Measures and Metrics of ML Data and Models to Assure Reliable and Safe Systems

Benjamin D. Werner, US Army DEVCOM
Benjamin J. Schumeg, US Army DEVCOM
Jon Vigil, OptTek Systems, Inc.
Shane N. Hall, Montana State University
Benjamin G. Thengvall, OptTek Systems, Inc.
Mikel D. Petty, University of Alabama

Key Words: artificial intelligence, machine learning, reliability, safety, data, models, assurance

SUMMARY & CONCLUSIONS

The US Army solicited partners through a Broad Agency Announcement to propose solutions under a Small Business Technology Transfer contract mechanism for the program "Metrics and Methods for Verification, Validation, Assurance and Trust of Machine Learning Models & Data for Safety-Critical Applications in Armaments Systems." OptTek Systems, Inc. and University of Alabama in Huntsville (UAH) were one of the selected proposals for Phase I. Under this contract agreement OptTek and UAH set the goal to research & develop (R&D) fundamental metrics & measures for the certification & qualification of ML training data sets & models. Of particular note, the use of a safety score calculated from the accuracy as well as a dedicated look at data quality have been demonstrated as reasonable approaches to the proposed topic. As the Technical Point of Contact for this effort, the US Army Combat Capabilities Development Command Armaments Center (DEVCOM AC) authored the topic and provided guidance on the effort to align with mission objectives. This paper is an exploration of the research and development conducted by OptTek and UAH within the framework of how it may be applied to the assurance of systems to be developed by the US Army and augment practices in reliability and safety.

1 INTRODUCTION

The design, development, and deployment of armaments systems requires adherence to the existing processes and procedure of the Army and Department of Defense (DoD). Army Regulation (AR) 770-3 [1] details the necessary requirements and approvals for a system to achieve materiel release prior to being released to the field. The evaluation community and the respective specialty engineering domains provide evidence to demonstrate that systems are safe, suitable, and supportable under this regulation. For traditional systems this has proven to be an effective risk management procedure. Integration of Artificial Intelligence and Machine Learning (AI/ML) into future systems may require adjustments and changes to existing regulations. The possible nondeterministic nature of AI/ML in development and operation, as well as the complexity of system models when fielded, make it challenging to adhere to current processes and procedures. Given the nature of the work at Combat Capabilities Development Command Armaments Center (DEVCOM AC) and the dedication to safety of armaments systems, engineers have been identifying gaps between current processes and future challenges, and the potential means to fill them, especially in the framework of a digital engineering ecosystem. As part of this effort, DEVCOM AC authored a Small Business Technology Transfer (STTR) proposal [2] to identify measures and methods to support the verification, validation, assurance, and trust of AI/ML enabled functions in safety critical systems. The proposal by OptTek, in partnership with University of Alabama in Huntsville (UAH), was one of the solutions chosen. This paper will explore some of the research conducted and proposed measures as well as how it may be applied to the existing tools and processes in the design, development, and release of materiel.

A key concern highlighted by DEVCOM AC stakeholder engagements [3] was the safety of future armaments systems where there is nondeterministic development of algorithms compared with the manual development of software. System safety for the DoD is dictated by MIL-STD-882E [4], which defines specific objectives for software to be determined safe to include assessing the level of rigor of the development. Analogous to a Capability Maturity Model Integration (CMMI) construct [5], the level of rigor for safety critical software flows through the requirements, architecture, design, code, and testing to ensure that there are controls where necessary and the end product is fully deterministic. A key aspect identified is the criticality of data used during development, to ensure that it is appropriate, sufficient, extensive, and understood.

2024 Annual Reliability and Maintainability Symposium (RAMS) | 979-8-3503-0769-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/RAMS51492.2024.10457615

The goal of this Phase I STTR project was to establish fundamental methodologies to verify and validate data sources and ML models that require high assurance in safety-critical functions. To address this goal, OptTek and UAH researched and designed measures, metrics, and methods for performing verification, validation, and assurance (VV&A) for ML models and training data. There are several data and ML model types with potential armament system applications. The Phase I work focused on the VV&A methodology for a subset of relevant data model types, with consideration for broader applicability.

