Topic and Tone in YouTube Climate Discussions

Vered Aharonson[1,2] and Or Bovan[3]
[1]University of the Witwatersrand, Johannesburg, South Africa
[2]Medical School, University of Nicosia, Nicosia, Cyprus
[3]School of Software Engineering, Afeka, Tel Aviv Academic College of Engineering, Tel Aviv, Israel

Vered.aharonson@wits.ac.za
orb5@mail.afeka.ac.il

Abstract: Comments are routinely used to measure engagement on social media and to infer the effectiveness of content. This matters particularly for content with educational and social value, such as environmental sustainability. Automated analyses predominantly examine the sentiment expressed in text, which can be unrelated to a video's message and misread engagement. We examine comments on YouTube climate-change videos using a fully automated process. Off-the-shelf natural language processing tools assign provisional labels for topic relevance (on-topic or off-topic) and for sentiment polarity (positive, neutral, or negative). Comment texts are represented by word patterns with per-video weighting adjusted to mitigate bias introduced by videos with disproportionately high comment volumes. A simple machine learning model, logistic regression, then predicts topic and sentiment. We report recognition accuracy on held-out comments and analyse the sentiment distribution within on-topic and off-topic comments. The analysis is performed separately for top-level comments that respond to the video and for replies that react to other users' comments. For top-level comments, recognition accuracies were 0.93 for topic and 0.90 for sentiment. For replies, topic accuracy was 0.94, and sentiment accuracy was 0.88. On-topic comments showed high frequencies of both positive and negative sentiments, while off-topic comments were predominantly neutral. Among the replies, on-topic neutral sentiment was less common and positive sentiment was more common than in the top-level comments. These findings show that a fully automated, topic- and tone-aware process can reliably extract engagement patterns from YouTube comments, provided that tone (sentiment) analysis is paired with topic verification and that top-level comments and replies are analysed separately. A disproportionate number of neutral tones in off-topic threads shows that analyses that ignore the topic underestimate engagement. In on-topic comments, replies are less neutral, showing a more constructive conversation. The methodology is general and can be adapted to other domains where distinguishing genuine topical engagement from background noise is important. Ultimately, this insight can guide content creators, educators, and policymakers in evaluating and fostering meaningful online discussions around critical topics.

Keywords: Automated Social Media Analysis, Topic Recognition, Sentiment Analysis, User Engagement, Climate Communication

1. Introduction

YouTube has become a prevalent platform for public discourse on climate change. Video comment sections reflect viewers' reactions and opinions, and provide a window into their engagement with the video's content. Measuring engagement in social media is challenging, however, and is often reduced to crude metrics like view counts or "likes." More sophisticated approaches use natural language processing to analyse comment sentiment as a proxy for engagement (Liu 2012), a technique that has become well established (Giachanou and Crestani 2016; Aharonson and Joselowitz 2024).

A key limitation of standard sentiment-based analyses is that they treat all comments as relevant to the video. Sentiment alone can be misleading if the comments do not respond to the video's climate message. Users may post generic praise, spam, jokes, or personal remarks (Joshi et al. 2017) that express emotion but have no bearing on climate change. For example, two comments posted under a climate-awareness video are: (1) "I love you all my sister's btw ireally miss concert ❤️" and (2) "It's sad that those chem-trails are in the back." The first comment is emotional (with a positive tone) but clearly off-topic—the user talks about siblings and a concert with garbled language, unrelated to climate. The second comment directly mentions an alleged climate-related phenomenon ("chem-trails") and is on-topic, albeit expressing a negative sentiment. An automated sentiment analyser might flag the first comment as positive engagement with the video and the second as negative. In contrast, a human reader knows the first is an irrelevant praise and that the second is an on-topic reaction. Distinguishing such cases manually at scale is, however, is infeasible—YouTube comment sections include millions of comments. Moreover, this task remains difficult for automated methods due to informal language and context nuances. Recent language models still struggle to understand colloquialisms, concatenated words (e.g., "ireally"), and implicit context, which can confound sentiment predictions.

