2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51802

Integrating Facial Emotion Analysis and Large Language Models to Investigate Student Engagement in Online Engineering Courses

Julianna Gesun, Embry-Riddle Aeronautical University

Julianna Gesun, Ph.D., is currently an assistant professor of engineering whose research broadly focuses on understanding and supporting the process by which engineering programs facilitate the environments for students to develop optimal functioning in undergraduate engineering programs. Her research interests intersect the fields of positive psychology, engineering education, and human development to understand the intrapersonal, cognitive, social, behavioral, contextual, cultural, and outcome factors that influence thriving in engineering.

Lingxiao Wang, Louisiana Tech University

Dr. Lingxiao Wang is an Assistant Professor of Electrical Engineering at Louisiana Tech University. He earned his Ph.D. in Electrical Engineering and Computer Science in 2021 and his M.S. in Electrical and Computer Engineering in 2018, both from Embry-Riddle Aeronautical University. His research interests include robotics, autonomous systems, and artificial intelligence.

Dr. Ghazal Barari, Embry-Riddle Aeronautical University

Dr. GHAZAL BARARI is an Associate Professor of Engineering at Embry-Riddle Aeronautical University Worldwide. With a Ph.D. in Mechanical Engineering from the University of Central Florida, her research pivoted from combustion and biofuel's reaction mechanism to engineering education and student success pathways. She studies AI applications in curriculum design, simulation-based learning, and students' engagement in online environments. She leads the Engineering Fundamentals program's ABET accreditation, serves as an editor for International Journal of Aviation, Aeronautics, and Aerospace, and has been recognized for her innovative teaching and service.

Elaina Hudiburg, Embry-Riddle Aeronautical University - Daytona Beach
Khan Raqib Mahmud, Louisiana Tech University

Integrating Facial Emotion Analysis and Large Language Models to Investigate Student Engagement in Online Engineering Courses

Introduction

Understanding and supporting student engagement in online and asynchronous engineering courses has become increasingly important as these formats expand across undergraduate programs. Engagement is commonly conceptualized as a multidimensional construct encompassing behavioral, emotional, and cognitive components [1], and is associated with persistence, academic performance, and development of professional and technical skills [2], [3]. In engineering education, where students regularly engage in complex, cognitively demanding problem-solving tasks, fluctuations in attention, frustration, boredom, and confusion can meaningfully shape learning trajectories and retention in gateway courses. In face-to-face classrooms, instructors often rely on observable cues such as facial expressions, gaze direction, and posture to infer students' understanding and adjust instruction in real time. However, in asynchronous online environments, these informal diagnostic signals are largely absent. As a result, disengagement or unproductive struggle during self-paced problem solving may remain invisible, limiting instructors' ability to provide timely support. Although learning analytics and self-report surveys are commonly used to assess engagement in online contexts [4], these approaches often capture only partial dimensions of engagement, such as behavioral traces without affective context, or retrospective perceptions that may not reflect moment-to-moment experience. Other research in engineering education has explored physiological indicators of emotions using electrodermal activity (EDA) to capture real-time affective and attentional states [3]. While this method offers rich data, it often requires specialized equipment and controlled settings that may limit scalability in authentic online engineering courses. In contrast, camera-based approaches embedded within existing learning environments may provide a lower-burden and more scalable pathway for examining affective and attentional signals during asynchronous problem solving.

Recent advances in artificial intelligence (AI) offer new possibilities for examining engagement and emotion signals during learning activities. Prior work has explored facial emotion recognition and sentiment analysis in educational settings [6]–[8]. Transformer-based architectures [9] and large language models (LLMs) [10] were developed to process long textual inputs for reasoning. Multi-modal models, e.g., vision-language models, integrate visual and textual inputs to produce the association between two information modalities [11]. Some studies have applied facial behavior analysis to assess engagement in synchronous or in-person classrooms [12], [13]. Yet, limited work has examined how such approaches might support engagement analysis in asynchronous engineering problem-solving contexts. This research explores a novel approach that integrates facial expression analysis using a vision-language model (VLM), student self-reports, and instructor observations to investigate emotions and engagement during short, embedded problem-solving tasks in fully online engineering courses. Rather than positioning AI-derived predictions as definitive measures, we explore their use as complementary indicators that, when triangulated with self-report and instructor perception, may support a more nuanced and scalable understanding of engagement in asynchronous engineering education.

This work is intentionally scoped to fully online undergraduate engineering contexts and addresses a practically significant challenge: developing methods to interpret affective and attentional signals in environments where real-time instructor observation is limited or unavailable. The contribution of this study is therefore methodological and context-specific. We integrate AI-assisted facial analysis with self-report and instructor observation to examine engagement within asynchronous engineering courses. The interdisciplinary research team brings complementary expertise in online engineering instruction, engineering education research, human development, and AI, enabling pedagogically grounded implementation, theoretically informed interpretation and methodological development.