2 SAFETY SCORE

One of the measures to come forward from the effort is the use of a safety score. While first proposed in a paper by Salman et al. [6] to represent risk in security and safety applications, this can be leveraged to put a numerical value to assist in quantifying what the actual risks of a model failure may be. The safety score builds off of the better known accuracy measurement but elevates it to assess the impact of effect of the accuracy, or lack thereof. In the application of a standalone model there is minimal risk and thus, accuracy may be an acceptable measure, but in the contextual application of an armament system models that provide inputs and execute functions can precipitate considerable risk. The safety score takes the accuracy and measures it by a weight intended to represent the cost (loss) of the outcome. The cost weighting can be viewed in the context of both reliability and safety, with both having objective standards in defense applications to avoid the subjectivity that may be associated with subject matter expert (SME) opinion as the safety score is originally intended.

accuracy $$\frac{tp + tn}{fp + fn + tp + tn}$$

safety score $$\frac{w_{tp}tp + w_{tn}tn}{w_{fp}fp + w_{fn}fn + w_{tp}tp + w_{tn}tn}$$

Figure 1: Accuracy and Safety Score Calculations

In the equations presented in Figure 1, tp, tn, fp, and fn are the four possible outcomes in a binary classifier (true positive, true negative, false positive, and false negative); with recall that $tp + tn + fp + fn = N$, the total number of classified instances. In the safety score formula, the outcome weights w_{tp}, w_{tn}, w_{fp}, and w_{fn} are application-specific. A SME estimates the cost (loss) of each outcome, and each cost is divided by the sum of the four costs to get the normalized weights.

Regarding system safety, there are clearly defined risk scoring criteria in MIL-STD-882E that can be associated with the cost or loss. This risk scoring criteria delineates the probability and severity scores and associated measures. Whether the model returns a true positive, false positive, true negative, or false negative that will in turn impact some function of the system and execute an action or maneuver or display information to a user to make a decision. Any of those outcomes carry hazards, with the false results likely being of higher severity. Dependent on the system and function the model is executing or feeding the false positive or the false negative may carry more weight than the other. Adjustments to the weights can be correlated to the possible hazards based on how the system is integrated or deployed, allowing for very specific analysis of the impacts of model failures on those hazards. It is imperative that this is understood by the design team and evaluators as it has significant impact to the system users and the context of how the system will be deployed.

In the reliability sense, the cost weighting may be viewed through a different lens. Rather than the risk coming from MIL-STD-882E, the weight of any failure can be traced back to the system's failure definition and scoring criteria (FDSC). The FDSC defines and categorizes the potential failures a system might suffer. Traditionally this FDSC is used to score test incident reports and assess if the incidents observed or recorded in test are scored as reliability failures. A reliability failure would be an incident that impacts the ability of the system to complete its intended functions or mission. In descending severity, failures can be categorized as system aborts, essential function failures, or non-essential function failures. As with system safety, having this knowledge going into the design of the system and an AI/ML model is critical to understanding how to best design a model to mitigate the impacts of model failures on the system reliability. Rather than MIL-STD-882E, for a reliability vantage the FDSC could be leveraged to determine the cost weighting. The weights calculated from the FDSC criteria, with subject matter guidance, would identify which model accuracy shortcomings drive system reliability.

3 DATA QUALITY

Foroni et al. [7] asserts that existing data quality measures are focused solely on discrepancies between the dataset being evaluated (the noisy or dirty dataset) and an assumed dataset that accurately and completely represents reality (the clean dataset). Such measures, it is argued, are useful but not sufficient for a full assessment of dataset quality, which must consider the effect of data quality on the task for which the dataset will be used. According to Foroni et al. [7], the effect that a given deviation from the notional clean dataset may have depends not just on the numerical value of a data quality measure but on how the type of deviation affects the performance of a specific ML algorithm for the intended task.

