Empowering Military Decision Support through the Synergy of AI and Simulation

Joost van Oijen
Royal Netherlands Aerospace Centre
THE NETHERLANDS

Joost.van.Oijen@nlr.nl

Pieter de Marez Oyens
Royal Netherlands Aerospace Centre
THE NETHERLANDS

Pieter.de.Marez.Oyens@nlr.nl

ABSTRACT

In military tactical operations, there is an increased demand for machine-based decision support capabilities to support human decision-makers. Artificial Intelligence (AI) and Modelling and Simulation (M&S) are key technologies for enabling these capabilities, showing considerable progress in recent years. AI technologies become increasingly proficient in cognitive tasks such as situation assessment, course of action planning, and teaming with humans. Advances in simulation achieve higher levels of realism in representing physical and behavioural elements of the battlefield. Furthermore, there is growing synergy between the technologies, as seen by applications of e.g. generative modelling and reinforcement learning.

In this paper, we present an overview of the multi-faceted and interdependent roles of AI and simulation for empowering decision support capabilities in military decision-making. Based on an analysis of decision support, we assess upcoming AI technologies and their potential impact. Subsequently, we identify considerations for military M&S platforms for exploiting these technologies. For illustrative examples, a case study is used in the domain of tactical air command & control.

The aim is to provide insights within the community on the impact of current AI advancements in enhancing military decision support, and the corresponding needs of M&S technologies, building upon existing R&D efforts in the field.

1.0 INTRODUCTION

Decision-making is a critical process in military tactical operations. It is where information is gathered and analysed, situations assessed, courses of action considered, and operational plans are made to achieve specific mission objectives. The role of decision-making for the military operator also changes, due to an increasingly technology-driven battlefield. This can be seen throughout all command levels and organizational roles, from analysist and commanders, to platform operators and soldiers on the ground. Better access to and use of data and information will improve the quality and agility of decision-making, enabling increased adaptiveness during operations and faster mission planning cycles. At the same time, the integration of unmanned systems will lead to increased autonomous sensing and acting capabilities at the edge, which require human oversight and orchestration [1]. These developments impose different demands for an operator, involving more strategic thinking with stricter time-constraints, while teaming with intelligent systems to achieve a faster OODA-loop. In this regard, to prevent information and cognitive overload for an operator, the use of intelligent decision support systems (DSS) are vital to accompany the human in this changing role [2].

Artificial Intelligence (AI) and Modelling and Simulation (M&S) are key technologies for enabling machine-based decision support, each showing considerable progress in recent years. AI technologies become increasingly proficient in cognitive tasks such as sensory processing, situation assessment, course of action planning, and teaming with humans. The potential impact of AI on the military decision-making process has regularly been assessed [3], [4]. Consequently, advances in simulation achieve higher levels of realism in representing the physical and behavioural elements of the battlefield, also in part driven by AI. Synthetic environments can be developed faster from real-world data sources to simulate representative mission environments [5], whereas Computer Generated Forces (CGF) to populate these environments are equipped with more realistic behaviour models to simulate friendly or adversarial behaviours [6]. Such AI-enabled simulation environments can be used for decision support to understand complex real-world mission environments and evaluate different courses of action against potential adversaries, in a safe and cost-effective manner [7], not only for mission planning and execution purposes but also for experimentation and training purposes [8].

There is also a growing synergy between AI and M&S technologies, accelerated by AI advances and fuelling opportunities for decision support. For instance, generative AI has seen tremendous progress in creating highly realistic visual or auditory content, whose underlying algorithms are explored for generating realistic virtual environments and human behaviours [9], [10]; task-solving algorithms continue to improve each year and can cope with more complex tasks in more complex environments with less training time, often by several orders of magnitude [11]; and due to recent breakthroughs in large language models (LLMs), this technology is rapidly being integrated into human-machine interfaces, and its potential for military decision support has already been envisioned in detail in commercial promotional videos [12]. In this evolving landscape, there is a continuous demand for R&D and experimentation on how some of these technologies can be tailored to and applied for military decision support. However, as a viable playground for military tactical environments, current military M&S platforms often lack some essential capabilities that hinder such AI experimentation on a broader scale.

This paper presents an overview of the interdependent roles of AI and simulation in military tactical decision support. Based on an analysis of decision support where we review related work, we conduct a horizon scan of AI technologies and their potential impact. Consequently, we identify a set of considerations for military M&S systems to facilitate the integration of AI technologies. The aim is to provide insights within the community on the impact of current AI advancements in enhancing military decision support and the corresponding needs of M&S technologies.

2.0 MILITARY DECISION SUPPORT: A BRIEF ANALYSIS

Military decision support is often discussed in terms of how (AI) technology can support a decision maker in its decision-making process, such as the well-known OODA-loop [3], [7]. In this section we present a comparable view of decision support, but with a focus on tactical C2. In a C2 context, the Dynamic OODA-loop (DOODA-loop) has been proposed as a general model of C2 that can be used for the strategic, operational and tactical levels, highlighting the adaptive qualities of mission execution [13]. Figure 2-1 illustrates a basic view of decision support, in alignment with the DOODA-loop from [14]. Complementary, the view highlights the decision-making process as a joint process between an operator and an (AI) system, requiring some form of teaming between the actors. Below, we describe the different functions in this process and the role of AI herein.

