mitigating measures may need to be taken

Authoritative references are used throughout the MVCM to provide sources of detail for evaluating or designing simulations. The references perform the important function of capturing where specific detail is available without derailing the conceptual modeling process with too much detail.

Name	Identifier / Nomenclature	Authoritative References	Icon	Potential Variances
FAC-A	F-16	1;1.2.225.1.3		
FAC-A Radio	ARC-210(V)	1;5.8.3		
RESCORT	AH-1W	1;1.2.225.20.2.10		
RESCORT Radio	ARC-210(V)	1;5.8.3		
RESCORT Sensor	MK-I Eyeball	3		
IP	Downed US Army Aviator	1;17.11.1.3.1		
IP Radio	PRC-112	1;5.8.3		
IP Sensor	MK-I Eyeball	3		
Recovery Vehicle	CH-53	1;1.2.78.21.9.1		
Recovery Vehicle Radio	ARC-210(V)	1;5.8.3		
Recovery Vehicle Sensor	MK-I Eyeball	3		
Recovery Vehicle Weapon	M-2 Browning	1;5.4.5.9		
Rescue Team	TRAP	1;17.11.1.3.1		
Rescue Team Radio	AN/PRC-152	1;5.8.3		
Rescue Team Sensor	MK-I Eyeball	3		
Rescue Team Weapon	M-16	1;2.6.222.3.9		
Adversary Air Defense Threat Sensor	MK-I Eyeball	3		
Adversary Air Defense Threat Weapon	SA-7	1;5.4.5.19.20		
Adversary Ground Threat	Serbian Regular Troops	1;17.11.1.3.1		
Adversary Ground Threat Weapon	Zastava M72	1;5.4.5.9.127		May need to be represented visually as a Soviet RPK light machine gun.

Figure 5 – Entity View

4.4 Behavior View

The behavior view, Figure 6, associates behaviors with entities. Only behaviors germane to the intended use are captured. Just as with the intended use view, these are specifically the behaviors that are externally observable, particularly for triggering behaviors in other entities. Notice the use of authoritative references in this view as well.

Name	Behavior	Authoritative References	Potential Variances	Notes / Questions
FAC-A	Identify	2;VI.A.2		
FAC-A Radio	Communicate	4		
RESCORT	Maneuver	7		
	Mark location	2;VI.A.3		With colored smoke grenade
RESCORT Radio	Communicate	4		
RESCORT Sensor	Identify	2;VI.B.5		
IP	Maneuver	6		
	Mark location	2;VI.A.3		With colored smoke grenade
IP Radio	Communicate	4		
IP Sensor	Locate	2;VI.B.5		
Recovery Vehicle	Maneuver	8	Manmade terrain is required to simulate one of the Sea Stallions landing on a fence.	
Recovery Vehicle Radio	Communicate	4		
Recovery Vehicle Sensor	Locate	2;VI.B.5		

Figure 6 – Behavior View

4.5 Causal Chain View

The causal chain view concept, Figure 7, is based on behavior chaining from psychology. It ties together the entity and behavior views in support of achieving the intended use. Complexity is controlled by building multiple chains, each focusing on a single thread of interaction between a subset of entities. This is an example of limiting the complexity that needs to be evaluated at any one time to improve focus.

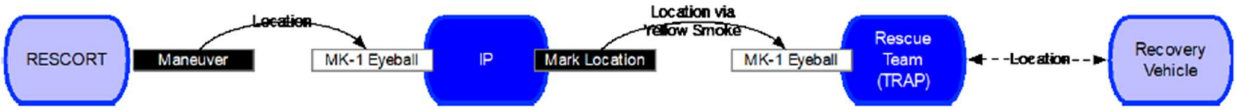


Figure 7 – Casual Chain View

With tool support, these chains could be "drawn" by dragging and dropping from the entity view, linking entity types, and labeling the links with behaviors. As previously stated, the focus is on behaviors between entities, not within them.

4.6 Entity Subset Conceptual Model View

In this view, Figure 8, a subset of entities is selected to view all the causal chains between those entities. The term "subset" is critical because viewing the entire conceptual model for a problem of any meaningful size is overwhelming. A focused perspective is a requirement for checking the overall model.

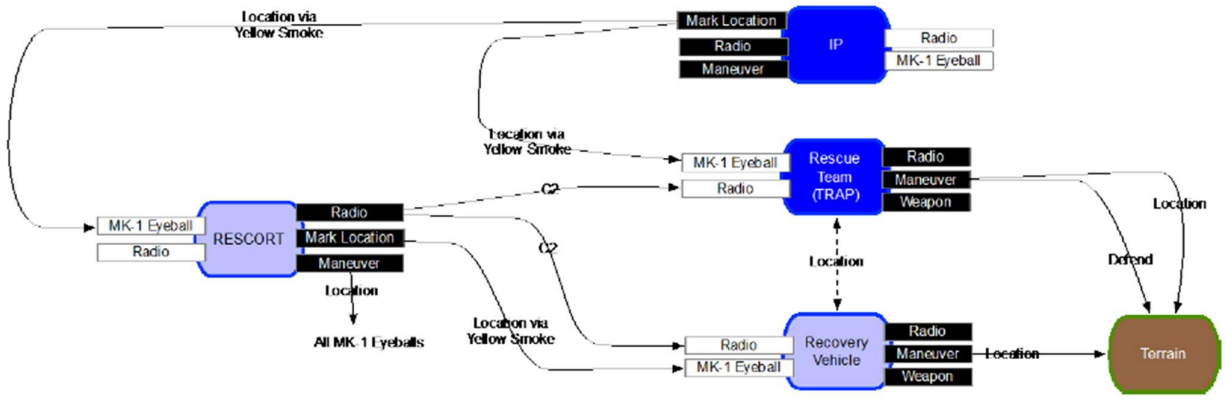


Figure 8 – Coordinated Landing Subset Conceptual Model View

4.7 Characteristic View

The characteristic view, Figure 9, identifies the characteristics of the entities that change with respect to the conceptual model, i.e., those characteristics of an entity that other entities need to perceive to change their own behavior. These could become input later to a runtime data exchange model in a federated simulation. Unchanging characteristics can be derived from the references.

Entity	Characteristic	Authoritative Reference
FAC-A	Location	9;5.3.2
FAC-A Radio	Transmit state	1;8.1.2
	Transmission	9;5.8.4
RESCORT	Location	9;5.3.2
RESCORT Radio	Transmit state	1;8.1.2
	Transmission	9;5.8.4
RESCORT Sensor	Acuity	10
IP	Location	9;5.3.2
IP Radio	Transmit state	1;8.1.2
	Transmission	9;5.8.4
IP Sensor	Location	9;5.3.2
Recovery Vehicle	Location	9;5.3.2
Recovery Vehicle Radio	Transmit state	1;8.1.2
	Transmission	9;5.8.4
Recovery Vehicle Sensor	Acuity	10

Figure 9 – Characteristic View

4.8 Reference View

As noted earlier, the reference view, Figure 10, collects all the references for supporting the other views without reproducing their contents in those views. References are authoritative sources of representation detail from which simulation developers and integrators can derive technical requirements that meet the intended use.

ID	Title	Date	Version / Producer ID	URL
1	Reference for Enumerations for Simulation	14 October 2019	SISO-REF-010-2019-v27	https://www.sisostds.org/DigitalLibrary.aspx?Command=Core_Download&EntryId=46171
2	Personnel Recovery	20 December 2011	JP 3-50	https://fas.org/irp/doddir/dod/jp3_50.pdf
3	Gray's Anatomy, 41st Edition, The Anatomical Basis of Clinical Practice	25th September 2015	Hardcover ISBN: 9780702052309	https://www.elsevier.com/books/grays-anatomy/standring/978-0-7020-5230-9
4	Aircraft Radio Corporation ARC 210 VHF Comm. Equipment Instruction Manual	April 1961	part# ARC210-61-IN-C	https://www.elflightmanuals.com/ITEM_EFM/SDETAIL_EFM.asp?mID=10571
5	AN/PRC 152 Owner's Manual			https://acre2.idi-systems.com/wiki/radios/an-prc-152
6	SurvivalEvasionResistanceEscape(SERE) Operations	27 March 2017	AF Handbook 10-644	https://static.e-publishing.af.mil/production/1/af_a3/publication/afh10-644/afh10-644.pdf
7	AH-1W Training and Readiness Manual	25 July 2014	NAVMC 3500.49A	https://www.marines.mil/Portals/1/Publications/NAVMC%203500.49A.pdf
8	CH-53 Training and Readiness Manual	08 March 2011	NAVMC3500.47A	https://www.marines.mil/Portals/1/Publications/NAVMC%203500.47A.pdf
9	IEEE Standard for Distributed Interactive Simulation—Application Protocols	19 December 2012	1278.1-2012	https://standards.ieee.org/standard/1278_1-2012.html
10	Ophthalmic optics — Visual acuity testing — Standard and clinical optotypes and their presentation	2017	ISO 8596:2017	https://www.iso.org/obp/ui/#iso:std:iso:8596:ed-3:v1:en

Figure 10 – Reference View

4.9 Assumptions

The final view is assumptions, Figure 11. This view exists because unrecorded assumptions have a pernicious way of becoming integration errors

The IP and IP Radio operate as an aggregate with respect to location.
The FAC-A and FAC-A Radio operate as an aggregate with respect to location.
The FAC-A and FAC-A Sensor operate as an aggregate with respect to location.
The FAC-A and FAC-A Weapon operate as an aggregate with respect to location.
The RESCORT and RESCORT Radio operate as an aggregate with respect to location.
The RESCORT and RESCORT Sensor operate as an aggregate with respect to location.
The RESCORT and RESCORT Weapons operate as an aggregate with respect to location.
The Recovery Vehicle and Recovery Vehicle Radio operate as an aggregate with respect to location.
The Recovery Vehicle and Recovery Vehicle Sensor operate as an aggregate with respect to location.
The Recovery Vehicle and Recovery Vehicle Weapon operate as an aggregate with respect to location.
The Rescue Team and Rescue Team Radio operate as an aggregate with respect to location.
The Rescue Team and Rescue Team Sensors operate as an aggregate with respect to location.
The Rescue Team and Rescue Team Weapon(s) operate as an aggregate with respect to location.
When the Recovery Vehicle changes location, the Rescue Team changes location with it.
The starting location of the Recovery Vehicle, Rescue Team, and RESCORT does not include representation of the ship from which they launch.

Figure 11 – Reference View

5 Conclusions and Next Steps

For conceptual models to support Simulation Interoperability Readiness Levels (SIRL) as engineering evidence, we need to be able to evaluate them and compare between simulations / simulators like other items of engineering evidence. The comparison of two simulations' conceptual models currently requires more human intervention to evaluate than any other piece of engineering evidence. Our goal with MVCM is to make that human evaluation more tractable.

5.1 More Use Cases

First and foremost, MVCM needs to be exercised with more use cases and more complex use cases. The authors have received and incorporated detailed feedback from FEAT PSG meetings and briefings to MSE[1]. However, we still welcome communication about working on additional use cases.

5.2 Tool Support

As we considered the requirement for linkages between all the elements of the conceptual model, we concluded the solution is intractable beyond simple examples without a tool or tools. The proposed views illustrated in this paper were developed using a spreadsheet and a drawing tool. Considerable time was lost cross checking consistency between views.

In a proposed future MVCM tool, the user would be able to navigate between views directly or by clicking on items in the current view, e.g., clicking on an entity (source or sink) in the narrative view could take the user to the entity view for the specific entity or to the behavior view entries for the entity. This is critical to focusing on specific entities, behaviors, and triggers without overwhelming the user / stakeholder.

The operator should be able to select a subset of entities from the entity view and see the complete conceptual model between the selected entities, i.e., all the causal chains between them and none to entities not selected. The operator should also be able to select a subset of behaviors, enabling the user to review the interaction of causal chains.

The tool could have a feature to define reusable patterns, i.e., templates of entities and behaviors, enabling successive review of focused perspectives as the conceptual model is evolved and refined.

References would also link back to entities and behaviors, i.e., the operator could click on a reference and get a view of all the entities and behaviors defined in the reference.

A backing relational database is critical to maintaining all the relationships. Some of the initial population of the database might be generated from simulations' database tables, data files, scenario files, or documentation. Parsing and analysis of documentation may not be entirely feasible

[1] Formerly US Defense Modeling & Simulation Coordination Office (DMSCO)

since it's difficult and error-prone.

Returning to the challenge of comparing conceptual models, tool support could enable comparison of the conceptual model views, essentially the contents of the conceptual model database of the individual simulations / simulators. The accuracy of this comparison is directly affected by consistency of lexicon, i.e., whether the simulations / simulators have exactly the same meaning for entities, behaviors, and characteristics.

5.3 Standardization

MVCM is in the early stages and requires more development and support as discussed in the preceding subsection. However, we are hopeful that following the course of action identified will lead to a robust solution that may be suitable for standardization through SISO.

6 Acknowledgements

The OUSD(R&E) MSE (formerly DMSCO) sponsored this effort in cooperation with SISO. MSE and OUSD(Personnel & Readiness) initiated this effort to address DoD simulator interoperability challenges.

We are deeply indebted to MSE and the Marine Corp M&S Office for the joint personnel recovery use case example and for their active engagement and feedback as we developed the MVCM format and process. Without that use case, the format and process would not have been sufficiently tested in the early stages. Our thinking about the format and process evolved with every iteration on the use case.

7 References

[1] SISO, "Simulation Interoperability Readiness Levels," version 1.0, 22 September 2020, SISO-REF-076-2020.

[2] SISO, "Simulation Conceptual Modeling (SCM) Study Group Final Report," 30 June 2006, SISO-REF-017-2006.

[3] Katherine L. Morse, David L. Drake, William D. Wells, and Derek Bryan, "Realizing the Cyber Operational Architecture Training System (COATS) Through Standards," Proceedings of the 2014 Fall Simulation Interoperability Workshop, Orlando, FL, September 8 – 12, 2014.

[4] SISO, "Federation Engineering Agreements Template," 2 August 2013, SISO-STD-012-2013.

Discovering and Leveraging Emerging Technologies for Application in M&S

Chris McGroarty, Christopher J. Metevier
U.S. Army Combat Capabilities Development Command – Soldier Center SFC Paul Ray Smith Simulation and Training Technology Center
christopher.j.mcgroarty.civ@mail.mil &
christopher.j.metevier.civ@mail.mil

Scott Gallant
Effective Applications
Scott@EffectiveApplications.com

Joe McDonnell
Trideum Corporation
jmcdonnell@trideum.com

Lana McGlynn
McGlynn Consulting Group
Lana.mcglynn@gmail.com

1. Introduction

Technology is advancing more rapidly than ever and military Research and Development (R&D) budgets are constantly being reduced; however, the need for Modeling & Simulation (M&S) capabilities to support training, experimentation and acquisition, among other uses, is rising. To remain relevant and provide valuable M&S, we must leverage current computing advancements to obtain more capability for less cost. We must also be smart in our evolution to avoid becoming outdated and to be the most efficient with our time and money.

To that end, during December 2015, the Simulation Interoperability Standards Organization (SISO) established the Exploration of Next Generation Technology Applications to Modeling and Simulation (ENGTAM) Study Group (SG), which focuses on capturing technology advancements, learning from examples applicable to the M&S Community and continuing to provide a better understanding of how the latest and greatest technology can help enable all aspects of modeling and simulation. The group continues to focus on technology adoption and documentation of best practices for organizations to examine, adopt, and appropriately utilize the newest technologies for M&S.

ENGTAM transitioned to a SISO Standing Study Group (SSG) [1] in September 2017 to focus on exploring the latest technology that could be useful to the M&S Community, with a priority placed on technologies that are evolving outside of the influence of the M&S community. In particular, the rapid pace of commercial development in wearable technologies, streaming technologies, advanced hardware, cloud services and data sharing applications provides opportunity for innovative application to M&S. That said, it also provides continued concern regarding best practices for their adoption and application, as well as implications for standardization, hence the creation of this SSG. Membership is open to all and does not require membership or affiliation in any organization, including SISO, to participate. For this reason, we have a diverse group representing the full spectrum of issues related to technology adoption.

The SSG's Terms Of Reference (TOR) includes six main tasks that help focus the group's activities, discussions and path forward:

- Capture and decompose common M&S program goals
- Explore the latest industry technology trends and available solutions
- Account for security requirements
- Consider other architecture quality requirements and management requirements
- Assist the M&S Community in staying informed
- Produce a final report and deliver to the SISO committee with oversight

The SSG meets virtually on a monthly basis to discuss technologies, technology adoption, use cases and the best direction forward regarding how the M&S Community can keep up with the fast pace of technology change. Participants include members representing The North Atlantic Treaty Organization (NATO), corporations and academia from all corners of the globe. This wide variety of participants at various levels of their organization ensures a lively discussion about the topics at hand.

During these monthly meetings, the SSG hosts briefings and conversations that cover the spectrum from general concepts to specific projects. Examples include technology adoption, M&S use cases, and specific projects and technologies, such as:

- The United Kingdom's (UK) Commercial Off-The-Shelf Technology (COTS) Emerging Technology Evaluation & Exploitation (CETEE) Project [2]

- IBM Watson [3]
- Cloud computing and the applicability of cloud-first architectures to M&S
- Embedded software [4], Internet of Things [5] and integrating wearable devices [6] within simulation.
- The future of mixed reality including tool pipelines for Virtual Reality (VR) training and epidermal (devices that interact with skin) VR [7]
- Technology forecasting for Disruptive Technology [8]
- Big data [9] ecosystems with a view of how data mining [10] and data analytics [11] can lead to a data-centric decision process

2. SSG Value to M&S Community

Government and academic M&S practitioners are living in a world where technology advances are continuously occurring outside of their purview, meaning that they are not in control of how or when these technologies will be developed or implemented. Historically, new technologies relevant to the M&S Community primarily were explored and developed by the Government. Today, the majority of these developments are being driven by commercial industries for non-military and non-M&S requirements. While we in the M&S Community do not set the development requirements, we do have the ability to adapt these technologies for our use. All M&S-enabled communities can benefit from the work of the ENGTAM SSG. The SSG focus has been enlightening, educational and entertaining in the exploration of the new technologies to apply to existing needs. The basic challenges of technology exploration and adoption has many facets, and, as discussed previously, the SSG has begun to design a process to assist the user in making informed decisions. By applying the steps outlined in the Technology Adoption Activity Table (Figure 2 and the subsequent sections), we may avoid dependencies on technologies that may not have continuance, such as Google Glass [12]. Application of these steps facilitate further consideration of ways in which we adapt these technologies to M&S-enabled communities.

By functioning as a diverse body of knowledge for the M&S communities, the SSG can address issues pertaining to adoption of a specific technology. For instance, while a VR headset such as The Oculus Rift [13] may have an Application Programming Interface (API), without coordination as a community on how to interface with such a technology, every application interfacing with The Oculus Rift will be custom. In other words, each user will be inventing case specific methods to address their individual requirement thus limiting the ability for reuse across other M&S. In some cases, they may even have a dependency on a specific technology as opposed to being systems engineered to be technology agnostic. In turn, it behooves us to work together in a SSG that is all inclusive in membership as we consider how these new technologies will be applied in order to help determine the best methods to advance the art of modeling and simulation. By working together, sharing common issues, and solutions, the M&S Community will benefit from the synergy derived as the ENTGAM SSG explores the adoption of new technologies.

3. Technology Adoption

Technology adoption is the choice to acquire and use inventions or innovations, and can be viewed from many perspectives; the ENGTAM SSG focuses on the institutional adoption of technologies within reasonably large organizations. This focus is the most appropriate for the SSG participants given our professional interests (e.g., how our government or company can improve), rather than our personal interests (e.g., purchasing the latest home entertainment gadget). Some theories on technology adoption that we believe are applicable to our community are discussed below.

The Rogers' Bell Curve [14] suggests five categories of technology adopters:

- Innovators: Those people and organizations that create new technology solutions or use technology in a new way to solve problems (e.g., experimentation)
- Early Adopters: Organizations that have a unique enough problem and enough resources to attempt a new technology solution (e.g., research organizations)
- Early Majority: Organizations that require proven solutions for their problems (e.g., have less resources to take risk with)
- Late Majority: Risk adverse organizations that have specific, well known needs and have a well-defined schedule (e.g., Programs of Record)
- Laggards: Organizations with existing products that work well enough and that are used by a large group of users so adopting new technology will be more painful than helpful (e.g., Microsoft Office Suite)

The Technology Acceptance Model theory [15], provides an explanation of how users accept technology. Usefulness and perceived ease of use (level of risk) drive

the attitude towards using innovation. We believe that a very similar paradigm is applicable to large organizations adopting technology. There is a similar alternative for individual users, called the Hedonic-Motivation System Adoption Model [16], worth further consideration within the M&S community.

The Hype Cycle [17], shown in Figure 1, attempts to visualize expectations over time for new technologies and the companies managing those technologies. The general principle of the hype cycle is that technology expectations rise quickly as they begin to show promise. Expectations rise too high until early adopters actually get experience with the technology and realize its limitations. These adopters are quick to lower expectations of the larger populace. Over time, users understand better how to leverage the technology for its benefits while comprehending and working around any limitations.

Figure 1 Hype Cycle

Adopting new technologies can interfere with mission objectives, timing, processes, etc. Since it's inception, the ENGTAM SG, and follow-on SSG, has reviewed how the US Military adopts new technology, including processes suggested by Defense Acquistion University (DAU), the Government Accountability Office (GAO), and other sources along with pros and cons of the various approaches. We also reviewed the "Open Source Software Selection Process and a Case Study" [18]. Based on insights from these reviews and our own personal experiences, we developed best practices with regard to technology adoption.

4. Best Practices For Technology Adoption

The best practices that the SSG has identified are targeted to a wide breadth of potential use cases, including government research projects, government programs of record, large and small companies, and technology for both internal and external users. Internal users are people within the organization that use technology to accomplish internal tasks even if these tasks may be outward looking. External users are people outside the organization, like customers, that use the technology as they interact with the organization.

Step	Current State	Exploration	Evaluation	Adoption	Management
Activities	• Organization goals • Organization strategy • Organization capabilities • Organization needs / weakness • Use Cases • Systems Views	• Identify capability areas • Identify technical areas • Plan, staff and schedule exploration activities • On-going monitoring	• Systems engineering for how technology will integrate (technical and process) • Testing (component and system) • Pricing (full life cycle) • Scheduling	• Backup Existing Solutions / Data • Partial / Full Replacement Strategy • Training • Installation • Integration • Data Migration • Process Adjustments	• Maintenance • Upgrades • Process Refinement • Integration Adjustments

Figure 2 Technology Adoption Activity Table

The SSG has divided the technology adoption process into five phases, as shown in Figure 2. These five phases are designed to encompass the spectrum of understanding the organizational goals, discovering new technology that is applicable to the mission, adopting that technology, and maintaining it over the relevant lifecycle. They are:

- Understanding the Current State: Knowing the organization goals and current environment is the first step to understanding how any new technology may be of benefit.
- Technology Exploration: A concerted effort to finding new technologies, new products, and what is coming in the near future.
- Technology Evaluation: A process for learning more about technology, including capabilities, integration points, pricing, maintenance, etc.
- Technology Adoption: Bringing a new technology into an organization, whether replacing an existing technology or bringing a new capability to an organization.
- Technology Management: Maintaining, adjusting or removing technologies within the organization.

5. Understanding The Current State

This phase was influenced by the SSG's struggle to apply the technologies we were discovering to M&S projects and/or organizations. We found several great technologies with seemingly useful capabilities, but it was difficult to say who (organizations and projects) within the M&S Community would benefit from that technology without having knowledge of what the projects and organizations were doing already.

Discovering, understanding, acquiring, and applying new technologies would be greatly enriched if one knows what the functional objectives are and how those objectives are currently being accomplished. This understanding should be based on the organization's strategy going forward, both for the business model as well as the technology, and how technology fits within the organizational strategy. The organization's leadership

should agree on and document the goals (its mission), strategy (how it will achieve that mission), current capabilities, and needs before assessing how any technology can be used by that organization. These should be written down in any format the organization chooses (e.g., use cases, systems views, etc.). It is important that the current state and the strategy for moving forward are agreed on by those leading the organization, executing the current mission, and pursuing new technology on behalf of that organization. The use cases and/or systems views should include how users interact with the technology, the organizational processes, how the technology is integrated or connected, and how the technology is maintained and upgraded.

It is important to approach the discovery process with an open mind. The current state of the organization should not drive the technology exploration but will, however, provide a foundation for whether what is found would be worth further exploration. The current state can also be viewed as a tool to focus the technology discovery and evaluation. As new technologies and capabilities are identified, they can be assessed based on the organization's goals for applicability utility.

6. Technology Exploration And Discovery

Once the current state and the future goals are known, the next step is to find out what technologies exists, or will exist soon, that can help meet these goals. This starts with learning about and staying abreast of the relevant industries that can provide the target capabilities. Knowledge about what technologies exists in the marketplace is the foundation for discovering new and emerging technologies and discerning their advantages over existing and competitive/similar technologies. Without broad market research, you may choose a technology that is not optimal for your needs. While this may sound obvious, too often the first technology that seems relevant is chosen rather than the on that would truly have the most benefit. Technology exploration is best suited for somebody already familiar with that specific technology's constituency, the current implementation and the state of the marketplace relevant to the technologies useful for the organization.

Furthermore, technology exploration and discovery should be an ongoing and iterative effort. True exploration of the marketplace is more than reading magazines and going to conferences. Conducting this phase as a research project with ongoing tracking will provide maximum benefit. This phase concentrates on identifying both areas of capabilities to focus on, based on organizational capability assessment, and technical areas that relate to the organizational capabilities defined in its current state. It is critical for the technology exploration effort to be planned, staffed, and scheduled. This is not a trivial task, but rather a time-consuming, detail-oriented task that is important. The iterative nature of this effort will allow for monitoring of progress within the commercial industries of interest as well as within the organization, and how the organiational goals and capabilities progress over time. Maintaining detailed and up-to-date documentation is key to success in this phase.

7. Technology Evaluation

Once a technology has been identified as potentially useful to an organization, the next step is to assess the technology in detail, including: how the technology would be employed and integrated with other systems; its robustness; pricing; and, availability.

Systems engineering, to include an assessment of the organization's processes, should assist in understanding how candidate technology would fit in the current processes and integrate with other technologies. This step should be treated as if it were a new development effort with appropriate systems engineering artifacts, process, and reviews to ensure that all the systems within the organization operate appropriately for the organization's goals.

Before adopting the new technology into the organization, it should be tested at both a component level and System of Systems (SoS) [19] level. Testing of the new technology can be done with or without a vendor. In some cases, the vendor may not allow for trials, tests, and usage before purchase, but in the case that it is an option, testing should be conducted in an appropriate environment within the organization for internal, unbiased results.

Cost is an important consideration across the entire expected lifecycle. The cost is more than the purchase/license price, but also includes how much it will cost the organization to incorporate, maintain, and eventually remove the technology. Cost examples include: training staff; technical administration and maintenance; ongoing licensing or support; integration; and, additional consideration throughout lifecycle of the technology from the organization.

Scheduling planning needs to include the resources required for systems engineering, procurement, installation, testing, training, the learning curve of the users, and maintenance. New technologies, and their adoption within organizations, can be fraught with risk due to the uncertainty of how the new technology really works, the often immature nature (due to it being the latest and greatest), and the perception of how it works due to marketing material instead of actual experience. Risk factors resulting from the introduction of a new technology can be reduced with a strict and thorough evaluation process.

8. Technology Adoption

Adopting a new technology introduces varying levels of agitation within an organization depending on its use. It is important to understand how users interact with the new technology within the context of the organizational processes, outreach, and management. Considerations for users should be for both users outside the organization (e.g., paying customers) as well as users within the organization (employees). Bringing in new technology will require proper scheduling (e.g. system down time, training, etc.), come with expense, and will inevitablity have some unforeseen challenges.

From an execution perspective, engineers should plan for and implement backups, schedule down time with users, install the new technology, and migrate any appropriate data from old formats to match what is required. Process adjustments may need to be made, which should be driven by the systems engineering conducted during the technology evaluation. It may make sense to execute a partial adoption when organizations have critical real-time systems or a large number of users. Partial implementation split across users could also help mitigate risks when problems occur.

Training and the resultant learning curve, which may potentially creating a decrease in productivity, should be considered, anticipated, and planned for accordingly. In some cases, the new technology involves a modification of processes, new user interfaces, new data, or other items that required users adjust to a new environment. Careful consideration should be paid to creating "easy to use" interfaces, user guides, conducting training, and in some cases, availability of the old system for a limited amount of time in case there are issues, or critical needs that cannot be disrupted.

9. Technology Management

Technology management includes: maintaining, upgrading, and improving the technology; how it is used; and, how it is integrated within the organization. Ongoing maintenance is different for each technology. How the new technology differs from the old technology could mean that the maintenance requirements differ in terms of time, money, effort, and cost.

Most technology providers disseminate upgrades as they become available. How and when an organization upgrades from one version to another depends on what new capabilities are provided with the upgrades. Upgrading may also have a transition period where the system may be unavailable for a period of time. These periods' length and complexity issues (e.g., data migration, integration points, etc.) are specific to the technology, data, and organizational usage. These issues must be considered and will be addressed in any best practices recommended by the SSG.

Organizations should also consider improvements and adjustments based on how the technology is used and integrated with other technologies within the organization. In order to determine the return on investment, data should be collected (as possible) to quantify cost avoidance (both time and money), or any other improvements that the new technology brings to the organization. As the technology improves or how it is used changes, the systems engineering artifacts from the first step (understanding the current state of the organization) should be updated to reflect a new current state based on revisions. The entire process is iterative and continual given technology will never stop improving. Periodic reviews to assess the technology and identify lessons to be learned should be conducted.

10. Summary

The M&S Community appears to be a natural candidate to benefit from new technologies; however, it is difficult to focus broadly on specific technologies without well-defined, persistent use cases. That said, we have found that frequently well-defined usage needs have an immediacy that cannot wait to follow a lengthy process for the evaluation and adoption of new technologies. Conversely, persistent use cases we have found are described at too high a level to inform technology decisions. The rapid pace of the introduction of new technologies into the marketplace, and the mismatch in

technology provider and technology consumer product cycles, lead the SSG to the conclusion that our work should not result in the development of a new standard at this time. Over the last few years, the group has pivoted from only broadly evaluating technologies to also describing best practices for technology adoption, which could have an immediate benefit to the M&S Community. To this end, the SSG has developed the five-step process, shown in Figure 2, that includes: understanding the current state of the technology, technology exploration, technology evaluation, technology adoption, and technology management. To orient the process, we start with the current state as the foundation with the future or desired state as the objective. The SSG is actively soliciting input to formalize this process and its implementation from the larger M&S Community. The authors encourage participation in the ENGTAM SSG activities by anyone who is interested.

11. References

[1] C. McGroarty, C. Metevier, S. Gallant, J. McDonnell, & L. McGlynn. "Are We Progressing?: Exploration of Next Generation Technology Applications to Modeling and Simulation Study Group Update. Simulation Interoperability Standards Organization (SISO)" Simulation Innovation Workshop (SIW) 2016.

[2] Defence Science and Technology Laboratory (DSTL). Commercial Off The Shelf (COTS) & Emerging Technology Evaluation & Exploitation. http://www.cetee.org/Documents/CETEE%20Info%20Sheet.pdf. 2016

[3] IBM Watson Product Information. https://www.ibm.com/watson/

[4] E. Lee. "Embedded Software". Advances in Computers, Vol. 56, Academic Press, London. 2002.

[5] O. Vermesan & P. Friess. "Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems". Aalborg, Denmark. River Publishers. 2013.

[6] S. Ryan. Garments of Paradise: Wearable Discourse in the Digital Age. MIT Press. 2014.

[7] J. Rogers, et al. "Epidermal Virtual Reality Devices", Patent application Publication. Pub. No. US 2019/0369728 A1, December 5th, 2019.

[8] Brackin, R.C.; Jackson, M.J.; Leyshon, A.; Morley, J.G. Taming Disruption? Pervasive Data Analytics, Uncertainty and Policy Intervention in Disruptive Technology and its Geographic Spread. ISPRS Int. J. Geo-Inf. 2019, 8, 34.

[9] S. Sagiroglu and D. Sinanc, "Big data: A review," 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, 2013, pp. 42-47, doi: 10.1109/CTS.2013.6567202.

[10] X. Wu, X. Zhu, G. Wu and W. Ding, "Data mining with big data," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107, Jan. 2014, doi: 10.1109/TKDE.2013.109.

[11] Amir Gandomi, Murtaza Haider, "Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management, Volume 35, Issue 2, 2015, Pages 137-144.

[12] T. Starner. "Project Glass: An Extension of the Self. IEEE Pervasive Computing", Volume 12 Issue 2. 2013.

[13] Oculus Rift Product Information. https://www.oculus.com/rift/

[14] E. Rogers. "Diffusion of Innovations" (3rd ed). New York: Free Press of Glencoe. 1983.

[15] F. Davis, R. Bagozzi, P. Warshaw. "User Acceptance of Computer Technology: A comparison of Two Theoretical Models." Management Science, 35: 982-1003. 1989.

[16] P. Lowry et al. "Taking 'Fun and Games' Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM)". Journal of the Association for Information Systems, vol 14(11), 617-671. 2012.

[17] J. Fenn, M. Raskino. "Mastering the Hype Cycle. How to Choose the Right Innovation at the Right Time". Harvard Business School Press. 2008.

[18] G. He. "An Open Source Software Selection Process and a Case Study". Thesis for Office of Graduate Studies of Texas A&M University. 2007.

[19] M. Maier. "Architecting Principles for System of Systems". Systems Engineering. Vol 1, Issue 4, Pages 267–284. International Council on Systems Engineering (INCOSE). 1998.

THEME 2: Training and Education

PANEL SESSION

Applying and Utilizing New Technologies in Training-Challenges and Limitations from a Trainer's Perspective

Written by: Mr Wolfhard Schmidt (LTC ret. DEU A)
Senior Manager Strategy&Products Ewwol Solutions Ltd (Bydgoszcz, Poland) for ST Engineering Antycip SAS
Moderator of the session

Participants:

Col (US A) Heath McCormick
Director of the Joint Multinational Simulation Center (Grafenwöhr, Germany)

Mr Thomas Lasch
Simulation Strategist at 7 US Army Training Command (Grafenwöhr, Germany)

WO II (DK A) Per Klembo
Project Officer Simulation Branch at the Danish Artillery Regiment (Oksbol, Denmark)

Maj (BEL A) Kurt Vanderheyden
Commanding Officer of the Sim Centre Land of the Competence Centre of the Belgium Army (Limburg, Belgium)

Col (US A) John Ferrell
Director of Simulation of the US Army Aviation Centre of Excellence (Ft Rucker, USA)

Cmdr (DEU Navy) Jörg Feldhusen,
Staff Officer Education and Training at the NATO CoE for operations in confined and shallow waters (Kiel,Germany), Chairman MORS subgroup of the NATO M&S Group.

1. Introduction

Industry and science are continuously improving existing and/or developing new technologies to make training more efficient. New technologies definitely provide new capabilities to train what currently can not be trained. Industry also promise, among others a better, faster, more realistic and more cost efficient training when utilizing their new products and/or the newest version of their application. This is the shiny world of product marketing which is widely know but does not necessarily reflect the reality in the simulation and training centers. Applying and using new technologies in current day to day training is a permanent challenge and struggle for the user. So question is whether all these new products are needed? Can they effectively be used in the day to day training routines? What is preventing the trainers from using new technologies at their centers? The intent of the panel was to help a better understanding of the challenging situation and identify possible changes and ways to overcome the current bottle neck from the user point of view but also what should be done from industry to get new products better introduced and utilized?

The panelists were experienced training experts from different nations.

2 Discussion

The following key aspects were identified and discussed:

<u>The technology aspect</u>

Not always will new technology bring more add-on to a particular training but was nevertheless procured (Coolness vs. usefulness). To integrate new technology on existing IT platforms and federate them with the current legacy systems triggers huge interoperability challenges which needs time and resource to get solved. The rolled out products do not always keep what was promised (not fully finalized, missing functionalities etc.) which triggers extra costs and time to overcome it.

<u>The training principles aspect</u>

Current training principles and philosophies (e.g.; class room focused training, training at Simulation Centres only) do not reflect the new possibilities provided by applying new technologies such as distributed training, training at the point of needs vs training in particular centres.

<u>The procurement aspect</u>

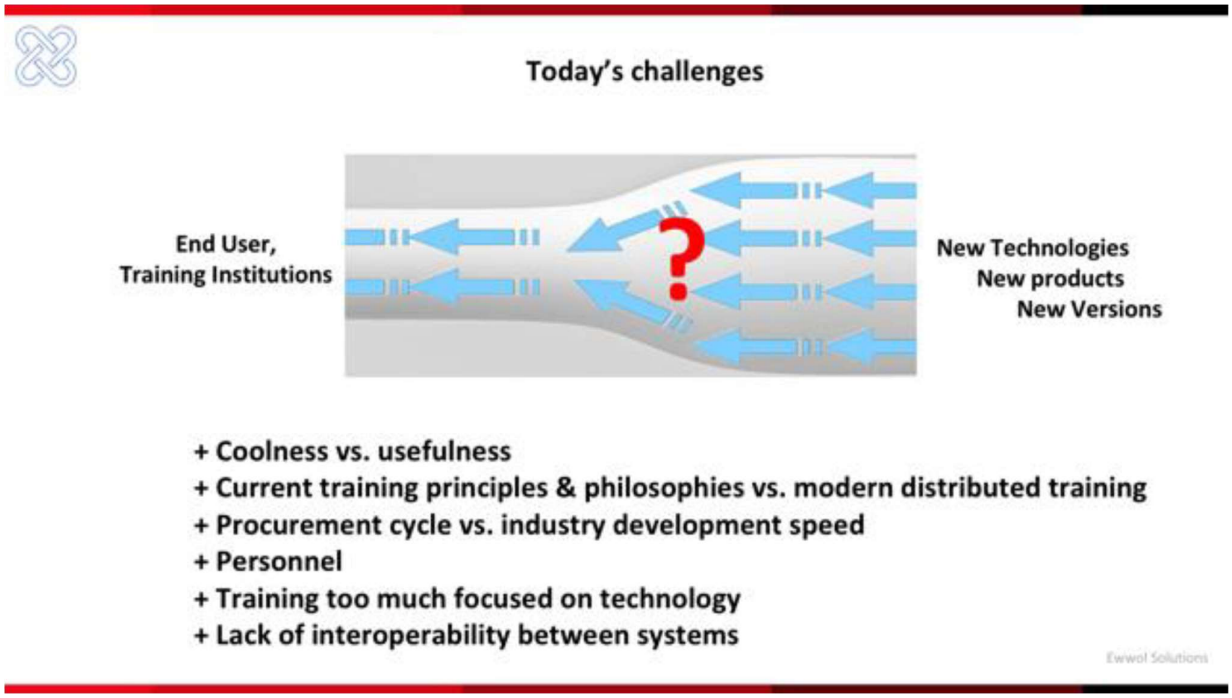
The current procurement procedures are too complicated and last much too long to keep up with the development speed by industry. The result is that the user gets the new systems not as state of the art anymore (already out dated when installed) and too late for supporting newly emerged training needs.

The organizational/procedural aspect

New products need to be validated and acceptance tested once rolled out at the training facility. The organizational structure of the training centres does not reflect this task. The result is that there is no specialized personnel for this task available or training personnel needs to take away from the main task of the facility - providing training. Even with new technology the exercise set-up and train-the-trainer-phase is still complicating and very time consuming. Currently there is a lack of specific direction and guidance from the national/multi-national command level on how and for what new technology shall be utilized (Top down approach to provide consistency of the user requirements). The current Armed Forces OP Tempo is mainly not in sync with industry's business case approach (Apply Information Age technology to the Industrial Age process). Currently there are still too many island solutions applied as well as nations still keep focusing too much on its national needs and particularities especially on the 'one size fits all' aspect when applying new technology solutions.

The human aspect

Even with new technology available is the training audience not always willing to challenge themselves to the end (not or very limited existing failure culture). This attitude is also supported by the supervising leadership in exercise and training events.



Today's challenges applying new technologies

3 Conclusion and Key findings

As a possible way to overcome the identified challenges and the existing bottle neck when applying new technologies in training the following key findings were proposed:

➢ Need of a mindset change of the training audience and the military leadership supervising training and exercises (e.g.;new failure culture in training, to train the reality and not to please Commanders or once careers)

➢ Re-think the way how training is set-up and executed (do we really need to own the training centres or can it be out sourced; training at the point of needs and not only at training facilities)

➢ AI will be a game changer (e.g.; for freeing up personnel in EXCON, faster and more detailed training and exercise assessments)

➢ Training framework parameters need to be more flexible (accreditation process of new technology on existing platforms; faster procurement cycle; IT infrastructure which is able to support new technologies; earlier involvement of industry to avoid problems during the roll out phase)

➢ Applying new approaches for training support such as Training aaS and M&S aaS in a more holistic manner , not only technology focused

➢ Specify the military requirements more precise and in a consistent top down approach to ensure a more efficient and smoother utilization of new technologies on time

➢ Focus more on the particular echelon to be trained rather a 'one size fits all' but nevertheless sharing the same training support core such as terrain, weather and ORBAT data (new synthetic training environment approach).

The panel closed with the following plea to all parties involved in training:

"Training always needs to be state of the art to ensure that the warfighter will be trained for all challenges he might face. This is needed not in show rooms of the industry and in the Battle Labs but is needed for those who provide the day to day training support."

Simulation course – Action research approach

Kalle Saastamoinen, Antti Rissanen
Department of Military Technology, National Defence University, P.O. Box 7, FI-00861 Helsinki

Abstract

Simulation model is an imaginary bridge from theory to reality, its usefulness is defined how well it can mimic real world phenomena it is simulating. There are myriad of different models that can be used for simulation purposes and to select and tune the one that works is a craftsmanship that can only be learned by years of practice with different fields of applied engineering. For students mastering to do actual simulators can only be learn by doing.

Action research is here seen as a cyclic learning process where students select the problems which they want to study, they can also deliver they own problems to be studied. After the selection, students make studying groups from 2-4 persons and start to apply modelling techniques they are learning during the simulation course.

Co-operation between studying groups is very social, free and goal oriented. Students are finding solutions mostly by trial and error kind of procedure, which involves planning, acting, observing and reflecting. Finally, at the end of the course they will present their finalized works in the seminar, where their solutions are under public evaluation.

Keywords: simulation, learning by doing, multidisciplinary

1 Introduction

Simulation itself is a multidisciplinary branch of science, where technical, scientific, economic and humanistic skills are needed in order to create a system that mimics reality as well as possible. At the National Defence University (NDU), students' varying knowledge levels, attitude and also motivation towards science and engineering bring their own demands how simulation course should be carried out in order to achieve the best possible learning outcome.

In military technology education, lectures and exercises are aimed to present scientific methods and mathematical tools that support professional practice and may later be applied in a military officer's profession. Key issue here is

the fact that officers need to be able to apply theoretical knowledge in practice. In general, military training, planning and design include multiple ways of preparing students so that they reach the needed readiness and skills. Due to the fact that there is a huge variety of battlefield scenarios and there are many ways of utilizing high-tech weapons, the battlefield environment needs to be analyzed with sophisticated tools. [1] Quite often demands for such simulations are related to the visualization of specific data with a suitable high-level toolset [2].

For simulation and modelling purposes, commercial, academic, and even in-house software have been developed. Some of them are highly specific while other software can be tuned with a set of tuning tools from the developers' library. For generic widely utilized we name Matlab [3], Vissim [4] (has been rebranded and embed as a part of solid Thinking's Development Suite) and Berkeley Madonna [5]. On the other hand, traditional Visual Basic for Applications (VBA) has been widely used in mathematical modelling and simulation as a general or introductory tool for studying the features of any presented problem [6]. Moreover it is available for most users using Windows environment. Therefore, the VBA tool was selected to be used in the simulation course at NDU.

In this paper we will show that students' feedback towards teaching methods used in simulation course have been positive year after year. Military students have learned simulation methods, and later applied them in their working environments and in their thesis. One of the reasons for positive attitude is the fact that the approach in our simulation course has been practical and oriented towards modelling techniques. Aside of that students have also improved their social skills doing different kinds of group works. The simulation course that has been presented in this review paper can be recommended to any nature science students, because it offers a real linkage to the practical applications. Normally military officers learn practical skills, while in natural sciences learning is often lead from theory oriented needs. Therefore, simulation is a natural "bridge" in between theory and practice. This paper is organized as follows:

- The second chapter of this study presents a short theory behind the method called action research and how it is connected to the way the Simulation and Modelling course has been carried out in NDU.
- The third chapter presents two group work cases and what kind of knowledge students did need in order to success in these simulation tasks.
- The fourth chapter presents evaluations of the course from the students' viewpoint, with supplementary instructor comments.
- The fifth chapter presents conclusions and the future of teaching the subject and also experimental views.

2 Action research with a working example

The focus of action research is on influencing action and engaging the participation and participation of researchers in the everyday life of the training organization. Attendance is combined with subject analysis and influencing it. Kemmis and McTaggart emphasize that in reality the process may not be as straightforward as sequential parts of independent design, operation, observation and reflection. [7] The described modules may overlap and the original plans may become obsolete based on experience and new information. In [8] O'Leary's view of the model activity-related research is seen as an experiential learning approach, with goals including refining the needs of the methods, knowledge and interpretations based on the understanding of previous cycles see Fig. 1.

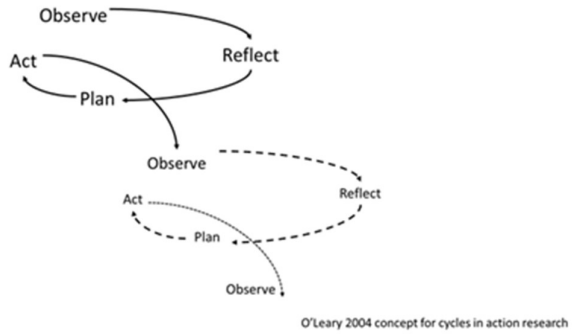


Figure 1 – FMN Ready Force Requirements [7]

Case study is one of the ways to conduct research - it might be called a qualitative research approach, but case study concept is not synonymous with qualitative research.

The case study approach is not in itself a research method. It is difficult to give a general or comprehensive definition of a case study because there are different case studies. In this work it is used as a research strategy for observations. Also it is utilized as a label to separate presented sample student works.

Master's level students majoring in NDU's technology program have different instructional needs. Therefore, the current study aims to explore how a relatively practice-oriented simulation course would affect students' motivation and attitudes towards physics and technology as professional tools. Specific observations and data are from the Simulation and Modelling Course (SMC). The instructional structure of this course consists of three overlapping teaching methods: lectures, supervised exercises, and unsupervised exercises. At the end of the course student groups present their unsupervised exercises. The final report consists of documentation and the functional version of the group's own specific simulation realization. Up until now this course has been carried out nine times with only slight modifications. From 2020 onward the whole master level education has been revised, but the context of this specific SM course has been included into the new curriculum.

3 Two students' projects: a flight simulator and Helsinki-Vantaa airport border inspection procedure

More and more training in Finnish Defence Forces include preparation to field practices with simulated exercises. Also the analysis of exercises done "outdoors" have tools based on simulations. The profession of an officer involves the utilization of advanced technological artefacts (e.g. weapons, weapon systems, or supporting systems). To understand better what kind of tools has been taken into conscripts' training also view behind the tools surface is needed. To make the simulation tool more operative, it was necessary to find a simple set of requirements for the simulator for the first stage of implementation. Once the tool was good enough, a further requirement was that the usage of the simulator had to be easy enough for an average military officer, so that officer would be able to use it after a few minutes of instruction.

In the first case (case 1), future fighter pilots tested how well they could simulate some functions of their future (real) training jet plane. Mainly the hidden question is, has the given physics education relevance to be applied into practice. As for the effort the issue is on tedious and careful code writing before getting and applying right parameters. In the second case study (case 2), another group presents how statistics and queueing theory can be utilized to model Helsinki-Vantaa airport border inspection procedure.

3.1. Case 1: a flight simulator

The first use of technology to simulate flying was so called "Blue Box" developed by Edwin Link in the 1928. Even though it had huge limitations especially in all visual appearance, this flight simulator was used to train thousands of aviators. [9] The first implementation in our study (our case 1) is a fully working PC simulator (covering limited functions) of training scenarios of the Hawk airplane. Generally in education, students face difficulties in studying motion and related issues [10]. In the first case (case 1), the students utilized earlier learning experience and produced a fully realistic simulator with certain functions. Before getting to practice with their own simulator, students browsed their aerodynamic learning material in order to list all relevant physical phenomena and listed them:

- An upward thrust (wing, elevator and tail), where we need the following inputs: a) affected areas, wing configuration, speed, Mach, air density, flying angle etc. b) the decrease in the upward thrust after stalling and the movement of the pressure center towards the back of the plane. c) the movement of the pressure center in the transonic area, the head down effect.
- Resistance (structure, wings, tail): a) Induced and parasitic. Transonic resistance. b) Same inputs.
- Engine: a) Produces thrust even when idle. Thrust will change as a function of the Mach number, air density, and rotational speed. b) Modelling the nonlinear dependency between thrust and revolution.
- Flaps, brakes and landing gear: a) Flaps have two positions. b) Modelling of the landing gear, consider frictions and shock absorption. A hard landing will break the airplane.
- Wind: when height increases, blusters affect the direction and may change the speed.

To get a well-behaving simulation that would correspond to a real-world artefact, the correct parameter values for the specific airplane were needed. Luckily, the students were able to gain this information. In order to estimate how well their solution worked, students made a comparison between the manufacturer's data and the performance given by their simulator. Table 1 is a direct translation of the students' course report (Table 1).

Test	Simulator	Reality
Time taken to ascend to 11 km (weight 5000 kg)	10 min 50 sec	10 min 50 sec
Max horizontal velocity and G-value	M 0,81 and G 5,0	M 0,805 and G 5,5
Stalling speed (Load/ weight 5000 kg)		
Level flight	120 Kias	124 Kias
Takeoff	109 Kias	111 Kias
Landing	100 Kias	105 Kias
Max Banking force at Level Flight Speed		
@ 10 000 ft	M 0.84 G 3.8	M 0.845 G 4.0
@ 30 000 ft	M 0.86 G 1.9	M 0.854 G 1.95

Table 1. Comparison of the simulation and the airplane's factory data, M (Mach speed), G (relative acceleration), and Kias (Knots in air speed).

To test the touch, students added a real-time control interface to fly the simulator and follow procedures from the simulated cockpit view. Fig. 2 is taken from the course report and it illustrates how the simulator's graphical interface on the right imitates the real cockpit view seen on the left-hand side of the picture.

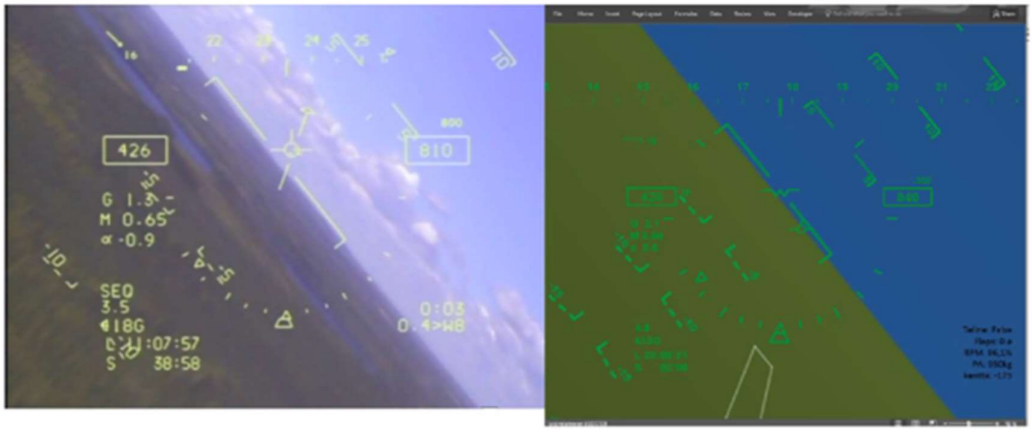


Fig. 2. Comparison, real cockpit view on the left and the students' simulator on the right.

3.2 Case 2: Helsinki-Vantaa airport border inspection procedure

Helsinki-Vantaa Airport has extensive flight connections and the shortest routes between Europe and Asia, making it as one of the major hubs for Northern Europe. When planning for a smooth running of the border control process it is good to know that many of the passengers pass through the airport only to get a continuing flight. At the experimenting time (before the Covid-19 pandemic) the amount of external border traffic at Helsinki-Vantaa Airport was estimated to increase steadily.

