Acquiring Maintainable AI-Enabled Systems

MAJ Iain Cruickshank and MAJ Shane Kohtz

Army Cyber Institute, United States Military Academy

iain.cruickshank@westpoint.edu and (845) 938-7566

shane.kohtz@westpoint.edu and (845) 938-9657

Major Iain Cruickshank – is a Functional Area 49 (Operations Research/Systems Analysis) officer in the U.S. Army. He is currently a senior research scientist at the Army Cyber Institute. He has previous assignments with the Army's Artificial Intelligence Integration Center, the 780th Military Intelligence Brigade, and the 101st Airborne Division. He holds a Ph.D. in Societal Computing from Carnegie Mellon University, which was obtained as a National Science Foundation Graduate Research Fellow, and an MS in Operations Research from the University of Edinburgh, which was obtained as a Rotary Ambassadorial Scholar.

MAJ Shane Kohtz – is a Functional Area 51 (Acquisition) officer in the U.S. Army. He is currently a cyber research manager at the Army Cyber Institute. He has previous assignments with the Missile Defense Agency, the Army's Program Executive Office Intelligence Electronic Warfare & Sensors, the 1st Infantry Division, and the 101st Airborne Division. He holds a MBA from the Naval Postgraduate School with a focus in Systems Acquisition Management. He is a member of the Army Acquisition Corps and holds a DAWIA Advanced certification in program management.

Abstract

The Army and other services are quickly entering into an age where many, if not all, acquisitions programs will need to contend with acquiring Artificial Intelligence (AI)-enabled systems. While there has been research on how to acquire the data or model for an AI-enabled systems, sustainment considerations have been overlooked. Given the importance of sustainment for any acquisition program of record – both in terms of cost and in terms of program effectiveness – it is imperative that the Army, and the rest of DOD, plan for AI-enabled system maintenance. To address this gap, this paper proposes a framework and practices that draw on best practices from industry, program maintenance, and Machine Learning Operations (MLOps) to integrate AI maintenance into a product support strategy and Life Cycle Sustainment Plan. The framework outlines necessary components for sustainable AI and considers varying levels of maintenance to reduce operation and sustainment costs.

Introduction

Technology on the battlefield will increasingly need to become data centric and automated to have a tactical advantage over adversaries' technologies; AI will be an integral part of future warfare (NSCAI, 2021). The United States Department of Defense (DOD) primary solution to this capability gap is a significant investment into Artificial Intelligence (AI) and, AI's primary driver, Machine Learning (ML). For example, in preparation for fiscal year 2023, the Department of Defense requested $1.1 billion to further research and development of the immature AI and ML technology (Department of Defense [DOD], 2022). AI will be part of many future systems that we will acquire and upgrade; by 2045 it will probably be a standard component of every major piece of military equipment (NSCAI, 2021). As these technologies mature, and are incorporated into systems and programs, they then need to be maintained. While the defense acquisition community has started considering data (Nagy, 2022), use cases (Guariniello, 2021), and hardware for AI-enabled systems, there is little to no thought on how the

sustainment of these AI-enabled systems will work for major programs. Thus, while the DOD has invested heavily into maturing AI and ML for future AI-enabled systems, its less clear how the defense acquisition community could maintain and sustain these AI-enabled systems.

This paper proposes a paradigm, along with recommendations for program offices, to utilize when planning the acquisition strategy of an AI-enabled program of record. We first outline the importance of maintenance planning in a program and why AI-enabled systems need maintenance. We then discuss the main considerations in planning for the maintenance of an AI-enabled system. These maintenance considerations are necessary to inform the strategy to meet sustainment requirements known as the Product Support Strategy (PSS) and Life Cycle Support Plan (LCSP) for a program of record (Office of the Under Secretary of Defense for Acquisition and Sustainment [OUSD(A&S)], 2021). The early planning for the acquisition logistics strategy prevents the possibility of a program breach or uncaptured costs later in the program. AI-enabled systems will become more prevalent on the future battlefield while the sustainment planning occurs now.

