A Data Strategy for Data-Driven Training Management: Artificial Intelligence and the Army's Synthetic Training Environment

Benjamin Goldberg
U.S. Army DEVCOM Soldier Center
Orlando, Florida
benjamin.s.goldberg.civ@army.mil

Randall Spain
U.S. Army DEVCOM Soldier Center
Orlando, Florida
randall.d.spain.civ@army.mil

Kevin Owens
The University of Texas at Austin
Austin, Texas
kowens@arlut.utexas.edu

Jeremy Lanman
U.S. Army PEOSTRI
Orlando, FL
jeremy.t.lanman.civ@army.mil

COL Paul Kwon
U.S. Army PEOSTRI
Orlando, FL
paul.o.kwon.mil@army.mil

Kevin Gupton
The University of Texas at Austin
Austin, Texas
kgupton@arlut.utexas.edu

Chris McGroarty
U.S. Army DEVCOM Soldier Center
Orlando, Florida
christopher.j.mcgroarty.civ@army.mil

Paul Butler
The MITRE Corporation
Orlando, FL
pvbutler@mitre.org

ABSTRACT

The U.S. Army's Synthetic Training Environment (STE) and supporting training and learning concepts define Artificial Intelligence (AI) as a functional requirement to optimize the use of simulation to support individual and team readiness requirements. A current limitation to technologic tools examining AI is access and proper management of meaningful data. Many AI methods are developed under controlled and isolated settings with limited use cases and data-points. These investments prove a methodology from a technology readiness standpoint, but often fail to meet the intent of having ready-to-transition AI services that create valid measures and drive calculated decisions. In this paper, we will present a strategy for defining data requirements and management to support an evolutionary approach to AI development and validation. How do we directly address this issue? Establishing a data strategy on standards, best practices, acquisition requirements, and mission threads can produce data repositories specifically implemented to drive AI maturation. This emphasizes collecting data with a purpose, and establishing explicit implementation guidelines that align to desired end-state AI capability. This position is explored at a high-level in the context of STE and future Programs of Record. We will present a framework based on AI services associated with adaptive training management, the type of functions each service provides, and the type of data bucket required to drive its utility. Services explored include building more objective assessments across multi-modal data and across training iterations; building personalized feedback and scenario adaptations that target strengths and weaknesses; creating recommender engines for guided training progression to maintain proficiency; and building realistic synthetic entities that enhance training fidelity. Each of these services demand careful consideration for data instrumentation and management. Beyond persistent storage, we will present recommendations for the capture, contextualization and retention of data to drive evolutionary maturation of each AI function.

ABOUT THE AUTHORS

Dr. Benjamin Goldberg is a Senior Research Scientist at the U.S. Army Combat Capability Development Command – Soldier Center. His research focuses on adaptive experiential learning with an emphasis on simulation-based environments and leveraging Artificial Intelligence to create personalized experiences. Dr. Goldberg holds a Ph.D. in Modeling & Simulation from the University of Central Florida and is well published across several high-impact journals and proceedings, including IEEE Transactions of Learning Technologies, the Journal of Artificial Intelligence in Education, and Computers in Human Behavior.

Dr. Randall Spain is a Research Scientist at the U.S. Army Combat Capability Development Command – Soldier Center. He holds a Ph.D. in Human Factors Psychology from Old Dominion University and an M.S. in Experimental Psychology. His research focuses on the design and evaluation of adaptive and intelligent training systems.

Kevin Owens is an Engineering Scientist at the Applied Research Laboratories: The University of Texas at Austin. After retiring from the US Navy, with years of experience as an instructor, curriculum designer and operational trainer, Kevin earned a B.S. in Workforce Education and Development from Southern Illinois University, and a M.S. in Instructional Design, Development and Evaluation from Syracuse University. Kevin has since spent over 20-years in industry and academia conducting DoD related learning engineering projects to improve education and training. He is currently working on Science and Technology projects related to the US Army's Synthetic Training Environment, researching a Competency-Based Experiential Learning model.

Dr. Jeremy Lanman is the Chief Technology Officer for the U.S. Army Program Executive Office for Simulation, Training and Instrumentation (PEO STRI) in Orlando, Florida. He earned a B.S. in Computer Science from Butler University, an M.S. in Software Engineering from the Embry-Riddle Aeronautical University, and a Ph.D. in Modeling & Simulation from the University of Central Florida. Also, he is a graduate of the Army Management Staff College, and a Harvard Senior Executive Fellow. Dr. Lanman serves as a graduate scholar and adjunct professor at the University of Central Florida, Florida Institute of Technology, and Embry-Riddle Aeronautical University.

COL Paul Kwon is currently assigned as the Clinical Advisor at PEO-STRI in Orlando, Florida. Responsibilities include medical advisor to the PEO and integrate medical expertise on the Simulation Training Environment. Additional duties include Adjunct Assistant Professor at USUHS (Epidemiology and Biostatistics), Affiliate Professor at the University of Central Florida School of Medicine, Reviewer for Journal of Military Medicine, Pediatric staff at Patrick SFB, and Level III acquisition work force qualified. COL Kwon has over 35 publications and presentations in national conferences and peer reviewed publications. He is the recipient of the 2008 Patient Choice Award and nominated for Best Pediatrician in the Western Regional Command.

Chris McGroarty is the Chief Engineer for Advanced Simulation at the U.S. Army Combat Capability Development Command – Soldier Center. His research interests include distributed simulation, simulation architectures, applications of Artificial Intelligence Technologies to simulation, novel computing architectures, innovative methods for user-simulation interaction, methodologies for making simulation more accessible by non-simulation experts, and the application of videogame industry technologies. He manages a variety of research efforts that mature, integrate, and demonstrate these technologies in a relevant Army and Department of Defense context. He received his Master of Science and Bachelor of Science in Electrical Engineering from Drexel University in Philadelphia, Pennsylvania.