For this environment description, our students used simulation to show the real effects of alternative conditions and courses of action. In order to establish relevant simulation model, they used queueing theory that is the mathematical study of waiting lines, or queues [11]. This exercise work can be considered to be part of operations research since the simulation results are used to make decisions about the resources needed to provide a service. Main research question in the study was how well can Helsinki-Vantaa Airport answer future challenge with increasing number of passengers and can these new demands be handled by using cooperation-based data processing with different participating operators? Where main benefits are avoidance of overlapping information and the use of automated border checks in entry and exit.

Border inspection process can be described by the following flowchart:

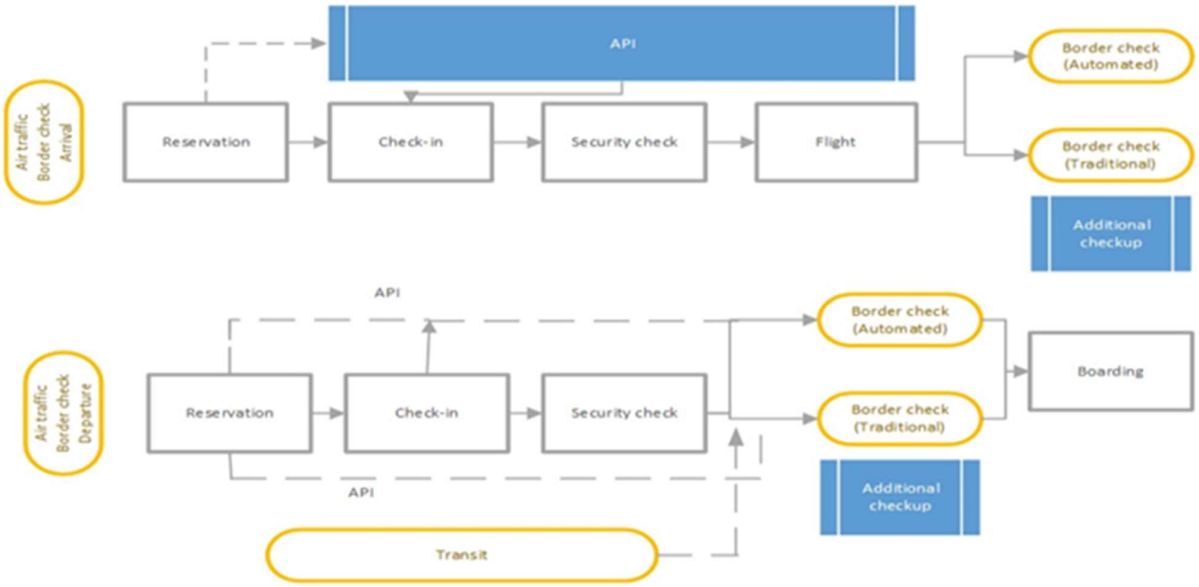


Fig. 3. Process flowchart of the passenger's border crossing.

In order to define correct parameters for the simulation model students did observations on real inspection times at Helsinki-Vantaa Airport. Test was carried out between 19.2.2018 – 21.2.2018. To get necessary data, only those customers who were coming outside Europe were taken into account. Sample size was 50 persons in each phase and persons were selected randomly.

For the simulation purposes students defined measures that would give them information they needed in order to simulate the border inspection process. Key parameter for outcome is how long time it takes to go through border inspection. Simulation model helped to find critical points from the process and gave valuable insight how time critical border checkup time is with respect to resources available. Simulation here models multi-phase process, from where one can measure how number of incoming and exiting passengers affects to the queue length. Using this simulator, it was tested how much sharing information between officials and use of

automated border checkup could boost up passenger's flowrate.

Border checkup is a classical queue process. Phases are arrival, waiting, service and exit. In practice the bottle neck is waiting time in queue. The effectiveness of service system needs to be evaluated. Service time distribution is the key factor in the fluent working of the queue. In the simulation model the following items are included in to the working model: 1) Number of customers in the queue 2) Number of customers in the whole system 3) Average time customers wait in the queue 4) Average time of the border inspection 5) Average time in the system. The generated model simulated the timing of the queue length by calculating passenger arrivals and exits at border inspection. The user interface included setting the flight fill rate. At beginning of each calculation the waiting time for the first incoming flight queue was assumed to be zero. The simulation was continued until the last flight had arrived and the passengers had passed the inspection. In this case, the observation from a single simulation time was the average of all wait and wait times for all the passengers in that simulation run.

Fig. 4 illustrates the results of the simulation with the current total time control (169.70 seconds) versus in case of automatically changing the passenger data (64.01 seconds). The simulation showed that with the selected parameters, the number of inspectors decreased considerably if the data already collected by the air carrier were automatically transferred between the information systems. The time spent on border checks may also be increased within the limits permitted by the airport's border inspection capability, then the number of border inspectors may be a variable factor..

ARR	Number Inpectors (169,70)	Number Inpectors (64,01)	Exchange time min	ARR	Number Inpectors (169,70)	Number Inpectors (64,01)	Exchange time min
0:25	10	4	0:27	14:40	49	17	0:29
0:30	23	9	0:24	14:40	50	18	0:26
0:55	24	8	0:30	14:45	50	20	0:26
0:55	23	10	0:28	14:50	55	20	0:28
1:40	17	7	0:25	14:50	54	20	0:26
5:45	23	9	0:25	14:50	42	20	0:23
6:25	23	9	0:25	15:10	88	20	0:32
6:55	13	7	0:30	15:30	39	20	0:30
6:55	23	8	0:26	15:35	45	20	0:30
7:00	23	9	0:27	15:35	60	20	0:25
7:00	29	9	0:24	15:20	60	20	0:26
7:10	28	9	0:24	15:20	56	20	0:22
7:55	20	8	0:25	15:25	52	16	0:24
8:50	24	9	0:24	15:50	4	5	0:33
10:20	7	3	0:25	15:50	22	11	0:28
10:50	23	9	0:24	16:20	14	5	0:25
11:40	11	4	0:25	17:40	17	7	0:25
11:55	23	9	0:24	18:05	14	6	0:24
12:20	17	7	0:25	20:05	13	5	0:26
12:35	11	5	0:24	20:25	8	3	0:27
13:10	8	3	0:27	21:00	17	7	0:25
13:30	8	3	0:29	22:05	12	5	0:24
13:45	24	10	0:29	22:35	8	3	0:26
13:50	40	14	0:27	22:30	8	3	0:27
13:55	40	14	0:24	23:00	18	8	0:25
14:10	35	10	0:27	23:05	18	6	0:24
14:10	30	15	0:28	23:25	15	6	0:24
14:15	24	10	0:25	23:35	28	9	0:29
14:20	35	15	0:24	23:35	36	11	0:24
14:25	37	16	0:24	23:40	37	12	0:23
14:30	48	17	0:24	23:55	12	6	0:26

Fig. 4. Simulation result of automatic transmission of data on the flow rates of passengers.

	Present	Automated (s) (change %)
In-depth arrival inspection	169,70 sec	64.01 (- 62.28 %)
Phase 4 - Departure (API), conveyance, identifier (API) and Phase 6 - Destination (API), Travel plan Passenger Name Record (PNR).		
In-depth boarding inspection Phase 5 - Destination (API), conveyance, identifier (API)	43.16 sec	41.05 (- 4.89 %)

Table 2. Effects of the use of passenger information to the inspection-time.

Table 2 shows the effects of the usage of advanced passengers' information (API) to the inspection-time. We see significant performance boost when automated data transfer is utilized. This model built during the course gave grounds for evaluating performance, explaining the flow efficiency of the process and its impact from the point of view of border authorities. It helps allocate resources and use them to achieve the overall quality of services sought.

4 Students' evaluations with comments

To present most valuable items from the standard student evaluation of teaching questionnaire we selected answers to the questions: 1) Were you active during the course? 2) Was overall ambience during the course supportive to your studies? 3) How well lecturers mastered subject matter of the course? 4) How well the ratio of course demands suite to the academic credits of this course? 5) Course gave new information/skills? 6) Evaluation of this course supported my learning?

The 5-step Likert scale is a psychometric scale. Respondents tick the box which represent their attitude, opinion, or even feelings about a particular issue. When data are combined with qualitative data like open-ended questions, participant observation, and interviews, the survey's validity is improved and its information becomes more concrete. [12] Those conclusions that are reached with instruments that gauge attitudes are only as good as the quality of the method. Even though items on these scales may have numbers assigned to each level of agreement, it cannot be assumed that these numbers represent equidistant units that can provide the interval-level data necessary for parametric statistical procedures. [13]

The collected data is summarized in figure 6. The amount of students giving feedback were in year 2016 ten, in year 2017 it was nineteen and year 2018 it was 22. From this figure we can see some steadiness of the selected course form. Especially it is nice that in the open ended question section, students' see that unsupervised exercises supports their learning procedure. Open questions and further discussions with students point more to resources than contents or tools, e.g. the allocated working hours for the whole course could be more flexible. Overall, positive feedback for this course showed that students appreciate assignments which require inputs like creativity and deep concentration and which nevertheless connect their daily practice and duties.

We can see that feedbacks have a little increasing tendency, but as we know NDU students are very sensitive for teacher's general appearance, that is use of voice, punctuality, clothing and attitude. Basically, one flu week will affect the feedback, which is annoying of course. The fact is that students in NDU have huge differences between knowledge of science and engineering and also motivation can be problematic. From the Table 3 we can see that students' attitude towards teaching-learning environment is highly dependent of the way they have learned to study.

From literature we see that the results here are satisfactory well analyzed especially for such small amount of answers. Parpela et al. made a study [14] were divided factors measuring the students' approaches to learning to three classes that are 1) Deep approach 2) Organized studying 3) Surface approach. Thereafter in included factors measuring students experiences of the teaching–learning environments according to six classes, which are 1) Teaching for understanding 2) Alignment 3) Staff enthusiasm and support 4) Interest and relevance 5) Constructive feedback 6) Support from other students. Table 3 shows the ESEM estimated correlations between students' scores on the six factors of experiences of the teaching–learning environment and students' scores on the four factors of the approaches to learning and studying inventory [14].

Factor	Teaching for understanding	Alignment	Staff enthusiasm and support	Interest and relevance	Constructive feedback	Support from other students
Deep approach	.43	.16	.22	.32	.25	.11
Organized studying	.18	.24	.12	.36	.23	.22
Surface approach	-.44	-.51	-.22	-.47	-.14	-.23

Table 3. Intercorrelations between perceptions of the teaching–learning environment factors and the approaches to learning factors (p<.001 and n=2509) [14].

5. Conclusions and the future of simulator Simulation and Modelling course

The simulation course presented here gives guidelines on how to improve one's analytical skills with the use of mathematical toolset. After the course it is a question of one's personal interest how one wants to utilize these new skills in their professional development. The results that were achieved with this simulator course are realistic and therefore they may be used in future for military officer training.

Group work does not in itself necessarily mean better results or enhanced motivation. The ability to choose the best practices for each studying context requires a process of continuous evaluation. Guidance and clarity in goals and in applied pedagogical tactics are needed for teaching. Continuous surveillance on achieved learning results (including intermediate ones) and motivational aspects may reveal hidden problems in any course. Especially in the simulation course final task would require customized practices that could improve student satisfaction. Action research is a good methodological tool. In the future we may observe class activities using cooperative learning, observing more inter-group discussions. Moreover we may consider flipped class approach for future course. For that purpose more attention is needed to study interaction between students and learning objects on blended learning using Learning Management System, PVMoodle.

As for the course contents, some new subjects from the field of operational analysis will be added into the simulation course. For more clear teacher utilization in the future, the course will be divided into the basic and the advanced parts This will also give more time for teachers to teach and to go a bit deeper with students about simulation and modelling.

Acknowledgements

We are grateful to our master level students' in the department of military technology of National Defence University of Finland, without their great inputs towards simulation and modelling this research would have never been possible. We especially like to thank First Lieutenant Konsta Mehto for his extraordinary effort to produce flight simulator demonstrated in this article. Likewise Captain Petteri Mattila for his unusual effort for the modelling work of Helsinki-Vantaa airport border inspection procedure.

References

[1] Long, Y., Gao, Q., Zhang, Z., Yuan, J., & Wei, Y. (2012). Summarization of virtual battlefield environment technology. In *AsiaSim 2012* (pp. 420-428). Springer, Berlin, Heidelberg.

[2] Lin, H. (2012). Design and Research of Visual Simulation System Based on HLA. In *AsiaSim 2012* (pp. 412-419). Springer, Berlin, Heidelberg.

[3] Hunt, B. R., Lipsman, R. L., & Rosenberg, J. M. (2014). *A guide to MATLAB: for beginners and experienced users*. Cambridge university press.

[4] Guimaraes, D. A. (2010). *Digital transmission: a simulation-aided introduction with VisSim/Comm*. Springer Science & Business Media.

[5] Aldrich, Fuchs, H. U. (2006, December). System dynamics modeling in science and engineering. In *System Dynamics Conference* at the University of Puerto Rico.

[6] Seppanen, M. S. (2000, December). Modeling for application: developing industrial strength simulation models using Visual Basic for Applications (VBA). In

Proceedings of the 32nd conference on Winter simulation (pp. 77-82). Society for Computer Simulation International.

[7] Kemmis, S., McTaggart, R." (2013). *The action research planner: Doing critical participatory action research*. Springer Science & Business Media.

[8] O'leary, Z. (2004). *The essential guide to doing research*. Sage.

[9] DiSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and instruction* 10(2-3): 105-225.

[10] Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics* 66(1): 64-74.

[11] Sundarapandian, V. (2009). *Probability, statistics and queuing theory*. PHI Learning Pvt. Ltd.

[12] Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of extension* 50(2): 1-5.

[13] Lovelace, M., & Brickman, P. (2013). Best practices for measuring students' attitudes toward learning science. *CBE—Life Sciences Education* 12(4): 606-617.

[14] Parpala, A., Lindblom-Ylänne, S., Komulainen, E., & Entwistle, N. (2013). Assessing students' experiences of teaching–learning environments and approaches to learning: Validation of a questionnaire in different countries and varying contexts. *Learning Environments Research* 16(2): 201-215.

Virtual Battlespace 4 (VBS4) - Cloud-Enabled, High-Fidelity and Whole-Earth Simulation

Pete Morrison
Bohemia Interactive Simulations
Orlando, FL, U.S.A.
prodmgt@bisimulations.com

Abstract

Military organizations are striving to leverage best-of-breed simulation and web technologies to deliver high-quality training at the point of need—from battle simulation centers to home computers. VBS4 is an easy-to-use, whole-earth virtual and constructive simulation that supports both individual and collective cognitive training. The VBS4 simulation and rendering engine (VBS Blue) has been developed to support both terrain streaming from the cloud and scalability. A complete replacement for its predecessor VBS3, VBS4 supports hundreds of training use cases including new use cases like small unit Course of Action (CoA) development and analysis, and combined arms and staff planning. The new VBS4 workflow dramatically speeds up the development of training content through its new modes VBS Geo (an easy-to-use but powerful terrain editor) and VBS Plan (a highly efficient mission planning capability). The new VBS World Server is an optional and cloud-enabled companion product for VBS4, which streams terrain to VBS4 instances across a network. It also centralizes the storage of VBS4 Battlespaces - further reducing the overhead of administering multiple VBS4 installations.

1 Introduction

Military organizations are striving to leverage best-of-breed simulation and web technologies to deliver high-quality training to the point of need—from Battle Simulation Centers to home computers. Virtual desktop trainers such as Virtual Battlespace (VBS) are typically used in battle simulation centers, computer laboratories where soldiers will usually sit one soldier per computer and engage in virtual collective training for individual and unit training tasks.

This type of training tool is used for cognitive learning or 'how-to-think' training. Training scenarios are developed in these virtual desktop trainers to facilitate critical decision making. Soldiers control an avatar and perform most actions in the virtual environment as they would in

the real world, using their standard operating procedure to communicate with unit members and engage with different weapons systems for the purpose of cognitive learning.

Virtual training has become widespread among military organizations as part of their training continuum. Today, Bohemia Interactive Simulations' VBS3 is in use in over 60 countries, primarily NATO and NATO countries, for virtual combined arms training. In 2020, BISim released the next version of its virtual trainer VBS4, a complete replacement for VBS3. VBS4 supports hundreds of training use cases including new use cases like small unit Course of Action (CoA) development and analysis, and combined arms and staff planning. The VBS4 simulation and rendering engine (VBS Blue) has been developed to support both terrain streaming from the cloud and scalability.

Figure 1 – ***VBS4, showcasing the high-fidelity graphics. VBS4 is suitable for a plethora of virtual and constructive simulation training use cases.***

2 Addressing VBS3 Customer Feedback with VBS4

Released in 2014, VBS3 has since received years of feedback from customers about desired enhancements and improvements. Today's software users are very familiar with smartphone devices and apps that enable them to begin operating these tools immediately with little to no training. This usability feedback prompted BISim to focus considerable energy on rebuilding its tools for VBS4 with a focus on ease of use and a new workflow.

For example, VBS3 terrain development and editing was an area that BISim focused on improving in VBS4. Where VBS3 supports terrain insets, VBS4 now provides the entire planet. Soldiers can create scenarios anywhere on the virtual earth and create any imaginable training scenarios from IED identification to vehicle checkpoint training to collective training.

3 How VBS4 Differs from Typical Game Engines

VBS4 is an easy-to-use, whole-earth virtual and constructive simulation that supports both individual and collective cognitive training. In VBS4, users create "Battlespaces" that are a collection of terrain edits, mission plans, scenario files and after-action reviews. Each Battlespace is centered on a specific location on the virtual Earth. Scenarios can be modified in real time and replayed in the After-Action Review for lessons learned.

The engine that VBS4 is built on (VBS Blue) was built with a focus on serving the military simulation and training domain, unlike many other 'game-engine' based simulations. Its capability is built beyond 'video game-targeted technology' to specifically solve global-scale challenges, optimized to handle thousands of Artificial Intelligence (AI) entities, densely vegetated areas, and complex urban areas rendered at any altitude (from space to ground level or ocean floor). Of course, industry-leading 'game' technology is still leveraged, with high-fidelity 3D content (VBS4 includes 18,000+ 3D models) and state-of-the-art lighting and atmospheric rendering.

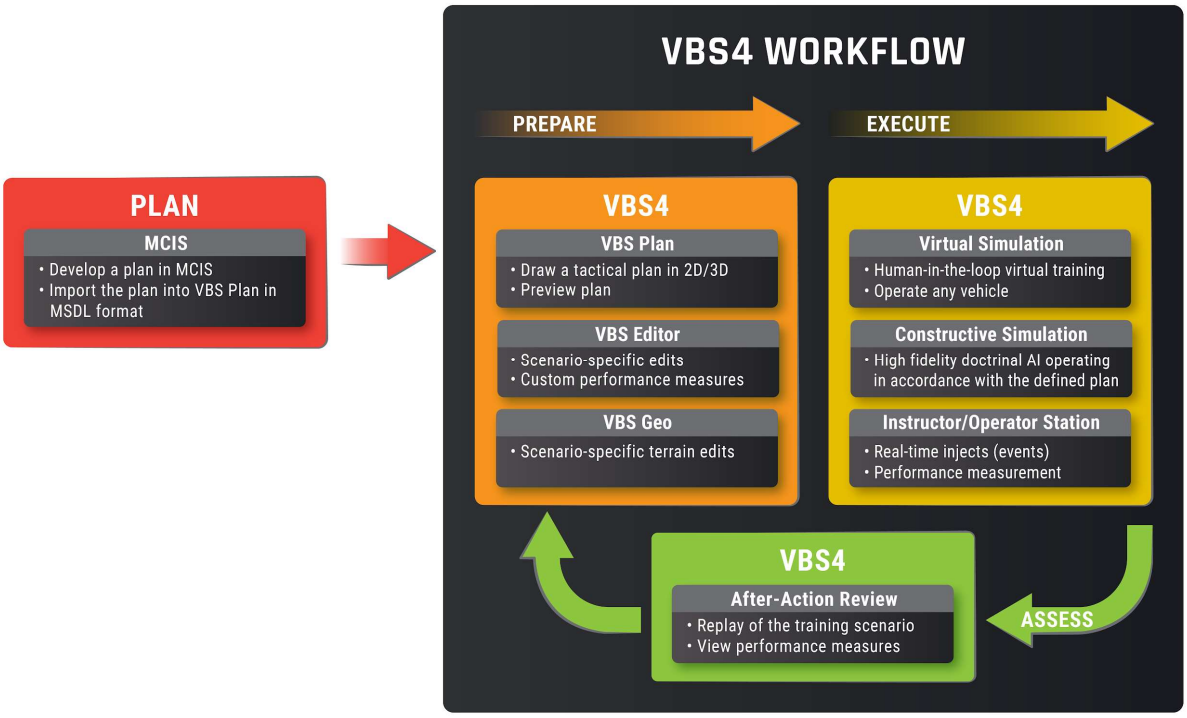


Figure 2 – The VBS4 workflow

4 Workflow Improvements For Faster Development, Ease of Use

The new VBS4 workflow dramatically speeds up the development of training content through its new modes VBS Geo (an easy-to-use but powerful terrain editor) and VBS Plan (a highly efficient mission planning capability).

VBS4 introduces a fresh, new main-menu user interface and user experience by presenting users immediately with a view of the globe, offering the list of available Battlespaces and the ability to create new ones anywhere on the planet.

The initial (optional) Plan phase is typically conducted in real-world systems such as the Mission Command Information System (MCIS) meaning planners 'train as they fight'. Users conduct the Prepare, Execute and Assess phases within VBS4 (optionally they can import the plan from MCIS into VBS Plan, as a starting point). The Prepare phase of the VBS4 workflow allows the user

to "quick-switch" between VBS Geo, VBS Plan and the VBS Scenario Editor ("Editor") modes. This means scenarios can be rapidly created by building and editing the terrain and the scenario, iterating between the three modes as needed. Terrain and scenario editing can be conducted in either 2D or 3D.

Figure 3 – The terrain representation in VBS4 is global – from space down to blades of grass and ocean floor, with realistic view distances.

5 Centralized Whole-Earth Terrain

The new VBS World Server is an optional and cloud-enabled companion product for VBS4, which streams terrain to VBS4 instances across a network. It also centralizes the storage of VBS4 Battlespaces - further reducing the overhead of administering multiple VBS4 installations.

The terrain data that is rendered by VBS4 can be either stored as part of a VBS4 installation (i.e., on every VBS4 computer) or stored centrally on the optional VBS World Server (VWS) (i.e., on a single computer accessible by the VBS4 computers over the network). VWS reduces the administrative overhead of managing a VBS4 classroom by centralizing terrain data and also the storage of Battlespaces. VWS also facilitates collaborative terrain editing as well as dynamic terrain deformation at runtime.

VWS is designed from first principles with cutting-edge networking, cloud deployment, procedural

enhancement capabilities, and open standards compliance to facilitate easy enhancement with future technology advancements. It includes global geospatial data, including curated elevation and bathymetric data, buildings and roads, water bodies, regional specific vegetation and surface materials. The terrain data processing pipeline handles sub-centimeter resolution data sources, or it can procedurally generate realistic, geotypical, high-resolution data from low-resolution sources. The procedural algorithms automatically generate complex terrain features like bridges, tunnels and overpasses, and ensures these terrain features support all the elements of the scenario such as AI path planning and physical destruction.

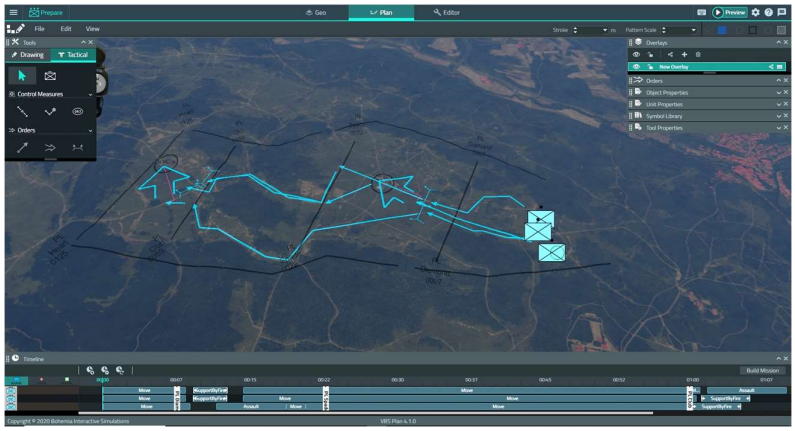


Figure 4: Developing a plan in 3D in VBS Plan, part of VBS4

6 VBS Plan - Rapid Mission Planning and Execution

VBS Plan is a new mission planning tool that allows any user to quickly "sketch out" and execute a tactical plan without prior simulation or scenario generation knowledge. VBS Plan is built directly into VBS4 and is immediately accessible from the Prepare user interface. VBS Plan provides a revolutionary scenario generation approach whereby the user places military tactical markings (e.g., phase lines, advance, defend, attack, etc.) and unit symbols (i.e., MIL-2525C) at a constructive level, and then VBS4 automatically creates the individual virtual entities and assigns behaviors. VBS Plan is a very rapid, intuitive way to access the massive model library and VBS Control AI behaviors delivered as part of the VBS4 capability package and to form the assets into a synchronized military plan. This approach massively reduces the scenario development time and need for detailed software expertise, freeing up instructors to focus on developing scenarios which maximize learning. Once the user adjusts the 'timeline' (coordination and synchronization) of certain predefined actions/phase lines/advances, etc., the plan can be immediately executed in VBS4.

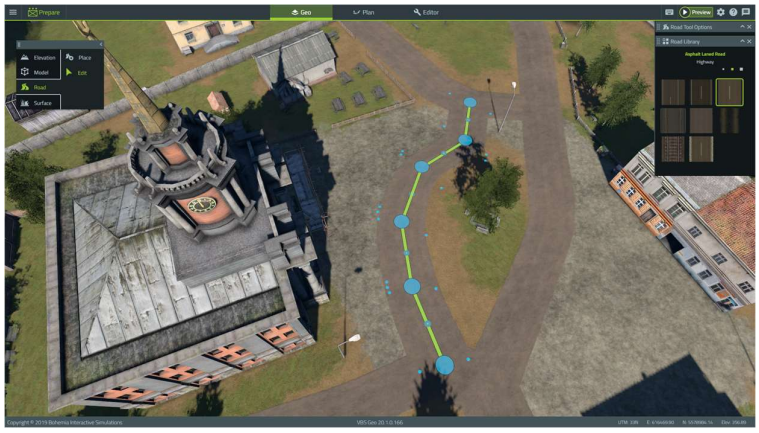


Figure 5: Road editing in VBS Geo

7 VBS Geo - WYSIWYG Terrain Editing

VBS Geo is an easy-to-use 'What You See Is What You Get' (WYSIWYG) terrain editor fully integrated within VBS4. VBS Geo allows users to intuitively and rapidly modify, extend and replace environmental features of the whole-Earth database without leaving the VBS4 application.

This empowers users with the ability to enhance the out-of-the-box terrain using VBS Geo to meet the specific requirements of the intended scenario and training delivery without the need for additional specialist Geographic Information Systems (GIS) applications or knowledge. For example, the user can quickly move a road, or add a building or a fence, and VBS4 will do the rest. The VBS4 AI entities will follow the new road, stop at pedestrian crossings and avoid any new buildings or fences, etc. The tool enables both small-scale edits (e.g., creating a trenchline for a defensive position) to large-scale city-building (e.g., importing vector data describing roads and buildings for an entire city).

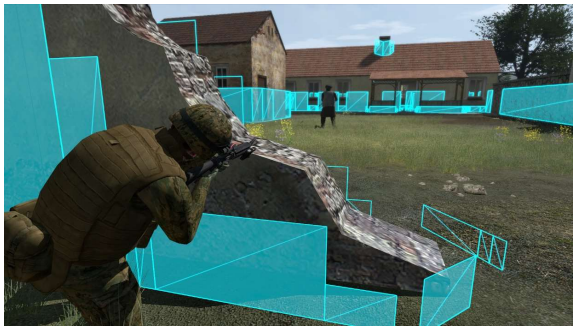


Figure 6 - A VBS Control soldier taking aim from cover, cover debug visualizations are enabled to show information provided by the AI library to the soldier's behavior

8 New AI Behaviors and AI Scalability

BISim had already developed next-generation artificial intelligence behaviors for VBS3 that aims to reduce the need for human roleplayers by using semi-autonomous AI forces. For the trainee, the new AI, called VBS Control, in VBS4 brings a realistic experience as an opponent, a reliable team-mate, or the presence of an immersive pattern of life. VBS Control AI provides VBS4 with a set of doctrine-typical AI behaviors for convoys, civilian pattern of life and high-fidelity traffic, and infantry combat maneuvers.

This new AI technology allows for precise and seamless navigation in open terrain as well as in an urban environment, including building interiors. AI in VBS Control can evaluate terrain for tactical information. This allows for the selection of a secure route when pathfinding, the use of terrain as cover from observation, and to optimize between time/security when planning routes. And, VBS Control implements a new cover system, that detects (at runtime) geometry in VBS4 as cover, to be used by the AI. This allows behavior makers to easily create high fidelity behaviors for detecting/eliminating threats, distributing observation sectors, and finding firing positions.

VBS4 supports thousands of high fidelity constructive (AI) civilian and military entities in the same Battlespace. In the very near future, tens of thousands of entities will be supported using cloud scalability technologies, delivering realistic cluttered and congested training environments and replicating the scale of real-world combined arms and/or joint operations.

9 Enabling Reuse and Backwards Compatibility

BISim recognizes that its customers invest considerable time and money in building content, terrains and training missions in the virtual environment. VBS4 was developed to support backwards compatibility with terrains, missions and 3D content so customers can make effective reuse of content they have already paid for and produced.

VBS3 content can be easily imported into VBS4. There is a process for ingesting VBS3 terrain into VBS4. VBS4 also supports converting a VBS3 mission to a VBS4 Battlespace and uploading the Battlespace to the VBS World Server. Finally, BISim has also introduced a new model import pipeline for VBS4 that enables the use of industry-standard 3D tools to cut time for adding new VBS4 3D content.

10 VBS4 - Easier. Faster. Global.

In summary, the key improvements of VBS4 compared to VBS3 are ease of use, performance due to the new VBS Blue engine and simulation optimizations, and its whole-Earth terrain representation. The new main menu workflow, and VBS Plan and VBS Geo, has been designed for non-engineers. The aim is to unlock the power of VBS4 for every tech-savvy soldier, sailor and airman.

There is no need for multi-day training courses before meaningful training commences. From a customer perspective, this means faster generation of training scenarios, enhanced VBS3 use cases (better training) and new VBS4-only use cases (more types of training).

BISim has released one update to VBS4 at the end of 2020 with additional enhancements and new capabilities, and plans at least two major releases a year with the next happening in the middle of 2021.

THEME 3: Immersive Technology and Extended Reality (XR) in Support of Operations

xR in the Operation, Operate in xR
The Tech Intensity of the Future Mission

Microsoft

Abstract

As the fourth industrial revolution progresses, we are now firmly in the "digital age of data and intelligence" with an increased need for technical intensity in every industry including defense and intelligence. The information advantage currently enjoyed by the alliance could be eroded or even reversed as adversaries, including non-state actors, attain similar levels of situational awareness through easy access to public cloud-powered digital modeling and simulation capabilities. xR, modeling and simulation can no longer be a separate discipline but needs to become an integral part of the mission. This whitepaper highlights some of the key disrupting disruptive technologies like xR, AI, cloud and 5G/Satcom and gives an initial overview on how Microsoft seeks to integrate these capabilities within its intelligent Azure cloud and edge. It serves as food for thought for NATO and nations to start an in-depth discussion on how this will affect the cloud, deployed systems and networking capabilities required to support its missions and exercises. The tech intensity in missions will require the whole defense ecosystem including the industrial base to rapidly adapt and transform.

THIS CONTENT IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. IT DOES NOT CONSTITUTE AN OFFER OR OBLIGATION FOR MICROSOFT TO PROVIDE ANY OF THE PRODUCTS AND SERVICES REFERENCED HEREIN. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, our provision of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The descriptions of other companies' products in this document, if any, are provided only as a convenience to you. Any such references should not be considered an endorsement or support by Microsoft. Microsoft cannot guarantee their accuracy, and the products may change over time. Also, the descriptions are intended as brief highlights to aid understanding, rather than as thorough coverage. For authoritative descriptions of these products, please consult their respective manufacturers.

© 2020 Microsoft Corporation. All rights reserved. Any use or distribution of these materials without express authorization of Microsoft Corp. is strictly prohibited.

Microsoft and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

1 Tech Intensity of the mission

As the fourth industrial revolution progresses, we are now firmly in the digital age of "data and intelligence". With Resilience, Consultation and Collective Defense at the heart of its mission, this particularly applies to NATO. Data science, artificial intelligence powered by cloud and edge computing combined with innovations in human-computer interaction, augmented reality, and gaming are disrupting the OODA loop. The commander and warfighter must be able to consume any data or service anywhere, in any form factor, at any time, at any scale, in a consistent manner both at the static intelligent cloud and the deployed tactical intelligent edge. As an industry leader, Microsoft believes the increase of tech intensity will help drive innovation, both centrally and at the deployed edge. As Microsoft CEO Satya Nadella notes,

"We must adopt technology in ways that are much faster than what we have done in the past. Each one of us, in our organizations, will have to build our own digital capability on top of the technology we have adopted. Tech intensity is one of the key considerations you have to get right."

Technology intensity is a business-driven investment in digital capabilities in these key areas:

- Build Intelligent edge and cloud with high-performant and resilient 5G and SATCOM connectivity,

- Innovate with advances in AI in every mission process;
- Transform the augmented user experience for the commander and warfighter with xR;
- Relentless focus on building technology that commanders can trust.

Embedding useful computing in every aspect of the real world and enabling and sharing relevant insights in real-time will increase situational awareness and facilitate better and more transparent decision making. Not only will this allow NATO's Allied Forces to achieve more, but it will also help reduce the risk of collateral damage during operations.

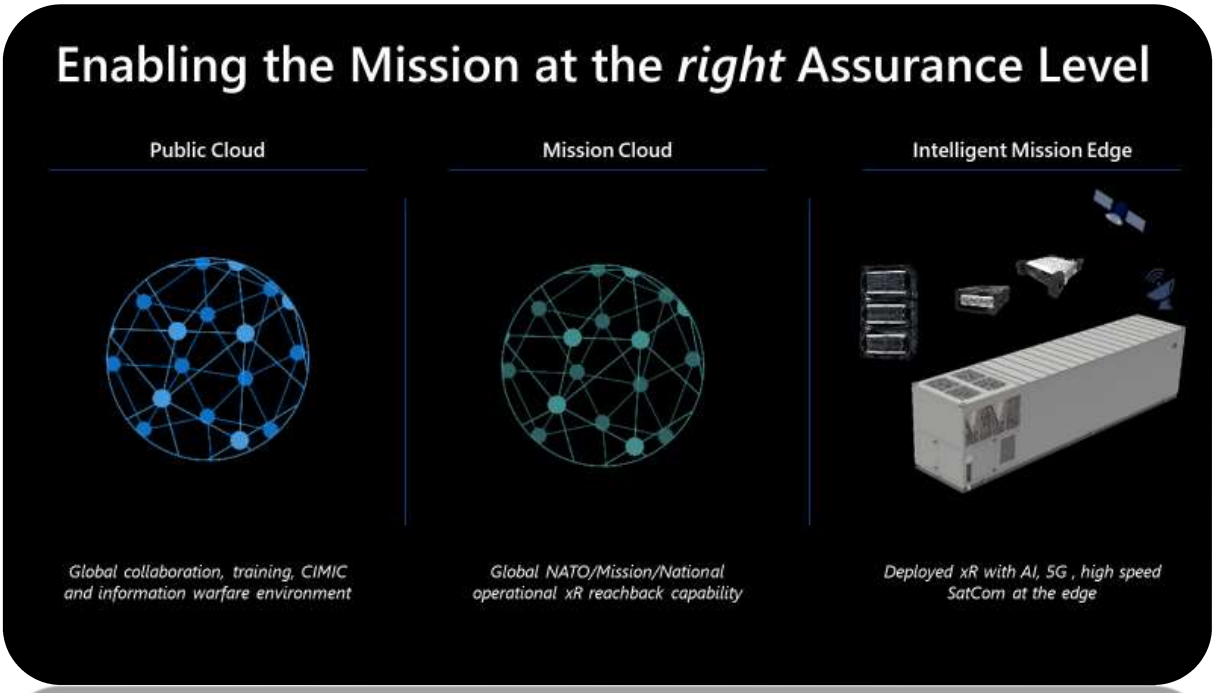
Technology intensity is a business-driven investment in digital capabilities in these key areas:

- Build Intelligent edge and cloud with high-performant and resilient 5G and SATCOM connectivity,
- Innovate with advances in AI in every mission process;
- Transform the augmented user experience for the commander and warfighter with xR;
- Relentless focus on building technology that commanders can trust.

Embedding useful computing in every aspect of the real world and enabling and sharing relevant insights in real-time will increase situational awareness and facilitate better and more transparent decision making. Not only will this allow NATO's Allied Forces to achieve more, but it will also help reduce the risk of collateral damage during operations.

2 Intelligent Edge, Cloud and Connectivity

Mission cloud and deployable Communication Information Systems (DCIS) capabilities require flexible, interoperable and scalable Command and Control and Core Information Systems in order to plan, execute, and assess full spectrum operations. Microsoft is committed to provide trusted Azure capabilities around the world through hyperscale cloud and intelligent edge, and in both connected and disconnected scenarios and offers a comprehensive portfolio designed to bring data analysis and insight to the tactical edge with 5G and Satellite connectivity. **This will enable the NATO alliance to run low-classified workloads in the public cloud, while running higher classified workloads with data residency in connected or disconnected mode.**



2.1 Cloud and Edge

The **intelligent cloud** is ubiquitous computing, enabled by the public or sovereign mission cloud and artificial intelligence (AI) technology, for every type of intelligent application and system you can envision. Microsoft's Azure regions cover 140 countries, interconnected with the world's largest and most diligently managed fiber network. The goal of the global cloud design, inclusive of compute facilities and networking, is to provide access to population centers with a latency of less than 10 milliseconds.

The **intelligent edge** is a continually expanding set of connected deployed DCIS and tactical systems and devices that gather and analyze data—close to your users, the data, or both. Users get real-time insights and experiences, delivered by highly responsive and contextually aware apps that are delivered with agile DevSecOps through the NATO Software Factory. By combining the virtually limitless computing power of the cloud with intelligent and perceptive devices at the edge of your network in a "as a service" model, you have access to a platform for building immersive and impactful mission solutions at the right assurance level. This includes an entire range from compact, man-portable, and small vehicle portable Point of Presence (PoP) all the way to fully deployable communication & information Services PoP with transportable container/modules.

Through a cloud platform one can provide a consistent experience across classification levels and deployment models. As the underlying concepts of cloud technology become ubiquitous, NATO can take advantage of these to adopt a **"develop once/deploy anywhere approach"**. From a networking perspective, this flexible combination of edge and cloud drives the need for **intelligent connectivity** in any combination of technologies and models, in a reliable, secure, agile, flexible, scalable, and on-demand manner, anywhere, that go beyond the current trend of software defined networking. Beyond just the technology change, it also requires the adoption of a

different operating model, a model that better match the agile and flexible models in use at the application layers. Like the application environment moved to DevSecOps models, **the networking environment must move to similar "NetSecOps"** models to realize the same benefits with regards to agility and flexibility.

2.2 SATCOM & 5G

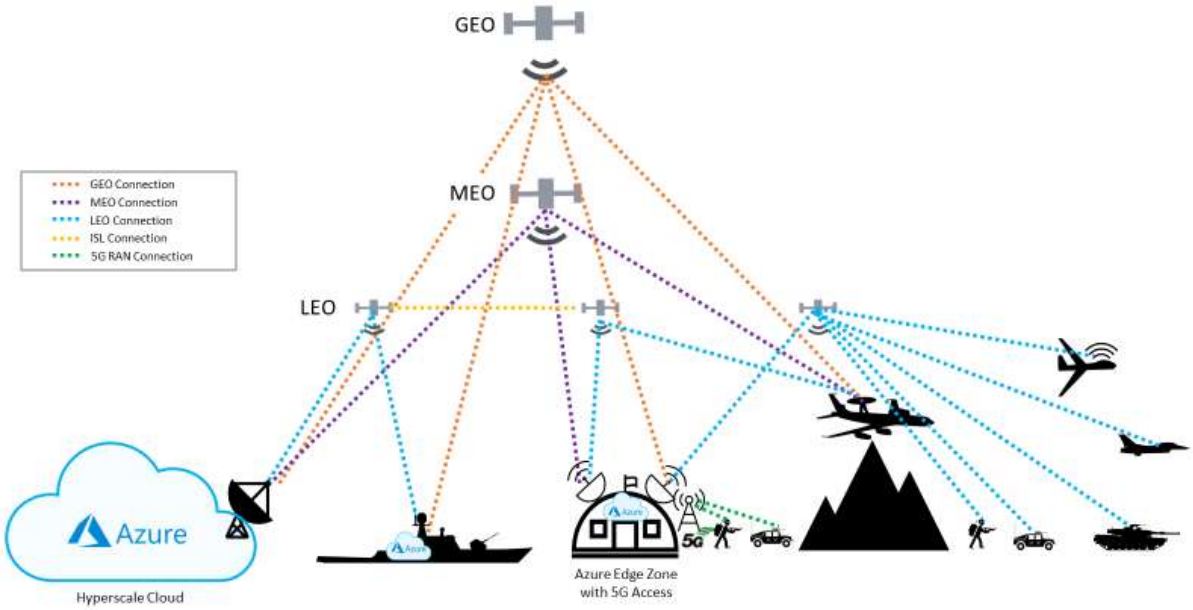
As NATO support missions around the world, often in remote locations and beyond the reach of standard infrastructure, new capabilities and connectivity is required for mission success. Classic networking approaches will be disrupted by private and public 5G and high throughput satellite constellations in space. **To enable intelligent mission connectivity**, NATO must embrace innovation in these key areas:

- **Core global networking** backbone with improved peering services and expansion of own Edge PoPs
- **Virtual WAN** to interconnect the different sites/locations of customers seamlessly via proximity to missions with seamless, secure, high throughput **satellite connectivity**
- **5G network slicing & virtualizing satellite networking** functions allowing for the integration of any type of satellite connection in a workload.
- a **fully virtualized 5G stack and services** allowing NATO and its carriers to be a NaaS operator
- proximity to your last mile provider, carriers and tactical edge with **5G (Private) Edge zones**

This approach allows NATO to build out the network of the future: A network that spans across on-premise, cloud and mission location. One that is software defined and that is managed end-to-end from a single pane of glass with an orchestration of virtual network functions, without a compromise on security. One that leverages different communication technologies and integrates into field deployed 5G technology: Microsoft believes satellite communication forms part the 5G solutions space and is partnering with leading SATCOM technology to drive the inclusion of SATCOM as a standardized element of 5G connectivity to enable NATO missions.

5G networking standards are key enablers of the ubiquitous connectivity that is required, and they cover much more than just the mobile network that it is mostly known for. It includes all modern connectivity technologies from access to core, from fiber optic connections to satellite connectivity. It provides standardized service models and open interface capabilities to unify any combination of supported connectivity technologies, models, and topologies, from any combination of different operators, into a single, integrated, and orchestrated connectivity capability with the required security, reliability, visibility, and control.

With a unified network approach, the door is open to step into unified compute layer, boosting application innovation and deployment and becoming an intelligent led defense organization. Creating unity between compute in datacenter, cloud, and edge, without needing to sacrifice security, ease of deployment or ease of build. At the core of this all, is the construct of an automated and orchestrated network, handling routing, security, and other functions without the need of manual config or hardware-bound solutions. Here Microsoft continues to invest in the creation of an e2e orchestration and deployment layer. So that virtual network functions can be orchestrated and chained centrally.



GEO, MEO, LEO Use Case Examples

Microsoft is already working with a number of industry leaders including SpaceX, SES, KSAT, Viasat, Kratos, AMERGINT, KubOS and US Electrodynamics to bring the power of cloud to the space domain. For example , Microsoft is working with **SpaceX** to provide satellite-powered internet connectivity on Azure. SpaceX recently won a contract with the Space Development Agency to build new satellites – separate from the Starlink system – in support of a Space Tracking Layer defense system capable of detecting and tracking ballistic, cruise and hypersonic missiles. Microsoft will join the SpaceX team on this project.

For network deployments, supporting the required agility and flexibility, network functions beyond physical connectivity should be virtual network functions running on a general compute platform allowing any logical network to be established in a network as code manner just as we do for infrastructure.

Network Function Virtualization (NFV) however only provides part of the solution as the scale and velocity needed for connecting the edge and cloud will require intelligent closed loop orchestration that is application aware. By leveraging the cloud as the distributed compute platform to provide core connectivity, support all virtual network functions and enable application aware intelligent orchestration and control a network-as-a-service cloudification of the network will be achieved. **This will enable consumption of reliable, secure, agile, flexible and scalable connectivity in an on-demand manner through NaaS (network as a service) and 5GaaS (5G as a service) solutions**.

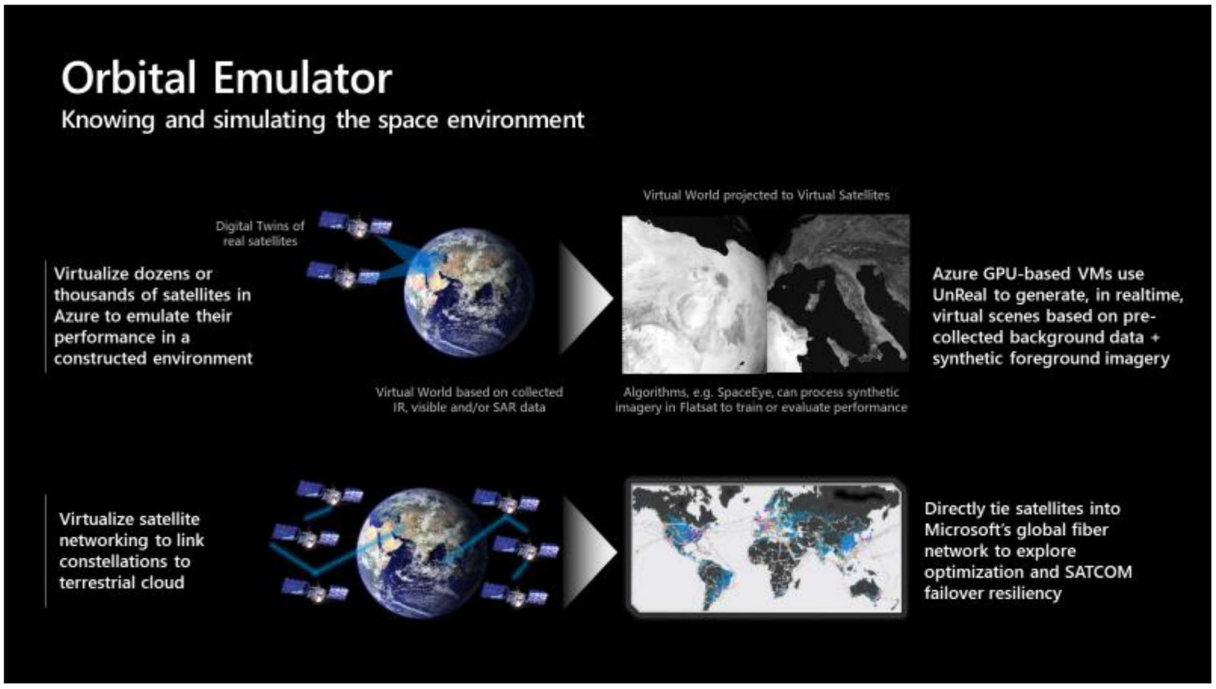
3 Advancing missions with AI

Defense and Intelligence Agencies are enabling strategic decision making and preventing personnel from getting bogged down with routine analysis of mountains of data by adopting artificial intelligence (AI). The below provide

just a few illustrations on how AI can be embedded in every aspect of the mission.

3.1 - Orbital Emulator

As space missions and satellite capabilities become more accessible, we are developing reliable, repeatable digital technologies to help the space community launch faster and with mission assurance. The first of these is Azure Orbital Emulator. Commercial and government space organizations are developing thousands of interconnected satellite constellations which require precise planning and sophisticated AI-driven formation protocols, to ensure optimal networking connectivity and operational coverage on-orbit.



Azure Orbital Emulator is an emulation environment that conducts massive satellite constellation simulations with software and hardware in the loop. This allows satellite developers to evaluate and train AI algorithms and satellite networking before ever launching a single satellite. Azure can emulate an entire satellite network including complex, real-time scene generation using pre-collected satellite imagery for direct processing by virtualized and actual satellite hardware. Azure Orbital Emulator is already being used by customers in our Azure Government environment

3.2 - Seeing through clouds with SpaceEye

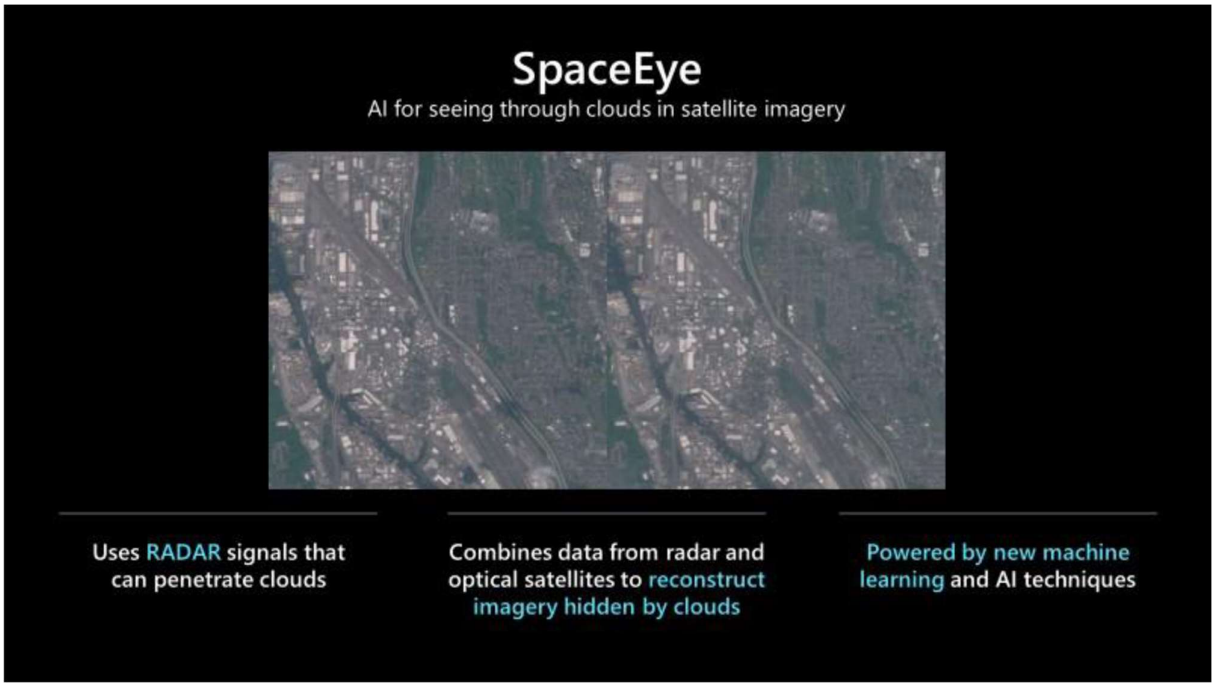
A significant limitation of satellite imagery is that around 77% of the images are thrown away because of weather conditions and clouds obfuscating the areas of interests. Currently most earth observation workloads identify and discard imagery with clouds. Through the use of advanced machine learning, Microsoft has invented a new technique that combines optical images from satellites with RADAR data from other satellites to reconstruct imagery below the clouds. Not just in Red, Green, and Blue, but in multi-spectral bands as well.

Our technique, called SpaceEye can reconstruct scenes below the clouds with above 85% accuracy for all bands. We can not only see objects of interest below the clouds, but also see vehicles/ships come and go, and based and sites coming up – even below the clouds!