Background

Maintenance planning in a program of record. Maintenance is one of the most critical aspects of a major acquisitions program. Maintenance considerations occur early in the lifecycle of a program of record, and early sustainment decisions have a long-term effect during the operations and sustainment phase of a program (Department of Defense [DOD], 2016). Why is sustainment planning important early in the acquisition lifecycle? The acquisition community has known for years that operation and sustainment costs account for the majority of a program's total ownership costs; in fact, 72% of the total ownership costs occur during the program's operation and sustainment phase (Schinasi, 2003). Figure 1 below illustrates how a program costs are distributed across an acquisition program's lifecycle. Operation and sustainment planning slightly improved in recent years. The O&S Cost Management Guidebook stated, "in the December 2014 Selected Acquisition Reports (SARs), on average, 67% of the reported costs are attributable to O&S" (DOD, 2016). Despite the slight improvement, most costs for a program remain during operations and sustainment.

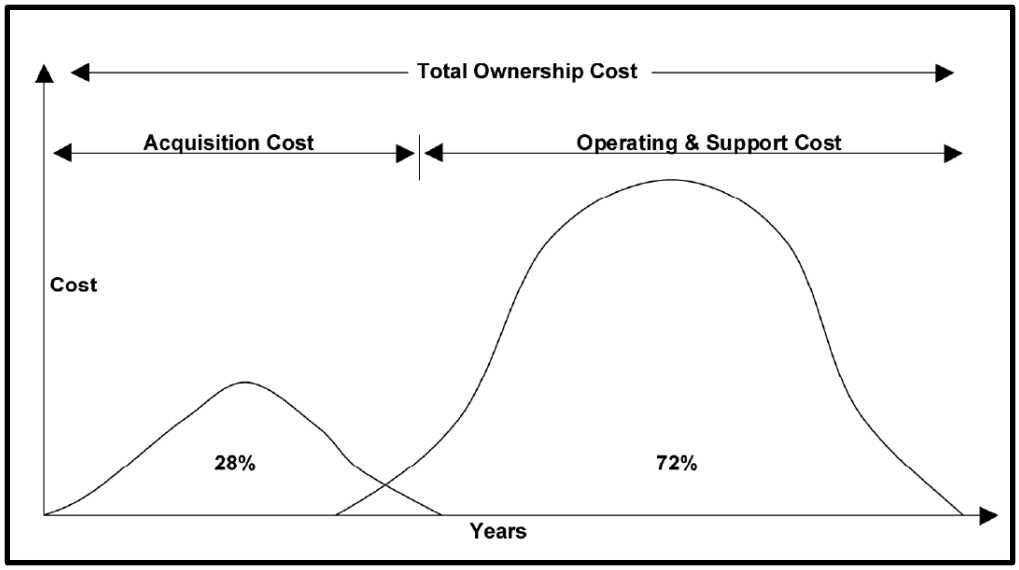

Figure 1. Nominal Life-Cycle Cost of Typical DOD Acquisition Program with a 30-Year Service Life. Source: Schinasi (2003).

In addition, when requirements are approved, nearly 85% of operation and sustainment costs are known with less than 10 % of the life-cycle costs spent (Schinasi, 2003). Figure 2 illustrates the importance of early planning with systems for AI/ML requirements. AI/ML capable systems are early in the technology maturation process with substantial investments, but the majority of sustainment costs are already determined. Program offices must proactively plan and determine the Product Support Strategy (PSS) at program inception and then the Life Cycle Sustainment Plan (LCSP) at the first acquisition milestone, Milestone A, even though the sustainment of AI enabled systems may be unknown currently (OUSD[A&S], 2021).

