Optimizing Readiness Through AI-Driven Analytics for Automated Training Insights

Svitlana Volkova, Summer Rebensky, Isabel Erickson, Louis Penafiel, Hsien-Te Kao
Aptima, Inc.
Woburn, MA
svolkova@aptima.com, srebensky@aptima.com, ierickson@aptima.com, lpenafiel@aptima.com, hkao@aptima.com

ABSTRACT

Currently, military training needs assessments and curriculum reviews rely on time-intensive manual analysis of actual student and supervisor surveys and interviews, introducing lag, inconsistency, and lack of scalability. For example, analyzing survey responses from just 100 students can take a team of 5 analysts over a month to complete manually. The proposed integrated AI analytics, which leverage real student and supervisor data, aim to fundamentally transform this process. In this paper we demonstrate novel AI-powered analytics that automatically analyze authentic student and supervisor survey responses to: (1) identify training gaps and effectiveness issues; (2) understand differences across career field needs; and (3) assist in reworking surveys and evaluation interviews for more actionable insights. Compared to traditional manual qualitative assessment approaches which can take months and involve large teams, these AI techniques can provide results in a matter of hours with minimal human involvement. Specific AI-powered analytics applied to real student data include topic modeling to automatically identify themes related to gaps and effectiveness; syntactic parsing and semantic embeddings clustering to extract detailed insights; connotation analysis to assess student opinions and experiences at scale; and large language models to generate redesign recommendations. In our example use case analyzing actual student surveys, these methods revealed specific pain points such as outdated equipment limiting realism and insufficient hands-on time. The integrated AI analytics fundamentally advance the state of the art, moving from periodic manual surveys toward dynamic AI-assisted assessment and improvements. This will provide agile, optimized expertise development to sustain readiness. Longer-term, the proposed AI tools will boost consistency, efficiency, and personalization of training via continuous automated analysis of human feedback. The paper and presentation will cover the methods used, the benefits and challenges of these approaches and feedback for government and industry on implementing similar processes within their own training evaluation programs.

ABOUT THE AUTHORS

Dr. Svitlana Volkova, Chief of AI at Aptima, leads the company's efforts in developing trustworthy and human-centric AI systems that address complex real-world challenges for the Department of Defense and other government agencies. Her research advances natural language processing and machine learning techniques, with a focus on graph neural networks, causal inference, and multimodal models. Dr. Volkova's pioneering work in AI-powered analytics has made significant contributions to explaining complex social systems and behaviors. A recognized leader in human-centered AI design, evaluation, and trustworthy AI systems, Dr. Volkova has spearheaded projects funded by DARPA, IARPA, SOCOM, DOE ASCR, and NNSA, advancing AI capabilities to support critical national security missions. Her expertise is reflected in over 70 peer-reviewed publications spanning AI, machine learning, and social media analytics. Dr. Volkova serves on program committees and review boards for top-tier AI conferences, including AAAI, ACL, EMNLP, and NeurIPS. As an advocate for diversity in technology, she is an active member of Women in Machine Learning. Dr. Volkova earned her PhD in Computer Science from Johns Hopkins University, where she was affiliated with the Center for Language and Speech Processing and the Human Language Technology Center of Excellence. Her forward-thinking approach and commitment to responsible AI development position her at the forefront of shaping the future of safe, secure, and trustworthy AI that align with human values and societal needs.

Dr. Summer Rebensky, Scientist, Aptima, Inc. has expertise focusing on human performance, cognition, and training in emerging systems. In her role at Aptima, Inc. she serves as the capability lead for Air Force Training, Learning, and Readiness technologies. She specializes in leading efforts to develop, test, and implement AR, VR, XR, and modern training solutions to improve and measure human performance. Dr. Rebensky has previous experience as a research fellow as a part of the Air Force Research Laboratory's Gaming Research Integration for Learning Laboratory

(GRILL) conducting research on interface designs for novel aviation use cases in the civilian and DoD world as well as and human-agent teaming designs. Her other experience includes leading Florida Tech's ATLAS research lab with grant and STTR research with the Air Force, Navy, and FAA; human factors assessments of a trainer aircraft for the F-35 with Northrop Grumman; and developing DoD courseware with Raytheon. Her research experience involves leveraging VR, AR, and modern technology to optimize human performance in training and operations, individualizing learning through gamification and adaptive technologies, human-agent teaming, and trust in AI. Dr. Rebensky received her BA in psychology, MS in aviation human factors, and PhD in aviation sciences with a focus in human factors from Florida Tech.