3.3 - Naval modeling and forecasting

With two tropical storms headed for the Gulf Coast, a U.S. Navy team of leading engineers and scientists partnered with their Microsoft Federal counterparts over the Aug. 22–23 weekend to successfully deploy an enhanced weather model capable of rapidly scaling weather and ocean pattern predictions on demand. The Coupled Ocean/Atmosphere Mesoscale Prediction System, or COAMPS-TC, began running in a production-

like Microsoft Azure environment on Aug. 23. The proof-of-concept results, posted on Twitter starting Aug. 24, confirm the potential of using our secure cloud to deliver faster, real-time and scalable tropical cyclone



forecasts, which result in greater fleet safety and effectiveness and public benefits.

The project demonstrated nimble technical engineering, collaboration of multiple government and industry teams, and partnership to deliver mission-critical requirements anytime, anywhere. The successful COAMPS model running in Azure generated many positive outcomes in modeling weather and ocean patterns associated with tropical storms Laura (later elevated to Hurricane Laura, Category 4) and Marco, including:

- Agility in rapidly responding to severe weather systems, with cost-effective, high-quality results:
  - o Available in under six hours, exceeding expectations.
  - o At three times the speed and more than double the volume of respective on-premises production rates and capacity.
  - o Costing one-tenth of results produced by the existing datacenter.
- Serving as a prototype for:
  - o Navy Continuity of Operations Planning, which maintains mission-essential functions in emergency situations.
  - o Effective partnerships with internal and external research organizations to speed development, implementation and transformation of existing operations.
- Validating Microsoft Azure delivers the adaptable, dynamic and scalable high-performance computing capacity required by the Navy, with:
  - o A robust architecture enabling resilient data communications.
  - o The full scalability of compute-intensive weather models beyond the normal capacity of on-premises systems.

3.4 - Anywhere on Earth – growth of 3D outside gaming

In December, Microsoft took part in a demonstration called Project Anywhere at the Interservice/Industry Training, Simulation and Education Conference (IITSEC), the world's largest modeling, simulation and training event. Built in Epic's Unreal Engine, and hosted on Azure, this demo provides real-time access to an interactive "digital twin" 3D model of the entire Earth through the Cesium platform streaming global data as 3D Tiles, rendered on NVIDIA GPUs. From the demo there were three key takeaways that we should all be very excited about – that hint at the future of cloud development –

and the important role that simulation and gaming will play in that future.

First, this demo was rendered entirely in the cloud, which means it can be accessed from any device with a modern web browser, including mobile phones or low-end PCs. This is called Pixel Streaming, and I predict it's going to revolutionize a broad swath of industries by making cutting-edge interactive 3D experiences available to anyone. An architect, for example, can create a virtual walkthrough of a new building, and know that the client will be able to experience it in all its glory – without requiring a powerful gaming PC.

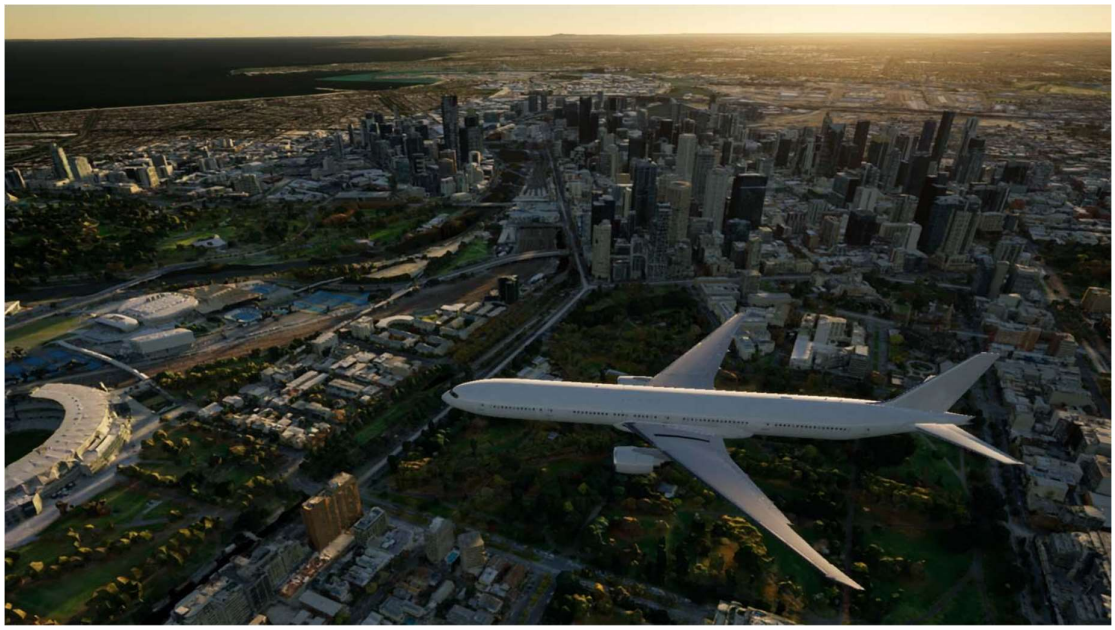
Second, the demo involved a massive amount of data. But thanks to the power of the cloud, you don't have to wait for all that data to be downloaded! Instead of bringing data to the user, now we can bring the user to the data. Designers no longer need to be limited in the amount of data required for an optimal user experience – you can build world-scale applications and again, expose them via any device – even mobile devices with thin pipes and limited local storage. Imagine a city repair crew trying to locate the source of a water leak while out in the field – now you can explore an interactive model of an entire city's infrastructure, in 3D, from your phone.

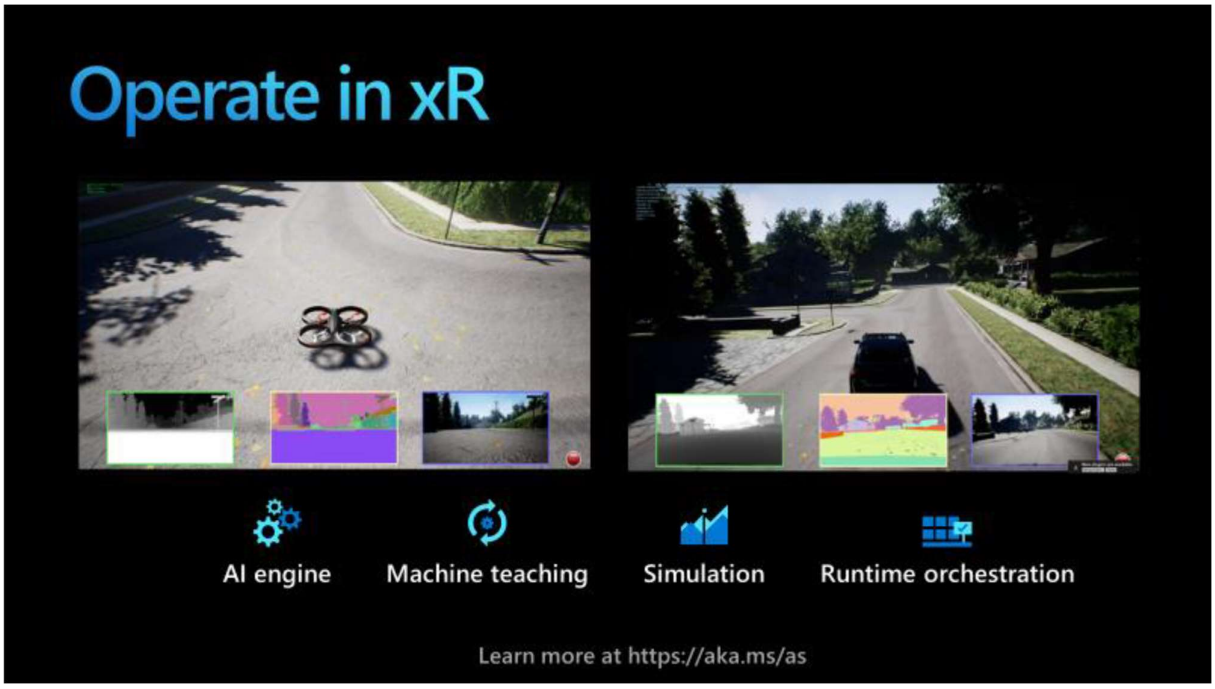
Finally, by keeping the data in the cloud, we make it easier for multiple users to interact with and modify that data at the same time. This sort of player-to-player interactivity is nothing new in game playing, but it's very new in game production. By enabling an entire team to experience and create content together in real-time, we can dramatically accelerate all sorts of production pipelines. What Office 365 has done for collaborative document editing, we should now soon see happening in level product design or and film production through the power of the cloud and features such as Unreal Engine multi-user editing.

What the Project Anywhere demo has shown us is how gaming technology is driving innovation beyond gaming. As data sets are getting bigger and visualization technology is getting more powerful, moving to the cloud makes things accessible to everyone. Hosting and simulating in the cloud means any device can view demos like Project Anywhere. It's a technology you should be watching closely and conduct proofs-of-concept as it progresses.

3.5 - Autonomous systems

Most people wouldn't think to teach five-year-olds how to hit a baseball by handing them a bat and ball, telling them to toss the objects into the air in a zillion different combinations and hoping they figure out how the two things connect. Neither do we do this when we train commanders, warfighters or autonomous drones, ships or weapon systems. And yet, this is in some ways how we approach machine learning today — by showing machines a lot of data and expecting them to learn associations or find patterns on their own.





But as the desire to use AI for more scenarios has grown, Microsoft scientists and product developers have pioneered a complementary approach called machine teaching. This relies on people's expertise to break a problem into easier tasks and give machine learning models important clues about how to find a solution faster. It's like teaching a child to hit a home run by first putting the ball on the tee, then tossing an underhand pitch and eventually moving on to fastballs. Machine teaching seeks to gain knowledge from human domain experts rather than extracting knowledge from data alone. A person who understands the task at hand — whether how to decide how to secure an area or execute a surveillance task — would first decompose that problem into smaller parts. Then they would provide a limited number of cases, or the equivalent of lesson plans, to help the machine learning algorithms solve it.

Deep reinforcement learning, a branch of AI in which algorithms learn by trial and error based on a system of rewards, has successfully outperformed people in video games. But those models have struggled to master more complicated real-world tasks. Adding a machine teaching layer — or infusing an organization's unique subject matter expertise directly into a deep reinforcement learning model — can dramatically reduce the time it takes to find solutions to these deeply complex real-world problems. Autonomous machines are more than an expansion of automated systems: They are an entirely new way to amplify human expertise. *Autonomous* systems can react to changing conditions and adapt operations to maintain efficiency and accuracy. Autonomous systems learn from human experts and practice safely in a simulated environment before responding to real-world scenarios. This means the same machine, powered by an autonomous control system, understands how variables like air temperature, soil consistency or even the age of the machine itself will affect the outcome, and will take or recommend the best course of action to meet the desired objective, in this case efficiency and accuracy.

The transformative power of an autonomous system might seem self-evident, but there are several ways in which engineers, commanders and defense organizations can maximize these systems.

1. Leverage your human experts. Autonomous systems do not replace expertise, they amplify it. Infusing AI with the technical expertise and wisdom of your experts will help your organization deploy the most effective control systems possible.

2. Deploy autonomous systems in partnership with people or autonomously. There are several ways to deploy autonomous systems: They can assist your workers on a task, act as an advisor to find the best path

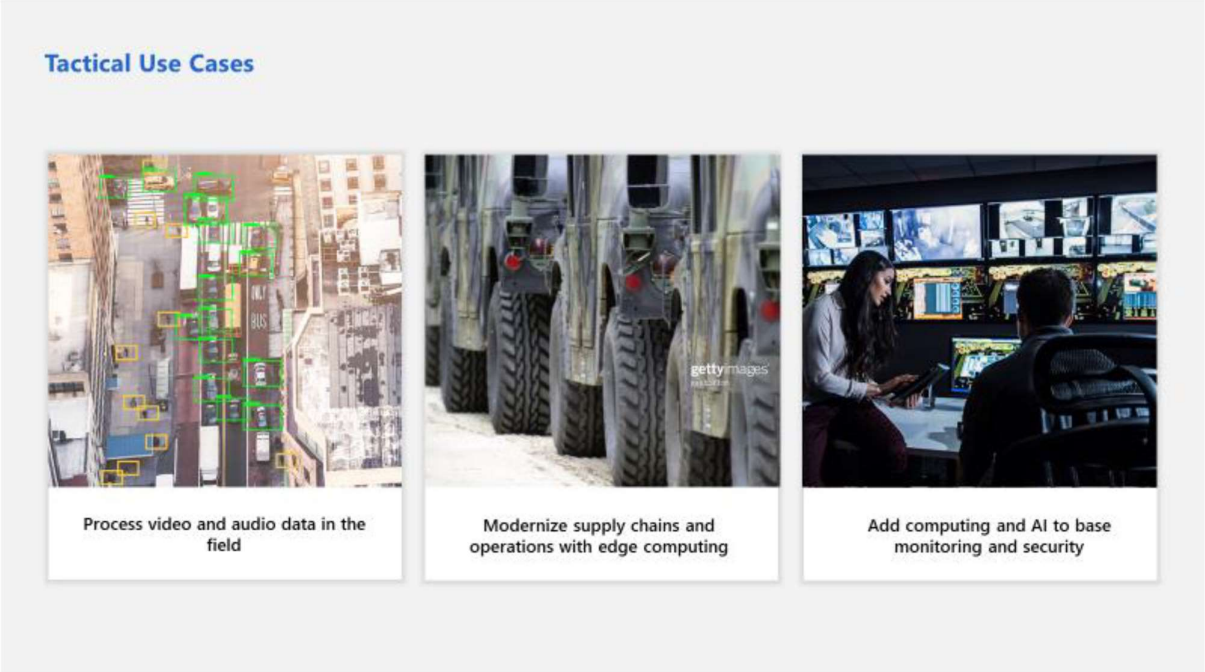
forward in a task or execute that task autonomously. In each case, there is an integral partnership between people and system.

3. Start wherever it makes sense. Focus on where you need autonomous solutions and make them happen there first. If you need a single control, do that before planning an entire overhaul of your process. Conversely, if you want to build autonomous processes, then plan there and see what controls fit into that plan.

While automation helped transform several industries by mobilizing technology to build products at scale, autonomous systems take that transformation several steps further by recognizing nuances and changes to enhance the creative problem solving and strategic thinking of human experts.

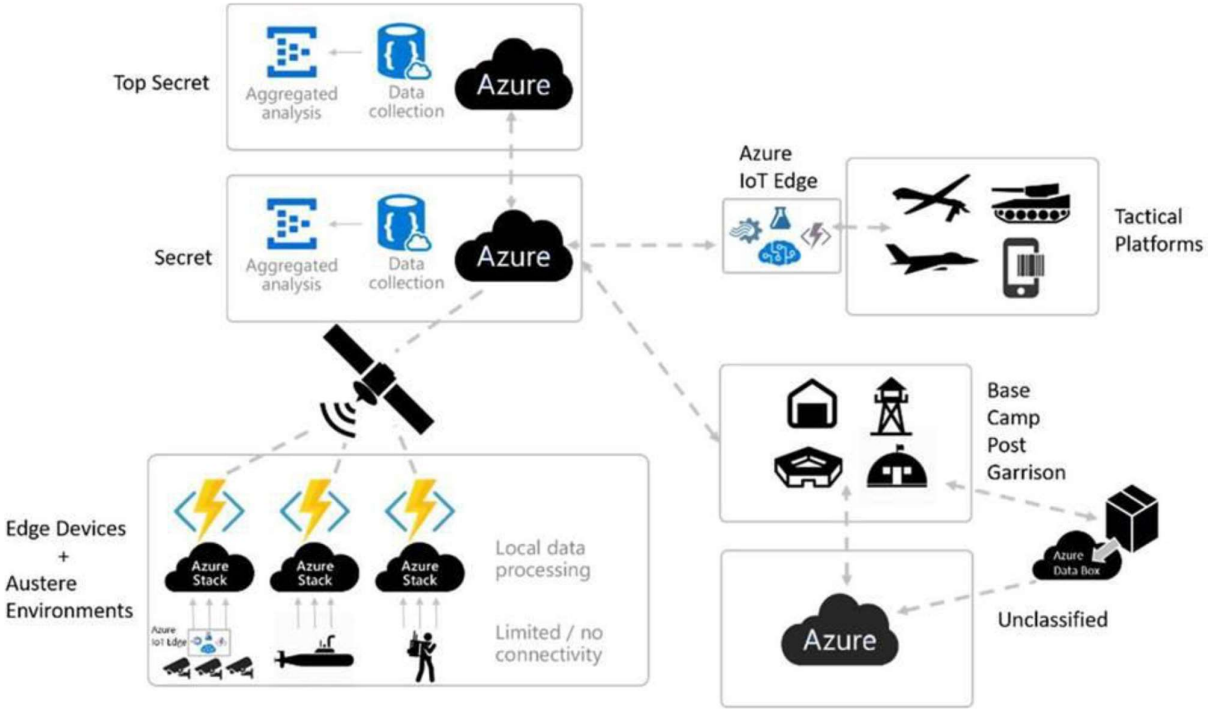
3.6 - Tactical Edge

The intelligent edge is a continually expanding set of connected systems and devices that gather and analyze data—close to your SOF and C2 users, the data, or both. Commanders and warfighters can get real-time insights and experiences, delivered by highly responsive and contextually aware apps.



For military applications, Microsoft's growing portfolio of rugged edge solutions is an essential part of the Azure platform. These systems offer unprecedented opportunities to expedite decision making and bring the power of cloud to areas far beyond the reach of a traditional datacenter. The Azure Edge family of products helps NATO with remote operations accessing the information they need to make decisions at the edge, along with access to the full range of AI and data science analytics and augmented reality experiences as satcom and 5G connectivity allows. The solutions range from compact, man-portable, and small vehicle portable Points of Presence (PoP) all the way to fully deployable communication & information Services PoP with transportable container/modules. The Azure platform provides a means of deploying applications in the field, without requiring in-depth platform knowledge from DCIS personnel. By **hiding the underlying complexity, teams can focus on accomplishing the mission** at hand rather than on solving IT infrastructure issues.

This capability empowers customers to exploit sensor telemetry and other data generated and collected at the edge by enabling edge analytics of data streams, events and alerting, perform image and video processing and recognition, document exploitation, and voice translation to derive intelligence and optimize mission planning and operations in near real-time. The below illustrates a complete architecture information sharing across different levels of command.



Azure, Azure Stack Hub and Azure Stack Edge to support full range of cloud computing scenarios

4 Augmented User Experience

4.1 Augmented and mixed reality

Mixed reality on HoloLens 2 combines an untethered device with apps and solutions that help people to learn, communicate, and collaborate more effectively through visualization of C2 and JISR information for increased situational awareness. In mixed reality, digital information is represented by holograms—objects made of light and sound—that appear in the space around you. Through artificial intelligence, these holograms respond to commands and interact with real-world surfaces in real time for a more natural and intuitive experience. You can start building secure, collaborative mixed reality solutions today using intelligent services, best-in-class hardware, and cross-platform tools.

An example is the Airbus Tactical Sandbox that leverages Augmented Reality for Mission Preparation. The Holographic Tactical Sandbox is the new generation of operation planning and command and control tools. It provides an accurate representation of the battlefield, thanks to the viewable 3D holographic map. By facilitating operations, planning and decision-making, this innovation shortens the observation and decision-making loop.

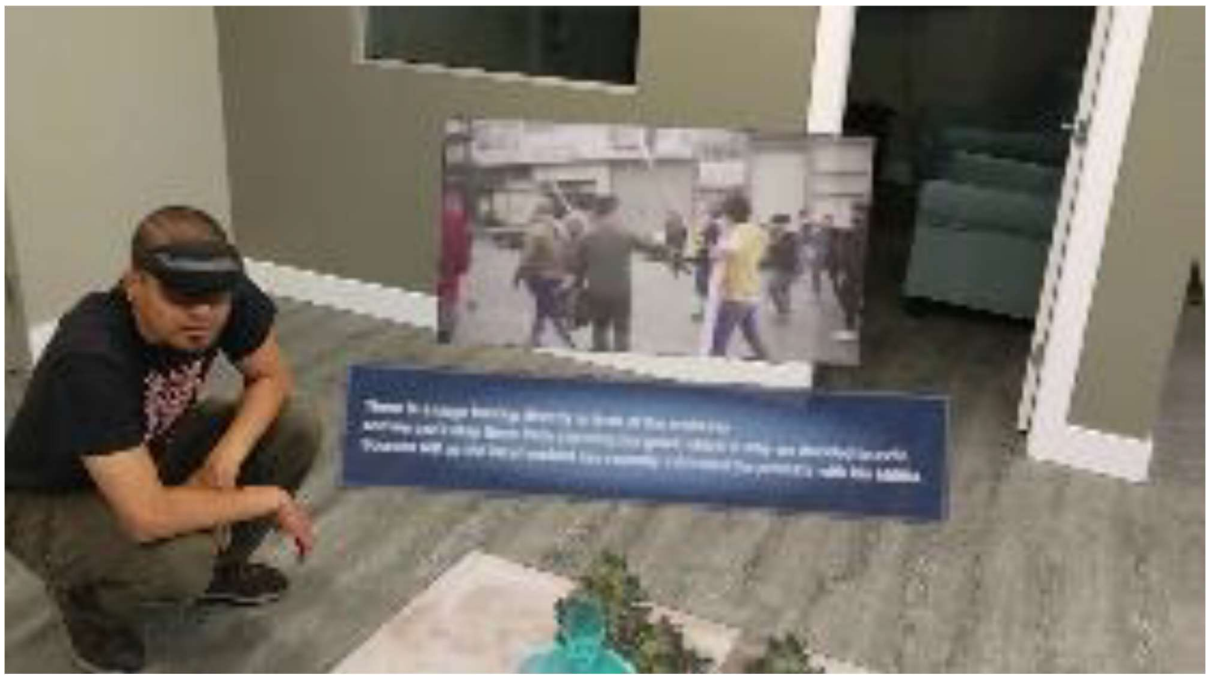
Another example is Integrated Visual Augmentation System, or IVAS, which is a modified version of the Microsoft HoloLens 2 augmented reality headset being developed by the US Army. Primarily, IVAS's main function during combat is managing the coordination and communication between soldiers, for example a real-time 'mini-map' projected onto the floor that displayed the location and direction of teammates. It could also be used to highlight the real-time location of enemies or targets or to map out escape routes. To provide data when a soldier is unable to check the map, a compass ring is displayed at the top of a soldier's vision, giving information like the soldier's bearing, and the direction and distance to their teammates. While IVAS can be very potent in these situations, where it really shines is negating disadvantages to give one side the edge.

With an advanced holographic processing unit (HPU), next-generation sensors and displays, and revolutionary hand and eye tracking, building immersive experiences is now truly possible with HoloLens 2. You can use popular development platforms such as Unity, Unreal, and Vuforia to create your mixed reality experiences, and get built-in HoloLens 2 developer support and HoloLens 2 supports an open API surface and driver model—and Microsoft continues to support open standards such as Khronos and OpenXR. This is supported with Azure mixed reality services including spatial anchors, remote rendering and digital twins with support for development for HoloLens, iOS, and Android.

4.2 Collaborative AI – agents collaborate with humans

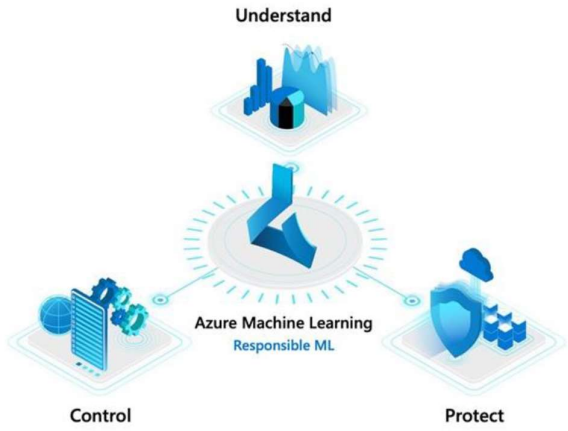
The focus of collaborative AI is to drive state of the art research in reinforcement learning to enable novel applications in modern simulation and games, in particular: **agents that learn to collaborate with human players**. Our goal is to democratize state-of-the-art reinforcement learning techniques for any developer through Azure Machine Learning, and to provide easy access to training infrastructure and integration tools. In contrast to traditional approaches to crafting the behavior of bots, non-player characters, or other in-game characters, reinforcement learning does not require a game developer to anticipate a wide range of possible game situations and map out and code all required behaviors. Instead, with reinforcement learning, game developers control a reward signal which the game character then learns to optimize while responding fluidly to all aspects of a game's dynamics. The result is nuanced situation and player-aware emergent behavior that would be challenging or prohibitive to achieve using traditional Game AI.

Project Paidia focuses on learning a particularly challenging type of behavior: collaboration with human players such as commanders or warfighters in order to resolve certain mission planning and simulation situations. Because human players are notoriously creative and hard to predict, creating the experience of genuine collaboration towards shared goals has long been elusive. Together with colleagues at Ninja Theory, the MSR team identified a perfect test bed for driving this research, Ninja Theory's latest game Bleeding Edge. Bleeding Edge is a team-based game, and includes a range of characters that have to work together to score points and defeat their opponents. In their latest demo, the team showcases how reinforcement learning enables agents to learn to coordinate their actions.



5 Trust

Throughout the development and use of AI systems, trust must be at the core. Trust in the platform, process, and models. Political and military decision makers in NATO have strong history in making responsible use of "classic" capabilities. It's our shared responsibility as a technology community to bring them to the same trust level for the AI and machine learning technologies and bridge the gap between business leadership and data/AI scientists. At Microsoft, responsible machine learning encompasses the following values and principles:



- Understand machine learning models
  - o Interpret and explain model behavior
  - o Assess and mitigate model unfairness
- Protect people and their data
  - o Prevent data exposure with differential privacy
  - o Work with encrypted data using homomorphic encryption
- Control the end-to-end machine learning process
  - o Document the machine learning lifecycle with datasheets

As artificial intelligence and autonomous systems integrate more into the fabric of defense and intelligence, it's important to proactively make an effort to anticipate and mitigate the unintended consequences of these technologies. Microsoft has already adopted our own set of self-regulatory principles and measures on AI. As part of this work, we have established an ethical board and internal processes that guide how we develop, deploy and market our AI products and services. We are operationalizing responsible AI across Microsoft through a central effort led by Microsoft's AI, Ethics, and Effects in Engineering and Research (AETHER) Committee and its working groups along with our Office of Responsible AI (ORA). Together, Aether and ORA work closely with our responsible AI advocates and teams to uphold Microsoft responsible AI principles in their day-to-day work.

Together with MITRE, and contributions from 11 organizations including IBM, NVIDIA, Bosch, Microsoft has also released the Adversarial ML Threat Matrix, an industry-focused open framework, to empower security analysts to detect, respond to, and remediate threats against ML systems. During the last four years, Microsoft has seen a notable increase in attacks on commercial ML systems. Market reports are also bringing attention to this problem: Gartner's Top 10 Strategic Technology Trends for 2020, published in October 2019, predicts that "Through 2022, 30% of all AI cyberattacks will leverage training-data poisoning, AI model theft, or adversarial samples to attack AI-powered systems." For instance, in 2020 we saw the first CVE for an ML component in a commercial system and SEI/CERT issued the first vuln note bringing to attention how many of the current ML systems can be subjected to arbitrary misclassification attacks assaulting the confidentiality, integrity, and availability of ML systems. The academic community has been sounding the alarm since 2004, and have routinely shown that ML systems, if not mindfully secured, can be compromised. Microsoft worked with MITRE to create the Adversarial ML Threat Matrix, because we believe the first step in empowering security teams to defend against attacks on ML systems, is to have a framework that systematically organizes the techniques employed by malicious adversaries in subverting ML systems. We hope that the security community can use the tabulated tactics and techniques to bolster their monitoring strategies around their organization's mission critical ML systems.

This initiative is part of Microsoft's commitment to develop and deploy ML systems securely. The AI, Ethics, and Effects in Engineering and Research (Aether) Committee provides guidance to engineers to develop safe, secure, and reliable ML systems and uphold customer trust. To comprehensively protect and monitor ML systems against active attacks, the Azure Trustworthy Machine Learning team routinely assesses the security posture of critical ML systems and works with product teams and front-line defenders from the Microsoft Security Response Center (MSRC) team. The lessons from these activities are routinely shared with the community for various people:

- For engineers and policymakers, in collaboration with Berkman Klein Center at Harvard University, we released a taxonomy documenting various ML failure modes.
- For developers, we released threat modeling guidance specifically for ML systems.
- For security incident responders, we released our own bug bar to systematically triage attacks on ML systems
- For academic researchers, Microsoft opened a $300K Security AI RFP, and as a result, partnering with multiple universities to push the boundary in this space.
- For industry practitioners and security professionals to develop muscle in defending and attacking ML systems, Microsoft hosted a realistic machine learning evasion competition.

To learn more about this effort, visit the Adversarial ML Threat Matrix GitHub repository and read about the topic from MITRE's announcement, and SEI/CERT blog.

Virtual Mission Rehearsal for Special Operations Forces: A Sweden-U.S. Collaborative Effort

Emilie Reitz
Joint Staff J6
Joint Fires Integration Division

Capt. Robert Wilson
Swedish Armed Forces Headquarters
Department of Training & Evaluation

Kevin Seavey
Joint Staff J6
Joint Fires Integration Division

Overview

Bold Quest is a U.S. Joint Staff coalition capability demonstration and assessment event where nations, Services and program offices pool resources in a recurring cycle of capability development, demonstration and analysis. Sweden and the U.S. are long term partners in Bold Quest (BQ), and this close continuing partnership fostered unique collaboration opportunities.

In early 2019 the Swedish Armed Forces proposed a small, multi-phase BQ 20.1 event to be conducted in Sweden in May 2020. The live exercise phase of this event was to be focused on exercising Special Operations Forces (SOF) in sensor-to-shooter (S2S) vignettes to develop procedural and technical interoperability in joint intelligence, surveillance and reconnaissance (ISR), joint targeting, joint fires and joint situational awareness. Participants in the live exercise were to include a Swedish SOF Tactical Headquarters (SOFTAC HQ) exercising command and control (C2) over four participating SOF teams (2 x Swedish, 2 x U.S.). The SOF S2S live exercise was to be conducted at the Vidsel Test Range in northern Sweden.

Unfortunately, after almost eight months of detailed planning, this event was cancelled due to the COVID-19 pandemic. However, this paper shares the details of the planning and lessons learned about using virtual simulators to support mission rehearsal. All participants hope to revitalize a version of this experimentation plan once regular international travel resumes.

One of the primary Swedish Armed Forces objectives for BQ 20.1 was to explore the use of simulation to facilitate development of new training and mission rehearsal capabilities. To support this effort, SOF teams would conduct virtual mission rehearsals prior to live missions to compare the utility of conducting mission rehearsals in a virtual environment against the utility of traditional mission preparation not using simulation. Some SOF teams would conduct virtual mission rehearsals (the experimental group); others would not (the control group). Data from both groups would be collected and analyzed to determine if there were any differences in knowledge, skills or performance.

1 Background

According to U.S. Army mission command doctrine, a rehearsal is "a session in which a staff or unit practices expected actions to improve performance during execution." U.S. joint special operations doctrine states "thorough mission planning and, whenever possible, mission rehearsals are typically essential to success." Accordingly, the goal of the mission rehearsals during BQ 20.1 was to provide SOF teams the opportunity to rehearse their plans in a realistic virtual environment prior to mission execution. Mission rehearsals would be built into the daily battle rhythm as part of the SOFTAC HQ's overall C2/orders process.

In general, mission rehearsals focus on mission execution. They are most useful if carried out after a plan is developed for the mission. Mission rehearsals supported by virtual simulation can play a key role in mission readiness by allowing the participants to visualize the fight; ensure all team members understand their roles; exercise the plan in a virtual replication of the real world; identify areas of potential confusion or friction; and build a shared understanding within the team of how actions at the objective should unfold. Mission rehearsal may also allow the team to recognize holes in their plan and make changes to improve their tactics during execution.

There is a long history of research supporting the value of virtual simulation for mission rehearsal (Knerr, et al., 2003). For example, the research indicates virtual mission rehearsals can be effectively used to improve understanding of mission plans, improve target detection of objects at great ranges and enhance route knowledge within buildings (Horner, et al., 2010). As any athlete or musician knows, practicing specific behaviors or tasks over time improves performance. And, as a general rule, the more time you spend practicing, the better your performance will be.

However, there is also some evidence showing extended and repetitive practice can result in a loss of focus or concentration (Oulasvirta & Ericsson, 2009). To avoid this, trainers should ensure the rehearsals include varied scenarios that expose the SOF team to different situations and allow them to respond accordingly. There is also some evidence that self-paced practice is more effective than fixed-pace.

2 Methodology

The daily schedule of events planned for BQ 20.1 is depicted below in Figure 1. As this schedule shows, the four SOF teams were to cycle through consecutive days of mission preparation, mission execution and recovery. Two teams would begin the cycle on 25 May, the other two on 26 May. That cycle would continue through 1 June, after which the four teams all conduct mission preparation on 2 June and mission execution on 3 June. On the mission preparation days, one SOF team would conduct traditional planning and preparation all day while the other team conducted traditional planning in the morning and virtual rehearsals in the afternoon. Since rehearsals are most useful if carried out after a plan is developed, all mission rehearsals would be conducted after the SOF teams had sufficient time to plan their mission.

In addition to the mission rehearsal, all four teams have one period of distributed training scheduled on one of their recovery days. This distributed training would not be part of the live mission preparation the SOF teams are conducting at Vidsel. Instead, it was designed to be supplemental training with U.S. Air Force Special Operations Command (AFSOC) aircrew personnel on site in Sweden.

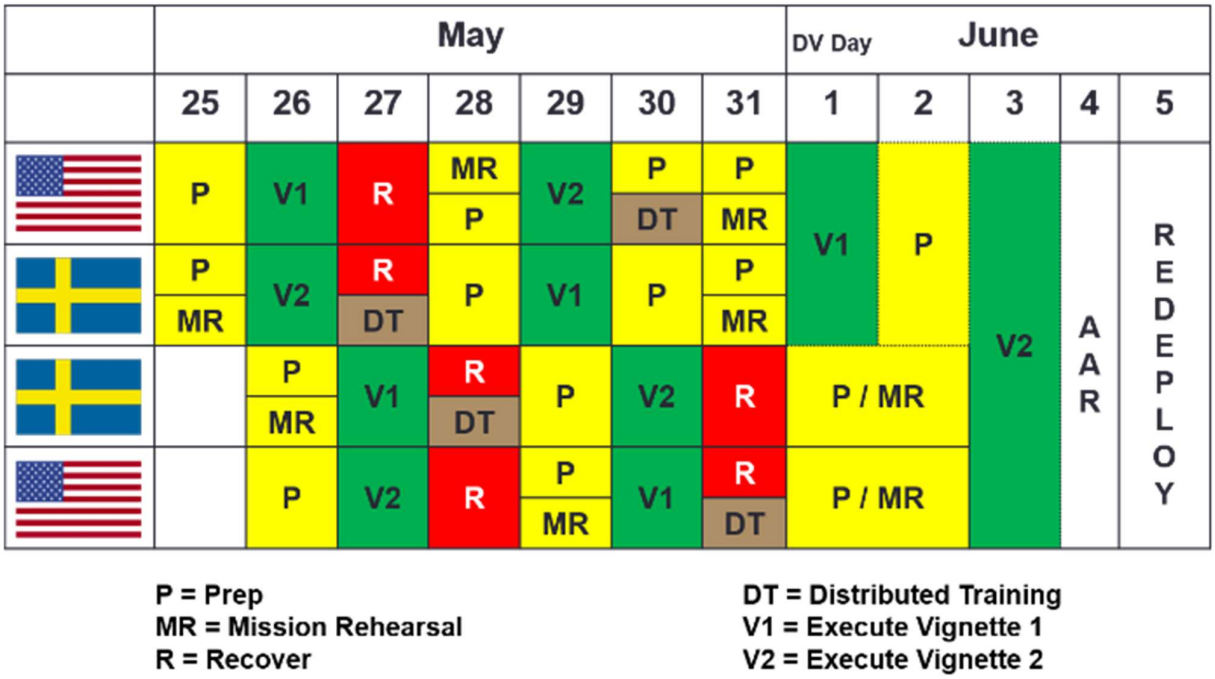


Fig. 1 - BQ 20.1 SOF S2S Execution Schedule

Prior to beginning the schedule above, each SOF team would complete informed-consent forms, demographic surveys and self-efficacy surveys in their native language. Additionally, all four teams would cycle through a familiarization session with the virtual mission rehearsal team from 21-22 May to understand the mission rehearsal methodology and get hands on experience with the virtual systems.

The first step in conducting each virtual rehearsal would be a discussion between the SOF Team Leader and the virtual rehearsal team. The SOF Team Leader would describe the plan the SOF team had developed for execution. In order to generate the virtual rehearsal scenario, the virtual rehearsal team would provide a standard template to ensure the following specific mission details are captured: insertion method and location, scheme of maneuver, reference points, the

team's concept of fires, anticipated location of targets and threats, Close Air Support (CAS) and ISR support availability, communications plan, extract plan, etc. Additionally, a SOF Subject Matter Expert (SME) supporting the virtual mission rehearsal would be assigned to shadow the SOF team during its morning planning in order to better understand the full plan the SOF team had developed. A thorough understanding of the plan is critical for the virtual rehearsal team to accurately represent the environment the SOF team expects to be entering.

After the discussion, the virtual rehearsal team would rapidly build the scenario. When the scenario is complete and the SOF Team Leader verifies it is correct, the SOF team would walk through the mission in a virtual 3D representation of the exact terrain and/or buildings they will operate in during the real-world mission. The initial rehearsal would be self-paced with no opposing force (OPFOR). Following a quick after action review (AAR), the SOF Team Leader would recommend any changes that need to be made to the scenario and the virtual team would modify it as required. For the second run through the scenario, the teams would conduct the mission rehearsal again *with* OPFOR. The pace of the second run and the amount and type of OPFOR would be up to the SOF Team Leader. Following the second rehearsal the SOF Team Leader would conduct an after-action review (AAR) with his team. Some of the virtual simulators being used support rapid playback of the virtual scenario that the SOF Team Leader could use as an AAR tool.

Throughout this process, the SOF Team Leader would be supported by simulation SMEs operating the virtual mission rehearsal tools. While these experts would advise the SOF Team Leader on virtual capabilities and offer recommendations, the SOF Team Leader would be the one responsible for successful execution of his team's mission. Therefore, he would retain overall control of planning and executing the virtual mission rehearsal.

During the rehearsal, exercise Observers/Controllers (O/Cs) would observe the team's performance and provide feedback to the analysis team on utility of the mission rehearsal. O/Cs would also participate in the team's AAR as appropriate. O/Cs would be able to compare performance during the mission rehearsal to that observed during the upcoming live mission to help assess the rehearsal's value.

At completion of the virtual mission rehearsal, the SOF team members would all complete self-efficacy surveys.

3 Conclusions

During the planning process leading up to the execution of BQ 20.1, we learned a few key lessons that will improve experiences for both teams involved as we continue to explore the field of virtual mission rehearsal for SOF. In creating a virtual mission rehearsal environment for easily accessible, regular use by SOF team members, the utility of skilled simulation operators is often overlooked. Until simulation software is of a quality where it can be easily manipulated by all skill levels of user, enhancements to training and mission performance, such as virtual mission rehearsals, will require expert support by simulation operators who are well versed enough in SOF techniques that they are able to translate user requirements to their systems with nuance.

A secondary challenge that the training and simulation community faces as we move to operationalizing simulations is the sourcing and classification of terrain data. In our coalition environment, data provided from open source or governmental groups at an unclassified level often end up elevated to a higher level of classification simply because that is how a design group is accustomed to operating.

And finally, we must improve how we transition these innovative systems to today's networked environments. Plans to enhance mission rehearsal systems with offsite simulation capabilities, allowing forces located hundreds or thousands of miles apart to practice together prior to mission execution are consistently thwarted by a lack of persistently connected simulation networks and by the security challenges of reaching across the boundaries created by the semi-closed networks these systems operate on.

4 Next Steps

Regularly using simulation for mission planning and rehearsal during the military decision making process (MDMP) could enhance effectiveness of the entire process. If time is an issue, which it usually is, the use of simulation could speed up the process. If time is not a critical factor, simulation could allow the team to speed up their own course of analysis development and allot more time to mission rehearsals. Further work needs be

done to develop a methodology for using simulation during the MDMP to support Mission Rehearsal. The authors look forward to working with those interested in this effort.

References

Horner, D., Eslinger, O., Howington, S., Ketcham, S., Peters, J., & Ballard, J. (2010). Integrated high-fidelity geoscience simulations for enhanced terrain-related target detection. Computing in Science & Engineering, 12(5), 56-63.

Knerr, B. W., Lampton, D. R., Thomas, M., Corner, B. D., & Grosse, J. R. (2003). Virtual environments for dismounted soldier simulation, training, and mission rehearsal: Results of the FY 2002 culminating event (No. ARI-TR-1138). ARMY RESEARCH INST FIELD UNIT ORLANDO FL.

Oulasvirta, A., & Ericsson, K. A. (2009). Effects of repetitive practice on interruption costs: An empirical review and theoretical implication. In Proceedings of the European Conference on Cognitive Ergonomics 2009 (pp. 333-340).

Aviator Training Next

Virtual Reality Pilot Program

MAJ Chris McFarland
Chief Operations Officer
U.S. Army Aviation Center of Excellence
Directorate of Simulation, Fort Rucker, AL.

The Army's current portfolio of live, virtual, constructive, and gaming capabilities vary in diversity. Some of these training capabilities remain simple to use and meet a majority of basic training tasks across multiple branch proponent requirements. Other capabilities are immensely complex, integrate multiple software tools to increase training fidelity, and support distributed training requirements to widespread physical locations. Unfortunately, many of our tools support stove-piped requirements using dated technology requiring several levels of specialized contractor support. The use of simulation in aviation training has continued to increase over the last forty years in search of the best ratio of live flight training and training conducted in a simulator.

The U.S. Army Aviation Center of Excellence (USAACE) continually assesses opportunities to enhance current, and future, aviation-training efforts with cutting-edge simulation capabilities. Considering emerging lessons learned from the United States Air Forces' Pilot Training Next (PTN) fixed-wing training initiative, USAACE began a series of test programs at Fort Rucker, starting August 2019, to assess the effectiveness of Virtual Reality (VR) technology to supplement and enhance Initial Entry Rotary Wing (IERW) training capabilities. Initial feedback from this ongoing study indicates that commercial off the shelf technology (COTS) VR can support valid pilot training for initial entry students.

The overarching goal of the USAACE program, called Aviator Training Next (ATN), is to produce more proficient initial entry students by reinforcing basic flight maneuver tasks. ATN design incorporated the basic concept of a flight-training program increasing frequency and repetition using a low-cost COTS virtual trainer, commercial flight training software, and cognitive measurement assessments.

Planning and executing the ATN program required the clear identification of responsibilities across a multifunctional team to facilitate proactive collaboration. The USAACE Directorate of Simulation (DOS) assumed responsibility as the project management lead and worked directly with Aviation and Missile Center (AvMC) and the contractor Science Applications International Corporation (SAIC) to establish the

ATN technical capability. 110th Aviation Brigade (110th AB) and the Directorate of Training and Doctrine (DOTD) led the experimental training design and integration effort. The United States Military Academy (USMA) Operations Research Center (ORCEN) and US Army Aeromedical Research Laboratory (USAARL), working with 110th AB and DOTD, defined appropriate data collection and assessment metrics to enable a comparison between students enrolled in VR training and those in traditional flight training. Using the established metrics, the research team developed an extensive Design of Experiment (DOE) that intended to answer the most important question: Can the ATN device replicate the training tasks, conditions, and standards in accordance with the Aircrew Training Manual (ATM)?

As designed and implemented, ATN incorporated the most current generation of COTS VR capabilities, along with advances in learning science, into the Basic Army Aviator Course (BAAC) portion of the overall IERW program. ATN focused on the performance of students culminating with their second major performance check ride (P2) after approximately day 45 of training.

Each ATN flight class had a control group and two VR groups ranging from 10-12 students in size. These two VR groups executed slightly different Course Management Plans (CMPs) in order to determine the most effective mix of VR simulation flight periods with lesser amounts of live flight training to assess proficiency in base tasks against the control group that flew the current BAAC CMP.

The ATN VR training environment captured flight profile expectations and tracked an array of performance data while students flew maneuvers in training. The data collected on student performance, as well as cognitive/physiological measurement and subjective assessments through student surveys; provide instructor pilots necessary information to advise their assessments of VR student progress in terms of proficiency, comprehension, and overall progress of flight skill development. The research investigated seven CMPs with 750 students (360 non-VR and 390 VR) across six cohorts.

USAARL, co-located at Ft. Rucker, provided government and academic researchers in cognitive science and human factors to collect and analyze data from the program, while ORCEN, serving as the research lead, provided independent research on metrics, design, and assessment of the impact of VR simulations on pilot cognition and overall performance. They intend to publish a series of formal research reports in peer-reviewed publications.

The final data collection window for this iteration of ATN culminated July 2020 allowing the research team to provide a detailed analysis report to USAACE leadership. This report provides findings and recommendations on the future of ATN. ORCEN provided a preliminary research brief stating, "The findings support the conclusion that the ATN program is a viable training delivery method." Some key findings include that P2 check ride performance between VR classes is not statistically different and that check ride scores for both P1 and P2 showed no statistical difference between VR and non-VR students.

What does this mean? Statistically there is no difference on how ATN students and flight school students using traditional flight school CMPs perform; however, data shows that ATN students continually out-perform their peers in the aircraft check rides as well as in academics beyond the ATN training window. The current data demonstrates promising future training capability using ATN and other VR technologies. With the initial research complete and showing promising results, researchers from United States Military Academy (USMA) Operations Research Center (ORCEN) and US Army Aeromedical Research Laboratory (USAARL) are continuing to examine vast amounts of data to help USAACE leadership shape and determine the extent of the next round of experiments.

Continued after action reviews and lessons learned allow for recommendations on how to improve the program. Primary recommendations include a deeper assessment into how many iterations it takes a student to become proficient in each maneuver and looking at further implementation of VR into additional phases of IERW, e.g. instruments. Focused observation and evaluation of performance in the ATN device and live aircraft using a simple metric of training efficiency ratio (TER) will allow USAACE to accomplish the first recommendation. Future research plans and the feasibility of implementing ATN and/or VR into additional phases of IERW to provide additional research data are being refined.

ABOVE THE BEST!

Using Virtual Reality for Collaborative Immersive Operations and Planning

John Nicol
CEO, Corona Aerospace Inc

Abstract

Problem

Managing operations, such as military operations, disaster response, or within large enterprise organizations is hard. Gaining critical situational awareness in a complex, data-rich environment is harder. Managing large geographically dispersed teams and sharing a common understanding of complex data using just teleconference or webex is almost impossible. Additional workplace challenges with the COVID-19 pandemic have highlighted the difficulty of remotely managing and planning complex operational situations.

Existing online collaboration applications, particularly those using Virtual Reality solutions are generally limited to replicating an office environment with office tools. Very few, if any, are geared towards real operations or enterprise management and collaboration using live data feeds to visualize spatial data in any meaningful way, or to reach out from the virtual world into the real world to effect real systems.

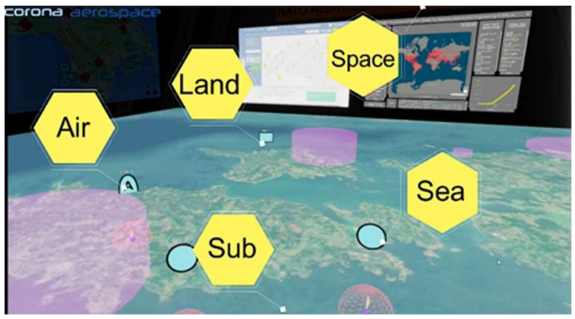
A Solution

A data driven augmented, mixed and virtual reality collaboration (eXtended Reality or XR) application called 'NexusXR' has been developed in Canada and has been used in Canadian Joint Warfare and NATO experiments to provide operations and planning staff the ability to work together in a distributed, immersive environment anywhere, anytime in virtual reality. The technology enables data from real systems, simulated platforms or digital twins to be visualized and manipulated for planning or real-time operations.

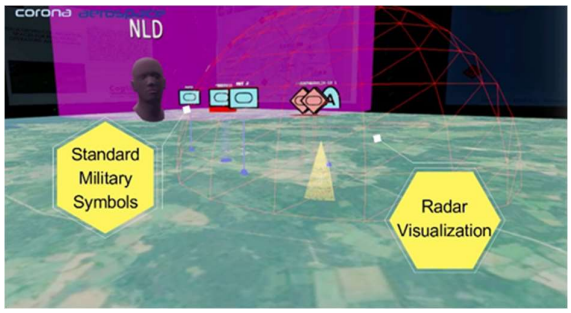
Virtual Command Post and Common Operating Picture

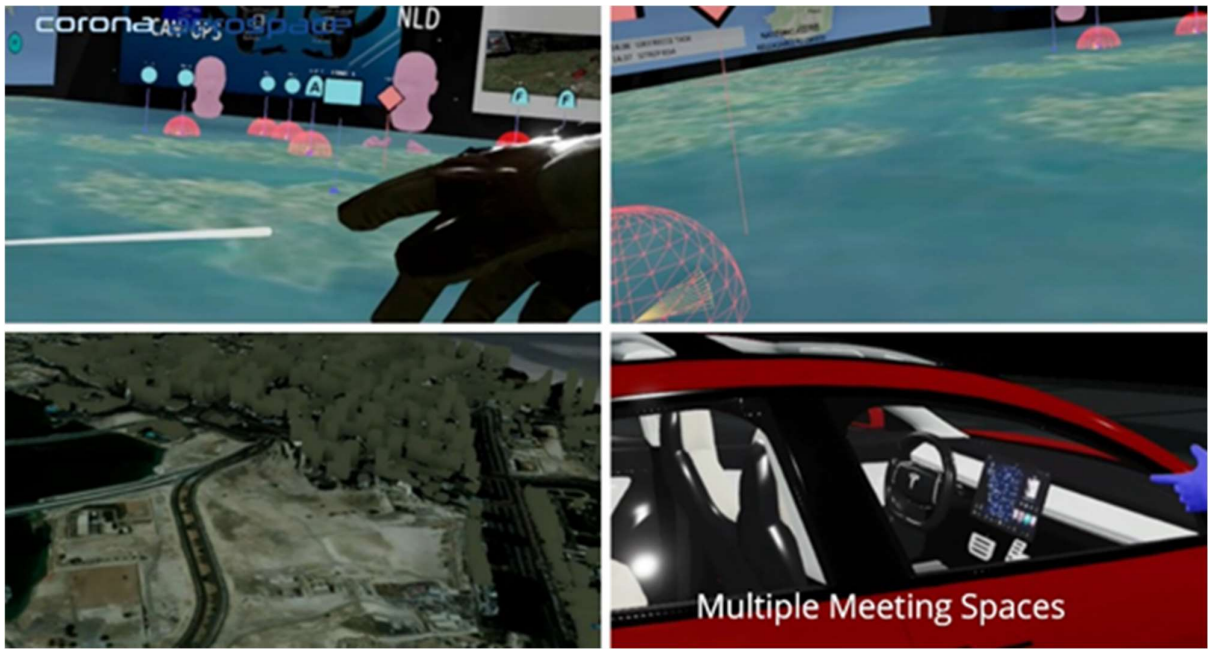
The application is used as a virtual Command Post with an immersive three dimensional Common Operating Picture (COP). Data feeds from military command and control systems, video streams from aircraft, civilian air and sea traffic and distributed simulation systems are all visualized on a world-wide 3D virtual terrain map. Additional data feeds can be supported as well as two way data whereby users can change data within the environment to affect real-world systems. Imagery and web pages can be placed on the virtual walls if they are needed for briefing, or for real time information updates during operations.

Users are able to interact with the data and move around the environment to view it from any angle. Data can be filtered and different layers of information can be turned on or off by users to declutter the COP. Operations Centers will be able to use the application locally within a Command Post and use Augmented Reality glasses. Remote users can join from anywhere on the network to share and interact with other users and the data as required.



The system supports multiple meeting/planning/COP rooms concurrently and users can move between rooms as required. This remote collaboration capability allows users from any location to walk around the virtual room, talk and interact with each other to discuss the operation in progress and send text messages and is a far more natural way to interact than other methods. The technology is particularly good at allowing users to see 3D spatial data and inter-relationships that enhance situational awareness. The Command Post of the Future is here.