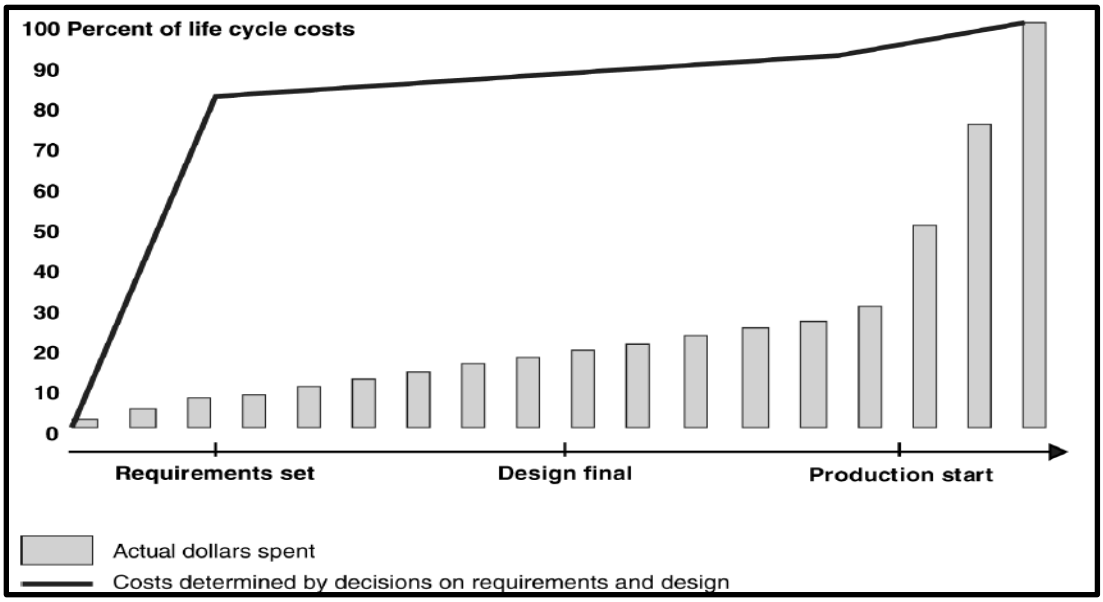

Figure 2. Percent of Operating and Support Costs Determined at Various Points in the Acquisition Process. Source: Schinasi (2003).

AI-enabled systems and their maintenance. AI-enabled systems, like any other piece of technology, require maintenance. An AI-enabled system consists of traditional software and, possibly, hardware, depending on the purpose of the system in addition to AI components of the system. AI components often require several hardware and software dependencies, often called a stack (Moore, 2018). Figure 3 illustrates the AI stack. One of the critical elements of the AI components, and, really, what makes the entire system an AI-enabled system are the ML models. The ML models enable the system to engage in automated behaviors and activities that typically require human levels of perception or reasoning; they are the 'brain' of the AI-enabled system. These ML models, much like every other component of the AI-enabled system, also require maintenance.

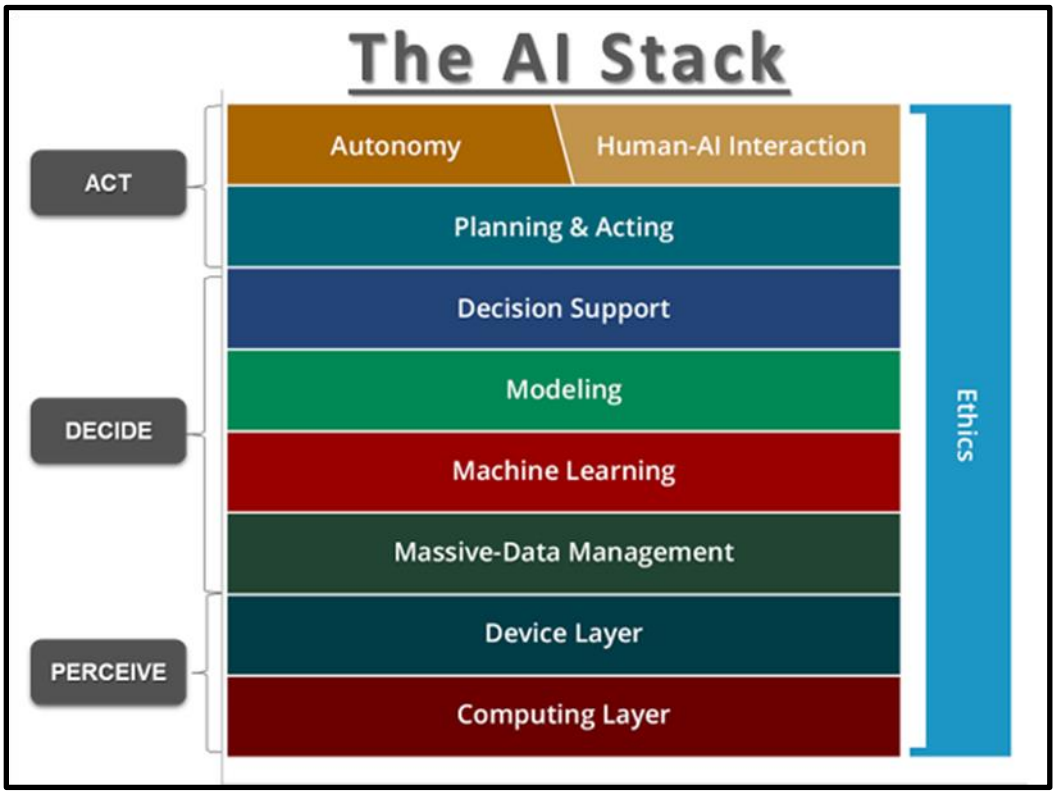

Figure 3. Carnegie Mellon University's AI stack, depicting the necessary components of an AI-enabled system. Source: Moore (2018).

