

## Measuring Engagement in VR Environments: A Pilot Training Case Study

**Benjamin Bell, Elaine Kelsey**  
Eduworks Corporation  
Corvallis, OR

benjamin.bell@eduworks.com  
elaine.kelsey@eduworks.com

**Benjamin Nye**  
University of Southern California  
Institute for Creative Technologies

Playa Vista, CA  
nye@ict.usc.edu

**Winston (“Wink”) Bennett**  
Warfighter Readiness Research Division  
711 HPW/RHA

Wright Patterson AFB, OH  
winston.bennett@us.af.mil

### ABSTRACT

The USAF has established a new approach to Specialized Undergraduate Pilot Training (SUPT) called Pilot Training Next (PTN) that integrates traditional flying sorties with VR-enabled ground-based training devices and data-driven proficiency tracking to achieve training efficiencies, improve readiness, and increase throughput. Eduworks and USC’s Institute for Creative Technologies are developing machine learning (ML) models that can measure user engagement during any computer-mediated training (simulation, courseware) and offer recommendations for restoring lapses in engagement. We are currently developing and testing this approach, called the Observational Motivation and Engagement Generalized Appliance (OMEGA) in a PTN context. Two factors motivate this work. First, one goal of PTN is for an instructor pilot (IP) to simultaneously monitor multiple simulator rides. Being alerted to distraction, attention and engagement can help an IP manage multiple students at the same time, with recommendations for restoring engagement providing further instructional support. Second, the virtual environment provides a rich source of raw data that machine learning models can use to associate user activity with user engagement. We have created a testbed for data capture in order to construct the ML models, based on theoretical foundations we developed for a related project reported at MODSIM 2019. We are running pilots through multiple PTN scenarios and collecting formative data from instructors to evaluate the utility of the recommendations OMEGA generates regarding how lapsed engagement can be restored. We anticipate findings that validate the use of ML models for learning to detect engagement from the rich data sources characteristic of virtual environments. These findings will be applicable across a broad range of conventional and VR training applications.

### ABOUT THE AUTHORS

**Dr. Benjamin Bell** is the president of Eduworks, where he leads simulation, training, human-machine interaction, and decision support development. His research has addressed the use of simulation for training and education across a spectrum of applications, including K-12, higher education, military, and national security training. He has held faculty positions, chief executive positions in industry, and leadership roles for several international conferences. He is an adjunct professor at Embry Riddle, holds a PhD from Northwestern, and is a University of Pennsylvania graduate.

**Elaine Kelsey** is a computational linguist with Eduworks Corporation, where she focuses on natural language processing and machine learning. She led development of Eduworks’ automated question generation, assessment, and competency alignment algorithms. She holds multiple bachelor’s and master’s degrees in computer science, linguistics, molecular biology, and biostatistics, and is working towards a MS/PhD in machine learning.

**Dr. Benjamin Nye**, Director of Learning Science at ICT, leads the Army’s Promoting Engagement in Virtual Learning Environments (ENGAGE) project and is co-PI on the ONR Personal Assistant for Life-Long Learning (PAL3) project. His research, recognized for excellence in tutoring and behavior simulation, focuses on technologies to promote learning, and has yielded over 20 peer-reviewed papers, 12 book chapters, 1 book, and 5 open-source projects.

**Dr. Winston “Wink” Bennett** is the Airman Systems Directorate Readiness Product Line Lead. He is a Senior Research Psychologist in the Warfighter Readiness Research Division, 711th Human Performance Wing Air Force Research Laboratory. Through more than 25 years of service in the Air Force research community, he has achieved international recognition as a leader in education, training, and performance measurement research. He has led numerous research products that have since become part of the operational military community and have significantly improved mission effectiveness. He pioneered training, education, and measurement technologies and transitioned research results to operational military, scientific and commercial communities that have produced ground-breaking training technology and research and serve as a foundation for other researchers and practitioners to follow. He is a Fellow of three research societies and the Air Force Research Laboratory.