Kevin Gupton is an accomplished Systems Architect with over 20 years of experience in data engineering and knowledge management. Currently, he holds the position of Engineering Scientist at the Applied Research Laboratories, The University of Texas at Austin. Kevin has contributed his expertise in system architecture and data management to various notable projects, including the Army Synthetic Training Environment (STE), Joint Federated Common Data Services (JFCDS), Integrated LVC Test Environment (ILTE), and NATO MSG-164 M&S-as-a-Service. He earned a B.S. in Mathematics and an M.S. in Computer Science from Texas A&M University.

Paul Butler is a Principal Systems Engineer with The MITRE Corporation. For the past five years, Mr. Butler has served as MITRE's PEO STRI Project Leader, providing technical and management leadership to training programs across PEO STRI's diverse PM SE, PM ST, and PM CT2 portfolios. Mr. Butler has nearly forty years of experience in developing systems architectures for Army modeling and simulation training solutions in live, virtual/gaming, and constructive domains. He earned his B.S. in Mathematics from The University of the South and M.S. in Applied Mathematics from Vanderbilt University.

A Data Strategy for Data-Driven Training Management: Artificial Intelligence and the Army's Synthetic Training Environment

Benjamin Goldberg
U.S. Army DEVCOM Soldier Center
Orlando, Florida
benjamin.s.goldberg.civ@army.mil

Randall Spain
U.S. Army DEVCOM Soldier Center
Orlando, Florida
randall.d.spain.civ@army.mil

Kevin Owens
The University of Texas at Austin
Austin, Texas
kowens@arlut.utexas.edu

Jeremy Lanman
U.S. Army PEOSTRI
Orlando, FL
jeremy.t.lanman.civ@army.mil

COL Paul Kwon
U.S. Army PEOSTRI
Orlando, FL
paul.o.kwon.mil@army.mil

Kevin Gupton
The University of Texas at Austin
Austin, Texas
kgupton@arlut.utexas.edu

Chris McGroarty
U.S. Army DEVCOM Soldier Center
Orlando, Florida
christopher.j.mcgroarty.civ@army.mil

Paul Butler
The MITRE Corporation
Orlando, FL
pvbutler@mitre.org

INTRODUCTION

Artificial intelligence (AI) has emerged as a transformative technology with the potential to revolutionize various fields, including military, medicine, and domains that require timely and accurate decision-making processes (Russell & Norvig, 2021). The Department of Defense (DoD) defines AI as "the ability of machines to perform tasks that normally require human intelligence" (Allen, 2020). In the context of this paper, we refer to AI as the utilization of intelligent systems that exhibit human-like cognitive abilities to perform complex tasks. These systems are designed to process and analyze vast amounts of data, extract valuable insights, and make informed recommendations or predictions based on goals and outcomes.

The U.S. Army and PEO STRI is increasingly using AI to improve its simulation, test, and training capabilities. These technologies can be used to create more realistic and immersive training environments, to personalize training for individual Soldiers, and to assess Soldier performance more accurately. However, to fully realize the potential of AI and machine learning (ML) for simulation, test, and training, the U.S. Army needs to develop a comprehensive enterprise data strategy. This strategy should address the following key areas:

- Data collection and storage: The Army needs to collect and store large amounts of data from a variety of sources, including simulations, training exercises, and real-world operations. This data will be used to train AI/ML models and to evaluate the effectiveness of training programs.
- Data governance: The Army needs to establish clear policies and procedures for managing data. This includes ensuring that data is accurate, secure, and accessible to authorized users.
- Data analytics: The Army needs to develop the tools and techniques to analyze data and to extract insights that can be used to improve simulation and training. This includes using AI /ML to identify patterns and trends in data, and to predict future outcomes.
- Data sharing: The Army needs to share data across its different simulation and training systems. This will allow for the development of more comprehensive and integrated training programs.

An enterprise data strategy for AI/ML in simulation, test and training would have several benefits for the Army, including:

- Improved training effectiveness: AI/ML can be used to create more realistic and immersive training environments, to personalize training for individual Soldiers, and to assess Soldier performance more accurately. This can lead to improved training effectiveness and readiness.

- Reduced costs: AI/ML can be used to automate many of the tasks involved in simulation and training, such as data collection, analysis, and reporting. This can lead to reduced costs through increased efficiency.
- Enhanced decision-making: AI/ML can be used to analyze data and to extract insights that can be used to improve decision-making. This can help the Army to make better decisions about training, operations, and resource allocation.

Critical to the success of any formal data strategy is understanding that AI-driven capabilities are rapidly advancing. To harness the full potential of AI to promote data-driven adaptive training services in the U.S. Army's Synthetic Training Environment (STE), a strategy must be developed to inform the collection, labeling, and management of data that will influence, and mature simulation to support individual and team readiness requirements. The purpose of this paper is to address this need by discussing data requirements and data management policies to support AI-driven services aligned to STE's adaptive training management tool (TMT) objectives. In the follow sections we describe the motivation for the broader enterprise-level data strategy the U.S. Army is establishing to support simulation-based training acquisition programs. Then, we discuss several prominent adaptive training functions from a capability standpoint (e.g., multi-modal assessment, conversational agents, pedagogical reasoning, competency recommender engines) and highlight AI-based techniques and methods used to enable these capabilities (e.g., machine learning, natural language processing, computer vision, reinforcement learning). These services align to the functions of STE's TMT, including building objective assessments across multi-modal data and across training iterations, building personalized feedback and scenario adaptations that target strengths and weaknesses, creating recommended engines for guided training progression to maintain proficiency, and building realistic synthetic entities that enhance training fidelity. Each of these services demand careful consideration for data instrumentation and management. We conclude with a set of recommendations for the capture, storage, contextualization (i.e., metadata and labeling) and retention of data to drive evolutionary maturation of each AI function. Aligning data specifications with the capability objectives outlined in STE will allow the U.S. Army to effectively communicate the essential components required for AI-driven adaptive training to the acquisition community.