This concept crosswalks with the need for system safety and reliability to ensure a full assessment of the operating conditions and associated stressors on the system (model). The concept of employment and the operation modes summary and mission profile can provide a starting point for how and where the system is expected to be operated, this can then feed into Failure Modes Effects and Criticality Analyses (FMECA) to generate potential failure modes for the expected conditions as well as an exploration of the boundaries of the model [8] to understand how well the model can perform the intended task. From the system safety parlance an Operational Design Domain Analysis (ODDA) also considers the operational conditions that the system will be deployed into and the

potential hazards that thus may be precipitated.

Budach et al. [9] reports an extensive empirical study of the effects of training data quality on ML algorithm performance. Six quantitative data quality measures were defined and used to quantify data quality; all were defined mathematically to have values in the interval [0, 1]. The data quality measures were:

- Consistent representation; degree to which features do not have multiple semantically equivalent values in the dataset.
- Completeness; ratio of non-missing feature values to number of samples in the dataset.
- Feature accuracy; deviation of feature values in the dataset from their true values.
- Target accuracy; deviation of target feature values in the dataset from their true values.
- Uniqueness; fraction of unique samples in the dataset.
- Target class balance; relative proportion of samples of each target class in the dataset.

By identifying and quantifying these measures on data, it enables analysis of a data set to reveal the inherent quality. Minimum required quality values can be defined for a given system requirement and implementation, and the corresponding data set quality can now be measured. Those minimum quality values can be defined based on perceived reliability of external sensors, data streams, and other sources of information utilized by the internal system for decisions and operations.

Additionally, manipulation and degradation of the data can take place to reduce a data set to a specific data quality value. Methods and procedures were developed as part of this effort to intentionally degrade the data to a given proposed quality value. This allows for a controlled simulation of degraded environmental performance, and therefore measurement of effectiveness and performance in that reduced situation. The results of system response to the manipulated data can inform the reliability and performance of that system in more realistic conditions – or at a minimum, highlight possible failure modes.

Biessmann et al. [10] is a survey of automated data validation methods for AI/ML. Methods for validating training data that will be input to AI/ML algorithms and methods for validating prediction data that has been output from AI/ML algorithms are both surveyed. The surveyed methods are placed in the context of a so-called "data pipeline", which is asserted to be a common component of AI/ML systems. Biessmann et al. [10] identifies several "data validation dimensions" and provides definitions of them that are conceptual, as opposed to the mathematical and precise definitions of data quality measures in Budach et al. [9], but that have some commonality with the measures in Budach et al. [9]. In Biessmann et al. [10] the data validation dimensions, and the methods that apply to them, are grouped into two broad categories: dimensions from general data management that are applicable to AI/ML models, and dimensions specific to AI/ML models. The general data management dimensions include:

- Data correctness
- Data consistency
- Completeness
- Statistical properties

The AI/ML-specific dimensions include:

- Predictive performance, e.g., precision
- Robustness to corrupted data
- Robustness to adversarial data
- Privacy, e.g., differential privacy and k-anonymity
- Fairness, e.g., does the AI/ML make predictions or decisions that treat humans fairly

Biessmann et al. [10] echoes the assertion of Foroni et al. [7] that the quality of data input to an AI/ML model can only be completely evaluated by considering the performance of the trained AI/ML model, stating "Data validation in the context of ML … has to take into account the current state of the downstream ML model …". Finally, Biessmann et al. [10] is unusual with respect to the other sources surveyed in that it includes ethical and legal issues, such as the privacy and fairness dimensions, with data quality issues.