## Measuring Engagement in VR Environments: A Pilot Training Case Study

**Benjamin Bell, Elaine Kelsey**  
**Eduworks Corporation**  
**Corvallis, OR**  
**benjamin.bell@eduworks.com**  
**elaine.kelsey@eduworks.com**

**Benjamin Nye**  
**University of Southern California**  
**Institute for Creative Technologies**  
**Playa Vista, CA**  
**nye@ict.usc.edu**

**Winston (“Wink”) Bennett**  
**Warfighter Readiness Research Division**  
**711 HPW/RHA**  
**Wright Patterson AFB, OH**  
**winston.bennett@us.af.mil**

### PROBLEM SUMMARY

To support a complex array of current and envisioned missions, the U.S. Air Force trains and educates a large uniformed workforce across diverse Air Force Specialty Codes (AFSCs). Early-stage training in many AFSCs presents Airmen and officer trainees with large corpora of foundational knowledge to master, delivered through a variety of modalities. This content is taught at technical training or flight schools in courses lasting up to several months. Because motivation is a necessary element for learning retention and force readiness, Air Force education and training stakeholders continue to look for ways to engage their constituencies, through gamified interactive learning, simulation, and the use of mixed-reality immersive environments.

To maintain an engaged, motivated cadre of Airmen and officers, OMEGA, via a service-oriented, general-purpose appliance operating in concert with the learning environment, will recommend interventions when end-users are showing signs of a lapse in engagement. These services will be accessible to any AF instructional system, which can automatically or through a human instructor locally effect OMEGA’s recommendations. OMEGA will enable AF and DoD learning environments to identify and adapt to detected lapses in engagement, promoting greater motivation, retention, and readiness.

### PROJECT CONTEXT: PILOT TRAINING NEXT

The test case for OMEGA, selected by the Air Force, is Pilot Training Next (PTN). PTN, now training its third iteration of pilots, aims to reduce the time and cost of undergraduate pilot training (UPT) by leveraging virtual and augmented reality, biometrics, AI, and data analytics. While PTN student pilots fly the same hours and sorties in the T-6 as do their legacy pilot training counterparts, the ground-based training is done using an immersive PC-based flight simulator (Lockheed Martin’s Prepar3D®), VR headset (HTC’s VIVE™ Pro), stick, throttle and rudder pedals, and a syllabus of PTN-specific scenarios (Figure 1).

PTN has maintained a heavy emphasis on data collection and analysis. Reliable and predictive metrics are needed not only to assure instructors and higher command that student pilots are achieving the same skills as their legacy training counterparts, but also because PTN students qualify for graduation on the basis of achievements rather than on calendar time. This fundamental change in how students progress requires an abundance of data that supplement instructor evaluations to ensure skill mastery.

During scenarios flown in the simulator, objective data is readily captured for every time interval, such as aircraft state (position, attitude, airspeed) and configuration (aileron, rudder and elevator deflections; flap and gear positions). Instructors can also monitor how the scenario is progressing overall and can provide verbal feedback in real-time.

Some important metrics though are less directly observable or measurable. Student engagement is widely accepted as a critical mediating factor in both learning retention (*e.g.*, Hu & Hui, 2012) and learning outcomes (*e.g.*, Chi & Wylie, 2014). The importance of engagement is not lost on instructors (though labels like *attention* and *focus* are more common in pilot training), and is often the basis for, or at least an element of, scoring situational awareness (SA). Two additional factors make engagement even more salient for PTN. First, instructors have less visibility into student engagement (due to the VR headset) than in conventional simulators. Second, a vision for PTN is for one instructor to be monitoring multiple students simultaneously. Indirectly-observable measures such as engagement will thus require some level of automation support to cue instructors when lapses are detected.



**Figure 1. PTN station showing VR headset, flight controls and external displays.  
(U.S. Air Force photo by Sean M. Worrell)**

## MODELLING ENGAGEMENT

In training sorties flown in the aircraft or in conventional simulators, an instructor pilot (IP) monitors the student pilot's performance. Maneuvers are evaluated by observing air speed, vertical speed, attitude, angle of attack, and so on. Instructors are also interested in situational awareness (SA), which they can assess by observing the student's ability to "stay ahead of the airplane": anticipating upcoming changes in heading, airspeed, or altitude; applying smooth control inputs to adjust bank angle, pitch and power; and maintaining proper scan of the flight instruments.