Ms. Isabel Erickson, Research Engineer, Aptima, Inc., brings a multidisciplinary background spanning biomedical engineering, human factors, data science, and project management. She uses her unique combination of skills across many R&D projects ranging from predictive heat stress monitoring, cognitive assistants for intelligence, surveillance, and reconnaissance (IS/R), and advanced natural language processing for social media and generative conversation analysis. She also brings knowledge of design control of highly regulated systems such as medical devices with software, mechanical, and electrical components through her 3 years of experience at Medtronic and GE Healthcare. She has an MS in industrial and systems engineering, and a BS in biomedical engineering and biomechanics from the University of Wisconsin-Madison.

Mr. Hsien-Te Kao, Associate Research Engineer, Aptima, Inc., brings a diverse array of skills to his role, encompassing computational social science, behavioral modeling, personalization, mobile health, and decision-making. His innovative research at Aptima leverages advanced machine learning analytics to explore multifaceted online information landscapes, ranging from narratives in conflict dynamics such as Ukraine-Russia to discussions on eating disorders and military video comments. Recently, he has contributed insightful findings on eating disorder online communities, presented at CHI 2024. He is a Computer Science PhD candidate at the University of Southern California.

Mr. Louis Panafiel, Research Engineer and Lead, Aptima, Inc., works with the Artificial Intelligence Technologies capability—a portfolio of research projects focused on modeling and understanding natural language discourse, patterns of life, and planning. His work spans the various fields of machine learning and artificial intelligence, such as natural language processing, reinforcement learning, and recommendation systems. His roles vary from algorithm development, project manager, and technical/proposal writing, among others. His leadership of internal research projects contributed to the funding of a DARPA AIE topic. Prior to Aptima, he conducted data science research at CERN's Large Hadron Collider and NASA's Jet Propulsion Laboratory. Mr. Penafiel holds an MS in physics from Cornell University and a BS in physics and mathematics from the University of California, Riverside. Through his work at Aptima, he has contributed to various conference presentations, such as Advanced Maui Optical and Space Surveillance Technologies Conference and SPIE and has a publication in Monthly Notice of the Royal Astronomical Society for one of his research projects.

Optimizing Readiness Through AI-Driven Analytics for Automated Training Insights

Svitlana Volkova, Summer Rebensky, Isabel Erickson, Louis Penafiel, Hsien-Te Kao
Aptima, Inc.
Woburn, MA
svolkova@aptima.com, srebensky@aptima.com, ierickson@aptima.com, lpenafiel@aptima.com, hkao@aptima.com

INTRODUCTION

As George Forbes, Director of HAF Digital Operations Directorate said, "We can shift from spending time doing manual tasks – like putting information into computers – and move to more cognitive techniques where we can analyze the data because the computer is doing much of the busy and manual work." (Air Force, 2023).

The U.S. military relies on effective training to ensure readiness and operational success. However, the current process for assessing training needs and reviewing curricula is time-consuming and manual, involving surveys and interviews that are analyzed by hand. This approach introduces lag, inconsistency, and a lack of scalability, limiting the agility and effectiveness of training improvements.

To address these challenges, we propose an integrated AI-driven analytics approach that transforms the training evaluation process. By leveraging advanced natural language processing (NLP) and machine learning (ML) techniques, our system automatically analyzes survey responses to identify training gaps, understand differences across career fields, and provide recommendations for improving surveys and evaluation interviews.

The proposed AI-powered analytics includes topic modeling to identify themes related to gaps and effectiveness in training, syntactic parsing and semantic embedding clustering to extract detailed insights to reason about training effectiveness (Blei and Lafferty, 2009; Volkova et al., 2021), connotation frame analysis to assess student opinions and experiences at scale (Rashkin et al., 2015; Rashkin et al., 2017), and large language models (LLMs) to generate redesign recommendations (Borders and Volkova, 2021; Horawalavithana et al., 2022). Through an example use case, we demonstrate how these methods can surface pain points such as outdated equipment limiting realism and insufficient hands-on time. Our approach represents a significant advancement in the state of the art, moving from periodic manual surveys to dynamic, AI-assisted assessment and continuous improvement. By providing agile, optimized expertise development, this system has the potential to greatly enhance military readiness. In the longer term, the proposed AI analytics will boost consistency, efficiency, and personalization of training through continuous automated analysis of human feedback.