ML models, despite their potential, still suffer from several issues that necessitate frequent maintenance. ML models, by their nature, learn correlations useful to a certain task from the data that is presented to them. Thus, these models could have performance issues if the data presented to the model when in use is different than the data it was trained on (i.e. Out-of-Domain Data problem) (Patruno, 2019). As an example of this, a computer vision ML model, which is meant to detect certain vehicles from a ground perspective, can fail when something as simple as the background, or biome, is different between the model's training data and where the model is used

(e.g urban versus rural setting). ML models can also suffer from issues like model drift (Talby, 2018), data drift (Evidently AI, 2021), concept drift (Patruno, 2019), or even changing of hardware, like sensors, which all greatly affect ML model performance. In addition to those issues which naturally arise, ML models can also be directly attacked via Adversarial ML, which will also seriously degrade ML model performance (Talby, 2018). Finally, it should be noted that many of these issues are unique to ML and ML-enabled systems; changing of something like the background of images does not affect the hardware or software of a traditional, digital system. Thus, ML models have their own inherent issues which necessitate maintenance for those ML models, which over and above the maintenance for traditional hardware and software systems.

While ML models suffer from several issues, which can greatly affect their performance, dealing with these issues frequently requires far less resources and know-how than the initial development of the ML model. Maintaining ML models in use in the real-world (i.e. model deployment) can often be handled with a collection of updating and monitoring processes, which are collectively part of the industrial ML paradigm of MLOps (Treveil, et al., 2020). MLOps, at its core, is a set of practices which aims to productionize ML systems (Treveil, et al., 2020). Figure 4 depicts the core components and relationships of MLOps. While the principles and practice of MLOps are still an active area of research, three practices that are a mainstay of MLOps are the monitoring of data and models in production, the continual updating of models in response to changes, and having model maintenance take place with model operation (Treveil, et al., 2020). These are an integral part of MLOps because they are how organizations and businesses can use ML models despite their inherent issues. Thus, key to the use of ML models in the real world and in production systems in the MLOps paradigm is having in place the right tools and practices to monitor an ML model and its data as well as the correct steps to update ML models, as close to operation as is feasible.

