Making sense of student feedback and engagement using artificial intelligence

Christopher Dann, Petrea Redmond, Melissa Fanshawe, Alice Brown, Seyum Getenet
School of Education, University of Southern Queensland, Toowoomba, QLD 4350, Australia

Thanveer Shaik, Xiaohui Tao, Linda Galligan, Yan Li
School of Mathematics, Physics and Computing, University of Southern Queensland, Toowoomba, QLD 4350, Australia

Making sense of student feedback and engagement is important for informing pedagogical decision-making and broader strategies related to student retention and success in higher education courses. Although learning analytics and other strategies are employed within courses to understand student engagement, the interpretation of data for larger data sets is more challenging and rarely pursued. This is concerning as data offers the potential for critical insights into engagement behaviour and the value students place on engagement. Artificial intelligence (AI) offers a revolutionary ability to make sense of data, with capacity for prediction and classification, by consuming vast amounts of structured and unstructured data sets. This paper reports on how AI methodologies (specifically, deep learning and natural language processing) were used to leverage labelled student feedback in terms of online engagement in five courses in a regional Australian university. This paper reinforces the value of AI as a viable and scalable multilayered analysis tool for analysing and interpreting student feedback, particularly for categorising student responses as to the types of engagement that they most valued to support their learning. The paper concludes with a discussion of suggested further refinement, including how the AI-derived data may add insights for informing pedagogical practice.

Implications for practice or policy:
• AI offers an ability to make sense of large data sets in higher education courses.
• Teachers can use student feedback data categorised into types of engagement by AI to support reflection on what students value in their courses.
• Educators and key stakeholders can use the insights AI analysed data offers for informing pedagogical practice and decision-making in higher education to enhance student experiences.

Keywords: student engagement, student experience, online engagement, artificial intelligence, natural language processing, higher education, regional university

Introduction

Over the last decade, the phenomenon and dimensions of student engagement in higher education have been widely researched and received significant attention, often motivated by being a key measure of the student experience both at the university level and through national surveys. Engagement has been linked to motivation, persistence, retention and success. Shaped by elements of engagement (Brown et al., 2023; Redmond et al., 2018), including cognitive, behavioural, collaborative, emotional and social, as well as theories of engagement (see Kahu, 2013; Kahu & Nelson, 2018), student engagement can be defined as “the energy and effort that students employ within their learning community, observable via any number of behavioural, cognitive or affective indicators across a continuum” (Bond et al., 2020, p. 3).

Delivery of coursework through online learning has experienced unprecedented growth in higher education, opening doors and increasing the accessibility of education to a diverse student cohort (Dyment & Downing, 2020; Muir et al., 2019). The number of online enrolments has globally surpassed those studying within the bricks and mortar of the university prior to COVID-19 (Calhoun et al., 2017) and has experienced even higher enrolment percentages since the pandemic (Thompson & Lodge, 2020).
Extensive research indicates that student engagement online positively influences learning, satisfaction and retention (Boulton et al., 2019; Kahu et al., 2020; Stone et al., 2019; Tight, 2020). Yet, at this point, the interpretation of data related to student engagement, particularly in larger courses, or big data sets, is not evident or explored, due to the challenge of using learning analytics and other strategies to better understand this phenomenon or aspect of student learning (Stojanov & Daniel, 2023). This is of concern as these data sets offer critical insights into both engagement behaviour and the value students place on engagement (e.g., social, emotional, cognitive, behavioural and collaborative) (Redmond et al., 2018), with the added benefit of informing pedagogical decision-making and broader strategies related to student retention and student success.

Although a relatively new discipline for teachers and researchers, learning analytics is increasingly being used to better understand students’ learning and the activities they engage in (Alzahrani et al., 2023). For example, some teachers have used the inbuilt functions of the learning management system to trace users’ digital footprint within the learning management system (Olipas, 2023). This process has enabled educators within the course to gain data about the ways students engage online, the types of course resources they access and activities they complete, as well as the ways they interact with others (such as within forums). Information gathered via learning analytics can be used by educators to predict student performance, visualise the data and provide information that can be used for proactive student support and interventions (Gašević et al., 2015).

Although learning analytics data and other strategies are employed within courses to understand student engagement, insights into student engagement for larger courses, or big data sets, are more challenging. Yet, potentially this data offers critical insights into both engagement behaviour, as well as the value students place on different forms of engagement (Gašević et al., 2015). These data can help inform pedagogical decision-making, as well as broader strategies related to student retention and student success.

