ENGINEERING Management 2026, 13(1): 105–123
https://doi.org/10.1007/s42524-026-4186-7

RESEARCH ARTICLE

Xinyi XU, Yuxuan MU, Jin XUE, Petr MATOUS

Large Language Models for analyzing project stakeholders' environmental opinions to support decision-making

Abstract In the global context of sustainable development, stakeholder concerns about the environmental impacts of infrastructure projects have become increasingly prominent, which can significantly influence the progress of projects. However, integrating changing environmental opinions into project decision-making remains a challenge due to the complexity, highly dynamic nature and volume of data. Large Language Models (LLMs) have emerged as transformative tools for efficiently and rapidly analyzing this type of data, offering new opportunities for enhancing decision-making processes. This research proposes a framework utilizing LLM for three major approaches in opinion analysis among stakeholders: sentiment analysis, stance analysis, and topic modeling. The framework has been applied to the case of the Scarborough Gas Project in Western Australia. A set of smaller models, including Neural Networks (NNs), Support Vector Machines (SVMs), Random Forest, Logistic Regression, and BERT, were fine-tuned using GPT-3.5 as a base and compared for performance in sentiment and stance analysis, with SVM achieving the highest accuracy rates of 83.90% and 87.55%, respectively. Integrating LLMs into topic modeling also significantly enhanced the interpretation of stakeholder environmental opinions by transforming keyword lists generated by traditional LDA methods into coherent narratives, reducing reliance on human interpretation, refining themes, and enabling a more comprehensive understanding of environmental, political, and legal issues. This study presents the first unified framework that integrates LLM embeddings with external classifiers to simultaneously analyze all three analytical tasks, to our knowledge. Central to the framework is the theoretically grounded Sentiment–Stance–Topic Matrix and Decision-Making Map, which systematically translate unstructured stakeholder input into prioritized engagement actions. By categorizing sentiment, stance, and topic configurations into targeted strategies, the framework offers structured, data-driven guidance for project decision-makers. This approach bridges gaps in traditional stakeholder analysis and provides a transferable decision-support tool, enabling more inclusive, responsive project governance aligned with global sustainable development goals.

Keywords online stakeholder engagement, large language models, decision-making, sentiment analysis, topic modeling, environmental discourse

Received Sep. 29, 2024; revised Jun. 10, 2025; accepted Jul. 2, 2025

Xinyi XU, Yuxuan MU, Jin XUE (✉), Petr MATOUS
School of Project Management, University of Sydney, Sydney, NSW 2006, Australia
E-mail: jin.xue@sydney.edu.au

This research was supported by the Engineering Vacation Research Internship Program at the University of Sydney.

1 Introduction

The importance of environmental conservation within the context of sustainable development has been increasingly acknowledged. As public awareness and concern for environmental issues rise, enhancing environmental performance has become a key goal in project management (Gao et al., 2022). Infrastructure projects, in particular, present significant environmental challenges, such as pollution, ecosystem destruction, and high greenhouse gas emissions (Büyüközkan and Karabulut, 2017; Cuppen, 2018). These well-documented issues are a major cause for concern, unsettling long-term ecological equilibria (Oyewole et al., 2023).

As a part of healthy democratic processes, members of the public often express concerns about large construction, energy, and infrastructure projects, questioning their environmental impact and project teams' environmental commitments. Such public skepticism, community backlash, and opposition from political and economic entities can present major barriers to project promoters' interests (Hysa and Spalek, 2019). The rapid spread of opinions on social media can further amplify the voices of different groups (Daemi et al., 2021).

Stakeholder engagement has been defined as "a broadly inclusive and continuous process between a project and those potentially affected by it" (Cundy et al., 2013). Stakeholder engagement aims to improve projects by better understanding the concerns of those who might be impacted, involving them in the decision-making process, and ensuring that their opinions, concerns, and feedback are considered. This collaborative process is integral to sustainable project management (Raineri and Paillé, 2016).

Analyzing stakeholder environmental opinions helps in identifying potential conflicts that may emerge around projects, as stakeholder perspectives toward environmental issues can significantly influence the social license and legitimacy of any public endeavor (Ershadi and Goodarzi, 2021). Incorporating stakeholder opinions allows project planners to develop more inclusive decisions, leading to improved environmental outcomes and stronger community relations (Gao et al., 2022).

Traditional methods of stakeholder engagement and opinion analysis, such as surveys, interviews, and public consultations, provide valuable insights but are often time-consuming, costly, and limited in capturing real-time, broad audience sentiments (Project Management Institute, 2022). Natural Language Processing (NLP) -based techniques like sentiment analysis, stance analysis, and topic modeling have been widely used in recent years as they allow more automated ways to filter, synthesize, and interpret stakeholder opinions (Harrin, 2021). These methods categorize stakeholder emotions and stances and uncover hidden themes within their discussion (Pang and Lee, 2008). However, such approaches that rely on traditional machine-learning techniques have been criticized for inaccuracy. Additionally, these come with high model training costs and a steep learning curve to achieve better performance.

Large Language Models (LLMs) like GPT represent a broadly accessible, cheap, and easy-to-implement advancement in processing written language, offering superior performance in tasks such as sentiment analysis and topic modeling (Andrus et al., 2022; Kim et al., 2024; Kwon et al., 2024). These models, pre-trained on vast amounts of text data, can be fine-tuned with labeled data for specific applications, capturing subtle linguistic nuances and complex semantic relationships more effectively than traditional methods (Devlin et al., 2019). By leveraging LLMs, researchers and practitioners can achieve higher accuracy, flexibility, and efficiency in analyzing stakeholder opinions.

This research proposes a comprehensive methodological framework that delineates how LLMs can be leveraged effectively across various stages of stakeholder opinion analysis, including environmental opinion data cleaning, sentiment analysis, stance analysis, and topic modeling. The application of this framework is illustrated in a case study of a real-world project, demonstrating LLMs' promise in enhancing stakeholder engagement and facilitating sustainable decision-making.

The specific objectives are:

1) To develop a comprehensive methodological framework for applying LLMs in various aspects of environmental opinion analysis (sentiment analysis, stance analysis, and topic modeling).

2) To validate the framework through a real-world case study, demonstrating its effectiveness in aggregating and analyzing environmental opinions from online sources.

3) To demonstrate how LLM-based opinion analysis can be systematically linked to project-level decision-making processes, enabling structured, data-informed engagement strategies.

The paper is structured as follows. Section 2 reviews current literature on LLM applications in sustainable infrastructure and stakeholder engagement. Section 3 presents the enhanced methodological framework. Section 4 reports the case study results. Section 5 discusses the findings and implications of the proposed framework for stakeholder engagement and project-level decision-making.

2 Literature review

2.1 Analytical methods in environmental opinion analysis

Analyzing stakeholders' opinions to facilitate decision-making in projects is challenging on multiple levels (Oyewole et al., 2023). Traditional qualitative manual coding methods include content and thematic analysis for pattern identification (Khanra et al., 2021; Lam et al., 2019; Praveena and Aris, 2021). Quantitative statistical analysis of correlates in survey data can detect drivers of environmental opinions and their trends (Jaiswal et al., 2018; Marquart-Pyatt, 2015; Project Management Institute, 2022). While these approaches are informative, they are labor-intensive and time-consuming, preventing real-time response to stakeholder feedback.

Further analytical techniques for stakeholder opinions have emerged based on NLP applications to suddenly ubiquitously available online data. These include sentiment analysis, topic extraction, and stance detection. Sentiment analysis, which usually categorizes emotions within text into positive, negative, or neutral tones (Pang and Lee, 2008), is a popular method for gauging public sentiment. Various studies have been conducted on public sentiment on environmental topics (Amangeldi et al., 2024), including climate change (Corbett and Savarimuthu, 2022; Effrosynidis et al., 2022), renewable energy transition (Bucur et al., 2024; Jeong et al., 2023, 2023; Mäntylä et al., 2018) and environmental policies (Fu et al., 2024).

Sentiment analysis can be combined with topic extraction. Topic extraction includes frequency-based word cloud construction and more sophisticated topic modeling approaches, such as those based on Latent Dirichlet

Allocation (LDA) (Blei et al., 2003; Laureate et al., 2023). These approaches have been applied, for example, to stakeholder concerns about megaprojects (Xue et al., 2020), construction projects (Lai and Kontokosta, 2019), and gas projects (Dokshin, 2021). Sentiment analysis, however, is generally not most effective in capturing the stance of individuals toward complex social or political issues.

Stance detection, by contrast, allows the identification of stakeholders who oppose an issue. It provides a complementary angle to sentiment analysis (ALDayel and Magdy, 2021). Compared to the techniques introduced above, stance analysis is relatively rare in the realm of project management. One reason for this rarity is its analytical complexity. Unlike sentiment analysis, which can divide texts into a dictionary of words tagged with sentiment levels, stance analysis requires a deeper understanding of the context of the texts being analyzed (Küçük and Can, 2020). Existing research utilizing stance analysis has focused on detecting public stance on political issues, such as the political discourse on social media ideology polarization (Bucy and Evans, 2022). Opinion stance polarity alone, however, does not fully reflect users' nuanced position toward complex issues (Druckman and Levendusky, 2019). Additionally, current stance analysis usually struggles with identifying implicit opinions, especially when contextual or background knowledge is required (ALDayel and Magdy, 2021). This can result in inaccurate stance classification, further complicating the interpretation of public opinions in digital discourse.

LLMs, like GPT, have revolutionized NLP, showing immense potential in tasks including sentiment analysis, stance analysis, and topic modeling. Traditional approaches in sentiment analysis and stance analysis, including rule-based systems and machine learning techniques like Support Vector Machines (SVMs) and Random Forests, have been supplemented by deep learning methods, notably Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) (Aljebreen et al., 2023). LLMs trained on vast amounts of unlabeled text data allow for fine-tuned application to specific sentiment analysis tasks with superior performance, as demonstrated by the BERT model (Devlin et al., 2019).