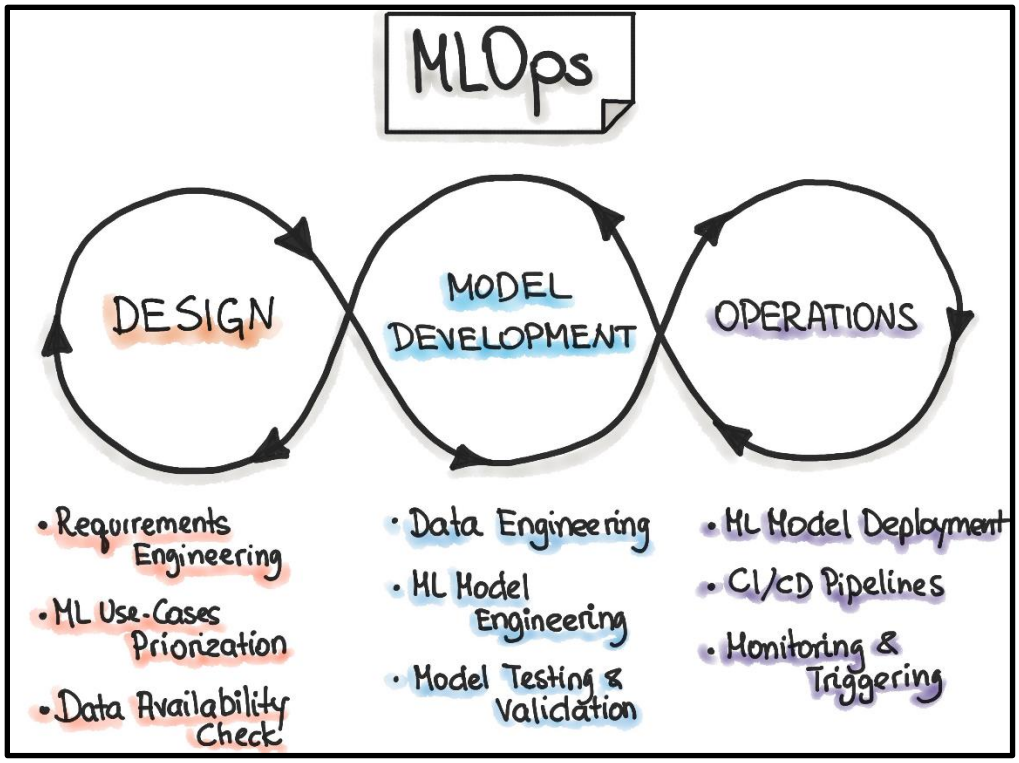

Figure 4. Core components of MLOps and their relationships. Source: Visengeriyeva (2023, March 30).

Of note in the MLOps paradigm is *model retraining*. Ideally, model retraining consists of running all of the steps required to train an ML model, but with a new dataset; model retraining should not require any changes to the code – just changes to the weights of the model (Patruno, 2019). This type of maintenance generally needs to occur anytime the data changes, and an updated training data set is available (Evidently AI, 2021). Thus, this type of maintenance generally comes in two forms, periodic and dynamic (Evidently AI, 2021). Periodic retraining is when there are known changes in the data that will occur, such as quarterly or yearly changes in business practices. Whereas dynamic retraining occurs any time there are changes in the data generation process, such as collecting in an adversarial environment (e.g. detecting credit fraud) or a naturally dynamic process (e.g. labeling objects in imagery). When it comes to dynamic retraining, it can occur on widely variant time scales depending on the ML application; some ML models need to be updated daily, while others need only be updated monthly or yearly (Evidently AI, 2021). Regardless of the frequency of ML model retraining, all experts on the subject of using ML models in the real world agree that this process is a must for any ML-enabled system. Thus, model retraining is a necessary part of any ML model and may need to occur daily.

Considerations for Maintaining an AI-enabled System

When it comes to taking AI-enabled maintenance into program planning, there are a couple of key considerations. These considerations should inform program offices when they perform a Product Support Business Case Analysis (PS BCA) that informs the PSS and LSCP (Department of Defense [DOD], 2022). The PS BCA evaluates potential alternatives for

sustainment to include organic, contractor, or a ratio mix of support that informs a decision on the program's sustainment strategy (Department of Defense [DOD], 2014). The PSS and LCSP are updated at each acquisition milestone; however, as highlighted earlier, nearly 85% of the sustainment costs are determined when requirements are set (Schinasi, 2003). An understanding of the requirements and maintenance "touch time" of AI/ML systems is imperative during the strategy development phase to properly plan and budget sustainment. This maintenance of ML models is in addition to all the hardware and software underlying the AI stack, which are necessary to run the ML models. Such a requirement can enable possible project scenarios wherein the ML model is a sub-product, or product-within-a-product, of a larger AI-enabled system. Overall, in addition to the maintenance requirements of software and any hardware, there are also requirements for the maintenance of the AI components that should address any intellectual property, data, and ML models.

Intellectual Property and Data. A critical component to the PS BCA, PSS and LCSP is a program's Intellectual Property (IP) Strategy for sustainment planning. DODI 5000.91 (Product Support Management for the Adaptive Acquisition Framework) states "the IP strategy identifies, and acquisition contracts should secure, sufficient technical data, manuals, and publications to enable informed Government decisions to acquire maintenance and repair through Government organic capability and/or contractor-provided solutions" (OUSD[A&S], 2021). The role of data rights is even more critical for AI enabled systems given the amount of maintenance required on a routine basis. Program offices may be unaware of the type of data required to conduct organic maintenance because AI is an emerging technology.

The Defense Federal Acquisition Regulation Supplement outlines government rights for data, which are unlimited rights, government purpose rights, or limited rights (General Services Administration [GSA], 2023). Program offices must understand these rights in acquisition planning and contract negotiation for AI/ML enabled systems. A recent RAND study noted that government program offices did not understand data rights, which had long term impacts on sustainment planning. Vendors would leverage the "proprietary" label and utilize court systems to maintain data rights in a weapon system for follow on sustainment. As a result, the government typically would not want to go through the elaborate court proceedings and thus acquiesce to the vendor's claims concerning data rights (RAND, 2021). The RAND case study highlights the importance of data rights when planning weapon system sustainment, and the lessons learned are imperative since AI-enabled systems require a substantial amount of touch time for maintenance.