Several theoretical frameworks for characterizing engagement informed our model of engagement. We created an inventory of nine relevant engagement and disengagement models from the literature that emphasize behavioral indicators (e.g., data from log files or from direct queries to the user) (Core, et al., 2016). These included Intrinsic vs. Extrinsic (Porter & Lawler, 1968); Two Factor Hygiene-Motivator Theory (Gawel, 1997); Motivators from Maslow's Hierarchy (Ibid); Achievement Goal Theory (Pintrich, 2000); D'Mello & Graesser's Engagement model (2012); and Baker's indicators of passive vs. active disengagement (Baker, Corbett, Roll & Koedinger, 2008). From this we synthesized a multi-timescale engagement and motivation model (Bell, Kelsey & Nye, 2019).

More recently, we refined the model to reflect the aviation focus of this project, preferring metrics associated with *event response* tasks (e.g., maneuvering to avoid a new hazard) and *monitoring* tasks (e.g., maintaining straight and level attitude). We incorporated significant research conducted to identify indicators of distraction and disengagement for accidents attributed to loss of control and airplane state awareness (Harrivel, et al, 2017; Wickens, 2005). A subset of these states is relevant to flight tasks performed in simulated environments:

- Attention vs. Distraction: Situational awareness was particularly reduced by the induction of diverted attention. Channelized attention or attentional tunneling also indicated loss of situational awareness (Harrivel, et al, 2017);
- Boredom and Distraction: Distraction can be characterized by any time without interaction with the system; engaged pilots interacted with the system to optimize performance, even when this was not required to meet performance requirements (Cummings, et al, 2013);
- Attentional Tunneling: Attentional tunneling is indicated by lack of interaction with one or more system elements, coupled with strong interaction with another element (Wickens, 2005);
- Vigilance: Diligence, distraction and daydreaming all lead to failures in practical monitoring tasks (Casner & Schooler, 2015).

The model resulting from this additional analysis is shown in Figure 2. The resulting expert model is based on eight input metrics which are used to compute three mid-level features: Performance, Efficiency, and Responsiveness. An overall composite measure of current engagement within a given data window is derived from these mid-level features. The relative weights and derivation methods shown in Figure 2 represent the initial trial conditions. We anticipate that these will be refined through additional testing.

Composite Measure of Engagement: Inverse weighted average of Performance, Efficiency, and Responsiveness			
Weight	Category	Metric	Calculation
15%	Performance (Over/Under)	Metric 1: Conf(Expected)	Conf(Expected) * % Tasks Missed (z-score) / Expected (z-score)
		Metric 2: % Tasks Missed (z-score)	
		Metric 3: % Expected (z-score)	
35%	Efficiency	Metric 4: Deviation from an ideal path (max(0,  z-score  -1) of closest distance)	Normalized, weighted composite
		Metric 5: Time to complete each segment (z-score)	
		Metric 6: % Resources (Fuel) Remaining (z-score)	
50%	Responsiveness (Reaction Time)	Metric 7: Reaction Time: How long after an 'event' does a detectable change occur? - Normalized by expertise level (% Expected) - Rapid response discounted (if incorrect)	Normalized, weighted composite
		Metric 8: Corrections - How long to correct after going off-course	

**Figure 2. Engagement Model**

## ENGAGEMENT METRICS AND VIRTUAL REALITY

A desktop flight simulator generates a rich set of data, including aircraft position, attitude and configuration. From such data, objective performance metrics can be calculated with some reliability. For instance, detecting when a student pilot lowers the gear while the airspeed exceeds the maximum gear-down speed is straightforward. To monitor engagement, however, requires aggregating observable measures to generate an indirect estimate of engagement. Our model, for instance, specifies eight such indirect measures.