Student evaluation of teaching feedback is often collected in higher education courses at end of the semester and comprises of both quantitative and qualitative data (Goldsmith et al., 2022). The qualitative feedback in text format presents student opinions and different perspectives of their learning experience. Manual analysis of qualitative feedback demands a tremendous amount of time and resources. To overcome this challenge, manual labelling of one semester’s qualitative feedback can help to train an artificial intelligence (AI) model and use the model to classify students’ feedback in future semesters.

AI offers a revolutionary ability to analyse data, with its capacity for prediction and classification, by mining vast amounts of structured and unstructured data sets, such as text, videos and audio recordings. Machine learning (ML), a subcategory of AI, focuses on developing algorithms and techniques that enable machines to have cognitive and predictive capabilities through learning and analysing large amounts of data (C. C. Aggarwal, 2018). ML is commonly seen in signal processing, facial recognition, product recommendations and predictive text. Natural language processing (NLP) is a type of ML that has been used across many disciplines to process human language to analyse, summarise and extract the opinions of vast clusters of people (Kastrati et al., 2021, Nazari et al., 2021). Computer-aided tools implementing NLP techniques use a programming language to encode natural language and speech through video and audio recordings or data (Tyagi & Bhusan, 2023). The coding enables NLP to understand student feedback, process it into categories and output predictive insights (Maimone et al., 2023).

This paper reports on using ML methodologies within AI to examine perceptions of student engagement in pre-existing student opinion surveys of engagement in a regional university in Australia. These surveys are an integral aspect of the evaluative process adopted by the university. For this study, data were collected at the end of each semester for every course within the university. Although course educators regularly used student opinion surveys to examine and improve their offerings, due to a large number of surveys and participants, data analytics was yet to be performed at the university level. For this study, it was determined that AI may be helpful to analyse large amounts of student data, which could be considered essential to understand student experience at course, programme, school and faculty and...
university levels concerning satisfaction and retention, along with insight into positive and negative experiences, which could inform future offerings.

Therefore, the research focused on two questions:

- Is AI (specifically, NLP) a viable tool to gather and analyse information about university-wide course experiences, particularly focusing on engagement preferences?
- How can this analytical data inform future offerings, pedagogical practice and decision-making particularly related to students' valuing of online engagement?

**Backgrounding key concepts**

**Why student engagement?**

Student engagement features as one of the key priority areas in higher education, an essential measure of quality (Quality Indicators for Learning and Teaching, n.d.), as well as used to predict student learning experiences and outcomes (Gay & Betts, 2020; Hussain et al., 2018). Student engagement has been identified as a critical indicator of student success, including student retention, persistence, course achievement, motivation and improved graduation rates (Ferrer et al., 2022; Kahu & Nelson, 2018). Alternatively, low engagement or disengagement has been found to negatively affect the quality of the student experience and learning outcomes (Higher Education Standards Panel, 2017). Given this, it is not surprising that student engagement has gained not only prominence as a measure of the quality of the student experience, but also a growing expectation for academics and higher education institutions to employ learning analytics and other research techniques to investigate the phenomenon of student engagement (G. Ramaswami et al., 2023).

The term *engagement* is complex and multifaceted, with definitions constantly evolving. Student engagement can be understood as a student’s psychological investment in, and commitment to, learning (Kim et al., 2019; Lee et al., 2019; Northey et al., 2018). Over the last 2 decades, there has been an increased interest in defining the nuanced term online student engagement (Lim et al., 2021; Redmond et al., 2018; Yates et al., 2021). In their guide *Enhancing Online Engagement in Higher Education*, Brown et al. (2023) proposed a working definition for online engagement as being “the regular and ongoing synchronous and asynchronous formal and informal activities, actions, energy, and behaviours that involve the learner within their learning environment and broader learning community, where the end goal is to enhance and achieve learning” (p. xiii).