ML Model Maintenance Considerations. There are a few different paradigms to approaching maintenance for ML models. Much like sustainment for other components of a system, the maintenance of an ML model can use both contract and organic service support alternatives. At the one end of the spectrum is the ML model maintenance being performed solely by contract. This means contractors would be responsible for all of the tasks of model maintenance including data and model monitoring, development of test and evaluation metrics, development of model retraining procedures, model updating (i.e. performing the model retraining procedures), model retirement and replacement, and model governance (i.e. making sure any ML model is meeting

necessary guidelines and regulations). A particular version of the contractor only approach in use is the ML-as-a-Service (MaaS) model. The MaaS model usually works through application programming interfaces (APIs), whereby the contractor has full responsibility for the model, to include initial development and maintenance, and a user just sends data to an API to use the ML model. This type of model is currently used by companies like OpenAI and by organizations like the XVIIIth Airborne Corps and often works on a pay-per-usage type of pricing scheme.

While the contractor-only approaches present the simplest approach to maintenance planning, they have serious pitfalls that must be considered. For the MaaS model, despite the simplicity of this model, much like any other pay-per-use pricing scheme (e.g. cloud services, SaaS), it can quickly become exorbitantly expensive if there is a lot of use of the service. Additionally, it requires connectivity back to the API to work. So, if the AI-enabled system is meant to work in austere environment or have a lot of usage on the ML-models, going through a MaaS model may be overly costly. Additionally, having contractors perform all the functions of ML maintenance ignores the hard-learned lessons behind the MLOps paradigm; namely the operation of the ML model has been separated from its maintenance and development. A primary reason why MLOps places the development and maintenance of ML models so close to the running of ML models is that these models require constant monitoring and frequent updating (Treveil, et al., 2020). In fact, one form of updating, model retraining, can occur as frequently as daily for an ML model in production in an adversarial and dynamic environment. As with our previous computer vision example of detecting objects from a ground perspective, the ML model would need to be, at a minimum, retrained every time the biome changes (e.g. moving from rural to urban) and every time an organization wants to detect a new or different set of objects. Conceivably, such a change in an ML model's operating environment could occur several times over the course of a single operation for a military unit. Thus, given the frequent nature of ML model maintenance, having contractors provide all this maintenance could be cost prohibitive.

At the other end of the spectrum is a service only solution, where servicemembers and DOD civilians are responsible for all of the aforementioned ML model maintenance tasks. While this certainly presents some potential for cost savings in terms of maintenance, the Army and DOD may lack the skill sets in house, in sufficient numbers, to perform some maintenance functions. This is especially true for maintenance functions like designing a test and evaluation scheme for both the ML model and its data as well as determining the right model retraining procedures (e.g. active learning, fine-tuning, etc.). These types of maintenance tasks often take a seasoned data scientist with domain area expertise and, often, advanced education. That said, some of the maintenance tasks actually require very little education and can be learned with suitable training. For example, actually performing model updates, given a guide to the model's retraining procedures, is trainable task that does not require an advanced educational background. Thus, planning to do the full spectrum of model maintenance in house may be infeasible, given constraints on in house ML expertise.

Conclusion & Recommendation

The acquisition of AI-enabled technologies that will be successful for military operations must have sustainment of their ML models taken into primary consideration. ML models have critical fragilities that require monitoring and updating. What is more the typical frequency of ML model retraining for dynamic and adversarial environments makes it prohibitive for this type of maintenance to be done by contractors. Fortunately, if an AI-enabled system is properly implemented, monitoring and retraining ML models can be a trainable task that can be performed in house. So, it is vital that we acquire AI-enabled systems that allow for this in house maintenance if that AI-enabled system is going to be useful for military operations. As such, we recommend a hybrid approach to ML model sustainment planning, that leverages expertise from contractors, but relies on servicemembers for execution of the maintenance. The following figure, Figure 5, details the sustainment tasks and which component should be responsible for them.