In the context of educational or persuasive content, such as climate change awareness videos, an emotionally charged off-topic comment does not represent genuine *engagement* with the message. Including off-topic comments, spam or personal chatter in sentiment analysis dilutes the signal of genuine engagement. An abundance of neutral off-topic comments could misleadingly suggest viewer apathy or balanced sentiment, when in reality, engaged viewers might be strongly positive or negative about the content. Thus, to reliably gauge engagement, one must verify that the tone of the comments reflects the topic at hand. Prior research has highlighted the importance of filtering on-topic comments to isolate sentiment toward specific subjects or aspects within text (Schouten and Frasincar 2016), but no large-scale application of this concept has been attempted.

In addition, the structure of conversation threads on YouTube may influence tone. Top-level comments are direct reactions to the video, whereas reply comments form discussions between users. These two levels of interaction could exhibit different engagement dynamics. In example, reply threads might become more emotional or, conversely, more supportive and nuanced as users build on each other's comments. Ignoring the distinction between top-level and reply comments might obscure these patterns (Iordanou et al. 2022).

From a communication and education perspective, engagement with mediated content is increasingly understood as a multidimensional construct that extends beyond surface-level affective reactions. Prior work in educational communication has shown that vocal and linguistic cues can significantly influence perceived credibility, charisma, and learner engagement, even when content is delivered asynchronously or via digital media (Katz-Navon et al. 2025). Crucially, this line of research highlights that emotional signals are meaningful only insofar as they are interpretable in relation to the instructional or persuasive intent of the message. Translating this insight to social-media discourse, sentiment expressed in user comments constitutes evidence of engagement only when it is demonstrably anchored to the topic under discussion.

To address these gaps, this paper introduces a fully automated topic-and tone-aware analysis for YouTube comments. Our approach uses a lightweight text classification model to predict those labels from comment text, incorporating a weighting scheme to balance contributions from videos with disparate popularity. Off-the-shelf natural language processing tools are used once-off to assign topic and sentiment labels for the model's training and evaluation. We apply this method to a large dataset of YouTube comments on climate change videos, separating top-level comments and replies to other comments. This work extends our prior study on climate-change video engagement (Aharonson and Joselowitz 2024), which used comment sentiment to infer video persuasiveness. In that study, sentiment trends were analysed without verifying the topic relevance of each comment. We hypothesise that the present topic-filtered approach, together with the examination of conversational dynamics in replies, can yield a clearer, more accurate picture of audience engagement with climate content.

2. Methodology

2.1 Data Collection and Preprocessing

We compiled a large dataset of YouTube comments from climate-change-related videos. Using the YouTube Data API (Google 2025a; Google 2025b) and an open-source Python script (Bouman n.d.), we scraped approximately 1.6 million comments from English-language YouTube channels focused on climate change and environmental topics. Both top-level comments - those directly responding to the video - and reply comments - those responding to other users - were collected, with metadata linking each comment to its video. Data privacy was respected by not retaining user identities. Basic preprocessing steps were applied to the comment text: conversion to lowercase, removal of non-alphanumeric characters (such as emojis and URLs), and tokenisation (breaking the text into words - "tokens"). Common stopwords were removed, and lemmatisation (converting words to their base, dictionary form) was applied. After preprocessing, each comment is represented as a minimal set of words (see Table 1) that captures its essential content. This cleaned, normalised text was used for subsequent analysis.