Methods

The study is implemented in fully online, asynchronous undergraduate engineering fundamentals courses delivered through structured, modular weekly learning environments. Planned implementation includes high-enrollment gateway courses such as Introduction to Engineering, Statics, Dynamics, Programming, Graphical Communication (CAD), and Fluid Mechanics, offered in multiple asynchronous sections each term. These courses were selected because they represent a broad range of introductory engineering problem-solving modalities and support short, time flexible activities that can be embedded into weekly modules with minimal disruption. At the university for which this study takes place, about 100 students enroll in these courses per term.

To collect engagement data while preserving normal course flow, a brief, module-aligned micro-task is embedded within the learning management system. Students complete a short problem-solving prompt with their camera enabled (no more than five minutes). Immediately afterward, students review their own recording and select emotions from a predefined list to indicate how they felt across segments of the activity (typically two to three minutes). This design enables synchronized collection of facial expression data and self-reported affect while maintaining student autonomy and minimizing additional cognitive load.

AI-Based Engagement Inference

Student self-reports are triangulated with AI-derived facial indicators and instructor perception data. The AI component employs a VLM, which can analyze visual inputs and generate textual outputs (see an example in Figure 1) [11]. In this study, the VLM analyzes cropped facial images to predict student emotion and an attention score intended to reflect engagement.

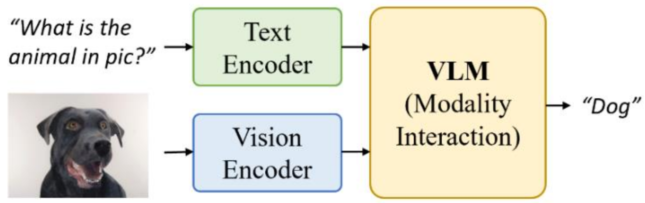

Figure 1. A VLM can analyze input images and texts to generate textual outputs

This preliminary development and evaluation use the publicly available DIPSER dataset, a multimodal classroom dataset designed for modeling student emotion and attention in in-person learning environments [14]. DIPSER includes synchronized student video, face bounding boxes, head pose metadata, and annotations of attention (1–5 scale) and emotion from expert raters and student self-reports. Although collected from in-person settings, DIPSER provides a benchmark for evaluating whether VLMs can infer affective and attentional states from facial cues relevant to learning contexts.

VLM Reasoning Workflow

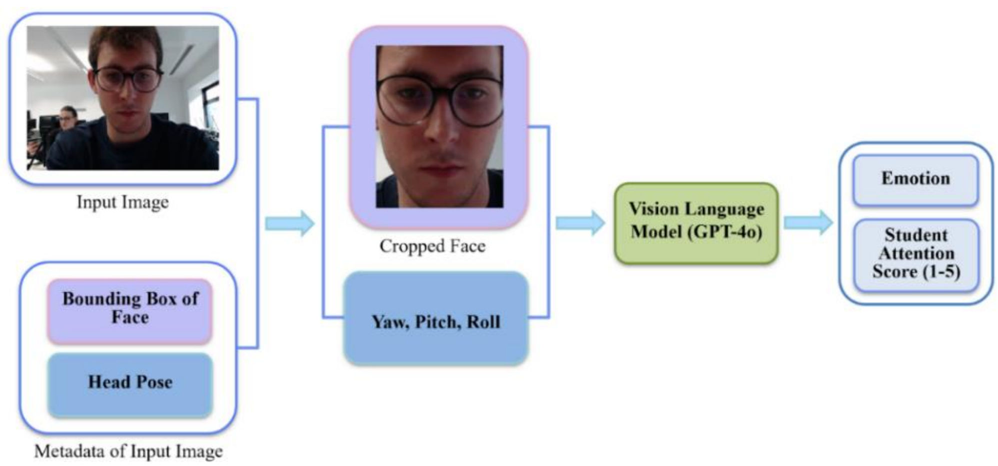

Figure 2. The workflow of the proposed VLM-based student emotion analysis.

Figure 2 illustrates the workflow of the proposed VLM-based method. Inputs of the VLM include a cropped facial image and the head's orientation data. Outputs include the predicted emotion and an attention score value (1-5) to indicate student engagement level. Using DIPSER, individual student camera frames and associated head pose metadata are provided as input to the VLM (using GPT-4o). The model outputs (1) an emotion label and (2) an attention score on a 1–5 scale. To reduce ambiguity, the original nine emotion labels are mapped into three affective categories based on affect polarity: Enjoyment, Relief, and Boredom. Emotion predictions serve as the primary indicator for attention estimation, with head pose used as a secondary cue to allow limited adjustment. Few-shot examples are included in the prompt to encourage consistency and interpretability of outputs.