In this paper, we present the methods used in our integrated AI analytics approach, discuss the benefits and challenges of these automated evaluation methods, and provide feedback for government and industry on implementing similar processes within their own training evaluation programs. We believe that this work represents an important step towards leveraging AI to enhance military training and readiness.

STATE OF THE ART

Currently, the analysis of surveys and interviews in military training assessments and curriculum reviews relies heavily on manual, time-intensive processes. Researchers and analysts typically employ qualitative data analysis software, such as NVivo, ATLAS.ti, or MAXQDA, to organize, code, and interpret the textual data collected from open-ended survey questions and interview transcripts (Woods et al., 2016). These tools allow users to manually assign codes or themes to specific segments of text, facilitating the identification of patterns and trends across the dataset.

While these qualitative data analysis tools provide a structured approach to organizing and analyzing textual data, they have several limitations. First, the coding process is largely manual, requiring significant time and effort from the analysts. This can lead to inconsistencies in coding, especially when multiple researchers are involved, and can limit the scalability of the analysis (Kaefer et al., 2015). Second, these tools primarily rely on the expertise and interpretation of the analysts, which may introduce bias and subjectivity into the findings. Third, the manual nature of the analysis makes it challenging to quickly adapt and iterate on the coding scheme as new insights emerge, leading to a lack of agility in the process.

In recent years, there has been a growing interest in leveraging natural language processing (NLP) and machine learning (ML) techniques to automate and enhance the analysis of qualitative data. Researchers have explored the use of topic modeling algorithms, such as Latent Dirichlet Allocation (LDA) (Blei et al., 2003) and Non-negative Matrix Factorization (NMF) (Lee & Seung, 1999), to automatically discover themes and topics in large collections of text data. Additionally, sentiment analysis and opinion mining techniques have been applied to gauge the overall sentiment and identify specific aspects or features mentioned in the text (Liu, 2012).

Our approach differs from previous methods in several key ways. First, we integrate multiple advanced NLP techniques, combining topic modeling, connotation frame analysis, and automated insight extraction for a comprehensive analysis. Second, our approach is specifically tailored to address the unique challenges of military training evaluation, going beyond general educational applications. Third, we employ connotation frames instead of simple sentiment analysis, providing a more nuanced understanding of opinions and power dynamics expressed in responses. Finally, our novel insight extraction method using semantic embedding and clustering preserves narrative coherence while identifying distinct themes, offering more actionable insights than traditional qualitative analysis methods. These innovations allow for a more detailed, context-aware, and actionable analysis of military training effectiveness compared to existing approaches in the field.

The adoption of these advanced NLP and ML techniques in the context of military training assessment and curriculum review has been limited. The proposed integrated AI analytics approach aims to address this gap by leveraging state-of-the-art methods, such as connotation frames (Rashkin et al., 2015) and transformer-based language models (Vaswani et al., 2017), to provide a more nuanced, contextual, and scalable analysis of survey and interview data. By automating the extraction of insights and generating actionable recommendations, this approach has the potential to significantly improve the efficiency, consistency, and agility of the training evaluation process.

ANALYSIS OF INTERVIEW FINDINGS

The following sections outline the three main analyses conducted to analyze interview data from 162 supervisors and 244 graduates from an Air Force base (provided by the Air Force base). Any references to career fields, air frames, or air forces bases have been anonymized by request of the collaborators and are replaced by generic terms within brackets when referenced. Interview questions prompted the graduates’ current supervisors and graduates’ themselves to reflect on how well the Air Force base prepared them for their following assignment. Questions included prompts related to knowledge, preparation, equipment currency, training, confidence, necessary supplemental training, negative impacts to training, and knowledge and skills required to perform the task. NLP and ML methods were used to automatically extract themes and findings from the qualitative responses to each question. Each approach, the methods used, and key insights are detailed in the sections below.

Topic Analysis

The topic analysis was performed using Structural Topic Modeling (STM), a variant of the Latent Dirichlet Allocation (LDA) method, a widely used generative probabilistic model for discovering abstract topics in a collection of documents (Blei et al., 2003). LDA is an unsupervised learning algorithm that assumes each document is a mixture of various topics, and each topic is characterized by a distribution over words (Blei and Lafferty, 2009). STM goes one step further by utilizing metadata about documents to improve the assignment of words to latent topics in a corpus (Roberts, 2019). The algorithm aims to discover the hidden topic structure in the document collection by iteratively assigning words to topics and updating the topic distributions across specified metadata. In the context of analyzing the interview data from the Air