BACKGROUND AND MOTIVATION

In an era of rapid technological advancement, the future operational environment presents complex challenges that demand innovative approaches to training, education, and talent management. When considering this from a readiness standpoint, the U.S. military must respond to an uncertain and volatile future by preparing our talent more fully, upskilling our staff continuously based on operational necessities, and enabling optimum performance through a training strategy that leverages technology and data-driven services grounded in learning science. To address these challenges, the U.S. Army continues to modernize its training capabilities through its STE program. Synthetic-based training involves immersive, realistic simulations that recreate real-life situations, generating valuable data from multiple sources. The data gathered from these environments can provide insights into individual and team performance (Sottilare et al., 2018). With STE set to leverage advancements in gaming and extended reality to meet the U.S. Army's collective training needs, there is increasing recognition of the pivotal role that data and AI will serve in delivering state-of-the-art training solutions. These solutions are envisioned to involve data-driven adaptive training management tools informed by best practices in intelligent tutoring system design and advancements in AI to enhance training effectiveness and competency development across individuals and team structures.

Decades of research show the potential of AI-driven training to address the Army's training and readiness needs. Meta-analyses examining the effect of adaptive instructional systems (AIS) and intelligent tutors have consistently shown enhanced learning outcomes across various domains. For instance, Kulik and Fletcher (2016) found that intelligent tutoring resulted in higher learning gains when compared to traditional instructional methods, with an average effect size of 0.8 – 0.9 standard deviations. The personalized feedback and tailored instruction offered by AIS enable learners to receive targeted support, address their specific knowledge gaps, and progress at their own pace when compared against peers engaging in traditional learning strategies (VanLehn, 2016).

For synthetic-based adaptive training and experiential learning, certain types of data and AI services will be required to drive the adaptive training management elements for STE (Figure 1). AIS and intelligent tutoring systems operate on common architectures that apply modeling techniques across three primary elements: (1) learner modeling (2) domain modeling, and (3) instructional modeling (Sottilare et al., 2012; Woolf, 2010). These functional models cover the primary data variables used to tailor instruction, but each element requires data-driven techniques to perform a variety of services to facilitate an adaptive experience. This involves monitoring interactions against a defined problem or scenario and measuring performance within the domain model to assess the quality of process, procedure, and behavior against an objective.

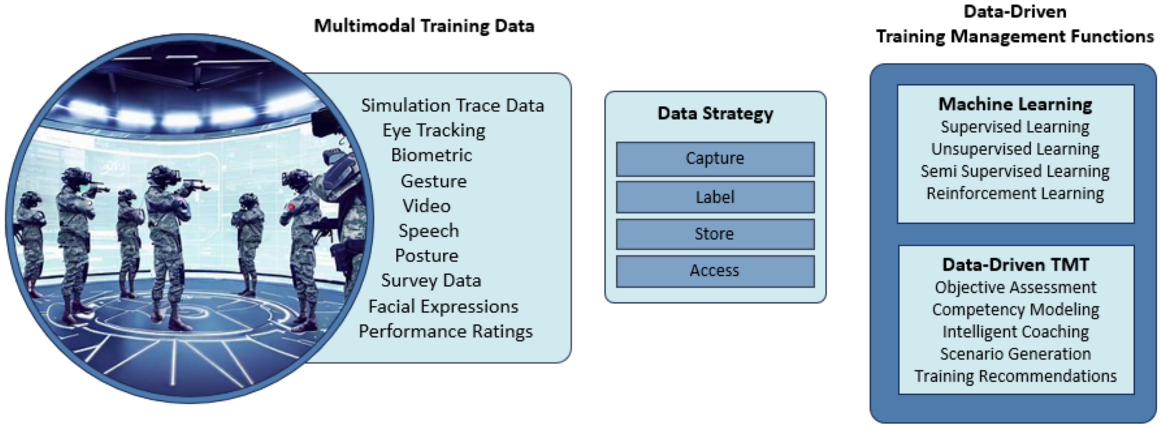

Figure 1. Data-Driven Training Management Functions

Utilizing these models, AI-driven training services can analyze learner actions during STE-based training events and provide real-time coaching guidance (Johnson & Lester, 2018; Spain et al., 2021). With a strategy to measure domain level performance, a pedagogical coaching model is used to the tailor the learning experience, aiming to find the right balance between challenge and feedback. This approach, known as micro-adaptation (Durlach & Spain, 2014), guides learners through the Zone of Proximal Development (Vygotsky, 1976), where they are challenged and supported to optimize learning. Progress is tracked over time, monitoring performance across different problems and scenarios, driving updates to long-term learning goals and competency objectives, tracked as knowledge, skill, and ability levels in the learner model. While the learner model is responsible for assessing proficiency over time, it also plays a key role in personalizing the training experience providing data that recommender engine type services can use to plan and structure the tailored and customized learning pathways.

AI-DRIVEN ASSESSMENT, COACHING, AND CONTENT GENERATION

In the following sections we discuss a set of core services to support adaptive training that will enable STE to provide objective assessments, intelligent coaching and assessment, rapid content creation and the ability to model, estimate, and predict performance over time. Ensuring the appropriate training data are recorded, labeled, stored, and available for AI/ML-driven functions will be critical for meeting the adaptive training goals of STE. The core set of services align with components of any adaptive training service to including data to support learner analytics, objective assessment, and competency modeling; data to inform pedagogical interventions and tutoring approaches, data to build training recommender engines and personalized learning pathways, dynamic scenario generation.