**Types of student engagement**

Literature and related interpretation of student engagement also reference elements or types of engagement, with these concepts, understood to be interconnected, dynamic and multidimensional (Malmberg et al., 2023; Pittaway, 2012; Weimer, 2016). Building on existing works, we introduced five elements of online engagement for higher education, considered crucial for effective learning and teaching (social, cognitive, behavioural, collaborative and emotional elements) (Redmond et al., 2018). Social engagement is understood by Redmond et al. to consist of students’ participation and investment in the learning environment, including those pursuits that extend beyond the educational or virtual classroom. Cognitive engagement includes students’ surface and deep thinking and relates to students’ focus on the complex ideas and skills of learning. Collaborative engagement is related to purposeful learning with others, including study groups, assessment group tasks and learning forums. Behavioural engagement is referred to students’ learning presence or students upholding online learning norms. Finally, emotional engagement relates to the affective component of learning, including a student’s attitudes and feelings towards learning.
Understandings of AI

The term AI is constantly evolving, currently defined by Holmes and Tuomi (2022) as a distinct area of study and advancement rather than merely an artificial form of intelligence, highlighting that useful and usable AI definitions rely on their specific applications and purposes. Davies et al. (2021) have suggested that AI is currently being perceived as a potential remedy for perceived challenges in the broader field of education. Remedy or not, the influence of AI in higher education is evident although there is a range of perspectives amongst teachers and students regarding its use (West et al., 2023). Eggert (2022) explored the opportunities of using AI in education to empower learning and teaching and provide future skills and lifelong learning and found AI can record student’s prior knowledge, emotional state or economic background and assist teachers to adjust their teaching according to student needs. In addition, AI has been used in education to track students’ learning habits and progression, grade assessments and determine the value of courses, and evaluate student opinions through course feedback (Chen et al., 2020). Although the capabilities of AI appear to offer benefits for teachers in higher education, there is a call to increase the understanding of the uses of AI technologies in education to benefit student engagement and success (Hrastinski et al., 2019). This is useful, as manual monitoring and tracking students’ feedback is traditionally time-consuming and demanding of financial resources (Stone et al., 2016).

One of the most widely used AI methodologies for analysing end-user data is NLP (Estrada et al., 2020). NLP has been widely used in assessment (Chen et al. 2020). This technology has the capability to detect typographical and grammatical errors, and as such, AI-enabled online tools can be used by students when completing assessment to ensure the accuracy of their submitted copy. However, at the same time, the increasing use of AI tools by students to complete assessment tasks has led to the need for AI governance within higher education institutions to monitor quality and academic integrity (Selvaratnam & Venaruzzo, 2023).

NLP can also be utilised to interpret large data sets of qualitative data, including feedback or opinions of end users (Estrada et al., 2020). NLP has the potential to read the feedback in many languages and understand the semantic meaning of data with training. In recent years, NLP has been applied to review items such as movies and books (A. Aggarwal et al., 2019). Topic modelling within NLP enables text documents to be read, summarised, annotated and categorised. Furthermore, it uses techniques such as contextual semantic tagging with parts of speech to understand the context of words.

Deep learning in ML is part of AI methodologies. Deep learning has a multi-layer neural network with processing layers to train new concepts and link to previously known concepts. Deep learning enhances NLP with concepts such as continuous bag of words, skip-gram models, convolutional neural networks (Li et al., 2018), recurrent neural networks, long short-term memory and gated recurrent units, which are different forms of deep learning techniques used in text classification (Prokhorov & Safronov, 2019; S. Ramaswamy & DeClerck, 2018). AI technologies, such as deep learning methods, have advanced to the point where certain algorithms, such as convolutional neural networks and recurrent neural networks, have become well-known for their capacity to analyse a wide range of data types, including audio, video and images (Dwivedi et al., 2023).

Before AI can understand qualitative information with its annotation and summarisation capabilities, there are important considerations for the data being input, including typographical errors, language, domain-specific words, sarcasm and ambiguity, along with student use of emoticons and special characters.

Methodology

This research was undertaken by a multidisciplinary team including education academics and AI experts in a regional university in Australia. To determine if AI (specifically, NLP) was a viable tool to gather and analyse course feedback, quantitative and qualitative content analysis was used to (a) examine the
possibility of implementing AI methodologies on student feedback; (b) extract student engagement and learning experiences in one course over multiple offerings; and (c) explore the transferability of this approach at a broader university-wide level in relation to better understanding the student perspective of their course experience. Content analysis is commonly used in education research because it can apply to both quantitative and qualitative studies (Kleinheksel et al., 2020).