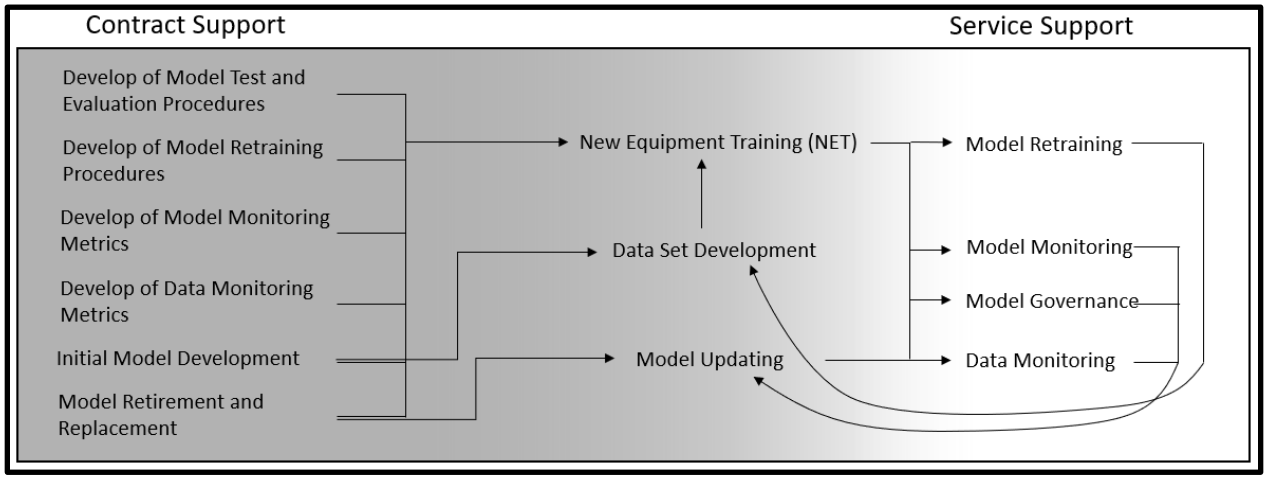

Figure 5. ML model sustainment tasks in a hybrid maintenance plan with associated dependencies between contractor and service maintenance tasks.

When it comes to the actual amount of effort expended on these maintenance tasks, those in the service support region are the equivalent of field-level maintenance (DOD, 2022). Those tasks are the ones most frequently done and the tasks that can address most issues with ML models in use. Whereas those within the contract support, namely model retirement and replacement, as well as some that are a shared task, like model updating, would be depot-level maintenance (DOD, 2022). These tasks should only be needed periodically and to address major issues with the ML model.

Along with our proposal of a hybrid maintenance model for AI-enabled systems, we also propose the following points be part of any program planning:

- *Data Rights*: program offices, looking to have ML models in their programs, may negotiate limited rights for implementation of the ML models since government operators would be doing the model retraining and monitoring. However, since the deliverables will most likely come from mixed funding, the program offices should, at a minimum, negotiate for government purpose rights of the technical data and deliverables. This

approach will give the program office flexibility in the future if they decide to change the sustainment strategy.

- *ML Model Touch-Time Analysis*: As has been mentioned within this paper, ML models, the brain of any AI-enabled system, require model retraining for various reasons. The amount of model retraining for any given ML model is highly context dependent; it can vary from daily retraining up to monthly or even yearly (Evidently AI, 2021). Thus, as part of the PS BCA, there needs to be a retraining requirements analysis. This analysis should, at a minimum, consider how often the data environment for the AI-enabled system predictably changes, whether it will be used in an adversarial environment (i.e. data environment where people generating the data attempt to change data generation patterns to fool the system), and how often the data generation process changes physical locations (i.e. a sensor moves from one geographic region to another). With the information from this analysis, a program office can have a much better estimation of the maintenance cost requirements. We also note that this type of analysis is fruitful grounds for future, impactful research.

In conclusion, as the Department of Defense invests heavily in emerging AI technology, the acquisition community must prioritize maintenance and sustainment considerations. Early and knowledgeable sustainment planning for a new technology such as AI and ML is imperative considering 85% of operation and sustainment costs are determined in the requirement development stage (Schinasi, 2003). This research proposes a new paradigm and provides a usable framework for the acquisition and sustainment strategy development of a maintainable AI-enabled system. ML models have critical fragilities that drive the need for substantial maintenance on AI-enabled systems. The proposed framework's maintenance considerations serve as a starting point for program offices to evaluate alternatives in the Product Support Business Case Analysis for informed decision-making on Product Support Strategy and Life Cycle Sustainment Plan. The necessary technical data, data rights, training, and a mix of organic and contractor maintenance support are important inputs when developing the Product Support Strategy. This research recommends a mixed sustainment strategy for contractor deliverables and depot-level maintenance while service members execute field-level maintenance for data monitoring and model retraining, monitoring, and governance. Future research can focus on maintenance touch time frequency in a complex operational environment to inform AI maintenance requirements further. Nonetheless, AI-enabled system maintenance sustainment planning is crucial and should start now.

