# Unveiling Sentiment in Online Peer Learning through Machine Learning and LLMs

**Karen DSouza[1], Pratibha Varma-Nelson[1], Shiaofen Fang[1], and Snehasis Mukhopadhyay[1]**
[1]Indiana University Purdue University Indianapolis
kmdsouza@purdue.edu, pvn@iu.edu, shfang@iu.edu, smukhopa@cs.iupui.edu

**Abstract**

To establish an inclusive peer-based online learning environment, we delve into the metrics of peer engagement and sentiment using machine learning (ML) and large language models (LLMs). Our work compares the capabilities of using ML algorithms and LLMs in education to decipher student sentiment and engagement.

## Introduction

The use of machine learning in the online education domain requires adaptability due to the dynamic environments in classes. Tracking online classroom behavior for the purpose of predicting engagement is challenging because of the lack of engagement metrics and standards to quantity them. The concept of student engagement involves student participation and capturing student attention. Attention may not translate into academic success because engagement does not always equate to effective learning (Zyngier 2008). On the other hand, positive student engagement leads to improved academic outcomes and lower dropout rates (Archambault et al. 2009). This research case study is based on cyber peer-led team learning (cPLTL) workshops in Chemistry and focuses on novel non-invasive techniques to extract engagement insights. Student engagement is a crucial part of the peer learning process. Peer education is effective when every student in the peer group participates, bringing to light diverse points of view of problem solving. This facilitates critical thinking, and healthy debates within the group discussion. Predicting student engagement helps identify well performing cPLTL groups and peer leaders needing additional support from the educators. In the cPLTL model, during the online recitation, educators are absent and unaware of the performance of the peer group. The machine learning model serves as a prediction tool to provide feedback to the educators about their peer groups. These insights about student engagement are necessary to assess the quality of the cPLTL progression over the course of a semester.



## Related Work

Although student engagement is multidimensional, it can be studied from three main angles – behavioral, affective, and cognitive (Conner 2016). Existing research on capturing engagement is limited and as such developing student engagement requires a deeper understanding of learning related sentiments (Wang and Degol 2014). Quantifying the multilevel student engagement is an impending challenge. Studies have shown that online student engagement can be effective through multiple modes of interaction during class (Dixson 2010). Traditionally, the engagement methods in online learning platforms using ML are non-invasive such as semantic analysis of the textual data (Toti et al. 2021). Other techniques use wearables to track stress, postures, student's gaze, hand movements, and emotions and predict engagement (Bustos-López 2022). The potential of applying LLMs to sentiment analysis has recently received much attention. LLMs perform better in cases of limited annotations, zero shot learning and binary sentiment classifications, however small language models are better suited for domain specific sentiment tasks (Zhang et al. 2023). Another preliminary study using 16 benchmark datasets showed that ChatGPT is a universal sentiment analyzer due to its zero shot capabilities (Wang et al. 2023).

## Method

We present two approaches to evaluate the sentiment of the students in cPLTL classes (1) Traditional ML (2) LLMs. We hypothesize that the analysis of a combination of both ML and LLMs will improve the effectiveness of engagement and sentiment metrics.

cPLTL transcripts consist of six to eight student speakers collaboratively discussing solutions to problem sets. Each transcript is a 15-minute text corpus of sentences which are domain specific to Chemistry. In both approaches, the transcripts are initially subject to data cleaning to remove stop words, and punctuation, converted to lower case, assigned parts of speech tags and lemmatization.

Our efforts to standardize engagement metrics rely on first using ML to evaluate sentiment polarity and subjectivity from TextBlob – a lexicon based sentiment analyzer, and SentiWordNet. We run the same prompts through Python APIs that are integrated into LLMs such as ChatGPT and Bard to extract the sentiment polarity. Lastly, we perform an evaluation of sentiment obtained from ML and LLM to deduce the engagement trends in the data.

One technique to gain insights is to evaluate prediction of scores of cPLTL classes is to input these features to ML regressors. It is a supervised approach since the outcomes were previously labeled by two human experts using a 1-5 scale. The scores are then averaged to maintain consistency. In the traditional ML method, we extract 13 textual features that include TextBlob polarity, TextBlob subjectivity, counts of positive, negative, and neutral words from TextBlob, Vader and SentiWordNet, SentiWordNet polarity and Vader compound metrics. The polarities reflect the sentiment of the group. We do not single out individual student's metrics because cPLTL is a group learning environment. Educators value group dynamics so that changes to a peer group can be made in successive cPLTL classes.

Using ChatGPT and Bard separately, we extract the polarity and tone of each transcript (N=35). We summarize the transcript using LLMs to determine the overall sentiment and other key descriptive lexical features of the peer group learning. Finally, in the combination method, we analyze both the ML and LLMs results to determine whether the engagement sentiment in the peer groups were consistent.

## Experimental Results

The results of the analysis from method 1 and 2 using LLMs are presented in Table 1. A sample comparison from two LLMs shows more informative educational insights than what was retrieved from traditional machine methods. Row 2 in Table 1 shows the confusion sentiment that was extracted by ChatGPT from zoom class 2. It is also reflected in the negative Bard polarity displayed on the same row.