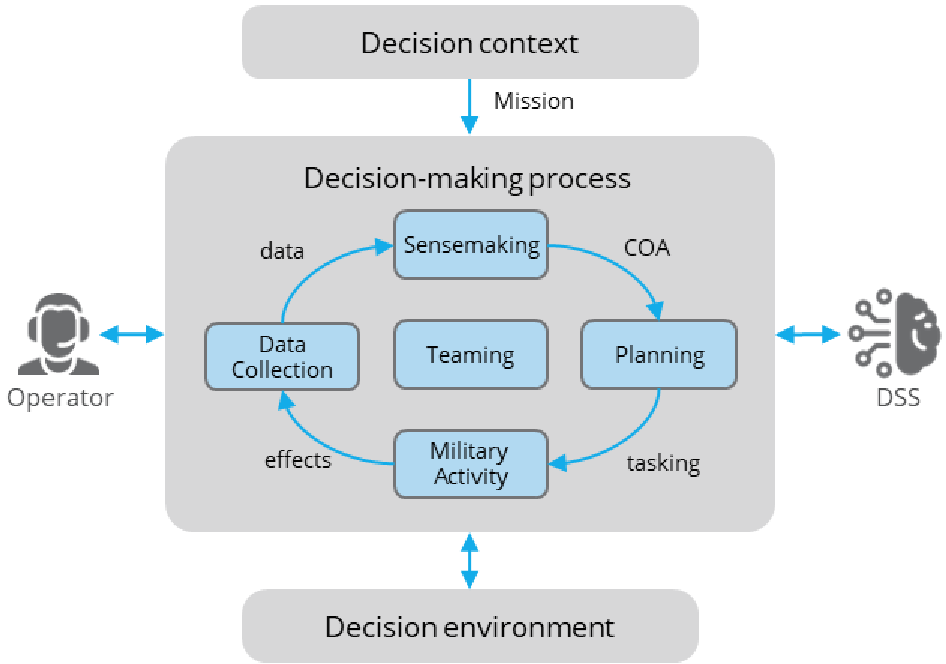

Figure 2-1: A military decision support system (DSS) view in a C2 context

Sensemaking: is the process that interprets data collected from the environment into actionable information and aids in determining necessary actions through the establishment of a course of action (COA). In terms of data analytics, it covers all levels of analytics: descriptive, diagnostic, predictive, and prescriptive [15]. Decision support is mainly concentrated around this process, in which we distinguish between providing support for establishing (1) situational awareness and (2) situation understanding. Where situational awareness governs the real-time perception and interpretation of what is happening, situation understanding provides a deeper analysis of why things are happening by forming a comprehensive grasp of the overall context, implications and potential courses of action (COA) concerning the mission.

One role of AI for decision support is using analysis methods to facilitate operators in quickly understanding and evaluating the situation through insightful information, hereby lowering the cognitive burden of information processing. Examples are plentiful and can relate to geospatial analysis [16], tracking, aggregating and annotating enemy units based on a priori intelligence [17], predicting enemy movements, or analysing ingress strategies such as helicopter landing zones [18]. Providing analytical results towards the operator is often mediated through some form of tactical *common operational picture* (COP).

Consequentially, AI can support operators in developing, testing, evaluating and weighing different hypotheses for effective COAs. These processes are common in military planning methodologies such as the Military Decision-making Process (MDMP) [19]. Two common paradigms of AI used are conventional computational experimentation and contemporary optimization-based methods, both dependent on Modelling & Simulation (M&S) technologies. Through computational experimentation, what-if scenarios can be run to measure and analyse different plans' expected effects [3]. Through techniques such as Monte Carlo-based search methods, large amounts of scenarios can be computed and analysed, and the results offered to the operator [1], [20], [21]. Alternatively, machine learning-based approaches, such as reinforcement learning (RL), gain traction and are based on self-learning algorithms to learn optimal behaviour policies (COAs in this context) through trial and error in interactive simulation environments [22]. Trained models can be used to optimise strategies or tactics of individual units or the coordinated behaviours between teams. These algorithms can support decision-makers in various ways, such as determining effective friendly COAs or the enemy's most likely or most dangerous COAs, Red Teaming, and tactics development to identify weaknesses in the tactics of allies or enemies.

Planning: is the process of translating a selected COA to an operational plan for action that can be effectuated into military activity through instructions, orders or tasking. It translates *what* needs to be done to *how* it is to be done [13]. In terms of decision support, similar AI technologies used for developing COAs could be used for generating plans. However, as planning can be regarded as a more low-level decision problem, different techniques may be deemed necessary. To give an example, a COA in an air mission may reflect the act of conducting a precision strike on a high-value enemy target, whereas an associated plan may involve specific routing strategies of strikers to minimize visibility from enemy radar. To facilitate different levels of AI optimization, different levels of abstraction may also be required for M&S environments. Common levels in military simulation are the *engineering*, *engagement*, *mission* and *campaign* levels, associated with different levels of details and aggregation [23].

Military activity and data collection: is about producing effects in the environment, and collecting data from the environment used for sensemaking. These processes fall outside the scope of a C2 decision-making process. However, depending on the capabilities of the presumed system from Figure 2-1, it may support these processes by e.g. delegating instructions through communication and coordination between (un)manned assets, or mediating with Intelligence Requirement Management and Collection Management (IRM&CM) processes.