Studies like Miah et al. (2024) showcase the effectiveness of combining LLMs with neural machine translation to handle multilingual sentiment analysis. LLMs have also been shown to effectively generate and analyze opinion markers and stance types, such as affective, relational, and epistemic stances (Lewandowska-Tomaszczyk and Liebeskind, 2024). For topic modeling, LLMs significantly outperform traditional approaches like Latent Dirichlet Allocation (LDA) by capturing complex relationships and linguistic subtleties (Jung et al., 2024). Studies have shown that LLM-based topic modeling can produce more coherent and granular topics, enabling a deeper understanding of stakeholder opinions (Jaiswal et al., 2018; Lam et al., 2019). For example, GPT-based approaches have been successfully used in summarizing public concerns about ESG investing, environmental issues, and policy implications through enhanced opinion extraction and analysis techniques (Jaiswal et al., 2018). Overall, LLMs are promising tools to increase the accuracy, flexibility, and efficiency of sentiment analysis, stance analysis, and topic modeling.

2.2 Stakeholder opinions and sustainable decision-making

Previous research has demonstrated the importance of integrating stakeholder opinions in shaping effective sustainability strategies and decisions (Gao et al., 2022; Swalih et al., 2024). For instance, a study by Engert et al. (2016) investigated how companies integrate stakeholder opinions into their sustainability strategies. They found that businesses that actively engage with stakeholders through regular consultations and transparent communication channels are better equipped to identify relevant sustainability issues and develop more effective strategies. A case study by Devine-Wright (2011) on wind farm projects illustrated how local environmental and social concerns shape stakeholder opposition through place attachment. This research underscored the importance of understanding and addressing stakeholder opinions in a timely manner to mitigate conflicts and facilitate project acceptance. Another notable study by Olander and Landin (2008) investigated the impact of stakeholder engagement on construction projects, revealing that early and continuous involvement of stakeholders helped identify potential environmental risks and community concerns, leading to more sustainable project outcomes. Reed (2008) also highlighted that comprehensive information inputs and proper process of stakeholder participation can strongly affect the quality of decisions in environmental management.

Recent studies have continued to support these findings. Freeman et al. (2021) emphasized that effective stakeholder engagement is crucial for organizational success and sustainability, particularly in identifying and addressing environmental, social, and governance (ESG) issues. A study by Morsing and Spence (2019) examined corporate social responsibility (CSR) communications, finding that transparent and consistent engagement with stakeholders fosters trust and better decision-making.

Despite repeated confirmation of the worthiness of strong stakeholder engagement (Chung et al., 2023), challenges and limitations remain, including struggles with effectively balancing diverse stakeholder interests, particularly when they conflict (Frooman, 1999). Methods for analyzing and integrating stakeholder opinions can be resource-intensive and complex. Engert et al. (2016) found that while companies benefit from stakeholder engagement, they also face difficulties in maintaining transparent communication channels and regular consultations due to resource constraints. Moreover, Burke and Clark (2016) highlighted that achieving widespread support and compliance is challenging, as decisions must reflect a broad spectrum of stakeholder values and concerns, and ways to aggregate these costs effectively are often unavailable.

In summary, while the integration of stakeholder opinions into sustainable decision-making processes has proven beneficial in enhancing decision quality and project outcomes, addressing the inherent challenges requires continuous refinement of engagement strategies and analytical methods.

3 Methods

This study proposes a methodological framework that is structured with three main components driven by LLMs: sentiment analysis, stance analysis, and topic modeling, each tailored to meet the specific requirements of sustainable project decision-making, as shown in Fig. 1. After environmental opinion data are collected from X (previously Twitter), sentiment and stance analysis are conducted using a supervised approach, which leverages LLMs for contextual embeddings. A smaller model is then trained on these embeddings to classify sentiments as positive, negative, or neutral and stance as for, oppose, or unsure. Additionally, key performance metrics are used to evaluate the effectiveness of this approach. The unsupervised method used for topic modeling combines keywords generated by Latent Dirichlet Allocation (LDA) with similar tweets. LLMs are then used to summarize these topics, providing a coherent understanding of the main themes discussed by stakeholders. Finally, these three analytics will be integrated into the 3-dimensional matrix, offering comprehensive guidance and decision mapping for various scenarios.

3.1 Data collection

The Scarborough Gas Project was chosen as the case study for this research due to its significant environmental impacts that have sparked substantial public discussion among a wide range of stakeholders. The project has generated extensive online debates, making it a rich source of varied perspectives and data for analysis.

Tweets on Platform X (previously Twitter) serve as the main source of data for this study, providing insights into stakeholder environmental opinions. According to Eemeren and Grootendorst (2003), opinions are 'subjective beliefs or judgments formed by individuals based on their personal experiences, values, or emotions rather than objective evidence'. Tweets, as expressions of personal thoughts and reactions, provide a direct reflection of stakeholders' subjective views on environmental issues, making them an ideal medium to analyze stakeholder environmental opinions.

Data were collected from X using the keyword 'Scarborough gas' project, with a timeframe from January 2021 to November 2022. Twitter API 2.0 was used to collect tweets, which were accessed via the 'tweepy' Python library and Python 3.1, with PyCharm and Google Colab. Through the full archive search function, tweets and metadata, such as author information and interactions, are all extracted for the following analysis.

All collected tweets were then filtered using the GPT-3.5 model to include only environment-related statements. A prompt instructed GPT-3.5 to read indexed tweets and determine if they were directly about the issues of the environment, outputting 'Yes', 'No', or 'Unsure' for each. These results were converted into a datasheet and matched with the original tweets for further analysis. Some examples are included in Table 1.

Within the prompt, all the tweets were provided with sequential indices, and at the beginning of the prompt, processing instructions were given. It asks GPT to 'read the given texts with sequential indices, tell me whether they are directly about the environment or not', and defines the output format to be a sequential index with

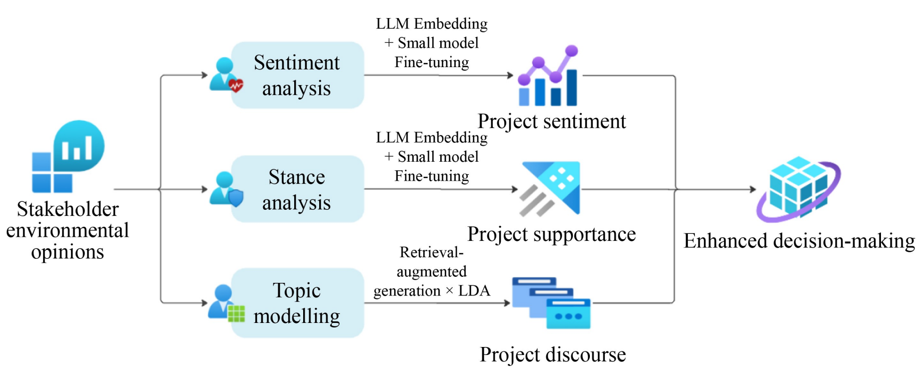

Fig. 1 Conceptual Framework for the three dimensions of analytics for enhancing project sustainable decision-making.

Table 1 LLM labeling sample for environmentally related tweets

Text	Label
The notion that a big 'carbon bomb' floats around the planet causing Global Warming is pure fantasy. Even IPCC (2021) does not spruik such nonsense.	Yes
Job vacancies! Gas Heating Engineer - #Scarborough Asset Data Analyst - #Scarborough Void Cleaner - #Scarborough Relief Responder (@ReachAndRespond) #Whitby Cyber Security Specialist - #Redcar or #Scarborough	No

'Yes', 'No', or 'Unsure' for each. The outputs with specific indices were then transformed into a datasheet and matched with the original tweet texts for filtration.

3.2 Sentiment analysis

Sentiment analysis is essential for gauging public sentiment toward infrastructure projects, particularly those impacting the environment. It provides insights into whether the general public's sentiments are positive, negative, or neutral, allowing decision-makers to understand a project's emotional landscape (Kwon et al., 2024). Traditional methods like SVMs and Naïve Bayes lacked context sensitivity and required extensive feature engineering (Das and Singh, 2023). LLMs address these issues by offering more accurate, context-aware sentiment analysis across diverse languages and complex emotional expressions.

As in Fig. 2, sentiment analysis is performed by first generating contextual embeddings for the text using the OpenAI API GPT-3.5. These embeddings are then input into traditional classifiers like SVMs, logistic regression, or neural networks, which are trained to classify sentiment as positive, negative, or neutral. The use of LLM-generated embeddings allows these models to effectively capture the nuanced meanings within the text, improving the accuracy of sentiment detection. By integrating LLMs with traditional models, this approach addresses the limitations of conventional methods, providing a more accurate and scalable solution for sentiment analysis in online environmental opinions.