Objective Assessment, Multimodal Learner Analytics, and Competency Modeling

Approaches for learning evaluation vary considerably across domains and environments, but current trends in this area focus on using multi-modal data sources and ensemble machine-learning algorithms to analyze learning behaviors to inform student prediction models. By using data from multiple input sensors (e.g., video, audio, eye tracking, facial expressions, text responses, simulation-trace data) multimodal learning analytics can be used with machine learning-driven approaches to assess and predict student behaviors and learning states in synthetic training applications. Multimodal learning analytics have been used to successfully model student engagement (Sabourin & Lester, 2013; Wu et al., 2016), conceptual knowledge (Nye et al., 2014; Pavlik et al., 2013), procedural knowledge (Vatral et al., 2022), and self-regulated learning behaviors (Azevedo & Gasevic, 2019). In the context of U.S. Army training, these techniques, coupled with dynamic synthetic training scenarios, enable timely predictions of Soldiers' current and future competency, and allow systems and instructors to provide adaptive scaffolding during training events.

A critical goal of performance assessment is to not only measure task performance but also measures the processes and underlying knowledge, skills and behaviors required to perform a task, and apply evidence-centered approaches that link constructs and competencies to observed states. To support this, multimodal learning analytics can leverage data the training system (i.e., trace data), sensor data gathered from biometric devices worn by the trainee, video and audio streams and traditional tests or performance rating forms. These data can be compiled using supervised, unsupervised, and semi-supervised ML techniques to detect, diagnose, and predict cognitive, psychomotor, and affective learner states (Hoque, 2016). The insights gathered through these analyses can support real-time or post-training analysis and influence training management functions including changes in instructional strategies and scenario adaptations and support competency and learner modeling, and dynamic scenario generation.

Coaching, Support, and Instructional Scaffolding

In addition to facilitating learner analytics and objective assessments AI can also be used to implement adaptive instructional support and scaffolding in STE. An important question that arises during a training or tutoring session is how to support a learner who has made multiple errors or has reached an impasse. The instructor, coach, or tutor, whether human or AI agents, must understand the misconceptions and mistakes and create plans to address them. In intelligent tutoring systems, *tutorial planners* are the component of the pedagogical model responsible for determining what actions to take to help a trainee overcome an impasse (Woolf, 2010) Tutorial planners utilize predefined rules or schemas to determine when and how to provide coaching and scaffolding. These instructional rules are often based on learning theories or expert input. Tutorial planners have historically been manually authored, but in recent years, ML techniques, including Reinforcement Learning (RL), have been used to craft data-driven *tutorial planning policies.* By analyzing student interactions with training content and modeling accumulated rewards, the planner can make informed decisions about when and how to deliver feedback, remediation, and coaching to maximize learning and retention. RL-based tutorial planners have been used in narrative-centered learning environments (Rowe & Lester, 2015; Sawyer et al., 2017); adaptive courses that support computer programing (Shen et al., 2018), and to support cognitive engagement during online training (Fahid et al., 2021; Spain et al., 2021).

Identifying the data requirements to support data-driven tutorial planners in AI-driven training systems is essential for developing robust adaptive training systems that can mimic human coaching and teaching practices. As previously noted, tutorial planning can be seen as a RL task in which the tutor (the AI agent) seeks to enact pedagogical decisions (i.e., actions) that will affect its environment (i.e., the trainee and his/her learning environment) to optimize student learning outcomes (i.e., rewards). By obtaining relevant representations of the learning states, actions, and rewards, the AISs and intelligent tutors can make informed decisions regarding adaptive instructional sequencing (Doroudi et al., 2019), modeling student problem-solving trajectories (Rafferty et al., 2016), and provide targeted remediation (Spain et al., 2021). Recent work with the U.S. Army's Generalized Intelligent Framework for Tutoring (GIFT), a component of STE's TMT, has focused on identifying the data requirements to support RL-driven coaching and instruction and can provide a valuable resource for introducing data for state, action, and reward representations, and strategies that can be followed to capture the needed data to support data-driven coaching and instructional support (Smith et al., 2022).

In addition to providing personalized feedback and support, advances in AI are creating rapid improvements in conversation agents that can engage in natural language-driven interactions with students. This type of learning, called interactive dialog, is highly effective because it requires trainees or students to engage ongoing conversations with the agent about a specific topic. When students answer questions, explain problem-solving steps, and discuss similarities and differences between concepts, they are more likely to understand and remember the material, leading to better learning outcomes (Chi & Wylie, 2014). Iterative dialogue-based instruction requires several functions. First, students need to interact and take turns speaking with the agent. They can respond to questions or prompts from the tutor. Second, the tutor, in turn, should simulate human conversation patterns. The tutor must understand the meaning of the student's input and provide an appropriate response and instructional strategy. And lastly, from a data management standpoint, the training system must be able to track the concept being tutored, student responses including bugs, misconceptions, or common errors associated with the training topic or concept, the prompts and instructional tactics used to remediate and correct student misconceptions and the success of these interventions.

Guided Learning Pathways

Another common AI service applied in AIS and intelligent tutors involves recommender engines to help select the next learning objective and learning pathway in support of that objective. This varies from tutorial planners as it accounts for long-term learning goals that can span across several tasks and competencies. These AI services are informed by tracked competency states and the mathematical models established to build a predictive metric of proficiency. In the context of experiential learning, a proficiency model considers three variables: (1) how Well someone performs over time; (2) how Hard the conditions that they performed against; and (3) how Often they performed a task or applied a competency. This WHO modeling approach (Hernandez et al., 2022) is grounded in evidence-centered design and is dependent on tracking longitudinal performance gains across several structured practice opportunities. When considering an established training plan, a recommender engine is designed to provide deeper insight into what exercises/problem sets, scenarios, and scenario configurations (e.g., conditions, difficulty, complexity) are appropriate for achieving a stated objective.