While privacy concerns are more prevalent with respect to social, medical, and financial systems, legal and ethical issues are of considerable interest to DEVCOM AC. Systems to be fielded by the Army must have undergone legal review for adherence to Law of Armed Conflict [11] and applicable treaties. Additionally, per a 2021 memorandum [12] from the Secretary of Defense, the DoD Ethical Principles will apply to all DoD AI capabilities, of any scale, for warfighting and business systems. Specifically with respect to the Equitable principle, tools, processes, and measures developed to evaluate the quality of data sets for use in defense applications may help inform adherence [13].

Data quality not only focuses on the data itself but also on the management and control of the data. Data Safety Guidance from Safety Critical Systems Club [14] contains a comprehensive set of guidelines and best practices for managing data in the context of safety-critical systems. As the source notes, "mistakes introduced in data, or the inappropriate use of data, within safety-related systems have been factors in a number of documented accidents and incidents". This source includes a section (Appendix J) devoted specifically to data associated with AI/ML systems [14]. The section identifies several categories of data], briefly mentions some approaches to data safety for AI/ML applications, and provides references for additional information.

All of these data quality aspects relate back to the challenges of assuring safety and reliability in AI/ML models. MIL-STD-882E requires various levels of rigor to take place as part of system development and analysis. This rigor can extend to all aspects of the system, to include data. An understanding of the impact the quality of the data to the possible hazards can assist in not only bounding the integration of the system with other systems but also the impact of the system when deployed to data-poor environments. Additionally, knowledge of data dimensions paints the picture of general data quality, completeness, and applicability. Identification of shortfalls in a specific dimension, when

coupled to hazards tied to that dimension based on scoring and traceability, allow for adjustments to be made in data sources to mitigate those hazards. Lastly, understanding the control and management of data ensures high quality data remains within the development pipeline through to deployment and employment of the system.

The ability to manipulate the data during training to assess model performance directly impacts the reliability of the final system. Sensor and data feed reliability can be adjusted to mimic potential data reliability issues from external sources, simulating possible real-world scenarios. Through the Failure Modes and Effects Analysis (FMEA), these performance shortfalls can be tied to future failure modes. Additionally, the FMEA can be expanded to include the dimensions of data as part of any analysis, to identify those respective failure modes. Limitations in a specific dimension can be traced to a future failure mode, and therefore potential overall system failure after integration. Mitigations to these dimensions can take place to reduce the risk of that specific failure.

4 MODEL ANALYSIS

Optimization-driven interrogation of trained models can also be used to assess model robustness and safety. One such assessment for classifiers is the counterfactual explanation-based robustness score (CERScore) [15]. The CERScore quantifies the minimum expected changes to a sample's features (model inputs) that result in a different classification decision from the model, i.e., the distance between a sample in one class and the nearest counterfactual. This measure can be used to quantify classifier robustness to incidental or adversarial perturbations in input data streams. A lower CERScore indicates that relatively small input perturbances can change the model's classification, indicating a less robust model. OptTek and UAH developed software using OptQuest metaheuristic optimization engine [16] to efficiently search for minimum-distance counterfactual examples and compute the CERScore for models with mixed numeric and categorical input features and this was demonstrated on a test model for foliage cover classification in wilderness areas. Additionally, the counterfactual search and robustness measure was modified to quantify distance between selected "protected" and "target" classes that define high-consequence decision boundaries where model misclassification carries a high level of risk, present a system safety hazard, or is otherwise undesirable.

Another method for assessing model safety is falsification, or identification of inputs or input sequences that cause a model to violate one or more safety conditions [17]. Optimization can be used to drive efficient falsification searches when safety criteria are quantifiable and testable through simulation, as is common with reinforcement learning models. OptTek and UAH developed software to perform metaheuristic optimization-driven falsification searches and demonstrated this process with an open-source neural network model trained to control the trailing car in a "follow-the-leader" driving simulation [18]. An assessment was performed to maximize the probability of an undesired turning maneuver, i.e., cases where the trailing car is in correct position behind the lead car with a straight bearing, but the model indicates it should turn. This search identified multiple falsifying input states where the model indicated the car should perform the undesirable turn, but these cases were determined to be outside of the feasible operating domain of the driving simulation.