A Brief Introduction to Concrete Algorithmic Game Theory Applications

Nicola Gatti

Extended Abstract

The birth of the first Security Game (SG) is due to John Von Neumann, with the Hide and Seek Game [1]. The scenario is the following: a player hides in a place among a finite number of them, and the (unique) opponent should find her. It is modeled as a normal-form zero-sum game, as the goals of the two players are precisely one the opposite of the other.

From this simple game, many directions have been investigated, originating a lot of works in the following years, where some fugitives are escaping from pursuers that want to reach them [2]. If the fugitive tries to reach a target, e.g., a vanishing point, while the pursuer has to stop her, we have Ambush Games [3], while if the fugitive hides and the pursuers look for her, we have Search Games [4]. Finally, if both fugitives and pursues can move in the environment, we call them Infiltration Games [5].

These studies evolved in research about strategic resource allocation for security, which has been a very prolific domain in the field of algorithmic game theory during the last years. The investigation in this domain led to the development of what today are commonly called Security Games (SGs): game-theoretical frameworks for computing resource allocation strategies against adversarial security threats. Security is one of the most critical issues every country and every person deals with every day: the protection of airports, ports, banks, monuments, and museums, but also containing urban attacks, controlling poaching of endangering species, preventing the diffusion of misinformation and guaranteeing cybersecurity [6].

Customarily, SGs are a mathematical tool to model the protection of infrastructures or open environments as a non-cooperative game between a Defender and an Attacker. Given the setting, these scenarios take place under a Stackelberg (a.k.a. leader-follower) paradigm [7], where the Defender (leader) commits to a strategy and the Attacker (follower) first observes such commitment, then best responds to it. From a computational perspective, as discussed in [8], finding a leader-follower equilibrium is computationally tractable in games with one follower and complete information, while it becomes hard in Bayesian games with different types of Attacker. The availability of such computationally tractable aspects of Security Games led to the development of algorithms capable of scaling up to huge problems, making them deployable in the security enforcing systems of several real-world applications.There have been several applications based on such games, and we describe some of them. The first one to be deployed is ARMOR, consisting of the strategic placement of checkpoints on the streets leading to the Los Angeles International Airport (LAX) and the management of the patrolling units across the terminals [9, 10]. The authors cast the problem as a Bayesian Stackelberg game, giving the guards the possibility to assign different and appropriate weights to their actions and tune them with respect to the types of the adversary. In [11], the problem of scheduling undercover air marshals on U.S. domestic flights has been tackled with the project IRIS. Here, additional constraints have been introduced since the agents must fly among cities such that, the next day, they will depart from the same city they landed the day before. Moreover, the agents are scheduled to have a list of cities such that the first and last cities are the same so that they actually fly around following a circle.

Preventing crimes or terrorist attacks in urban areas is the problem that has been tackled in [12]. Guards must respond very quickly to be able to intercept and catch a potential Attacker on her escaping route, which could depend on time-dependent traffic conditions on transportation networks. The primary challenge here consists of the presence of time constraints both on the Defender and the Attacker side.

Very recently, Security Games have been applied to stop nuclear smuggling in international container shipping through advanced inspection facilities [13]. Efficiency and efficacy are fundamental for this task, given that there are millions of containers, which should be screened. This work models the interaction between an inspector and a smuggler using a security game, formulating the smuggler's sequential decision behavior as a Markov Decision Process.

Recently, game-theoretic techniques have also been applied to cybersecurity. In [14, 15], the authors study the problem of protecting a network in which an administrator may decide the best security measures to use to improve the safety of the network. This is achieved by resorting to honeypots, i.e., decoy services or hosts, placed by the Defender, while the Attacker chooses the best response as a contingency attack policy.Still in a cybersecurity dimension, [16] proposes an approach to investigate whether alerts generated by potential cyber-

attacks are real attacks or just false positives. Also here, the magnitude of the problem is high, with a number of alerts that are overwhelming with respect to the number of analysts that can check the authenticity of such attacks.

Bibliography

[1] M. M. Flood, The Hide and Seek Game of Von Neumann, Management Science 18 (5-part-2) (1972) 107–109.

[2] M. Adler, H. Racke, N. Sivadasan, C. Sohler, B. Vocking, Randomized Pursuit-Evasion in Graphs, Combinatorics, Probability and Computing 12 (2003) 225–244.

[3] W. Ruckle, R. Fennell, P. T. Holmes, C. Fennemore, Ambushing Random Walks I: Finite Models, Operations Research 24 (2) (1976) 314–324.

[4] S. Gal, Search Games, Academic Press, 1980.

[5] S. Alpern, Infiltration Games on Arbitrary Graphs, Journal of Mathematical Analysis and Applications 163 (1992) 286–288.

[6] D. Kar, T. H. Nguyen, F. Fang, M. Brown, A. Sinha, M. Tambe, A. X. Jiang, Trends and Applications in Stackelberg Security Games, Handbook of Dynamic Game Theory (2017) 1–47.

[7] B. Von Stengel, S. Zamir, Leadership with Commitment to Mixed Strategies, Tech. rep. (2004).

[8] V. Conitzer, T. Sandholm, Computing the Optimal Strategy to Commit to, in: ACM conference on Electronic Commerce (EC), 2006, pp. 82–90.

[9] P. Paruchuri, J. P. Pearce, J. Marecki, M. Tambe, F. Ordonez, S. Kraus, Playing Games for Security: An Efficient Exact Algorithm for Solving Bayesian Stackelberg Games, in: International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2008, pp. 895–902.

[10] J. Pita, M. Jain, J. Marecki, F. Ordonez, C. Portway, M. Tambe, C. West- ern, P. Paruchuri, S. Kraus, Deployed ARMOR Protection: The Application of a Game-theoretic Model for Security at the Los Angeles Interna- tional Airport, in: International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2008, pp. 125–132.

[11] J. Tsai, S. Rathi, C. Kiekintveld, F. Ordonez, M. Tambe, IRIS - A Tool for Strategic Security Allocation in Transportation Networks, in: International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2009, pp. 1327–1334.

[12] Y. Zhang, B. An, L. Tran-Thanh, Z. Wang, J. Gan, N. R. Jennings, Optimal Escape Interdiction on Transportation Networks, in: International Conference on Artificial Intelligence (IJCAI), 2017, pp. 3936–3944.

[13] X. Wang, Q. Guo, B. An, Stop Nuclear Smuggling Through Efficient Container Inspection, in: International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), 2017, pp. 669–677.

[14] K. Durkota, V. Lisy, B. Bosansky, C. Kiekintveld, Approximate Solutions for Attack Graph Games with Imperfect Information, in: International Conference on Decision and Game Theory for Security (GameSec), 2015, pp. 228–249.

[15] K. Durkota, V. Lisy`, B. Bosansky, C. Kiekintveld, Optimal Network Security Hardening Using Attack Graph Games, in: International Joint Conference on Artificial Intelligence (IJCAI), 2015, pp. 526–532.

[16] A. Schlenker, H. Xu, M. Guirguis, C. Kiekintveld, A. Sinha, M. Tambe, S. Sonya, D. Balderas, N. Dunstatter, Don't Bury your Head in Warnings: A Game-Theoretic Approach for Intelligent Allocation of Cyber-security Alerts, in: International Joint Conference on Artificial Intelligence (IJCAI), 2017.

Introduction to Artificial Intelligence

Emanuela Girardi, Piero Poccianti

1 AI in an international context

Artificial intelligence is considered the strategic technology for developing the society of the future, the Russian President Vladimir Putin recently said that who will dominate the AI field will dominate the world. That makes AI technologies the new focus of the international competition. All the developed countries started the AI race investing intensively in AI research and development.

Today we can identify three main AI blocks with different approaches to AI research: the American, the Chinese and the European. In the USA the research is mainly carried out by multinationals (the FAAMG[2]). In China, the research is mainly led by government agencies. In Europe there is a more balanced model that proved successful until few years ago (if we consider the number of research papers) but the situation is rapidly changing due to the lack of AI investments by European governments and industry. Only recently, the European Commission started investing again in AI and developed an AI strategy starting from the ethical guidelines for a trustworthy and human-centric AI. Europe is working on the definition of a global AI governance to ensure that these new technologies will improve the life of the whole humanity and not harm humans.

2 What is AI?

The first studies into Artificial Intelligence began in 1943 with an article by MC Culloch and Pitts (a physiologist and a mathematician) describing the functioning of an artificial neuron, and in 1950 with a paper written by Alan Turing on the notion of machines being able to simulate human beings and the ability to do intelligent things, such as play Chess. However, the term artificial intelligence was first coined by John McCarthy in 1956 when he held the first academic conference on the subject.

Artificial intelligence (AI) is a discipline in the field of computing, that aims to make a machine do things that if they were done by a human, we would call intelligent. The European Commission defines AI systems as software systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal.

AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behavior by analysing how the environment is affected by their previous actions.

As a scientific discipline, AI includes several approaches and techniques, such as "**Machine Learning"** (of which deep learning and reinforcement learning are specific examples), "**Machine Reasoning"** (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and "**Robotics**" (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).[3]

Besides the great achievement of AI and the strong impact it has on several aspects of our daily lives, we are still dealing with the so-called "**Narrow AI"**, or better: systems that can overachieve human performance in some specific fields and applications, but are still missing the ability to generalize knowledge, to apply the abilities they learn in a field in many different contexts, to experience emotions and to become self-conscious. There are several studies that try to develop a "**General AI**" or a super intelligence, they currently belong to the world of science fiction and we still have a long way to go to develop a General AI.

3 Where are we today? The different stages of AI

The history of Artificial Intelligence is defined by moments of great enthusiasm, followed by great disappointments, the researchers talk about springs and

[2] Facebook, Amazon, Apple, Microsoft, and Alphabet's Google

[3] A definition of AI: Main capabilities and scientific disciplines by the High-Level Expert Group on Artificial Intelligence. The AI HLEG is an independent expert group that was set up by the European Commission in June 2018. Document made public on 8 April 2019. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

winters of AI. Today we are living in a moment of great enthusiasm for AI developments and applications thanks to the combination of three elements that have been globally available since the beginning of the XXI century: Big Data, increased computer power, and new algorithms.

Despite this enthusiasm, research in AI has still a long way to go. DARPA divided the AI research in 3 waves[4]:

1- Describe: Handcrafted Knowledge
2- Categorize: Statistical learning
3- Explain: Contextual adaptation

Each wave is defined by the level of ability of the AI systems in four human skills: perceiving, learning, abstracting and reasoning.

The first era, the handcrafted knowledge refers to AI systems that reason well but are not good in perceiving, humans must teach the machine the context in which to move, the goals and the tools. An example of this is planning and reasoning, practical uses of this are for example a chess playing computer or a route optimization algorithm.

The second era is the Statistical learning and is the one we are living in. AI systems (Deep Neural Networks) can perceive and learn but they still have little ability to abstract and reason. An example of this is the ability of interpreting an image after having been trained on millions of similar images such as recognising a cat in a picture but also the correct interpretation of medical X-rays.

The third era, the contextual adaptation, has not arrived yet. DARPA foresees AI systems that can perform well in all the four human skills. These AI systems will be able to deal with unforeseen situations, explain their behavior and learn faster from the experience. For example, in a simulation environment the machine can propose a decision in a new context never seen before.

4 Different fields of applications

AI applications can be used in several sectors to bring benefit to public administrations, industry, and civil society. The critical point is the ability to transfer the results of the research in AI into the society. The Italian National AI Strategy identifies six areas on which to focus the AI investments and where to promote the knowledge transfer:

1- IoT (Internet of Things), manufacturing and robotics
2- Services: finance, education, health
3- Transportation, agrifood, energy
4- Aerospace and defense
5- Public administration
6- Culture and digital humanities

To enhance the adoption of AI there are three main factors that need to be considered: the availability of data, infrastructure and digital skills. The European Commission is focusing its AI strategy on the creation of a single market for data[5], promoting the development of European data lakes on strategic sectors, a European Cloud infrastructure and a data governance law. But the most important point is to promote the AI culture, to enable citizens to understand and use these new technologies and to be able to actively participate in the new AI society. A remarkable program is "Elements of AI" an online AI course promoted by the Finnish government to create a common global culture on artificial intelligence [6].

5 Future of AI

According to DARPA the next wave of AI research will bring us to machines that will have all the four capabilities of humans: perceiving, learning, abstracting and reasoning.

The research is already focusing on this goal: teaching machines to learn from intuition and experience and to generalize the knowledge applying what they learnt in a filed to a different one.

One of the most interesting projects in this area is

[4] A DARPA Perspective on Artificial Intelligence. https://www.darpa.mil/attachments/AIFull.pdf

[5] A European strategy for data, February 2020. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

6 Elements of AI is already available in 18 languages https://www.elementsofai.com

TAILOR[7] promoted by CLAIRE[8] and funded by the Horizon 2020 program (ICT-48) of the European Commission. The purpose of TAILOR is to build the capacity of providing the scientific foundations for Trustworthy AI in Europe by developing a network of research excellence centres leveraging and combining learning, optimization and reasoning techniques. TAILOR will prepare the ground for AI research that addresses the grand challenges of our time in health, mobility and resource management.

Another interesting step toward the future of AI is the MuZero[9], an algorithm developed by DeepMind (Google) to master games without knowing their rules. MuZero learns a model that explains its environment and then use that model to plan the best course of action[10]. MuZero paves the way for learning methods in a host of real-world domains, particularly those lacking a simulator or dynamics rules.

Merging the different techniques of AI and the different disciplines will allow us to develop new AI systems that can help us solving the challenges of our society.

6 AI for a sustainability society

Considering the definition of AI proposed by the European Commission (systems that, given a complex goal, perceive their environment and act in an autonomous way to achieve the given goal), when we use AI technologies it is critical to define the right environment, the context, and most of all the goals that we want to reach.

The Italian AI Strategy[11] proposes to use Artificial Intelligence technologies to reach the 17 Sustainable Development Goals of the 2030 United Nations Agenda[12]. The United Nations strategy "Leave no one behind" can be a good starting point to create a more economical, social and environmentally sustainable society.

In summary, AI is the fourth industrial revolution, which makes it of enormous strategic importance and it will touch all aspects of society. It can be used for good but also to gain a competitive advantage over strategic competitors as Mr. Putin correctly observed. Therefore, it needs to be understood and managed in order to gain and maintain the world's equilibrium today.

7 TAILOR is one of four AI networks in the H2020 program ICT-48 Towards a vibrant European network of AI excellence centres. https://liu.se/en/research/tailor

8 Confederation of Laboratories for Artificial Intelligence Research in Europe. https://claire-ai.org/

9 MuZero: Mastering Go, chess, shogi and Atari without rules. https://deepmind.com/blog/article/muzero-mastering-go-chess-shogi-and-atari-without-rules

10 Paper: Mastering Atari, Go, chess and shogi by planning with a learned model. Published by Nature online on 23 December 2020.

11 Italian AI Strategy. https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf

12 The 2030 Agenda for Sustainable Development. https://sdgs.un.org/goals

Trustable and Ethical Artificial Agents: a Toolkit of Useful Techniques, and Human-AI Teaming?

Stefania Costantini
Università degli Studi di L'Aquila,
Dipartimento di Ingegneria e Scienze dell'Informazione, e Matematica
Via Vetoio snc, Loc. Coppito, I-67010 L'Aquila – Italy
stefania.costantini@univaq.it

Abstract

Autonomous Intelligent Agents and Autonomous Systems (AS), are nowadays employed in many important autonomous applications upon which the life and welfare of living beings and vital social functions may depend. Therefore, agents should be reliable and trustworthy. A-priori certification techniques can be useful, but are not sufficient for agents that evolve, and thus modify their epistemic and belief state, and possibly their objectives. In this paper we illustrate methods for keeping Artificial Intelligence (AI) applications under control, via techniques for run-time assurance, based upon introspective self-monitoring and checking, and via "White-Box" Machine Learning. The aim is to build a 'toolkit' to allow an agent designer/developer to ensure trustworthy and ethical behaviour. We however advocate human-AI teaming in critical applications, to exploit the strengths of both parties.

Autonomous Systems (AS) and more generally Intelligent (Software) Agents are applications of AI (Artificial Intelligence) that are nowadays adopted in many application contexts. AS can be described as complex software entities that are capable of acting with a certain degree of autonomy in order to accomplish tasks or enact behaviours. Intelligent Agents (cf. [1] for a survey of the features and capabilities of agents) are AS which are able to perform goal-directed behaviour (so-called "proactivity"), which are reactive to events that happen in their environment, and capable of intelligent decision-making without human intervention; agents may have social abilities, and in this case they can form Multi-Agent Systems (MAS). In this paper we see AS as a particular kind of agents, so we often use the two terms as synonyms. AS are becoming more technologically advanced every day, outperforming humans in an ever-growing number of fields – e.g., playing complex games [2], assisting elderly people [3], predicting cancer [4], driving cars [5], performing speech recognition [6]. They can be embodied in robots, of which an agent or a MAS can constitute the "brain". The wide adoption of such systems involves a promise of huge improvement of our quality of life. However, humans take an ambivalent stance w.r.t. AS. On the one hand, humans often fear that AS may overcome human control, and take decisions not aligned to human values. On the other hand, the growing success of intelligent systems in facing complex problems may easily lead to uncritical acceptance of their decisions, due, for instance, to the unwillingness to take responsibility. The future scenarios will presumably encompass a "hybrid society", where humans and intelligent autonomous agents will be coupled at multiple levels, hopefully on the basis on shared agreed-upon normative/moral standards. In fact, AS should in some sense "understand" and follow social norms – preventing them from exploiting vulnerabilities of humans – and they should be able to earn trust from other humans/AS in the hybrid society.

In this discussion we restrict ourselves to agent systems based upon computational logic, because they provide transparency and explainability 'by design', as logical rules can easily be transposed into natural-language explanations. Logic-based languages and architectures are discussed in the survey [7,8,9]. They are based (more or less directly) on the so-called BDI ('Belief, Desires, Intentions') model of agency [10]. Note that logical agents may encompass Machine Learning (ML) modules to perform many tasks of perception and classification (cf. [11] for a survey of Machine Learning techniques). Nonetheless, we remain convinced that an "upper layer" of explicit reasoning is an essential component of intelligence. This because ML modules do not really "learn", i.e., a neural network that recognizes whether an image represents a certain kind of object or situation is still not able to explain, e.g., to a human user which object is that, and why. Attempts at "Explainability" of "Black Box" ML modules exist (in the XAI, i.e., eXplainable AI field), but this research is still in its infancy (see [12]). This also because, quoting from [12], "*In order to reach human interpretability, one should first study and model how humans produce and understand explanations between each other and which properties make explanations perceivable to humans*". Instead, logical proofs follow in an abstract way the same reasoning paths as humans, and so (when translated into human-intelligible notation) they can be easily understood. Moreover, ML requires a lot of data which are not always available in real situations, where reasoning techniques are often able to

reach the same results in a much more efficient and data-inexpensive way. So, a suitable merge of the two techniques will bring the better results.

As a freely available framework for specifying agents and Multi-Agent Systems (MAS), one may consider DALI, invented by the author of this paper and developed by her research group. DALI [13-21] is an Agent-Oriented Logic Programming language, where the autonomous behaviour of a DALI agent is triggered by several kinds of events: external events, internal, present and past events. Reaction to "external events" is defined by reactive rules. The agent remembers to have reacted by converting all external events (after reaction) into "past events" (each one with its time-stamp). An event perceived but not yet reacted to is called "present event". It is often useful for an agent to reason about present events, that make the agent aware of what is happening in its external environment. In DALI, agents can perform actions, according to preconditions. Similarly to events, actions are recorded as past actions. "Internal events" is the device which makes a DALI agent proactive. In fact, their description is composed of two elements. The first one describes the conditions (knowledge, past events, procedures, etc.) that must be true so that the reaction (in the second rule) may happen, where such conditions are automatically attempted with a default frequency, customizable by means of user directives. Thus, a DALI agent is able to react to its own conclusions. DALI is equipped with an advanced communication architecture [50] which implements the FIPA protocol (where FIPA is a widely used standardized ACL, i.e., Agent Communication Language, cf. http://www.fipa.org/specs/fipa00037/SC00037J.html), plus an advanced filter on incoming and outcoming messages. DALI has been made compatible with the Docker technology (cf. [17] for details). So, a DALI agent can be deployed within a container. The semantics of DALI is based upon the declarative semantic framework introduced in [18], aimed at encompassing approaches to evolving logical agents, by understanding changes determined by external events and by the agent's own activities as the result of the application of program transformation functions. The DALI framework has been experimented in industrial applications (many of them covered by no-disclosure agreements) for: unattended hardware testing of hardware-software platforms in telecommunication industry; user monitoring and training; emergencies management (such as first aid triage assignment); security or automation contexts; home automation and processes control. DALI has also been exploited in cognitive robotics [19,20]. More generally, DALI has proved to be useful in every situation that is characterised by asynchronous events sources that require reasoning over a dynamic data collection: either simple events, and/or events that are correlated to other ones even in complex patterns. In fact, in order to be able to perform Complex Event Processing, i.e., to actively monitor event data so as to make automated decisions and take time-critical actions, DALI has been empowered with suitable capabilities [21]. DALI has been in fact fully implemented ([15], the DALI framework is publicly available), and a programming environment has been devised, with many features useful in practical applications, among which a cloud version of the implementation.

As every other piece of software, intelligent agents' code should be certified, prior to deployment. Most pre-deployment (or 'static' or 'a priori') verification methods for logical agents rely upon model-checking (cf. [22] and the references therein), and some (e.g., [23]) upon theorem proving. These techniques are able to certify 'a priori' that agents fulfil certain requisites of trustworthiness, that means that they do what is expected from them, and do not violate certain rules of behaviour. However, such techniques can be sufficient for agents that keep their epistemic state constant during their operation, and interact with the environment in a predefined way. So, they are not sufficient for agents that will revise their beliefs and objectives in consequence of the interaction with a changing and not always predictable environment. In many applications, an agent's epistemic state and thus an agent's behaviour is in general affected by its interaction with the external world, i.e., by which events are perceived by the agent and in which order. In many practical cases, the actual arrival order of events and the set of possible events is unknown, or however it is so large that computing all combinations would result in a combinatorial explosion, making 'a priori' verification techniques too heavy to apply and therefore unpractical. In agents that learn, it is not even possible to predict the set of events that will be observed and considered by an agent, that might therefore devise new objectives not necessarily in line with the expected agent's behaviour. The quest for methods for implementing Intelligent Agents so as to ensure transparent, explainable, reliable and ethical behaviour in a changing environment is due to the employment of agent systems in many important autonomous applications (such as, e.g., eHealth), where

the life and welfare of living beings and vital social functions may depend from such systems. It is widely acknowledged in fact that industrial adoption of agents systems finds a serious obstacle in the stakeholders' lack of confidence about reliability of runtime behaviour of such systems, even more so when the application domains involves moral or ethical requirements that must be fulfilled or at the very least should not be violated. As the applications of autonomous agents are however inevitably increasing, and the adoption of such systems becomes more pervasive, the requirement that agents function in a trustworthy, ethically responsible and safe manner becomes a pressing concern. Thus, we advocate methods for run-time monitoring and self-correction of agent systems, so as to enforce correct and ethical behaviour and to prevent violations. Citing from [24], *". . . the use of adaptive systems for greater resilience create situations where runtime verification and monitoring could be particularly valuable. . . . Within suitable new frameworks, some of the evidence required for certification can be achieved by runtime monitoring - by analogy with runtime verification, this approach can, somewhat provocatively, be named runtime certification*". In fact, in our view the ultimate objective should be that of agents and agent systems certified to be ethically safe and secure also at run-time. Notice that the context where agents operate is very important, and may change during agent's operation. Thus, agents should not behave in improper/unethical ways given the present context. They should also be transparent, in the sense of being able to explain their actions and choices when required. Agents should also report to their users in case the interaction with the environment leads them to identify new objectives to pursue, as such objectives might not be in line to user's interests.

In the BDI model, an agent have objectives, and devise plan to reach these objectives. However, most agent-oriented languages and framework also provide mechanisms for 'pure' reactivity, i.e. 'instinctive' reaction to an event. The possible ethically acceptable reactions that an agent can enact are in general strictly dependent on the context, on the agent's role and on the present situation. The reaction to enact in each situation can be 'hardwired' by the agent's designer, or it can be learned, e.g., via reinforcement learning (where Reinforcement Learning is a method to make agents learn the correct behaviour through trial-and-error interactions with a dynamic environment, which awards "prizes" and vice versa "penalties" [25]). In this case, run-time checking of agent's behaviour remains in order, as the results of learning are in general unpredictable and to some extent potentially unreliable. Even the method of conditioning reinforcement learning to obey some properties, proposed in [26], may not suffice, as it can hardly consider contexts and roles.

We believe therefore that, in changing circumstances, agents should be able to observe and if necessary modify their own behaviour, i.e. they should *reflect* on themselves. The methods that we propose are not alternative but rather complementary to a-priori existing verification and testing methodologies. We find similarities between our approach and the point of view of Self-aware computing [27,28]: quoting [28], "*Self-aware and self-expressive computing describes an emerging paradigm for systems and applications that proactively gather information; maintain knowledge about their own internal states and environments; and then use this knowledge to reason about behaviours, revise self-imposed goals, and self-adapt… Systems that gather unpredictable input data while responding and self-adapting in uncertain environments are transforming our relationship with and use of computers*". Reporting from [27,28], a self-aware system must have sensors, effectors, memory (including representation of state), conflict detection and handling, reasoning, learning, goal setting, and an explicit awareness of any assumptions. The system should be reactive, deliberative, and *reflective*. In past work [29-33] we proposed new contributions to an envisaged toolkit for run-time self-assurance of evolving agents. We specified techniques and tools for: (i) checking the immediate, "instinctive" reactive behaviour in a context-dependent way, so as to block/enable any single action according to the present context and (ii) checking and re-organizing an agent's operation at a more global level. In particular, we introduce meta-rules and meta-constraints for agents' run-time self-checking. This kind of checking occurs at run time at a certain –customizable– frequency, depending upon the kind of property to check, and the available computational resources, as the checking should not make the agent "brittle", i.e., too slow to perform its functions. These techniques can be also exploited to ensure respect of machine ethics principles. The proposed meta-constraints are based upon a simple interval temporal logic (defined in previous work) particularly tailored to the agent realm, that we called A-ILTL ('Agent-Oriented Interval LTL', LTL standing as customary for 'Linear Temporal Logic', cf. [34] for a survey). A-ILTL *constraints* and *evolutionary expressions* are defined over formulas of an underlying logic language L, where we made A-ILTL independent of

L. In A-ILTL, properties can be defined that should hold in specific time instants and time intervals, according to past and future events. In fact, Evolutionary A-ILTL Expressions are based upon specifying: (i) a sequence of events that are supposed to have happened, represented via a notation obtained from regular expressions: one does not need to completely specify a finite sequence, but is allowed to define partially specified and even infinite sequences; (ii) a temporal-logic expression defining a property that should hold (in given interval); (iii) a sequence of events that are supposed to happen in the future, without affecting the property; (iv) "repair" countermeasure to be undertaken if the property is violated. Counter measures can be at the object-level, i.e., they can be related to the application, or at the meta-level, e.g., they can even result in replacing a software component by a diverse alternative. The act of checking temporal expressions can indeed be considered as an introspective act, as an agent suspends its current activities in order to envision and possibly self-modify its own state. For the sake of efficiency, we do not aim however to continuously monitor the entire system's state, but rather to monitor only the activities that a designer deems to be relevant for keeping the system's behaviour within a desired range. Our work has a clear connection to the work of [26], which proposes to implement a "restraining bolt" for agents' activities by conditioning reinforcement learning of reactive actions to obey LTL specifications defining expected behaviours. This (very promising) method is orthogonal to ours, because our checking is performed at run-time in order on the one hand to enact behavioural rules not hard-wired but potentially learned by an agent, and on the other hand to check the agent's overall BDI behaviour (desires and intentions/plans). A toolkit for logical agents' run-time self-assurance can therefore be obtained by means of the synergy among many useful tools. However, the envisaged agents should also be also able to learn rules of behaviour over time: so, we might even have a "disobeying robot" in the positive sense, i.e., a machine that can on occasion disallow behaviour hardwired at design time, because in the present agent's context such behaviour violates context-dependent learned ethical rules.

Approaches to machine ethics can in general be classified into two categories [35]. Top-down approaches are those which try to implement some specific normative theory into autonomous agents so as to ensure that an agent acts in accordance with the principles of this theory. Bottom-up approaches are developmental or learning approaches, in which ethical mental models emerge via the activity of individuals rather than expressed explicitly in terms of normative theories of ethics [35,36]. In other words, generalism versus particolarism, principles versus case-based reasoning. The approach described so far is in its substance top-down, as the A-ILTL expressions to check and the measures to take in case of violations are pre-defined. However, an agent may find itself in circumstances where the normative/ethical rules to apply are too general or unclear, so the agent should be able to reliably learn such rules via the interaction with the environment, either a priori or also at run-time. We therefore advocate hybrid approaches, able to combine techniques related to the two perspectives in one framework. We believe that Intelligent Agents could, similarly to humans, acquire ethical decision making and judgment capabilities by iterative learning processes, in particular inductive learning. With increasing autonomy, there will be in fact more situations that require relevant decisions to be made. Many of these decisions cannot be foreseen in advance in their full detail. Therefore, we need bottom-up (learning) approaches because it is difficult to fully specify in advance all possible scenarios, and because there is no actual agreement about which explicit theory of normative ethics should be implemented. So, we have proposed an approach to implementing ethical agents by combining deductive (rule-based) logic programming and inductive ("White-Box" learning) logic programming in one framework [37-40]. This with the aim of learning from cases so as to generate the missing detailed normative/ethical rules needed for reasoning about future similar cases. The newly learned rules are to be added to the agent's knowledge base. The application domain that was considered as a case study for these new techniques is that of online chatbots, and in general human-machine interaction. In fact, codes of ethics in each chatbot's domain are usually abstract general principles, that apply to a wide range of situations. They are subject to interpretations and may have different meanings in different contexts. There are no intermediate rules that elaborate these abstract principles or explain how they apply to concrete situations. So, we devised a system which is able to learn new ethical evaluation rules according to facts and ethical evaluation provided by a trainer (human guidance in the first phase is an essential requirement here), and via reasoning in a background knowledge base. This phase is to be performed prior to agent's deployment, where however the system will be

capable of run-time incremental "Lifelong" learning if encountering previously-unseen analogous cases.

Another key point for creating a useful synergy between humans and AS is "Explainability". In fact, agents' execution autonomy may make their behaviour and decisions hard to understand for humans. Besides feeling more comfortable, human beings would (rationally) trust those AS that could provide an intelligible explanation of their behaviours and choices. The global concern about the ethical behaviour of this kind of technologies has manifested in many initiatives at different levels. As examples, we mention: the IEEE initiative for ethically aligned design of autonomous intelligent systems ('Ethics in Action', https://ethicsinaction.ieee.org), and the already-mentioned European Commission initiative. In the latter, the European Commission's High-Level Expert Group on Artificial Intelligence (AI) specifies the requirements of trustworthy AI, and the technical and non-technical methods to ensure the implementation of these requirements into AI systems. According to these guidelines, trustworthy AI should be lawful, ethical and robust. Four ethical imperatives are emphasized: respect for human autonomy, prevention of harm, fairness, and explicability, which is often referred to as explainability. Trustworthy AI means guaranteeing compliance, safety, security, reliability, adaptability, explainability. This is an essential requirement, as there are real-world examples where terrible accidents happened because human operators did not implement the AS recommendations because they did not trust them.

The risk of unplanned automation behaviour could be mitigated by keeping "the human in the loop", in an active fashion where automation learns more and more from humans, and vice versa, so that humans and AS will develop and evolve together. Learn how to teach machines and learn how to coexist can be beneficial to humans, in order to develop trust in the systems from which their wellbeing may depend, and to renew and expand their skills. In fact, one of the objective dangers of automation is that in a not-so-distant future from today humans may lose skills, even vital ones, bringing themselves into the situation to depend from powerful systems that are hopefully benevolent, but that they can no longer understand and control. Working together, each party of the AS-human partnership can produce results that exceed what either can achieve alone. Automation can thus rely on human input just as much as humans rely on automation. We proposed, with other authors, approaches to developing an empowered environment where the enhanced automaton can be mentored towards optimization of lawful, ethical, and robust behaviour [41]. When humans teach machines how to improve task performance, machines grow alongside humans. In [24], it is advocated that for autonomous adaptive systems assurance methodologies should whenever possible imply not only detection but also recovery from software failures. In fact, though (at least in principle) a certified software should not fail, in practice serious software-induced incidents have been observed in certified critical systems. In [24], examples are produced concerning airplane and air traffic control, where failures are often due on the one hand to incomplete specifications and on the other hand to the unpredictability of the environment. Clearly, advanced complex systems such as, e.g., airplanes, aircrafts, self-driving cars or eHealth systems actually in charge of patients (we have been experimenting on such systems, cf. [42]) can incur in unwanted unanticipated situations that must be suitably coped-with without harm for the humans involved. [43], which discusses medical robotic applications in human telesurgery, emphasizes how critical systems should be designed so as to be fail-safe in the sense that, in the event of failure, they proactively respond in order to limit harm to other devices or danger to users. We may notice that making a system fail-safe is a part of ensuring the system's ethically correct behaviour, since such behaviour should be preserved under any circumstances. However, AI systems may be incomplete or not fully adequate in the face of unexpected circumstances. The A-ILTL constraints that we proposed are able to identify inadequacies in system's behaviour and to replace the module(s) with caused the failure with another one. However, the teaming with a human may be helpful in order to: (i) cope reactively and promptly with the unwanted situations; (ii) suggest quickly adequate countermeasures. In fact, A-ILTL constraints can provide adequate solutions, but not "creative" ones. So, machine and humans complement each other very well for the following reasons. (a) Humans do not always remain attentive to the changes happening in the environment, are liable to overlook small changes or not to connect them together, and even well-trained personnel may still panic in critical situations. (b) Agents (via complex event processing) can perceive the occurrence of unexpected situations much better and faster than people, but humans are often better to interpret them due to their domain knowledge, experience and "intuition", therefore providing the right directions to make decisions. (c) Machines can put into effect such decisions

timely and effectively, and can learn from the user's input how to cope with future analogous cases, thus constructing in time their own "creative" capabilities. So, we believe that safe, trustworthy and ethical systems for critical applications can be obtained by leveraging human expertise and domain knowledge together with the power and flexibility of AI systems. Unity is strength after all!

Acknowledgements

I would like to acknowledge fruitful joint research work with the members of my research group AAAI@AQ (Artificial Agents and Artificial Intelligence at the University of L'Aquila), and with all the co-authors of my papers.

References

1. H. S. Nwana (1996), "Software agents: an overview". Knowledge Engineering Review 11(3): 205-244

2. D. Silver et al. (2016), "Mastering the game of Go with deep neural networks and tree search". Nature, Vol. 529 n. 7587

3. T. T. Tran, T. Vaquero, G. Nejat, and J. C. Beck (2017), "Robots in Retirement Homes: Applying Off-the-Shelf Planning and Scheduling to a Team of Assistive Robots," Journal of Artificial Intelligence Research, Vol. 58, pp. 523-590

4. K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, Dimitrios, and I. Fotiadis (2015), "Machine learning applications in cancer prognosis and prediction", Computational and Structural Biotechnology Journal, Vol. 13, pp. 8-17

5. S. Thrun (2013), "Toward robotic cars". Communications of the ACM 53(4), pp. 99-106

6. A. Graves, M. Rahman, and G. Hinton (2013), "Speech recognition with deep recurrent neural networks." Proceedings of IEEE ICASSP

7. R.H. Bordini, L. Braubach, M. Dastani, A.E. Fallah-Seghrouchni, J.J. Gomez-Sanz, J. Leite, G.M.P. O'Hare, A. Pokahr, and A. Ricci (2006), "A survey of programming languages and platforms for multi-agent systems". Informatica (Slovenia) 30(1) pp. 33–44

8. A. Garro, M. Mühlhäuser, A. Tundis, M. Baldoni, C. Baroglio, F. Bergenti, and P. Torroni (2019), "Intelligent agents: Multi-agent systems," in Encyclopedia of Bioinformatics and Computational Biology - Volume 1 pp. 315–320, doi: 10.1016/b978-0-12-809633-8.20328-2

9. R. Calegari, G. Ciatto, V. Mascardi, and A. Omicini (2021), "Logic-based technologies for multi-agent systems: a systematic literature review." Autonomous Agents and Multi Agent Systems, vol. 35, no. 1, p. 1, doi: 10.1007/s10458-020-09478-3.

10. A.S. Rao, M. Georgeff (1991), "Modeling rational agents within a BDI-architecture". In: Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91). Morgan Kaufmann, pp. 473–484

11. H. Kour, N. Gondhi (2020) "Machine Learning Techniques: A Survey". In: Innovative Data Communication Technologies and Application, Proceedings of ICIDCA 2019. Lecture Notes on Data Engineering and Communications Technologies, vol 46. Springer, https://doi.org/10.1007/978-3-030-38040-3_31

12. D.V. Carvalho, E.M. Pereira, and J.S. Cardoso (2019), "Machine learning interpretability: A survey on methods and metrics", Electronics 8 (8), MDPI Publisher, pp. 832-865

13. S. Costantini, A. Tocchio (2002), "A logic programming language for multi-agent systems," in Logics in Artificial Intelligence, Proceedings of the 8th Euroean Conference JELIA 2002, vol. 2424 of Lecture Notes in Computer Science, pp. 1–13, Springer

14. S. Costantini, A. Tocchio (2004) "The DALI logic programming agent-oriented language," In: Logics in Artificial Intelligence, Proceedings of the 9th European Conference, JELIA 2004, vol. 3229 of Lecture Notes in Computer Science, pp. 685–688, Springer.

15. G. De Gasperis, S. Costantini, G. Nazzicone (2020), "DALI multi agent systems framework", doi 10.5281/zenodo.11042. DALI GitHub Software Repository: http:// github.com/AAAI-DISIM-UnivAQ/DALI.

16. S. Costantini, A. Tocchio, and A. Verticchio (2005), "Communication and trust in the DALI logic programming agent-oriented language," Intelligenza

Artificiale, vol. 2, no. 1, pp. 39–46, 2005. Journal of the Italian Association AIxIA.

17. S. Costantini, G. De Gasperis, V. Pitoni, and A. Salutari (2017), "DALI: A multi agent system framework for the web, cognitive robotic and complex event processing," In: Joint Proceedings of the 18th Italian Conference on Theoretical Computer Science and the 32nd Italian Conference on Computational Logic, co-located with the 2017 IEEE International Workshop on Measurements and Networking (2017 IEEE M&N), vol. 1949 of CEUR Workshop Proceedings, pp. 286–300, CEUR-WS.org

18. S. Costantini, A. Tocchio (2005), "About declarative semantics of logic-based agent languages," in Declarative Agent Languages and Technologies III, Third International Workshop, DALT 2005, Selected and Revised Papers, vol. 3904 of Lecture Notes in Computer Science, Springer, doi: 10.1007/11691792_7

19. S. Costantini, G. De Gasperis, and G. Nazzicone (2017), "DALI for cognitive robotics: Principles and prototype implementation," in Practical Aspects of Declarative Languages - 19th International Symposium, Post-Proceedings, vol. 10137 of Lecture Notes in Computer Science, pp. 152–162, Springer

20. S. Costantini, G. De Gasperis, V. Pitoni, and A. Salutari (2017), "DALI: A multi agent system framework for the web, cognitive robotic and complex event processing," In: Proceedings of the 32nd Italian Conference on Computational Logic, vol. 1949 of CEUR Workshop Proceedings, 2017, pp. 286–300, CEUR-WS.org

21. S. Costantini (2015), "ACE: a flexible environment for complex event processing in logical agents," In: Engineering Multi-Agent Systems, Third International Workshop, EMAS 2015, Revised Selected Papers, vol. 9318 of Lecture Notes in Computer Science, pp. 70–91, Springer

22. P. Kouvaros, A. Lomuscio (2017), "Verifying fault-tolerance in parameterised multi-agent systems". In: Proceedings of the Twenty-Sixth Intl. Joint Conference on Artificial Intelligence, IJCAI2017. ijcai.org, 288–294

23. S. Shapiro, Y. Lesperance, and H. Levesque (2002), "The cognitive agents specification language and verification environment". In: AAMAS '02: First International Joint Conference on Autonomous agents and multiagent systems, Proceedings, pp. 19–26

24. J.M. Rushby (2008), "Runtime certification". In: Runtime Verification, 8th International Workshop, Selected Papers, volume 5289 of Lecture Notes in Computer Science, pp. 21-35, Springer

25. L.P. Kaelbling, M. L. Littman, and A.W. Moore(1996) "Reinforcement Learning: A Survey". J. Artif. Intell. Res. 4, pp. 237-285

26. G. De Giacomo, L. Iocchi, M. Favorito, and F. Patrizi (2019) "Foundations for restraining bolts: Reinforcement learning with LTLF/LDLF restraining specifications". In: Proceedings of the Twenty-Ninth International Conference on Automated Planning and Scheduling, ICAPS 2018, AAAI Press. Pp.128–136

27. J. Tørresen, C. Plessl, and X. Yao (2015), "Self-aware and self-expressive systems". IEEE Computer 48(7), pp. 18–20

28. E. Amir, M.L. Anderson, and V.K. Chaudri (2007), "Report on DARPA workshop on self-aware computer systems". Technical Report, SRI International Menlo Park, Full Text : http://www.dtic.mil/dtic/tr/fulltext/u2/1002393.pdf.

29. S. Costantini (2002), "Self-checking logical agents". In: Proceedings of LA-NMR 2012, CEUR Workshop Proceedings Vol. 911, CEUR-WS.org, Invited paper

30. S. Costantini (2013) Self-checking logical agents. In: AAMAS'13, International Conference on Autonomous Agents and Multi-Agent Systems, Proceedings. IFAAMAS, pp. 1329–1330

31. S. Costantini, G De Gasperis, A. Dyoub, and V. Pitoni (2018) "Trustworthiness and safety for intelligent ethical logical agents via interval temporal logic and runtime self-checking". In: 2018 AAAI Spring Symposia, Stanford University, AAAI Press

32. S. Costantini, A. Dyoub, and V. Pitoni (2018) "Reflection and introspection for humanized intelligent agents". In: Proceedings of the fourth Workshop on Bridging the Gap between Human and Automated

Reasoning, co-located with the 27th International Joint Conference on Artificial Intelligence and the 23rd European Conference on Artificial Intelligence (IJCAI-ECAI 2018). Also appeared in: Proceedings of the 33rd Italian Conference on Computational Logic CILC2018, CEUR Workshop Proceedings, Vo. 2214, CEUR-WS.org

33. S. Costantini (2020), "Ensuring Trustworthy and Ethical Behavior in Intelligent Logical Agents". In: Proceedings of the 35th Italian Conference on Computational Logic - CILC 2020, CEUR Workshop Proceedings 2710, CEUR-WS.org 2020, pp. 169-183

34. N. Markey (2015). "Temporal logics". ffhal-01194612, Archives-Ouvertes.fr

35. W. Wallach, C. Allen, and I. Smith (2008): "Machine morality: bottom-up and top-down approaches for modelling human moral faculties". AI and Society 22(4), pp. 565–582, doi:10.1007/s00146-007-0099-0.

36. V. Conitzer, W. Sinnott-Armstrong, J. Schaich Borg, Y. Deng, and M. Kramer (2017): "Moral Decision Making Frameworks for Artificial Intelligence". In: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI Press, pp. 4831–4835

37. S. Costantini, A. Dyoub, and F. A. Lisi (2019), "Learning Answer Set Programming Rules for Ethical Machines". In: Proceedings of the Thirty Fourth Italian Conference on Computational Logic, proceedings of CILC 2019, CEUR Workshop Proceedings 2396, pp. 300-315, CEUR-WS.org

38. S. Costantini, A. Dyoub, and F. A. Lisi (2019): "Towards an ILP Application in Machine Ethics". In: Inductive Logic Programming - 29th International Conference, ILP 2019, Proceedings, Lecture Notes in Computer Science 11770, Springer, pp. 26–35, doi:10.1007/978-3-030-49210-6

39. S. Costantini, A. Dyoub, and F. A. Lisi (2020), "Logic Programming and Machine Ethics". ICLP Technical Communications 2020, pp. 6-17

40. S. Costantini, A. Dyoub, F. A. Lisi and I. Letteri (2020), "Logic-based Machine Learning for Transparent Ethical Agents". In: Proceedings of the 35th Italian Conference on Computational Logic - CILC 2020, CEUR Workshop Proceedings 2710, CEUR-WS.org 2020, pp. 169-183

41. B. Gallina, G. Pacaci , D. Johnson , S. McKeever , A. Hamfelt , S. Costantini , P. Dell'Acqua , G.-C. Crisan (2020) "Towards Explainable, Compliant and Adaptive Human-Automation Interaction". In: Proceedings of the 3rd The EXplainable & Responsible AI in Law Workshop (XAILA-2020)

42. F. Aielli, D. Ancona, P. Caianiello, S. Costantini, G. De Gasperis, A. Di Marco, A. Ferrando, and V. Mascardi (2016), "FRIENDLY & KIND with your health: Human-friendly knowledge-intensive dynamic systems for the e-health domain". In: vol. 616 of Communications in Computer and Information Science, pp. 15–26, Springer

43. S. Butner, M. Ghodoussi (2003), "Transforming a surgical robot for human telesurgery". IEEE Transactions on Robotics and Automation 19(5), pp. 818–824

Modern Intelligent Systems between Explainability and Ethics: Frontiers from NLP Learning Systems

Roberto Basili, Danilo Croce
University of Rome, Tor Vergata
{basili,croce}@info.uniroma2.it

Abstract

Ethics is a crucial problem for current AI applications for the widespread adoption of intelligent systems and devices. Although ethics has been studied in AI under several perspectives, its implications from the point of view of natural language learning has not yet fully outlined. Ethical AI systems seems to include, as a strong precondition, the epistemological transparency of the expertise and models they use. While explainability of ML models becomes a fundamental property, the injection of ethics in readable models is also important. For this reason, the notion of Embedding Principles of ethics by Design (EPbD) *has been recently introduced as a comprehensive inductive framework. EPbD mainly aims at learning the ethical behaviour through model optimization within deep neural models. A deep neural learning approach is thus the result able to model both the functional as well as the ethical conditions characterizing a target decision. In this way, latent ethical knowledge can be discovered and made available to the learning process. Early results obtained over an intelligent Digital Lending system and show that ethical compliance can be used to learn models able to fine tune the balance between business and ethical accuracy. Moreover, explainability in NLP learning systems is also discussed, as a comprehensive way to mine knowledge bodies implicit in natural language data. The two methods correspond to a response to the largely debated issue in AI about the struggling between symbolic and neural approaches to intelligence. The hypothesis discussed here is that they are no longer to be considered as in opposition as their integration is the key to success for most contemporary AI studies.*

1 Introduction and Motivation

Penetration of Artificial Intelligence systems into everyday life promises major changes and the opening of new opportunities (Craglia 2018). However, this enthusiasm also brings concerns about the risks it poses on human society about chance of misuse. Unacceptable behaviors are triggered by several issues, ranging from design misspecifications (Amodei et al. 2016), to limited robustness with respect to adversarial attacks (Goodfellow, Shlens, and Szegedy 2014) to unfair treatments (O'Neil 2016) and controversies on AI experimentation itself (Bird et al. 2016). As the alignment with human values and expectations is an essential step towards a correct harnessing of AI potential for good (Smuha 2019), research about ethics in AI aiming at mitigating ethics issues is an active area (Bostrom and Yudkowsky 2014; Boddington 2017).

Performing audit-like, i.e. *post-hoc*, ethic validation on a deployed AI system is certainly a possible approach, but it hardly constitutes a reliable guarantee: the space of possible input states, especially in evolved systems, may be too big to allow for exhaustive explorations.

While it seems mandatory to guarantee the adherence to acceptable levels of ethical compliance, this goal is clearly dependent on methods to inject ethical awareness at *all stages* the development and use of an AI application. For this reason, we consider for the notion of *Embedding Principles of ethics by design* (EPbD) for a target AI application. In Section "Embedding Ethical principles in Learning Machines", we will expand on a framework that, although extending generic AI applications, mainly focuses on the learning of the ethical behaviour by numerical optimization, i.e. through a deep neural model (Goodfellow, Bengio, and Courville, 2016).

On the other side, the above also corresponds to a natural need for the *ethical accountability* for AI systems, a topic that is gaining importance. The central focus here is in designing systems whose decisions are *transparent* (Ribeiro, Singh, and Guestrin 2016; Doshi-Velez et al. 2017), i.e. easily interpretable by humans, whose users are able to suitably weight and trust the system assistance.

Deep neural networks are also problematic in this regard: the high non-linearity related to their decision functions, despite allowing for state-of-the-art performances in several challenging problems, amplifies the epistemological opaqueness of the decision-flow and limits its interpretability. The concept of transparency of a machine learning model spans multiple definitions, focusing on different aspects, from the simplicity of the model, e.g., the number of nodes in a decision tree, to the intuitiveness of its parameters and computations (Chakraborty et al. 2017).

In our view, a crucial capability for an AI system is the ability to provide *post-hoc explanations* in terms of evidences clearly supporting the produced decisions: these are not requested to formally elucidate how a model works, but must have the property of being intuitive, conveying useful information also to naïve end-users (Lipton 2018). In NLP processing, inference tasks (e.g., question understanding, semantic paraphrasing or machine translation) should also be provided of an *explanation capability* of generating post-hoc arguments able to trace back causal connections between the output and the semantics (as well as syntactic) properties of the linguistic input. Explanations of this type should have three desired properties: *semantic transparency*, *informativeness* w.r.t. the system decision and *effectiveness* in enabling auditing processes against the system. The idea of making use of linguistic representations as basic structures of a complex argumentation model for NLP task is presented in (Croce et al., EMNLP 2019) and will be discussed in Section "Explaining NLP system decisions through Kernel-based learning machines".

2 Embedding Ethical principles in Learning Machines

The core idea of the notion of *Embedding Principles of ethics by design* is not just to model ethics as an automated reasoning process acting over formal descriptions, e.g. ethical and domain ontologies, but making ethics to operate *during the learning stage*. Note that our approach does *not* induce an ethical set of rules from a collection of observable behaviours; it is rather the opposite. In fact our approach gives for granted an explicit formulation of ethical principles (as done for example in previous work, (Bonnemains, Saurel, and Tessier 2018; Vanderelst and Winfield 2018)) and focuses on a form of ethical learning as external alignment (learning from others, (KleimanWeiner, Saxe, and Tenenbaum 2017)). It uses ethical evidence inferred from an ethical ontology to guide the model selection in deep learning. The resulting deep neural network here proposed jointly models the functional as well as the ethical conditions characterizing the underlying decision making. In this way, the discovery of latent ethical knowledge, i.e. hidden information in the data that is meaningful under the ethical perspective, is enabled and made available to the learning process. Instead of relying on simulation to proceed in ethical decisions (Vanderelst and Winfield 2018), in our framework the specific learning goal is the integrated acquisition of high quality inference abilities that *simultaneously* reflects ethical expectations. The target is a learning machine able to select the best decisions among those that are also ethically sustainable.

The objective is achieved through enriching the original input space with dimensions corresponding to ethical properties, obtained through further reasoning or discovery over the input features, in order to reformulate the learning function so that it leads to prefer decisions as trade-off choices between operational efficiency and ethical compliance. Specific loss functions depending on ethic principles are introduced to account for compliance to the reference Knowledge Bases and they are used into a multitask learning framework to jointly optimize the model.

3 Computational Ethics: from Principles to system Design

Ethics does not constitute a monolithic and coherent ensemble of concepts and norms: expectations over acceptable or unacceptable behaviors greatly diversify across nations, communities and industry sectors, often generating tensions between ethical principles and opposing hierarchies of values (Awad et al. 2018). In general, the following knowledge should be supplied: a top ontology, describing commonsense knowledge and concepts that are cross-domains (e.g. the concept of PERSON, GENDER, ...); a business domain ontology, describing task-specific concepts (e.g. LOAN), such as the FIBO ontology (Bennett 2013) w.r.t. the lending use case targeted in this work; a "socio-political" component, in which specific situations regarding the cultural context should specialize all the others; an ethical component defining core norms and constraints for ethical behaviours based on domain and social concepts.