Content analysis can be described as a method to statistically analyse textual data (Mayring, 2000). She goes on to comment that both inductive and deductive coding are possible and are useful for empirical studies because they follow rules of analysis which are controlled step by step rather than working at a more holistic level. The focal point of qualitative content analysis is systematically categorising textual data in order to make sense of it (Miles & Huberman, 1994). The data analysis quantitively presents the qualitative data, for example, the number of times students made negative comments is graphed within sentiment distribution (see Figure 3).

This study commenced after receiving approval from the university ethics committee (ETH2023-0793). The approved ethics protocol permitted the online and anonymous collection and use of historical data within a specified range of dates.

The data included in this study was obtained from student feedback in evaluation of teaching surveys undertaken by this regional university after each course offering. Upon completion of the course, students were offered the opportunity to provide voluntary, anonymous course feedback on a Likert-style scaled score and qualitative feedback on aspects of the course. The student evaluation questions required students’ opinion on three questions: “What were the best aspects of this course?”, “What aspects of this course are most in need of improvement?” and “Is there anything else you want to tell us about this course?” Data from the student responses help to determine different elements of students’ engagement in their learning experience. Also, the student comments reflect their positive or negative experience and opinion on teaching and learning services being provided within the course.

In this study, the role of AI methodologies, including NLP and deep learning methods within the educational context, formed the basis of the inquiry into the use of and management of student feedback. To train the machine, student feedback was retrieved from five Semester 1 courses which were mandatory in the programme, which included 383 students’ feedback sentiments. To test the machine, data was retrieved from Semester 2 offerings of the same courses, with 311 students’ feedback sentiments.

Conceptual framework

The online engagement framework (OEF) for higher education (Redmond et al., 2018) informed and guided the conceptual framework for this study. The OEF proposed five elements of engagement that impacted students’ experience in online learning: behavioural engagement, cognitive engagement, social engagement, emotional engagement and social engagement, as illustrated in Table 1.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>OEF for higher education (Redmond et al., 2018, p. 190)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online engagement element</td>
<td>Illustrative indicators</td>
</tr>
<tr>
<td>Social engagement</td>
<td>Building community</td>
</tr>
<tr>
<td></td>
<td>Creating a sense of belonging</td>
</tr>
<tr>
<td></td>
<td>Developing relationships</td>
</tr>
<tr>
<td></td>
<td>Establishing trust</td>
</tr>
<tr>
<td></td>
<td>Thinking critically</td>
</tr>
<tr>
<td></td>
<td>Activating metacognition</td>
</tr>
<tr>
<td></td>
<td>Integrating ideas</td>
</tr>
<tr>
<td></td>
<td>Justifying decisions</td>
</tr>
<tr>
<td></td>
<td>Developing deep discipline understanding</td>
</tr>
<tr>
<td>Cognitive engagement</td>
<td></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Online engagement element</th>
<th>Illustrative indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distributing expertise</td>
</tr>
<tr>
<td>Behavioural engagement</td>
<td>Developing academic skills</td>
</tr>
<tr>
<td></td>
<td>Identify opportunities and challenges</td>
</tr>
<tr>
<td></td>
<td>Developing multidisciplinary skills</td>
</tr>
<tr>
<td></td>
<td>Developing agency</td>
</tr>
<tr>
<td></td>
<td>Upholding online learning norms</td>
</tr>
<tr>
<td></td>
<td>Supporting and encouraging peers</td>
</tr>
<tr>
<td>Collaborative engagement</td>
<td>Learning with peers</td>
</tr>
<tr>
<td></td>
<td>Relating to faculty members</td>
</tr>
<tr>
<td></td>
<td>Connecting to institutional opportunities</td>
</tr>
<tr>
<td></td>
<td>Developing professional networks</td>
</tr>
<tr>
<td>Emotional engagement</td>
<td>Managing expectations</td>
</tr>
<tr>
<td></td>
<td>Articulating assumptions</td>
</tr>
<tr>
<td></td>
<td>Recognising motivations</td>
</tr>
<tr>
<td></td>
<td>Committing to learning</td>
</tr>
</tbody>
</table>

**Procedure**

The method followed a four-step process, as described below and illustrated in Figure 1.

*Step 1: Pre-processing the data collected*

After obtaining ethical approval to use the feedback data of the student evaluation of teaching surveys, the course examiners worked with two other academics to analyse and hand-coded the data from five courses (383 students' feedback sentiments), using illustrative indicators to guide the coding into the OEF five engagement elements, as illustrated in Table 1. It is estimated this manual coding of data process took approximately 15 hours.