2.2 Feature Representation

Each preprocessed comment was represented as a feature vector. We adopted a bag-of-words representation using individual words as features (Joulin et al. 2017),where each unique word in the text defines a dimension in the feature vector. To avoid an excessively large and sparse feature space, we limited the vocabulary to the most informative terms. We introduced a weighting scheme inspired by the Term Frequency times Document Frequency (TF-DF) scheme (Rathi and Mustafi 2023). This approach down-weights terms that are very common in high-comment-volume videos, under the assumption that such terms (for example, a specific video presenter's name or a catchphrase heavily repeated in one video's comment section) are less useful for generalisable classification. Very rare terms (occurring in only a few comments) were discarded. We determined an appropriate vocabulary size by experimenting (balancing performance and sparsity). Following the weighting and normalisation, we obtained the final feature representation for classification.

2.3 Classification

2.3.1 Model

We trained a simple, explainable classifier to predict each comment's topic label (on-topic/off-topic) and sentiment label (positive/neutral/negative) from its feature vector (section 2.2). We chose a logistic regression (LR) model for both tasks. This simple linear classifier has been found to perform competitively in text classification tasks when combined with appropriate feature representations. In particular, its accuracy performance compared favourably with more complex models for tasks such as text sentiment or topic analysis, while remaining efficient and interpretable (Wang and Manning 2012; Joulin et al. 2017). Two separate logistic regression classifiers were trained: one for topic (binary classification) and one for sentiment (tri-class classification). We implemented the classifiers using the open-source scikit-learn library (Pedregosa et al. 2011), utilising the LIBLINEAR solver for efficiency (Fan et al. 2008). Model hyperparameters, such as the strength of regularisation, were tuned based on validation performance to prevent overfitting, given the high dimensionality of the feature space.

2.3.2 Labels and Training

The training of a classifier needs ground truth indication, or *labels*. In our context, these are the topics and sentiments of the comments. As manual labelling by humans ('gold-standard labels') is not scalable for this amount of comments, we applied machine-generated approximations of ground truth ('silver-standard labels'). Each comment was labelled with a topic relevance tag and a sentiment tag. For topic labelling, we used the ClimateBERT classifier (Webersinke et al. 2022) to determine whether each comment's text is on-topic (i.e., it pertains to climate change or the content of the video) or off-topic. ClimateBERT is a publicly available language model pre-trained on climate change discourse, which makes it adept at recognising climate-related terminology and context. We applied the model using the default parameters from Webersinke et al. 2022. If the model's confidence that the comment is climate-relevant exceeded this default threshold, the comment was labelled *on-topic*. Otherwise, it was labelled *off-topic*. For sentiment labelling, we employed the VADER sentiment analyser (Hutto and Gilbert 2014). VADER is a lexicon- and rule-based tool designed for social media text, known for its effectiveness in detecting positive, neutral, and negative sentiment in short informal messages. VADER provides a sentiment score for each comment. Based on the score, we assigned each comment one of three sentiment labels: *positive*, *neutral*, or *negative*. As in the topic labelling, VADER's recommended thresholds were used: scores above 0.05 for positive, below -0.05 for negative, and in-between for neutral labels (Hutto and Gilbert 2014). This automated sentiment labelling captures the general tone or emotional polarity of each comment. The labels from ClimateBERT and VADER were treated as the ground truth for the LR model training and evaluation. We split the labelled dataset into training and test sets. We reserved 20% of the comments as a held-out test set to evaluate the model's generalisation, and trained on the remaining 80%.

2.3.3 Evaluation

We report performance using the macro-F1 score, which is robust to class imbalance. The macro-F1 metric computes the harmonic mean of precision and recall (F1) for each class independently and then averages them, giving equal weight to each class. This provides a balanced evaluation of how well the classifier performs across the three sentiment categories and the two topic categories, without being dominated by the majority classes. We compute macro-F1 for the sentiment task (averaged over positive, neutral, and negative) and for the topic task (averaged over on-topic and off-topic). All reported results are on the held-out test set, using the automated labels as the ground truth. We read a small, randomly selected sample of 50 comments to get a sense of their content.

3. Results

The collected dataset (section 2.1) comprises roughly 1.6 million YouTube comments from climate-related videos. Of these, 56.6% (about 0.9 million) are top-level comments directed at the video, and 43.4% (about 0.7 million) are reply comments in threaded discussions.