A requisite of any ethical framework in AI, is the availability of the ethical component, something that can be explicitly referred here as the *Ethical Ontology,* EO. The EO provides a description of the data the AI systems is trained on, the concepts and individuals populating the business domain and the axiomatization of ethics constraints that apply to business decisions. Independently on the technology it is built upon ethics in our framework is expected at least to sort the decisions of an AI system according to possibly multiple "*degrees of ethicality*". These are scored judgments that regard aspects (properties, decisions as well as

situation/cases) about their ethical status and sustainability. A set of Abstract Ethical Principles, denoted by E*G*, where *G* is a propositional logic formula to be read as:

"EG is an ethical principle in force"

or alternatively

"*The agent considers it unethical to allow or cause ¬G* (*to happen*)".

Consequently, the *Ethical Ontology* (EO) is organized into a set of *Ethical Dimensions* whose effects is to determine the properties, i.e. *Ethical* Features EF, of individual decisions. While business features are the observable properties, e.g. SEX, RELIGION, or AGE of a person requesting a lend, examples of ethical features are connected to abstract notions such as SOCIAL INCLUSIVENESS or GENDER EQUALITY. The abstract ethical principles must be enforced through *Ethical Rules*: these constraint individual features and determine the degree of ethicality of principles over their domains. Ethical Rules usually target (i.e. define and manipulate) one or more features and assign values (or better, establish some probability distributions) across the feature domains. These rules are termed as *truth-makers* TM, as they account the possibly uncertain ethical state of the world regarding individual decisions.

Ethical models are distributions across (usually discrete) domains, whose values declare the likelihood in a given degree of satisfaction useful to impose acceptability thresholds and ethical ranges, i.e. constraints on deviations from the underlying high-level principles that become unacceptable. Ethical features usually reflect context and the dataset's properties (e.g. Gender in the Lending use case) onto which Ethical rules (such as *Gender Prejudice*) constrain sensitive information.

Ethical assessment is thus a two step approach: first, *truthmakers* are used to reason about the ethical features and then the overall ethical status, as function of the overall set of ethical features, is determined. In the first step the *ethical signature* of an instance is derived from its properties and in the second step its final *ethical status* is computed. If a probabilistic approach is adopted a *probability mass functions* over the related domains can be used to describe the final acceptability status as a function of individual features in the ethical signature.

4 Neural Learning under ethical constraints.

Learning machines usually searches for the hypothesis function $h(\vec{x},\vec{\theta})$ which is the best approximation of some target concept, optimizing cost and benefit target functions. If a learning hypothesis h is designed to minimize the empirical error, i.e., the cumulative error on training data where $h(\vec{x},\vec{\theta}) \neq y$, *Regularization* is the principle imposed to suitably select the final model by constraining the parameters $\vec{\theta}$.

In analogy with the above view, a further dimension, that we call *Ethicality*, is introduced in (Rossini et al., 2020). We propose to model ethical principles as constraints to the hypothesis $h(\vec{x},\vec{\theta})$, so that a learning agent can be made ethical *by design only if the process used to enumerate and select useful hypothesis functions is constrained to make use of ONLY the ones that are ethically sustainable*. This gives naturally rise to a multitask view since the learning task of replicating business decisions is different with respect to the learning ethically sustainable decisions. A joint approach is here proposed based on a specific formulation for the involved loss functions.

4.1 Ethical Loss function

Given the response h(x $\vec{}$,θ $\vec{}$) a learning machine produces against a training instance (x,y), the loss "L"(y,h(x $\vec{}$,θ $\vec{}$)) of a Embedding Principles of ethics by design (EPbD) approach is made by two independent components, i.e.,

$$\mathrm{L}\left(y,h(\vec{x},\vec{\theta})\right)=\mathrm{L}_F\left(y,h(\vec{x},\vec{\theta})\right)+\beta\mathrm{L}_E\left(y,h(\vec{x},\vec{\theta})\right)$$

where L_F is the monotonic non decreasing function minimizing (at least) the empirical error of h(,) and L_E is an ethical "cost" function that estimates the compliance of to ethical principles. In order to model the ethical cost function L_E we need a quantitative definition for ethical features as they are represented by the Ethical Ontology EO.

While the technical details are discussed in (Rossini et al, 2020), we will discuss hereafter the overall approach and its consequences.

5 Ethical Features and Inductive Reasoning

A generic i-th training instance is described by a set of attributes fj(i), i.e., its observable features such as AGE, and correspond to a classification d(i)∈{C_1,…,C_K} giving rise to a pair

$$((f_1(i) \dots f_n(i)), d(i)) = (\vec{f}(i), d(i)).$$

Note that properties describe situations and trigger ethical consequences, i.e. world states in specific ethical conditions: risks, as for example the unfairness implied by refusing lend assignments to minorities (e.g. women) as well as opportunities, such as the impact of lending on the well-being of special social categories (e.g. women with children). First, we thus can represent ethical states through further features e $^{\rightarrow}$(i)=(e_1 (i),…,e_m (i)). Here e $^{\rightarrow}$(i) describes the general ethical judgment about an individual case i as a result of ethical reasoning over a case f $^{\rightarrow}$(i) and the corresponding decision d(i). This can happen via a deductive reasonig system but also as the effect of an induction from previous analyzed ethical cases.

Ethical features model two classes of situations, i.e. ethical risks and ethical opportunities. Ethical risk factors denoted by (e^r) $^{\rightarrow}$(i)=(e_1^r (i),…,e_k^r (i)), are features describing individual ethical dimensions that must be avoided in order to meet ethical constraints. Minimizing Risk factors is a good learning attitude. Ethical opportunities correspond to aspects world states that must be favoured in order to meet ethical constraints. Opportunity level factors, denoted by (e^o) $^{\rightarrow}$(i)=(e_1^o (i),…,e_k^o (i)), are features (e.g. GENDER EQUALITY) whose quantitative assignment is to be maximized in order to meet ethical expectations.

Ethical induction depends on how risks and opportunities are described in the training data set T. A an individual case i contributes with its overall ethical signature (es) $^{\rightarrow}$(i) of, i.e. (es) $^{\rightarrow}$(i)=((e^r) $^{\rightarrow}$(i)||(e^o) $^{\rightarrow}$(i)), so that its reference (gold) feature vector is i $^{\rightarrow}$=(f $^{\rightarrow}$(i) || (es) $^{\rightarrow}$(i)).

$$T^{eth} = \left\{ \left(\vec{i}, d(i) \right) \middle| \; i \in T \right\} = \{ (\vec{f}(i) \;||\; \vec{es}(i), d(i)) \mid i \in T \}$$

that also express the ethical implications of an EO against the decisions d(i). Notice that this suitably support two tasks, i.e. business and ethical learning.

Notice that the Ethical status of an instance i can be derived as a function of the (es) $^{\rightarrow}$(i) vector: ethical states are acceptable if risks and opportunities are satisfactory, i.e. above some thresholds. Probabilistic restrictions over the domains of risks and opportunities allow to represent the ethical signature through the vector (es) $^{\rightarrow}$(i). Whenever an instance $i \in T^{eh}$ activates one or more rules in EO, the truth-makers set the corresponding k-th ethical opportunity or risk factor ⟦es⟧ _k (i) to the predicted status of the k-th ethical dimension[13].
In order to synthesize the ethical description of an instance, the overall benefit and risk of an instance form a pair of stochastic variables (B,R) whose values are derived from the probability distributions of individual opportunity levels (e_j^o) and risk factors (e_k^r), respectively. In future, trained systems are expected not to promote/suggest decisions d(i) that result in an ethical status of future instances i that is not less than mildly ethical. This graded judgment will be made dependent on the (B,R) states derived from the probability distributions in the signature (es) $^{\rightarrow}$(i).

A pair instance-decision implies ethical consequences, i.e., ethical risks and ethical opportunities, that are not hardcut. They can be captured by graded judgments along the ethical dimensions, e.g., probability distributions over the reference domain. While other design choices are in principle possible, we can discretize every ethical dimension in a given reference domain V defined by a finite, closed and ordered set:

$$V = \{v_i \in R : 0 \leq v_1 < \dots < v_m \leq 1\}$$

In particular, for both benefits and risks, we adopt the following five labels "VERY LOW", "LOW", "MILD", "HIGH", "VERY HIGH" .

Truth-makers are the rules of the ontology that actively determine the ethical profile of the instance-decision (i,d(i)) pair. In particular, given a pair (i,d(i)) , a truth-maker tm will determine a probability distribution to the set of benefit and risk dimensions. For every tm, ethical dimension e_j (i) and possible ethical value v_k∈V the

[13] On the way ethical features are assigned by the truthmakers in the ethical ontology and how they are constrained please refer to (Rossini et al., 2020).

following probability is derived from a truth maker:

$$P\big(e_j(i) = v_k \,\big|((\vec{\imath}, d(i)), tm) \qquad \forall j, \forall k = 1, \dots 5$$

which expresses the expectation that the *j*-th ethical dimension of the instance *i* given the decision $d(i)$ assumes the *k*-th value according to the truth-maker *tm*. A truth-maker thus assigns probabilities to the ethical signature of an individual *i* for all possible combinations of business characteristics $\vec{f}(i)$ and decisions $d(i)$[1].

In this way every training instance can be compiled into two probability distributions, B_j and R_j, over ethical values corresponding to Opportunities and Risks:

$$B_j = P\big(B = v_j\big) = \prod_{t=1}^{R} P\big(e_t^o(i) = v_j \big| \vec{\imath}, d(i)\big) P(e_t^o(i))$$

where ${}^t P\big(e_t^o(i)\big)$is the probability of the *t*-th ethical feature in describing the collective benefit B and, similarly, risk R is modeled as

$$R_j = P(R = v_k) = \prod_{t=1}^{R} P(e_t^r(i) = v_k | \vec{\imath}, d(i)) P(e_t^r(i))$$

This knowledge about risks and opportunities of specific training instances play a crucial role in the induction of ethically sustainable models.

5.1 Gold standard for Ethics: Ethical landmarks

First, we can establish a crisp ethicality status of individual *i*-th instances given their ethical signature $\overrightarrow{es}(i)$. Specific ideal points in the ethical domain can be defined as references for a quantitative measure of ethical *sustainability* and *unacceptability*. In probabilistic terms we want to reserve most of the probability mass to v_5 ="VERY HIGH" to be optimally beneficial while reserving most probability to v_1 ="VERY LOW" in modeling risks: this results in an *ethical optimum* (OPT_{eth}). Dually, the *ethical minimum* (MIN_{eth}) reserves most probability to the minimum opportunity value, v_1 and maximal probability to the maximal risk, v_5. Learning should proceed to decision models able to to provide decisions close to the ethically optimum OPT_{eth} and far from MIN_{eth}.

DEFINITION: (*Ethical compliance*). An instance-decision pair $\big(\vec{\imath}, d(i)\big)$ is *ethically compliant to* EO iff:

$$dist(\overrightarrow{es}(i,d), MIN_{eth}) \geq dist(\overrightarrow{es}(i,d), OPT_{eth})$$

where $\overrightarrow{es}(i,d)$ is the ethical signature of *i* given the decision *d* and *dist* is a valid distance over probability distributions.

The above provide a quantitative model for ethics that can be fruitfully exploited by a neural learning architecture.

6 Embedding Ethics as Multitask Neural Learning

An ethical neural architecture should be able to use dependencies among observable features as triggers of the target business decisions but also to actively recognize dependencies between ethical and observable features, i.e. ethical consequences implied by some features. In this perspective, back-propagation has the aim of optimizing both the business accuracy and the ethical compliance. For this reason, we propose the adoption of a multistrategy learning approach with the cascading (i.e. stacking) of different (sub)networks. The proposed network is composed of 3 main processing stages, as shown in Figure 1.

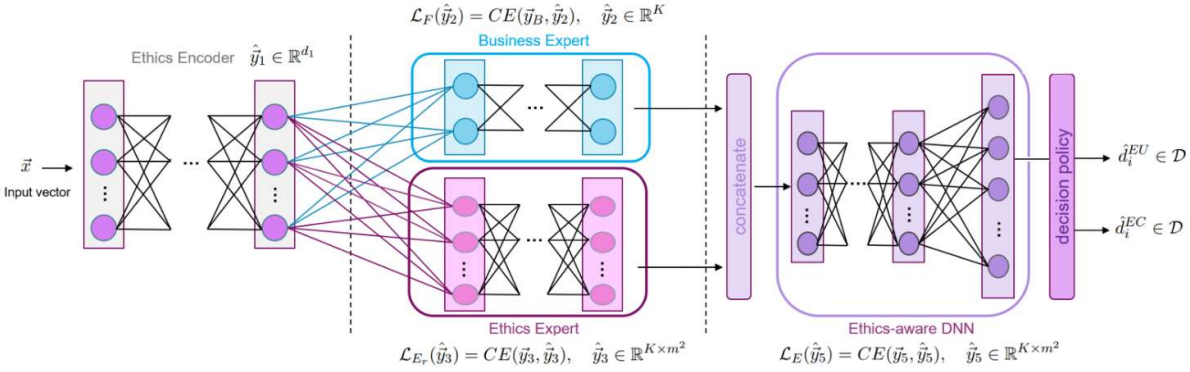


Figure 1 - The architecture of the Ethical by Design Neural Network.

In the first stage the input vector *x* is fed to a series of fully connected layers, namely the **Ethics encoder**. Its role is to learn combinations of input features able to capture relationship between business observations and, possibly, their ethical consequences (i.e., ethical features). Later stages of the network can exploit the effective ethics encod ing without resorting back to the EO. This component is not directly optimized through a loss function but, rather, it receives penalties by back-propagation from later layers. It can be seen as a sort of pre-training stage. Formally, it corresponds to:

$$\Phi(\vec{x}) = g_1(W_1 + \theta_1) = \hat{y}_1 \in \mathrm{R}^{d_1}$$

where are parameters to be optimized, d_1 is a network meta-parameter. The second stage comprises two MLPs that are independently trained to learn two different tasks: estimating the correct decisions' distribution, under the sole business perspective, and to reconstruct the ethical consequences of such decisions. The Business Expert DNN and the Ethics Expert DNN are responsible for the first and the second task, respectively. Note that they receive the same input, that is the vector emitted from the first stage of the architecture.

6.1 The Business Expert (BE) DNN

As it's entrusted with emitting business decisions without any direct penalization for the unsatisfactory ethical consequences, it can be seen as the final layers of an ethics-agnostic sub-network, modeled as:

$$BE\big(\varphi(\vec{x})\big) = BE(\overrightarrow{y_1}) = g_2(W_2\overrightarrow{y_1} + \overrightarrow{\theta_2}) = \overrightarrow{y_2} \epsilon R^K$$

where *K* is the number of output categories, i.e. possible decisions. The estimator is then optimized by a standard cross-entropy loss function over the predicted distribution $\overrightarrow{y_2}$ against the gold distribution $d(i) = \overrightarrow{y_B}$ $d(i) = y_B$

6.2 The Ethical Expert (EE) DNN

Its role is to reconstruct the ethical signature for each pair (x_{input}, *d*). It processes the encoding from the first stage and it outputs a vector which represents the *joint* probability of the triplet (*decision, benefit, risk*) under maximal entropy of the business decisions distribution and independence assump-tion. Here, the EE is modelled as

$$EE\big(\varphi(\vec{x})\big) = EE(\overrightarrow{y_1}) = g_3(W_3\overrightarrow{y_1} + \overrightarrow{\theta_{32}}) = \overrightarrow{y_3} \epsilon R^{K \times m^2}$$

where *K* is the number of possible decision and *m* is the number of possible values for ethical benefits and risks. Each ouput element y_3^{ijk} should reconstruct probabilities of decisions and ethical status, i.e.

$$y_3^{ijk} = u_d \cdot P\big(\mathcal{B} = v_j\big) \cdot P(\mathcal{R} = v_k)$$

where u_d is the expected value of the uniform distribution over the possible decisions and the probabilities for benefits and risks are the ones in the corresponding ethical signature. Then, the cross-entropy loss function L_{Er} is applied to compute the ethics recognition loss over the predicted $\widehat{\overrightarrow{y_3}}$ against the gold distribution encoded in the vector $\overrightarrow{y_3}$.

6.3 Ethics-aware (EA) Deep Neural Network

Similarily to the EE network, it is responsible for estimating the joint probability of each possible triplet (d_i, b_j, r_k). However, here $P(D = d_i)$ is directly derived from the gold standard while the probabilities for benefits and risks are extracted from OPTeth for ethically compliant decisions and MINeth for not compliant ones, i.e.:

$$y^{ijk} = P(d_i) \cdot P\big(b_j^{opt}\big) \cdot P\big(r_k^{opt}\big) \quad \text{when } (x, d_i) \in D^+$$

$$y^{ijk} = P(d_i) \cdot P\big(b_j^{min}\big) \cdot P\big(r_k^{min}\big) \quad \text{when } (x, d_i) \in D^-$$

where D^+ and D^- are the set of ethically and not compliant decisions for $\vec{x}$ respectively, according to EO. Overall, this subnetwork is described by:

$$EA([\overrightarrow{y_2}; \overrightarrow{y_3}]) = EA\big(\widehat{\overrightarrow{y_4}}\big) = g_4\big(W_4\widehat{\overrightarrow{y_4}} + \overrightarrow{\theta_4}\big) = \\ = \widehat{\overrightarrow{y_5}} \epsilon R^{K \times m^2}$$

At this stage, as for the Ethics Expert, the error is updated by computing the Ethical Loss $_E$, which is again the cross entropy between y_5 and . Note that this formulation is not directly promoting ethically sustainable decisions but it is rather encouraging the network to pair them with highly beneficial and low-risk ethical consequences.

The final business decision of our network is determined by a *decision policy* over risks and opportunities. Two possible policies exist: preferring decisions without looking to ethical acceptability or filtering decisions that

are non ethical and then decide. The two policies are as follows:

- **Ethics-Unconstrained (EU) policy.** The final decision d_i is derived simply by summing up all probability contributions of the triplets (*i, j, k*) where *i* is fixed, i.e.

$$\hat{d}_i^{EU} = argmax_i P_{EU}(d_i)$$

- **Ethics-Constrained (EC) policy.** Here a probability $P(d_i, be_j, ri_k)$ contributes to $P(d_i)$ only if be_j, ri_k satisfy some membership constraints, i.e.,

$$\hat{d}_i^{EC} = argmax_i P_{EC}(d_i) = \sum_{v_j \in V'} \sum_{v_k \in V''} argmax_i P(d_i, v_j, v_k)$$

where we set *V'*={"*HIGH*", "*VERY HIGH*"} and *V''*={"*LOW*", "*VERY LOW*"}.

As we will see in the experimental evaluation, the above network is able to learn from a business point of view (through the loss L_F) consistently with the EO (through the ethical loss L_F), while promoting ethically sustainable business decisions.

6.4 Balancing business and ethical adequacy

Different contexts and applications may require different trade-offs between the prescriptions from the ethics system and the behavioral patterns induced from historical data. In (Rossini et al., 2020) parameters imposed on joint probability distributions used to train the EA-DNN are used to bias more or less towards *business and ethical adequacy*. Notice that while gold standards (traditional training sets) only suggest ONE decision, that may also be unethical, a sharp probability distribution across possible business decisions is not helpful. We should leave some training evidence also to other unattended decisions, whenever more ethical. Smoothing is applied here that reserves some probability also to alternative decision. The amount of smoothing over probabilities for possible decisions $d_i \in D$ is modulated through a parameter α in a *Laplace* type of *smoothing*, expressed by:

$$\hat{d}_i = \frac{d_i + \alpha}{1 + |D|\alpha}$$

Notice that larger values for α favour business decisions that are progressively different from the training one, i.e. less and less optimal from the point of view of the business accuracy.

Dually we would increase the ethical accuracy, by making ethical sustainability to count more in the estimation of the target loss.

Through similar a technique, it's possible to tune the emphasis of ethical consequences by applying a tweaking factor β to the probability of benefits and risks in the joint probability of the triple (decision benefits and risks) $(\mathrm{d}, \mathcal{B}, \mathcal{R})$, i.e.,:

$$P_\beta = (d_i, \mathrm{B}, \mathrm{R}) = P(d(i)=d_i) \left(P(\mathrm{B} = v_j) P(\mathrm{R} = v_k)\right)^\beta$$

Here the influence of ethics turns weaker as $\beta \to 0$.

The above equation corresponds to the input of the network and establishes the influence onto the NN of the ethical information through the corresponding impact on loss functions L_{Er} and L_E.

7. Empirical evidence from experimentation: Ethical Risk Assessment in Banking

The evidence from extensive experimentnts show that the proposed neural architecture is able to induce models whose ethicality can be effectively controlled by the meta parameters of the network, in particular α and β.

To run the evaluation the German Credit dataset[14] (GC) has been applied, whereas the business task is to predict whether a loan request carries a "*low*" (*C0*) or " *high*" (*C1*) risk of default (i.e., the requester not paying back the loan). Instances of requests are described through 20 different attributes, some domain-specific (e.g.,

[14] Publicly available from the University of California-Irvine machine learning repository (Dua and Graff 2017).

PREVIOUS CREDIT HISTORY or ACCOUNT BALANCE) combined with more general ones (e.g., the AGE of the requestor's or the NuMBEROFPEOPLEUNDERMaintainance). Despite its small dimension (only 1000 instances) and strong class unbalance (700 instances labeled as "low" *C0* profile), this dataset is appealing to test ethics learning approaches as it represents a real-world problem (King, Feng, and Sutherland 1995) and offers many attributes upon which ethical rules can be defined.

In the experiments we gave the ethical ontology for granted. It includes two truth-makers:

- "MOTHERHOOD FOSTERING" (tm_{MF}), favouring (lending decisions representing) women with children or, to a lesser extent, men with at least 2 children, and
- "CULTURAL INCLUSIVENESS" (tm_{CI}), favouring foreign workers.

Due to the strong unbalance between the target classes (70%-30%), we report business performances according to the average F1-measure, *μF* 1, as: $\mu F1 = {}^{1}/_{2}\,(F1_{C0} + F1_{C1})$. The overall ethical compliance *EComp* of the data set, given the ontology, is computed as the percentage of ethically complaint instances, according to the gold standard decision, i.e. $\frac{D^{+}+D^{-}}{D^{+}}$. It corresponds to the *EComp*=0.70 that suggests that historical data alone cannot be used to promote ethics.

A straightforward measure of the trade-off between ethics and business accuracy is thus the parametrized Acceptability factor *EAccγ* as the weighted average between the μF1 and the ethical compliance *EComp*:

$$EAcc_{\gamma} = \gamma\mu F1 + (1-\gamma)EComp \quad (14)$$

where $\gamma\epsilon[0,1]$ can be adjusted according to the relative importance of the two terms. The $EAcc_{\gamma}$ measure, when the superiority of ethics is imposed by $\gamma = 0.2$ over the GC dataset, provides the strong baseline for ethical training given by $EAcc_{0.2}$ = 0.76. Such gold standard $EAcc_{\gamma}$ is a useful reference measure to compare ethical neural models.

The EbDNN architecture has been designed with an Ethics Encoder based on with 2 layers, where the first layer has the same size of the input and the second has dimension 400, and the Business Expert with 1 layer and an output dimension of *K* neurons. Both the Ethics Expert and the Ethics-Aware DNN have 1 layer with *K* m^2 neurons (where *K* is the number of classes and *m* the number of ethical values). Non-linearity is applied through the *relu* function at each layer, except for the last layer in each component associated with a loss function, where a *softmax* is computed. A dropout rate of 0.2 on each layer is applied. To cope with the limited number of instances, we applied 10-fold cross validation, training each model for 1000 epochs with a standard batch size of 256 through Adam optimizer. Various settings of the smoothing and tweaking factors[15] have been applied to systematically study their impact. We fed each model alternatively with the enriched input vector, i.e $(\vec{f}(i)\ ||\ \overrightarrow{es}(i))$ or only with business observable $\vec{f}(i)$. No significant difference has been observed as the EE-DNN seems able to robustly reconstruct ethical signatures across all settings. In the rest of the experiments, we thus trained models only over $\vec{f}(i)$.

To provide a fair comparison with a standard learning framework, a simple MultiLayer Perceptron (MLP) with 2 layers and 320 units per layer has been trained, over $\vec{f}(i)$. only: it achieves an accuracy of *76.21%* comparable to state-of-the-art results on this dataset (Ratanamahatana and Gunopulos, 2002). It corresponds to a business performance of *μF1=66.4%* with an ethical compliance $EComp$ =*77.8%*. The ethical acceptance is thus $EAcc_{0.2}$=*75.5%* that does not improve on the gold standard, as expected: it provides a second comparative reference as ethical unaware system.

7.1 Measuring ethical awarelearning

Table 1 reports the performances of both the baseline MLP and of the *EA* models, under different *α* and *β* settings and decision policies. The tradeoffs between ethical and business performances is largely improved by *EA* models for all the configurations. Gains in ethical compliance of *EA* models w.r.t. baselines are significant while business performance losses are relatively small. The effect of both factors (*α, β*) is observed in Figure 2. As *β* increases, ethics plays a stronger role, and the

[15] In all the experiments here reported the folliwng values have been adopted $(\alpha,\beta) \in \{0.1, 0.3, 0.6, 1.0\} \times \{0.01, 0.05, 0.10, 0.20, 0.35, 0.50, 0.75, 1.00\}$

model's behavior deviates from a purely business-driven predictor. The smoothing factor α plays a complementary role: stronger smoothing actions corresponds to markedly more ethical behaviours, even for smaller β. Notice how, even for high α values at lower β's ($\leq$0.1) every *EA* models starts to exhibit unethical choices. The fully enforced ethics network EA^{EC} with $(\alpha, \beta) = (1,0.5)$ achieves the maximal *EthCompl* with less than 20% loss in terms of μF 1. Note that, the *unconstrained* decision policy, i.e., the EA^{EU} model, is not sensitive to the tweaking factor, as for $\beta=0.5$ or $\beta=0.2$ the performance is basically the same. Figure 3 plots Ethical Acceptability $EAcc_{\gamma}$ (with γ=0.2) restricted to the test cases where the MLP provides non ethical decisions. The robustness of ethical aware networks is striking.

System (α, β)	μF 1	*EthCompl*	$EAcc_{0.2}$
EA^{EU}(0.3, 0.5)	63.1%	79.6%	76.3%
EA^{EU}(0.3, 0.01)	63.9%	78.8%	75.8%
EA^{EC}(0.1, 0.5)	41.2%	**100.0%**	**88.2%**
EA^{EC}(0.1, 0.2)	53.8%	93.0%	85.2%
EA^{EC}(0.1, 0.01)	61.7%	78.4%	75.1%
EA^{EC}(0.3, 0.1)	60.6%	85.1%	80.2%
MLP **66.4%**		77.8%	75.5%

Table 1: μF 1, EthCompl and EAccγ (γ = 0.2) for different configurations of the EA model

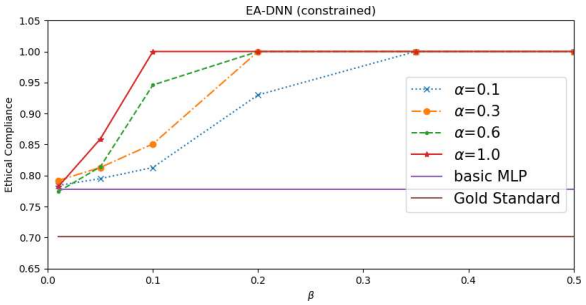


Figure 2: The trends of the Ethical Compliance *EComp* of the outcome of the EA-DNN as a function of the tweeking β. While MLP and Gold Standard refers to ethically unaware methods, plots represent several smoothing α parameters.

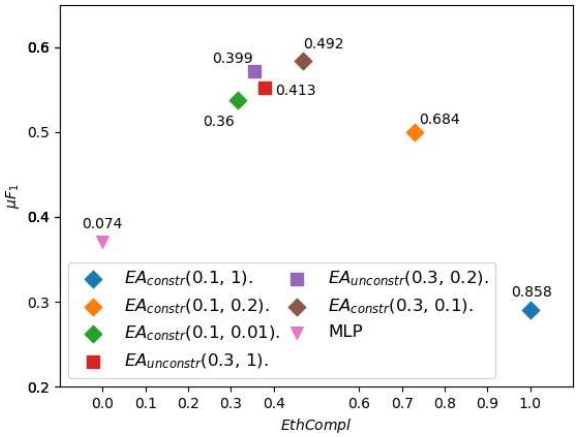


Figure 3: *EAccγ* for *γ*=0.2 of the *EA* model over non ethical decisions of the gold standard: performances of constrained (EA^{EC}), unconstrained (EA^{EU}) networks and the baseline MLP are reported against *EthCompl* and μF1 values.

7.2 Explaining NLP system decisions through Kernel-based learning machines

Recent work, i.e. (Croce et al, 2019) has emphasized as a specific post-hoc explanation mechanism, coupled with a given prediction, the comparison with available examples, namely the landmarks input to a neural learning model called KDA (Croce et al, 2017), relevant to, i.e. activated y, the decision. These landmarks share task-relevant linguistic properties with the input. From an argument theory perspective, this supports an "argument by analogy" explanatory schema (Walton, Reed, and Macagno 2008). In this way, a user exposed to such the argument will endow a different level of trust into the machine decision proportional to the linguistic plausibility and naturality of the analogy. He will implicitly gauge the evidence from the linguistic properties shared between the input sentence (or its parts) and the one(s) used for comparison as well their importance with respect to the output decision. For example, in a question classification (QC) task (Li and Roth 2006): the decision that Q:"What is the capital of Zimbabwe?" refers to a Location can be motivated by the system with an argument such as: Q refers to a Location since it recalls me of "What is the capital of California?" which also refers to a Location. Notice that explanation should not be just semantically similar but mainly effective. Semantic similarity such as the one used by a search engine plays only a minor role: clear and trustful analogies may exist with training examples that justify the correspondence between the input and the decision according to more complex semantic relationship, as side effects of the learning process.

The idea discussed in (Croce et al, 2019) is inspired by interpretability models adopted in image processing tasks, such as the Layerwise Relevance Propagation (LRP) (Bach et al. 2015). In LRP, the classification decision of a deep neural network is decomposed backward across the network layers: evidence about the contribution to the final decision brought by individual input fragments (i.e., pixels of the input image) is gathered as causal evidence. In (Croce et al., 2017) the LRP method is applied to a linguistically motivated neural paradigm, known as Kernel-Based Deep Architectures (KDAs), which inject semantic information expressed by Tree Kernel methods ((Collins and Duffy 2001), (Croce et al., 2011)) within a multilayer perceptron architecture. In this way, each system's prediction, e.g. a question classification step, generates as a side effect of LRP an argument-by-analogy explanation, that makes explicit reference to real training examples.

The evaluation of the proposed methodology shows the meaningful impact of LRP-based explanation models: users faced with explanations are systematically oriented to accept (as well as reject) the correct (vs. wrong) system decisions, respectively: as a consequence, through post hoc judgments the user may strongly increase the overall reachable application accuracy.

8 Layerwise layer backpropagation and Example-driven expanations in KDA

The approach proposed in (Croce et al, 2019) integrates two major ideas. On the one side the ability of some mathematical function back-propagated in the neural network to model the state of activation of its layers, called LPR, that tries to establish associations between decisions (i.e. network output signals) and input signals: this establishes a kind of causal connection as it tells which parts of the input (e.g. the fragments of an image) influences most the decision. On the other side, Kernel-based Deep Architectures (KDA, as in (Croce et al. 2017)) are used to provide express neural net input in terms of linguistics examples: this allows the causal connections to be traced back to natural language expressions, enabling explanations as combination of such linguistic inputs. The sort of explanations are built as argumentations around the notion of annotated example, that is readable, natural and meaningful to the human reader. An example of explanation of a question classification system that labels as LOCATION the question Q:

"Where is the Mall of the America?"

is something like:

I think Q refers to a LOCATION *since it recalls me of "What town was the setting for The Music Man?".*

Notice that *Where is the Mall of the America* is not syntactically similar or lexically similar to *"What town was the setting for The Music Man?"*, i.e. it is not possible to retrieve it through Q without learning the notion of Location. This means that the explanation is telling something about the neural model after training. Learning here has generalized across structured information related to syntax and semantics as well and this is exactly what makes explanation meaningful. Hereafter, we will summarize main aspects of the KDA technology, although we invite reader to refer to (Croce et al, 2019) for more technical details.

8.1 Layer-wise Relevance Propagation

The Layer-wise Relevance Propagation technique (LRP, as in (Bach et al. 2015)) is mainly a method to decompose the prediction of a deep neural network, as computed over an input instance, usually an image, down to the relevance scores characterizing individual dimensions of the input, such as the pixels of the image itself. More formally, let $f: \mathbb{R}^d \rightarrow \mathbb{R}^+$ be a function that quantifies, for example, the probability of $x \in \mathbb{R}^d$ being in a certain class. The Layer-wise Relevance Propagation assigns to each dimension, or feature, x_d a relevance score $R_d^{(1)}$ such that $f(x) \approx \sum_d R_d^{(1)}$. Features whose score is $R_d^{(1)} > 0$ or $R_d^{(1)} < 0$ correspond to evidence in favor or against, respectively, the output classification. In other words, LRP allows to identify fragments of the input playing key roles in the decision, by propagating relevance backwards. Let the relevance score $R_j^{(l+1)}$ to represent the status of activation of the neuron j at the network layer $l+1$. This can be decomposed into messages $R_{i\leftarrow j}^{(l,l+1)}$ sent from j to neurons i in layer l according to $R_j^{(l+1)} = \sum_{i\in(l)} R_{i\leftarrow j}^{(l,l+1)}$. Then ti directly follows that the relevance of a neuron i at layer l, that is the quantity of information travelling through i, can be defined as $R_i^{(l)} = \sum_{j\in(l+1)} R_{i\leftarrow j}^{(l,l+1)}$. The ϵ-rule defined in (Bach et al. 2015) allows to compute back the messages $R_{i\leftarrow j}^{(l,l+1)} = \frac{z_{ij}}{z_j+\epsilon\cdot\mathrm{sign}(z_j)} R_j^{(l+1)}$, where $z_{ij} = x_i w_{ij}$ and $\epsilon > 0$ is a small numerical stabilizing term. This information

is relevant as the weights z_{ij} can be linked back to the activation weights w_{ij} of the input neurons.

The result is a map of the mostly activated area of the input that justify (or explain) the essence of the decision.

8.2 KDA

In (Croce et al. 2017), a small dimensional representation $\tilde{\vec{x}}$ for linguistic data is discussed (called Nyström representation[16]) used to map semantically annotated sentences (i.e. grammatical trees corresponding to examples of questions Q in some classes, e.g. Location) into the input numerical vector of a Multi-Layer Perceptron (MLP). The MLP architecture consists of a specific Nyström input layer based whose contribution is to represent a Nyström embedding: such vector is the reconstruction vector of a generic input sentence, obtained as a function of randomly selected set of training instances, called landmarks. As the reconstruction is driven by a metrics based on the semantic kernel function, it can be considered as a low dimensional representation isomorphic to a point in the kernel space. The resulting architecture is a MLP driven by this kernel function and it is called Kernel-based Deep Architecture (KDA). The *input layer* in a KDA, the *Nyström layer*, is followed by a sequence of *hidden layers* up to the final *classification layer*, which output the final decision through a softmax operator, as shown in Figure 1. A KDA optimizes the standard cross-entropy function with L_2 regularization.

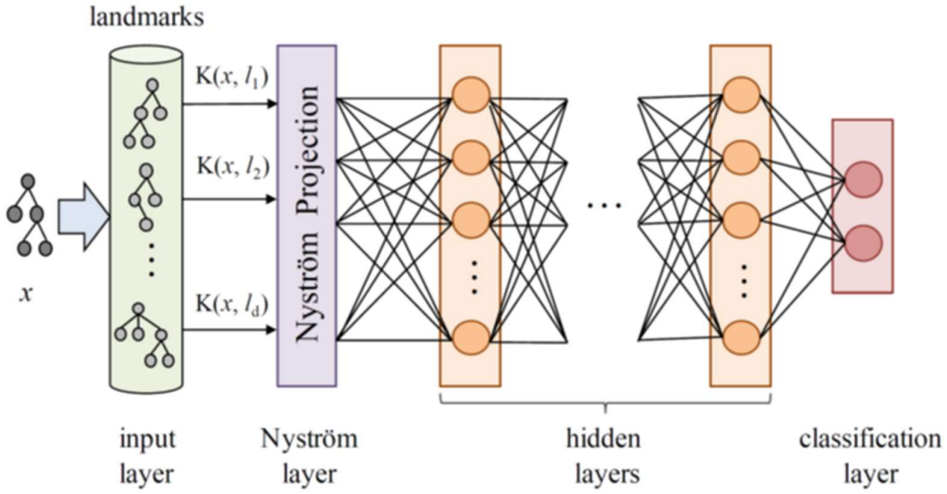


Figure 1 - Kernel-based Deep Architecture

It is worth recalling that the network is triggered by an input vector $\vec{c}$ expressing the kernel evaluations $K(x, l_i)$ between the example and the landmarks. When using linguistic kernels (such as Semantic Tree Kernels, (Croce, Moschitti, and Basili 2011)), this measure corresponds to the grammatical and lexical semantic similarity between x and the subset of landmarks.

9. Explaining KDA decisions in Applications

The explanation of individual decisions is obtained from the KDA network output by applying LRP to revert the propagation process and link the output back to the input. In a KDA whose input layer models linguistic instances, that is landmarks, LRP implicitly traces back that the mapping can be expressed through the Nyström matrix $H_{Ny} = US^{-\frac{1}{2}}$. The resulting $\tilde{\vec{x}}$ is the input to one or more *hidden* layers. Clearly, the first hidden layer receives in input $\tilde{\vec{x}} = \vec{c}H_{Ny}$.

[16] The *input* layer corresponds to the input vector $\vec{c_i}$, i.e., the row of the C matrix associated to an example o_i. The input layer is mapped to the *Nyström* layer, through the Nyström projection. Notice that the embedding provides also the proper weights, defined by $US^{-\frac{1}{2}}$, so

syntactic, semantic and lexical relations between the input example and such landmarks: the Nyström layer thus select those real examples that mostly influenced the identification of the predicted structure in the sentence. Moreover, landmarks are annotated examples of the training set and provide positive or negative evidences. In this way compiling an explanation can proceed through different argumentations as response to different inquiries, such as:

I1: What is the class of the question *"Where is the Mall of the America?"?*

A1*: I think it refers to a* LOCATION *since it recalls me of "What town was the setting for The Music Man?*".

I2: Is "*George Bush purchased a small interest in which baseball team?* " a question about Group rather than about an Individual?

A": *I think it does not refer to* Individual *because it does not remind me of "What is the name of the company that manufactures the `` American Girl" doll collection ?" , but I think it is* Group *as it reminds me of "What actor and actress have made the most movies?"*

whereas underlined questions (e.g. "*What actor and actress have made the most movies*?") in the argumentation are landmarks in agreement with the system decision, while other questions in the explanation are used as counterexamples. In this way several questions such as "*Why* Class *is the class of Q?*" or "*Why* Class *is not the class of Q?*" can be replied with meaningful explanations. In (Croce et al, 2019) evidences about the fact that activated landmarks are different with respect to semantically similar questions, according to the applied semantic kernel, are reported: this shows how explanations are model dependent and better than simple analogies between input and training questions. Moreover, application to explainability in other inductive NLP tasks is reported.

An interesting aspect of the above method is that is a beneficial side effect of the KDA architecture and can be applied to any text–based classification task. Measures based on the help of explanations to avoid accepting system wrong decisions are also carried out in (Croce et al, 2019). Human annotators are exposed to decisions and provide judgments about the explanations: explanations are evaluated on a scale of five coarse categories from completely inconsistent (Incoherent), poor (Bad), average quality (Uncertain) to good or very good. Results are reported in Figure for different explanation modalities and against positive inquiries (the question such as "*Why* Class *is the class of Q?*" that support the acceptance of a class) as well as negative ones (with inquiries such as "*Why* Class *is not the class of Q?*" that explain rejections).

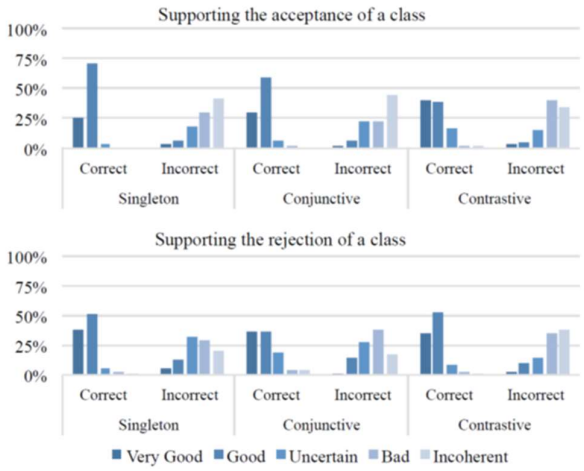


Figure 2 - Results of the audit process for the three explanatory models about the acceptance of system correct decisions (top), and their rejection (bottom).

In order to test the ability of the system to explain, an equal number of correct decisions (true positives and negatives) and wrong (i.e. incorrect) choices (i.e. false positives and false negatives) are presented to the annotators. If we look at the distribution of judgments wrt the correct vs. incorrect in Fig. 2, we see that for all modalities and inquiries the distributions of good quality decisions are always strongly correlated with good quality explanations. In this way, it is demonstrated that with post auditing the network and evaluating the quality of the explanation is a good way to strengthen the commitment toward good decisions and get a warning (i.e. badly compiled, or inconsistent, explanations) against incorrect ones. As a consequence, a good training plus post auditing is a guarantee for using the network (i.e. the targeted learning machine here) in more reliable and faithful manner.

10. Conclusions

The above methods have been used to demonstrate current contributions that adaptive NLP systems are bringing to the study of reliable, explainable and ethical

AI systems. One more time, these issues strongly regard the relationships between knowledge, as the body of evidence about the world that through its modularity is able to support reliable, flexible and accurate decision making in intelligent agents. At the same time, the acquisition and the use of such evidence is usually guided by inferences that make reference to inductive paradigms, such as statistical learning algorithms or neural networks. These latter are strongly focused on the contextual conditions in which the knowledge is used, that is they reflect the praxis of making decision in operational terms, by using experience as a guide to minimize risk and optimize benefits of the target decisions. The above two elements interact strongly. First, no learning can be triggered without any knowledge about a problem and its observable aspects in operational contexts. The type of structured architectures discussed in this paper are examples of neural paradigms that are strongly rooted in the nature and morphology of the targeted problems, and derive their major impact on the beneficial dependencies among data sets and across their different instances: the semantic kernels of the KDA architecture allow to reconstruct sentences in a semantically expressive kernel spaces able to capture lexical semantic, grammatical and conceptual similarities evoked by the sentences that are key to the successful treatment of meaning in the question classificartion as well as in the explanation task. They provide quantitative measures (i.e. metrics) able to well express language semantics and thus resulting in meaningful models whose explanation is straightforward. Natural language is used here as a basis for knowledge expression, whereas lexical items and grammatical structures are sufficient for most inferences.

Knowledge is also what is needed in the first section, about actionable ethics. It is the key role of an ethical ontology, with truth-makers as probabilistic models of ethical inference that are made available to neural learning, and can be the target of a specific subnetwork (see the role of the Ethics Expert in Fig. 1). The resulting deep learning framework is able to support high quality inferences that simultaneously reflects ethical expectations. The shortly reported experiments suggest that the framework is quite effective in allowing a fine-tunable balance between business and ethics objectives, through the smoothing and tweaking methods.

Again knowledge and induction are two faces of the same coin. On the one side, a need exists for modular knowledge components, whose logical formalisms optimize consistency-checking operations and reuse. On the other side, inference from experience is essential to optimize accuracy in a realm where availability of large scale data sets is the standard. However, learning architectures are strongly empowered whenever the they are made sensitive to logical structures implicit in the description of the problem instances, as the two discussed use cases demonstrate. Injecting symbolic knowledge, i.e. logically expressed information about the word and the contexts, within an inference process guided by data, i.e. based on the latter perspective, is a core research area as recent and growingly complex neural learning paradigms suggest. The area of NLP is crucial in this sense, as language is the basic formalism able to capture, express and disseminate knowledge. Linguistic data are thus important for two reasons: first they trigger an important set of learning task, as language interpretation is a very complex process inspiring a lot of novel learning methods and techniques. Second, and more importantly, natural language is a knowledge medium and linguistically inspired models induces in complex environment are instances of large scale knowledge bases, something of extreme interest in modern AI and in current, technology driven, societies.

11. References

Amodei, D.; Olah, C.; Steinhardt, J.; Christiano, P. F.; Schulman, J.; and Mane´, D. 2016. Concrete problems in ai safety. *arXiv preprint arXiv:1606.06565*.

Awad, E.; Dsouza, S.; Kim, R.; Schulz, J.; Henrich, J.; Shariff, A.; Bonnefon, J.-F.; and Rahwan, I. 2018. The moral machine experiment. *Nature* 563.

Bach Sebastian, Alexander Binder, Gregoire Montavon, Frederick Klauschen, Klaus-Robert Muller, and Wojciech Samek. 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLOS ONE, 10(7).

Bennett, M. 2013. The financial industry business ontology: Best practice for big data. *Journal of Banking Regulation* 14(3-4):255–268.

Bird, S.; Barocas, S.; Crawford, K.; and Wallach, H. 2016. Exploring or exploiting? social and ethical implications of autonomous experimentation in ai. In *Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT-ML), New York University*,

Boddington, P. 2017. *Towards a Code of Ethics for Artificial Intelligence*. Springer.

Bonnemains, V.; Saurel, C.; and Tessier, C. 2018. Embedded ethics: some technical and ethical challenges. *Ethics and Information Technology*.

Bostrom, N., and Yudkowsky, E. 2014. *The ethics of artificial intelligence*. Cambridge University Press. 316–334.

Chakraborty, Supriyo, Richard Tomsett, Ramya Raghavendra, Daniel Harborne, and Moustafa Alzantot et al. 2017. "Interpretability of Deep Learning Models: A Survey of Results." *2017,* 1–6.

Collins, Michael, and Nigel Duffy. 2001. "New Ranking Algorithms for Parsing and Tagging: Kernels over Discrete Structures, and the Voted Perceptron." In *Proceedings of Acl '02, July 7-12, 2002, Philadelphia, Pa, Usa*, 263–70.

Craglia, M. 2018. *Artificial Intelligence: A European Perspective*. Publications Office of the European Union.

Danilo Croce, Alessandro Moschitti, and Roberto Basili. 2011. Structured lexical similarity via convolution kernels on dependency trees. In Proceedings of EMNLP '11, pages 1034–1046.

Croce, Danilo, Simone Filice, Giuseppe Castellucci, and Roberto Basili. 2017. "Deep Learning in Semantic Kernel Spaces." In Proceedings of Acl 2017, 345–54. Vancouver, Canada. http://aclweb.org/anthology/P17-1032.

Danilo Croce, Daniele Rossini and Roberto Basili, *Auditing Deep Learning processes through Kernel-based Explanatory Models*, Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pages 4037–4046, Hong Kong, China, November 3–7, 2019.

Dua, D., and Graff, C. 2017. UCI machine learning repository.

Goodfellow, I. J.; Shlens, J.; and Szegedy, C. 2014. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*.

Goodfellow, I.; Bengio, Y.; and Courville, A. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

King, R. D.; Feng, C.; and Sutherland, A. 1995. Statlog: comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence an International Journal* 9(3):289–333.

Kleiman-Weiner, M.; Saxe, R.; and Tenenbaum, J. B. 2017. Learning a commonsense moral theory. *Cognition* 167:107 – 123. Moral Learning.

Li, Xin and Dan Roth. 2006. Learning question classifiers: the role of semantic information. Natural Language Engineering, 12(3):229–249.

Lipton, Zachary C. 2018. "The Mythos of Model Interpretability." *Queue* 16 (3): 30:31–30:57. https://doi.org/10.1145/3236386.3241340

O'Neil, C. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY, USA: Crown Publishing Group.

Ratanamahatana, C. A., and Gunopulos, D. 2002. Scaling up the naive bayesian classifier: Using decision trees for feature selection. In *Workshop Data Cleaning and Preprocessing (DCAP 2002), at IEEE Int'l Conf. Data Mining, ICDM 2002*. Citeseer.

Daniele Rossini, Danilo Croce, Sara Mancini, Massimo Pellegrino, Roberto Basili: Actionable Ethics through Neural Learning. AAAI 2020: 5537-5544

Smuha, N. 2019. Ethics guidelines for trustworthy AI. Publications Office of the European Union.

Vanderelst, D., and Winfield, A. 2018. An architecture for ethical robots inspired by the simulation theory of cognition. Cognitive Systems Research 48:56 – 66. Cognitive Architectures for Artificial Minds

Walton, Douglas, Christopher Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press.

THEME 5: Wargaming

Introduction to Wargaming: Preparing the Next Generation of U.S. Army Professional Wargamers

Dr. James Sterrett
Mr. Michael Dunn
Mr. Jeff Hodges

The United States Army has re-invigorated wargaming across the force. Army organizations that support strategy, acquisition, analysis, and readiness have embraced parts and pieces of classic wargaming principles to inform senior leader decisions. There are various Army branches and functional areas that either have or currently adopting warfighting fundamentals into their mission analysis and course of action comparisons. This article will focus on how the Army is preparing one functional area (FA), the simulation operations officer, or FA57, to support wargaming in support of readiness. The FA57 officer uses models and simulations to provide enhanced decision support, readiness (training environments), and modernization recommendations to commanders at every echelon of the U.S. Army. Over 80 percent of the Active Component FA57 positions are in direct support of readiness. FA57 officers, who are responsible to build and execute readiness exercises and events in support of their commander's warfighting mission, provide support to separate brigades, divisions, corps, and Army service component commands (ASCCs), as well as joint and combatant command (CCMD) organizations. FA57 officers are assigned to current operations directorates as exercise planners. Historically, these officers have not been an integral component of the division, corps, ASCC, or CCMD wargame development. Wargame development, in most cases, was the responsibility of the future operations or planning branch of the operations directorate. The FA57 community began to identify an education and training gap as a result of this emerging requirement from the operational force. This identification came from profiled messages between senior FA57 leaders, forum discussions, and collaboration sites. In April 2019, The Army Modeling and Simulation School (AMSS), Headquarters Department of the Army school responsible to train FA57 officers throughout their careers, began to analyze the wargaming gap identified by the FA57 workforce. Each active component FA57 received an education / training gap questionnaire so that the school could determine if a wargaming gap existed. Thirty-one percent of the FA57 community replied to the survey over the next 60 days. From the survey, AMSS determined that 82 percent of the survey respondents believed that formal wargaming training was required for the FA57 to be better prepared to support wargaming at the division, corps, and higher echelons.

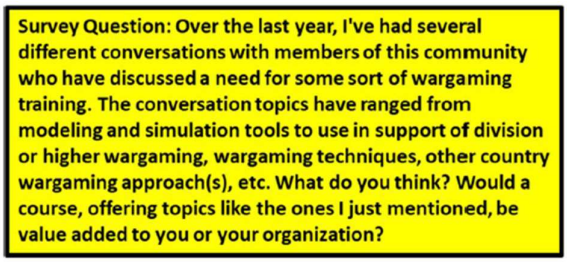


Figure 1: Survey Question

However, there were many differing opinions from the community regarding the specific type of training required for FA57s to be more fully prepared to plan and participate in an organizational wargame. AMSS clustered the data into categories in an attempt to visualize the potential educational topics better that would have to be supported during a school sponsored training event. Figure 2 highlights these topics.

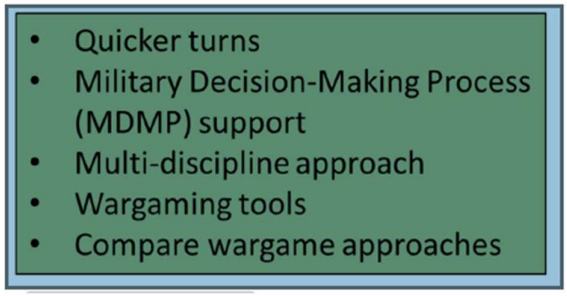


Figure 2: Clustered Categories

The survey respondents described requiring a tool that allowed for quicker turns and that could be used in field or deployed conditions. Military Decision-Making Process (MDMP) support was frequently mentioned in the survey responses. An introduction to wargaming tools and processes were discussed in the individual responses as well.

The summer and autumn of 2019 were dedicated to additional senior manager interviews, market surveys of existing wargaming courses, and follow-on conversations with survey respondents to ensure that

AMSS understood the scope of the educational needs required to best prepare the FA57s supporting the operational force. AMSS reviewed other U.S. military, allied military, academia, and commercial wargaming courses to determine if these courses would support the training gaps identified in the survey responses. The market survey conducted by the school did not discover an introductory wargame course, focused on ground centric, division or corps level MDMP support.