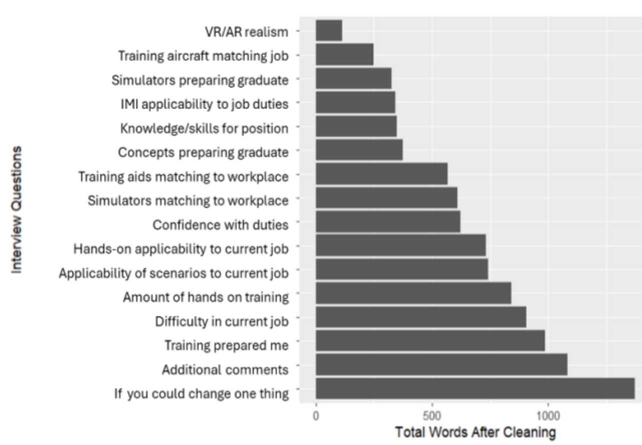

Figure 1. Word Count per Question after Cleaning

Force base, STM was applied to automatically identify themes related to training gaps and effectiveness issues across the survey responses. Before topic modelling, text data was automatically cleaned: punctuation removed; uninformative words filtered; words stemmed to normalize tense.

Survey Comment Statistics

The findings could be utilized by stakeholders to identify which questions would be most valuable to include in future surveys to reduce survey burden and fatigue on the operational units. It can also be used as a preliminary analysis to determine which questions are the highest value to dig into further using NLP methods discussed later in this paper. In addition, these findings might indicate the most and the least problematic areas. As seen in Figure 1, questions that prompted students whether VR/AR was realistic and if the training aircraft matched the job had the least words in their answer. However, the additional comment question and if you could change one thing had more words in their responses. For example, besides open-ended responses like “If you could chance one thing” and “additional comments”, student focused on describing difficulties in current jobs and elaborated on how training prepares them.

Topic Analysis Findings

In Figure 2, the topic prevalence is denoted as "(gamma)" following standard practice in topic modeling. Gamma represents the probability distribution over topics for each document in the corpus. In other words, it indicates the proportion of words in a document that are associated with each topic. This is distinct from a simple frequency count of words in the document, as it considers the probability of each word belonging to a specific topic based on the overall distribution of topics across the entire corpus.

The topic analysis uncovers seven main topics across the questions in order or prevalence: (1) On-the-job (OTJ) was difficult, (2) Adding hands on time, (3) Wanting help with concept and theories, (4) the [aircraft] simulation, (5) the IMI student instructor workflow, (6) [Aircraft] scenario, and (7) Similar outdated equipment at [Air Force Base]. The visualization of the topic prevalence in Figure 2 can be view horizontally as well as vertically. For example, for the question related to “Training aircraft matching job” the topics of “[Aircraft] scenario”, “[Aircraft] simulation]”, and “Simulator equipment outdated at [Air Force Base]” are present. This analysis quickly shows that scenarios and simulations may be outdated, or that simulators and scenarios help address physical equipment that is currently outdated. The visualization presented here informs key stakeholders of potential training gaps or challenges that necessitate either future investigation in those areas—through NLP analysis or targeted follow-up interviews—or through funded improvements to training. Some methods for qualitative analysis of themes are described in later sections.