*Step 2: Using AI to analyse text (NLP techniques)*

The AI team used the manual coding from the academics to train the algorithm to code student comments into the five areas of the OEF. Before this, NLP techniques were used to clean the data, including text cleaning, tokenisation, stopwords removal and stemming adopted for text pre-processing of student feedback.

*Step 3: Deep learning modelling*

After the text pre-processing, the AI team used the manual coding of the student comments from the five courses as labelled data to train the NLP model to extract the students’ sentiment so that the OEF was “known” to the computer. Sentiment extraction is a process that uses individual lexicons to identify the typical sentiment of the writer. The degree of personal opinion and factual information in a text is measured by subjectivity. In this study, the TextBlob technique, a Python programming language module (https://textblob.readthedocs.io/en/dev/), was used to estimate the student feedback's subjectivity and extract the student feedback's sentiment. This sentiment was calculated after the comment has been positioned within the OEF detailed in Table 1.

*Step 4: Visualisation of the data*

Once all of the data were processed and the key information was extracted, the AI model was programmed to display information for the OEF’s five elements and the responses' polarity.
Results

Accuracy of AI in interpreting data

By setting threshold values to the numerical data from the deep learning model results, the student comments could be classified into five engagement areas. The deep learning model classification performance was evaluated using a confusion matrix (Ting, 2017), an ML evaluation metric to check classification performance that provides the classification results. To test the machine learning, we used 311 unlabelled students’ feedback sentiments from Semester 2 courses. The evaluation technique provided several correct and incorrect classifications with count values for each engagement area. The performance metrics precision, F1-score, recall and balanced accuracy are computed based on these values. Although the overall accuracy of the deep learning model was 76%, multi-label (multiple engagement areas in a student comment) classifications were evaluated by verifying the model performance in classifying individual engagement areas. Hence, the model performance in each engagement area is presented in Table 2.

<table>
<thead>
<tr>
<th>Engagement area</th>
<th>Precision</th>
<th>F1-score</th>
<th>Recall</th>
<th>Balanced accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioural</td>
<td>0.65</td>
<td>0.51</td>
<td>0.42</td>
<td>61.1</td>
</tr>
<tr>
<td>Cognitive</td>
<td>0.56</td>
<td>0.54</td>
<td>0.52</td>
<td>63.9</td>
</tr>
<tr>
<td>Emotional</td>
<td>0.86</td>
<td>0.72</td>
<td>0.62</td>
<td>93.1</td>
</tr>
<tr>
<td>Social</td>
<td>0.56</td>
<td>0.58</td>
<td>0.60</td>
<td>72.2</td>
</tr>
<tr>
<td>Collaborative</td>
<td>0.87</td>
<td>0.70</td>
<td>0.58</td>
<td>56.1</td>
</tr>
</tbody>
</table>

Precision metrics estimate the fraction of correct classifications of an engagement among all correct classifications made by the deep learning model. For example, behavioural engagement received 0.65 precision, meaning 65% of the classifications made by the model were correct. Recall is the model’s sensitivity in which a proportion of each engagement classification the model makes belongs to that engagement. To explain in simple terms, recall for behavioural engagement is 0.42, meaning that 42% of the classifications belong to behavioural engagement, with the remaining 58% being missed in classification. The F1-score is a simple harmonic mean of precision and recall which supports comparing two classifiers and provides an overall metric of the model performance. Balanced accuracy mean for
each engagement area explains how well the classifier model can classify. The results can be further improved by adding more manually labelled data and making the labels balanced.

**Visualisation**

In this section, the visualisation results of the deep learning model are presented. A prototype was built to classify student feedback into five engagement areas and visualise the results at different hierarchical levels of a university, such as faculty, school, programme and course levels for each semester. As discussed in the Methodology section of this paper, the deep learning model was trained with manually labelled data from the five courses. The distribution is shown in Table 3.

<table>
<thead>
<tr>
<th>Engagement area</th>
<th>Number of records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioural</td>
<td>98</td>
</tr>
<tr>
<td>Cognitive</td>
<td>114</td>
</tr>
<tr>
<td>Collaborative</td>
<td>25</td>
</tr>
<tr>
<td>Emotional</td>
<td>118</td>
</tr>
<tr>
<td>Social</td>
<td>28</td>
</tr>
</tbody>
</table>

Data was entered into the model for testing. The findings are presented for each hierarchical level below.