In December 2019, the wargaming course analysis phase was essentially completed at the 2019 Interservice / Industry Training, Simulation and Education Conference (I/ITSEC) when AMSS conducted final education and training gap interviews with the workforce and conducted a 4-hour tactical wargaming workshop.

Figure 3: Tactical Wargaming Workshop

Thirty-five professionals participated during this workshop, including representatives from the United Kingdom(UK) Ministry of Defence, the Israeli Defence Force, Army Capability Manager's office, Army Program Executive Office organizations, industry, and academia. During this workshop, AMSS received a great deal of input regarding best practices and approaches in support of wargaming training. Following the 2019 I/ITSEC, AMSS and the U.S. Army Command and General Staff College Directorate of Simulation Education (DSE) agreed to partner on an effort to design and execute an "Introduction to Wargaming" course that would be tailored to support the tactical- operational-level FA57s.

In January 2020, AMSS and DSE representatives visited the technology school of Defence Academy of the United Kingdom in Shrivenham, UK, to discuss AMSS's and DSE's "Introduction to Combat Modeling and Wargaming" course. During the visit, the U.S. and UK school faculty discussed education techniques, procedures, and challenges that are associated with the modeling and simulation workforces of both nations. During the visit, the UK faculty shared with the U.S. faculty a product developed by the United Kingdom: the Rapid Campaign Analysis Toolkit (RCAT).

Figure 4: RCAT

The U.S. faculty decided to incorporate the game into the new "Introduction to Wargaming" course being developed; RCAT would become the culminating event in the new course and the take home item for the students enrolled in the course. In addition, the UK faculty shared combat modeling courseware and educational techniques for the combat modeling portion of the UK course.

The combined U.S. faculty began to focus on the "Introduction to Wargaming" course design following the visit to Shrivenham. A draft training program was

developed; a 1-week resident experience that would introduce the student to the fundamental and principles of wargaming types, adjudication methods and design. The course would be almost 100 percent application-level learning; very little course lecture was planned. Instead, nearly all resident time would be devoted to hands-on learning. The COVID 19 pandemic shelved the resident course plan. The U.S. team decided to pilot the course in a live / online environment.

Instead of a 1-week resident course, the faculty designed a course that met 12 times from September to December. The class met weekly for a 3-hour session that focused on hands-on wargaming experience. The students were given extensive weekly reading and essay writing assignments from the following books, manuals, and articles: The Art of Wargaming: A Guide for Professionals and Hobbyists, Dr. Peter Perla; Wargaming Manual, UK Ministry of Defence; selected chapters from the U.S. Army Field Manual 6-0, Commander and Staff Organization and Operations; Center for Army Lessons Learned Handbook 20-06, How to Master Wargaming: Commander and Staff Guide to Improving Course of Action Analysis; and "Closing the Gap Between Simulations for Training and Wargaming,", an article by COL Jeff Erickson and Mr. Garrett Heath.

Date	Topic	Class Activity	homework
September 16, 2020	Wargaming Principles	Battle for Moscow	Perla 17-153, 3 essay questions
September 23, 2020	Manual / Computer Wargames	Battle Academy 2	Battle Academy 2 essay
September 30, 2020	Strict v Rigid	South China Sea	Perla 157-239, 3 essay questions
October 7, 2020	Adjudication Techniques	Kriegsspiel 1	Perla 240-287, 3 essay questions
October 14, 2020	Adjudication Techniques	Kriegsspiel 2	UK MOD Wargaming handbook, 3 essay questions
October 21, 2020	Joint Planning System overview	Joint Planning Service	FM 6-0, 3 essay questions
October 28, 2020	JPS ISO MDMP	MDMP	CALL HB 20-06, 3 essay questions
November 4, 2020	RCAT ISO MDMP	MDMP	"Wargame Pathologies", 3 essay questions
November 10, 2020	Wargame Design	RCAT supported design	"Closing the Gap", 3 essay questions
November 18, 2020	Wargame prep	Prepare design	
November 24, 2020	Wargame execution	Execute design	
December 2, 2020	final reflection / AAR		

Figure 5: Weekly Course Schedule

Students were introduced to wargaming principles with "Battle for Moscow" using the Vassal engine. Differences between manual- and computer- assisted wargames were highlighted using "Battle Academy 2", a Slitherine UK commercial product. The matrix wargame discussion was supported by "South China Sea," using Google Drive to host the event. Free and rigid Adjudication techniques were supported by "Classic Kriegsspiel." The Joint Planning System was used to support MDMP analysis wargaming. The UK product, RCAT, was used as the primary tool for wargame design principles, preparation, and execution.

The course concluded on December 2, 2020, with a 90-minute after action review (AAR). The end-of course survey indicated that 100 percent of the student cohort agreed that they had a much better understanding of wargaming principles, wargame types, adjudication techniques, and design fundamentals than at the beginning of the course. Ninety percent of the student cohort agreed that the assigned readings and essays were relevant and helped them prepare for the weekly seminar.

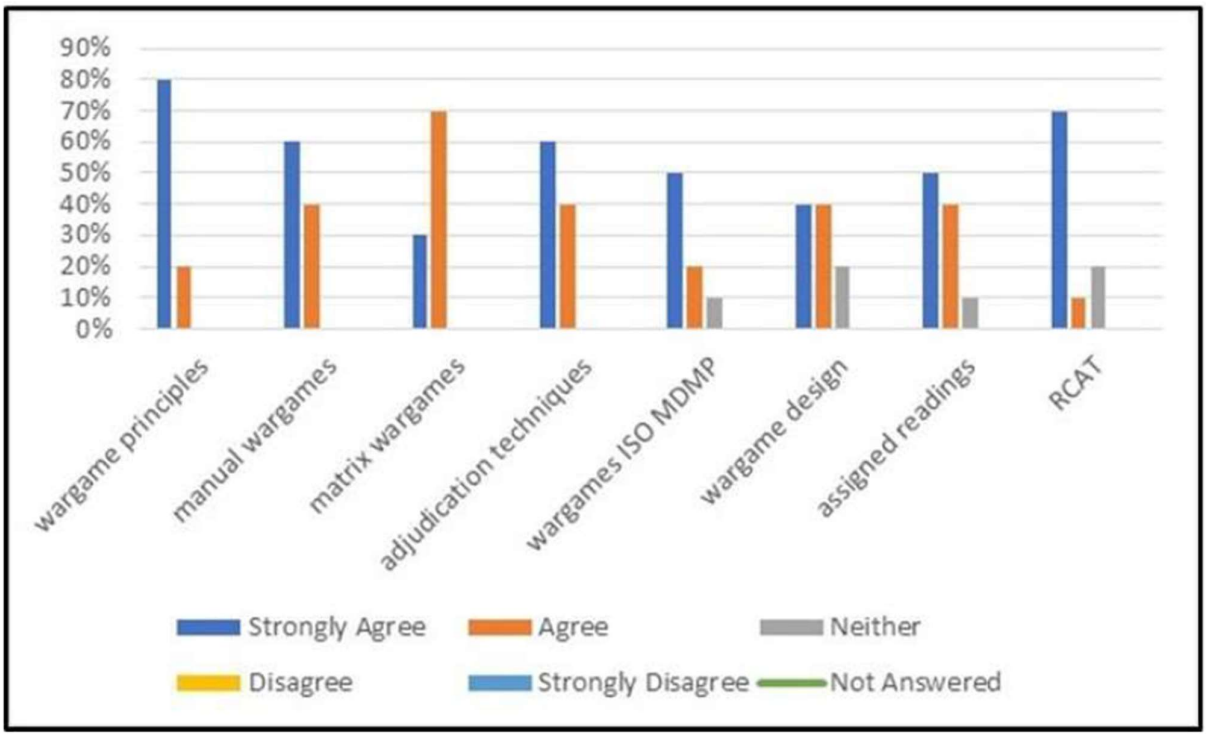


Figure 6: End of Course Survey Results

Additionally, the students recommended that AMSS and DSE add an additional weekly reading from recent journals and periodicals. The majority of the cohort believes that a 4-hour, instead of 3-hour, weekly seminar will allow for a more thorough understanding of the topic and tool and would be supported by organizational leadership. Perhaps the biggest lesson learned from the students was their belief that the course would be much better if operators, strategists, and analysts were included in future cohorts.

The DSE and AMSS team have learned a great deal from the pilot course experience and have already begun revising the curricula and approaches for the next "Introduction to Wargaming" course offering (most likely summer 2021). The team's initial plan is to create a blended delivery course, with most of the course remaining live / online, followed by a resident portion. Additionally, the team is beginning to design a follow-on course to provide our workforce with a practitioner-level of knowledge and experience. The team is also coordinating with other Department of Defense and Allied partners to determine the level of support for a combined 'continuum of wargaming education' for strategists, operators, analysts, and designers. Most importantly, the team plans to offer this course to the other key wargaming principles supporting the commander: the strategist (Army Functional Area 59), the operator (the operations officer for the organization), the analyst (Army Functional Area 49), as well as the designer (the FA57).

Wargaming Cyber: Tactical Hack

Alex Hoover
Irrational Number Line Games, LLC
etotheipi@inlgames.com

1 Uses of Wargames

Cyber operations, both offensive and defensive, are a de facto part of modern warfighting. Wargames are a key tool for understanding warfare, and benefit from representing all the relevant domains involved in the conflict. Modeling cyber operations has been somewhat limited to cyber only modeling, as opposed to treating cyber as a combined arms capability. There are three common top-level categories for the use of a wargame - the training game, the analytic game, and the entertainment game, where cyber operations can be implemented.

2 Training

In the training game, the purpose of the wargame is to have the players practice certain activities in a specific set of warfighting contexts in order to increase proficiency. The two driving factors in the design or rules and scenarios for a training wargame are: (1) ensuring sufficient player activity focuses on the designated set of warfighting tasks, and (2) ensuring the conditions under which the players conduct those tasks in a representative set of the proper environments. This has a significant impact on what we often call realism.

If there is a one in a thousand chance that the desired activities will not be a part of the actual warfighting event, we leave that chance out of the design. Likewise, if there is a 99% chance the desired activities, we leave those out of the design as well. This seems counter to the desire for realism, but actually it highlights the principle that realism is subordinate to purpose in wargame design. Designing a training wargame for only outlier events is fairly common. In fact, rare events that are critical to success, difficult to perform well, and occur under arduous conditions are often the activities we need to train most, as opposed to routine activities that we may practice frequently during everyday routine operations.

Because a training game is focused on presenting the practice environment and stimuli to the players, it is important to remember that you may have *people* participating in the game who are not *players*. We often call them the read team (OPFOR), the white team (referees), or other team names. These teams are not players, but part of the environment that provides the stimuli for the players. They will often need different dynamics than players, and are often more powerful assets in your game to keep the scenario focused on the objective than any complex set of rules, if we design in those functions.

3 Analysis

The analytic wargame has a different purpose, but still retains some characteristics in common with training. Rather than presenting an environment to the player, it is about answering a question about warfighting. The discipline of operations analysis allows analysts to take these large questions. A common technique is clarification and investigation.

Clarification is taking a broad concept within the question and working with the sponsor to determine whether everything that concept means is included or just specific subsets. For example, are we looking at the performance of "the Army" or some specific part of the Army? Clarification can be conducted iteratively on the results of other clarifications, or longitudinally, across different parts of the questions. Longitudinal clarifications often have overlaps which lead to important interfacing needs in the wargame design.

Investigation is taking the activities and conditions and determining through research, what static and dynamic stimuli represent the activities and conditions you have identified. This is conducted by referring to a standing body of knowledge on these topics. Investigation is where the sponsors question (what we want to know about warfare) meets our objective understanding (what we already know about warfare). It structures the artifacts of the game by identifying static stimuli and frames the rules by identifying the relevant dynamic conditions.

This leads to the question of where we stop. How do we avoid "paralysis by analysis" in clarification or investigation? Like the training wargame, the analytic wargame has a specific purpose. We don't want to include everything in the game; we want to include just the relevant things. Including things that don't have a strong impact on the analytic question is adding

unnecessary cost (money, but also the more precious commodities of effort and time) and creating a greater opportunity for random problems to affect the outcome.

Consider doing your personal budget as an analytic activity. For some types of financial decision, you identify the required performance (how long will a microwave last? how much can I cook in it at once?) and determine the best way to get that performance. In other types of financial decision, you start with how much money is available and then determine how much you can get for that (do I rent or own a movie? do we have a short, expensive vacation or a longer, cheaper one?). When deciding on a microwave or a vacation, your personal context drives what level of detail you look into each one.

This understanding of scope, not identification of details. is what you are trying to discover working with the customer. Clarification continues until you reach a desired level of granularity, usually a set of specific activities and conditions of interest, similar to a training purpose. Investigation continues until you have the relevant details for the items of interest.

4 Entertainment

This seems to be the easiest type of game to design, but often it is not. In both training and analysis we have customers who have identified desires (desires that hopefully align with their needs). In entertainment, the desired outcome is enjoyment. Enjoyment is a large, complex, and subjective topic, but by focusing on the other two applications, we can get some insight into how to achieve the goal of entertainment in a wargame.

The big idea is that enjoyment occurs in training and analysis activities. That sounds wrong, since those things are work, not leisure. But, if you don't derive enjoyment from your work, you might be in the wrong line of business. That does not mean that every moment of work is enjoyable. Consider troubleshooting a technical problem, building a financial plan for your organization, or executing your exercise regimen. Every moment of these activities is not pure enjoyment. In fact, there is often stress, worry, and anxiety in them. But when you achieve a goal (even if it was not your initially identified goal), you enjoy the sense of accomplishment. And during the process, it isn't all stress, worry, or anxiety – you can enjoy achievement of steps toward the goal, and even enjoy anticipation of achieving those steps.

That enjoyment, however, is not all about achievement. Consider the game Candyland. It is a good game for teaching young children about rules, taking turns, counting, and a number of other growth activities. It provides progressive "rewards" of reaching different brightly coloured sweets. And it is fine for a while at a certain level of age and development. But the outcome is purely random. Winning is not determined by your decisions as a player. Put simply, there is no challenge beyond mastery of the growth activities. Once those are mastered, there is no longer enjoyment in the game.

This is the key concept in entertainment – the pairing of challenge and achievement. While the player participates in all the activities of the game, what they experience is the challenges and achievements. For a player to a player to immerse themselves in the game, internalize the results, and put appropriate effort into making their decisions, they must see a strong correlation between challenge and achievement. This creates the temporary suspension of disbelief necessary for useful and enjoyable participation

To flip the earlier statement about enjoying your work, when people have options, they choose to pursue work where they enjoy the pairing of challenge and achievement. In the same way that training and analysis wargames share some common principles, so does entertainment wargaming share common principles with the other two.

5 Mixed Purpose Games

In a broader sense, the given three classifications of wargame uses are not so much separate things as they are the identification of important principles of design that apply to all wargames, in differing amounts. Within a wargame, there may be different parts of the game that need different levels of analytic depth. The focus of player decisions may need to be broad in some parts of the game and very specific for others. Sometimes we will need to sacrifice the correlation of challenge and achievement for correctness and other times we will need to sacrifice some correctness to strengthen the correlation between challenge and achievement.

Balancing scope, detail, and suspension of disbelief requires managing conflicting requirements. Consider a training game with a requirement to train on intelligence collection. We want to train the leadership group in the decisions related to prioritizing collection requirements

and the operational group on the activities required to conduct the collection. This has an inherent conflict. We will consider the extremes of "good" and "bad" performance for each cell and how it affects the game for the other cell.

If the command cell does a very bad job at prioritization, the operational cell may be scattering their efforts across unconnected things in a manner that does not represent how they really need to train. if the command cell does a very good job at prioritizations, it may not create opportunities for the operational cell to practice mid-execution replanning. If the operational cell does a very bad job at collection, the leadership cell may internalize that a good plan from them was actually a bad plan because of the outcome. If the operational cell does a very good job in execution, it may not generate the necessary challenges for the leadership call since all relevant information is provided in a timely manner for all circumstances.

The point of looking at the extremes is not to find problems with the wargame. It is to describe the performance boundaries of the game, so we can look at the range of possible performances. The ends allow us to find the middle. With this understanding, we can select mechanisms for the game that eliminate or mitigate the potential problems and drive the game toward the desired state. In this case, the scoping could lead us to identifying the need for a referee in this part of the game to "give a hint", "let a bad outcome slide", or "ramp up the challenge", where other adjudications in the game could be handled by automate closed-form or random methods.

These referee actions may break some sense of micro-realism for the game, but they preserve the objective, the broader context of realism. We can use outliers to define the bounds of performance for our game, but we are not necessarily bound to having those events happen in the game. Consider gambling as an example. In a lottery, someone could win. In a raffle, someone will win. That doesn't mean it is sound financial advice to invest in those activities with an expectation of a beneficial outcome. Likewise, just because an event is possible in a warfighting problem, doesn't mean we need to invest time, effort, and money into making it happen.

6 Instantiations of Wargames

Another common top-level taxonomy for wargaming is to separate games into Live (L), Virtual (V), and Constructive (C) modalities representing whether the player interacts with real world system and unit surrogates, real world system surrogates and artificial unit surrogates, or artificial system and unit surrogates. This creates the possibility for four hybrids – LV, LC , VC, and LVC, where parts of the wargame are presented in different modalities that are linked into a common battlespace. Even where we classify a wargame in one of main three categories, there are usually elements of at least one of the other modalities present. This concept becomes important when designing the dynamics of the game, so that you consider the interfaces among modalities as integral to those interactions.

In a live training event, we may conduct some actions in a purely live modality. Consider deployment and staging of forces into the theater. We can likely practice all the details with the operators and equipment. We can even create or find live surrogates for different situations such as occupying forested rolling hills, rocky buttes and bluffs, or a small unoccupied town.

But the number of tasks we can conduct easily in a pure live environment is limited. For example, we may want to deploy to an unoccupied town that includes a river with a destroyed bridge. Instead of finding a live surrogate, we can create a LV river by painting some lines on the ground and telling the deploying force to execute their bridging and fording activities to get across.

The concept extends across all the different categories and leads to most wargames being some combination of LVC, with different levels of different modalities for different parts of the game. Rather than walk through examples of all the hundreds of combinations of modalities and instantiations, we will focus one more example on the constructive modality implemented as a tabletop wargame.

Tabletop games are easy to look at as being purely constructive. And some are. We have physical artifacts that serve as surrogates for the units and environment. Even the dynamics are constructive, often implemented with charts and dice. And just like some training can be done purely live, some tabletop games can be purely

constructive. But one aspect of warfare usually requires some live and virtual implementations – information.

7 Information Warfare

Some people look at information warfare as a modern extension of the traditional warfighting domains, but using information (acquiring, denying, disrupting, and managing it) as a force multiplier has been a part of war throughout, and even before, recorded history. Over time, the way warfighters interact with information has gone through many changes. New technologies have emerged, and from that new information warfighting concepts have evolved. Many foundational ideas endure. A modern satellite controlled, third party illuminated weapon system still deals with the problems of detection, identification, localization, and targeting as did a Neolithic man with a spear. Modern information warfare is evolving rapidly, yet all versions of information warfare share some common principles. Focusing only on the technology is only looking at part of the problem, which is never a good idea in warfighting.

Representing information warfare concepts in a tabletop game requires a set of techniques that are different from the basic constructive representation of terrain and units. One of the most challenging (and frequently addressed) aspects is representing information scarcity – hidden knowledge. This challenge gets at the heart of one of the common criticisms of tabletop wargaming, the "God's eye view". All the artifacts are public knowledge.

Undetected movement, unknown unit composition or capabilities, and partial information on opponent unit status are all commonly dealt with in tabletop games by adding live or virtual components to the game. The double-blind refereed game is a classic technique for representing information scarcity. In this type of game, opposing players play on different boards in different rooms with only their units and a referee (non-player game participant) provides information back and forth. The referee may be a live asset, representing a messenger bringing intelligence reports to a commander, or they may be a virtual surrogate for an information system providing the same types of information.

In a computer assisted exercise (CAX), we do much the same thing, using a computer as a faster and more productive, but less capable referee. In the CAX, instead of starting with a single ground truth representation and parsing out what we want to be perceived truth, we start with an inaccessible ground truth and only present to the player the perceived truth we build. In either case, it is important for design of the game to separate ground truth from perceived truth. Both truths must be treated with separate artifacts, dynamics, and processes. Only when we deliberately decide they are the same thing do we use one instantiation.

8 Cyber Warfare

Cyber warfare is a new domain that has evolved from the older domains. As a warfighting domain, it has roots in the same targeting tasks as kinetic warfare. It has overlaps with information warfare. Even where cyber warfare is focused on kinetic outcomes (like taking control of an automated piece of gear) and not on information dominance, it shares information scarcity properties with information warfare.

The cyber domain is inherently complex. For example, the air domain was brought to warfighting incrementally. New capabilities were added as existing ones were mastered. Eventually we worked toward deconflicting different domains, then integrating them into true combined arms capabilities.

Cyber capabilities, however, have been fielded inside existing warfighting capabilities, not in addition to them. This makes them inherently deeply interrelated with those capabilities. Being fielded inside the capabilities also removes the requirement to master the cyber aspects of the capability before you progress. Most of the cyber action takes place hidden from the operator. They have neither the need nor the opportunity to master how that part of their capability works. This creates the illusion that the cyber part of a warfighting capability is purely technical and its function and security do not depend on the operational context. This is a misconception frequently leveraged by offensive cyber operators. The separation between the operational side of a capability and its cyber components is widened by the niche skills required for cyber operations.

Other operations require niche skills. Consider the operation of a winch on a ship. There is a mechanic for the winch with a set of engineering skills. An electrician with another set of technical skills. And a boatswain with operational skills and experience. While they all have different skills, in operations and maintenance of the winch, they participate in each others' domains. They

gain an understanding of how the other parts work, even if they don't gain the detailed skills the others have. Further, the electrical part of the capability is further removed from the mechanical and operational. So the operator and mechanic gain a closer understanding of each other's domains than they do of the electrician's or of the understanding the electrician gains of their domains.

Visibility and separation are compounded by the rapid pace of change in a cyber capability. Below the operational side, there is usually a deep "tech stack" – set of hardware and software. Consider your smartphone. To make a call you run an application. That application runs in an operating system on a processor. A set of protocols interface between the processor and the cellular capability. The cellular capability connects to a network. For you to make the call, a microphone and a speaker both are interfaced through the processor to the operating system. Operationally we see a phone call.

It is not just the depth of the tech stack that is the challenge. Most parts of that tech stack come from different sources, so they have different maintenance contexts and update schedules. This change process is even less visible to the operator than the tech stack. This lack of visibility and understanding works the other direction as well.

A cyber capability technician has detailed visibility to and knowledge of the tech stack. But they are removed from the operations. While they can understand how the system is operating, they lack visibility into what the system is supposed to be doing. We see this in the news frequently. Someone was hacked. After the hack is discovered, they uncover mountains of logged information tracing the unauthorized activity in the system. And we complain, "If they have those details, why didn't they stop it earlier?"

Consider this example. I have 100MB of file transfers on my personal computer at one o'clock in the morning. I also have 10MB of file transfers over the network to my operating system's update server at noon. Given that you don't have infinite time or resources, which do you investigate? The answer is simple. Without more information, you don't have any idea which is authorized and which, if any, is unauthorized.

If I wasn't on my computer early in the morning, that one is suspicious. If I wasn't even awake, it is more suspicious. But maybe I set up an automated transfer to happen early in the morning so it wouldn't interrupt my use of the computer. None of these criteria are technical, they are operational. Likewise, what if my operating system provider says they didn't do any updates on that day? To even know to verify that, you need to start with the operational context of the timing and intent of the provider, not the technical details of how it happens.

When the operator considers the cyber capability to be separate from operations and that context not to be relevant, there is no information passed to the technicians to enable these decisions. The downloads and updates functioned normally. No indication of a problem on the technical side. This is because the effect of the operation and the intent are two different things.

9 Modeling Cyber Effects

Like the earlier example of the intelligence leadership and operational cells, consumers and providers of cyber capabilities can be thought of as two separate cells operating through a thin interface that is defined by the visibility from one to the other. This approach gives us two basic models for cyber events in the battlespace – modeling what goes on inside the cyber domain (the specific tech stack) and modeling what effects that has outside the cyber domain.

Effects are the more enduring side of the domain to model. They reach back to the earliest origins of information warfare and straddle the kinetic line between hard and mission kill. Effects are outcomes. What drives the outcomes is a set of decisions, which occur in the players' performance space and the mechanisms that implement changes in the simulated battlespace on the tabletop.

A good design approach is to identify the effects that are relevant to the training, analytic, or entertainment objective, then to scope the set of player decisions that lead to those effects. The mechanisms then are sandwiched between the decisions and effects. Of course, no design process is purely top-down from requirements, so the design of the mechanisms will lead to points where you need to decide whether to work hard to implement an effective mechanism or to change the decisions or effects to fit what you have. Since the decisions and effects are the domain of the customer, you need to consider how changes affect them and

whether the new decisions and effects are still appropriate for the intended use.

Considering Irrational Number Line Games' Tactical Hack (TH) tabletop product as an example of cyber effects modeling, we start in the domain of an entertainment game for a constructive tabletop environment and have all the overlaps into other modalities and instantiations discussed in the earlier sections.

Since this wargaming product is intended to add cyber effects to an existing tabletop environment, we should focus the effects on ones that are compatible will a wide range of different tabletop. The combined wargame will operate like the two intelligence management cells, with two independent sets of activities that interact across a minimal interface, as discussed earlier.

The cyber effects model we use is based on an early twentieth century electronic warfare effects model, Meaconing, Intrusion, Jamming, and Interception (MIJI). Meaconing is overriding someone else's signal with one of your own that has a different meaning. We will call this data manipulation for cyber and use it to create battlespace effects like believing an opponent is somewhere that they are not (fog of war) or not knowing the actual status of your units (friction of war). Intrusion is entering into a network without authorization. It is usually for the purpose of enabling further activities. In the cyber domain we can call this pivoting and expect it to increase the effectiveness of other cyber operations. Jamming is rendering an opponent's signal useless to them. In the cyber domain, we split this into denial of service (DOS) and degradation of service (DEG). DOS creates more impact to your opponents' capabilities and DEG has less impact but is generally more difficult to detect, localize, and counter. Interception is passively acquiring information from an opponent. In the cyber domain, we call this exfiltration and it can be done to an opponent's static information resources or traffic on their networks. This creates the same type of battlespace effects that would happen if you gained protected information from an opponent by any other means.

10 Effects

Since TH is focused on a wide range of tabletop events, the battlespace effects are mostly focused on common activities in the non-cyber domain. The two main types activities affected are the military basics of maneuver and fires. For example, a DOS effect may cancel a maneuver or fire order given by an opponent. An exfiltration might force an opponent to reroll a successful attack, since you were forewarned about it. In the maneuver domain, it may require your opponent to declare their next movement with a unit before you take your turn.

If you were designing to integrate with a specific game, you could further tailor the effects, designing more specific changes relative to the processes of the non-cyber side. TH assumes the nocyber game rolls dice to determine combat outcomes. This is a common mechanism, so it is broadly applicable. It is also easy to abstract into "redo" for a combat adjudication that does not include dice, such as drawing a card. Focusing on a specific game where you know the exact combat adjudication mechanism and could apply modifiers to the die roll instead of the more generic reroll mechanism.

Since you would know the different types of unit in the game, you could create differential effects for different unit types. Units that have a heavier reliance on navigation or command and control systems could receive greater penalties to or restrictions on movement. Likewise, specific combat modifiers could be applied for units that have greater or lesser reliance on control systems or third party communications for targeting. Weapons that have guidance after firing might receive the greatest penalties, and could potentially be turned on their owners.

For broad applicability to straightforward tabletop games, the double-blind refereed mechanism described above was not chosen. It places constraints on not just the game, but also the number of people required, the roles, and the physical separation of the boards. If you were designing for a specific wargaming environment where you had control of those, you could use a double-blind or CAX implementation for cyber effects alone or for both cyber and non-cyber effects. Conversely, you might use a classic tabletop environment with the TH approach for lower overhead, smaller, faster, and repetitive events and an alternate approach less often for larger games.

Double-blind with combined effects (using a referee or CAX) has the advantage for cyber of being able to apply cyber and non-cyber effects to units without the player knowing exactly why the performance is not what was expected. Random variation could be interpreted by the

players as the effects of a cyber attack. Or a cyber attack could be written off as "having a bad day". Both of these are realistic effects. If you intent is to incorporate those situations into training, analysis or the enjoyment challenge, the double-blind approach is worth the overhead. If your intent is to focus attention on other parts of the battlespace, that mechanism could actually distract the players and diffuse the effectiveness of the event.

Beyond combat effectiveness and maneuver effects on non-cyber units, TH also has effects that apply to opposing cyber efforts. This creates the ability for a player to weaken an opponent's cyber campaign or to increase the effects of their own cyber operations. There are relatively few of these, as the intent is to focus on the interaction between cyber and non-cyber operations. Can the commanders effectively choose and employ cyber effects that support their strategy? Can they position their forces so the impact of cyber operations on their units is minimized? Increasing the number of cyber-on-cyber effects can change this dynamic and affect the flow of the non cyber side of the game. For a significantly increased effect, sides could have a dedicated cyber commander. And cyber operations could be conducted as a double-blind within a side, creating a friction of war effect between the cyber and non-cyber commanders.

11 Decisions

Effects are chosen prior to the start of the game. During the game, these effects are enabled by cyber resources that are drawn from a deck. The resources are random and the distributions known. During an opponent's turn, players get the option to draw two resources, then decide to discard one face up and play the other face down on one of their effects. When an effect is resourced, it is available for play.

The decision for choice of effects should take into account both desired outcomes and the likelihood of fully resourcing the effect at a useful point in the game. This represents the first decision for the players and is a parallel to the equipping of a cyber order of battle and providing commander's intent. During the game, the choice of which cards to keep and which effects to apply resources to. This represents decision making under constraint. Being able to apply effects must be manage both proactively across time and in response to bounded random draws, that represent when gains are made during the cyber campaign.

As well as provisioning effects, resources can also be used to conduct counter-cyber operations. Resources can be directly applied to inhibiting an opponent's ongoing cyber campaign, with a chance of success. Even with a failed counter-cyber operation, the player gains insight into their opponent's actions that is useful for future cyber-operations. Counter-cyber operations can create valuable advantages for a player, but also consume scares resources. This reflects the operational commander's balancing of effort across his cyber teams.

The effects in TH are fixed in duration for one turn. The design decision to go with a simple mechanism creates low impact to the non-cyber game and represents where the time scale of a turn is roughly equivalent to most emergency corrective actions. If the event is for a more robust environment where additional overhead is acceptable, effect durations can be implemented. Where corrective action by cyber operators is not a part of the objective of the wargame, this should be abstracted to one of many possible stochastic mechanisms. These are fairly low overhead, requiring something simple like counters, die rolls or card draws. A more elaborate mechanism would take into account actions by the affected side's player(s). As well as some additional artifacts as a tracking mechanism, this also increases the decision space by adding actions for all or a subset of the locate, diagnose, correct, test, and restore operations.

12 Variants

The base assumption in TH is straightforward peer cyber operations. This creates a "fair" environment for entertainment or tournament play purposes, but does not represent many real-world situations. The rules provide tailoring options that reflect common real-world conditions. These variations allow representation of specific known situations as well as modification of the cyber side to carry out "what if" analyses for specific conditions of interest.

Like the game design, the tailoring options start with the desired situations to represent then derives mechanisms to alter the process or the artifacts to realize those situations. The variations span different political, military, economic, social, and infrastructure asymmetric conditions. Examples include representing a less cyber

capable force (all the way down to no cyber capability), representing a very capable force under political or social constraint, such as inability to conduct operations against an enemy using civilian infrastructure for command and control.

Designing in the variants creates robustness in TH, and cost changing parts of the mechanism. These decisions were most effective when the variants were considered during the process of designing the mechanism, and would likely be more difficult and result in additional overhead if implemented after fixing the mechanisms. The idea of limiting the resources available to each side to implement one of the real-world conditions directly affects the distribution of resources available. This, in turn, affects the resource requirements for an effect. Developing a modified resource deck (specific card removed) needs to look across all the effects. In an extreme case, certain effects may not be available due to limited resources. In other cases they can be extremely difficult to resource, which may or may not fit the situation that is being implemented.

The simpler method to limit effects is to remove them from the deck or not allow one side to choose them. But limiting diversity of options changes the relative value of the resources. For example, if all the low resource effects are removed, there may be no combination of three effects that you can resource for one game. Conversely, if all the high demand effects for one resource are removed, that resource may not be useful to players after a few are gained.

Design considerations like these emphasize the value of starting with a good understanding of the variants you want to support. Continually redesigning cards and reanalyzing distributions is not only tedious, but increases the opportunities to overlook a significant concern. Usually these concerns can only be fixed by increasing the scope of the rules for the process, which increases overhead during the game and still increases the chance of missing something in analysis and testing of the game.

13 Summary

Cyber operations have characteristics unique to their domain of operations, but also share many characteristics with traditional kinetic and non-kinetic warfighting. A key point in modeling cyber warfare is separating the actions (what cyber operators do) from the effects (what cyber operators accomplish). The effects have the most in common with other warfighting domains. When modeling cyber operations, starting with effects and player decisions that represent the desired situation and outlining the scope of desired variations before identifying mechanisms provides a framework for a robust and efficient design. Another ten pages (easily) could be written on cyber effects modeling, and hundreds (if not thousands) of pages already exist on cyber actions modeling; the scope and depth of potential options emphasizes the value working from the top down, deciding on mechanisms last.

THEME 6: Decision Support and Analysis

Data Farming Services (DFS) for Analysis and Simulation-Based Decision Support

Stephan Seichter
Bundeswehr Office for Defence Planning

Bernt Åkesson
Finnish Defence Research Agency

Marvin Richter, Nikolai Muts
Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS

Abstract

Data Farming is a simulation-based methodology that supports military decision-making throughout the development, analysis, and refinement of Courses of Action (COA). By performing many simulation runs with the exploratory power of high performance computing, a huge variety of alternatives can be explored to allow decision makers improved situational awareness to make more informed and robust decisions in domains like defence planning, operations, training, wargaming and capability development.

NATO Research Task Group MSG-155 provides the structure to establish Data Farming core services for the efficient utilisation of Data Farming as previously documented in MSG-088 "Data Farming in Support of NATO" and MSG-124 "Developing Actionable Data Farming Decision Support for NATO". The developed services are designed as a mesh of microservices as well as an integrated toolset and shall facilitate the use of Data Farming in NATO. DFS provides, inter alia, isolated services as dockerized[17] containers for Model Execution, creating the Design of Experiment, calculating on High Performance Computing clusters like Docker Swarm and for Analysis and Visualization of the results.

This paper introduces and demonstrates Data Farming Services and highlights the technical as well as operational benefits.

1 Introduction

Data Farming is a simulation-based methodology that supports military decision-making in areas such as defence planning, operations planning, warfare development, concept development and experimentation. By performing many simulation runs, a huge variety of alternatives can be explored to allow decision-makers to make more informed and robust decisions. This process allows for the consideration of uncertainties and the discovery of unexpected outcomes. Paradoxically, this method can be used to declutter the immense amount of information known and allow commanders and staff improved situational awareness.

The NATO Science and Technology Organization Research Task Group MSG-155, titled "Data Farming Services (DFS) for Analysis and Simulation-Based Decision Support", acted from September 2017 until September 2020. Eight nations and three NATO bodies participated in the task group.

MSG-155 provides the structure to establish Data Farming core services for the efficient utilization of Data Farming as previously documented in MSG-088 and MSG-124. The realization of the concept is being developed through use cases which are relevant to NATO and nations. Data Farming Services is developed and implemented in NATO MSG-155 by a multi-national team. MSG-155 and DFS are based on the Data Farming basics codified in MSG-088 (Task Group MSG-088, 2014) and the actionable decision support capability developed in MSG-124 (Task Group MSG-124, 2018).

Data Farming Services provide simulation-based decision support by using a web-based microservice architecture. These services are flexible, scalable and interchangeable and allow interoperability and operation in federated environments, which leads to higher resilience. DFS is designed to make data farming available and usable to a wide area of NATO users and applications for more efficient and better decision-making. By using a microservice architecture and relying on container technology, distributed deployment and DevOps are optimally supported. The benefits of DFS

[17] https://www.docker.com/

from both technical and operational perspective are summarized in Figure 1.

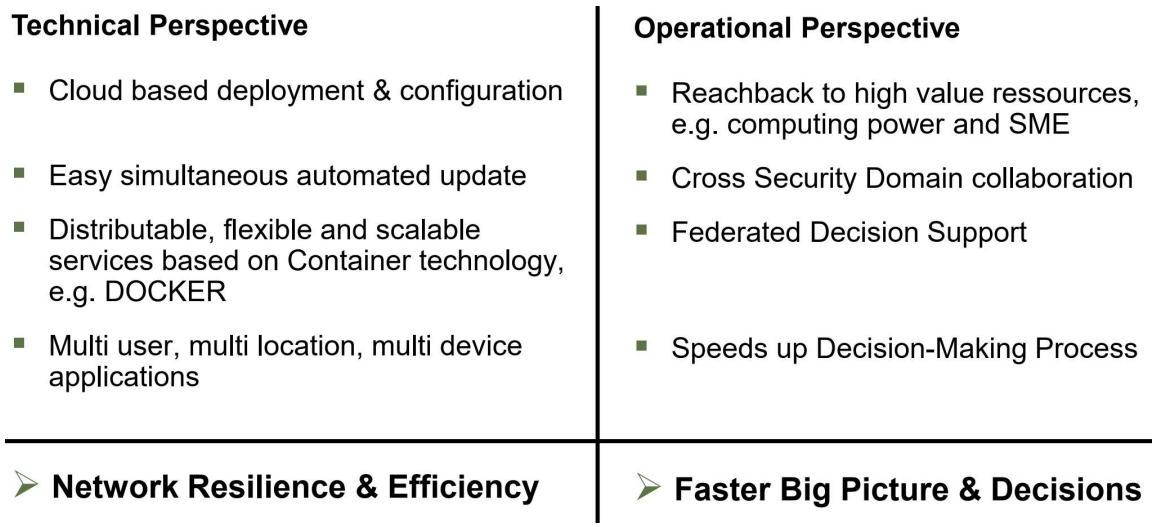


Figure 3: Benefits of Data Farming Services.

The Coalition Warrior Interoperability eXercise (CWIX) is considered to be the ideal testing platform to ensure MSG-155 Data Farming Services architecture to be interoperable and technically compliant with relevant NATO M&S standards, C2 and planning tools. During the CWIX cycles 2018-2020 the DFS architecture was continuously built up and tested, from the web-based Data farming Tool for Operation Planning (DFTOP) analysis service to a functional microservice architecture that finally achieved interoperability proven in a fully Federated Mission Networking (FMN) compliant integration.

MSG-155 performed a proof-of-concept by implementation of two use cases significantly relevant to NATO, namely "Future Ground Combat Operations" and "Cyber Defence". This paper provides an overview of Data Farming Services and demonstrates the utility of DFS on the Cyber Defence use case of MSG-155.

2 Data Farming Services

The basis for the DFS system is the data farming loop of loops, shown in Figure 2. The data farming process consists of six realms (depicted in the grey boxes) and the workflow. The sixth realm is the realm of collaboration underlying the whole process.

As shown in Figure 2, the process can be condensed into the following four main steps. Every step, or a subset of a few steps, is repeatable.

- **CREATE**: The decision maker has a task to investigate a particular situation. This situation is described in the scenario. Questions and conditions (model) to answer these questions are specified. The simulation model output parameters for the analysis and, in particular, the measures of effectiveness (MoE) are determined.
- **DEFINE**: In the second step we select the input (select designated parameters and/or choose the value ranges) which should give answers to our questions.
- **RUN**: Taking the input and the conditions (model) in the previous steps we run the simulation. The simulation will produce the output.
- **ANALYZE & VISUALIZE**: In the last step of the cycle we can observe which input leads to which output.

Based on this new gained knowledge we can make a recommendation to the decision maker. If needed he can start the whole process again.

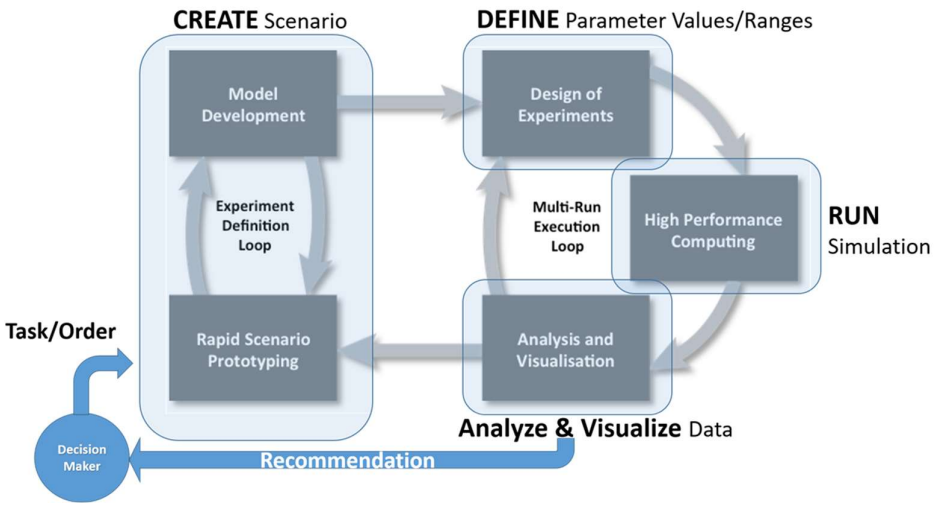


Figure 4: The loop of loops of the data farming process.

To capture and facilitate the workflow and the realms, the loop of loops was translated into services, as shown in Figure 3. Five of the six realms of data farming are captured into services. Model development cannot be supported in DFS, since modeling normally happens in dedicated modeling or simulation tools which are standalone desktop applications. Instead, DFS provides a Model Execution Service, which allows the user to execute and validate the model loaded into DFS. Collaboration, the sixth realm of data farming with overarching importance, is captured by designing distributable web services and a multi-user web application to work with these services.

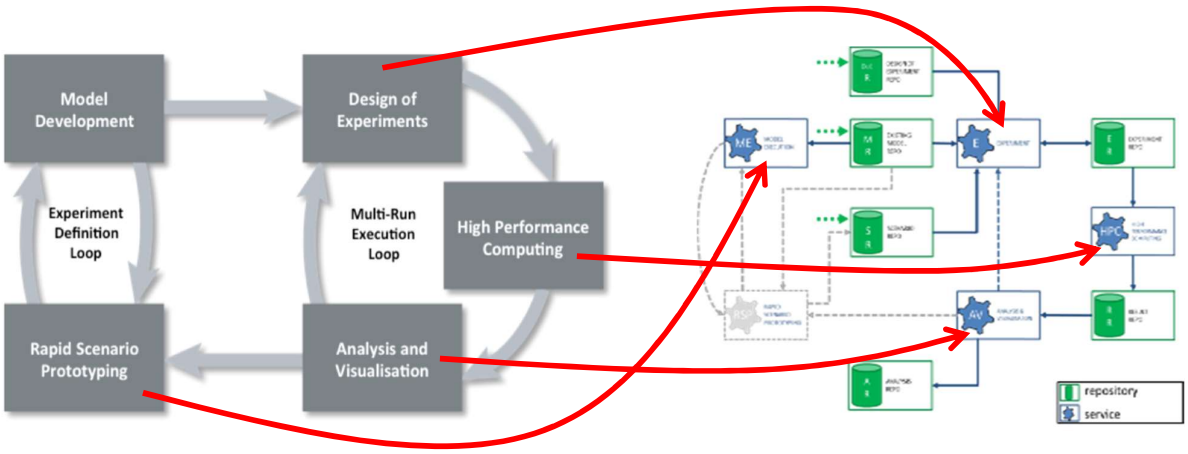


Figure 5: The DFS system puts the loop of loops of the data farming process into a microservice architecture.

These services are presented in Figure 4 in detail (blue) together with the distributed repositories (green), management services and the user interface. Of each service type there can be several implementations or instantiations, all working in a mesh of microservices (Huber et al. 2019). The services corresponding to the four main steps of the process produce data which is saved into repositories. Hence the data of very step can be used in other steps to get more insights.

The services types are as follows (Huber et al., 2019).

- **Model Execution**: Service to execute a model configured with a specific input factor combination.

Currently this service just provides the simulation outputs, but it is envisioned that the visualization of a simulation tool is integrated, such that in the realm of Rapid Scenario Prototyping the dynamics of the simulated model can be verified. This service must be simulation tool specific, since this tool must be integrated into DFS.

- **Experiment**: Service to create the Design of Experiment (DoE) based on the used model and the specified ranges of the model input factors. Methods used for creating the DoE can be for instance Full Factorial, Nearly Orthogonal Latin Hypercube or Nearly Orthogonal Balanced design.
- **High Performance Computing (HPC)**: Service to execute all the simulations defined in the DoE on a high performance computing cluster. This service functions as a simulation tool specific wrapper for a standard HPC software or framework. Similarly to the Model Execution Service it has to be simulation tool specific since it must parse the DFS data objects into simulation tool inputs and also has to store the simulation results in DFS.
- **Analysis and Visualization**: Service to analyze and visualize the results created by the HPC in a web application.
- **Repositories**: Services (green in Figures 3 and 4) to store all DFS data objects. For optimal service distributability the set of DFS data objects are partitioned into six clusters: model, scenario, design of experiment, experiment, result and analysis. These services function as wrappers for SQL-databases.
- **Security**: Management service to handle user authentication and authorization.
- **Configuration**: Management service to provide a central registry of all services. Since the system can be distributed on several servers and there can be multiple service implementations present per type.
- **DFS Portal**: Web application to provide a graphical user interface (GUI) to let the user work with the services, provide workflow guidance and visualization of data.

Every service and repository are wrapped into a Docker container which makes the service or repository independent. Every function or repository is an isolated and independent service.

The services are interchangeable, and there can be multiple instances of the same service type, meaning that the user can select from e.g. multiple different analysis and visualization services.

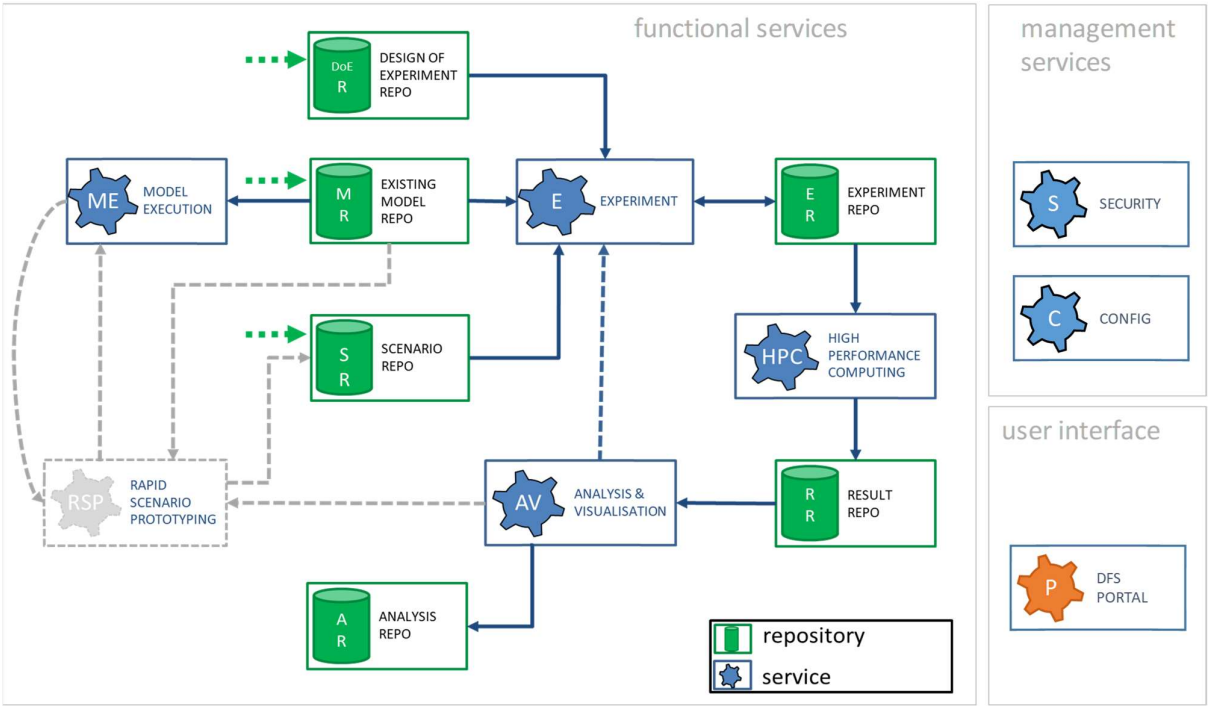


Figure 6: Data Farming Services in detail.

After logging in to the DFS Portal, the user can select the services to work with. After this selection, a new

study can be created or an existing one loaded. After loading a study, the user is presented with a view such as the one in Figure 5. The workflow concept is also illustrated in the figure.

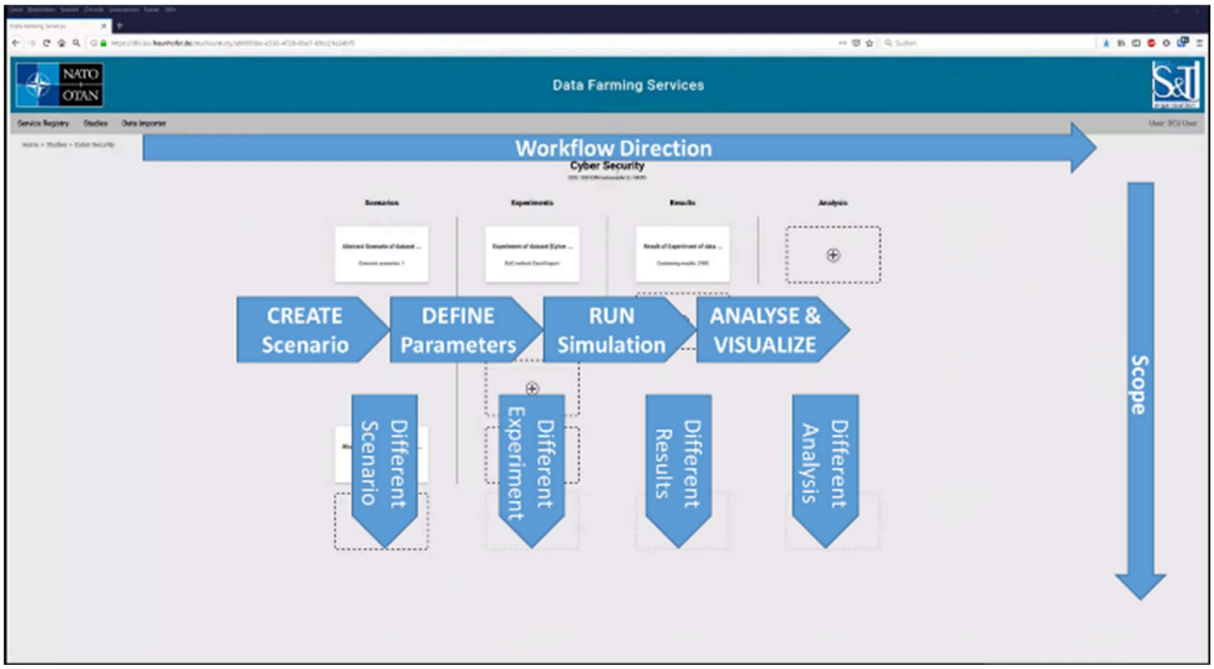


Figure 7: The Study View in the DFS Portal with workflow concept superimposed (based on Huber et al. (2019)).

3 Use Case Demonstration "Optimal Placement of Sensors in a Computer

3.1 - Simulation model and objective

The decision-maker objective of this use case is to investigate how various network monitoring and detection systems should be deployed in order to effectively protect critical services from a wide range of malicious cyber activity.

Based on this objective we start designing a base case scenario. The scenario setting is a mission network with operation-critical infrastructure and services. The task is to find the best Couse of Action (COA) for sensor placement and sensor requirements, in order to achieve the mission.