Scenario Generation and Fidelity

Realistic and dynamic scenarios tailored to training tasks and conditions are crucial for effective training. By leveraging AI in the generation of scenarios and simulation assets (behaviors, events, effects), the U.S. Army can create more immersive training environments that accurately simulate operational conditions and behaviors, while reducing the effort for engineering simulation systems and authoring scenario content. AI can support several critical functions to support exercise design and fidelity including generation of operational condition and situations; generation of entity behaviors; generation of group (or collective) behaviors; and dynamic scenario adaptations. We briefly highlight each of the functions below:

- **Generating Operational Conditions and Situations:** AI can provide the opportunity to generate diverse operational conditions and situations that reflect real-world complexities. By analyzing historic data, environmental features, and mission objectives, these algorithms can create relevant scenarios tailored to training objectives and to the competency/proficiency of the learner. AI-supported generation ensures that training simulations capture the intricacies and challenges faced in real military situation, while lowering the time and expertise required to author a complex training environment.

- **Generating Entity Behaviors:** Simulating realistic behaviors of entities such as friendly forces, opposing forces, and civilian populations is a central part of simulation-based training. The availability of high-fidelity training experiences is also contingent on simulated entity behaviors, as increased autonomy reduces the reliance on training-support players within the scenario. By analyzing intelligence reports, recordings of real operational behaviors, and cultural factors, AI algorithms can accurately model and simulate the behaviors of different entities with true-to-life cultural, doctrinal, and competency characteristics and therefore enhance training effectiveness and realism.

- **Generation of Group (or Collective) Behaviors:** Generating group and collective behaviors is another core function that AI could enhance. Training scenarios involve complex interactions among groups of entities, ranging from coordinated opposing forces to urban areas busy with civilian life. AI techniques such as multi-agent systems, enable the simulation of group behaviors by modeling dynamics and interactions within scenarios. By leveraging AI, we can simulate the coordinated actions of friendly forces or adversarial strategies employed by opposing forces. This capability allows for realistic (i.e., validated) and dynamic scenarios that facilitate effective training.

- **Importance of Adaptive Scenario Fidelity and Validity of the Operational Environment:** AI-supported scenario fidelity plays a critical role in delivering immersive and realistic training experiences. By dynamically adapting scenarios based on learner performance and evolving training objectives, AI algorithms ensure that training remains engaging, relevant, and effective. This adaptability allows for better training outcomes by challenging learners appropriately and replicating real-world conditions. However, it is necessary to validate and verify the AI-generated scenario fidelity against the operational environment and training objectives. While AI improves training effectiveness and efficiency by rapidly generating a wide range of scenarios, comprehensive data collection and ethical AI implementation are necessary to address potential biases or unintended consequences.

DATA INSTRUMENTATION AND MANAGEMENT

To effectively harness the potential of AI services for learner analytics, performance assessment, instructional support, scenario generation and behavior representation, training programs—and the training enterprise as a whole—need robust data instrumentation and management practices. Table1 provides a list of data requirement function, data sources, and methods that aim to support three critical TMT functions. By aligning the data requirements of each AI service and specifying the data sources, collection responsibilities, timing, and collection methods, training programs can ensure the availability of necessary data to support learner analytics, performance assessment, instructional support, and scenario/behaviors generation. This full-scale data instrumentation and management approach facilitates the integration of AI services and maximizes the value derived from training data for enhanced training outcomes.

Table 1: Data requirements for TMT functions and AI services

TMT Functions	Data Required	Data Sources	Data Collection Responsibility	Timing of Data Collection	Data Collection Methods
Learner analytics and objective assessment	Training logs, simulation state and events, recorded audio and video, voice and digital message communications, physiological measures (e.g., eye tracking data), performance metrics, learner feedback, and objective assessment results	Integrated training systems, simulation platforms, audio and video recording devices, communication systems, sensor devices, and assessment tools	The training organization or institution overseeing the program, trainers, instructors, or designated evaluators responsible for assessment	Continuous data collection during training sessions and exercises, along with assessment exercises and post-training evaluations	Automated data collection mechanisms integrated into training systems, sensors, audio/video recording tools, trained evaluators, assessment tools, and performance tracking systems
Instructional support, scaffolding	Learner profiles, performance history, assessment results, training logs, contextual information, simulation state and events	Integrated training systems, learner profiles and record stores, assessment tools, simulation platforms	Training administrators, instructors, or AI-supported systems responsible for delivering instructional support	Continuous data collection during training sessions and exercises	Automated data capture from integrated systems, learner input, instrumentation, and simulation platforms
Guided learning Pathways	Learner profiles, performance history, competency estimates, Index of available training scenarios	xAPI assessment statements, competency frameworks, learner model, scenario metadata	Proponent defining competency frameworks, training designers	Post-scenario input of recorded assessments	Automated data capture from integrated systems, instrumentation, and simulation platforms
Scenario and behaviors generation	Historical data, operational context, mission objectives, intelligence reports, cultural data, scenario design guidelines	Historical training data, operational databases, intelligence reports, cultural databases	Training organizations, subject matter experts	Prior to scenario creation and scenario iteration processes	Data mining and analysis, expert input, structured interviews, knowledge elicitation techniques, and contextual information gathering

RECOMMENDATIONS FOR ESTABLISHING A DATA CAPTURE, CONTEXTUALIZATION, AND RETENTION STRATEGY

To harness the full potential of the AI services described above, training programs must adopt a proactive and action-oriented data management approach early in the fielding of their systems. Training programs should prioritize collecting data at scale, capturing a wide range of multimodal sources, and using adaptive instructional services and data reporting standards to track assessments and contexts across training experiences and environments (Goldberg et al., 2021). A systematic approach ensures that the data collected represents the diversity of training tasks, conditions, and training strategies employed to meet a defined objective. It also establishes a data repository that can be accessed by the research community to help in the maturation of the AI data functions described above.