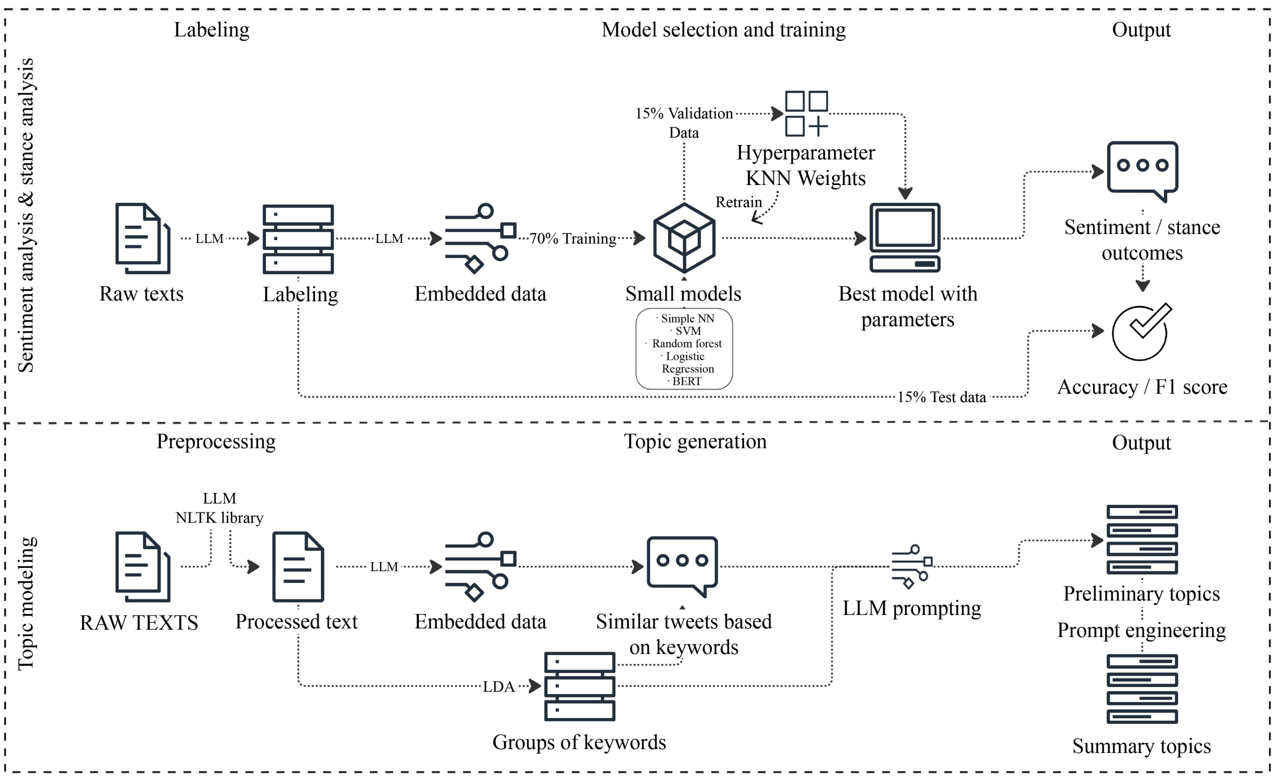

Fig. 2 LLM-based method flowchart for sentiment analysis, stance analysis, and topic modeling.

3.2.1 Setup and labeling

The sentiment analysis utilized the entire data set of 3445 rows. The data was split into three subsets: 70% for training, 15% for validation, and 15% for testing. This commonly applied divisions aim to ensure that there is sufficient data for effective model learning while allocating enough data for validation and testing purposes. It balances data availability for training, hyperparameter tuning, and performance evaluation under resource-intensive cases, as recommended in previous studies (Ma et al., 2024; Mathew et al., 2021; Roy et al., 2021). Sentiment labeling centered on analyzing the tone of each tweet, categorizing them as positive, neutral, or negative, with illustrative examples provided in Table 2.

3.2.2 Modeling

The modeling process began with the generation of contextual embeddings for the tweets using the OpenAI API. These embeddings served as input to five models for sentiment classification: Simple NNs, SVMs, Random Forest, Logistic Regression, and BERT. Then, for each model, hyperparameter tuning was performed to identify the optimal settings for maximizing model performance. Next, K-Nearest Neighbors (KNN) weight optimization was carried out to determine the ideal weight for integrating KNN predictions with those of the top-performing model. Finally, the selected model was retrained using a combination of the training and validation data sets, integrated with KNN, and evaluated on the test set to assess its effectiveness.

Among the above steps, KNN plays a crucial role in improving classification accuracy. This method identifies the closest tweets (neighbors) based on similarity, with closer neighbors having more influence on the sentiment prediction. The labels of these neighbors were then weighted inversely by their distances to compute KNN prediction probabilities. These probabilities were combined with the probabilities from the LLM model using a weighted average. The combined probabilities determined the final sentiment prediction, effectively leveraging both the neighborhood information from KNN and the contextual understanding from the LLM, with the formulas as follows.

Let:

- E be the embedding of a tweet.
- $K = \{k_1, k_2, ..., k_n\}$ be the set of KNN labels for the nearest neighbors.
- $D = \{d_1, d_2, ..., d_n\}$ be the set of distances to the nearest neighbors.
- P_{LLM} be the LLM probability distribution for sentiment classes.
- ω be the weight for KNN in the combined probability.
- c be the sentiment class (positive, neutral, negative).

The KNN probability distribution P_{KNN} can be calculated as:

$$P_{\mathrm{KNN}}[c] = \frac{\sum_{i=1}^{k} \frac{1}{d_i} \times \mathbf{1}(k_i = c)}{\sum_{i=1}^{k} \frac{1}{d_i}}$$

where $\mathbf{1}(k_i = c)$ is an indicator function that is 1 if the ith nearest neighbor's label k_i is equal to class c, and 0 otherwise.

Next, the combined probability distribution P_{combined} is given by:

$$P_{\mathrm{combined}} = \omega \times P_{\mathrm{KNN}} + (1-\omega) \times P_{\mathrm{LLM}} .$$

Finally, the combined prediction $\hat{y}$ is the class with the highest combined probability:

$$\hat{y} = \mathrm{argmax}_c P_{\mathrm{combined}}[c] .$$

3.2.3 Output and evaluation

Two primary metrics were used to evaluate the performance of the sentiment analysis models: accuracy and F1 score. First, the accuracy provides an overall measure of correctly predicted instances out of all instances.

Table 2 Sentiment and Stance Label Example

	Text	Label
Sentiment	Join us online tonight for a supporter event guaranteed to be more exciting than a barrelful of particularly whimsical monkeys. @TomCBallard is MC. #SayNoToScarborough is the campaign. 4pm AWST is the time.	Positive
	The notion that a big 'carbon bomb' floats around the planet causing Global Warming is pure fantasy. Even IPCC (2021) does not spruik such nonsense.	Neutral
	Here we go again. Solve the climate crisis by approving a project that will emit over three times Australia's current emissions by itself. Who says there is intelligent life on earth?	Negative
Stance	@LindaAshton2 @FromeneKa @davobob @elliemail @E49Gillian Labor just approved the Scarborough gas, so we will be right.	For
	The notion that a big 'carbon bomb' floats around the planet causing Global Warming is pure fantasy. Even IPCC (2021) does not spruik such nonsense.	Unsure
	Join us online tonight for a supporter event guaranteed to be more exciting than a barrelful of particularly whimsical monkeys. @TomCBallard is MC. #SayNoToScarborough is the campaign. 4pm AWST is the time.	Against

However, it can be misleading in imbalanced data sets, as it does not account for differences in the model's ability to handle minority classes (Huang and Ling, 2005). Secondly, the F1 Score is a balanced measure that accounts for both false positives and false negatives, making it particularly useful in scenarios with imbalanced data. It is the harmonic mean of precision and recall. Precision quantifies the proportion of all predicted positives (including false positives) that were actually positive. Recall captures the proportion of correctly identified positive instances among all actual positive instances in the data, including those that were not identified (Goutte and Gaussier, 2005). By using both accuracy and the F1 score, a comprehensive evaluation of the model's performance is ensured (Bello et al., 2023; Elmitwalli and Mehegan, 2024).

The Accuracy and F1 scores are calculated as follows:

Let:

- True Positive (TP): Correctly predicted positive cases; the model identifies an actual positive as positive.
- True Negative (TN): Correctly predicted negative cases; the model identifies an actual negative as negative.
- False Positive (FP): Incorrectly predicted positive cases; the model identifies an actual negative as positive.
- False Negative (FN): Incorrectly predicted negative cases; the model identifies an actual positive as negative.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}.$$

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall}.$$

$$precision = \frac{TP}{TP + FP}.$$

$$recall = \frac{TP}{TP + FN}.$$

3.3 Stance analysis

Stance analysis is crucial for understanding public opinion on infrastructure projects, particularly on environmental issues. It goes beyond sentiment analysis by identifying whether the expressed opinions support or oppose a specific issue (Küçük and Can, 2020). This deeper understanding of stakeholder positions helps decision-makers gauge public support or resistance to projects, enabling more informed and responsive decision-making processes. Stance detection has evolved from feature-based methods like SVMs, which require labor-intensive feature engineering, to deep learning models (RNNs, CNNs) that handle sequential data but demand large data sets and high computational power. Traditional methods struggle with nuanced context, word order, and the complexities of social media text (Küçük and Can, 2020). LLMs like GPT improve stance detection by understanding context, capturing word relationships, and eliminating the need for extensive feature engineering. Despite challenges with certain content types, LLMs offer more accurate, reliable, and adaptable stance analysis across diverse topics and domains (Kim et al., 2024).

In this study, stance analysis is performed by first generating contextual embeddings for the text using the OpenAI API GPT-3.5. These embeddings are then fed into various small models, such as SVM, logistic regression, and neural networks, which are trained to classify the stance as supportive, opposing, or neutral. The use of LLM-generated embeddings allows the smaller models to effectively capture the nuanced meanings within the text, improving the accuracy of stance detection. Additionally, a KNN approach is incorporated to refine the predictions by considering the closest neighbors in the embedding space, further enhancing the model's performance. This combined approach leverages the strengths of LLMs and traditional models, providing a powerful and efficient solution for stance analysis in online environmental opinions.

3.3.1 Setup and labeling

Similar to sentiment analysis, the stance analysis utilized the entire data set of 3445 rows. The data was split into three subsets: 70% for training, 15% for validation, and 15% for testing. The stance labeling focused on determining whether the tweet is for, against, or remains unsure of the infrastructure project; some examples are shown in Table 2.

3.3.2 Modeling

The modeling process for stance analysis followed the same procedure as sentiment analysis. It began with generating contextual embeddings for the tweets using the OpenAI API. These embeddings were then fed into five small models to classify the stances, including the Simple NN, SVM, Random Forest, Logistic Regression, and BERT. Each model underwent hyperparameter tuning to identify the optimal settings for performance, with the KNN approach incorporated to enhance classification accuracy. The embeddings were used to find the nearest neighbors for each tweet, and the labels of these neighbors informed the final stance classification. The optimal KNN weight was determined to balance the contributions from both the KNN and the LLM predictions. Finally, the best model was retrained using both the training and validation sets, incorporating the KNN-weighted embeddings, and tested on the test set to evaluate its performance comprehensively.

3.3.3 Output and evaluation

The evaluation metrics for stance analysis followed the same approach as sentiment analysis. The accuracy and F1 scores are used to assess model performance. Accuracy measures the proportion of correctly predicted instances, providing an overall performance measure, but it can be misleading with imbalanced data sets. The F1 score, the harmonic mean of precision and recall, offers a balanced evaluation by considering both false positives and false negatives. Precision indicates how many predicted positive instances are truly positive, while recall measures how many actual positive instances are correctly identified. Using both metrics ensures a comprehensive evaluation, balancing overall correctness with sensitivity to class imbalances.