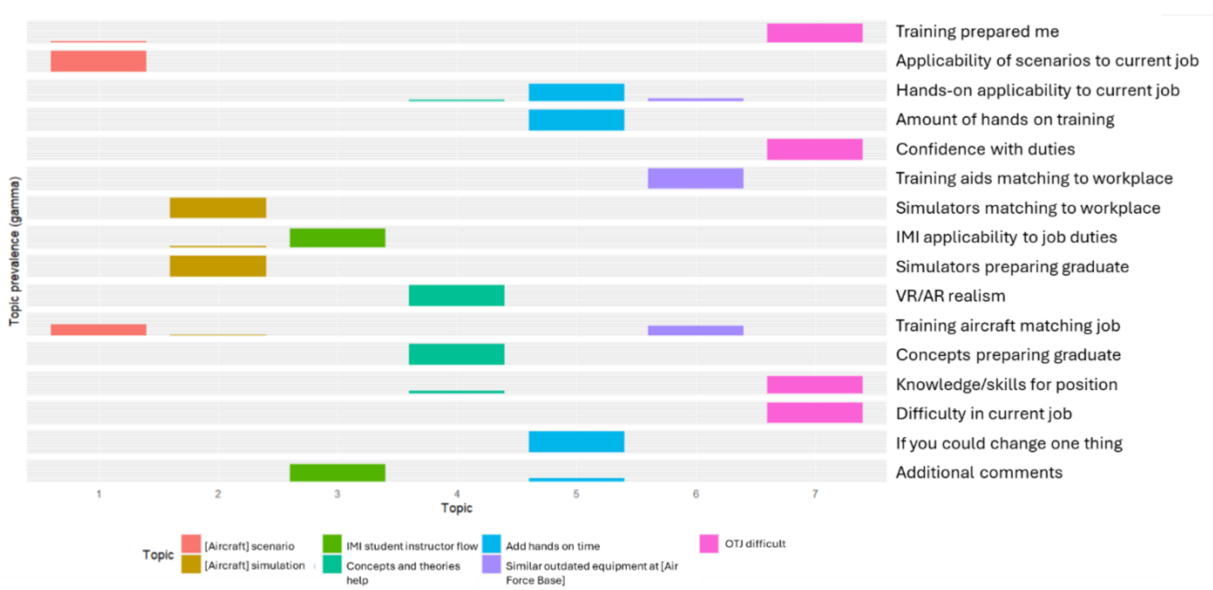

Figure 2. Topic Prevalence by Survey Question.

Connotation Frame Analysis

The connotation analysis approach used in this study goes beyond simple sentiment analysis (Liu, 2012) by leveraging connotation frames, as introduced by Rashkin et al. (2015). Connotation frames provide a more nuanced and fine-grained understanding of the opinions and experiences expressed in the survey responses, capturing the subtle shades of meaning that may be overlooked by traditional sentiment analysis methods. Connotation frames are a collection of transitive verbs that reveal attitudes towards their subjects/objects: perspectives of writer/reader; value of subject/object; state of subject/object and effect on subject/object. For example, in the sentence "The instructor praised the student", the verb "praised" has a positive connotation on both the instructor (the subject of the sentence) and the student (the object). However, in the sentence "the scenario failed the student" the verb "failed" has positive connotation on the student but negative connotation on the scenario.

Connotation frames are designed to capture the implied sentiment and power dynamics between the subject and object in a sentence (Rashkin et al., 2017). The approach involves identifying the subject-verb-object (SVO) structure in each sentence and analyzing the relationship between these elements. By considering the connotative meanings of the words used, the method can determine whether the subject is perceived as having more or less power than the object, and whether the perspective expressed is positive or negative. The key advantages of using connotation frames over simple sentiment analysis are:

1. **Contextualized sentiment:** Connotation frames consider the relationship between the subject, verb, and object in a sentence, providing a more contextualized understanding of the sentiment expressed (Rashkin et al., 2015).
2. **Power dynamics:** By analyzing the implied power dynamics between the subject and object, connotation frames can capture subtle nuances in the opinions and experiences shared by the respondents (Rashkin et al., 2017).
3. **Fine-grained analysis:** Connotation frames enable a more detailed and fine-grained analysis of the survey responses, helping to identify specific aspects of the training that are perceived positively or negatively by the trainees.

Connotation Frame Analysis Findings

In the context of analyzing the Air Force base interview data, connotation frames were used to assess student opinions and experiences at scale. The connotation frame analysis provides a distribution over the perspective/targeted sentiment score for each question. The Y axis represents if the relation between the subject and object were positive or negative. As it is wider at the bottom, this would indicate that the subject and object of the sentence responses were spoken negatively more frequently. If it is wide only on one side, that would indicate only the subject or object was spoken to negatively. On the flip side, if the gourd-like shape is wider at the top, that would indicate that responses were generally more positive, with even shapes indicating mixed responses. Figure 3 shows the connotation frame analysis findings.

Questions that received more negative comments related to the training aids and aircraft matching the workplace and difficulties in the current job. Stakeholders could glean insights that these are the areas where trainees are experiencing the most negative experiences. Diving into qualitative comments uncovered responses such as "Trainers were good at [Air Force Base] but just not [Aircraft] specific". Connotation analysis provide the opportunity to prompts graduates and supervisors specifically on specific aspects of training such as time, content, mediums and promptly see briefly the elements that are responded to positively or negatively. If done at scale, the aspects of training spoken to most negatively can be prioritized, additionally if administered longitudinally, future interviews or surveys can see if the responses have shifted to a more positive direction as a means of return on investment. The same technique can also be applied to other areas of service member surveys such as a Command Climate Survey to identify which areas of climate are currently having negative experiences particularly with sexual harassment or diversity. With a large enough sample, connotation frame analyses could be completed by different groups of career field, identity, and rank.