Human-machine teaming: refers to the collaborative interaction between the operator and the system, where each actor contributes its unique strengths and capabilities to enhance overall performance, decision-making, and problem-solving. True collaborative decision-making demands that involved actors are mutually predictable and directable, and maintain a common ground [24]. However, the necessity of such requirements depends on the level of integration and the role of AI in the decision-making process. In [4], an assessment is given of different levels of human-AI collaboration, incorporating AI qualities such as transparency, explainability, initiative, adaptiveness, and Theory of Mind.

It should be noted that there are also reservations on the capabilities of AI in C2 decision-making and that human judgement remains essential. In [25], it is argued that AI cannot reliably compliment or replace the human role in apprehending strategic environments, as military decision-making environments are non-linear, complex and uncertain, and tactical commander's qualities of initiative, creativity and empathy remain critical. In [3], it is further recognized that AI will not have the capability (in the near future) to automate the role of a decision-maker in a C2 system, establishing orders without a human-in or -on the loop, and that instead of a more gradual introduction of AI seems plausible.

3.0 AI HORIZON SCAN

The field of AI is rapidly developing in recent years. This section describes a selection of developments and assesses their impact on military decision support. Mainly, we focus on those techniques that support the processes of *sensemaking*, *planning* and *teaming* from the previous section.

Large Language Models

Large language models (LLM) are deep neural networks trained on vast amounts of text data to understand and generate human language. Models such as GPT-3.5 have emerged as crucial tools with significant implications and continue to impact various sectors. Their ability to comprehend context and perform language-related tasks at a human-like level makes them highly valuable in driving societal progress [26]. These models demonstrate their potential to streamline processes, extract insights from vast data, and provide intelligent solutions. More potential lies in the realm of multi-modal LLMs (MLLM). These models process, besides text, other modalities such as images or visual and auditory content in video, allowing the model to grasp a more extensive understanding of the world that is not solely based on text [27].

LLMs provide a natural method for human-machine interaction. Moreover, they exhibit *teaming* characteristics where they can build common ground with users through intentful machine interaction. Through AI alignment techniques, LLMs can be trained to ensure that results align with human preferences, values and intentions [28]. Their ability to analyse different data types and reasoning about them in human language makes this class of models perfect candidates for assisting military commanders in *sensemaking*. They can help assess the feasibility and risks associated with different COAs by developing more informed strategies, refining operational plans, and evaluating potential outcomes. Moreover, users can interact with these models to understand their reasoning process and identify arguments behind decisions.

For the acceptance of LLMs in professional and high-risk domains such as the military, the models must be robust and trustworthy. I.e., knowing if a model's output is factually correct according to the task it is trained on and is not *hallucinating* (generating nonsensical or unfaithful content) [29]. To this end, a significant effort is put into analysing the causal reasoning capabilities of LLMs. LLMs outperform state-of-the-art causal reasoning algorithms in important reasoning topics such as counterfactual inference and graph discovery, making remarkably few mistakes (albeit important ones) while continuing to improve [30].

Reinforcement learning

Reinforcement learning (RL) is a machine learning paradigm where agents learn optimal decision-making strategies (c.f. policies) through interaction with an environment. RL has shown significant progress in the last decade, demonstrated by the evolution of game-playing agents able to master complex strategic games such as Go or Starcraft II. Although initial advances in hardware computing have played a large role in the progress of RL, nowadays, many innovations focus on learning efficiency through improved learning strategies. Efficiency is improving yearly by several orders of magnitude, incrementally building upon learning enablers such as short-term and episodic memory, exploration strategies, and meta-learning [31].

RL techniques can support *sensemaking* and *planning* processes due to their strategic and tactical planning capabilities. Related research in the military domain has demonstrated RL for COA optimizations in battle situations [21], [22]. Current demonstrations take place in austere environments compared to the intracity and complexity of real-world battle environments. Though with the prospect of above improvements in learning efficiency, demonstrations in increasingly complex environments are expected, where AI-driven strategies can be used as blueprints for actual military engagements.

Besides environment complexity, RL algorithms commonly struggle to find policies in environments with sparse rewards. Solutions that may seem easy to judge by humans (e.g. through common sense or intuition) can be challenging to grasp for machines. To address this problem, RL with human feedback (RLHF) has recently been proven to be highly effective in improving RL systems [32], [33]. RLHF uses human-generated feedback to allow learning from human expertise. This approach has already succeeded in domains where defining explicit reward functions is challenging, such as language generation, game-playing, and autonomous driving [33]–[36]. From a military perspective, such an approach benefits strategic planning problems by using the experience of expert decision-makers and subject matter experts.

Semantic graphs

A semantic graph is a graphical model that captures semantic relationships between concepts in a graph-like structure. A knowledge graph (KG) is a specific type of semantic graph used for knowledge representation. KGs are useful for various application fields such as question answering-systems (QnA), recommender systems, information retrieval, and data governance and analytics, being applied in domains such as medical, cyber security, finance or journalism.