3.4 Topic modeling

Topic modeling helps to automatically identify and summarize key themes and concerns from large volumes of text data, such as tweets or forum discussions (Wallach, 2006). For environmental opinions on infrastructure projects, topic modeling enables stakeholders to quickly grasp the main issues being discussed, such as environmental impact, community support or opposition, and specific areas of concern like air quality or wildlife preservation. This understanding is vital for decision-makers who need to consider public sentiment in their planning and communication strategies (Jiang et al., 2016). Traditional models struggle with topic evolution, word order, and semantic relationships. Balancing model performance with computational complexity remains difficult, as parametric models limit flexibility, while non-parametric models increase complexity (Abdelrazek et al., 2023; Vayansky and Kumar, 2020).

The proposed solution using LLMs addresses the challenges of traditional topic modeling by incorporating contextual embeddings that capture nuanced relationships between words, leading to more coherent and interpretable topics. This approach effectively handles the noisy, sparse, and dynamic nature of social media data, which traditional models struggle with. Additionally, LLMs overcome the limitations of the "bag of words" assumption (Wallach, 2006) by considering word order and context, and they balance high performance with manageable complexity, making them suitable for tracking topic evolution without the computational burden of more complex models.

3.4.1 Pre-processing

As in Fig. 2, this method starts with pre-processing the entire data set of 3445 rows of tweets. The preprocessing step is crucial to ensure clean, standardized text for enhanced contextual understanding, robustness, and efficient handling of complex tasks like misspellings and emojis (Alam and Yao, 2019). The data are first preprocessed using the following prompt with OpenAI API: "Preprocess the following tweet for text analysis by performing lowercasing, removing punctuation, tokenizing, removing stopwords, correcting misspellings, handling emojis, and lemmatizing. Return only a single line of the final processed text, without any labels or additional text: '{tweet}'". After that, additional stopword removal is conducted using the NLTK library for comprehensive cleaning.

3.4.2 Topic Generation

Then, the topic generation process (Fig. 3) begins with the traditional application of Latent Dirichlet Allocation (LDA). The LDA model identifies 20 topics, each represented by the top 20 words, which provides an initial set of keywords for each topic (Jelodar et al., 2019). However, to refine these topics and make them more representative, embeddings were generated from the preprocessed tweets by the OpenAI API to find tweets similar to those identified by LDA. These embeddings help contextualize the topics by considering the semantic relationships between words. Based on similarity, the top 40 tweets for each topic were selected.

Finally, the selected similar tweets were combined with the LDA-extracted keywords, using LLM prompt engineering to produce more coherent and refined topics that were insightful and representative of the underlying data. The following prompt was used: "Based on the following keywords and tweets, provide a specific and concise topic that summarizes the main idea in less than 15 words. The topic should be clear and insightful, reflecting the key points discussed in the tweets. Start your response with 'Topic:'."

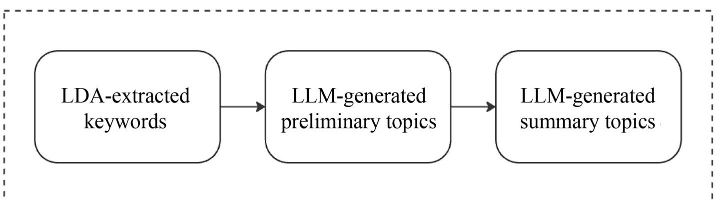

Fig. 3 A brief workflow diagram for the three-step topic evolution.

3.4.3 Summary topic generation

After generating these refined topics, further summarization was carried out to resolve any overlaps and to produce more concise and meaningful topics with prompt engineering in LLM. The prompt was: "For all the topics you have already generated, can you re-summarize them into more concise and meaningful topics, including more details? There are no requirements for the number of topics. Ensure each of them makes sense."

This structured approach of combining LDA with LLM-enhanced embeddings addresses the limitations of traditional topic modeling by ensuring the topics are not only statistically derived but also contextually insightful, providing a deeper understanding of the themes discussed by stakeholders.

3.5 Stance-sentiment-topic opinion matrix

This Matrix visually represents the relationship between three key variables in text data: sentiment (positive, negative, neutral), stance (for, against, unsure), and topic. After finishing the above steps, the embeddings of all the summarized topics are again generated, matching each tweet's embeddings through the OpenAI API.

With these topics, sentiment, and stance classifications, a 3×3 matrix was constructed. Each tweet was plotted as a dot within the matrix, with its position determined by the combination of its sentiment and stance. Dots were color-coded according to the assigned topic, enabling clear visualization of how topics are distributed across different sentiment and stance categories. The matrix was visualized using Python's matplotlib library.

4 Results

4.1 Sentiment analysis

The sentiment results for the Scarborough case study are shown in Fig. 4. It reveals a predominance of neutral and negative sentiments across the data set. Specifically, 68.2% of the data are categorized as negative, while 25.8% reflects neutral sentiment. Only 5.9% of the data are marked as positive. This distribution highlights the predominance of negative sentiment within the data set.

The performance of various models for sentiment analysis was evaluated using the test set. The models included Simple NN, SVM, Random Forest, Logistic Regression, and BERT. The optimal parameters, KNN weights, test accuracy, and F1 scores for each model are summarized in Table 3.

The SVM and Logistic Regression models achieved the highest test accuracy of 84.33%. The SVM also had a high F1 score of 83.90%, indicating that it effectively balances precision and recall in the classification of sentiments. The Simple Neural Network also performed well, with a test accuracy of 83.95% and an F1 score of 83.41%.

The Random Forest model had the lowest performance among the tested models, with a test accuracy of 77.37% and an F1 score of 73.44%. The BERT model, although powerful, showed a slightly lower performance in this context, with a test accuracy of 79.69% and an F1 score of 76.68%.

Overall, the SVM and Logistic Regression models demonstrated the best performance in sentiment analysis, making them suitable choices for accurately classifying the sentiments expressed in the tweets.

4.2 Stance analysis

The sentiment results for the Scarborough case study are shown in Fig. 4. The distribution of stances reveals that the majority of data points (73.7%) are against the discussed topics—about 13.9% fall into the "unsure" category, indicating a degree of uncertainty or indecision. Meanwhile, 12.4% of entries express a "for" stance, reflecting support for the topics. This distribution highlights the predominance of opposing viewpoints within the data set.

The performance of various models for stance analysis

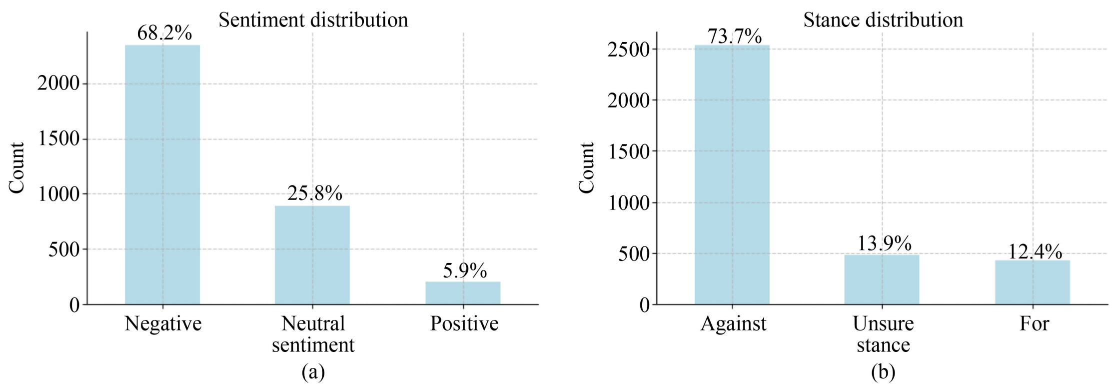

Fig. 4 Sentiment (a) and Stance (b) Distribution for the case project.

Table 3 Comparison of model performance in sentiment analysis and stance analysis

	Model	Optimal parameters	KNN optimal weight	Test accuracy	F1 score
Sentiment	Simple NN	hidden dimension = 64, learning rate = 0.01	0.2	83.95%	83.41%
	SVM	C = 100, kernel = 'rbf'	0.0	84.33%	83.90%
	Random forest	max_depth = 20, min_samples_split = 2	0.1	77.37%	73.44%
	Logistic regression	C = 10, solver = 'liblinear'	0.0	84.33%	83.45%
	BERT		0.7	79.69%	76.68%
Stance	Simple NN	hidden dimension = 64, learning rate = 0.01	0.2	83.95%	82.95%
	SVM	C = 10, kernel = 'rbf'	0.0	88.01%	87.55%
	Random forest	max_depth = None, min_samples_split = 10	0.4	84.53%	83.07%
	Logistic regression	C = 10, solver = 'liblinear'	0.0	87.62%	87.25%
	BERT		0.6	83.17%	82.08%

was evaluated using the test set. The models included Simple Neural Network (NN), Support Vector Machine (SVM), Random Forest, Logistic Regression, and BERT. The optimal parameters, KNN weights, test accuracy, and F1 scores for each model are summarized in Table 3.

The Support Vector Machine (SVM) model achieved the highest test accuracy of 88.01% and an F1 score of 87.55%, indicating its effectiveness in accurately identifying the stance expressed in the tweets. Logistic regression also performed well, with a test accuracy of 87.62% and an F1 score of 87.25%.

The Simple Neural Network (NN) achieved a test accuracy of 83.95% and an F1 score of 82.95%, showing good performance in stance classification. The Random Forest model had a test accuracy of 84.53% and an F1 score of 83.07%, making it a reliable option.

The BERT model, although powerful, showed slightly lower performance in this context, with a test accuracy of 83.17% and an F1 score of 82.08%.

Overall, the SVM and Logistic Regression models demonstrated the best performance in stance analysis, making them suitable choices for accurately classifying the stances expressed in the tweets.

4.3 Topic modeling

Using LDA, 20 initial topics were generated, with examples shown below (Table 4), each represented by the top 20 words.

The initial topics generated by LDA were then refined to be more insightful and representative by LLMs. This process involved using embeddings to find similar tweets and combining them with the LDA keywords to input into the OpenAI API. Example topics after this refinement process are shown in Table 5. It can be observed that topics 9, 14, and 20 involve overlaps, which need to be further summarized in the next step.