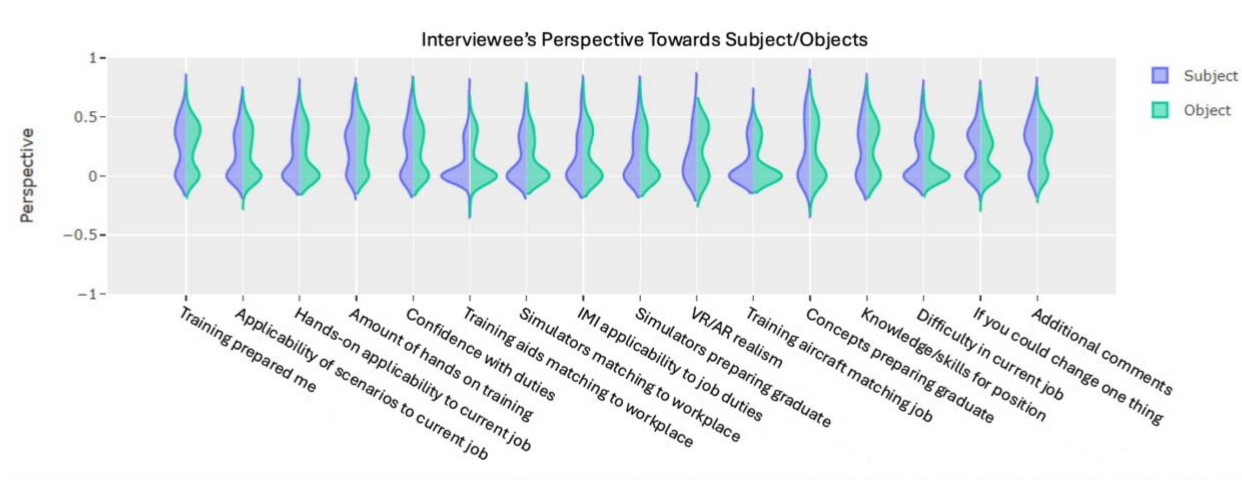

Figure 3. Connotation Frame Analysis by Survey Question.

Automated Insight Extraction

Unlike the topic analysis that intends to extract topics from unstructured data, our novel insight extraction approach combined qualitative responses across all interview questions. The methodology of insight extraction focused on extracting distinct thematic elements from comments based on word associations – word embeddings (Mikolov et al., 2013). In contrast to traditional topic analysis, our method comprehensively considers each comment as containing multiple ideas. It addresses conflicting statements by parsing each comment into discrete subject-verb-object structures. Each parsed unit is then embedded using a LLM that captures sentence semantics and clusters them based on semantic similarity, thereby preserving narrative coherence over a strict subject-verb-object interpretation. The clustering selection is designed to maximize the separation and cohesion of narrative clusters and minimize small, isolated narrative clusters. Figure 4 visually represents the extracted narratives in a semantic embedding space using the TSNE (Van De Maaten and Hinton, 2008), illustrating closely related, divergent, and sub-narrative themes.

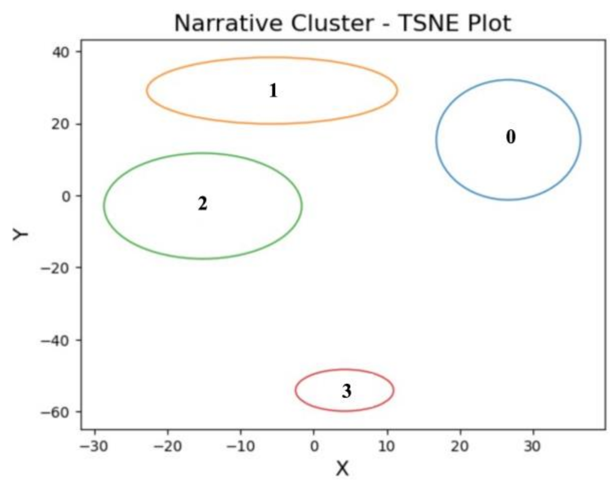

Figure 4. Narrative Clusters for Graduate Responses

Graduate Comment Findings

Graduate comments resulted in four unique clusters or themes. Each cluster represents a commonly shared theme across graduate comments. The X and Y axes represent a lower 2D projection of the high dimensional semantic embedding space by TSNE. The descriptions that emerged from the narrative cluster were then run through an LLM to rephrase the insights into a sentence-like structure. The TSNE cluster plotting can be seen in Figure 4 and the insights are presented in Table 1.