**Faculty level**

The deep learning model classification of machine-generated data is presented in Figures 2 and 3. In Figure 2, all three pie charts are independent and present the engagement and sentiment classification for one of the faculty levels. In the engagement chart, 54.7% of the student feedback at the faculty level is classified as emotional engagement, and the other areas are labelled with different colours. The positive engagement chart presents positive sentiment of student feedback at the faculty level, in which 56.7% is positive for an emotional engagement at the faculty level. Similarly, 57.1% of student feedback is negative in emotional engagement in the negative engagement at the faculty level.

![Figure 2. Faculty level – engagement and classification](image)

The actual count of the positive and negative engagement at the faculty level can be seen in Figure 3. The bar presents a positive and negative sentiment distribution for the model analysed data from the collective courses. It is clear that there is more positive feedback in all engagement areas, but the negative sentiment is more focused on interpreting the concerns causing the negative feedback. The bar chart also explains the deep learning model efficiency in the classification task.
Students were more likely to report on positive and negative elements of emotional engagement within the course, more so than social, cognitive, behavioural or collaborative engagement. These results may indicate students place importance on elements of emotional engagement, including motivation to engage within higher education courses.

![Faculty-Level Sentiment Analysis](image1)

**Figure 3. Faculty level – sentiment distribution**

**School level**

Similar to the faculty-level analysis, the pie charts present the proportion of each engagement area at the school level. Figures 4 and 5 present the results of one school in the study, where emotional engagement dominates other areas. According to the deep learning model results, social and collaborative engagements are minimal at the school level. This may indicate the need to examine social and collaborative engagement opportunities, such as building a community and sense of belonging within the course offerings, along with opportunities to learn with and from peers.

![School Level Engagement and Classification](image2)

**Figure 4. School level – engagement and classification**
Figure 5. School level – sentiment distribution

Programme level
One of the programme’s engagement and sentiment classifications is presented in Figures 6 and 7. Negative engagement is found only in behavioural and emotional engagement. These two areas of engagement relate to the way the course is designed to support movement through the course and support the development of agency and commitment to learning. Using students’ feedback to develop these skills within the course may also assist student engagement throughout their programme (Brown et al., 2024). It is interesting to note, the pie charts show no social engagement in student feedback at the programme level. This data indicates an opportunity to redesign courses to include all elements of online engagement, including social engagement, which is essential for students in developing community and creating a sense of belonging and togetherness within their programme (Brown et al., 2021; Redmond et al., 2018).

Figure 6. Programme level – engagement and classification

The bar charts in Figure 7 present the distribution of positive and negative sentiments in the student feedback on the programme.
Course level
The deep learning model results are filtered at the course level for one of the courses, as shown in Figures 8 and 9. There is no negative sentiment for the course in any of the engagement areas. Social and collaborative engagement are absent at the course level.

The bar chart in Figure 9 shows that positive emotional engagement is high compared to other engagement areas. As a result of using AI to make sense of student feedback in this course, additional social and collaborative elements to develop relationships between students could be added to future offerings as a potential opportunity to increase engagement.
The AI’s analysis resulted in the classification of student comments into five engagement areas. Although the overall accuracy of the deep learning model was 76%, it is important to consider multi-label classifications. Each engagement area’s performance metrics, including precision, F1-Score, recall and balanced accuracy, provide insights into the model’s strengths and weaknesses. For example, a precision of 0.65 for behavioural engagement indicates that 65% of the model’s classifications in this area were correct. However, a recall of 0.42 suggests that 42% of behavioural engagement was correctly classified, leaving 58% missed.

The challenge lies in translating these pie charts into actionable strategies for educators. If negative engagement is primarily found in behavioural and emotional engagement, educators should focus on addressing issues related to these areas. For example, they might consider implementing targeted interventions to improve student self-regulation and emotional well-being. It is essential to closely examine the specific comments within these categories to understand the underlying issues and develop effective strategies.

The visualisation results provide a hierarchical view of engagement areas at faculty, school, programme and course levels. At the faculty level, emotional engagement appears dominant, but negative sentiment is also notable. This suggests that although students are emotionally engaged, there are concerns that need attention. Positive feedback is more widespread but should not overshadow areas that require improvement.