By 2024, PEO STRI aims to have an Enterprise Data Strategy (EDS) that defines the vision and direction, contains plans and guidelines, and provides resources and tools that enable programs to chart their data implementations across the data lifecycle. The EDS will align with the DoD Data Strategy (U.S. Department of Defense, 2020), a key component of the Department's Digital Modernization program, supporting the National Defense Strategy (NDS) (U.S. Department of Defense, 2022) by enhancing military effectiveness through access to accurate, timely, and secure data. The PEO STRI EDS will improve communication of system data structures and associations within and across programs. Additionally, the EDS will enhance interoperability and reduce inaccuracies as training systems are developed and integrated. Developing this data architecture will greatly expedite integration with other Army programs of record internal and external to PEO STRI. These benefits will be realized in STE and when Army Live training ranges and Mission Training Complex data analytics teams use ML and artificial intelligence applications to find patterns in data that fuel decision making and enhance training effectiveness. In the following sections we provide several recommendations for aligning with the goals of the EDS and for establishing best practices to support AI-driven training management capabilities discussed above.

Recommendation 1: Develop a Strategy to Collect and Preserve Multimodal Data

As a starting point, we recommend targeting data sources that align across the cognitive, psychomotor, and affective learning domains, and which have been used extensively in the literature to model key indicators of learning and skill development (see Table 2). To support the development of ML models that can classify learner states, initial formative assessments can be derived from observer driven bookmarks that are labeled, timestamped, and aligned to a scenario task (i.e., experience event), which enables rapid retrieval and filtering of data to feed AI models. Table 2 lists candidates for initial training environment instrumentation and can support an array of data-driven assessments that leverage the breadth of ML techniques. These data represent common candidates to support modeling of task performance across several domains that involve different sets of knowledge, skills and behaviors required to be proficient. As an example, consider tracking movement and coordination using video from an infantry squad versus a surgical unit team; each team requires different interpretations of their behaviors in relation to their environment, but both domains can leverage cameras to drive computer vision and audio to drive natural language services. This can result in generalizable techniques to use standard data types to obtain features used to train ML and to infer performance. While mature data capture methods might not currently exist, collecting training data under a controlled management strategy will support rapid development and validation of AI services as data is collected, retained, and analyzed iteratively against an established context.

Recommendation 2: Data Labeling for Context Mapping

To maximize its value, multimodal data must be labeled. We recommend using structured metadata and labeling techniques to align with "experience events" built within a broader mission context (Hernandez et al., 2022). This will facilitate the alignment of data sources to support performance assessment at the task, sub task, and step level from which AI modeling techniques can be used to detect and predict cognitive, affective, behavioral states and training outcomes. Establishing rich context within a task experience event is critical and must include who was trained (e.g., individual and team associations), what tasks were trained and under what conditions, when was a task initiated and when was a task completed within a broader scenario context, what competencies are applied within a task context (via front-end analysis), and what were the task outcomes with formative and summative measures of performance. To support this context mapping, the scenario creation process must account for configuring these experience events to enable data management. This involves defining explicit triggers and events associated with a task's initialization, and then use training management services to track the evolving mission context based on when and where an experience event takes place.

Data labeling will introduce new requirements and assumed workflows. Using AI services to help automate the data labeling process will be pivotal for in minimizing the burden placed on trainers and training operators, allowing them to focus on their primary responsibilities of training. In addition, we recommend interactive exercise control tools to drop bookmarked observations on a data time log and AAR tools to support post-scenario reviews and labeling in a controlled manner that removes the time burden during training execution.

Table 2. Examples of Multimodal Data Sources and Associated Learning Indicators

Data Type	Cognitive	Affective	Psychomotor/ Behavioral
Simulation trace data: Interaction logs capture detailed activities in the synthetic environment. These logs provide micro-behavioral data logged at millisecond intervals, assessing student problem solving, coordination, and environment interaction such as navigation, menu selection, and text entries. System events like adaptive feedback are also recorded. All actions and events are automatically time-stamped and organized chronologically.	Problem solving skills, Intellectual skills, Cognitive workload, Procedural knowledge	Boredom, Frustration	Individual actions and performance, Interaction data (menu selections, navigation, text entries), Team actions and performance
Video: First, second and third-person video data are valuable for assessing individuals and teams in physical and virtual environments. By combining video input with open-source software and AI-driven multimodal learning analytics, we can support video-based performance assessment for collective skills. This approach enables the detection and classification of cognitive and affective states. Note: Video can be obtained by physical and virtual camera configuration, supporting extensibility across virtual, mixed reality, live environments.	Problem solving skills, Intellectual skills, Visual attention, Cognitive workload	Engagement, Boredom, Frustration, Confusion	Individual performance, Team actions and coordination, Collaborative learning behaviors, Posture
Speech: Audio data and recordings that be analyzed using manual or natural language processing-driven approaches to support the assessment individual and team-level processes and outcomes.	Intellectual skills, Cognitive strategies, Reasoning, Team cognition	Engagement, Frustration, Confusion, Team cohesion, Team trust	Team coordination, Information exchange, Leadership, Oral presentations
Biometric: Sensor data captures various biometric measurements such as electrodermal activity, heart rate, heart rate variability, and other relevant metrics. These measurements can offer insights into affective states and changes, serving as correlative evidence of specific cognitive and affective states like stress, arousal, and workload.	Cognitive workload	Arousal, Engagement	Movement, orientation, activity level
Eye tracking: Active data captures reflections of infrared light through the pupil and off the back of the eye's retina, providing measures of visual attention, point of gaze, blink rates, eye scanning patterns, and pupil diameter.	Cognitive load, Visual attention, Vigilance	Engagement	Areas of interest (i.e., fixations), Scan patterns
Motion capture data: Sensor-driven motion capture data to be used to measure posture and gestures.	Correlates of cognition and attention	Engagement, Boredom, Frustration, Stress	Team coordination for movement, Non-verbal communication
Input device trace data: User actions captured from input devices like mouse, hand-controller, and weapon/system activity. These data offer insights into user interactions with computer-based training software, including VR, AR, Live, and gaming-based systems.	Rate and/or fixation of keypresses, controller or mouse-movements, control actions which could be indicative to cognitive state.		Interaction data (menu selections, navigation, text entries)
Performance ratings: Training performance data including formative and summative assessments, field observations, Situational judgment test, peer assessments, tacit knowledge test, STX lanes ratings, and other forms of data that could serve as dependent variables (I.e., outcomes of interest).	Intellectual skills, Judgement, Reasoning, Procedural knowledge, tacit knowledge	Cross cultural competence, cohesion, trust, confidence	Coordination, Communication, Procedural knowledge, Tacit knowledge, Skill proficiency
Survey data: Self report data related to constructs of interest. These data can be used to correlate features with labels and outcomes of interest.	Cognitive load, Attention level,	Stress, arousal, confusion, trust, engagement, efficacy, motivation,	Ratings of task performance, leadership, team cohesion, teamwork, team processes