Two additional falsification searches identified possible action sequences (operational trajectories) by the lead car that resulted in unsafe model behavior. The first assessment searched for lead car action sequences that minimized the worst-case distance between lead and trailing cars over a simulated interval. This identified multiple trajectories where the model collided with the lead car. The second assessment searched for lead car action sequences that maximized the trailing car's distance from the road centerline. This identified trajectories that caused the model to leave the road.

Model risks or safety violations identified using these techniques can be further investigated for both the severity and likelihood of failure to inform the overall assessment of system safety. Additionally, robustness and falsification results may be useful for improving model safety and reliability if used to refine the model training process and/or restrict model operation to exclude domains where known failure occurs.

5 CONCLUSION

DEVCOM AC continues to research and develop the methods and means for ensuring safe, suitable, and supportable systems regardless of the type of system deployed. While AI/ML may bring challenges to the current process required under AR 770-3 [1], it is possible to improve that process through continued research and understanding of assurance technologies. The fundamental methodologies proposed by OptTek and UAH as part of their STTR Phase I research and development effort provide additional means and methods for DEVCOM AC to understand the safety and reliability of future AI/ML models through analysis of data and development.

The ability to quantify and measure an aspect of safety allows an independent assessor to better understand the inherent risk of a system for a given use or deployment. The weights and adjusted score permit subject matter experts in safety and reliability to give an appropriate amount of concern to certain model outputs, when considering the broader impacts of those results. The identified hazards and their trace back to results inform the appropriate weights needed to meet a specific threshold of safety. Assessments of reliability – especially when integrated into a larger system – can be determined when weighing and scoring critical accuracy aspects of the model that are interfaced with the broader system.

Tied closely to that is the perceived reliability for the given data set and performance. Controlled and measurable data manipulation can quickly identify performance failures during development and training, allowing for reliability engineers to effectively assess how that result can impact system effectiveness. Identification of data set features and dimensions

can identify new failure modes, allowing for additional corrective action to mitigate those possible failures.

This should not be considered an exhaustive review as OptTek and UAH did provide additional methods and measures as part of their STTR Phase I research. This research was highlighted to show the correlation to ongoing efforts within DEVCOM AC along with alignment to existing DoD processes and procedures. This research, combined with the requirements of MIL-STD-882E and AR 770-3, will continue to provide the methods and means to assuring safety and reliability of materiel released by the Army.