Similar patterns emerge at the school and programme levels, with emotional engagement often prevailing. However, at the programme level, negative engagement is observed in behavioural and emotional aspects. Educators should investigate these issues to enhance the overall student experience.

At the course level, positive emotional engagement stands out, but social and collaborative engagement are notably absent. The absence of negative sentiment is encouraging, but educators can explore ways to promote social and collaborative interactions in courses.
These pie charts serve as valuable tools for educators to identify areas of concern and improvement in student engagement. Educators should interpret these charts in conjunction with specific comments to develop targeted strategies for enhancing the student experience and addressing any negative engagement issues effectively. Additionally, continued data collection and analysis can further refine these strategies over time.

**Discussion**

In this study, AI methodologies such as NLP and deep learning methods, were adopted in a higher education context to understand student engagement, based on their feedback to the student evaluation of teaching surveys completed at the end of the course each semester. Sentiment analysis, which is one of the fields of NLP that evaluates student opinion (A. Aggarwal et al., 2019), was performed on the student responses to extract positive and negative opinions towards the university educational infrastructure and their learning experience. Classification of student engagement was based on the OEF (Redmond et al., 2018), which consists of behavioural, cognitive, collaborative, emotional and social engagement elements.

A deep learning model was trained to identify these elements using a set of manually labelled student responses from 383 responses to the student evaluation of teaching surveys. The deep learning model was then able to classify the engagement elements accurately in unlabelled student responses in a subsequent semester’s set of data from the same course. Based on this deep learning model and the utilised data, a prototype was further developed to visualise student engagement across the university at faculty level, school level, programme level and course level. The prototype was able to retrieve the most frequent words in positive and negative feedback for the various elements of engagement, at each of these levels. As a result, data would offer granular data to support deep reflection (Redmond et al., 2021) by key stakeholders, including academics, departments, faculty and university administrators.

This project drew from existing studies, research and knowledge of online engagement by the project team, as well as the specialist knowledge of and skills in AI technologies by a number of the project team. It proved achievable for the AI team to develop and train an AI prototype tool to accurately code unlabelled student feedback in terms of online engagement and also produce a visualisation that can be used in the reflective practice of academics and decision-making of administrators in a university. What follows is a discussion of key insights gained from this project.

The first research question addressed was the following: “Is AI (specifically, NLP) a viable tool to gather and analyse information about university-wide course experiences, particularly focusing on engagement preferences?” Although AI has been suggested as a tool to assist in analysing student feedback (Chen et al., 2020), the AI methodology proposed in this study was applied to understand its possible role in a specific educational context. Techniques employed as part of this study, such as NLP and deep learning, proved that AI is a viable tool to analyse university-wide course experience. In this research, 383 pieces of labelled data were used to train the proposed deep learning model, which we anticipate can potentially be applied by entire university cohorts of student feedback with a reasonable degree of precision and accuracy.

This study reinforces the importance of pre-processing the collected raw data to eliminate the typographical errors commonly found in student feedback as part of the data cleaning process (Hagiwara & Mita, 2019). The adoption of an educational framework has been shown to be useful if there is clean labelled data that has been collected. Data cleaning is the most costly and time-consuming aspect of an AI project, and the data sets used were commonly processed by enabling spelling correctors like Grammarly into the feedback system (Ghufron & Rosyida, 2018). However, domain-specific language, sarcasm and ambiguity continue to be a challenging part of NLP, particularly in uncovering the latent semantic meaning of the feedback (Shaik et al., 2022). This study addressed this challenge using named entity recognition, rule-based, statistical, deep learning and transformers modelling. Additionally, to process emoticons and special characters that the students used to express their sentiment in feedback,
a multimodal approach, which is an NLP approach (A. Aggarwal et al., 2019), successfully converted the emojis to their corresponding unicode, or in the case of an image, processed these to determine the sentiment.

Insights from the study indicate that the detailed labelling of the initial course data sets enabled the AI model to have high accuracy in the areas of the OEF that had the most labels. The emotional area had the least labels, and as such, its accuracy was less than other areas, where large numbers of labels were evident. However, we note that data still needs to be examined for internal bias of those who label the original data sets. This study did not intend to undertake an examination of the generation of unbiased labelling, but internal bias needs to be considered in future work in this area.