Preliminary Results

Preliminary experiments were conducted using the DIPSER dataset to evaluate the feasibility of the proposed VLM-based workflow. DIPSER includes student self-labeled emotions and attention scores. We consider these self-labeled emotions and attention scores as the ground truth. We evaluated the LLM model's predictions by comparing predicted emotion labels and attention scores with the dataset's self-labeled values, which were treated as ground truth.

When facial cues were strong and unambiguous (e.g., smiles or drooped eyes), the model's predicted emotion and attention scores aligned with available ground-truth labels. As shown in Figure 3, clear smiles and alert facial features were associated with predictions of enjoyment and higher attention scores, while flat affect and drooped eyelids were associated with boredom and lower attention.

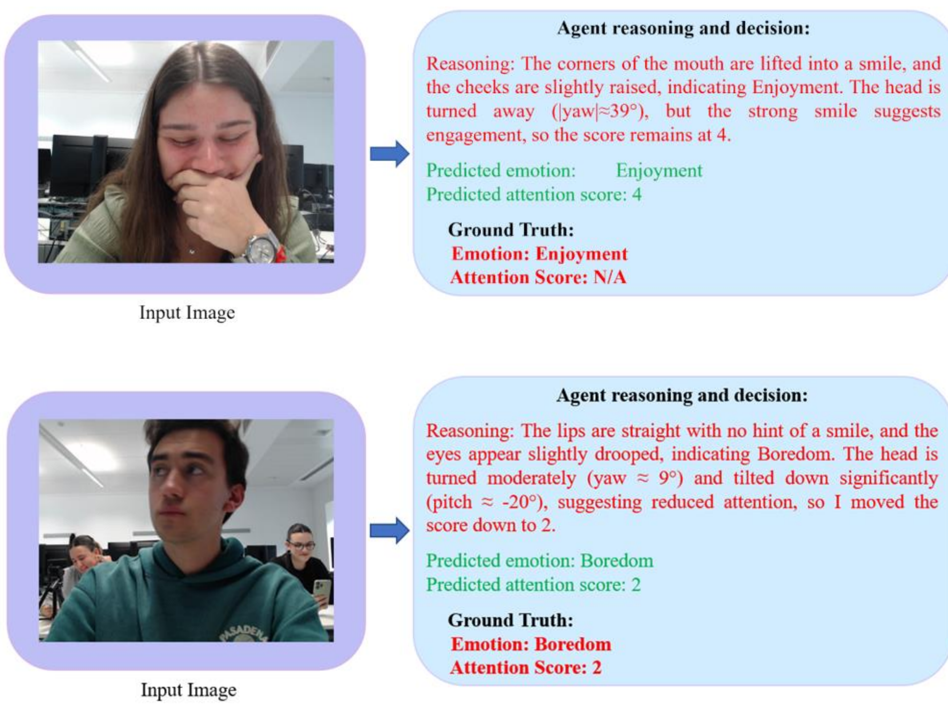

Figure 3. Representative cases where the model's predictions align with the available ground truth from the participant

However, mismatches occurred when facial expressions were subtle or when emotions with similar visual characteristics (e.g., "Shame" versus "Enjoyment") were difficult to distinguish from a single frame (see Figure 4). Even for human observers, it is difficult to accurately distinguish emotions based on a single-frame photo. This highlights the limitations of emotion analysis via purely visual observations and underscores the importance of triangulating AI-derived predictions with other information modalities, such as student self-reports and instructor observations. These preliminary results emphasize feasibility rather than performance optimization and motivate further validation in applied online and asynchronous engineering education learning contexts.

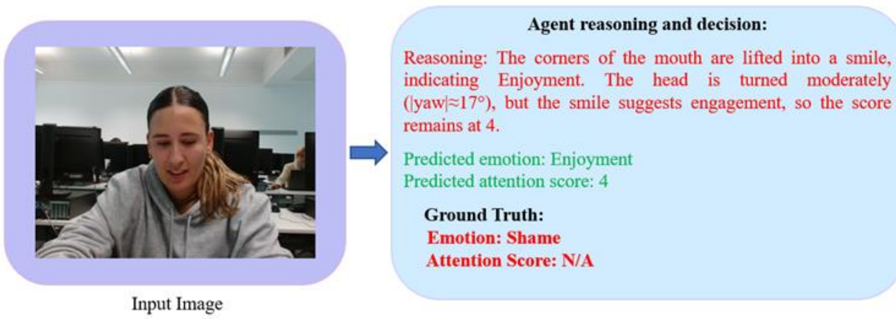

Figure 4. A mismatched example, where the model predicts "Enjoyment", given the smile on the student's face but the ground truth emotion label (i.e., self label) from the participant is "Shame".