The addition of a VR head-mounted display (HMD) adds additional data points that could be incorporated as part of a suite of metrics to monitor engagement levels. Typical VR headset and sensors can capture head position and movement; higher-end devices, such as the VIVE Pro Eye, can capture eye tracking data. The VR environment thus adds to the already rich data stream available from the simulator. This apparent abundance of data, however, does not solve the problem of developing reliable measures of engagement. Several challenges for interpreting the data remain, including, non-exhaustively:

1. Understanding which data points are relevant to engagement;
2. Setting proper coefficients representing how each data point should be weighted;
3. Distinguishing between and properly applying a single data point  $x$  observed at time  $t$  compared with a trend of how  $x$  behaves over some interval (*e.g.*, from  $t - 5$  seconds to  $t + 5$  seconds).
4. Incorporating the velocity of the change in a data point, for instance, how abrupt an aileron deflection or throttle movement the student applied.

A principal emphasis of this work is to explore the role that machine learning models could play in interpreting simulator and VR device data in order to develop measures to drive our conceptual model of engagement. This machine learning approach is summarized in the next section.

## MACHINE LEARNING MODEL

We employ machine learning to allow OMEGA to develop more accurate predictive associations between raw data inputs and higher-level aggregated engagement metrics. This section describes the techniques and architecture of the OMEGA machine learning component. Our design leverages the underlying data streams available from Prepar3D to provide better predictive power in situations where there is limited access to interpreted data (e.g. when interpreted metrics of event occurrence, event success/failure, and efficiency are not available). To achieve this, we employ three methods in combination:

1. **Standard machine learning.** We use standard machine learning techniques to attempt to accurately predict engagement and disengagement in input metric sequences. These approaches are attractive because they enjoy fast estimation methods with low run-time, and therefore can provide near-instant feedback to instructors. Based on the features and data available in Prepar3D, we have selected Support Vector Machine (SVM) and Binomial Regression as the most applicable approaches. These techniques are most powerful in cases where sequence classification is not strongly context-dependent. For OMEGA, however, we expect some context-dependence in the data. For example, rapid adjustments of heading, altitude and airspeed may represent recovery from a period of inattention if these maneuvers occur between waypoints, but may represent an attentive reaction if observed during an event requiring active response (e.g. a heading change when passing a waypoint). To mitigate this risk and to improve the model, we deploy two additional “deep learning” layers of machine learning that are more robust to sequence classification in context-dependent data.
2. **Deep Learning - Bidirectional LSTM (long short term memory).** We use bi-directional LSTM (a type of recurrent neural network) to produce improved results in sequence classification problems that are heavily context-dependent. In this case, determining whether a given sequence of composite low-level metric readings represents disengagement is likely to be highly context-dependent, for instance a climb at full throttle versus a climb at normal cruise speed. Bidirectional LSTMs, which consider the ‘context’ of both the preceding and following time slice data when predicting disengagement, are used in OMEGA to reduce the incidence of false-positive detection of disengagement in this environment. The tradeoff for improved disengagement and inattention detection is the high resource and time cost of maintaining bidirectional LSTM in OMEGA. Recurrent Neural Networks (RNNs) like LSTM are can be difficult to train due to memory-bandwidth-bound computation limitations. For this reason, we have selected a second deep learning approach in case the performance requirements present too much risk for OMEGA.
3. **Attention-based Modeling.** The alternative deep learning approach we selected as the third layer in OMEGA’s machine learning stack is Attention-based Modeling. Attention-based models are sequence-to-sequence models designed to improve performance of RNN-based approaches. We use attention-based modeling as the third layer as a possible performance throttling mechanism in cases where bidirectional LSTM is too resource-intensive for OMEGA to be effective in real time.

## TRAINING THE MODEL

Data for training the machine learning components derives from experimental subjects who fly a pre-selected set of PTN scenarios in a data collection station that mirrors most of a PTN simulator, namely, the simulation software, stick and throttle, and VIVE Pro HMD and sensors. The data collection station also includes a dedicated application for the experimenter to monitor each scenario, interact with the subject, and time-stamp relevant events. Figure 3 shows an experimenter and subject during a data collection session.