Further summarization was applied through LLM prompt engineering, refining the topics from Table 5 into more concise and meaningful summaries. Five of the final summarized topics are presented in Table 6.

4.4 Stance-sentiment-topic opinion matrix

For the following opinion matrix in Fig. 5, a significant portion of the tweets (63.2%) fall under the "against" stance with negative sentiment, which covered most of the topics. This indicates a strong opposition to the Scarborough Gas Project, with negative sentiment dominating discussions. In contrast, neutral sentiment tweets are more evenly distributed across multiple topics, with Topic 1 (general environmental and climate impact of the Scarborough Gas Project) accounting for 9.3%.

Positive sentiment tweets with a "for" stance are less frequent, making up only 7.2% of the total tweets. These primarily center on Topic 5 (approvals and controversies surrounding the project. The visualization highlights that the majority of discussions are characterized by opposition and negative sentiment, with supportive perspectives and positive sentiment playing a minor role.

5 Discussion

This proof-of-concept study demonstrates the potential of employing LLM for systematic online stakeholder environmental opinion analysis to aid decision-making in projects. The high accuracy, precision, and recall of the sentiment analysis, stance analysis, and topic modeling suggest this proposed systematic approach could effectively provide insights for stakeholder engagement targeting according to the stakeholders' sentiment and project support.

5.1 Model performance

The models compared in this study exhibit distinct strengths and limitations, making them suitable for

Table 4 LDA topic keywords examples

Topic Number	Top 20 keywords
1	gas, scarborough, Woodside, project, new, field, gas field, pluto, scarborough gas, emission, climate, fuel, fossil, fossil fuel, labor, go, lng, beetaloo, development, investment decision
2	gas scarborough woodside project scarborough gas gas project scarborough gas project woodside scarborough woodside scarborough gas wa new art rock art rock australia emission billion burrup development pluto
3	gas scarborough project climate labor gas project green light scarborough pluto gas pluto gas light scarborough pluto pluto gas project green pluto target blow climate target ultra labor climate labor climate target
4	gas scarborough scarborough gas project gas project scarborough gas project woodside reef barrier barrier reef climate court halt bid bid halt new damage bn bn scarborough gas bn scarborough
5	gas scarborough project new coal climate gas project beetaloo woodside emission scarborough gas coal gas australia narrabri fuel fossil fossil fuel labor beetaloo narrabri like

Table 5 Insightful topics leveraging LLM and LDA

Topic Number	Topic
1	Woodside's Scarborough Gas Project: Climate Impact and Investment Decision
2	Impact of Woodside's Scarborough Gas Project on Indigenous Rock Art and Emissions
3	Ultra-polluting Scarborough Pluto gas project jeopardizes Labor's climate target green light
4	Legal challenges aim to halt billion-dollar Scarborough gas project damaging Great Barrier Reef
5	Impact of Gas Projects on Climate Change and Emissions
6	Woodside's Scarborough Gas Project Heightens Fossil Fuel Emissions and Climate Concerns
7	Urgent Action Needed to Stop Scarborough Gas Project's Climate Destruction
8	Opposition to Woodside's Scarborough gas project in Western Australia due to environmental and climate
9	Opposition to Woodside's Scarborough Gas Project due to Climate Impact
10	Environmental impact of Woodside's Scarborough Gas Project in Australia
11	Legal Injunctions to Protect Great Barrier Reef from Woodside's Scarborough Gas Project
12	Environmental Impact of Scarborough Gas Project in Australia
13	Opposition to Woodside's polluting Scarborough gas project in Australia.
14	Opposition to Woodside's Scarborough Gas Project for Climate Conservation
15	Controversial Woodside Scarborough Gas Project in Australia Raises Climate Concerns
16	Controversy Surrounding Scarborough Gas Project in Western Australia
17	Woodside and BHP's Scarborough Gas Project receives final go-ahead in Western Australia
18	Labor Support for Scarborough Gas Project In Australia
19	Woodside Scarborough Gas Project: Controversial Fossil Fuel Expansion Amid Climate Crisis
20	Opposition to Woodside's Scarborough Gas Project for Climate and Environmental Concerns

different applications. SVM stands out for its robustness and high accuracy in both sentiment and stance analysis, particularly for small-to-moderate data sets, and is well-suited for real-time monitoring tasks (Cortes and Vapnik, 1995). Logistic regression offers similar performance with low computational requirements, making it ideal for resource-constrained scenarios and situations requiring interpretability. BERT is good at capturing nuanced textual relationships, but is more computationally expensive, making it valuable for complex tasks such as legal or policy analysis (Devlin et al., 2019). Random Forest is effective for exploratory analysis, especially when feature importance is a priority, but is less suited for high-dimensional data or real-time use (Breiman, 2001). Simple Neural Networks, despite their lightweight architecture, perform competitively and scale well, making them suitable for evolving data sets like social media trends.

5.1.1 Project sentiment

The results demonstrate that the combination of the SVM classifier with LLM embeddings can significantly enhance the performance of sentiment analysis, achieving an accuracy of 84.33% and an F1 score of 83.90%, which is shown in Table 3. This outcome surpasses the benchmarks reported in the existing literature. For instance, in a summary of selected approaches to Textual Sentiment Analysis (Das and Singh, 2023), many prevailing sentiment analysis methods struggle to achieve accuracy and F1 scores above 80%. The success of this approach highlights the potential of advanced LLM embeddings to overcome the limitations of traditional sentiment analysis

Table 6 Final summarized topics

Topic Number	Topic
1	**Woodside's Scarborough Gas Project:** Environmental and Climate Impact: Summarizes overall impact on environment, carbon emissions, and climate change
2	**Indigenous Rock Art and Emissions:** Consequences of Woodside's Scarborough Gas Project: Focuses on specific impact on indigenous heritage and emissions
3	**Legal Challenges and Environmental Concerns:** Halting the Scarborough Gas Project: Discusses legal efforts to stop the project and its effects on the Great Barrier Reef
4	**Opposition and Urgent Action:** Stopping the Climate Destruction of Scarborough Gas Project: Highlights the need for immediate action to prevent environmental harm
5	**Labor Support and Fossil Fuel Expansion:** The Scarborough Gas Project: Examines political backing for the project and its implications in the climate crisis

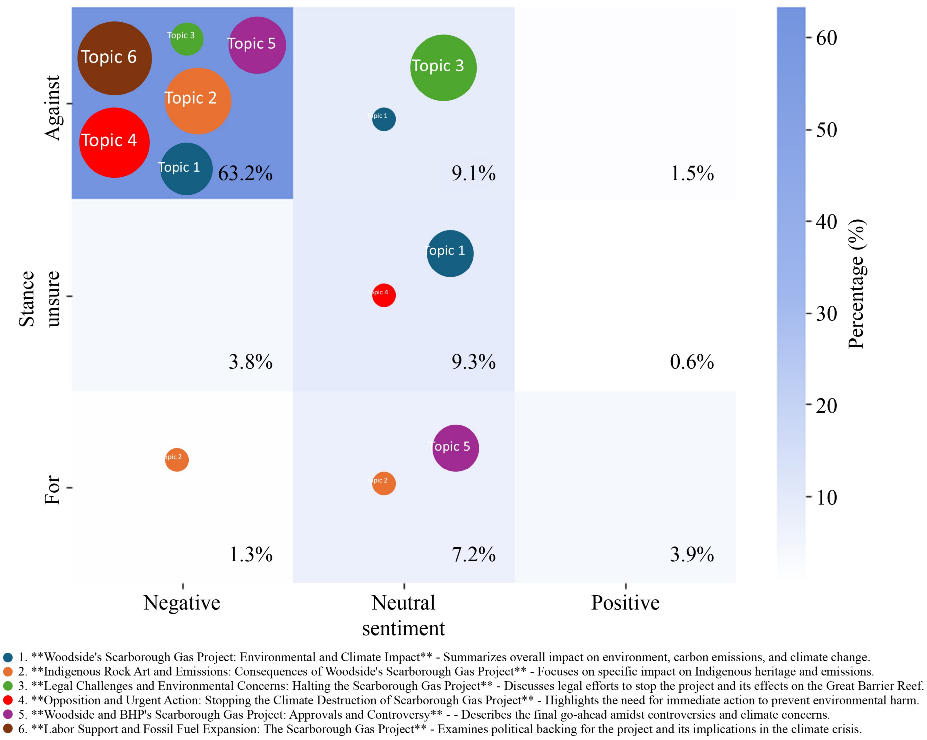

Fig. 5 Stance-sentiment-topic opinion matrix. The 3 × 3 matrix links sentiment (*x*-axis), stance (*y*-axis), and topics (colored circles). Circle size shows topic frequency, and cell percentages reflect tweet proportions for each sentiment-stance pair.

techniques, which often fail to capture the complexity of textual data.

He et al. (2022) emphasized the inadequacy of commonly used sentiment analysis Python toolkits, including TextBlob, Vader, StanfordNLP, etc., with an evaluation of 11 commonly used sentiment analysis tools on seven social media data sets, only getting the average weighted F1 scores across various data sets remains below 0.6. The substantial improvement demonstrated by this model suggests that integrating LLM embeddings with SVM classifiers can significantly enhance the reliability of sentiment analysis, even when the quality of the data set is suboptimal.

The improvement is particularly critical in contexts where sentiment analysis informs policy decisions. Inaccuracies in these analyses could lead to misguided actions, especially when dealing with environmental impacts where stakeholders can understandably be highly sensitive. Accurate sentiment analysis helps in predicting potential conflicts by identifying strong negative feelings among stakeholders regarding project impacts (Bucur et al., 2024). When these concerns are detected early and accurately, project teams can take proactive steps, such as conducting more detailed environmental impact assessments or adjusting project plans, to prevent conflicts from escalating (Gao et al., 2022). Early response to stakeholder concerns reduces the likelihood of protests or significant opposition and, importantly, creates more sustainable project outcomes.