Table 1 provides three examples of raw comments and the results of their preprocessing (section 2.1). The table illustrates that informal and emphatic language (e.g. all-caps "NEED", repeated punctuation "!!!", emoticons) is cleaned in this process, yielding a simplified representation. This opertion puts the focus on content words (e.g. “need act”, “informative”, “complete nonsense”) while discarding stylistic noise, which can help the automated analysis.

Table 1: Examples of raw comments and their processed (normalized) text.

Raw Comment	Processed Comment
“We NEED to act on this!!!”	*need act*
“Great video, very informative :)”	*great video informative*
“This is complete nonsense.”	*complete nonsense*

The initial vocabulary used to feature the topic and tone classification (section 2.2) included several tens of thousands of words. Following weighting and normalisation, the final feature representation for classification included 3000 words.

The automated tone and topic labelling (section 2.3.2) yielded the distribution of positive/neutral/negative depicted in Figure 1a, and the frequencies of on-topic vs. off-topic comments shown in Figure 1b. Most comments were identified as not relevant to the climate topic. Positive sentiment appears to be the most common label overall, followed by negative, then neutral. These observations apply when all comments are considered together. However, as we will see, this overall distribution masks important differences between on-topic and off-topic comment subsets.

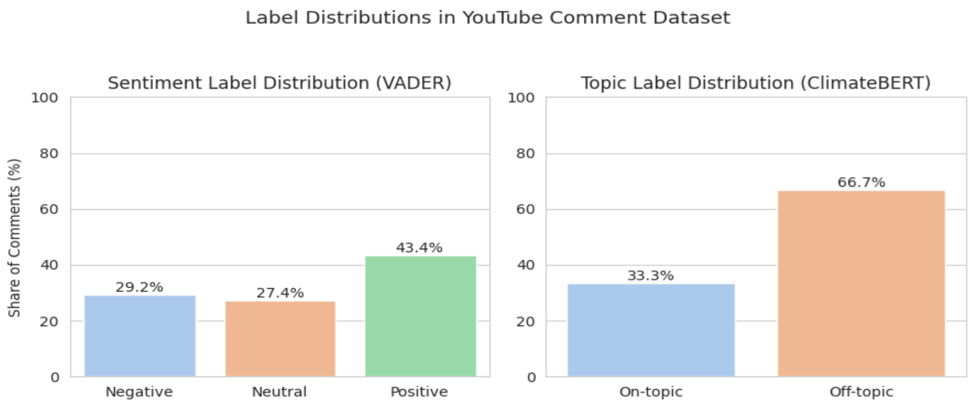

Figure 1: Distribution of sentiment and topic labels in the dataset. (a) percentage of comments labelled as positive, neutral, and negative). (b) percentage of comments labelled as on-topic and off-topic.

The topic LR classifier (section 2.4) achieved a macro-F1 score of 0.93 in distinguishing climate-related comments from off-topic comments. The sentiment LR classifier yielded a macro-F1 score of 0.90 in classifying comments into positive, neutral, or negative tone. Performance on reply comments was similar: the topic classifier yielded a macro-F1 score of 0.94 on the replies, and the sentiment classifier’s macro-F1 score on replies was 0.88.

Figure 2 illustrates the sentiment breakdown for on-topic vs. off-topic comments, separated into top-level and reply categories. The distribution of sentiment labels differs between on-topic and off-topic comments. Among comments labelled on-topic, we observe high frequencies of both positive and negative sentiments. In contrast, comments labelled off-topic were predominantly neutral in tone. Sporadic reading of the comments yielded that indeed, on-topic comments actively express opinions or emotions about the climate content—viewers praising the video or the cause (positive sentiment) or expressing frustration, fear, or disagreement (negative sentiment) regarding climate change issues. Many off-topic comments are simple statements or unrelated discussions that neither endorse nor criticise the video's message (e.g., users tagging friends, making jokes, or talking about an unrelated subject).