**Figure 3. Data collection PTN station (background), experimenter station (foreground). Photo by the authors.**

For purposes of creating training data for the machine learning models, experimenters are trained in a protocol to (1) time-stamp lower-intensity and higher-intensity segments of a scenario, to help the models account for workload in processing measures of user activity; and (2) engage the subject in conversation at specific points during a scenario. Conversing with the subject acts as a surrogate for disengagement. We posit that loss of attention, or distraction, will be statistically detectable in the simulation log files. Specifically, we anticipate three possible types of deviations:

- 1) **Response Time:** Most dominantly, we anticipate that subjects' time to respond to changes in the environment will be slower and/or less precise when engaging in conversation. Specifically, we anticipate a longer duration with no response after an event that requires a maneuver (*e.g.* heading change), followed in some cases by an initial control input that is more abrupt, more prone to overcorrect, or may even be in the wrong direction.
- 2) **Performance:** We anticipate more likely failure to accomplish scenario goals (*e.g.*, missing required waypoints).
- 3) **Efficiency:** We anticipate that periods of distraction will tend to be less efficient, due to the above issues and due to less precise control over the aircraft (*e.g.*, slower damping of over-correcting heading changes).

Subjects were recruited from flying clubs in the Corvallis, OR and Los Angeles regions. Subjects qualified for the study through meeting either flying hour criteria or flight simulator experience criteria. Each subject was given a practice period with the PTN station and then asked to complete six PTN scenarios.

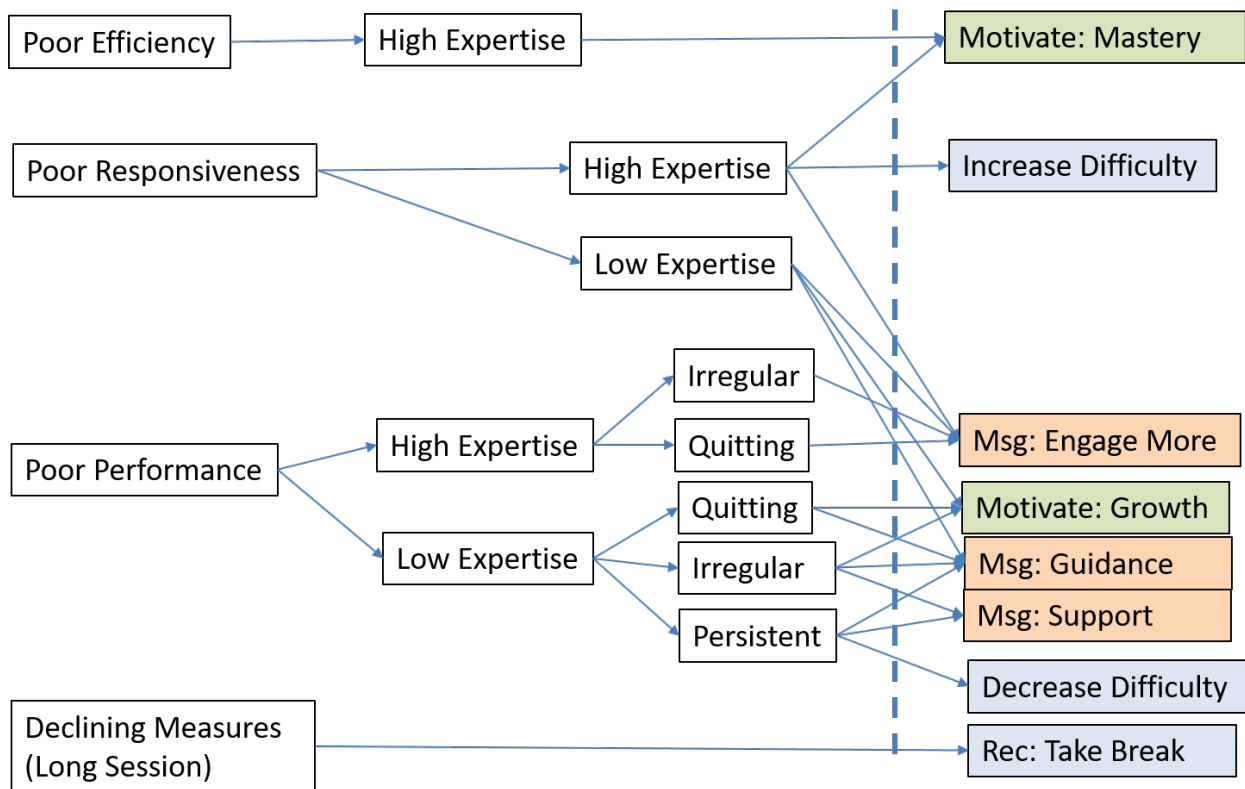
The collected data will be used to train the machine learning model, comparing both the predictive power and the latency and resource requirements for each of our deep learning modeling techniques.