KGs are well suited to support the *sensemaking* process through their ability to translate, organize and structure data into information and knowledge. In the military domain, they can represent battlefield situation awareness and support C2 [37]. They assist in multi-source intelligence analysis, allowing effective visualizations, searching and querying of information. At the same time, operators can derive meaningful conclusions and make more informed decisions in response to evolving situations.

More recently, Graph Neural Networks (GNNs) have gained much traction in the deep learning community, where a prime application domain is the decision sciences [38]. GNNs are deep learning models designed to work with graph-structured data, like KGs. These models help to predict missing information, classify and cluster knowledge, and enhance reasoning capabilities [39]. Furthermore, generative AI models for graphs have made significant progress in recent years, transforming the field of network analysis and prediction. These generative models, which often rely on GNNs, allow the generation of realistic and diverse graph samples to represent real-world intricate network patterns and have demonstrated potential in capturing complex relationships and structures in social networks and biological systems [40]. Such models can provide a powerful toolset for military operators to assist in developing scenarios and plausible adversarial strategies, enhancing war gaming exercises and simulation-based training.

Foundation models

The evolution of foundation models represents a pivotal advancement in AI. Foundation models are large-parameter machine learning models pre-trained on extensive datasets to succeed in a range of different tasks. Their adaptability and generalization capabilities enable customization for specific tasks through a process called transfer learning. Foundation models go beyond traditional task-specific approaches and can be used as a ‘general-purpose’ starting point for developing specialized applications. They have the potential to automate processes and supporting decision-making in different industries and sectors, such as healthcare, education, law, robotics, and the military [41].

In military operations, foundation models offer significant promise, especially when dealing with the inherent challenges of limited and unlabelled data. This scarcity of data can hinder the effectiveness of conventional machine learning methods that rely on extensive labelled datasets. Foundation models provide an innovative solution to this predicament. Leveraging their pre-trained knowledge, they can extract insights from external data sources and apply this to learning new, data-scarce tasks with minimal adaptation. This ability empowers foundation models to excel in decision-support tasks for the military, offering actionable insights and informed recommendations even in situations where traditional machine-learning approaches would struggle. Military decision-makers can utilize these models as powerful tools for generating operational strategies, predicting outcomes, and optimizing resource allocation, enhancing the quality and effectiveness of military operations. However, there are also risks involved. Due to the emergent properties of these models, there is a lack of understanding of how they work, what they are capable of, and when they fail [42]. Furthermore, bias in these models propagates to downstream applications built on them.

4.0 USE CASE

This section describes a use case example of a military decision support capability in the domain of tactical air command and control (C2). In this domain, the role of a Battle Manager provides a C2 capability to airborne and ground units and is responsible for overseeing and coordinating various aspects of a military operation. It includes activities such as developing mission plans, allocating resources, managing an effective use of airspace, identifying potential threats and response strategies, maintaining rules of engagements, tasking units, and assessing battle damage.

Decision support system

We formulate an example as a tactical decision support system (DSS) that can support the Battle Manager in analysing and developing different (friendly or enemy) COAs. Such a support system can develop response strategies for (actual, predicted or hypothetical) arising threat situations, forming COAs and operational plans for offensive and defensive military response by units in the field. Figure 4-1 illustrates the context of such a DSS. The left side shows an example sketch of a human-machine interface (HMI) for the operator, including a COP and a decision aid interface. The COP is used to display the real-time battlefield picture, annotated with various analytical insights, in this instance, for an incoming threat situation. The decision aid shows proposed COA options for which unit is most suited to engage the incoming threats, including argumentations and the option to effectuate a COA. In the context of the C2 process described in section 2, the DSS here would support the sensemaking process using AI-based models such as those described there. Initial concepts for this example of decision support have been developed in collaboration with military operators, though a detailed description is not within the scope of this paper. The right side of the figure sketches the application scope in which the DSS can be deployed, which we describe next.

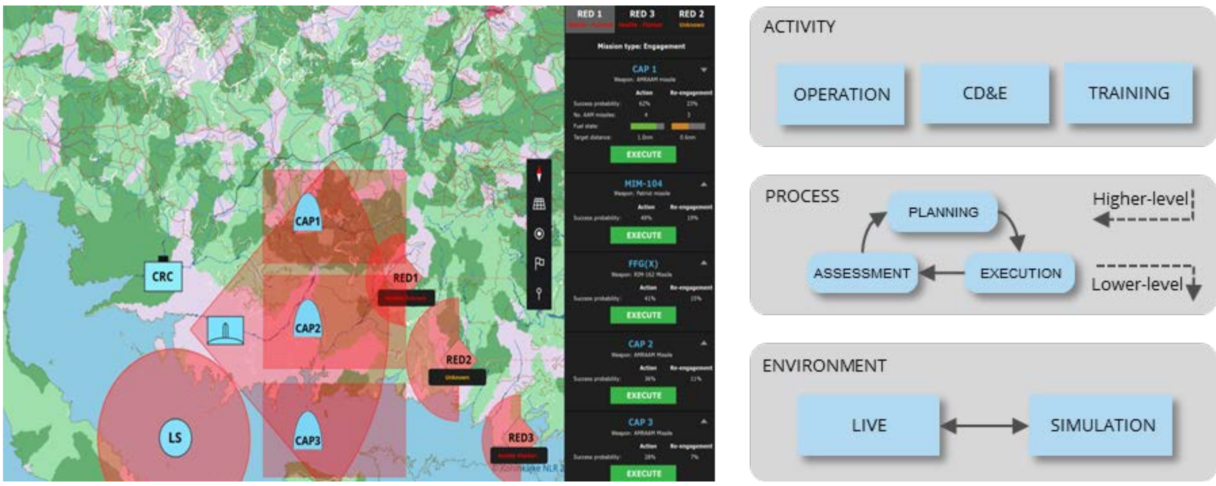

Figure 4-1: Tactical decision support system with an example interface sketch on the left and various application contexts on the right.