Both top-level comments and replies show similar polarity distribution pattern. On-topic comments, however, have a higher incidence of both positive and negative sentiment, whereas off-topic comments are largely neutral. The on-topic replies contain relatively more positive sentiment and less neutrality compared to on-topic top-level comments.

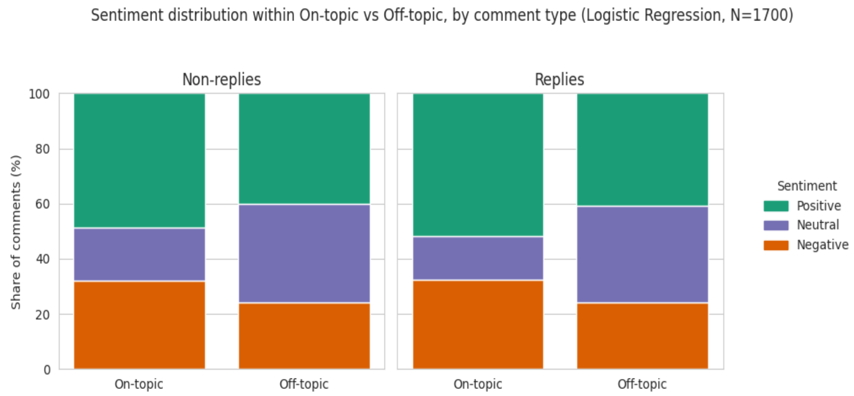

Figure 2: Sentiment distribution for on-topic vs. off-topic comments, for Top-level comments only and for Reply comments only. Each bar shows the percentage of comments with positive, neutral, or negative sentiment in the on-topic and off-topic subsets.

4. Discussion

The results highlight the importance of incorporating topic relevance when analysing social media comments for engagement insights. Our findings illustrate that without topic filtering, a large fraction of comments appeared neutral, potentially leading analysts to underestimate audience engagement or interest. By first separating out irrelevant (off-topic) comments, we obtain a clearer picture: most viewers who comment *about* the climate topic do so with a strong sentiment (either positive or negative), indicating true engagement, whether in support or dissent. Off-topic comments, in contrast, carry little meaningful sentiment about the video and thus inflate the neutral category. This confirms that emotional tone can only be reliably used when we ensure the tone is directed at the intended topic. In essence, topic filtering is similar to the aspect-specific sentiment analysis (Schouten and Frasincar 2016) in that it focuses on sentiment relevant to the video's subject matter. This targeted approach allows content creators and researchers to distinguish between genuine responses to the content and background chatter.

From a theoretical standpoint, these findings demonstrate how engagement-oriented communication principles can be operationalised in large-scale automated analysis (Katz-Navon et al. 2025). By explicitly separating topic relevance from emotional tone, the present study translates established insights from engagement and educational communication theory into a computational framework suitable for social-media research. This distinction allows sentiment to be interpreted as a meaningful indicator of engagement only when it is directed at the communicative intent of the content, rather than at peripheral or unrelated interaction.

Another insight from our study is the difference in engagement between one-off comments and threaded discussions. Top-level comments and reply threads serve different roles in the discourse, and our analysis suggests that replies tend to foster a more constructive or positive tone on the climate topic. The on-topic replies contained a higher proportion of positive sentiment compared to on-topic top-level comments. This phenomenon may reflect community-building behaviour: users often reply to add information, show support, or politely debate, leading to collaborative problem-solving or encouragement in the replies (Iordanou et al. 2022). In contrast, top-level comments can include drive-by negative reactions or off-topic tangents that are less likely to prompt continued discussion. From an engagement perspective, this suggests that conversations (as seen in reply chains) may be where more productive, solution-oriented exchanges occur, even on a contentious topic like climate change. It highlights the value of not only measuring how many people react to a video, but also examining *how* they interact with each other in follow-up discussions.