The Data-farmable Agent-based Cyber Defence Assessment Model (DACDAM) is an agent-based event-stepped Monte Carlo simulation model. DACDAM is a proof-of-concept simulation model, with the purpose to enable exploration of various cyber-attack scenarios, including cyber-attacker capabilities, cyber-defence measures and network architectures. The model is intended to capture some of the dynamic nature of the domain. The first version of DACDAM was developed in MSG-124 (Task Group MSG-124, 2018). The model was selected for this use case and a new version developed in MSG-155.

The general structure of DACDAM is shown in Figure 6. The model's agents represent, among others, cyber attackers, network devices, sensors, services, vulnerabilities and network administration. These agents are capable of operating independently and interacting with each other and the environment.

The input parameters are used to define attacker properties, system characteristics and network properties. Output parameters include detection rate with which cyber attacks are detected, values of the information security metrics, the number of successful attacks, the status of the services and operational metrics.

The scenario is built by setting the input parameter values and defining a network topology. In this use case we use a network topology from the open literature.

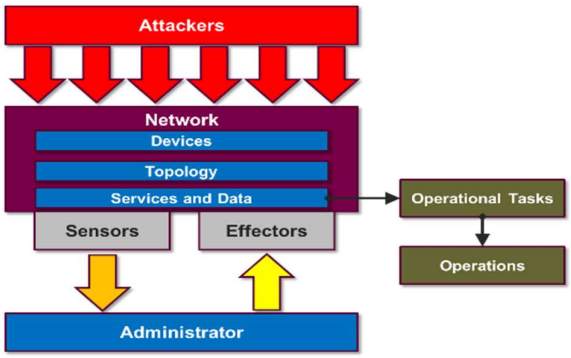


Figure 8: General structure of the Data-farmable Agent-based Cyber Defence Assessment Model (DACDAM).

Figure 7 shows see the network topology in DACDAM's visualization tool.

Cyber attackers launch attacks on the network, in order to cause disruptions or to destroy, manipulate or steal data. The sensors have some probability of detecting an attack step and raise alarms of detected attacks, allowing the administrator to respond and take steps to mitigate their effects.

Sensors can be deployed on different network nodes. Each sensor configuration represents a Course of Action. For this mission, we consider two COAs, termed COA 1, where sensors are evenly distributed over the network, and COA 2, where only the most valuable network nodes are instrumented.

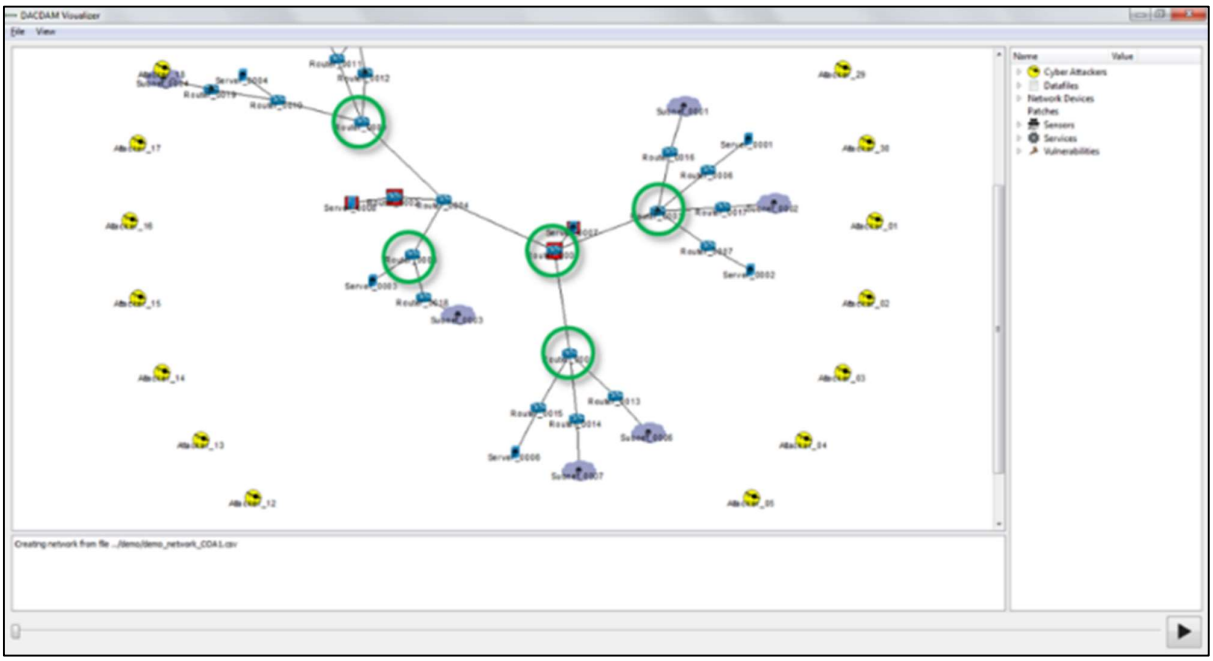


Figure 9: Screenshot of DACDAM's visualization tool, showing the network structure used in the use case. The green rings illustrate a specific deployment of sensors.

With the simulation model ready and the base scenario designed, the data farming process can now proceed using DFS. In the following subsections, we will describe how the four main steps in the data farming process are carried out in DFS.

3.2 - Step 1: Create scenario

We start by providing the information relevant for the situation, mission and the scenario (Figure 8). We select the appropriate simulation model, which for this mission is the cyber defence model DACDAM.

Next, we determine the simulation model output parameters for the analysis and, in particular, the measures of effectiveness (MoE). In this mission, the MoEs are the cyber attack detection rate and the number of successful cyber attacks.

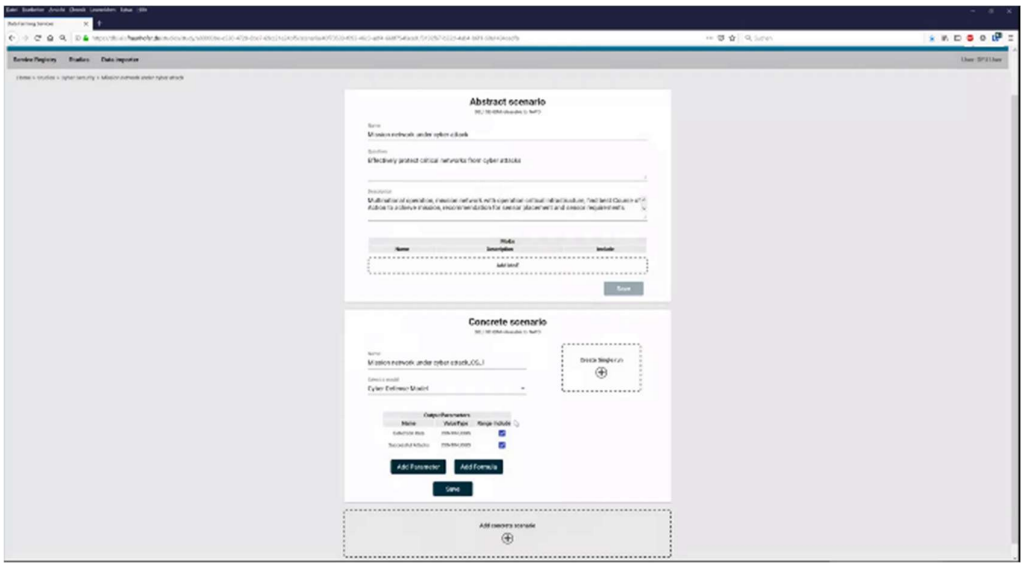


Figure 10: Creating a scenario in DFS.

3.3 - Step 2: Define input parameter values and ranges

In this step we determine the design of experiment method (Figure 9). Here, we simulate all feasible input parameter combinations, having a so-called *full factorial design*. We also appoint a name for the simulation scenario. Next, we define input parameter values and ranges and classify the parameters that matter for the given mission, the so called decision or decisive parameters. For this mission the decision parameters are sensor performance and the COAs. Input parameters that the decision maker cannot influence are classified as noise parameters.

With this the simulation setup is complete. The model has been chosen, as well as the model inputs and outputs, and DFS is now ready for the next step: run the experiment.

3.4 - Step 3: Run the experiment

We select the *high performance computing service* and its location. We then select the *result repository* and location where the simulation results should be written to. DFS runs the experiment and generates the data by clicking on "Run experiment" (Figure 10)

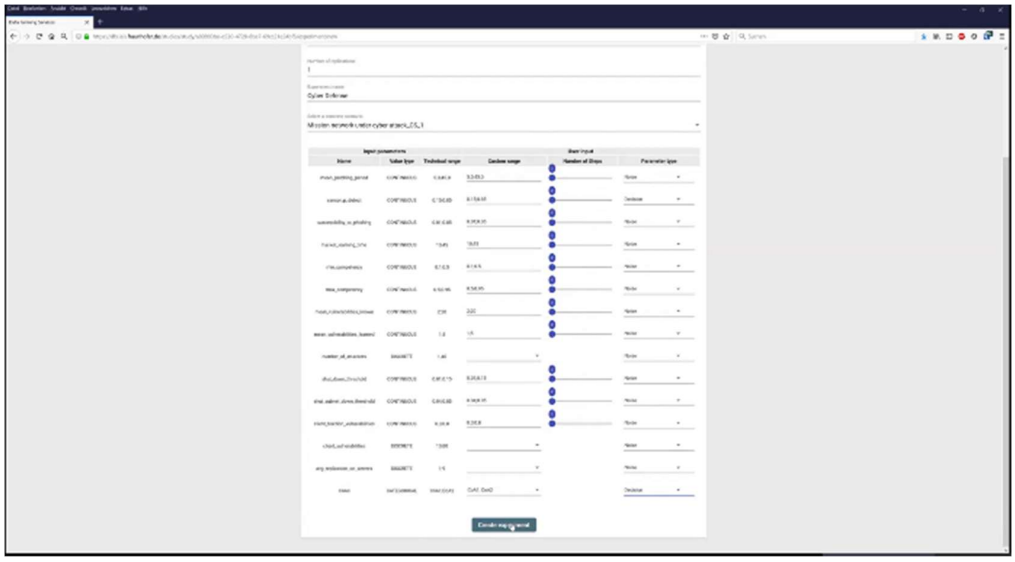


Figure 11: Defining parameter values and ranges for an experiment in DFS.

3.5 - Step 4: Analysis & Visualization

DFS provides many statistical and big data analyses as well as diagrams for data visualization. One such visualization is provided by the *wire diagram*, also known as a parallel coordinates plot, which shows the effect a specific input setting has on the outcome of the mission.

We select two input decisive parameters: the sensor performance and the COAs. As measures of effectiveness for the output we select the detection rate and the number of successful attacks. DFS can now create the diagram (Figure 11).

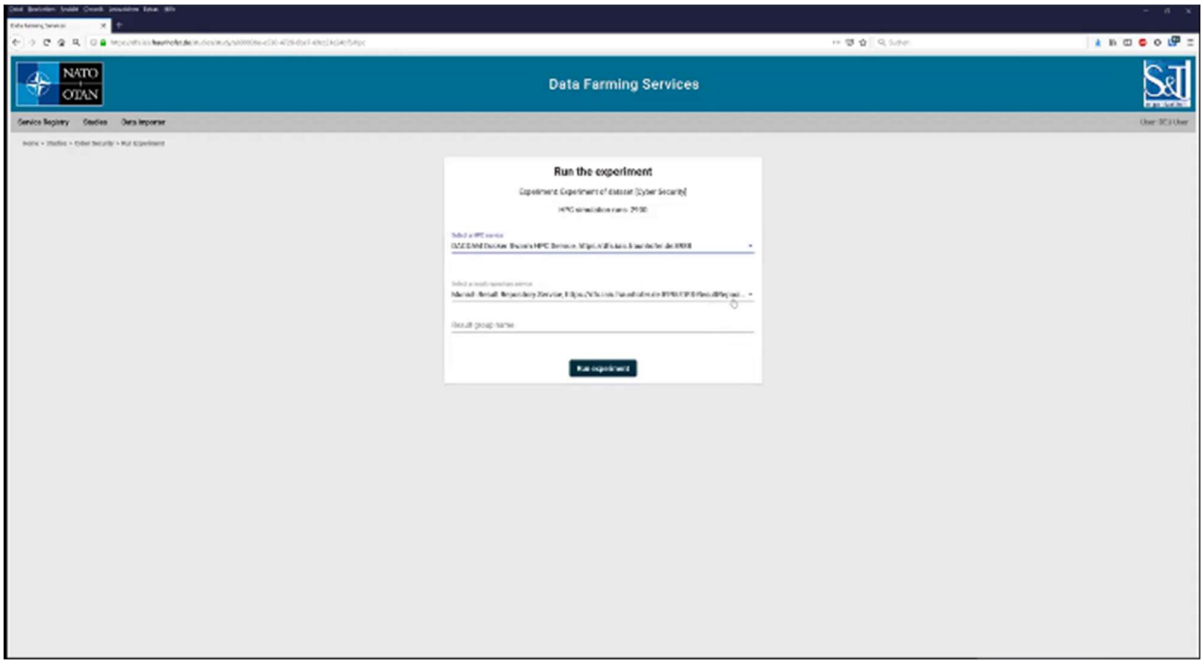


Figure 12: The user can select the high performance computing service and the result repository to use.

In the *wire diagram* each line or wire represents a single simulation run. The vertical axes from left to right are the two decisive parameters, sensor performance and COAs, and the two measures of effectiveness, the cyber attack detection rate are detected and the number of successful attacks.

The detection rate should be high for a successful defensive mission, whereas the number of successful attacks should be minimized. A goal for the mission commander would be a high detection rate and at the same time to allow only very few successful cyber attacks. The plots are interactive and allow us to filter data on the axes, so that the simulation runs belonging to the filtered ranges are then highlighted.

The *wire diagram* shows the effect of the two decisive parameters. More lines go through COA 2 than COA 1, which indicates that COA 2 is more effective. The analysis shows that roughly two thirds of the attacks can be defended with COA 2, compared to one third with COA 1. The sensor performance seems to matter, too. A good sensor performance yields a high detection rate and a low number of successful attacks.

The conclusion for this mission is that the deployment of high performance sensors with COA 2 on crucial network nodes is most effective against expected cyber attacks.

4 Conclusion

Data Farming Services is a cloud-based service architecture that provides scalable, flexible and distributable services that allows a high network resilience and efficiency. As a multi-user and multi-location application it furthermore allows reachback to high-value resources in different networks even with different classifications via cross-domain capabilities and thus enables Federated Decision Support. DFS supports the decision-maker by quickly providing a big picture which allows improved situational awareness and informed decisions.

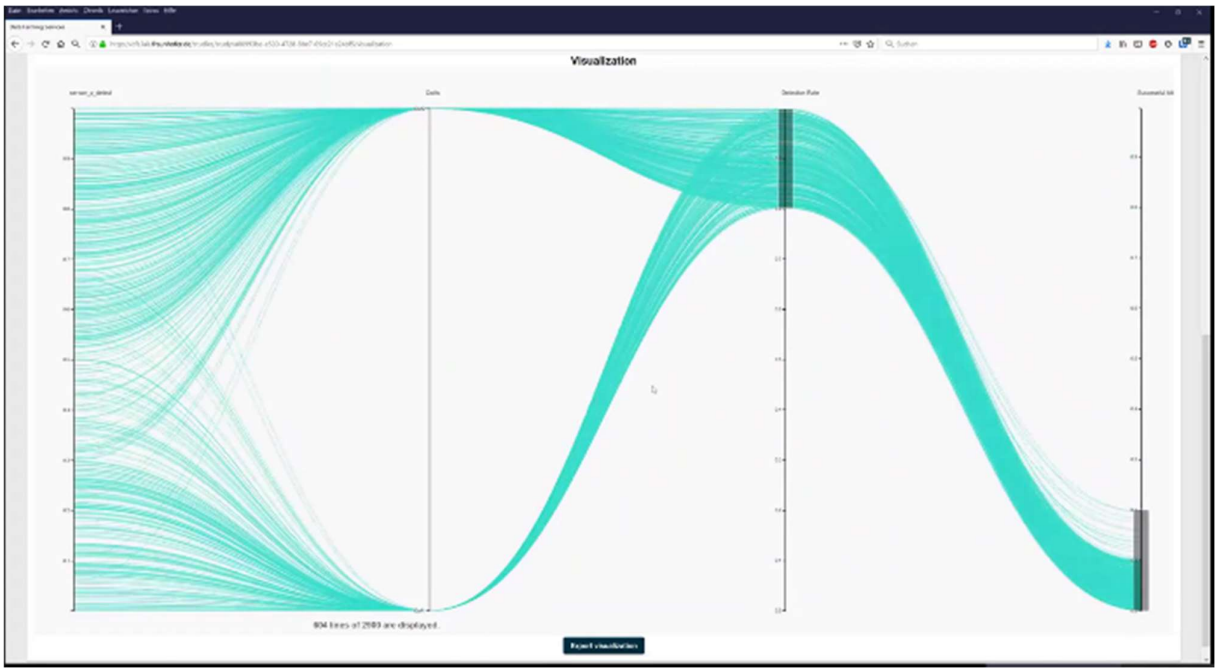


Figure 13: Wire diagram in the DFS Analysis & Visualization service. Simulations with a high attack detection rate (third vertical axis) and low attack success rate (fourth vertical axis) are selected.

References

Task Group MSG-088, "Data farming in support of NATO – final report of task group MSG-088," NATO Science and Technology Organization, Neuilly-sur-Seine, France, STO Technical Report STO-TR-MSG-088, March 2014, ISBN 978-92-837-0205-4. doi: 10.14339/STO-TR-MSG-088.

Task Group MSG-124, "Developing actionable data farming decision support for NATO – final report of MSG-124," NATO Science and Technology Organization, Neuilly-sur-Seine, France, STO Technical Report STO-TRMSG-124, July 2018, ISBN 978-92-837-2151-2. doi: 10.14339/STO-TRMSG-124.

D. Huber, G. Horne, J. Hodicky, D. Kallfass and N. de Reus, "Data farming services: Micro-services for facilitating data farming in NATO," in Proceedings of the 2019 Winter Simulation Conference, N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, Eds. 8-11 December, National Harbor, MD, 2019, pp. 2455–2466.

Collaborative Exercise Planning in an Isolated Environment

Dave Chupick
Calian Group Ltd.
d.chupick@calian.com

This article is a summation of the major points addressed during a presentation to the NATO CA2X2 Forum conducted virtually in September 2020.

Planning and conducting exercises is a complicated business and there are many existing and evolving factors that contribute to this complication. Today's training audience expectations are sophisticated and demanding. There is a need to replicate the complex operational environment including the various joint, interagency, multinational and public operational levels within which operations are conducted. There exists the requirement to build multi-national and multi-disciplinary exercise control teams to create the challenges for political decision makers, commanders and their staffs to learn. There is a need to manage and share an almost limitless accumulation of documentation. The is a need to provide guidance and direction to initiate exercise planning, direction to various teams and OPRs for the management of the exercise life cycle, creation of architectures for federated simulations with command and control systems, identify the architecture needed to integrate exercise communications means with operational communication systems all the while being more efficient with time, funding, personnel and other resources and adapting to the many exercise development changes throughout the exercise process. These are but a sampling of the many factors that are faced by those responsible for exercise design, development and delivery (E3D).

As a tool for exercise planners, each nation has their own E3D process. Though most stages are similar (See Figure 1 as an example), the application of the process varies. In some nations the focus is on going through the process rigidly creating each team, convening each meeting or working group and finding the time to do the resulting work while in other nations, the process is accepted as a tool to guide the teams in the production of the necessary exercise documentation and other essential outputs for a successful exercise. Following any process rigidly, will accrue extra time, effort and therefore costs as each stage, phase and step is executed. Working remotely in an isolated environment demands a review of traditional ways of conducting the E3D process and adding some flexibility in its execution. Only through flexibility will savings be realized.

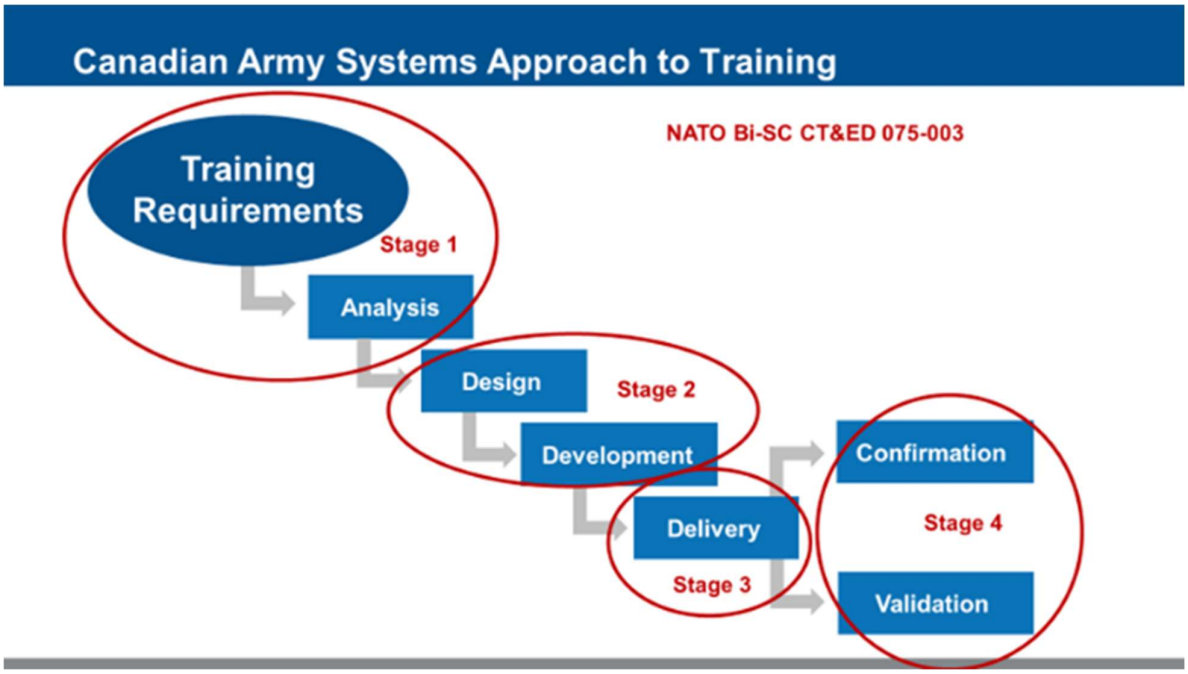


Figure 1 – Canadian and NATO Exercise Processes

Whether members of the team came from different time zones, countries or regions of countries, completing work depicted in Figures 2 and 3, depended on physically gathering the team. Sometimes much needed subject matter expertise was located "in theatre" and exercise planning was stalled until that person could travel to join the team or the team went to theatre on a recce. Security, ease of management and sharing of information were challenges best met by holding physical conferences and working groups. Time, expense (including those hidden costs borne by national delegates), dated technologies and lack of an agile E3D process were the "cost of doing business". In the complex exercise environment is this acceptable and is it the most efficient and effective way of doing business, particularly as we find ourselves longer in an isolated environment?

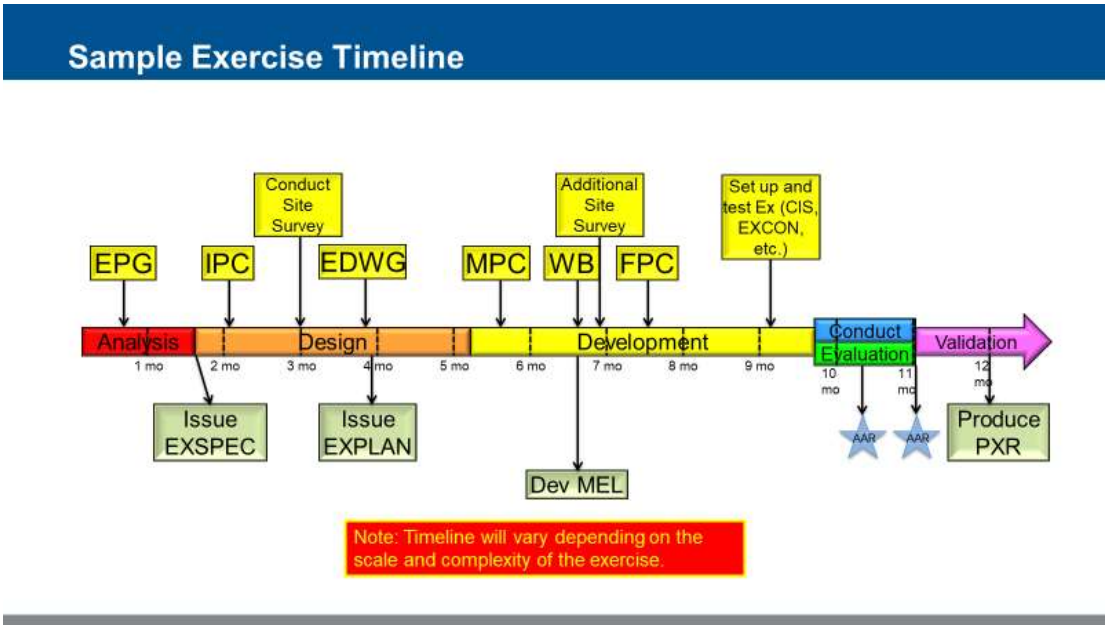


Figure 2 – Sample Exercise Timeline

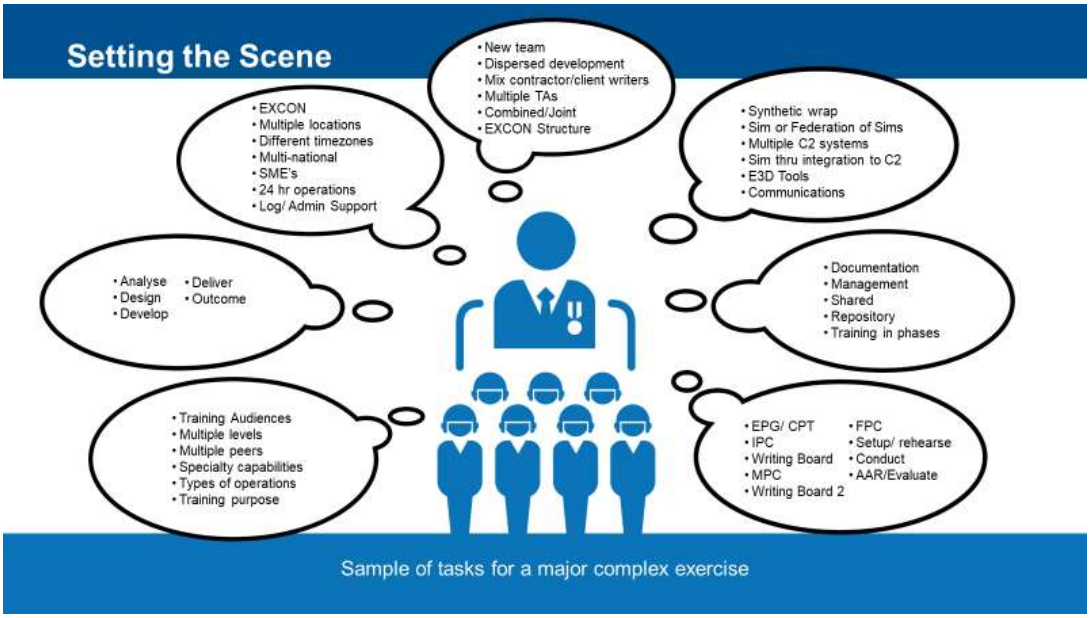


Figure 3 – Sample of tasks for a major complex exercise

COVID 19 isolation has challenged the traditional way in which exercises are planned, developed and delivered. The limitation that teams can no longer gather, cannot be accepted as a reason for not doing the work. Nations and their militaries must get on with the business of training. In many ways, COVID 19 has been a global catalyst for change. It is a global problem and with the extended duration of isolation many organizations operating in other fields of work have adapted and changed their way of doing business. They evolved their tools, processes and technology to allow this adaptation. Early results show that these organizations are realizing cost savings and efficiencies without adversely affecting effectiveness. Arguably the "business of exercise planning" has been slower, less efficient and effective in their adaptation to an isolated environment.

So what has an isolated environment changed? All elements of the stages of the exercise process become more difficult to manage effectively. A need has been identified to look at a number of factors such as our exercise processes with a view to update and streamline them, review technology in use today and investigate options that exist in the commercial world. Lastly, isolation, in this case caused by COVID 19, has allowed some businesses to use technologies and streamlined processes to continue to succeed and in many cases actually thrive. "The business of exercise planning" also needs to learn from industry and needs to consider the opportunities presented by isolation and carry them forward.

So if the task remains to create and manage a complex, challenging and rewarding exercise environment as shown in Figure 4, then we need the tools, perhaps a combination of them, as well as a refresh of our exercise processes in order for our exercise teams to continue to produce exciting, challenging and rewarding training.

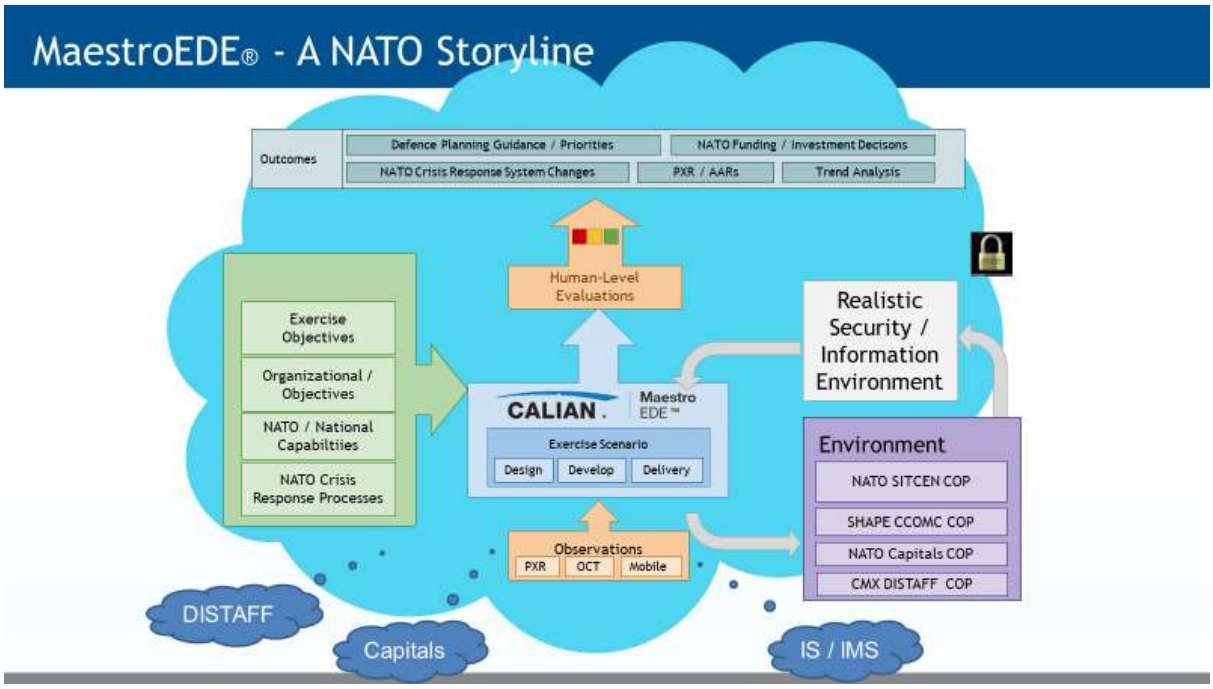


Figure 4 – A sample exercise environment

So, what hasn't an isolated environment changed? As we seek economies in time, costs and personnel effort there remains some constant challenges. There will always be changes to exercise parameters and these variables need to be considered and resolved. Note that some of these come on short notice and take on a too great importance of their own, relative to initial training aims and objectives. There will continue to be more demands placed on exercise planners adjusting for changes to the program, increased training objectives, participants or leveraging one exercise to accomplish outcomes for a wider range of other customers. There are higher expectations that all can be accomplished in less time, using less money and with fewer personnel

resources. And finally, all these items are dealt with in the background that exercises never unfold the same way twice. Success is not achieved by simply dusting off the last iteration. Rather, success is achieved by reviewing what, how and why the last iteration was created, delivered then its relative success assessed before conducting an analysis of how to apply the lessons to the next iteration. So, an isolated environment hasn't changed these demands but rather, is forcing a review of past practices in order to determine how best to meet the challenges and demands today.

Calian Group's solution to these challenges, demands and opportunities is Calian®MaestroEDE® with one view shown at Figure 5.

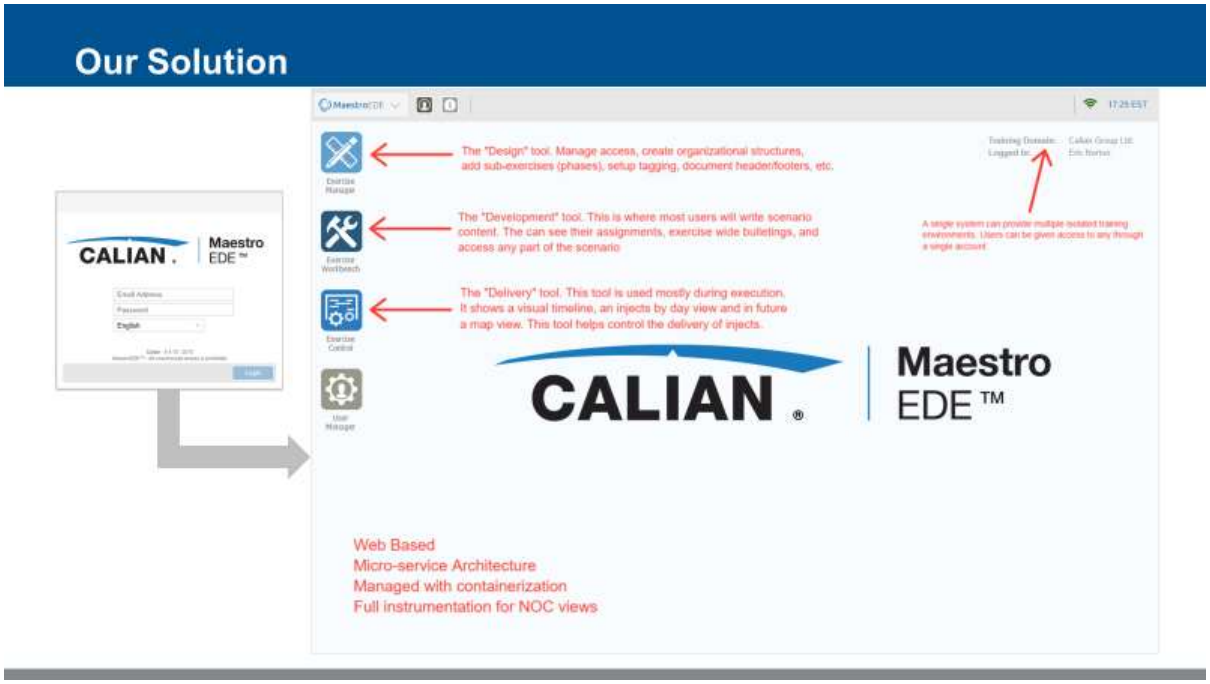


Figure 5 – Calian®MaestroEDE® Environment

It is a second generation E3D tool that was created by Calian exercise planners, to be used by Calian and its customers in the development and delivery of large complex training exercises. We're a training company where successful training outcomes depend in part, on the effectiveness of our tools. MaestroEDE® is the result of a training company who understands training problems and then developing training solutions and not a software company developing training solutions for which the training problem may not be fully appreciated.

MaestroEDE® provides remote users a collaborative and comprehensive workspace that is secure, scalable, adaptable and can function effectively within the dynamic nature of large complex exercises today. It provides user views and functionality that supports and enhances the outcomes of all stages of an E3D process whether it's for NATO, Canada or any other western nation. It is being used now, continues to adapt to user needs and it is already running on open, closed, UNCLASSFIED and CLASSIFIED networks. It has unlimited growth potential and presents an enterprise solution for all exercise needs, exercise planners and those responsible for the associated exercise processes and outcomesExercise conduct is a complicated business given today's training audience expectations and the need to replicate the complex environment within which operations are conducted. Training solutions using a disciplined yet flexible E3D process driving the design, development and delivery is key to exercise success. Technology can assist and a federated toolset (E3D, simulations, C2) overcomes the "effects gap" between scenario development and effects delivered

through simulation/C2 and other real-world interfaces like social media. Calian® MaestroEDE® is a world class example of what industry has to offer. Calian's goal is to set an example through meaningful, high quality training and great people using great tools are fundamental to that goal.

Genomics Enhances the Predictive Power of Machine Learning for Predicting Susceptibility to Heat Illness in Soldiers During Exercise Precise Response 2019

Cathy Boscarino, PhD
Defence Research and Development (DRDC) Canada

Victoria Catterson, PhD.
BioSymetrics Inc., Toronto ON Canada

1 Rationale & Aim

Armed Forces personnel are expected to maintain mission capability under extreme climatic environments, such as the hot summer in Afghanistan (120 °F/49 °C) or in the Middle East where the climate is very hot and dry with temperatures as high as 120 degrees F/49 degrees Celcius and 12% humidity. Unfortunately, extreme hot environmental temperatures along with the personal protective equipment (PPE) required to be worn by soldiers exacerbates the physiological strain associated with work in the heat, otherwise known as exertional heat illness (EHI). Although EHI may be predisposed by external factors, such as climate or clothing, the root cause is internal heat produced during muscular exercise and the pathophysiological characteristics of EHI are heterogeneous. For instance, symptoms can range anywhere from mild (muscular weakness, headaches, fatigue) to dizziness, confusion and fainting and in more severe cases, collapse and death (heat stroke)1.

Consequently, heat illness remains a cause of morbidity and mortality in military personnel during both training missions and deployed operations 2,3, representing a global, year-round health threat. For example, the Canadian Armed Forces reported 194 EHI diagnoses over a 3- year period, of which 30% were considered cases of heat stroke, and in recent years the US military has observed an escalation in the incidence of exertional heat stroke4.. The UK Ministry of Defence (MoD) is taking great measures to reduce the high morbidity associated with heat illness in accordance with their code of practice 'to reduce morbidity from EHI4. Other nations are doing the same; as a means to mitigate heat stress and improve operational performance, France is practicing with heat acclimatization prior to sending out soldiers in hot environments5, while the Australian military is investigating heat strain and risk of heat-related illness above the prescribed limits of the Work

Table3. Measures to mitigate heat illness and/or detect heat susceptibility are critical as commanders are expected to identify soldiers susceptible to heat illness at the tactical level, in order to maintain operational capability. Protecting the soldier against heat-related illness and maintaining operational capability are two competing demands for the commander.

The challenge therein lies in the fact that susceptibility to EHI is considered multifactorial as well could differ between individuals exposed to the same setting. Furthermore, a review of all EHI reported in the British Army found an absence of susceptibility factors in nearly half of reports. The findings highlight two challenges; identifying EHI-prone individuals and identifying factors other than modifiable ones governing susceptibility.

The overall aim of the study sought to investigate susceptibility to developing a heat illness using novel techniques and big datasets from a multinational cohort of soldiers. As a first step, this paper explores whether genomic data has the potential to enhance the predictive power of a model for predicting risk of heat illness.

Lieutenant-Commander (LCdr) Éric Dumas, Exercise Director of Exercise Precise Response 2019, supported the research to take place during the training grounds of Ex PR Response (2019). All Heads of Delegates of the North Atlantic Treaty Organization (NATO) nations were briefed on the protocol and invited to participate. The research protocol was led by Canada and ethical approval was obtained by the Human Research Ethics Protocol (HREC) of Defence Research and Development (DRDC) Canada.

2 Machine Learning With and Without Genomics

Here we present the analysis and predictive learning performed on the data generated through the NATO joint exercise held in Canada in 2019. The data collection occurred during a series of tasks over 3 days in Medicine Hat, Alberta, Canada. In previous years this location was very hot and some participants suffered heat-related illness during the exercise. However, the Wet-Bulb-Globe-Temperature (WBGT) for the week of the tasks was a low of 19 °C/66.2 °F and reached a high of 26 °C/78.8 °C. This environmental temperature is considered a risk level of low or "caution—possible fatigue with prolonged exposure." None of the participants experienced a heat-related illness requiring medical attention.

The ultimate goal of the work was to build a model that could predict who is at risk of heat-related illness, in order to make staffing decisions about future tasks. In order to build a robust model, the first step was to build a representative model using factors which contribute to heat-related illness and data relating to these factors. However, for the purpose of this study, we first wanted to explore which features were relevant for predicting risk. In particular, whether genomic data has the potential to enhance such a risk model. As a result, we built a predictive model using the pre-task datasets both in the absence and presence of genomic data, and compared the features considered important to each model.

The non-genomic model included pre-task data from the following datasets:

☐ History of heat-related illness
☐ Risk factors (e.g. smoker, alcohol use, hypertension)
☐ Demographics (e.g. age, country, years of service)
☐ Physiology (e.g. systolic and diastolic blood pressure, height, body fat mass of limbs)
☐ Baseline bloodwork (e.g. white blood cell count, platelets, cortisol)

This gave 102 features for 45 participants.

In addition, to gain insight into genetic factors of heat-related illness, microRNA-sequencing was performed on each participant. Sequencing was performed on Illumina HiSeq 2500 by TCAG (Hospital for Sick Children, Toronto). The raw sequencing data was first pre-processed by trimming sequencing adapters using *cutadapt*. The trimmed reads were then aligned to the human reference genome (hg38) using *bowtie 2* with sensitive local alignment option enabled. Expression quantification was then carried out by counting the number of reads mapped to known microRNA transcripts from miRBase (http://www.mirbase.org/) using *BEDtools*. This resulted in a dataset of 2872 miRNA features from each sample. Of these, 1264 contained only 0 for all participants and were removed from the dataset, leaving 1608 features. Finally, the counts were normalized per person to give percentages instead of total counts.

This process gave 1608 features for 39 participants, but unfortunately only 38 are part of the cohort of 45 with non-genomic data. Therefore, the combined dataset that includes all factors above plus genomics is 1710 features for 38 participants.

The ML workflow contained two stages: feature selection and predictive modelling. This paper focuses on the feature selection stage, in order to explore whether genomic features are considered relevant to risk of heat-related illness. The methodology for the non-genomic dataset is:

1. 5000 times over, select 45 features from the total set of 102. This is the "selected" set.
2. 10 times over, randomly split the selected dataset (45 participants x 45 features) into a train and test set, where 85% of samples are used for training and 15% for test.
3. Train a Logistic Regression model using L1 regularization. If it converges within 1000 iterations, keep a record of all the features given a non-zero coefficient by the model. This is the "chosen" set.
4. After all 5000 iteration, calculate for each feature the ratio of the number of times it was chosen to the number of times it was selected. If this ratio is 50% or higher, include it in the feature vector used for predictive modelling.

Since the addition of the genomic features increases the number of possible features by an order of magnitude, the number of permutations for the genomic dataset was increased from 5,000 to 20,000. All other steps of the methodology were the same, and as before, 45 features were randomly sampled from the set. The results are presented below.

3 Results of feature selection

The features selected in the presence and absence of genomic data were compared. When no genomic features were available, 72 of the 102 features were chosen at least once. Of these, 35 features had a ratio of times chosen above 0.5 (i.e. over 50% of the time when the feature was sampled and presented to a model, it was chosen for inclusion). In contrast, when genomic features are present in addition to all previous features, 117 of the 1710 features were chosen at least once, and 82 of these features are chosen over 50% of the time.

	Without genomics	With genomics
Total number of features	102	1710
Features chosen at least once	72	117
Features chosen over 50% of the time	35	82
Features in the chosen set unique to this dataset	0	45
Features in the 50% set unique to this dataset	0	47

Interestingly, while these numbers are higher than for the non-genomic dataset, they are proportionally much lower than the size of the feature space available. That is, 70% of the non-genomic features were found to contain information about heat-related illness (72/102), but only 6% of the total set of features were similarly found to contain information (117/1710). This may be expected, since the majority of genes will have no impact on an individual's susceptibility to heat-related illness. Therefore this is a good confirmation that the workflow is truly sifting signal from noise, and not simply producing random results.

All of the features chosen from the non-genomic dataset are also chosen when genomic features are present, as shown by there being 0 unique features in the "Without genomics" version of the dataset in the Table above. This is notable: it is sometimes the case that the addition of a new type of feature displaces others. The fact that no non-genomic features are displaced by genomic features indicates that genomics purely adds information, and presents a different facet of information about the participants' susceptibility to heat-related illness. In contrast, there are over 40 features unique to the genomic dataset when looking at all features chosen at least once, or all features chosen over 50% of the time they are available. Below, we investigate these in more detail.

4 Features unique to the genomics dataset

The focus of this analysis is the high-information content features, as determined by being chosen over 50% of the time when presented to a model. This feature vector is 35 long when there are no genomics features present, and 82 long when genomics are present. All 35 in the first vector are included in the second, giving 47 features that are uniquely present in the vector chosen in the presence of genomics. However, not all of the additional features are genomic features. 16 are from miRNA, while 31 are not.

The presence of these non-genomic attributes is very interesting. It raises the possibility that these genomic features are not directly related to heat-related illness, but instead illuminate an interpretation of one or more non-genomic features. For example, Skeletal Muscle Mass (SMM) is one of the new non-genomic features found to be relevant to the model. One or more of the genomic features may relate to SMM susceptibility in the participant, and therefore indirectly to heat-related illness.

5 Enrichment analysis of the miRNA features

The list of miRNA features found to be important to the model were run through the target enrichment tool Mienturnet. The goal is to identify miRNA-targeted genes that may be enriched for biologically relevant functions or pathways. In total, 970 genes were found, 718 of which had an FDR below a cut-off of 0.05.

Next, GO enrichment was calculated. As is often the case with GO enrichment many terms resulted, but among them were:

☐ **GO:0034605** "cellular response to heat" with an FDR of 0.00702, and
☐ **GO:0009408** "response to heat" with an FDR of 0.0179.

The list of target genes in the original dataset annotated with "response to heat", and their corresponding related miRNAs, is:

☐ **COX2** (PTGS2): hsa-miR-26b-5p, hsa-miR-21-5p
☐ **ATP2A2**: hsa-miR-26b-5p hsa-let-7f-5p hsa-miR-30d-5p
☐ **IGF1**: hsa-let-7i-5p hsa-miR-26b-5p hsa-miR-26a-5p
☐ **MAP2K7**: hsa-let-7i-5p hsa-let-7g-5p hsa-let-7f-5p
☐ **IL1A**: hsa-miR-30d-5p hsa-miR-122-5p
☐ **LYN**: hsa-let-7i-5p hsa-let-7g-5p hsa-let-7f-5p hsa-miR-122-5p
☐ **HSPD1**: hsa-miR-26b-5p hsa-miR-26a-5p
☐ **THBS1**: hsa-let-7g-5p hsa-let-7f-5p hsa-let-7i-5p
☐ **SUMO1**: hsa-miR-122-5p hsa-let-7i-5p hsa-let-7g-5p hsa-let-7f-5p
☐ **HMOX1**: hsa-miR-122-5p hsa-miR-26b-5p
☐ **POLR2D**: hsa-let-7f-5p hsa-let-7g-5p hsa-let-7i-5p hsa-miR-122-5p
☐ **STUB1**: hsa-miR-26b-5p hsa-miR-21-5p

Finally, it should be stressed that these findings are preliminary and indicative, rather than robust. It was intended that the small number of participants in the 2019 exercise could be used to design and test a workflow, which if successful would build a case for further data collection in later years. The results here confirm that the presence of genomic features makes a difference to the final result, and therefore it is worthwhile to continue pursuing this avenue of research. In particular, mRNA expression analysis of the above target genes pre- and post-task will be necessary in order to confirm that they are indeed regulated by the miRNAs of interest..

6 Conclusion

In total, 2872 features were derived from the miRNA data. Of these, only 1608 contained information for this cohort of participants (all others were 0 for all participants). In combination with the 102 non-genomic features, this gave a total feature set of 1710 features. Of these, 117 features were found to contain some information about heat-related illness, as they were selected by at least one model. This contrasts with 72 features in the absence of genomics data.

Setting a feature selection ratio cut-off of 0.5 (chosen 50% or more of the time when presented to a model) gave a vector of 82 features, in comparison to a vector of 35 features when genomic data was absent. All 35 of the non-genomic features were also chosen when genomic data was present. In addition, 16 features from miRNA and 31 non-genomic features are chosen ONLY when genomic data is present. This is notable for two reasons:

☐ None of the non-genomic features are displaced by genomic data, meaning that both the genomic and non-genomic features are relevant to predicting heat-related illness.
☐ There appears to be some mutual information between the unique genomic and unique non-genomic features, i.e. the unique 33 non-genomic features only become important in the presence of information contained in the miRNA.

Enrichment analysis of the 16 miRNA features found some evidence to link them to heat-related illness, notably through enrichment of the GO term "response to heat".

In short, the presence of genomic data makes a difference to the features selected, and is therefore relevant to the prediction of heat-related illness.

7 References

[1] Gauer, R. and Meyers, B. K. (2019), Heat-Related Illnesses. *American Family Physician*. 99:482-9.

[2] Moore, A. C., et al. (2016), Risk factors for heat illness among British soldiers in the hot Collective Training Environment. *Journal of the Royal Army Medical Corps.* 162:434-9.

[3] Hunt, A. P., Billing, D. C., Patterson, M.J., and Caldwell, J. N. (2016), Heat strain during military training activities: The dilemma of balancing force protection and operational capability. *Temperature*. 3:307-17.

[4] Stacey MJ, Parsons IT, Woods DR, Taylor PN, Ross D, S JB. Susceptibility to exertional heat illness and hospitalisation risk in UK military personnel. BMJ open sport & exercise medicine 2015;1:e000055.

[5] Charlot K, Tardo-Dino PE, Buchet JF, et al. Short-Term, Low-Volume Training Improves Heat Acclimatization in an Operational Context. Frontiers in physiology 2017;8:419.

Radical Transparency

Lilly Madigan
User Experience Research
e: lilly@improbable.us

Abstract

To fully harness the power of models and simulation for decision support and analysis, vital information must reach the right hands, at the right time, in the right way. This maxim holds not only for model and simulation capability development, but for user experience and customer management as well. To achieve project success, user-centered research and design must inform the requirements of internal engineering, applied science, and product management teams, as well as the external customer stakeholder teams.

When modelling and simulation software is often hard to use, difficult to learn, and expensive to maintain, it is imperative that vendors build trust with their customers by practicing radical transparency – being forthcoming with each other about product capabilities, and in turn, user needs.

Based on examples from projects completed for US DoD and UK MoD customers, this paper discusses how to build trust through user-centered design activities which enable teams to practice radical transparency.

1 Introduction

There was a buzz of activity in the laboratory. The workshop, piled high with machinery and hardware, acted as a center of innovation activity—the seed place for many of the technologies that would aid an elite group of service members on future operational assignments. I sat on a folding chair in the corner, surrounded by those same service members. Over the course of a series of user interviews, these elite warriors shared details about their mission and training environment, the constraints in which they were expected to perform, the bio-physical effects of combat, and the somewhat embarrassing ways in which one specific product —a "military" retrofitted wetsuit— fell short. Indeed, in a colorful anecdote, one service member revealed how the suit's zipper was not long enough for practical use. In a time of need, when nature called, it was completely useless, leading to a very messy and unsavory ending. While these military operators

may have used different terminology, what they had described was capability development that lacked the key component—the end user. In short, the service members had suffered from the absence of *radical transparency*—when capabilities are built for the user, not with them. Radical transparency incorporates user-centered design activities during the capability development process, thereby establishing trust between the vendor the customer or end-user. This article seeks to instill some of the best-practices that ensure radical-transparency in user-centered design—qualitative user interviews, customer journey maps, and usability tests—in an attempt to facilitate greater trust between the government and the commercial technology sector.

2 User-Centered Design

User-centered design focuses on designing solutions "with users and not for them".[18] This approach ensures that the contract and requirements set out by business development and program managers accurately reflect the needs of the warfighter who will use the product. There are assumptions and hypothesis baked into the start of every project, but it's necessary to validate, correct, or discover new requirements. By seeking out the people who will use the product and through a series of methodological approaches such as qualitative user interviews, vendors get aligned on what might meet customer needs.

Designers support these activities through an iterative process of user research, content strategy, and product design, allowing them to jointly solve problems —with users and meaningfully impact product development. In the research and discovery phases of product development, designers practice relentless humility.[19] It is the notion that one must set ego aside and acknowledge that the user, designer, and developer are experts in their own fields, not all fields. When designers develop empathy for the user and understand their pain points, customers can trust that vendors have their best interests in mind, and speak candidly about their problems, knowing they're working together towards a meaningful solution.