Table 1. Cluster Descriptions

Cluster #	General Theme	LLM Enhanced High-Level Theme Summarizations
Cluster 0	Improving Training Quality	• It is necessary to maintain and improve quality control over existing courses. • There needs to be more hands-on experience within these courses so students will understand what it takes to do their job when they graduate.
Cluster 1	Repetition and Hands-On Experience	• I like more. • More repetition would help me feel comfortable and prepared for the course. • Provide students with more opportunities or time in class that is hands-on/repetitive if they do.
Cluster 2	More In-Depth	• I think the (Interactive Multimedia Instruction) IMI should be more in depth
Cluster 3	Differences between training and operational equipment	• No hands-on training • Not enough time with an assigned airframe. • Most people learned on different models than their assignment, and there's a gap between general aviation fundamentals to modern systems and aircraft.

The four clusters that emerged highlighted the need to improve training quality, increase the amount of hands-on experience, provide more in-depth interactive experiences, and the challenge of differences between the training and operational equipment. The topic analysis can only show a collection of word association in terms of word clouds. Our TSNE cluster plot visualizes the number of *unique* themes from the response. In addition, the clusters also visualize how semantically similar or visually close the topics are to one another. The comment summary from cluster zero could be utilized by a training evaluator to determine that hands-on training large concern and spoken to uniquely from cluster one and two. Clusters one through two are close to one another because each cluster speaks to wanting more of something (the subject verb structure and purpose is similar). However, cluster zero is conveying graduates desire for more quality training related to courses, cluster one is conveying more instances of hands-on training, and cluster two is conveying more in Interactive Multimedia Instruction (IMI). Cluster 3 has a larger distance from the other clusters as the topic shifts from “wanting more” to “experiencing differences”.
In Table 1, ChatGPT summarizations of the statements within a cluster can be gleaned. For example, LLM summarized takeaways posited that “there needs to be more hands-on experience within these courses, so students understand what it takes to do their job when they graduate” for cluster zero. Training stakeholders could utilize this information a few ways. First, training evaluators could utilize the clustering to perform “pulse-checks” of each cluster topic. Distributing 1-2 question surveys to graduates to dig into suggested improvements. Based on the student clusters shown here, one could be targeted on specifically ranking what parts of the courses could have more hands-on training with a second pulse-check targeted on ranking what air frame components should be first updated to provide the highest immediate training value. Instructors could utilize this information to provide hands-on training more frequently within classes or pursue independent learning platforms so students can perform more hands-on training on their own schedule. Training directors can use these insights from the TNSE cluster and subsequent evaluation pulse-checks to estimate and prioritize funding requirements for the following fiscal year.

SUMMARY AND OPERATIONAL IMPACT

The topic analysis, performed using Latent Dirichlet Allocation (LDA), automatically identified seven main themes related to training gaps and effectiveness issues, including difficulties with on-the-job training, the need for more hands-on time, and outdated equipment. The visualization of topic prevalence enables stakeholders to quickly identify potential training gaps and challenges. The connotation analysis, utilizing connotation frames, captured nuanced opinions and power dynamics in survey responses, revealing that questions related to training aids, aircraft matching the workplace, and current job difficulties received more negative comments. This approach allows for the prioritization of negatively perceived aspects of training and targeted allocation of resources. The automated insight extraction method combined responses across all interview questions and employed t-distributed Stochastic Neighbor Embedding (TSNE) for topic modeling. By parsing comments into subject-verb-object structures and using a Large Language Model (LLM) to cluster semantically similar narratives, four main clusters were identified: improving training quality, increasing hands-on experience, providing more in-depth interactive experiences, and addressing differences between training and operational equipment.