5.1.2 Project stance

Results of stance analysis, also highlighted in Table 3, show that the SVM classifier again delivered superior performance, with an accuracy of 88.01% and an F1 score of 87.55%. These results are especially notable compared to the F1 scores reported in previous studies, where supervised stance analysis typically yields scores between 65% and 70%, with only a few instances surpassing 75% (ALDayel and Magdy, 2021). The considerable improvement in stance analysis highlights the effectiveness of utilizing LLMs to tackle challenges such as nuanced context, word order, and the complexities inherent in stance expression—areas where traditional methods often struggle.

Improving the accuracy of stance analysis is vital for understanding the support and opposition dynamics among stakeholders (Taylor and Rosca, 2024). This precise insight allows decision-makers to more accurately assess the project's feasibility and its likelihood of acceptance in the future. By effectively identifying and analyzing stakeholder stances, project teams are empowered to make informed decisions, proactively addressing potential opposition while strengthening support, ultimately leading to more successful project outcomes (Ki et al., 2020).

5.1.3 Topic modeling

The integration of LLMs in topic modeling significantly improved the interpretation of stakeholder concerns across iterations. Initially, the traditional LDA-based method provided only a list of keywords, such as "gas, Scarborough, project, emission, climate, labor …" as listed in Table 4, which, while relevant, required human interpretation to extract meaningful insights. Without this manual effort, the keywords alone often lacked the context needed to fully understand the nuances of stakeholder discourse, limiting their utility for efficient decision-making.

With the introduction of LLMs, the interpretation capabilities markedly improved. The combination of LLM-generated contextual embeddings with LDA keywords transformed these disjointed word lists into coherent narratives. For instance, topics evolved into themes like "Ultra-polluting Scarborough Pluto gas project jeopardizes Labor's climate target green light," as summarized in Table 5. This shift enabled a clearer understanding of the key environmental issues stakeholders were discussing, effectively enhancing the depth of analysis without relying on human interpretation.

In the second iteration, further refinement using LLM-based summarization through prompt engineering reduced the number of topics, making the summaries more concise and focused. As shown in Table 6, themes like "Labor Support and Fossil Fuel Expansion: The Scarborough Gas Project: Examines political backing for the project and its implications in the climate crisis" offer a much sharper focus on the political aspects of stakeholder concerns and their direct implications for the climate crisis.

The LLM-enhanced topic modeling approach ultimately enabled a more comprehensive and accurate understanding of complex stakeholder opinions by successfully identifying key themes and topics discussed by stakeholders, which is key in project management (Li et al., 2023). The final summarized topics offered a coherent understanding of the primary concerns, including the environmental and climate impacts of the project, the implications for Indigenous heritage protection, legal challenges, and political support. Moreover, decision-makers can now choose the level of topic granularity based on their specific needs, ensuring that the analysis is as broad or detailed as required for informed decision-making. These findings illustrate the multifaceted utility of LLMs in extracting and summarizing complex stakeholder opinions, ensuring that critical issues are thoroughly considered for informed, sustainable decision-making and conflict prevention (Elsawah et al., 2020).

5.2 Decision support diagram through stance-sentiment-topic mapping

The decision support diagram, derived from the Opinion Matrix in Fig. 6, is designed to facilitate project decisions by providing a structured analysis by combining the three analytics. It categorizes stakeholder environmental opinions into different quadrants and offers actionable strategies for stakeholder engagement. Some quadrants have been merged where similarities exist, resulting in a total of six decision-making strategies for the original nine sentiment-stance quadrants.

5.2.1 Against-negative: Strongly opposed and emotionally resistant

The most challenging scenario for project leaders occurs when stakeholders express both strong emotional opposition and firm disapproval of the project. These groups are often driven by significant concerns about the environmental impact and may organize resistance.

The Scarborough Gas Project has met with substantial opposition due to its environmental and climate impacts. Environmental groups, such as the Australian Conservation Foundation (ACF), argue that the project will contribute significantly to global greenhouse gas emissions—over 1.37 billion tonnes of CO_2 over its lifetime—posing a threat to critical ecosystems like the Great Barrier Reef. This opposition reflects a strong negative sentiment and opposing stance from stakeholders concerned about its contribution to climate change and the destruction of marine habitats. Additionally, topics include legal challenges and the impact on Indigenous heritage sites like the Murujuga rock art. These combined environmental and cultural impacts have fueled legal actions and broader public opposition, making it a focal point for climate activism in Australia (Smail, 2022).

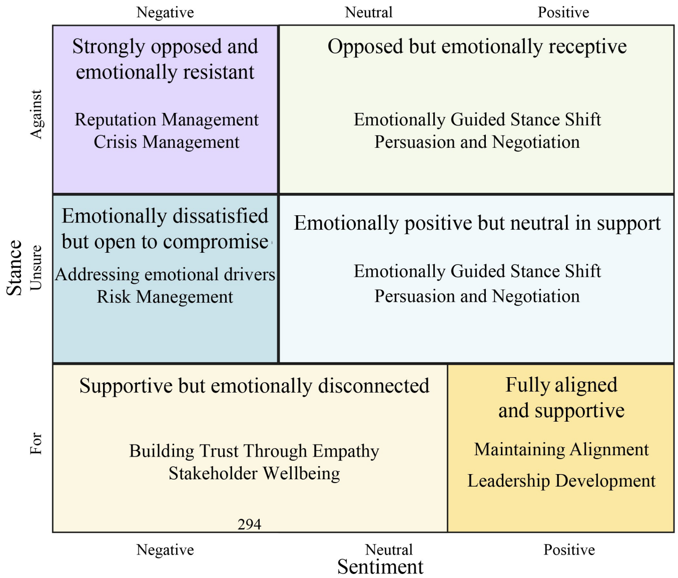

Fig. 6 Decision-making map for stakeholder engagement strategies based on environmental opinions.

Sample Tweets in this quadrant:

· "Here we go again. Solve the climate crisis by approving a project that will emit over three times Australia's current emissions by itself. Who says there is intelligent life on earth?"

· "Cool, do Woodside's Scarborough to Pluto Gas project next Tanya, the emissions will equal 15-16 coal large mines which is why the ACF are taking it court, on the grounds of the impact on the Great Barrier Reef. #GBR #auspol."

Decision-makers must prioritize a crisis management approach, recognizing that the environmental concerns raised in this quadrant are both urgent and valid. Immediate action to mitigate the environmental damage, through tangible changes or compromises, should be a top priority. While addressing these issues may begin to rebuild trust, reversing public opposition will require a long-term, sustained commitment to environmental responsibility and transparent communication (Buysse and Verbeke, 2003). Failure to act decisively risks exacerbating public dissent and further eroding stakeholder confidence.

5.2.2 Against-neutral/positive: Against but emotionally receptive

Some stakeholders may hold firm opposition based on specific environmental grounds coming from news, reports, or experts, but remain emotionally neutral or receptive, offering an opportunity for dialog. Decision-makers should focus on connecting with their values that resonate with their environmental concerns, responding in a rational way that hits back to the grounds, or adjusting the project based on the environmental grounds they provided (Bu et al., 2025).

However, it is crucial to differentiate between emotional receptiveness and emotional positivity that stems from active opposition. For instance, stakeholders are emotionally positive because they are actively engaged in opposing the project, not because they are receptive to changing their stance. Decision-makers must recognize this distinction and tailor their engagement strategies accordingly, referring to the solutions under the Strongly Opposed and Emotionally Resistant scenario.

· Join us online tonight for a supporter event guaranteed to be more exciting than a barrelful of particularly whimsical monkeys. @TomCBallard is MC. #SayNoToScarborough is the campaign. 4pm AWST is the time.

5.2.3 Unsure-negative: Emotionally dissatisfied but open to compromise

Stakeholders in this category may express environmental frustration or skepticism about the project but remain neutral in their formal position. While they are dissatisfied, primarily due to perceived environmental risks, they are not entirely opposed to the project and might be persuaded by the right approach. Decision-makers can persuade these stakeholders by addressing their environmental concerns directly and offering clear, transparent communication on how the project aims to mitigate environmental harm (Reed, 2008).

In this category, there may also be cases where dissatisfaction is nuanced and masked by neutral language. For example, some stakeholders may sound neutral but harbor strong reservations, requiring decision-makers to probe deeper to address their environmental concerns effectively.

5.2.4 Unsure-neutral/positive: Emotionally positive but unsure in support

In some cases, stakeholders may express positive emotional sentiment toward environmental protection but are uncertain about fully opposing the project. Similarly, it is crucial to distinguish between emotional receptiveness and emotional positivity, which stem from skepticism, under this category (Vining, 1987). For example, in this case, stakeholders may appreciate the legal actions taken by environmental groups like the ACF, yet this appreciation may arise from shared concerns about environmental impacts rather than an outright rejection of the project.

Sample tweets in this quadrant:

· "An environment group has launched a legal bid to halt a $16 billion gas development in Western Australia, arguing the effect of its greenhouse gas emissions on the Great Barrier Reef will be significant and should be assessed under national environment law."

· "The Australian Conservation Foundation has made a move to stop works on @WoodsideEnergy's Scarborough gas project…"

The engagement strategy should focus on building trust and providing clarity on the project's environmental impact while addressing their concerns without alienating them, in this case. Ideally, this scenario presents an opportunity for decision-makers to convert positive sentiment into more concrete support (Fulton et al., 2015).

5.2.5 For-negative/neutral: Supportive but emotionally disconnected

Stakeholders may believe in the project's benefit but feel emotionally disengaged or even mildly dissatisfied about other aspects, such as certain government officials and business leaders in this case. For instance, while Australia's Resources Minister Madeleine King supports the project, citing its importance for energy supply to Asia-Pacific countries like Japan and Republic of Korea, she acknowledges the broader climate concerns tied to its greenhouse gas emissions. Supporters like her often advocate for the project's necessity in the short term energy transition but emotionally disengage from or downplay the longer-term environmental risks it poses.

This emotional detachment of stakeholders must not be overlooked (Bu et al., 2025). While they may express practical support for the project, addressing their environmental concerns is still crucial. Fostering a sense of engagement by actively listening to and responding to these concerns can help secure long-term support (Holifield and Williams, 2019).

However, some tweets may fall into this category inaccurately due to sarcasm or indirect opposition. For example, stakeholders may express support ironically, referencing others' support while subtly criticizing the project's environmental shortcomings. Analysts need to recognize such potential misclassification and ideally consider the tone and context of these communications.