Application scope

The sketched DSS with decision aid technology for COA development is not limited to being only used during live mission operations and can be employed across activities of *mission operation*, *CD&E*, and *personnel training* [8]. First, in mission operation, it can be integrated into mission *planning*, e.g. to develop, analyse and compare different game plans based on what-if scenarios; into *execution*, e.g. to track COAs as they are unfolding, updating game plans based on new potentially unexpected situations, and proposing adaptations to in-mission plans; or into *debrief*, e.g. to collect lessons learned based on assessments of executed COAs, or based on what-if scenarios performed in hindsight for pivotal situations. Second, in CD&E, the technology can be used as a valuable support tool for experimentation through simulation in Battlelabs, e.g. for supporting the development of new doctrine or tactics; or for testing and evaluating the capabilities of new or future platforms or weapon systems in simulated mission environments. Finally, the capability could address a training gap, recognized for mission training at the higher tactical (command) level [8]. Trainees can train on tactical decision-making through wargaming activities through simulation-based training supported by machine-based COA analysis and development.

These examples illustrate that the decision aid technology applies across different military activities and contexts and is usable in both live and simulated mission environments. In the next section, we zoom in on the role of mission simulation in the DSS and discuss some key considerations for integrating and exploiting AI technologies herein.

5.0 AI-ENABLED MISSION SIMULATION

In the DSS sketched in the previous section, mission simulation plays a two roles. On the one hand, the simulation is used as real-time human-in-the-loop environment where the human operator can observe the environment and interact with military units that can represent live, virtual or constructive entities. On the other hand, the DSS uses the simulation environment to find, assess and generate solutions to decision problems using AI algorithms. Numerous M&S platforms are capable of mission simulation to fulfil the first role. For instance, government- or commercial-off-the-shelf Computer Generated Forces (CGF) platforms are broadly used in military training or Battlelab environments to simulate intricate mission environments. However, many of such platforms do not provide adequate flexibility to fulfil the second role. In this section, we describe the rationale for this.

Figure 5-1 depicts a reference model for a DSS capable of AI-enabled mission simulation. It shows a human operator supported in its decision-making process by various decision support services in a shared mission environment. These services may use a variety of AI models to facilitate sensemaking and planning problems. We highlight a set of considerations for operationalizing such AI models, focusing on the role of mission simulation. These considerations relate to the AI's (1) context of operation, (2) data requirements, (3) training environment, and (4) operational environment.

Context of operation

Deployed AI models address a particular problem in the mission environment, such as analysing, classifying, predicting or optimizing behavioural activities, for particular forces, groups or individuals. Designing, training and deploying these models successfully is always done in a particular *mission context*. This mission context provides the requirements, constraints and assumptions for the models and the environment they need to operate in. A context pertains to e.g. (1) expected geographical locations, conditions and threat types and behaviours that can be encountered, (2) rules and constraints on blue force's operation, such as following specific doctrine, operating procedures, rules of engagements, or other rules derived from mission planning, and (3) knowledge about the adversary's operational capabilities that may be known through Intel or lessons learned from previous operations.

It is vital to consider this contextual information when developing and training AI models. As it can contain dynamic, mission-specific elements, models may require re-training or fine-tuning to make them fit for purpose and tailored to a particular mission. Understanding the context and scope (and thus potential limitations) of AI models is crucial for an operator in judging the value of the outcomes of the models.

Mission simulation can be used here to reflect a representative operational environment scope for AI models during their training loop. To be an accurate representation, a simulated environment must reflect the mission context through the simulation of military entities (i.e. CGFs), equipped with appropriate physical and behavioural models that can operate in the required environment. In M&S systems, this context of the mission environment can be made available through existing or predefined scenarios and CGF models. However, developments in generative AI modelling will allow for more automated methods of generating appropriate mission environments and behavioural activity, fed by existing data and knowledge about the mission context.