Recommendation 3: Use Existing Data Standards to Capture Data

To ensure effective data capture, we recommend adopting existing data standards that provide a structured framework for organizing and representing information. These standards include the Experience API (xAPI; Walcutt & Schatz, 2019), which enables consistent tracking and communication of learning experiences across different platforms. The xAPI specification also emphasizes interoperability, providing policies for disparate training environments to output performance data in a controlled manner for persistent storage and longitudinal modeling. This enables a program like STE to interoperate within a broader ecosystem of training resources that support overall training progression to readiness. Additionally, standards like SensorML and the OGC Sensor Observation Service (SOS) facilitate the capture of low-level and raw event data from various sensors and devices. Furthermore, time series data can be effectively captured using standards such as the OGC TimeseriesML, allowing for the storage and retrieval of time-stamped data for detailed analysis and visualization. By leveraging these standards, organizations can enhance interoperability, streamline data capture processes, and enable comprehensive analysis across diverse data sources and domains.

Recommendation 4: Develop and Pilot Test Data Strategy Models

We recommend developing a focused, phase-based plan to testing data flow models. The PEO STRI pathway to creating the EDS began with developing a focused, phase-based STE Data Architecture – an ongoing pilot effort – that aligns, documents, and codifies the STE ecosystem data across its data sources, architecture, interfaces, technology stack, enabling extension and re-use by other programs in the STE family. The STE Data Architecture focuses on building data flow models, not just diagrams. Diagrams are static and diagram elements are generally superficially defined. Models built with data and software architecting tools such as Structurizr and MkDocstrings and published in the STE Platform Development Kit, are required because they are data-driven and provide a richer set of relevant information about the data flows and structures, and systems that produce and consume them. When considering ML and RL-based policies to guide adaptive training, it is also recognized that models and policies are designed to evolve and optimize over time. This requires acknowledgement that early implementation of AI training management functions may start with a limited set of pedagogical actions that will increase over time as the system learns new techniques through exploration vs. exploitation trade-offs. Here exploration refers to trying new actions whereas exploration refers to using only the actions a system knows will be supportive of student learning outcomes.

Recommendation 5: Protection, Privacy, and Ethical Guidelines

Ensuring the protection of data and privacy, as well as adherence to AI ethical guidelines, is of paramount importance. Ethical considerations for AI are a key element in a more reliable and safe operations under specific development and testing standards. Currently, there no formal ML verification processes that exist today. As a result, DoD officially adopted five Ethical Principles for Artificial Intelligence (DoD, 2020) together with an Executive Steering Group, that include the responsible development and deployment and use of AI, promoting equity and reducing unintended bias in AI capabilities, supporting traceability and transparency, maintaining reliability, and the ability to monitor, govern and control or disengage systems that demonstrate unintended behavior (OUSD, 2023). Further, training programs must establish clear requirements and policies to safeguard sensitive data and mitigate privacy risks. Robust security measures, anonymization techniques, and privacy controls should be implemented to address data protection concerns (Hampton & DeFalco, 2022). Additionally, compliance with AI ethical guidelines is essential to mitigate biases and promote transparency and accountability in AI development and deployment.

CONCLUSION

The U.S. Army's STE and supporting training management tools recognize the increasing role AI will play towards optimizing the use of simulation to support individual and team readiness requirements. Collecting, labeling storing, and managing access data to support AI driven functions and algorithms is essential for meeting this vision. To harness the full potential of the AI services training programs must adopt a proactive and action-oriented data management approach early in the fielding of their systems (Federal News Network, 2022). Training programs should prioritize collecting data at scale, capturing a wide range of multimodal sources, and using adaptive instructional services and data reporting standards to track assessments and contexts across training experiences and environments (Goldberg et al., 2021). Using a systematic approach will ensure that the data collected represents the diversity of training tasks, conditions, and training strategies employed to meet a defined objective. It also establishes a data repository that can be accessed by the research community to help in the maturation of the TMT functions described above. Moving forward, an effort should be made to collect scenario, conditions, performance, and effectiveness data from synthetic and live training environments to support a comprehensive data strategy. An enterprise data strategy for AI/ML in simulation and training will provide improve training effectiveness, reduce costs, and enhance decision-making.

REFERENCES

Allen, G. (2020, April). *Understanding AI Technology: A concise, practical, and readable overview of Artificial Intelligence and Machine Learning technology designed for non-technical managers, officers, and executives*. Retrieved from https://www.ai.mil/docs/Understanding%20AI%20Technology.pdf

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, *96*, 207-210.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Department of Defense (DoD) (2020, January 24). *DOD Adopts Ethical Principles for Artificial Intelligence, U.S. Department of Defense*. Retrieved from https://www.defense.gov/News/Releases/Release/Article/?2091996=dod%E2%80%90adopts%E2%80%90%20ethical%20principles%E2%80%90for%E2%80%90artificial%E2%80%90intelligence

Doroudi, S., Aleven, V., & Brunskill, E. (2019). Where's the reward? a review of reinforcement learning for instructional sequencing. *International Journal of Artificial Intelligence in Education*, *29*, 568-620.