5.2.6 For-positive: Fully aligned and supportive

In ideal circumstances for project proponents, stakeholders express both emotional and logical support for the project (Perlaviciute et al., 2018). However, in the case of the Scarborough Gas Project, this category remains limited. A small proportion of them argue that natural gas, emitting less CO_2 than coal, is vital for the energy transition. However, these views still have opponents asserting that the project contributes to significant greenhouse gas emissions.

Even among supporters, maintaining this alignment will require constant engagement and transparency, especially regarding environmental sustainability. For instance, while some stakeholders support the project for its economic contributions, they may still expect reassurances that Woodside is taking steps to minimize environmental impacts. In an ideal case, this level of positive alignment can also provide an opportunity to further strengthen stakeholder relationships and bolster the project's public image, turning stakeholders into advocates (Buysse and Verbeke, 2003).

5.3 Limitations

Despite the promising results, this study has some limitations that warrant attention in future research. One of the main challenges is that the accuracy of sentiment and stance analysis depends largely on the quality and size of the data set. While smaller data sets allow for more efficient analysis due to faster model training, they may limit the overall accuracy. Expanding the data set to include other platforms beyond X, such as forums and traditional media, could improve representativeness and offer a more comprehensive view of stakeholder opinions.

A high proportion of negative sentiment and opposition in the data set creates challenges for model training and evaluation due to imbalanced data. Although performance metrics like F1 scores were used to mitigate this, unbalanced data sets can skew the results, leading to biased interpretations. Further research is needed to develop more advanced techniques for handling such imbalances, which may include re-sampling methods, algorithm adjustments, or the incorporation of more sophisticated machine-learning strategies. Additionally, there are inherent biases in LLMs that may influence the objectivity and impartiality of the analysis (Navigli et al., 2023). Ongoing model refinement and continuous monitoring are essential to reduce the risk of biased outcomes, particularly when handling sensitive topics.

While the proposed framework demonstrates strong performance in analyzing sentiment and stance in public statements, we refrain from direct comparisons with results of analyses that used other social media data sets in previous studies and other types of data. However, the inherent properties of LLMs offer approaches to analyzing diverse data of various structures. Extensive pretraining on diverse data sets and the ability to capture subtle linguistic nuances contribute to a degree of transferability even after domain-specific fine-tuning (Brown et al., 2020). This characteristic reduces the impact of generalizability concerns compared to traditional models.

However, domain-specific fine-tuning remains a crucial avenue for future research, which enhances LLMs' precision and robustness in specialized contexts (Lu et al., 2025). Future research should explore further enhancements to the framework's adaptability and depth through techniques like transfer learning, bias mitigation, and hyperparameter tuning (Dogra et al., 2024). Additionally, exploring improved LLM architectures and integrating external knowledge bases can further refine model performance, enabling the capture of more nuanced stakeholder opinions across diverse data sources.

6 Conclusions and contribution

This study presents a theoretically grounded, methodologically innovative, and practically impactful framework for applying LLM to the analysis of online stakeholder environmental opinions and supports strategic decision-making. Through sentiment analysis, stance analysis, and topic modeling, the framework provides a robust and multifaceted approach to opinion analysis, with its effectiveness demonstrated through real-world case studies with satisfactory accuracy. Moreover, this LLM-integrated approach equips project managers with detailed and actionable insights, thus enhancing sustainable decision-making.

6.1 Methodological contribution

Methodologically, this study proposes a hybrid LLM-embedding with the external classifier architecture that resolves persistent challenges in stakeholder opinion analysis, including data sparsity, class imbalance, and low interpretability of end-to-end models. The framework delivers improved accuracy in both sentiment and stance detection, while the automated topic modeling uncovers in-depth, meaningful topics. This comprehensive approach allows for a more nuanced, efficient interpretation of stakeholder environmental opinions, providing richer insights than previously achieved. Despite some limitations, the framework presents a highly promising tool for future research and practice, and with the continuous refinement of LLM techniques and expanded data sources, it can further elevate the quality of environmental opinion analysis, driving more inclusive and sustainable project decision-making.

6.2 Theoretical contribution

Theoretically, this work establishes a conceptual link between LLM-enabled opinion analysis and stakeholder engagement theory. It introduces the Sentiment–Stance–Topic Matrix as a structured lens for interpreting complex stakeholder discourse and maps it onto a Decision-Making Map with well-defined engagement strategies. This data-driven framework directly informs risk management and stakeholder prioritization in sustainable project planning, offering a scalable method to translate unstructured online opinions into theory-informed, action-oriented interventions. By operationalizing abstract stakeholder constructs with AI-driven analytics, the framework extends the boundaries of traditional environmental decision-making, enabling more responsive, inclusive, and accountable strategies that bridge the gap between technical insight and real-world governance.

6.3 Practical contribution

The study provides a powerful toolkit for project managers, enabling fast and precise conversion of quantitative online stakeholder opinion analysis into qualitative action guidance through a decision-making map. By integrating sentiment, stance, and topic modeling, the toolkit offers actionable insights that help decision-makers understand stakeholder opinions and refine their strategies. This toolkit supports targeted mitigation strategies, prioritizing efforts to build trust and address opposition, ultimately fostering informed decision-making and promoting sustainable project outcomes.

Acknowledgements The early-stage results of this study were published in the 12th IPMA Research Conference proceedings.

Competing Interests The authors declare that they have no competing interests.

References

Abdelrazek A, Eid Y, Gawish E, Medhat W, Hassan A (2023). Topic modeling algorithms and applications: A survey. Information Systems, 112, 102176

Alam S, Yao N (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. Computational & Mathematical Organization Theory, 25(3): 319–335

ALDayel A, Magdy W (2021). Stance detection on social media: State of the art and trends. Information Processing & Management, 58(4): 102597

Aljebreen M, Alabduallah B, Asiri M M, Salama A S, Assiri M, Ibrahim S S (2023). Moth flame optimization with hybrid deep learning based sentiment classification toward ChatGPT on Twitter. IEEE Access: Practical Innovations, Open Solutions, 11: 104984–104991

Amangeldi D, Usmanova A, Shamoi P (2024). Understanding environmental posts: Sentiment and emotion analysis of social media data. IEEE Access, 12: 33504–33523

Andrus B R, Nasiri Y, Cui S, Cullen B, Fulda N (2022). Enhanced story comprehension for large language models through dynamic document-based knowledge graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence, 36(10): 10436–10444

Bello A, Ng S C, Leung M F (2023). A BERT framework to sentiment analysis of tweets. Sensors, 23(1), 133

Blei D M, Ng A Y, Jordan M I, Lafferty J (2003). Latent dirichlet allocation. Journal of Machine Learning Research EBSCOhost. Available at the website of ebsco.com

Breiman L (2001). Random Forests. Machine Learning, 45(1): 5–32

Brown T B, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D M, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33: 1877–1901

Bu Z, Liu J, Liu J (2025). Impact of emotions on the behavioral strategies of PPP project stakeholders: An RDEU evolutionary game analysis. Engineering, Construction, and Architectural Management, 32(4): 2242–2271

Bucur C, Tudorica B, Andrei J V, Dusmanescu D, Paraschiv D, Teodor C (2024). Sentiment analysis of global news on environmental issues: Insights into public perception and its impact on low-carbon economy transition. Frontiers in Environmental Science, 12: 1360304

Bucy E P, Evans H K (2022). Media-centric and politics-centric views of media and democracy: A longitudinal analysis of political communication and the international journal of press/politics. Political Communication, 39(2): 254–265

Burke J J, Clark C E (2016). The business case for integrated reporting: Insights from leading practitioners, regulators, and academics. Business Horizons, 59(3): 273–283

Buysse K, Verbeke A (2003). Proactive environmental strategies: A stakeholder management perspective. Strategic Management Journal, 24(5): 453–470

Büyüközkan G, Karabulut Y (2017). Energy project performance evaluation with sustainability perspective. Energy, 119: 549–560

Chung K S K, Eskerod P, Jepsen A L, Zhang J (2023). Response strategies for community stakeholder engagement on social media: A case study of a large infrastructure project. International Journal of Project Management, 41(5): 102495. Q1

Corbett J, Savarimuthu B T R (2022). From tweets to insights: A social media analysis of the emotion discourse of sustainable energy in the united states. Energy Research & Social Science, 89: 102515

Cortes C, Vapnik V (1995). Support-vector networks. Machine Learning, 20(3): 273–297

Cundy A B, Bardos R P, Church A, Puschenreiter M, Friesl-Hanl W, Müller I, Neu S, Mench M, Witters N, Vangronsveld J (2013). Developing principles of sustainability and stakeholder engagement for "gentle" remediation approaches: The European context. Journal of Environmental Management, 129: 283–291

Cuppen E (2018). The value of social conflicts. Critiquing invited participation in energy projects. Energy Research & Social Science, 38: 28–32

Daemi A, Chugh R, Kanagarajoo M V (2021). Social media in project management: A systematic narrative literature review. International Journal of Information Systems and Project Management, 8(4): 5–20

Das R, Singh T D (2023). Multimodal sentiment analysis: A survey of methods, trends, and challenges. ACM Computing Surveys, 55(13s): 1–38

Devine-Wright P (2011). Public engagement with large-scale renewable energy technologies: Breaking the cycle of NIMBYism. Wiley Interdisciplinary Reviews: Climate Change, 2(1): 19–26

Devlin J, Chang M W, Lee K, Toutanova K (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds), In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1. Association for Computational Linguistics, 4171–4186

Dogra V., Verma S., Kavita, Woźniak M., Shafi J., Ijaz M F (2024). Shortcut learning explanations for deep natural language processing: A survey on dataset Biases. IEEE Access, 12: 26183–26195

Dokshin F A (2021). Variation of public discourse about the impacts of fracking with geographic scale and proximity to proposed development. Nature Energy, 6(10): 961

Druckman J N, Levendusky M S (2019). What do we measure when we measure affective polarization? Public Opinion Quarterly, 83(1): 114–122

Effrosynidis D, Sylaios G, Arampatzis A (2022). Exploring climate change on Twitter using seven aspects: Stance, sentiment, aggressiveness, temperature, gender, topics, and disasters. PLoS One, 17(9): e0274213