In addition, we have not compared the proposed LLM-based method with existing emotion prediction models on the same dataset. This is preliminary work on adapting LLMs to predict student emotions and attention scores using images. To the best of our knowledge, this is the first work that implements LLMs on student emotion and attention score predictions, as many existing works use Convolutional Neural Networks (CNNs) [6]-[8] for this application. Compared to CNNs, LLMs can analyze the input images using common sense (or semantic analysis), whereas CNNs rely on the training on high quality, well labeled dataset.

Conclusions, Implications, and Future Work

This study explores the feasibility of integrating facial expression analysis, LLM-based emotion and attention inference, student self-reports, and instructor observations to examine engagement in asynchronous online engineering courses. Future work will deploy the proposed workflow within fully online and asynchronous engineering courses to examine alignment across data sources, refine attention inference methods, and explore how engagement insights can inform instructional design.

Ultimately, this research aims to contribute scalable, theory-aligned methods for understanding and supporting student engagement in online engineering education. First, the proposed approach demonstrates how engagement data can be collected in asynchronous courses without requiring real-time observation or substantial changes to course structure. Second, triangulating AI-derived indicators with student self-reports and instructor observations may help instructors identify patterns of confusion, frustration, or disengagement that are otherwise invisible in asynchronous environments. Such insights could inform targeted instructional interventions, including revising problem scaffolding, adjusting pacing, or providing additional support resources. Finally, by grounding engagement measurement in established theory, this approach supports responsible use of AI in education. Rather than treating AI predictions as definitive measures, our study positions them as complementary signals that inform reflective teaching practice.

References

[1] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School Engagement: Potential of the Concept, State of the Evidence," *Rev. Educ. Res.*, vol. 74, no. 1, pp. 59–109, 2004.

[2] L. Kallemeyn, G. Baura, F. Fils-Aime, J. Grabarek, and P. Livas, "Engineering Curriculum Rooted in Active Learning: Does It Promote Engagement and Persistence for Women?," in *2021 ASEE Virtual Annual Conference Content Access*, 2021.

[3] I. Villanueva, B. D. Campbell, A. C. Raikes, S. H. Jones, and L. G. Putney, "A multimodal exploration of engineering students' emotions and electrodermal activity in design activities," *Journal of Engineering Education*, vol. 107, no. 3, pp. 414–441, 2018.

[4] N. Bergdahl, M. Bond, J. Sjöberg, M. Dougherty, and E. Oxley, "Unpacking Student Engagement in Higher Education Learning Analytics: A Systematic Review," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, p. 63, 2024.

[5] Q. Liu, G. Evans, and O. Pan, "Engineering Students' Engagement and Learning Outcomes: A Typological Approach," in *2024 ASEE Annual Conference & Exposition*, 2024.

[6] D. Yang, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, "An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment," *Procedia Comput. Sci.*, vol. 125, pp. 2–10, 2018.

[7] J. A. Ballesteros, G. M. Ramírez V, F. Moreira, A. Solano, and C. A. Pelaez, "Facial Emotion Recognition Through Artificial Intelligence," *Front. Comput. Sci.*, vol. 6, p. 1359471, 2024.

[8] H. Zhu, P. Hu, X. Tang, D. Xia, and H. Huang, "NAGNet: A Novel Framework for Real-time Students' Sentiment Analysis in the Wisdom Classroom," *Concurr. Comput.*, vol. 35, no. 22, p. e7727, 2023.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All You Need," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, 2019, pp. 4171–4186.

[11] A. Radford, J.W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever. "Learning Transferable Visual Models

From Natural Language Supervision," in *International conference on machine learning*, 2021, pp. 8748–8763.

[12] P. Buono, B. De Carolis, F. D'Errico, N. Macchiarulo, and G. Palestra, "Assessing Student Engagement From Facial Behavior in On-line Learning," *Multimed. Tools Appl.*, vol. 82, no. 9, pp. 12859–12877, 2023.

[13] Z. Atiq and R. Batra, "How Do First-year Engineering Students' Emotions Change While Working on Programming Problems?," *ACM Transactions on Computing Education*, vol. 24, no. 2, pp. 1–30, 2024.

[14] L. Márquez-Carpintero, S. Suescun-Ferrandiz, C. Lorenzo Álvarez, J. Fernandez-Herrero, D. Viejo, R. Roig-Vila and M. Cazorla. "DIPSER: A Dataset for In-Person Student Emotion and Engagement Recognition in the Wild," *Science Data Bank*, https://arxiv.org/abs/2502.20209. (accessed January 2026).