Our methodology employed a relatively simple automated analysis that can yield robust results when appropriately designed. Powerful, advanced deep-learning models like ClimateBERT tend to be resourceintensive and function as ‘black boxes’. They require substantial computational power (hence more energy and even water for cooling), which is arguably at odds with the goal of environmental sustainability. Our simple, explainable logistic regression classifiers reliably predicted each comment’s topic label (on-topic/off-topic) and sentiment label (positive/neutral/negative) from its feature vector (section 2.3). Both classifications achieved high accuracy in replicating the output of more complex models. Notably, this performance required careful feature design, TF×DF weighting, vocabulary curation, and a large volume of training data. Our results thus suggest that for tasks of this nature, simple linear models can be surprisingly effective, corroborating observations from prior work on text classification baselines (Joulin et al. 2017; Wang and Manning 2012). Simplicity offers advantages: the models are computationally efficient and interpretable. One can inspect feature weights to understand which words are most indicative of off-topic content or of positive vs. negative sentiment in the context of climate discussions. This transparency could be useful for moderators or researchers seeking to diagnose the drivers of particular sentiments (for example, which keywords commonly trigger negative reactions).

There are several limitations to our current study. First, automated labelling of comments allows us to scale to millions of comments, but it may introduce label noise, i.e., misclassifications by the tools. We mitigated potential noise by using high-quality, domain-specific models and by focusing on broad categories (topic on/off, three-level sentiment) where automated tools are reasonably reliable. Even state-of-the-art tools for the domain, however, are not perfect. The ClimateBERT model might misclassify some comments that use sarcasm or indirect references, and the VADER analyser might misread complex sentiment, especially in the presence of humour or sarcasm. For instance, a sarcastic comment like "Great, another tax to save the planet... *sigh*" could be tagged as positive by a naïve lexicon-based algorithm due to the word "great," when in reality it is negative. Our analysis does not explicitly detect sarcasm or irony, which remains an open challenge in sentiment analysis (Joshi et al. 2017). Additionally, comments could be borderline on-topic or off-topic, and without human adjudication, we rely on the classifier’s threshold. In future work, we intend to compile a human-annotated sample of comments to audit the accuracy of the automated labels and adjust the system accordingly. This would provide a measure of confidence in the reported engagement patterns and help calibrate any biases in the initial labelling.

Another limitation is the modelling of reply context. Our current approach treats each comment in isolation, but a reply often draws meaning from the comment it responds to. Incorporating conversational context (for example, using the parent comment’s content as additional input to the sentiment classifier) could refine sentiment detection—especially for sarcasm or for understanding whether a reply’s sentiment is directed at the video or at another user. Thread-level analysis might also reveal how sentiments evolve in back-and-forth discussions (e.g., does a negative comment trigger correcting replies that are positive in tone?). Integrating context could be achieved by more advanced models, such as sequence-based deep learning or transformer models, which can handle multi-sentence inputs.

Our methodology can be generalised beyond YouTube climate change videos. To apply it in another domain (i.e., public health campaigns or educational content on any topic), one would substitute an appropriate topic classifier for the domain of interest (similar to how we used ClimateBERT for climate content). The sentiment analysis component (VADER or a suitable alternative) can remain largely the same, or be fine-tuned if domain-specific slang or tone differs significantly. The overall framework thus provides a richer understanding of user engagement and advances evidence-based evaluation of social media impact for content creators, educators, and policymakers.

5. Conclusions

This paper present a novel, fully automated process to analyse the topic and tone of YouTube comments in the domain of climate change communication. The analysis shows that on-topic discussion around climate videos is characterised by strong emotions, both positive and negative, and that focusing on these on-topic comments is crucial for accurately assessing a video’s impact. Off-topic commentary, if unfiltered, can mask a proper engagement and skew interpretations toward neutrality. Differentiating between initial comments and reply threads uncovered that interactive discussions tend to adopt a more constructive tone, a promising sign for community building in climate communication. These insights can inform how social media content is evaluated and moderated. The approach demonstrates a scalable way to sift signal from noise in online discourse. Its potential to clarify users' engagement patterns can guide more effective communication strategies and foster meaningful conversations.