Elmitwalli S, Mehegan J (2024). Sentiment analysis of COP9-related tweets: A comparative study of pre-trained models and traditional techniques. Frontiers in Big Data, 7: 1357926

Elsawah S, Hamilton S, Jakeman A, Rothman D, Schweizer V, Trutnevyte E, Carlsen H, Drakes C, Frame B, Fu B, Guivarch C, Haasnoot M, Kemp-Benedict E, Kok K, Kosow H, Ryan M, van Delden H (2020). Scenario processes for socio-environmental systems analysis of futures: A review of recent efforts and a salient research agenda for supporting decision making. Science of the Total Environment, 729: 138393

Engert S, Rauter R, Baumgartner R J (2016). Exploring the integration of corporate sustainability into strategic management: A literature review. Journal of Cleaner Production, 112: 2833–2850

Ershadi M, Goodarzi F (2021). Core capabilities for achieving sustainable construction project management. Sustainable Production and Consumption, 28: 1396–1410

Freeman R E, Dmytriyev S D, Phillips R A (2021). Stakeholder theory and the resource-based view of the firm. Journal of Management, 47(7): 1757–1770

Frooman J (1999). Stakeholder influence strategies. Academy of Management Review, 24(2): 191–205

Fu L, Tu X, Liao J (2024). Asymmetric impacts of climate policy uncertainty, investor sentiment on energy prices and renewable energy consumption: Evidence from NARDL and wavelet coherence. Journal of Environmental Management, 367: 122057

Fulton E, Boschetti F, Sporcic M, Jones T, Little L R, Dambacher J, Gray R, Scott R, Gorton R (2015). A multi-model approach to engaging stakeholder and modellers in complex environmental problems. Environmental Science & Policy, 48: 44–56

Gao X, Zeng S, Zeng R, Shi J J, Song R (2022). Multiple-stakeholders' game and decision-making behaviors in green management of megaprojects. Computers & Industrial Engineering, 171: 108392

Goutte C, Gaussier E (2005). A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In D. E. Losada & J. M. Fernández-Luna (Eds). Advances in Information Retrieval, Springer. 345–359

Harrin E (2021). Social Media for Project Managers. Project Management Institute

He L, Yin T, Zheng K (2022). They may not work! An evaluation of eleven sentiment analysis tools on seven social media datasets. Journal of Biomedical Informatics, 132: 104142–104153

Holifield R, Williams K C (2019). Recruiting, integrating, and sustaining stakeholder participation in environmental management: A case study from the great lakes areas of concern. Journal of Environmental Management, 230: 422–433

Huang J, Ling C X (2005). Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 17(3): 299–310

Hysa B, Spalek S (2019). Opportunities and threats presented by social media in project management. Heliyon, 5(4): e01488

Jaiswal A, Samuel C, Kadabgaon V M (2018). Statistical trend analysis and forecast modeling of air pollutants. Global Journal of Environmental Science and Management, 4: 427–438

Jelodar H, Wang Y, Yuan C, Feng X, Jiang X, Li Y, Zhao L (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. Multimedia Tools and Applications, 78(11): 15169–15211

Jeong D, Hwang S, Kim J, Yu H, Park E (2023). Public perspective on renewable and other energy resources: Evidence from social media big data and sentiment analysis. Energy Strategy Reviews, 50: 101243

Jiang H, Qiang M, Lin P (2016). Finding academic concerns of the Three Gorges Project based on a topic modeling approach. Ecological Indicators, 60: 693–701

Jung H S, Lee H, Woo Y S, Baek S Y, Kim J H (2024). Expansive data, extensive model: Investigating discussion topics around LLM through unsupervised machine learning in academic papers and news. PLoS One, 19(5): e0304680

Khanra S, Dhir A, Kaur P, Mäntymäki M (2021). Bibliometric analysis and literature review of ecotourism: Toward sustainable development. Tourism Management Perspectives, 37: 100777

Ki C W, Chong S M, Ha-Brookshire J E (2020). How fashion can achieve sustainable development through a circular economy and stakeholder engagement: A systematic literature review. Corporate Social Responsibility and Environmental Management, 27(6): 2401–2424

Kim S, Kim K, Wonjeong Jo C W (2024). Accuracy of a large language model in distinguishing anti- and pro-vaccination messages on social media: The case of human papillomavirus vaccination. Preventive Medicine Reports, 42: 102723

Küçük D, Can F (2020). Stance detection: A survey. ACM Computing Surveys, 53(1): 1–37

Kwon O H, Vu K, Bhargava N, Radaideh M I, Cooper J, Joynt V, Radaideh M I (2024). Sentiment analysis of the United States public support of nuclear power on social media using large language models. Renewable & Sustainable Energy Reviews, 200: 114570

Lai Y, Kontokosta C E (2019). Topic modeling to discover the thematic structure and spatial-temporal patterns of building renovation and adaptive reuse in cities. Computers, Environment and Urban Systems, 78: 101383

Lam J, Cheung L, Wang S, Li V (2019). Stakeholder concerns of air pollution in hong kong and policy implications: A big-data computational text analysis approach. Environmental Science & Policy, 101: 374–382

Laureate C D P, Buntine W, Linger H (2023). A systematic review of the use of topic models for short text social media analysis. Artificial Intelligence Review, 56(12): 14223–14255

Lewandowska-Tomaszczyk B, Liebeskind C (2024). Opinion events and stance types: Advances in LLM performance with ChatGPT and Gemini. Lodz Papers in Pragmatics

Li Y, Frans V F, Song Y, Cai M, Zhang Y, Liu J (2023). SDGdetector: An R-based text mining tool for quantifying efforts toward Sustainable Development Goals. Journal of Open Source Software, 8(84): 5124

Lu W, Luu R K, Buehler M J (2025). Fine-tuning large language models for domain adaptation: Exploration of training strategies, scaling, model merging and synergistic capabilities. Npj Computational Materials, 11(1): 84

Ma Z, Mir R, Dalton C A, Godfrey K E (2024). Choosing appropriate regularization parameters by splitting data into training and validation sets—Application in global surface-wave tomography. Seismological Research Letters, 95(5): 3029–3041

Mäntylä M V, Graziotin D, Kuutila M (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. Computer Science Review, 27: 16–32

Marquart-Pyatt S (2015). Public opinion about the environment: Testing measurement equivalence across countries. International Journal of Sociology, 45(4): 309–326

Mathew E S, Tembely M, AlAmeri W, Al-Shalabi E W, Shaik A R (2021). Artificial intelligence coreflooding simulator for special core data analysis. SPE Reservoir Evaluation & Engineering, 24(04): 780–808

Miah M S U, Kabir M M, Sarwar T B, Safran M, Alfarhood S, Mridha M F (2024). A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. Scientific Reports, 14(1): 9603

Morsing M, Spence L J (2019). Corporate social responsibility (CSR) communication and small and medium sized enterprises: The governmentality dilemma of explicit and implicit CSR communication. Human Relations, 72(12): 1920–1947

Navigli R, Conia S, Ross B (2023). Biases in Large Language Models:

Origins, inventory, and discussion. Journal of Data and Information Quality, 15(2): 10:1–10:21

Olander S, Landin A (2008). A comparative study of factors affecting the external stakeholder management process. Construction Management and Economics, 26(6): 553–561

Oyewole M O, Komolafe M O, Gbadegesin J T (2023). Understanding stakeholders' opinion and willingness on the adoption of sustainable residential property features in a developing property market. International Journal of Construction Management, 23(2): 358–370

Pang B, Lee L (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1–2): 1–135

Perlaviciute G, Steg L, Contzen N, Roeser S, Huijts N (2018). Emotional responses to energy projects: Insights for responsible decision making in a sustainable energy transition. Sustainability, 10(7): 2526

Praveena S, Aris A (2021). The impacts of COVID-19 on the environmental sustainability: A perspective from the southeast asian region. Environmental Science and Pollution Research International, 28(45): 63829–63836

Project Management Institute (2022). Guide to the Project Management Body of Knowledge (PMBOK guide) and the Standard for project management (7th ed edition). Project Management Institute

Raineri N, Paillé P (2016). Linking corporate policy and supervisory support with environmental citizenship behaviors: The role of employee environmental beliefs and commitment. Journal of Business Ethics, 137(1): 129–148

Reed M (2008). Stakeholder participation for environmental management: A literature review. Biological Conservation, 141(10): 2417–2431

Roy R B, Rokonuzzaman Md, Amin N, Mishu M K, Alahakoon S, Rahman S, Mithulananthan N, Rahman K S, Shakeri M, Pasupuleti J (2021). A Comparative Performance Analysis of ANN Algorithms for MPPT Energy Harvesting in Solar PV System. IEEE Access, 9: 102137–102152

Smail E (2022). Australian Conservation Foundation Launches Legal Action Against Woodside. Climate News Australia. Available at the website of climatenewsaustralia.com/australian-conservation-foundation-launches-legal-action-against-woodside/

Swalih M M, Ram R, Tew E (2024). Environmental management accounting for strategic decision-making: A systematic literature review. Business Strategy and the Environment, 33(7): 6335–6367

Taylor K M, Rosca E (2024). Toward a moral approach to stakeholder management: Insights from the inclusion of marginalized stakeholders in the operations of social enterprises. International Journal of Operations & Production Management. 44 (10): 1831–1858

Eemeren F H van Grootendorst R (2003). A Systematic Theory of Argumentation: The Pragma-dialectical Approach. Cambridge University Press

Vayansky I, Kumar S A P (2020). A review of topic modeling methods. Information Systems, 94: 101582

Vining J (1987). Environmental decisions: The interaction of emotions, information, and decision context. Journal of Environmental Psychology, 7(1): 13–30

Wallach H M (2006). Topic modeling: Beyond bag-of-words. In: Proceedings of the 23rd International Conference on Machine Learning, 977–984

Xue J, Shen G Q, Li Y, Wang J, Zafar I (2020). Dynamic stakeholder-associated topic modeling on public concerns in megainfrastructure projects: Case of Hong Kong–Zhuhai–Macao Bridge. Journal of Management in Engineering, 36(6): 04020078