The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

References

Camm, F., Whitmore, T. C., Weichenberg, G., Sheng, T. L., Carter, P., Dougherty, B., Nalette, K., Bohman, A., Shostak, M. (2021). Data Rights Relevant to Weapon Systems in Air

Force Special Operations Command (Report No. RR-4298-AF). RAND. https://www.rand.org/pubs/research_reports/RR4298.html

Department of Defense. (2014). *DOD Product Support Business Case Analysis Guidebook.* https://www.dau.edu/tools/Lists/DAUTools/Attachments/127/Product-Support-Business-Case-Analysis-(BCA)-Guidebook.pdf

Department of Defense. (2016). *Operating and Support Cost Management Guidebook.* https://www.dau.edu/tools/Lists/DAUTools/Attachments/126/Operating-and-Support-Cost-Management-Guidebook.pdf

Department of Defense. (2022). *Defense Budget Overview: United States Department of Defense Fiscal Year 2023 Budget Request.* https://comptroller.defense.gov/Portals/45/Documents/defbudget/FY2023/FY2023_Budget_Request_Overview_Book.pdf

Department of Defense. (2022). *Product Support Manager Guidebook.* https://www.dau.edu/tools/Lists/DAUTools/Attachments/129/Product-Support-Manager-(PSM)-Guidebook.pdf

Evidently AI. (2021, July 01). *When to Retrain an Machine Learning Model? Run these 5 checks to decide on the schedule.* Retrieved from KDNuggets: https://www.kdnuggets.com/2021/07/retrain-machine-learning-model-5-checks-decide-schedule.html

General Services Administration. (2023, March 1). *Defense Federal Acquisition Regulation Supplement (DFARS) 227.7103-5 Government rights.* https://www.acquisition.gov/dfars/227.7103-5-government-rights.

Guariniello, C., Balasubramani, P., DeLaurentis, D. (2021). A System-of-Systems Approach to Enterprise Analytics Design: Acquisition Support in the Age of Machine Learning and Artificial Intelligence. *Proceedings of the Nineteenth Annual Acquisition Research Symposium*, 205-217. https://dair.nps.edu/handle/123456789/4519.

Nagy, B. (2022). Tips for CDRLs/Requirements when Acquiring/Developing AI-Enabled Systems. *Proceedings of the Nineteenth Annual Acquisition Research Symposium*, 218-241. https://dair.nps.edu/handle/123456789/4587.

National Security Commission on Artificial Intelligence. (2021). *Final Report*. https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf.

Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD[A&S]). (2021, November 4). *Product Support Management for the Adaptive Acquisition Framework* (Department of Defense Directive 5000.91). Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500091p.PDF?ver=qk1slCU3Y0c1acIDocWyJA%3d%3d.

Moore, A., Hebert, M., Shaneman, S. (2018). The AI stack: a blueprint for developing and deploying artificial intelligence. *Proceedings of SPIE Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR IX*.

https://www.ri.cmu.edu/publications/the-ai-stack-a-blueprint-for-developing-and-deploying-artificial-intelligence/.

Patruno, L. (2019, June 10). *The Ultimate Guide to Model Retraining.* Retrieved from mlinproduction.com: https://mlinproduction.com/model-retraining/#:~:text=Quickly%20changing%20training%20sets%20might,require%20monthly%20or%20annual%20retraining.

Schinasi, K. V. (2003). *BEST PRACTICES: Setting Requirements Differently Could Reduce Weapon Systems' Total Ownership Costs.* Government Accountability Office.

Talby, D. (2018, June 5). *Lessons learned turning machine learning models into real products and services.* Retrieved from https://www.oreilly.com/radar: https://www.oreilly.com/radar/lessons-learned-turning-machine-learning-models-into-real-products-and-services/

Treveil, M., Omont, N., Stenac, C., Lefevre, K., Phan, D., Zentici, J., . . . Heidmann, L. (2020). *Introducing MLOps.* New York: O'Reilly.

Visengeriyeva, L., Kammer, A., Bär, I., Kniesz, A., Plöd, M., Eberstaller, S. (2023, March 30) *MLOps Principles*. https://ml-ops.org/content/mlops-principles.