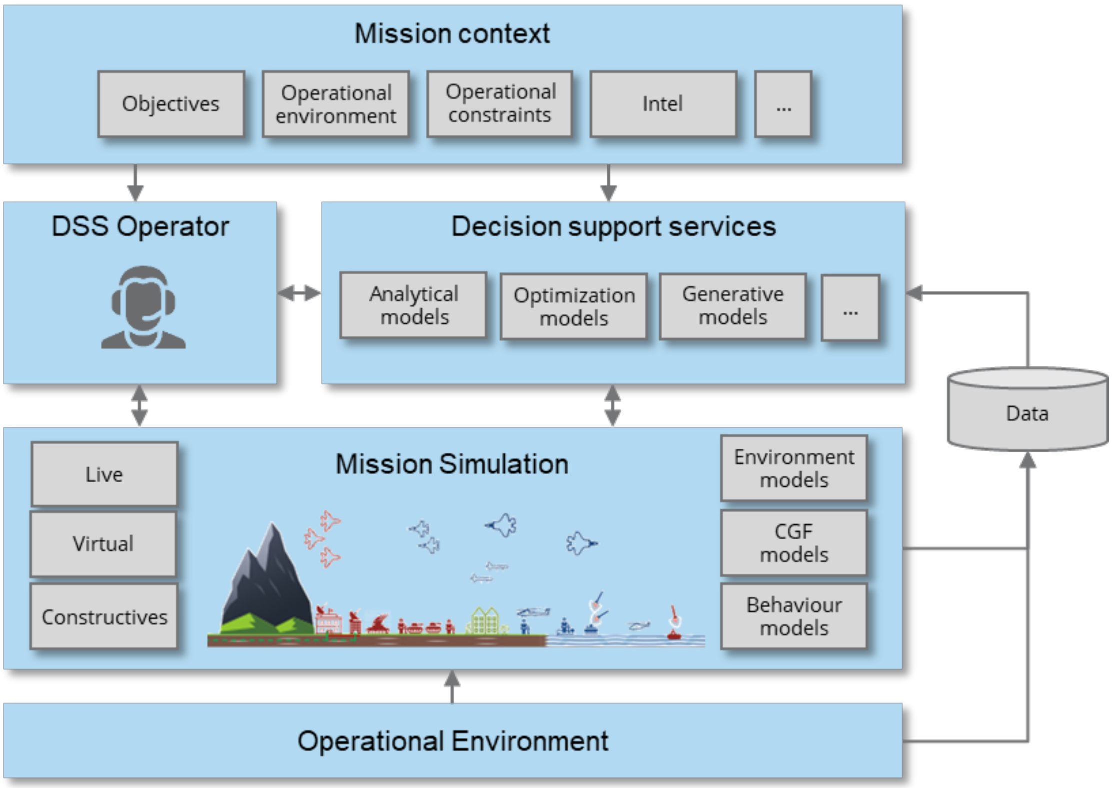

Figure 5-1: AI-enabled mission simulation in the context of a DSS

Data requirements

Many AI models rely on the availability of large datasets. For instance, consider machine learning-based classification or prediction models for behaviour analytics, or data-driven behaviour modelling approaches such as behaviour cloning or generative AI. In the military domain, behavioural data is often a scarce resource. Historical data on battle scenarios is limited, while collected training or operational data is often treated as sensitive and classified information. A common approach in AI for data-scarce problems is using data augmentation using synthetically generated data.

Mission simulation is ideally suited for this role. Through the simulation of constructive entities, large amounts of (labelled) time-series datasets can be generated for entity behaviours in various mission scenarios, battlefield configurations and environmental conditions. Fit for purpose synthetic data generated through mission simulation can be used to improve the quality of behaviour models through transfer learning techniques or the augmentation of real-world human data.

AI training environment

Certain categories of AI models require an interactive training environment. For instance, RL algorithms explore the consequences of actions in a learning environment in a trial-and-error fashion, or search algorithms such as Monte-Carlo tree search (MCTS) play out simulation runs to explore the effectiveness of different courses of action and the utility of certain environment states.

Mission simulation is suited to provide a development and training environment for such AI models. However, adopting the role of an AI training environment comes with strong demands for the configurability and computational capabilities of the simulation. On the one hand, the environment should be algorithmically configurable to establish a high-value training environment through scenario generation and adaptation techniques, required to facilitate smart learning strategies or heuristic searches for AI models. On the other hand, strong computing abilities are needed to facilitate the training of AI models in a reasonable time span [6]. Training time is especially relevant when AI models need updating for operations in new mission contexts. Computing abilities include faster-than-real-time (FTRT), headless execution, high-performance computing (HPC) and parallelization methods. Although innovations in AI algorithms continue to improve training efficiency, they are still sample-inefficient compared to human learning.

Digital twin environment

When AI models are used in deployment for real-time decision support, their inputs should reflect the current state of the operational environment. The mission simulation can reflect this state when it can act as a digital twin of the environment [43]. A digital twin environment is continuously updated with real-time data and enables AI capabilities for streaming analytics or real-time (re-)planning.

When mission simulation is used as a training environment and operational environment for AI models, the representational gap between the AI's source and target domain can be minimized, leading to more efficient transfer and interoperability between these environments. Many M&S systems, like those with live-virtual-constructive (LVC) capabilities, can function as a digital twin environment. However, there may be a discrepancy between the level of detail represented in the digital twin environment and the level required by different AI models, depending on whether they operate more on the engineering, engagement, mission or campaign level. In order to bridge this gap, one can benefit from M&S systems capable of multi-resolution simulation [44].

Concluding

These considerations for AI-enabled mission simulation originate from our experience and lessons learned from various former and current Defence research studies. As the global paradigm in state-of-the-art methods for (predictive) task modelling shifts from traditional to AI-based (deep learning) methods, it is pertinent to adopt a readiness for theoretical learning and practical adaptation. This last point has proven difficult in some of the more established military M&S platforms that originally have a strong basis for traditional (non-AI) modelling, although increased awareness can be seen throughout the industry.