When we compared the sentiment of zoom class 2 to the machine learning polarity extracted by SentiWordNet, it was also negative. There appears to be a good coherence in the sentiment extracted from method 1 and 2. The polarity from SentiWordNet is neutral for zoom class 2. This is because we set the limit to neutral for polarity greater than negative 0.5. It was in fact a negative polarity.

| **Zoom Class** | **ML Polarity** | **Bard Polarity** | **ChatGPT Sentiment** | **ChatGPT Summary** |
|---|---|---|---|---|
| 1 | Positive | Positive | Positive | Collaboration |
| 2 | Neutral | Negative | Confusion | Seeking clarification |
| 3 | Positive | Positive | Neutral | Collaboration |

Table 1: Sample comparison of ML polarity from SentiWordNet, text polarity from Bard, sentiment and summary extracted from LLMs such as ChatGPT using transcripts of three cPLTL zoom classes indicating educational insights such as collaboration and confusion.

## Discussion

The trends in polarity were backed by the summarization features of LLMs. LLMs such as Bard and ChatGPT summarize large volumes of text in a few seconds thus saving time and effort of running the documents through large ML programs. Educators can view the sentiment for every 15 minute block to ascertain whether the scores are improving. From table 1, it is observed that the polarity of the sentiment changed from negative in zoom class 2 to positive in the next zoom class 3. Confusion and uncertainty are viewed as positive situations of learning (Robison 2009). Therefore, this data point is metric of positive collaborative learning.

LLMs excel in text processing tasks. We were able to process lengthy transcripts easily to generate sentiment polarity and summarizations using APIs that integrated with Python. The limitations and biases of using ChatGPT and Bard were that both gave biased answers for long structured sentences. It appears they performed better on short sentences in the transcript. Sentiment in the educational domain from transcripts is a highly subjective exercise. Therefore, it is important to evaluate the results from both ML algorithms and LLMs.

## Conclusion

In the evaluation of domain specific educational transcripts, the efficacy of sentiment and engagement metrics depends on the tool being used. However, with the widespread use of generative AI tools, a combination of traditional machine learning methods together with generative AI is impactful. As an alternative solution, we used ChatGPT and Bard tools to extract the lexical polarity. A comparative approach showed that both traditional machine learning and generative AI provide consistent deeper insights into online learning progress. Generative AI presents a future potential to improve the performance of online learning platforms especially in case studies such as cPLTL where AI tools are not readily available.

## Ethical Statement

IRB approval is in place for this study. All participants are informed of the recordings before registering for the class as well as before the weekly recordings are turned on. The study is voluntary. All personal identifiable information is removed from the transcripts. The recordings are stored in a secure drive with access granted to the researchers only. The algorithms used for sentiment analysis are included in the standard Python package for natural language processing called natural language tool kit (NLTK). The study is in the preliminary stage of research. This means that no decisions have been taken based on the sentiment of the students. There are no ethical concerns that may arise out of bias and unfair assessments.

## References


Archambault, I.; Janosz, M.; Fallu, J. S.; and Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of adolescence*, *32*(3), 651-670.

Bustos-López, M.; Cruz-Ramírez, N. ; Guerra-Hernández, A. ; Sánchez-Morales, L. N. ; Cruz-Ramos, N. A.; and Alor-Hernández, G. (2022). Wearables for Engagement Detection in Learning Environments: A Review. *Biosensors*, *12*(7), 509.

Conner, T. (2016). Relationships: The key to student engagement. *International journal of Education and Learning*, *5*(1), 13-22.

Dixson, M. D. (2010). Creating effective student engagement in online courses: What do students find engaging? *Journal of the Scholarship of Teaching and Learning*, 1-13.

Robison, J., McQuiggan, S., & Lester, J. (2009, September). Evaluating the consequences of affective feedback in intelligent tutoring systems. In *2009 3rd international conference on affective computing and intelligent interaction and workshops* (pp. 1-6). IEEE.

Toti, D.; Capuano, N.; Campos, F.; Dantas, M.; Neves, F.; and Caballé, S. (2021). Detection of student engagement in e-learning systems based on semantic analysis and machine learning. In *Advances on P2P, Parallel, Grid, Cloud and Internet Computing: Proceedings of the 15th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC-2020) 15* (pp. 211-223). Springer International Publishing.

Wang, M. T.; and Degol, J. (2014). Staying engaged: Knowledge and research needs in student engagement. *Child development perspectives*, *8*(3), 137-143.

Wang, Z.; Xie, Q.; Ding, Z.; Feng, Y.; and Xia, R. (2023). Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv preprint arXiv:2304.04339*.

Zhang, W.; Deng, Y.; Liu, B.; Pan, S. J.; and Bing, L. (2023). Sentiment Analysis in the Era of Large Language Models: A Reality Check. *arXiv preprint arXiv:2305.15005*.

Zyngier, D. (2008). (Re) conceptualizing student engagement: Doing education not doing time. *Teaching and teacher education*, *24*(7), 1765-1776.