### **ADAPTIVE RECOMMENDATIONS MODEL**

OMEGA processes detected engagement levels to generate adaptive recommendations to help an instructor restore lapsed engagement. During a scenario, based on combinations of different state signals, OMEGA will generate a set of intermediate inferences. These inferences include, for example, whether poor performance is due to consistently bad results versus irregular behavior or inconsistency (*e.g.*, carelessness). The model employs both the basic state model and the aggregated inferences as inputs to calculate a scoring ranking for different adaptive interventions.

Our model considers three levels of outcomes: performance, responsiveness, and efficiency, each representing a distinct dimension of quality. Performance represents the basic ability to complete the assigned tasks, based on the performance criteria for those tasks (e.g., following a set of waypoints). Responsiveness represents the speed and effectiveness for a learner to adjust to new tasks or requirements (e.g., if a waypoint is moved, how quickly does the user adjust heading). Efficiency represents lean and strategic use of resources to complete a scenario (e.g., faster completion times).

These quality criteria can each be thought of as building upon each other: a learner must adjust heading to a new waypoint or else there is no way to determine responsiveness. Likewise, efficiency is impossible if the user is not responsive enough to stay on course. This means that only some of these factors should be addressed with certain types of learners (e.g, high vs. low expertise). For example, if a user is failing to master proper take-off procedures, critiquing fuel efficiency would be out of place and of no training value. On the other hand, an otherwise high-performing student pilot who is drifting off-course or leaving assigned altitudes may benefit from noting a need for improved in-flight checks.



**Figure 4. High-Level Policy for Adaptive Interventions in PTN**

The right-hand side of Figure 4 shows the interventions proposed for PTN. These include three distinct types of interventions: Messaging (Information about the task), Motivation (Context about the task and learning goals), and Recommendations (Suggestions on different tasks or breaks to improve learning). The left-hand side of Figure 4 outlines a high-level policy for when specific interventions are expected to be appropriate for users with different skill levels and in different states. These connections between intervention types and student states do not represent the actual model. Instead, they represent key dynamics that the model will produce. However, since the actual state space to calculate an effective intervention policy is too large to easily convert into a short graph, this model captures the key behaviors that the intervention model will be tested against, to ensure it behaves reasonably versus what would align to theoretical frameworks for engagement and responding to disengagement. The intervention types are outlined in Table 1.

**Table 1. Intervention Types for PTN**

Category	Intervention Type	Description
Messaging	Engage More	A suggestion or warning that signs of decreased engagement were noticed, with a suggestion on how to improve
	Guidance	A useful hint about how to do better on the specific scenario or skills
	Support	Affective feedback to help a learner who is struggling
Motivate	Mastery	Information about how better training can improve later real-life performance
	Growth Mindset	Feedback about how sustained effort leads to improved skills and outcomes
Recommend	Increase Difficulty	Suggest that the user might need more challenging scenarios to engage with
	Decrease Difficulty	Suggest user might benefit from trying some easier scenarios before this one
	Take Break	Suggest that user returns after taking a break, due to decreasing performance

## GENERATING ADAPTIVE RECOMMENDATIONS

We propose two distinct methods for generating adaptation recommendations based on the internal state of the models used to measure engagement. The first approach is much less computationally-intensive and will produce recommendations with lower latency. However, given the highly contextualized nature of the input data stream, we expect the second approach to produce more accurate results. Work is currently in-progress for testing the trade-offs between timeliness and quality under different simulation conditions.

In the first approach, we use the calculated values from the expert model of engagement (performance, efficiency, and responsiveness) as inputs to a machine learning classification model. Using the labeled data set produced during the data collection trials discussed above, we will apply several traditional machine learning modeling techniques to the classification task, where the outputs are the available set of adaptations available in the PTN training environment. We will use both Naive Bayes and Support Vector Machine (SVM) approaches. Given that the original input metrics include variables that are highly interdependent, we expect that SVM will yield superior results. SVMs have been demonstrated to predict the likelihood of learner withdrawal from online courses, for example (Kloft, Stiehler, Zheng & Pinkwart, 2014).