Thus, we aim to provide insight into some of the technical aspects of the 'AI-readiness' of military M&S platforms, addressing the crucial part simulation systems play in future-proofing support capabilities. Still, enabling systems to support AI tools does not directly lead to development efficiency as expenditures to develop data- and model-training pipelines require innovative solutions. Nevertheless, the improvements state-of-the-art AI models will bring are particularly beneficial.

6. CONCLUSION

In this paper we discussed the roles and interdependencies of AI and M&S technologies in the context of military decision support. If the first half of the paper we started with an analysis and review of military decision support in a C2 context (section 2), followed by a horizon scan of AI technologies, reflecting on their potential role and impact for decision support (section 3). In the second half of the paper we focused on a use case example of a tactical DSS and discussed how decision aid technologies can be used in a variety of military activities and contexts (section 4). This was followed by an assessment of AI-enabled mission simulation where we discussed considerations for the operationalisation of AI in M&S systems (section 5).

The accelerating developments in AI hold great promise for military decision support, providing many opportunities, but also posing challenges and considerations for operationalization. In this paper we focused on one of these considerations, namely the synergy with, and role of (mission) simulation. One the one hand, simulation supports AI development by offering a data source and training environment, able to reflect military missions and activities. On the other hand, AI supports simulation development through advances in generative AI where representative scenarios, environments and simulation (behaviour) models can be generated. Finally, the simulation as a digital twin provides a conduit between AI models and real-world military operational environment.

There is much active research on other aspects of AI operationalization in the military domain that we did not address. Consider for example themes such as robustness, verification and validation, transparency and explainability, data quality and governance, and ethical and legal aspects. Developing and successfully operationalizing fit for purpose AI for military decision support requires a holistic approach where all such themes need to be considered when AI models are integrated in DSSs. A plausible evolution of more intelligent DSSs is the incremental development and integration of AI models that offer specific analytical, decision aid, and teaming functions, hereby advancing from more isolated, pragmatic levels of support to more interconnected, conceptual levels of support where a shared understanding of the decision problem and environment can be established between man and machine. We believe that M&S systems will play an important role in this development, supporting the experimentation, validation and evaluation of DSS technology for military operators in current and future operational environments.

REFERENCES

[1] A. MacAllister, P. Rupp, J. Garrison, G. Hellstern, D. Javorsek, and P. Chu, “Using Machine Learning for Battle Management Analysis,” in *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, I/ITSEC, 2021.

[2] A. Kaklauskas, “Intelligent Decision Support Systems,” in *Biometric and Intelligent Decision Making Support*, Cham: Springer International Publishing, 2015, pp. 31–85. doi: 10.1007/978-3-319-13659-2_2.

[3] J. Schubert, J. Brynielsson, M. Nilsson, and P. Svenmarck, “Artificial intelligence for decision support in command and control systems,” in *Proceedings of the 23rd International Command and Control Research & Technology Symposium: Multi-Domain C2*, 2018, pp. 18–33.

[4] K. Van Den Bosch and A. Bronkhorst, “Human-AI cooperation to benefit military decision making,” NATO, 2018.

[5] US Army Combined Arms Center, “Synthetic Training Environment (STE) White Paper.”

[6] J. van Oijen and A. Toubman, “Are We Machine Learning Yet? Computer Generated Forces with Learning Capabilities in Military Simulation,” in *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, I/ITSEC, 2021.

[7] P. Kerbusch, B. Keijser, and S. Smit, “Roles of AI and Simulation for Military Decision Making,” in *Proceedings IST-160 Big Data & Artificial Intelligence for Military Decision Making*, STO, 2018.

[8] Specialist Team MSG-189, “AI Augmented Immersive Simulation in Training and Decision Making