REFERENCES

1. Department of the Army, "Army Regulation 770-3, Type Classification and Materiel Release," 16 July 2021. [Online]. Available: https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN30603-AR_770-3-000-WEB-1.pdf.
2. Army SBIR, "A22B-T002 Verification, Validation, Assurance, and Trust of Machine Learning," Department of Defense, 2022. [Online]. Available: https://www.armysttr.com/22b-topics.
3. B. J. Schumeg, B. D. Werner, T. M. Mills and D. M. Bott, "Roadmap Development to Reduce Risk Associated with the Deployment of Artificial Intelligence Enabled Systems," in *2023 Annual Reliability and Maintainability Symposium (RAMS)*, Orlando, 2023.
4. Department of Defense, "MIL-STD-882E: System Safety," 2012.
5. ISACA, "What is CMMI?," [Online]. Available: https://cmmiinstitute.com/cmmi/intro. [Accessed 2023].
6. T. Salman, A. Ghubaish, D. Unal and R. Jain, "Safety Score as an Evaluation Metric for Machine Learning Models of Security Applications," *IEEE Networking Letters,* vol. 2, no. 14, pp. 207-211, 2020.
7. D. Foroni, M. Lissandrini and Y. Velegrakis, "Estimating the extent of the effects of data quality through observations," *Proceedings of the 2021 37th IEEE International Conference on Data Engineering,* pp. 1913-1918, 2021.
8. B. D. Werner and B. J. Schumeg, "Leveraging Traditional Design for Reliability Techniques for Artificial Intelligence," in *2022 Annual Reliability and Maintainability Symposium (RAMS)*, Tuscon, 2022.
9. L. Budach, M. Feuerpfeil, N. Ihde, A. Nathansen, N. Noack, H. Patzlaff, F. Naumann and H. Harmouch, "The Effects of Data Quality on Machine Learning Performance," 2022.
10. F. Biessmann, T. Rukat, J. Golebiowski, D. Lange and P. Schmidt, "Automated Data Validation in Machine Learning Systems," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering,* vol. 44, no. 1, pp. 51-65, 2021.
11. Department of the Army, "Army Regulation 27-53, Legal Review of Weapons and Weapon Systems," 2019. [Online]. Available: https://armypubs.army.mil/epubs/DR_pubs/DR_a/pdf/web/ARN8435_AR27-53_Final_Web.pdf.
12. K. Hicks, "Implementing Responsible Artificial Intelligence in the Department of Defense," Department of Defense, Washington, DC, 2021.
13. B. D. Werner, B. J. Schumeg, T. M. Mills and E. V. Velilla, "An Assurance Case for the DoD Ethical Principles of Artificial Intelligence," in *2023 Annual Reliability and Maintainability Symposium (RAMS)*, Orlando, 2023.
14. Safety Critical Systems Club, Data Safety Initiative Working Group, "Data Safety Guidance, Version 3.4, SCSC-127G," 2022. [Online]. Available: https://scsc.uk/scsc-127G.
15. S. Sharma, J. Henderson, & J. Ghosh, "Certifai: Counterfactual for robustness, transparency, interpretability, and explanations fairness of artificial intelligence models", *arXiv preprint arXiv:1905.0785,* 2019.
16. M. Laguna, "OptQuest," *Optimization of Complex Systems,* 2011. [Online] Available: https://www.opttek.com/white-papers/
17. A. Corso, R. Moss, M. Koren, R. Lee, & M. J. Kochenderfer, M. J., (2021), "A Survey of Algorithms for Black-Box Safety Validation of Cyber-Physical Systems", *Journal of Artificial Intelligence Research*, Vol. 72, pp. 377-428. October 2021.
18. I. Greenberg, Y. Chow, M. Ghavamzadeh, & S. Mannor, "Efficient risk-averse reinforcement learning", *Advances in Neural Information Processing Systems*, *35*, 32639-32652. 2022.

BIOGRAPHIES

Benjamin Werner
US Army DEVCOM Armaments Center
Picatinny Arsenal, NJ 07806 USA

e-mail: benjamin.d.werner2.civ@army.mil

Mr. Werner is an assurance lead in DEVCOM Armaments Center's Quality Engineering and System Assurance Directorate. A large part of this role includes the integration across engineering domains to include quality, reliability, and safety and collaborating across the respective communities to assess and mitigate risks in AI enabled systems. He also served in Acting roles as System Safety Engineering Chief and Reliability Engineering Branch Chief. Prior to his time at the Armaments Center, he was a reliability engineer at the Army Materiel Systems Analysis Activity (AMSAA), now the DEVCOM Analysis Center. He earned his bachelor's degree in Mechanical Engineering at Pennsylvania State University and his master's degree in Systems Engineering at Johns Hopkins University.