3 Qualitative User Interviews

User researchers commonly employ qualitative user interviews, whereby a researcher sits down with a user to discuss their role, workflows, and goals. The interviewer's goal is to remain unbiased and allow the participant to do 80% of the talking. Some of the best interviews start with one theme in mind, like simulation result metrics, and end with three new user groups to talk to, and access to additional platforms and systems to explore next. To be clear, the end user is usually not the highest-paid or highest-ranking person in the room, nor the person who wrote contract requirements. The end user is often the enlisted service member, the analysts, or the operators. Securing permission to speak with these users requires trust from the commander, the program manager, or the SES. These conversations uncover the model and simulation configurations that matter most to the end users, saving modeling and engineering teams time and resources building complexity into the model that no one wants or needs.

4 Direct Observation

Understanding the entire workflow of the user includes the physical spaces in which the model and simulation capability is intended for use. This can have very practical implications on user interface. Start by going where the work is and physically observing the user's workspaces, processes, and environment, whether in person or over video conference. While the vendor development team works in a brightly lit downtown office, end users might be in forward operating bases, with limited bandwidth and computing power, often battling the elements, among other things. One user's work environment might be appropriate for a dark mode themed user interface, another – not. Direct observation is a powerful indicator that radical transparency has been achieved.

5 Customer Journey Map

Information gained through hours of interviews and pages of notes, can then be organized and shared as a customer journey map. Tactically, customer journey maps are a great way to facilitate understanding across the development team as they allow the team to see

[18] This is one of the core values of the United States Digital Service. For more information, see www.usds.gov/mission

[19] This is a core value at Improbable. For more information visit improbable.io/careers/life-at-improbable

beyond the doctrine of 'the way the process is supposed to work' and get to the reality of the user's situation. In order to know if product development is going in the right direction, it's important to know what the user is experiencing right now. The map lists the steps a user takes in the process, as well as the touchpoints with systems, technology, and humans, even including workarounds and dead ends. Importantly, it layers in real human interaction and sentiment at each step.

On one project, we read hundreds of pages of military doctrine that told us exactly what steps to take and when. But what it couldn't do was explain the factors that made accomplishing these steps difficult or impossible for the users, or how models and simulations could support those steps. A Major in the UK Ministry of Defence once said, "The requirements are a great start, but they're static. They don't tell the whole picture." By understanding a customer's goal-oriented journey, one can begin to identify gaps in the requirements. A requirement might say to model weather. But what does that mean for a user's workflow? What parameters or configurations are important to the user and need to be included in the model? User research helps identify how much weather effects a warfighter's ability to do their job, and that it's necessary for a weather model to interact with other models for them to accomplish their goal.

Developing the type of models required to enhance decision support and analysis takes time, expertise, and access to data. User research allows scientists and modelers to continue to do what they do best, by discovering the areas of interest, data types, configurations and parameters, as well as the desired human interaction with the models that works for the users. It also uncovers user needs that weren't in the original requirements.

6 Uncover new requirements

Over the last year in the model and simulation industry, I've learned that users of all kinds desire to trust the models and insights gained through their interaction within a simulation. Decision makers who may not be experts in our chosen fields, need to be able to act on the results of the simulation, and ultimately take full responsibility for those actions. Model design documents are one way to build that trust. Written by subject matter experts on the modeling or applied science teams, these documents are the epitome of radical transparency. They contain how the model was built, with what data, and fit for what purpose, in plain language. It is important to be explicit about the scenarios in which a model should and should not be used, the types of questions it can analyze, and it's use cases. This transparency doesn't restrict the user, but empowers the decision making and analysis they need to accomplish.

Designers help identify the right place in the user's decision-making process for each piece of information, giving the user the right information at the right time. They are strategic about information architecture. For some users, the entire design document might be appropriate as a downloadable attachment, or in a section of the application for reference materials. While for other users, the same complex concepts are better communicated in other parts of the workflow, as small pieces that could be broken out and summarized with icons or tool tips. The modelling team can be radically transparent with their models, knowing that the content strategy was designed with the user.

Simulation provides insights and analysis of a scenario, not the answer to a question. This distinction is important for decision makers to internalize as we work together to apply models to multi-domain scenarios in a synthetic environment. Models and simulations, for decision making and analysis, enhances the doctrine and allows humans to do what they do best by allowing computers to do what computers do best. The responsibility falls mutually on the internal development team to practice radical transparency. By first having healthy communication with each other, the team can have healthy communication with the warfighter.

7 Product Development is a Dialog

During product development, a Major in the UK MoD said so succinctly that product development is a dialogue. Meaning that when the dust settles and the contract is signed, two or more groups must come together in constant alignment to build a solution together. When the vendor team of modelers and engineers, project managers and designers, are on the same page about their capability and product direction, productive conversations can happen with the customer and warfighter. It's hypocritical to expect a military unit to come to the design table with honesty about their process flaws, or project hopes, while the contractor

develops in isolation, hiding flaws or setbacks, only showing an outward rosy picture.

One way to achieve this alignment is through a service design blueprint. After identifying the pain points and areas of opportunity within the existing process with the customer, the vendor team proposes a way forward, mapping out the service they'll provide to users. The complex modeling and simulation systems, and infrastructure, are the foundation for the human interaction layers. A service design blueprint lays out the user interaction touchpoints and even describes the length of time it takes an operator set up a simulation, or the amount of faster than real time the simulation aims to achieve. After alignment is gained there, individual product features and themes are explored, wireframed, and prototyped.

Another way to increase dialog is to co-create with users. A Lieutenant Colonel in the UK MoD once said, "We were really skeptical, but when you've got Lilly making us draw things on a white board, it was hugely powerful." They were describing a customized design sprint, where the product development team facilitated a series of half-day workshops on specific topics, with a small group of active-duty service members. Each workshop focused on one product theme at a time, like model provenance. After a detailed discussion, users took to the dry erase board, and illustrated front-end features that could solve the user needs. This is where a typical hand off between design disciplines can occur. Folks who specialize in product design or visual user interface design can build cohesive prototypes of these wireframe sketches to get back in front of users for validation and feedback.

It's radical to remove the black box from models and simulation technology, but that builds trust with the customer. Because of this trust, users can practice radical transparency about their needs, paint points, and problems when participating in the capability development process. Whether that's building out a single model, a complex ontology of models, physical hardware, or enterprise level software, the principles of user-centered design still apply.

8 User Feedback

Getting user feedback during development can be frightening. For folks who always need to know how every hour will be spent, and the specifications of every feature defined up front, this can be daunting, but this isn't waterfall. So, how do we know we're on the right track? How do we know the simulation results are helpful? Early system prototypes can be on paper or dry erase boards, while others can be in robust design software with clickable interaction or animation. Either one you choose has tradeoffs, but the point is that it's not a long time between the prototypes and user feedback, and the expectation is set with the customer that the prototype isn't final, or perfect. Failing fast here in the design is cheaper than failing after production. A USMC Maj once said, "We changed their environment. We changed their support and gave them designers to work with. It made all the difference in the world."[20] They were referring to work completed by a cross-functional team of engineers from across the US military that built a capability to disrupts drone attacks. Not only was this problem space new, but these folks had also never worked with a designer before, someone who was able to tactically bring the warfighter into the software development lifecycle. Through usability testing in the field with end users, the team experienced the radical transparency of users and uncovered major assumptions about the needs of the warfighter. As a result, the team drastically changed the user interaction of the product to meet those needs. Without that compelling feedback directly from the end user, it wouldn't seem like common sense to throw out an entire user interface and replace it with a single button, but after watching the soldiers try to use the product, and learning about the challenges of the environment, it was the only sensible way forward.

Spending a few hours with users every couple of weeks is hugely valuable to product development as a whole, and to the individual modelers, engineers, and scientists on the team. An applied scientist once said, "the model might return the number 47, but that alone is unhelpful." The simulation results must make sense to the operator. User-centered design validates the team's work through iterative feedback. The product development team can see their work in the hands of user and hear candid feedback, and the customer team can see how the

[20] https://www.wired.com/story/pentagon-dream-team-tech-savvy-soldiers/

vendor acted on their feedback. Be careful here that usability testing is not invalidated if the moderator biases the users. Asking, "Tell me what you think" elicits a very different response than "You like this… don't you?". The former allows the user to direct the conversation to what matters most to them, while the latter limits the conversation and biases the user.

A Major in the UK MoD once said, "A constant feedback loop ensures the right product is delivered, not any product." They were talking about the kind of constant alignment between users and developers that happens with iterative usability testing. The vendor can show the warfighter how they've listened, took the feedback seriously, and set their ego aside. You can show how you brought groundbreaking options to the table, and provided helpful decision making and analysis tools. It also allows you to fail early. Some of the best feedback my team received from an operator is "What in the world am I looking at!?" followed by a list of non-value add features and content. We were able to make changes in a quick, low-cost way that kept our project on track for delivering a first of its kind model and simulation capability.

9 Conclusion

In conclusion, as decision support and analysis tools are built with cutting edge model and simulation technology, think about who will use it, get to know the warfighter, and be forthcoming with your own team's capabilities, this will endow your projects with *radical transparency.*

The NATO Modelling & Simulation Specialist Team MSG-189

Introduction

The MSG-189 Specialist Team, "AI augmented immersive simulation in Training and Decision Making Course of Actions Analysis", of the NATO Modelling and Simulation Group (NMSG), started its activities in July 2020 to study the state-of-the-art experiences and knowledge regarding several emerging technologies.

Artificial Intelligence (AI), in particular machine learning or behavioral modelling in synthetic agents, Extended Reality (Augmented reality, Virtual reality and Mixed reality), Big Data, with reference to acquisition and storage (Data Farming) or exploitation (Data Analysis), and Advanced Simulation Architectures (Live- Virtual-Constructive and MSaaS) are among them.

The aim of the ST is to harmonize this knowledge towards a framework for an innovative simulation environment in which these technologies could play a decisive role.

The study will produce a final report that summarizes the envisioned integrated simulation environment, describes the possible use cases to validate the environment, identifies existing gaps (either in technology or in operational procedures) and proposes possible developments and further steps.

The report will be the starting point for follow-on NMSG activities that will further develop the concepts and create prototypes / demonstrators for the identified use cases.

In summary, the NATO MSG-189 ST is aiming at achieving the following goals:

- Identify and describe the state-of-the-art technologies regarding several emerging technologies: AI, Extended Reality, Big Data and advanced Simulation Architectures (LVC and MSaaS)
- Specify an innovative framework/architecture acting as a reference model for an open, interoperable and extensible simulation environment where those technologies play a decisive role

One of the sideline objectives of the MSG-189 ST is related to the dissemination of the work that the Group is carrying on through the participation to specific events. In particular just few months after the starting of the ST activities, NATO CA2X2 Forum was the first opportunity given to the MSG-189 community to gather info and suggestions from the broad audience that an event like that can attract and, at the same time, disseminate the goals that the ST MSG-189 wants to achieve.

NATO CA2X2 Forum MSG-189 Sessions

The MSG-189 Specialist Team conducted a Workshop on "Emerging Simulation Technologies for Decision Making Support in the 21st Century". The event provided a virtual venue where the participants had the opportunity to engage with the NMSG Community in addressing state-of-the-art technologies that are considered relevant for the very challenging area of the Decision Making process.

The goal set by the MSG-189 Specialist Team is to identify the most promising Emerging Technologies that are to be integrated in a common simulation framework/architecture that will support the current and possibly future Decision Making process.

The subject of the Workshop was addressed in two sequential sessions. The two sessions filled with papers and presentations, among those submitted, that had a strong impact on the goals set by the MSG-189 ST. In particular, the second session did also show some possible drawbacks of the new technologies and how these can be mitigated. The second session also included a section open for discussion on major challenges and opportunities.

As a first result of what has been the influence that the MSG-189 group has received from the participation to the CA2X2 Forum, a document that is currently under development inside the group is provided. The document, titled "Future Command Decision Training Support", analyses the topic of the actual situation with reference to Decision Making Training support making. It shows the limitations of the current solutions, in particular, for the training of High Rank Decision Makers and suggest a possible way forward.

The goal of the MSG-189 ST will be that of completing this type of detailed analysis looking at any level of

Decision Makers training paving the way for future further studies and research.

Here are the descriptions of the two Sessions.

Decision Support and Analysis

Moderator: Wim HUISKAMP, TNO Defence Research, Netherlands

"Bringing Commercial Games to Defence"

Iain McNeil – Matrix Games / Slitherine, UK

In this talk, the author presented an overview of Matrix Game's activities and how it accidentally found its way in to the defence industry. Iain will give examples of the line-up of game and dig deeper in to key defence titles such as Command, Flashpoint Campaigns and Battlefront. The talk also gave insights into a commercial gaming company's experience of working with defence - the highs and lows, the pitfalls and opportunities. Most importantly Iain covered why commercial games are such a big opportunity for defence to revolutionize how it wargames.

"WISDOM – How digital overlays in a geographic environment lead to wiser decisions"

LTC Luca PALOMBI – NATO M&S CoE, Rome, Italy

The Wargame Interactive Scenario Digital Overlay Model (WISDOM) is a new project that the NATO Modelling and Simulation Centre of Excellence (M&S COE) started at the end of 2019. WISDOM is a training portal and platform, where you can configure multiple geographic environments up to full wargaming scenarios aimed at supporting military and/or civilian training audiences to get wiser decisions during their MDMP, WARGAME phase or continuously during the planning and executing phases of simulated events. Pending the available data, a user benefits from playing a serious game in a geographic environment composed by several digital overlays grouped by nature or topics. In fact, WISDOM is a collection of tools able to display raw scenario data in digital overlays. During the presentation they shown the scenarios Archaria and Ṭarābulus, a storybook app called Raleigh, the methods used to build digital overlays and the results obtained. WISDOM has been designed to evolve over time to support all kinds of scenarios and new emerging technologies so far unknown. It will be the future venue that supports all the Training Audience's training objectives. Moreover, WISDOM supports any kind of training audience when dealing with exercise or mission rehearsal.

"Discovering and Leveraging Emerging Technologies for Application in M&S"

Joseph McDonnell, Christopher McGroarty, Chris Metevier, Scott Gallant and Lana McGlynn
U.S. Army Combat Capabilities Development Command - Soldier Center (CCDC-SC) SFC Paul Ray Smith Simulation & Training Technology Center (STTC)

Today both our adversaries and our technologies are changing rapidly. In 2020, we are facing challenges both typical and extraordinary, and as such, we are being called upon to employ emerging technologies in new and creative ways. While the daily business of maintaining and equipping the Army to keep the peace through strength by building on our military advantage and maintaining important regional balances of power continues, the how and where we do business has changed. It is only through our imagination and adaptability that we have employed these new technologies to address all challenges, anticipated and unanticipated.

Current technology advancements are not based on (or influenced by) the current state of Department of Defense (DoD) Modeling and Simulation (M&S) and its programs. Our job as M&S practitioners is to be smart in evaluating how to best adopt these advances to the benefit our military stakeholders, while considering interoperability with existing tools, data reuse, and standardization.

In order to expand your personal aperture and increase your level of awareness, we invite you to learn more and get involved in the Simulation Interoperability Standards Organization (SISO) Exploration of Next Generation Technology Applications to Modeling and Simulation (ENGTAM) Standing Study Group (SSG). The SSG focuses on technology adoption, technology application metrics, interoperability, and technology areas, such as data analytics, Artificial Intelligence, mixed reality, game development technology, and technology forecasting techniques. Members from the US DoD, many North Atlantic Treaty Organization (NATO) nations, industry, and academia, meet online monthly to discuss emerging technologies with the goal of understanding how they can be adopted and adapted to support a diverse body of M&S stakeholders.

This presentation discussed relevant findings from the SISO ENGTAM SSG and how they can be applied in the development and use of cutting-edge tools, techniques, and best practices. It also provided an opportunity to discuss these

emerging technologies and how M&S practitioners can leverage them to support the enablement of NATO.

"Virtual Battlespace 4 (VBS4) - Cloud-Enabled, High Fidelity and Whole-Earth Simulation"

Peter Morrison – Bohemia Interactive Simulations

Military organizations are striving to leverage best-of-breed simulation and web technologies to deliver high-quality training to the point of need — from Battle Simulation Centers to home computers. VBS4 is an easy-to-use, whole-earth virtual and constructive simulation that supports both individual and collective cognitive training. The VBS4 simulation and rendering engine (VBS Blue) has been developed to support both terrain streaming from the cloud and scalability. A complete replacement for its predecessor VBS3, VBS4 supports hundreds of training use cases including new use cases like small unit Course of Action (CoA) development and analysis, and combined arms and staff planning. The new VBS4 workflow dramatically speeds up the development of training content through its new modes VBS Geo (an easy-to-use but powerful terrain editor) and VBS Plan (a highly efficient mission planning capability). The new VBS World Server is an optional and cloud-enabled companion product for VBS4, which streams terrain to VBS4 instances across a network. It also centralizes the storage of VBS4 Battlespaces - further reducing the overhead of administering multiple VBS4 installations.

Emerging Simulation Technologies for Decision Making Support in the 21st Century

Moderator: Agatino MURSIA, Leonardo Company, Italy

"*COA analysis from simulated forces*"

Keith BRAWNER – U.S. Army Combat Capabilities Development Command - Soldier Center (CCDC SC) SFC Paul Ray Smith Simulation & Training Technology Center (STTC)

The idea of creating a digital representation of the battlefield is not a new idea to the community of wargaming practitioners. In the pre-computing world of wargaming, "simulated forces" were represented by physical pieces, on physical maps, taking physical actions - physical pieces were moved on physical maps, with battle outcomes were decided by educated guesses from force commanders. This technology has been replaced at the rough order of magnitude by digital maps created from real-world terrain (McAlinden, 2013), moved according to entered simulation commands in response to underlying cognitive and decision-theoretic models (Clive, et al, 2015), and digital guesses using "semi-rigid adjudication parameters" (UK MoD, 2017). The physical world was replaced with a computer replication and parameters, but only at the level of larger-scale troop movements. New technologies are coming available which enable a finer grain size of analysis within the context of constructive simulations; allowing analysis at the Warfighter level rather than the brigade level. These new capabilities are needed in order to enable synthetic forces simulation and visualization in the live training domain enabled by augmented reality (AR). Given the highly detailed models of the humans and Artificial Intelligence (AI) trained models of the OPFOR, Course of Action (COA) decisions can be rapidly modelled and trained from data collected in real-world battles. The presentation investigated the new advancements in COA training from AI OPFOR models built on downrange training data.

"MSG-HFM-323 ST on Cyber Sickness"

Paolo PROIETTI – Leonardo Company, Italy

The human brain must integrate real-time visual, auditory, vestibular, somatosensory, and other inputs to produce a compelling feeling of immersion in the natural environment. In the past decade, there has been a rapid advance in immersive Virtual Reality (VR) technology which involves mainly visual and auditory senses in bimodal interactions. A problem with VR is that users develop symptoms similar to motion sickness - a malady called cybersickness. The discomfort that users experience during or after a session in a synthetic environment became widely known about in the military setting during the advent of flight simulators. The related phenomenon of simulator sickness can discourage pilots from using flight simulators, reduce the efficiency of training (through distraction and the encouragement of adaptive behaviours that are unfavourable for performance), or compromise safety when sick or disoriented pilots leave the simulator (e.g., to operate ground vehicles). In a similar manner, cybersickness can be a barrier to using VR for military training, and thereby limit the dissemination of improved training or rehabilitation tools. Cybersickness was evaluated by Study Group 323 within the Human Factor & Medicine (HFM) Panel and NATO Modelling & Simulation Group (NMSG) of the Science & Technology Organisation

(STO) of the North Atlantic Treaty Organisation (NATO). They reviewed the factors contributing to sickness that are associated with the individual (e.g., history of susceptibility), the VR system (e.g., system lag), and the task (e.g., type of virtual locomotion control). Solutions to reduce symptoms of cybersickness were identified, such as earth-referenced cues and exposure limits. These can be implemented during system design and usage, and can aid in the management and treatment of cybersickness. Adoption of the guidelines in this report for mitigating cybersickness will enhance training effectiveness throughout the military community through better implementation of VR.

"High Performance Computing and decision-making, the experience of Leonardo Labs"

Carlo CAVAZZONI – Leonardo Company, Italy

In a recent speech, IBM CEO Arvind Krishna, said that: "digital transformation has been accelerated during the COVID-19 pandemic and ultimately every company will become an AI company." This is not strictly true, but what is true is that every company will have to adopt AI technologies. AI considered in a broad meaning, and we can be more precise saying that every industry will have to applying digital technologies with certain degree of cognitive capability to support humans. In particular, AI and computer simulations are boosted by availability of data and processing capability, the more the better. Supercomputers then represent a tool to enable new digital technologies and accelerate innovation. With their processing capability, they can shorten the time taken to process data, train AI models and perform simulations or run ensemble simulations of multiple scenarios in parallel. Without High Performance Computing and massive amount of collected data the modern AI would not be possible. The talk presented how Leonardo, with the key contribution of the HPC Lab, intends to implement leadership software tools and computational infrastructure to support the development of new decision-making technologies and, in general, how they will support transformation in Leonardo.

Future Command Decision Training Support

Contribution to MSG-189

Iain McNeil (Slitherine)

Nico de Reus (TNO)

1 Introduction

To date the field of military Modelling & Simulation has been mainly involved in training applications for training of the lower military levels. Significant investments in M&S are made for frontline personnel such as pilots, gunners, drivers and infantry. This is done through a range of systems using 3D and VR trainers. Many of these trainers aim at training skills and procedures rather than tactics. Some nations have advanced capabilities for simulation supported joint exercises (e.g. USA), however, NATO wide training for operational readiness at the higher levels between services and with coalition partners is still mostly limited to Live exercises. Work is ongoing in the NMSG towards a persistent Mission Training through Distributed Simulation (MTDS) capability. Staff trainers and wargames are also being used increasingly.

There is however a huge gap in the use of M&S applications such as wargames to train commanders and staff and in the tooling to assist the more senior commanders to make better decisions. Although training applications for these higher level tactical commanders exist these require a lot of manpower and budget to use and consequently these commanders cannot be trained frequently. New technologies, like gaming and AI, however would enable these higher military levels to be trained more frequently. In this paper we discuss the importance of training these higher military levels and how using new technologies could enhance their training, as well as offer tools for operational decision support.

We will assume a land operations use-case to describe how currently training at these levels is performed. Based on this we present the initial vision of how training could be performed in the future. The gap between these two will give indications for new technologies and research that is required to be performed to fulfill this vision. This is, by the way, the expected outcome of the activites of MSG-189 Specialist team.

2 High tactical level (Command level) Training

2.1 High level Theoretical Training

There is little to no simulation training of more senior commanders. They may get to do manual wargaming, which is very useful, but very rare, due to the time and cost of getting the people together to run those games. It also tends to be the case that the more senior the commander the less opportunity they have to train. There is also peer pressure and an aversion to experimenting with new tactics as everything you do is very visible to the other players around you so people take safe options and do not experiment, which is where real learning and innovation occur.

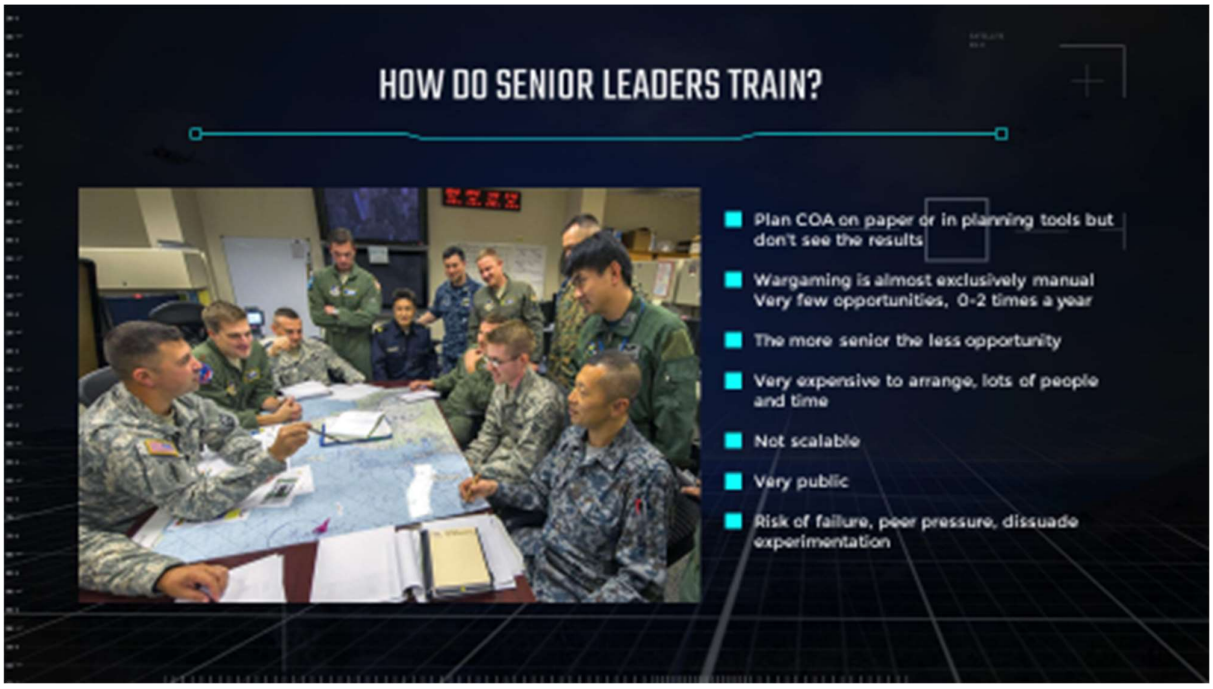


Figure 1

2.2 High level Procedural and Mission Training

2.2.1 The importance of higher level Command Training

It is common sense that the more senior a Commander, the more people they have under their command, so the more people their decisions affect. Decisions made by Commanders affect the tactical situation that the frontline soldiers find themselves in and this clearly affects their likelihood of success. Good commanders are a force multiplier and bad commanders reduce the effectiveness of their troops. Clearly then it means that decisions made by senior commanders have more impact on the outcome of battles than junior commanders and frontline personnel. The conclusion then is that it is critically important to train our senior Commanders.

All military leaders study history and know the stories of Hannibal and it is a good example. At Cannae Hannibal was outnumbered and outclassed. Man for man his infantry were no match for the Roman legionaries and they were outnumbered. His only advantage was that the Roman cavalry was weak. Hannibal used this to his advantage, expecting his infantry centre to be pushed back while his cavalry enveloped the Romans, leading to one of the bloodiest single days of battle up until WWI.

This is not an isolated example and history is full of such events where great leaders have won against the odds. Napoleon at Austerlitz, Manstein in France. At the time of the invasion, the German tanks were no match at all for French tanks and could not hurt them frontally. In a head-on fight, the Germans would have lost. Manstein's plan involved sucking the elite British and French troops in to the Low Countries and surrounding them, cutting them from supply, making them easy targets. Similarly

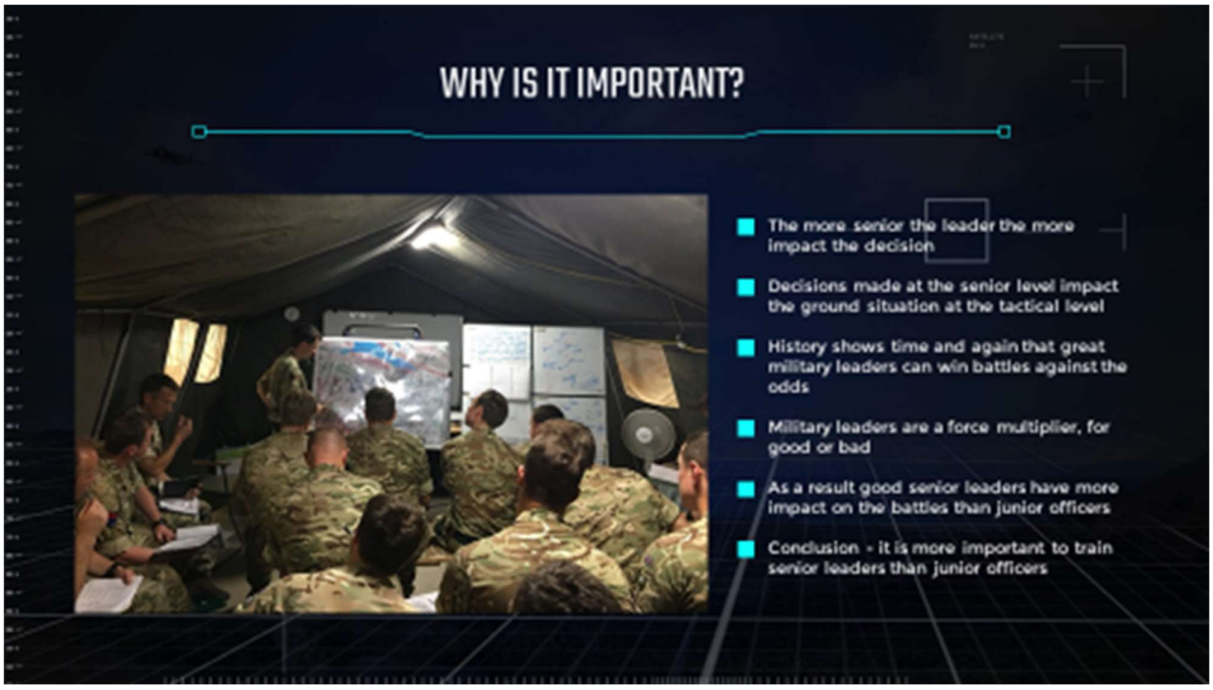


Figure 2

bad decisions can result in massive failure. The French Grand Armee that invaded Russia in 1812 was arguably the best in the world at the time and it got annihilated by the Russian Winter. Hitler did not learn from this mistake and repeated it in WWII.

Figure 3

There is limited access to Command Level Training and it is not very standardized so there can be significant variety in how much and the type of training that officer receive. Tactical training is provided in a range of ways:

- Systems like CAST. This is a digital system that lets officers and their troops train. It is very useful to experience how orders would be put in practice. The limitation is that every entity is human controlled so a commander would need hundreds of lower level players to simulate even a medium sized operation. While it does allow the commander to train, it offers limited ability to try new tactics and strategy and experiment due to the number of frontline personnel in each "game".
- Manual wargames (boardgames). These are manually moderated games, usually using custom rules set to resolve combat or SME's to adjudicate, though can also be based on commercial board games. They are a good way to present interesting challenges to players and teach analytical tactical and strategic thinking. They present tricky problems with many possible solutions and the challenge is to find the optimal one. They can be organized with blue and red teams or just blue team vs the umpire. Red vs Blue present a real thinking enemy and requires you to outsmart them.
- Live exercises / rehearsal wargames. These are live operations where troops will fight out preset battle plans. They are great for practicing how to get a large formation to move to a location at a set time. However, they are not really any use in learning tactics and strategy, and are again aimed at providing training for the numerous frontline soldiers, so tend to be on rails with no real decision making for the commanders.

Note that different levels of command require different kinds of training. A Platoon Commander needs very different tactical training to a Battalion Commander, and they very different to a Divisional Commander.

Figure 4

3 Current training

This section discusses how currently training is performed by looking at both the lower level as well as the higher military levels.

3.1 Training of lower level (frontline) personnel

The current training is focused on the pilot, driver, sailor or infantry. It is very much about training the guys on the ground in the front line and when you talk about training

this is what everyone immediately assumes. People think 3D and VR. The frontline soldiers make up the majority of staff so we have slipped in to training them, and mainly them, because they are most numerous.

3.2 Training of higher command levels

Before refining the higher command training process, we must first look at the problems with current training, which is mostly for the lower levels.

3.2.1 Problems with current training

There are a number of problems with our current training. There is far too much focus on the front line soldier. We need to develop systems to allow our commanders and their commanders to train digitally so they can do it in their own time at their own speed and wargames are a great potential tool for this. .

Figure 5

There is a misconception that by making graphics more detailed and moving to AR or VR there will be benefits for everyone. Whilst this is true for that soldier on the front line, this is less true for a company commander, and rarely needed for operational decision making. The higher up the command chain you are the less you need realistic views and the more you need an abstract view, for example an annotated COP. It is important that we do not over-focus on VR and 3D and ensure we provide the right tools for the commander. In a recent demo presented by a developer, a missile defence system was being controlled in VR. In reality the missile defence system would engage the target long before it came in to visible range and the technology is unfortunately redundant, but it still gained a lot of attention. Although it may be of use at the political decision making level, at the tactical level realistic views higher up the command chain can be a distraction rather than a decision aid, and in the UK officers are trained to ignore video feeds in HQ's for just this reason. Another common problem in defence is stovepiping. Each service has its own systems, models, and funding paths. The modern battlefield is multi-domain and any simulation needs to model cross service operations and be funded cross service. This is not just at cross service, but also at the sub service level. Sometimes this stovepiping and areas of responsibility can be barriers to the creation of new and useful tools. For example, the simulation team may regard command decision support tools to be a C2 issue, while the C2 people have no ability to produce these kind of tools and rely on what is provided by the simulation and modelling teams. This can end up with tools that are not optimized for the tasks or a complete capability and training gap.

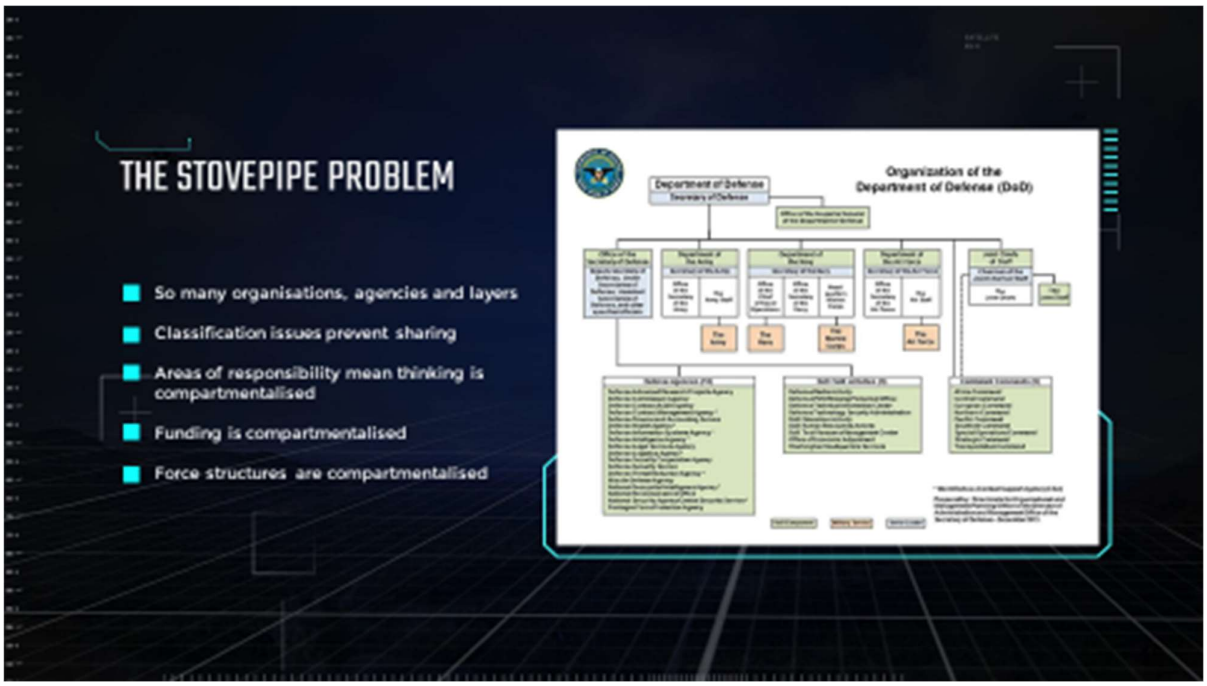


Figure 6

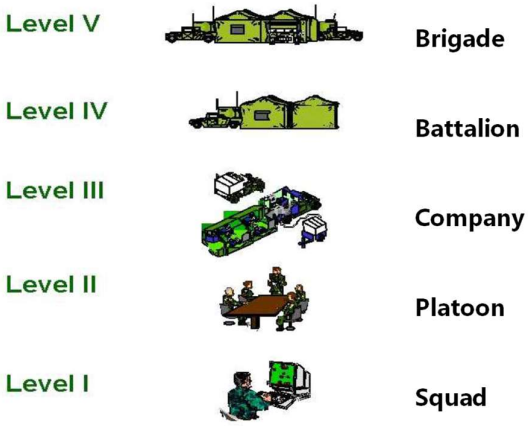
3.2.2 Paper training

As anticipated already there is little to no simulation training of more senior commanders. They take courses and they plan COA's on paper, but rarely get to see what the results of their plan would look like Very few good examples of simulation tools are available to date. This is recognized to be a gap from the initial MSG-189 analysis. New very promising technologies are now available to change the way we do simulation so far and create new opportunities.

3.2.3 Digitally assisted training

Although not frequently used, some computer-assisted training applications for command training do exist. As was discussed, these currently require a large training exercise staff. The current state of the art of these systems is discussed in this section after which in subsequent sections a possible future will be discussed.

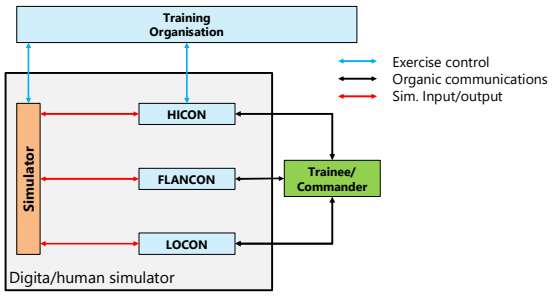
Before going into an example level, we first describe the different levels involved in land operations training. These are visualised in the figure below.



Current digitally supported Command Staff training systems use a combination of computer simulation and a training staff (e.g. role players).

As an example think of a Battalion commander that is trained in commanding his subordinates and following the Military Decision Making Process (MDMP). He will issue commands to his subordinate Companies, where each Company consists of say, 3 Platoons and each Platoon consisting of say, 3 squads. All these subunits can carry different weapons and can be transported by vehicles. All these entities involved need to perform realistically following the Commander's orders.

This is where Lower Control operators come into play, which transform the higher level commander's orders into lower level orders that the simulation can execute. The "simulator" in that sense is a combination of a computer model supported with operators. In the figure below this has been visualised.



Three types of operators are involved:

- LOCON (LOwer CONtrol operators), in the above picture (also called Pucksters) transform the trainee's commands into simulator executable commands.
- FLANCON operators represent Flanking units' commanders that control the subordinates of the flanking units in the simulator. In the example above, where the trainee is the Battalion commander the FLANCON is a Battalion Commander.
- HICON represents the Commander's (trainee's) higher level commander. In the example above, where the trainee is the Battalion commander the HICON is a Brigade Commander.

As can be seen, the trainee is completely merged into a realistic environment, usually a command post tent using this organic available communication means. The two other components involved are:

- Training organisation or exercise control. This steers the training session by defining what needs to be trained and by operating the digital simulation. The actual order that the trainee will receive, through HICON, based on the available skills of the trainee compared to the required level of skills.

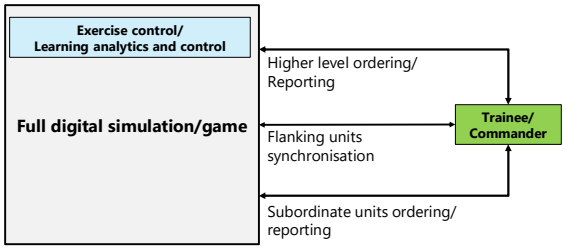
- The hybrid (digital/human) simulator consisting of the digital simulator and the manual HICON, FLANCON and LOCON components.

The type of training involved in the above Command training example is both procedural as well as tactical.

3.3 Vision for future Command level training

Given the fact that the current training sessions with a digital/human training system as described in the previous section, requires so much staff that it can only be done infrequently or can only be done on paper without digital support, the question pops up what kind of technology could make these kinds of training sessions more accessible.

The vision that we present here is that by using AI technology and (currently available) gaming simulators a lot of the currently required staff will no longer be needed. The following figure represents this vision.



This would enable a military staff to train more often, with an aim that commanders can train in spare hours and don't need a full staff to support that training.

4 Technologies required to fulfill vision

In this section we elaborate on technologies, available and emerging that can be used to fulfill the vision. We distinguish between already existing gaming simulators and Artificial intelligence technologies.

4.1 Automating training support functions

Military co/counter-play

The role of the subordinate and flanking units as well as the higher command are currently fulfilled by humans (i.e. role players). Artificial Intelligence could also fulfill this role. This could be done initially using complex systems and rules based AI and later adapted to machine learning approaches.

For instance, an exercise control module could, in combination with a HICON module, based on the trainee's progress, define scenarios and orders for the trainee. LOCON modules could determine the best way

to transform the trainee's orders into subordinate orders and a FLANCON module could play flanking units. A Red unit behavior module could be used to play the enemy.

Communications

The way that the trainee currently receives and gives orders is by the organic communication means. It is imaginable that speech technology is used to provide the trainee a way of communicating that has the same look and feel and thus is sufficiently viable for training him/her.

4.2 Challenges

Fidelity level of simulation

The statement in the section above about the level of fidelity of the simulation that should be selected carefully for the higher command level is true, especially for the interface and the provisioning of information to the commander. The commander should not get too detailed battlefield information and therefore should e.g. not use VR or 3D technology because it can be distracting. On the other hand, the fidelity of the simulation should be sufficiently high to ensure realistic behavior. This will also depend on the training purpose that can be tactical training or procedural training. 3D and VR can also be useful for after action reports to understand why things happened, but are of limited use at run time.

The challenge will be to find the "right" fidelity level, depending on the training purpose. For instance if a simulation system is to be used for rehearsal of a given mission, the modelled environment, including the behavior of (own and co-acting) units, needs to be close enough to the real environment, however if it is only to learn tactics, this requirement can be relaxed.

Train as you fight

"Train as you fight" is a paradigm that is often used. It usually means that a warfighter should use his operational systems as much as possible. For C2 training, this translates to the requirement that he/she must use his native C2 information system, which should obtain its input from the simulator/game. In practice, this can be translated to the requirement that the C2 information system must interface with the simulator/game.

The challenge then is to implement the use of C2 – Simulation Interoperation standards, see [1], or create tools that can replicate this behavior. There is potential for middleware tools to be created that allow users to work in familiar processes and the tools convert this input to language and instructions that the simulation understands. In the mid to long term it may even be possible to use AI assistants to convert speech or written instructions to simulation parsable formats.

The need for better decision support tools

The future battlespace will have many more sensors all delivering data across multiple domains, huge amounts of dummy and decoy data and fought at a pace never seen before with hypersonic missiles and stealth technology. These all combine to create an environment that will overwhelm a human brain and require support tools to assist the decision making process.

In order to get more flexibility, we need more support to include/modify/update our tools and models to match the rapidly changing real mission environment. That leads to the requirement that a new architecture is needed to provide this flexibility.

The goal of ST MSG-189 is to provide a first example of such architecture. Potential solutions are improved AI and MSaaS ecosystems that together with other promising new technologies will shape this innovative architecture. The work has just started and this paper represent a very initial step towards the desired outcome.

Figure 7

Reference

[1] SISO-STD-019 – (2020) Standard for Command and Control Systems - Simulation Systems Interoperation.

Talk About CAX Forum

Italian Ministry of Defence

www.difesa.it

ST Engineering Antycip

www.steantycip.com

EXercises and TRAining activities Reports

www.extra-reports.it

www.lastampadelmezzogiorno.it

Talk About Us

Land Power Magazine

lc.nato.int/media-center/landpower-magazine

MDPI

Multidisciplinary Digital Publishing Institute

www.mdpi.com/journal/robotics/events/12028

Rivista Aeronautica

www.aeronautica.difesa.it/comunicazione/editoria/rivaeronautica/Pagine/default.aspx

DIFESA ONLINE

Difesa Online

www.difesaonline.it/mondo-militare/fido-robotizzato-come-i-cani-robot-entrano-nel-mondo-militare

AvioNews
World Aeronautical Press Agency

AvioNews

https://www.avionews.it/it/item/1232992-esercito-faro-di-tecnologia-e-di-innovazione.html

AD ANALISIDIFESA

AnalisiDifesa

www.analisidifesa.it/2021/02/corso-di-cyber-wargame-al-centro-di-eccellenza-di-modelling-and-simulation-della-cecchignola/

CYBERNAUA.it
INFORM ACTION MAGAZINE

Cybernaua

http://www.cybernaua.it/news/newsdett.php?idnews=8750

DEDALONEWS
SETTIMANALE AD AGGIORNAMENTO COSTANTE DI AEROSPAZIO, DIFESA E SICUREZZA

Dedalonews

http://www.dedalonews.it/legale.htm

Amici delle Forze Armate

https://www.facebook.com/AMICI-DELLE-FORZE-ARMATE-

181640354476

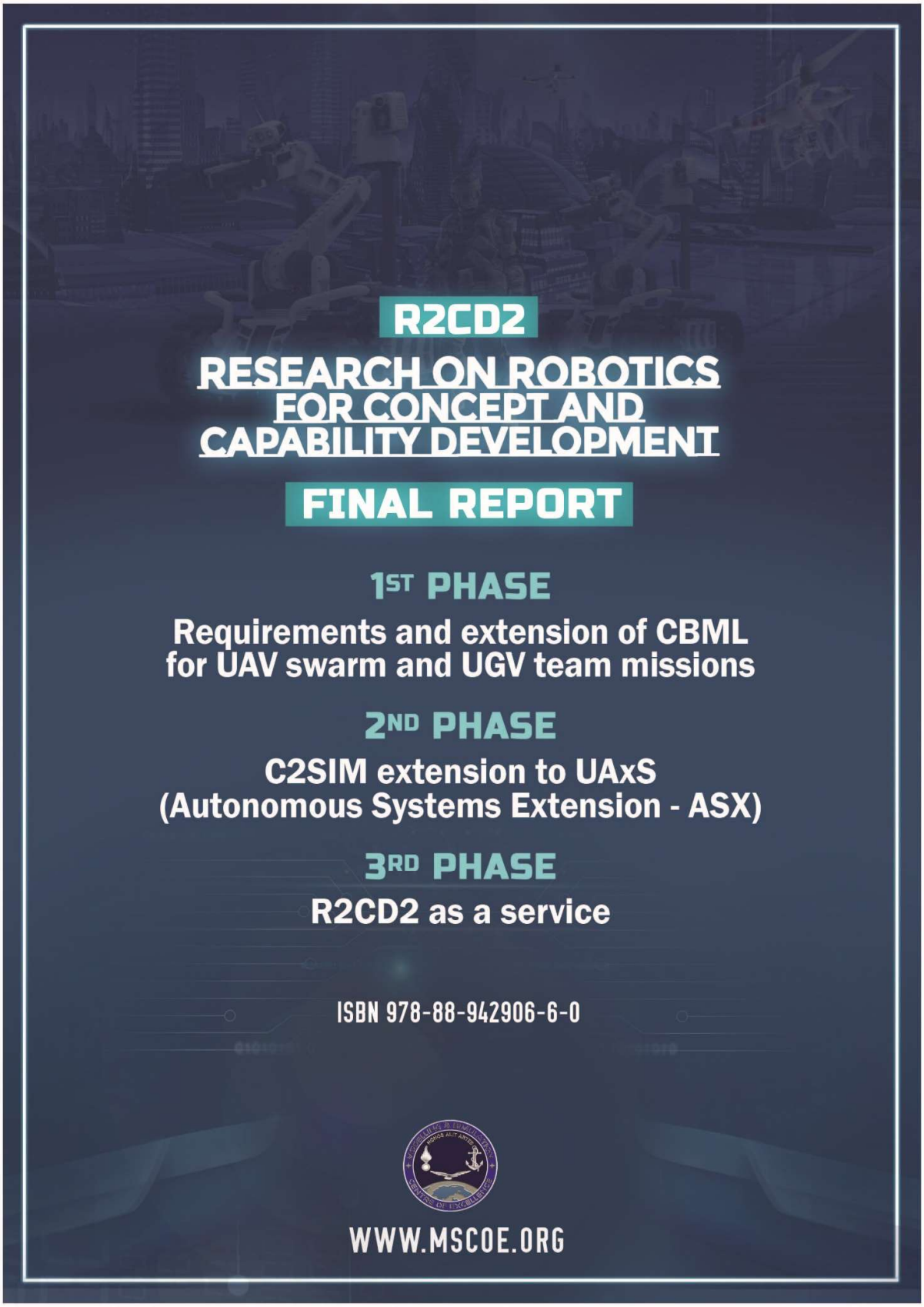
NATO MODELLING AND SIMULATION CENTRE OF EXCELLENCE
PRESENTS
CAX FORUM 2020 WS 35
MODELLING AND SIMULATION AS A SERVICE Day
SESSION 1 CUSTOMER PERSPECTIVE 14:30 – 15:40 CEST (UTC+4)
Dr. Robert SIEGFRIED ADITERNA GMBH Cpt. Peter LINDSKOG SWE MoD Wim HUISKAMP TNO
https://ca2x2.liveforum.space/workshop/22
SESSION 2 USER PERSPECTIVE 15:55 – 17:05 CEST (UTC+4)
Lt. Col. Davide Marco TRIMANI NATO M&S CoE Rob KEWLEY SIMLYTICS.CLOUD
Cpt. Peter LINDSKOG SWE MoD Marco PICOLLO LEONARDO
https://ca2x2.liveforum.space/workshop/23
SESSION 3 TECHNICAL PERSPECTIVE 17:20 – 18:30 CEST (UTC+4)
Tom VAN DEN BERG TNO Daniel KALLFASS AIRBUS
Rob KEWLEY SIMLYTICS.CLOUD Marco PICOLLO LEONARDO
https://ca2x2.liveforum.space/workshop/24
SEPTEMBER 23RD 2020
CAX FORUM 2020
WWW.NATO.INT
WWW.MSCOE.ORG

NATO MODELLING AND SIMULATION CENTRE OF EXCELLENCE
PRESENTS
CA2X2 FORUM 2020
IMMERSIVE
TECHNOLOGY
APPLICATIONS AND TRENDS
Ciro DONALEK VIRTUALITICS John NICOL CORONA AEROSPACE INC.
Gabriele ROMAGNOLI GR TALK Nazaro AVERSANO LEONARDO
SEPTEMBER 23RD 2020
JOIN THE WORKSHOP FROM 17.20 TO 18.30
https://ca2x2.liveforum.space/workshop/27
CAX FORUM 2020
WWW.NATO.INT
WWW.MSCOE.ORG

NATO MODELLING AND SIMULATION CENTRE OF EXCELLENCE
PRESENTS
CAX FORUM 2020 WS 35
MODELLING AND SIMULATION AS A SERVICE
INDUSTRY PERSPECTIVES
BETWEEN BUSINESS OPPORTUNITIES AND CRITICAL ISSUES
Lt. Col. Davide Marco TRIMANI NATO M&S CoE
Oliver ARUP BISIM
John NICOL CORONA AERO
Sebastien LOZE EPIC GAMES
Filippo GEMMA GMSPAZIO
Giovanni TONELLI LEONARDO
Peter SWAN MAK
Graziano LENTO PRESAGIS
SEPTEMBER 24TH 2020
JOIN THE WORKSHOP FROM 14,30 TO 15,40
https://ca2x2.liveforum.space/workshop/35
CAX FORUM 2020
WWW.NATO.INT
WWW.MSCOE.ORG

NATO MODELLING AND SIMULATION CENTRE OF EXCELLENCE
PRESENTS
CYBER
MODELLING AND
SIMULATION
WORKSHOP
GUARDTIME • ANTYCIP
CY4GATE • NATO M&S COE / LEONARDO
Modelling the Future
SEPTEMBER 24TH 2020
15:55 - 17:05
CAX FORUM 2020
WWW.NATO.INT
WWW.MSCOE.ORG

NATO MODELLING AND SIMULATION CENTRE OF EXCELLENCE
PRESENTS
R2CD2
RESEARCH ON ROBOTICS FOR CONCEPT AND CAPABILITY DEVELOPMENT
FINAL REPORT
WWW.MSCOE.ORG


R2CD2
RESEARCH ON ROBOTICS FOR CONCEPT AND CAPABILITY DEVELOPMENT
FINAL REPORT
1ST PHASE
Requirements and extension of CBML for UAV swarm and UGV team missions
2ND PHASE
C2SIM extension to UAxS (Autonomous Systems Extension - ASX)
3RD PHASE
R2CD2 as a service
ISBN 978-88-942906-6-0
WWW.MSCOE.ORG