These techniques do not account for the contextual nature of the data stream, instead analyzing each time slice as a separate case. As was the case for detecting engagement levels, therefore, the determination of an appropriate adaptation will depend highly on the context in which the disengagement event occurs, as well as on the environmental conditions being simulated.

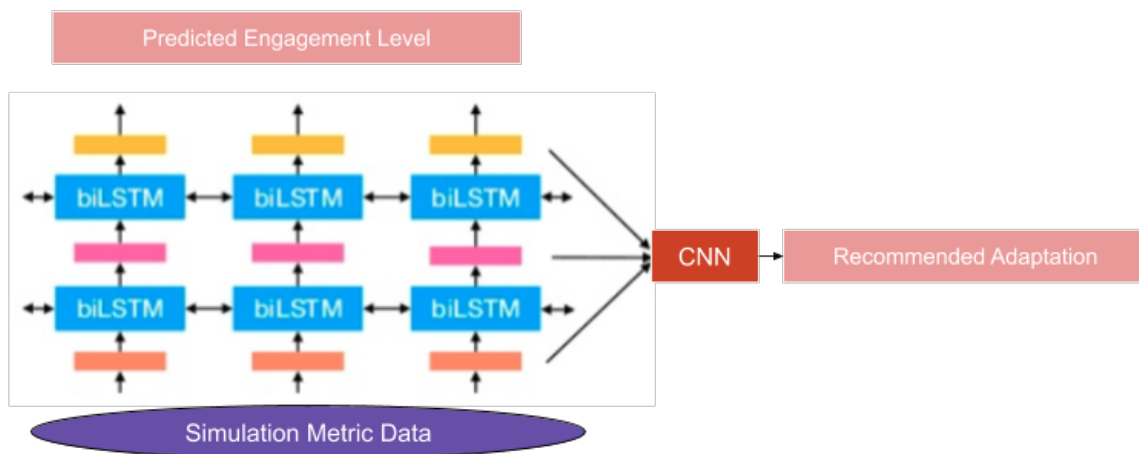
We define *context* as the data stream of pilot behaviors and actions preceding and following the time slice of interest, and *environment* as the set of conditions that obtain for that particular segment of the simulation (e.g. aircraft attitude, airspeed, vertical speed, status of systems). In order to fully account for the contextual nature of both the disengagement event detection and the recommendation of an appropriate adaptation, we are developing a second, more powerful model for capturing temporal information and learning high-level representations hidden in the metric data stream based on Artificial Neural Networks (ANN).



Until recently, much of the research in applying ANN to the interpretation of sensor data streams focused on traditional neural network approaches, such as feed-forward neural networks (FFNN) and deep convolutional neural networks (CNN). Al-Shabandar, *et al.* (2018) employed a range of machine learning models including ANNs to investigate factors driving student motivation in massively-online open courses (MOOCs). Recent success of recurrent neural networks (RNN) with long short-term memory (LSTM) in other applications has led to promising trials of this approach in several problem areas involving the use of sensor data streams to predict highly contextual operational states. RNNs have been used for, among other applications, associating motivation and engagement of students with outcomes in MOOCs (Piech, *et al.*, 2015).

More recently, architectural variations of this approach have been explored that also allow for the incorporation of operational conditions into the predictive model. These approaches use several bidirectional LSTM (BD-LSTM) models and a final FFNN layer to integrate both the contextual information encoded in the data stream (representing the sensor data) and the available operating context and environmental data. The model we have designed adapts this approach to the interpretation of Prepar3D data for (1) predicting the level of pilot engagement and detecting disengagement events; and (2) using contextual information about the nature of the disengagement to predict the most appropriate adaptation to recommend to the human instructor in the context of PTN training.