Ethics Declaration

This study analysed publicly available YouTube comments using automated tools and did not involve human subjects or private data. Therefore, no formal ethical approval was required.

AI Declaration

The authors used ChatGPT (an AI language model) for minor editing of the manuscript.

References

Aharonson, V. and Joselowitz, J. (2024) "On Presenters and Commenters in YouTube Climate Change Videos", in *Proceedings of the 11th European Conference on Social Media (ECSM)*, pp. 9–12. doi:10.34190/ecsm.11.1.2076.

Bouman, E. (n.d.) *youtube-comment-downloader* (Python package), version 0.1.76. [Software]. Available at: https://pypi.org/project/youtube-comment-downloader/ (Accessed: 2025-10-01).

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. and Lin, C.-J. (2008) "LIBLINEAR: A Library for Large Linear Classification", *Journal of Machine Learning Research*, 9, pp. 1871–1874.

Giachanou, A. and Crestani, F. (2016) "Like It or Not: A Survey of Twitter Sentiment Analysis Methods", *ACM Computing Surveys*, 49(2), pp. 1–41.

Google (2025a) "YouTube Data API v3: Comments: list". Google Developers Documentation. (Online) Available at: https://developers.google.com/youtube/v3/docs/comments/list (Accessed: 2025-10-01).

Google (2025b) "YouTube Data API v3: CommentThreads: list". Google Developers Documentation. (Online) Available at: https://developers.google.com/youtube/v3/docs/commentThreads/list (Accessed: 2025-10-01).

Hutto, C. J. and Gilbert, E. (2014) "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text", in *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 8(1), pp. 216–225. doi:10.1609/icwsm.v8i1.14550.

Iordanou, K., Aharonson, V., Christodoulou, V., Karpasitis, C., Joselowitz, J., Lilford, B., and Muraleedharan, S. (2022). Collaborative Learning in YouTube: Under Which Conditions Can Learning Happen or Fail to Happen?. In Proceedings of the 15th International Conference on Computer-Supported Collaborative Learning-CSCL 2022, pp. 569-570. International Society of the Learning Sciences.

Joshi, A., Bhattacharyya, P. and Carman, M. J. (2017) "Automatic Sarcasm Detection: A Survey", *ACM Computing Surveys*, 50(5), pp. 1–22. doi:10.1145/3124420.

Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T. (2017) "Bag of Tricks for Efficient Text Classification", in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2, pp. 427–431.

Katz-Navon, T., Aharonson, V. and Malachi, A. (2025) "The charismatic lecturer's voice: explainable machine learning models", International Journal of Educational Methodology, 11(4), pp. 479–493.

Liu, B. (2012) *Sentiment Analysis and Opinion Mining*. Springer Nature. https://doi.org/10.1007/978-3-031-02145-9.

Pedregosa, F. et al. (2011) "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Rathi, R.N., Mustafi, A. (2023) "The importance of Term Weighting in semantic understanding of text: A review of techniques", *Multimed Tools Appl,* 82, pp. 9761–9783. https://doi.org/10.1007/s11042-022-12538-3

Schouten, K. and Frasincar, F. (2016) "Survey on Aspect-Level Sentiment Analysis", *IEEE Transactions on Knowledge and Data Engineering*, 28(3), pp. 813–830. doi:10.1109/TKDE.2015.2485209.

Wang, S. and Manning, C. D. (2012) "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification", in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2, pp. 90–94.

Webersinke, N., Kraus, M., Bingler, J. A. and Leippold, M. (2022) "ClimateBERT: A Pretrained Language Model for Climate-Related Text", arXiv preprint [cs.CL]. doi:10.48550/arXiv.2110.12010.