Both administration and teaching perspectives can be informed by the resultant visualisation of a finely tuned NLP model. NLP assists educational institutions to process their vast student responses to student evaluation of teaching surveys with less effort (Kastrati et al., 2021). It can also analyse student opinion in terms of positive and negative sentiment, as well as the types of online engagement evident in student feedback (Shaik et al., 2023). Rather than the manual handling, the ability to train a computer to read and surface entire cohorts of students’ feelings and show their types of engagement with the click of a button adds value in two key areas. Firstly, it adds value to the reflective discussion for all staff who seek to improve their online engagement; secondly, it adds insights into administrative decisions about where and how resources might be deployed to enhance online engagement (Tao et al., 2023). With training, the deep learning model can classify huge volumes of unlabelled data and assist educators by reducing their efforts and time in doing repetitive tasks.

In answering the second research question “How can this data inform future offerings, pedagogical practice, and decision-making particularly related to students' valuing of online engagement?”, NLP has widespread applications across products and service-based applications to analyse end user feedback, including categorising student responses as to the types of engagement that they most value to support their learning. Eggert (2022) reinforced the value of NLP data to enhance the student learning experience, personalised learning management systems and teacher training. Based on the results from the prototype discussed in this study, this type of data offers great potential for educational institutions, including the value it offers teachers in reflecting on their pedagogical practices.

Insights from this type of data analysis also helps teachers and other stakeholders to make informed decisions on future policies and practices. It is of value at the various levels of a university, particularly in viewing online engagement trends evidenced in student perceptions and feedback in data sets. The associated sentiments of these trends may enable decision makers-and educators to enhance their course, programme and/or programme offerings.

Finally, access to such fine-grained visualisation of student feedback using NLP can also be used by teachers and course teaching teams to inform their teaching and learning choices. Each researcher in the research team found value in viewing the data and the insights it offered to their engagement strengths and weaknesses. The administrators in the research team also saw opportunities to understand how they could best apportion resources to enhance the students’ experience.

Conclusion

In this study, the resultant data gathered through NLP, and particularly visualisation of the data (see Table 3), afforded the course team insights into the prevalence of each engagement type and student perspectives of this engagement reflected in student feedback responses. Pedagogical approaches or practices being implemented in an educational institution can be evaluated based on their students’ feedback (Sbaffi & Zhao, 2022). The results from this prototype can reflect the trends of student feedback and their engagement towards the practices being implemented in the university. The prototype can show students’ positive or negative sentiment towards the practices. This data enables management to analyse and change their strategic decisions to enhance student learning experiences.
Future research in the use of AI tools based on educationally labelled data needs to consider the inbuilt bias that may be present in the labelling process, the balancing of the data categories and the collection of data for labelling so that it is clean and usable by those who develop the AI tool. How to present data and research findings requires further thought about how data can be shared and interpreted in ways accessible to teachers and key stakeholders to inform practice and decision-making. This research is limited because the data comes from one regional university in Australia. Collecting more data from other universities and across schools and faculties for initial labelling will assist in improving the accuracy and robustness of the outcomes.

In conclusion, NLP and AI processes are often driven by data science with a positivist research view. This research has brought a multidisciplinary team together with a predominately educational lens rather than a purely scientific view of data, which sets the scene for more education-driven development in the field of AI and NLP. In this study, NLP made it possible for AI to understand human language, listen to student opinions and feedback in the educational context and make sense of course feedback data. The use of AI reduces resources and the time required to read and understand vast amount of student feedback being generated at the end of each semester in the university. NLP has the potential to read student responses and understand their latent semantic structure in terms of sentiment. This enables opportunities for educators and educational administrators to gain insights into large volumes of data if they have access to a labelled set of initial data to train new models.

Acknowledgements

The authors would like to acknowledge the assistance of Dr Jacqui McDonald in project-managing this research.

References


Brown, A., Lawrence, J., Axelsen, M., Redmond., Turner, J., Maloney, S., Galligan, L. (2024). The effectiveness of nudging key learning resources to support online engagement in higher education courses. Distance Education, 45(1), 83–102. https://doi.org/10.1080/01587919.2024.2303491


**Corresponding author:** Thanveer Shaik, Thanveer.Shaik@unisq.edu.au

**Copyright:** Articles published in the *Australasian Journal of Educational Technology* (AJET) are available under Creative Commons Attribution Non-Commercial No Derivatives Licence (CC BY-NC-ND 4.0). Authors retain copyright in their work and grant AJET right of first publication under CC BY-NC-ND 4.0.