The model is composed of several stacked layers of ANNs. The first BD-LSTM network extracts latent features from the multiple metric data streams describing pilot behavior. The second BD-LSTM network extracts latent features from the metric data stream describing aircraft movement, and the third BD-LSTM network extracts higher level features describing the operational environment. These layers are stacked with a final neural network layer to predict pilot disengagement level and events, depicted schematically in Figure 5. The states of the internal layers of these BD-LSTM networks are then used as the input into a separate recommendation model to predict the most appropriate adaptation in a given context. The recommendation model layer is a CNN network, which will be trained using the labeled data set from the pilot trials.



**Figure 5. Adaptation recommendation stacked neural network architecture**

## PROJECT STATUS

We are currently concluding the data collection phase, anticipating data from twenty-five subjects, each flying six PTN scenarios. Results from the machine learning model development will be presented during the conference. A formative evaluation using Air Force instructor pilots to provide feedback to OMEGA's recommendations will immediately follow the model development.

## CONCLUSION AND FUTURE WORK

Simulation enhanced with VR provides immersive training that promises to advance learning outcomes and retention. A key factor in achieving positive results is learner engagement, which, in contrast to directly observable or objectively measurable factors, is challenging to assess. In some instances, the VR environment itself can obscure from instructor view some cues relevant to learning engagement. OMEGA addresses this gap by using machine learning models to develop predictive associations between simulation events and learner actions on the one hand, and learner engagement on the other. OMEGA also incorporates a model of adaptive interventions to remedy engagement lapses, and employs machine learning to develop associations between the context and environment of the engagement lapse and the optimal intervention to recommend to an instructor.

Our results will provide concept validation to establish more general-purpose, service-oriented appliance that client learning applications can employ for detecting lapses in engagement and motivation, and for recommending adaptive interventions. OMEGA can thus address a need, across the service branches, to ensure that simulation-based training, and training incorporating VR, results in engaged and motivated warriors, using adaptive instruction and providing data to help training managers track the efficacy of new technologies and paradigms.

## REFERENCES

- Al-Shabandar, R., Hussain, A. J., Liatsis, P., and Keight, R. (2018). Analyzing learners behavior in MOOCs: An examination of performance and motivation using a data-driven approach. *IEEE Access*, 6, 73669-73685.
- Baker, R. S., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction*, 18(3), 287-314.
- Bell, B., Kelsey, E., and Nye, B. (2019). Monitoring Engagement and Motivation Across Learning Environments. In *Proceedings of the 2019 MODSIM World Conference*. Norfolk, VA.
- Chi, M. T., and Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, 49(4), 219-243.
- Core, M.G., Georgila, K., Nye, B.D., Auerbach, D., Liu, Z.F., DiNinni, R. (2016). Learning, adaptive support, student traits, and engagement in scenario-based learning. In *IITSEC*, 2016.
- Casner, S. M., & Schooler, J. W. (2015). Vigilance impossible: Diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. *Consciousness and cognition*, 35, 33-41.
- Cummings, M. L., Mastracchio, C., Thornburg, K. M., & Mkrtychyan, A. (2013). Boredom and distraction in multiple unmanned vehicle supervisory control. *Interacting with Computers*, 25(1), 34-47.
- D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145-157.
- Gawel, J. E. (1997). Herzberg's theory of motivation and Maslow's hierarchy of needs. *Practical Assessment, Research & Evaluation*, 5(11), 3.
- Harrivel, A. R., Stephens, C. L., Milletich, R. J., Heinich, C. M., Last, M. C., Napoli, N. J., ... & Pope, A. T. (2017). Prediction of cognitive states during flight simulation using multimodal psychophysiological sensing. In *AIAA Information Systems-AIAA Infotech@ Aerospace* (p. 1135).
- Hu, P. J. H., and Hui, W. (2012). Examining the role of learning engagement in technology-mediated learning and its effects on learning effectiveness and satisfaction. *Decision support systems*, 53(4), 782-792.
- Kloft, M., Stiehler, F., Zheng, Z., and Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. *Proc. EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs* (60-65).
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., and Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in neural information processing systems* (pp. 505-513).
- Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning & achievement. *J. of Educational Psych*, 92(3), 544.
- Porter, L. W., & Lawler, E. E. (1968). Managerial attitudes and performance.
- Wickens, C. D. (2005). Attentional tunneling and task management. In *2005 International Symposium on Aviation Psychology* (p. 812).