Course of Action Analysis," NATO STO.
[9] G. Long, "Building a World With Deepfake Content - Who Needs Real Data," in *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*, I/ITSEC, 2022.
[10] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," *arXiv preprint arXiv:2304.03442*, 2023.
[11] S. Kapturowski *et al.*, "Human-level Atari 200x faster." 2022.
[12] "Palantir AIP." www.palantir.com/platforms/aip/
[13] B. Brehmer, "Command and control as design," *Proceedings of the 15th International Command and Control Research and Technology Symposium*, 2010.
[14] E. Jensen, "Sensemaking in military planning: a methodological study of command teams," *Cognition, Technology & Work*, vol. 11, pp. 103–118, 2009.
[15] I. H. Sarker, "Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective," *SN Computer Science*, vol. 2, no. 5, p. 377, 2021.
[16] N. M. de Reus, P. Kerbusch, M. Schadd, and M. Ab de Vos, "Geospatial analysis for machine learning in tactical decision support," STO-MP-MSG-184. NATO, 2021.
[17] S. Ahlberg *et al.*, "An information fusion demonstrator for tactical intelligence processing in network-based defense," *Information Fusion*, vol. 8, no. 1, pp. 84–107, 2007.
[18] G. K. Sewnath and J. A. Crijnen, "How much is too much? Levels of A.I. Explainability within Decision Support System's User Interfaces for improved decision-making performance," in *Proceedings MSG-184 Towards Training and Decision Support for Complex Multi-Domain Operations*, 2021.
[19] U. D. of the Army, "ADP 5-0: The operations process." 2018.
[20] N. MSG-124, "Developing Actionable Data Farming Decision Support for NATO," 2018.
[21] M. Sommer and M. Rüegsegger, "Deep Self-optimizing Artificial Intelligence for Tactical Analysis, Training and Optimization," in *NATO NMSG Symposium: Towards Training and Decision Support for Complex Multi-Domain Operations*, Amsterdam, the Netherlands, 2021.
[22] L. T. Doll, M. Behm, J. Brendecke, and D. Kallfass, "From the Game Map to the Battlefield–Using DeepMind's Advanced Alphastar Techniques to Support Military Decision-Makers," in *NATO NMSG Symposium: Towards Training and Decision Support for Complex Multi-Domain Operations*, Amsterdam, the Netherlands, 2021, pp. 14–1.
[23] M. A. Gallagher, D. J. Caswell, B. Hanlon, and J. M. Hill, "Rethinking the hierarchy of analytic models and simulations for conflicts," *Military Operations Research*, vol. 19, no. 4, pp. 15–24, 2014.
[24] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, "Ten challenges for making automation a" team player" in joint human-agent activity," *IEEE Intelligent Systems*, vol. 19, no. 6, pp. 91–95, 2004.
[25] J. Johnson, "Automating the OODA loop in the age of intelligent machines: reaffirming the role of humans in command-and-control decision-making in the digital age," *Defence Studies*, vol. 23, no. 1, pp. 43–67, 2023.
[26] W. X. Zhao *et al.*, "A Survey of Large Language Models." arXiv, Jun. 29, 2023. Accessed: Aug. 22, 2023. [Online]. Available: http://arxiv.org/abs/2303.18223
[27] S. Yin *et al.*, "A Survey on Multimodal Large Language Models." arXiv, Jun. 23, 2023. Accessed: Aug. 24, 2023. [Online]. Available: http://arxiv.org/abs/2306.13549
[28] Wikipedia, "AI Alignment." https://en.wikipedia.org/wiki/AI_alignment
[29] J. Kocoń *et al.*, "ChatGPT: Jack of all trades, master of none," *Information Fusion*, vol. 99, p. 101861, Nov. 2023, doi: 10.1016/j.inffus.2023.101861.
[30] E. Kıcıman, R. Ness, A. Sharma, and C. Tan, "Causal Reasoning and Large Language Models: Opening a New Frontier for Causality." arXiv, May 08, 2023. Accessed: Aug. 17, 2023. [Online]. Available: http://arxiv.org/abs/2305.00050
[31] DeepMind, "Agent57: Outperforming the human Atari benchmark," Sep. 2023. https://www.deepmind.com/blog/agent57-outperforming-the-human-atari-benchmark
[32] W. Bradley Knox and P. Stone, "TAMER: Training an Agent Manually via Evaluative Reinforcement," in *2008 7th IEEE International Conference on Development and Learning*,

Monterey, CA: IEEE, Aug. 2008, pp. 292–297. doi: 10.1109/DEVLRN.2008.4640845.
[33] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep Reinforcement Learning from Human Preferences," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Aug. 24, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html
[34] D. M. Ziegler *et al.*, "Fine-Tuning Language Models from Human Preferences." arXiv, Jan. 08, 2020. doi: 10.48550/arXiv.1909.08593.
[35] N. Stiennon *et al.*, "Learning to summarize with human feedback," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2020, pp. 3008–3021. Accessed: Aug. 24, 2023.
[36] D. Cohen *et al.*, "Dynamic Planning in Open-Ended Dialogue using Reinforcement Learning." arXiv, Jul. 25, 2022. doi: 10.48550/arXiv.2208.02294.
[37] X. Guo, L. Chen, Y. Li, Q. Wang, and W. Guan, "A Review of Knowledge Graph Research in Military Domain," in *Proceedings of the 10th International Symposium on Computational Intelligence and Industrial Applications*, 2022.
[38] A. Keramatfar, M. Rafiee, and H. Amirkhani, "Graph Neural Networks: A bibliometrics overview," *Machine Learning with Applications*, vol. 10, p. 100401, 2022
[39] Z. Ye, Y. J. Kumar, G. O. Sing, F. Song, and J. Wang, "A Comprehensive Survey of Graph Neural Networks for Knowledge Graphs," *IEEE Access*, vol. 10, pp. 75729–75741, 2022
[40] Y. Zhu *et al.*, "A Survey on Deep Graph Generation: Methods and Applications." 2022.
[41] S. Yang, O. Nachum, Y. Du, J. Wei, P. Abbeel, and D. Schuurmans, "Foundation models for decision making: Problems, methods, and opportunities," *arXiv preprint arXiv:2303.04129*, 2023.
[42] R. Bommasani *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.
[43] K. Aggarwal *et al.*, "Enabling Elements of Simulations Digital Twins and its Applicability for Information Superiority in Defence Domain," in *Proceedings of the MSG-197 Symposium: Emerging and Disruptive Modelling and Simulation Technologies to Transform Future Defence Capabilities*, 2022.
[44] K. Lee, G. Lee, and L. Rabelo, "A systematic review of the multi-resolution modeling (MRM) for integration of live, virtual, and constructive systems," *Information*, vol. 11, no. 10, p. 480, 2020.