The automated AI-driven analytics presented in this paper have the potential to revolutionize training evaluation and improvement processes within the Department of Defense. Traditional, manual analysis of this type of open-ended survey data could take a team of 5 full-time employees (FTEs) around 5 months to complete, whereas the use of NLP and ML techniques allows a single FTE to analyze the data and share the discovered insights within just hours. By leveraging advanced NLP and ML techniques, these methods can significantly reduce the time and effort required for analyzing survey and interview data while providing more consistent, unbiased, and actionable insights. The topic analysis and connotation analysis approaches enable the rapid identification of training gaps, effectiveness issues, and areas of negative sentiment, allowing decision-makers to prioritize resources and interventions accordingly. The automated insight extraction method offers an actionable understanding of the themes and narratives present in the feedback data, facilitating targeted improvements to training content, delivery, and equipment. Implementing these AI-driven analytics at scale across various DoD training programs can lead to more agile and responsive curriculum development, ultimately enhancing the readiness and performance of military personnel. By continuously monitoring and adapting to the needs and experiences of trainees, the DoD can ensure that its training programs remain relevant, effective, and aligned with the evolving operational landscape.

FUTURE AREAS

Future research will focus on enhancing LLM summarization capabilities for TSNE plots to provide more accurate and contextually relevant insights, as well as extending the application of AI-driven analytics to other areas, such as personnel management and command climate analyses, to identify issues related to diversity, inclusion, and workplace satisfaction. Integrating automated analysis with test and evaluation processes will enable the assessment of training rework effectiveness and interventions. The development of modular, user-friendly dashboards will allow stakeholders to easily access and interpret the results of AI-driven analytics, enabling data-driven decision-making at all levels. Exploring the integration of additional data sources, such as performance metrics and operational readiness indicators, will provide a more comprehensive assessment of training effectiveness and its impact on mission success. Investigating the use of real-time data collection methods, such as speech recognition and computer vision, will enable continuous, unobtrusive monitoring of training progress and identification of improvement opportunities. Collaborating with DoD partners to establish best practices and guidelines for the ethical and responsible deployment of AI-driven analytics in training evaluation and improvement processes will be crucial. By continuing to invest in and refine these AI-driven analytics capabilities, the DoD can position itself at the forefront of data-driven training optimization, ensuring that its personnel are equipped with the knowledge, skills, and abilities needed to succeed in an increasingly complex and dynamic operational environment.

REFERENCES

Air Force. (2023). Seeing into the future: AF looks to AI for data analysis. Secretary of the Air Force Public Affairs. https://www.af.mil/News/Article-Display/Article/3411334/seeing-into-the-future-af-looks-to-ai-for-data-analysis/

Rashkin, H., Bell, E., Choi, Y., & Volkova, S. (2017, July). Multilingual connotation frames: a case study on social media for targeted sentiment analysis and forecast. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 459-464).

Borders, T. L., & Volkova, S. (2021). An Introduction to Word Embeddings and Language Models (No. INL/EXT-21-61935-Rev000). Idaho National Lab.(INL), Idaho Falls, ID (United States).

Volkova, S., Glenski, M., Ayton, E., Saldanha, E., Mendoza, J., Arendt, D., ... & Greaves, M. (2021). Machine Intelligence to Detect, Characterize, and Defend against Influence Operations in the Information Environment. Journal of Information Warfare, 20(2), 42-66.

Blei, D. M., & Lafferty, J. D. (2009). Topic models. In Text mining (pp. 101-124). Chapman and Hall/CRC.

Rashkin, H., Singh, S., & Choi, Y. (2015). Connotation frames: A data-driven investigation. arXiv preprint arXiv:1506.02739.

Sameera Horawalavithana, Ellyn Ayton, Shivam Sharma, Scott Howland, Megha Subramanian, Scott Vasquez, Robin Cosbey, Maria Glenski, and Svitlana Volkova. 2022. Foundation Models of Scientific Knowledge for Chemistry: Opportunities, Challenges and Lessons Learned. In Proceedings of BigScience Episode #5 -- Workshop on Challenges

& Perspectives in Creating Large Language Models, pages 160–172, virtual+Dublin. Association for Computational Linguistics.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of machine learning research, 9(11).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b). Efficient estimation of word representations in vector space. In Proceedings of the International Conference on Learning Representations (ICLR) Workshop.

Woods, M., Paulus, T., Atkins, D. P., & Macklin, R. (2016). Advancing qualitative research using qualitative data analysis software (QDAS)? Reviewing potential versus practice in published studies using ATLAS.ti and NVivo, 1994–2013. Social Science Computer Review, 34(5), 597-617.

Kaefer, F., Roper, J., & Sinha, P. (2015). A software-assisted qualitative content analysis of news articles: Example and reflections. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research, 16(2).

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755), 788-791.

Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Roberts, M., Stewart, B., Tingley, D. (2019) stm: An R Package for Structural Topic Models. Journal of Statistical Software, October 2019, Volume 91, Issue 2