Durlach, P. J., & Spain, R. D. (2014). *Framework for instructional technology: methods of implementing adaptive training and education*. ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES FORT BELVOIR VA.

Fahid, F. M., Rowe, J. P., Spain, R. D., Goldberg, B. S., Pokorny, R., & Lester, J. (2022). Robust Adaptive Scaffolding with Inverse Reinforcement Learning-Based Reward Design. Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, 204–207.

Federal News Network (FNN) Online (2022, April 14). For the Army, getting to great decisions starts with data fabric. Retrieved from https://federalnewsnetwork.com/cme-event/federal-insights/ask-the-cio-army/

Goldberg, B., Owens, K., Gupton, K., Hellman, K., Robson, R., Blake-Plock, S., & Hoffman, M. (2021). Forging competency and proficiency through the synthetic training environment with an experiential learning for readiness strategy. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Orlando, FL*.

Hampton, A. J., & DeFalco, J. A. (Eds.). (2022). The Frontlines of Artificial Intelligence Ethics: Human-centric Perspectives on Technology's Advance. Routledge.

Hernandez, M., Goldberg, B., Robson, R., Owens, K., Blake-Plock, S., Welch, T., & Ray, F. (2022). Enhancing the Total Learning Architecture for Experiential Learning. In *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*.

Hoque, M. E. (2016). Three domains of learning: Cognitive, affective and psychomotor. The Journal of EFL Education and Research, 2(2), 45-52.

Johnson, W. L., & Lester, J. C. (2018). Pedagogical agents: Back to the future. AI Magazine, 39(2), 33-44.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. Review of educational research, 86(1), 42-78.

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. International Journal of Artificial Intelligence in Education, 24, 427-469.

Office of the Under Secretary of Defense for Policy (OUSD) (2023, January 25). DOD DIRECTIVE 3000.09, Autonomy in Weapon Systems. Retrieved from https://www.esd.whs.mil/DD/.

Pavlik Jr, P. I., Brawner, K., Olney, A., & Mitrovic, A. (2013). A Review of Student Models Used in Intelligent Tutoring Systems. In Design Recommendations for Intelligent Tutoring Systems: Volume 1 – Learner modeling.

Rafferty, A. N., Brunskill, E., Griffiths, T. L., & Shafto, P. (2016). Faster teaching via pomdp planning. *Cognitive science*, *40*(6), 1290-1332.

Rowe, J. P., & Lester, J. C. (2015). Improving student problem solving in narrative-centered learning environments: A modular reinforcement learning framework. In *Artificial Intelligence in Education: 17th International Conference, AIED 2015, Madrid, Spain, June 22-26, 2015. Proceedings 17* (pp. 419-428). Springer International Publishing.

Russel, S. & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th Edition). Pearson: Hoboken, NJ.

Sabourin, J. L., & Lester, J. C. (2013). Affect and engagement in Game-Based Learning environments. IEEE transactions on affective computing, 5(1), 45-56.

Sawyer, R., Rowe, J., & Lester, J. (2017). Balancing learning and engagement in game-based learning environments with multi-objective reinforcement learning. In *Artificial Intelligence in Education: 18th International Conference, AIED 2017, Wuhan, China, June 28–July 1, 2017, Proceedings 18* (pp. 323-334). Springer International Publishing.

Shen, S., Mostafavi, B., Barnes, T., & Chi, M. (2018). Exploring induced pedagogical strategies through a Markov decision process framework: Lessons learned. *Journal of Educational Data Mining, 10*(3), 27-68.

Smith, A., Spain, R.D., Rowe, J., Goldberg, B., Lester, J.: Formalizing Adaptive Team Feedback in Synthetic Training Environments with Reinforcement Learning. In: Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym10). p. 117 (2022).

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring (GIFT). *Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate (ARL-HRED)*.

Sottilare, R. A., Burke, S., C., Salas, E., Sinatra, A. M., Johnston, J. H., & Gilbert, S. B. (2018). Designing adaptive instruction for teams: A meta-analysis. International Journal of Artificial Intelligence in Education, 28, 225-264.

Spain, R. D., Rowe, J. P., Goldberg, B. S., Pokorny, R., Hoffman, M., Harrison, S., & Lester, J. C. (2021). Developing Adaptive Team Coaching in GIFT: A Data-Driven Approach. In *TTW@ AIED* (pp. 10-16).

Spain, R., Rowe, J., Smith, A., Goldberg, B., Pokorny, R., Mott, B., & Lester, J. (2022). A reinforcement learning approach to adaptive remediation in online training. The Journal of Defense Modeling and Simulation, 19(2), 173-193.

U.S. Department of Defense, DoD Data Strategy, September 30, 2020, DOD Data Strategy (defense.gov)

U.S. Department of Defense, 2022 National Defense Strategy, October 7, 2022, https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational psychologist, 46(4), 197-221.

Vatral, C., Biswas, G., Cohn, C., Davalos, E., & Mohammed, N. (2022). Using the DiCoT framework for integrated multimodal analysis in mixed-reality training environments. Frontiers in Artificial Intelligence, 5.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Walcutt, J.J. & Schatz, S. (Eds.) (2019). Modernizing Learning: Building the Future Learning Ecosystem. Washington, DC: Government Publishing Office. License: Creative Commons Attribution CC BY 4.0 IGO

Woolf, B. P. (2010). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann.

Wu, C. H., Huang, Y. M., & Hwang, J. P. (2016). Review of affective computing in education/learning: Trends and challenges. British Journal of Educational Technology, 47(6), 1304-1323.