Benjamin Schumeg
US Army DEVCOM Armaments Center
Picatinny Arsenal, NJ 07806 USA

e-mail: benjamin.j.schumeg.civ@army.mil

Mr. Schumeg is a lead software quality engineer within Armament Center's Quality Engineering & System Assurance Directorate. His role includes researching Test & Evaluation Verification & Validation capabilities for Artificial Intelligence, Machine Learning, Automation, and other technologies. Mr. Schumeg currently co-leads the Army AI Software Safety Sub- Group focused on the Test & Evaluation/Verification & Validation of AI systems. He earned his bachelor's degree in Computer Engineering from Pennsylvania State University, and his master's degree in Computer Engineering from Stevens Institute of Technology.

Jon Vigil
OptTek Systems, Inc.
2241 17th Street
Boulder, CO 80302 USA

e-mail: vigil@opttek.com

Jon Vigil is a Senior Data Scientist at OptTek Systems, Inc. He has over 8 years of experience in aerospace system concept analysis and optimization. He obtained a B.S in Mechanical Engineering from the University of New Mexico and a M.S. in Aeronautics/Astronautics from Stanford University. Jon previously worked for Sandia National Labs as a mechanical and aeronautical engineer. Jon has extensive experience with simulation and integrating software tools with optimization and machine learning frameworks.

Shane N. Hall, Ph.D.
Montana State University
Jabs Hall 248, PO Box 173040
Bozeman, MT 59717 USA

e-mail: shane.hall1@montana.edu

Shane N. Hall is an Assistant Professor of Management at Montana State University. His expertise is in linear, integer, and combinatorial optimization models and methods with extensions to metaheuristic solution methods. He has led multiple research and development projects with the United States (US) government as a principal investigator and performs analytic consulting supporting many organizations across the US Department of Defense. Prior to MSU, Shane was a Principal Analyst at OptTek Systems, Inc. after serving over twenty years as an active-duty officer and analyst in the US Air Force. He obtained a Ph.D. in Industrial Engineering from the University of Illinois at Urbana-Champaign, a M.S. in Operations Research from the Air Force Institute of Technology in Dayton, Ohio, and a B.S. in Mathematics from Brigham Young University in Provo, Utah. He has authored over twenty peer-reviewed publications related to optimization and simulation in the military and healthcare domains.

Benjamin G. Thengvall, Ph.D.
OptTek Systems, Inc.
2241 17th Street
Boulder, CO 80302 USA

e-mail: thengvall@opttek.com

Benjamin G. Thengvall is Chief Operating Officer at OptTek System, Inc. He is an expert in the areas of mathematical modeling, real-time optimization software and services, transportation and scheduling problems, agent-based and discrete-event simulation, and simulation optimization and analysis. He has spent his career providing innovative software solutions to complex real-world problems through mathematical modeling, simulation, and metaheuristic techniques. His experience spans projects in both the commercial and government spheres. Prior to joining OptTek, he was an Operations Research Scientist at Navitaire, Inc. (now Taleris), a leading operations research technology firm serving the airline industry. He received master and doctoral degrees in Operations Research and Industrial Engineering from the University of Texas at Austin and a B.S. in Mathematics from the University of Nebraska-Lincoln. He is the author of numerous journal articles and book chapters and holds multiple patents related to software and optimization.

Mikel D. Petty, Ph.D.
University of Alabama in Huntsville
301 Sparkman Drive, OKT N353
Huntsville, AL 35899 USA

e-mail: pettym@uah.edu

Mikel D. Petty is a Principal Research Scientist at the University of Alabama in Huntsville's Information Technology and Systems Center and a Professor Emeritus of Computer Science at UAH. He received a Ph.D. in Computer Science from the University of Central Florida in 1997. Dr. Petty has worked in modeling and simulation research and development since 1990 in areas that include verification and validation methods, simulation interoperability and composability, human behavior modeling, multi-resolution simulation, and cybersecurity simulation. He has published over 255 research articles, chapters, and papers. He served on National Research Council and National Science Foundation committees on modeling and simulation, is a Certified Modeling and Simulation Professional, and served for five years as Editor-in-Chief of the scholarly journal SIMULATION: Transactions of the Society for Modeling and Simulation International.