

MultiClimate: Multimodal Stance Detection on Climate Change Videos

Jiawen Wang^{◇,*} Longfei Zuo^{◇,*} Siyao Peng^{◇,†} Barbara Plank^{◇,†}

[◇]Center for Information and Language Processing, LMU Munich, Germany

[†]MaiNLP & MCML, LMU Munich, Germany

{jiawen.wang, zuo.longfei}@campus.lmu.de {siyao.peng, b.plank}@lmu.de

Abstract

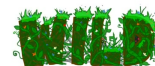
Climate change (CC) has attracted increasing attention in NLP in recent years. However, detecting the stance on CC in multimodal data is understudied and remains challenging due to a lack of reliable datasets. To improve the understanding of public opinions and communication strategies, this paper presents MultiClimate, the first open-source manually-annotated stance detection dataset with 100 CC-related YouTube videos and 4, 209 frame-transcript pairs. We deploy state-of-the-art vision and language models, as well as multimodal models for MultiClimate stance detection. Results show that text-only BERT significantly outperforms image-only ResNet50 and ViT. Combining both modalities achieves state-of-the-art, 0.747/0.749 in accuracy/F1. Our 100M-sized fusion models also beat CLIP and BLIP, as well as the much larger 9B-sized multimodal IDEFICS and text-only Llama3 and Gemma2, indicating that multimodal stance detection remains challenging for large language models. Our code, dataset, as well as supplementary materials, are available at <https://github.com/weryjw/MultiClimate>.

1 Introduction

As climate change (CC) gains global attention, measuring human stance towards CC becomes increasingly important. Numerous large language models (LLMs) and deep learning models have been developed to address these challenges. These models can help detect public opinions and assist stakeholders to improve decision-making, thus providing valuable insights into public perception regarding climate change (OpenAI, 2024; Meta AI, 2024; Dosovitskiy et al., 2021; Alayrac et al., 2022).

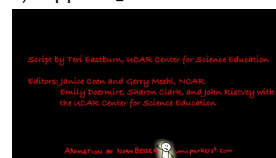
Stance detection is a task to determine whether authors of a document support, oppose, or take a neutral stance toward a specific target (Mohammad et al., 2016; Hardalov et al., 2022; Weinzierl and

*Equal contributions.



(a) Anytime you hear the word [0:05, left, Neutral] wild you can bet it's referring to something uncontrollable and unrestrained [0:10, right, Oppose].

ATTRIBUTION RESEARCH



(b) But sit tight attribution research may crack the nut and science researchers in this emerging field are [1:55, left, Support] on the case [2:00, right, Neutral].

Figure 1: MultiClimate sample annotations with aligned video frames and transcript sentences.

Harabagiu, 2023b). It enhances information management by efficiently categorizing diverse opinions. Identifying varying public opinions helps promote societal understanding and communication, thus reducing conflicts and enhancing public discourse. Previous research explored stance detection on climate change (Maynard and Bontcheva, 2015; Vaid et al., 2022; Upadhyaya et al., 2023a,b), but focusing on text-only data.

Images and videos can shape public awareness of climate change by vividly illustrating its social impacts. Recently, multimodality gained significant traction for connecting CC to discourse (Weinzierl and Harabagiu, 2023b), as images and videos largely impact CC perceptions (Wang and Liu, 2024). Investigating stance through multimodality also becomes essential. However, due to the lack of datasets, developing multimodal models for stance detection on climate change remains challenging, as previous studies mainly focused on language texts rather than the visual modality.

To fill this gap, this paper proposes the first open-sourced dataset, MultiClimate, that integrates both

visual and textual modalities, specifically 4,209 image frames and transcripts, to label stances in 100 CC-related YouTube videos (§3). §4 evaluates text-only, image-only, and multimodal models on MultiClimate stance detection. §5 concludes the paper and proposes future directions.

Our results show that the text-only BERT model outperforms image-only models on multimodal stance detection, and the best performance is achieved by fusing models from both modalities. We further experiment with 9B-sized large language and multimodal models and illustrate that these larger models deliver unsatisfactory zero-shot results, much lower than our state-of-the-art (SOTA) fusion models. Fine-tuning a large multimodal model brings about some improvements, but it is resource-heavy and encounters the Green NLP problems.

2 Related Work

Multimodal Stance Detection. Stance detection has mainly concentrated on textual analysis (Küçük and Can, 2021a; Lan et al., 2024), with a significant focus on the stance expressed in social media platforms like Twitter (Taulé et al., 2018; Conforti et al., 2020). Yet, a recent trend arose that gradually includes images and videos in stance detection (Küçük and Can, 2021b; Carnot et al., 2023).

Current multimodal stance detection datasets emphasize different aspects such as communication frames and trending topics like COVID-19 (Taulé et al., 2018; Weinzierl and Harabagiu, 2023a; Liang et al., 2024). These data rely solely on static images or extract the first frame from a video or GIF for visual input. Despite these advancements, research on stance detection using visual input remains limited. Our MultiClimate dataset addresses climate change topics by utilizing frames from full videos as visual input and transcripts as text, allowing for a more comprehensive information coverage.

Stance Detection on Climate Change. Stance detection on CC (Fraile-Hernandez and Peñas, 2024) aims at determining whether a given document expresses a supportive, opposing, or neutral attitude toward whether CC is a real concern. Recent CC stance detection studies particularly focused on social media texts (Vaid et al., 2022), especially Twitter (Conforti et al., 2020; Upadhyaya et al., 2023a).

Regarding modeling, for instance, Vaid et al. (2022) offered Fast-Text (Bojanowski et al., 2017)

	Videos	Frame-Transcript Pairs			Total
		SUPPORT	NEUTRAL	OPPOSE	
Train	80	1,449	1,036	887	3,372
Dev	10	204	83	130	417
Test	10	194	73	153	420
Total	100	1,847	1,192	1,170	4,209

Table 1: Overview of MultiClimate statistics.

and BERT (Devlin et al., 2019) variants for stance tasks. Upadhyaya et al. (2023a) proposed a framework that utilizes emotion recognition and intensity prediction to discern different attitudes in tweets about climate change. CC stance detection is also applied to downstream tasks such as fake news detection (Mazid and Zarnaz, 2022). However, none of the models above handles multimodal inputs.

3 The MultiClimate Dataset

We propose a new open-source MultiClimate dataset, the first climate change stance detection corpus on multimodal data. To the best of our knowledge, Mendeley provides a closest dataset to ours (Aharonson, 2023). However, it only provides a simple Excel file with 168 YouTube links and some basic statistics on each video.¹ Most of these videos are not under the Creative Commons license and none is annotated with stance labels.

Our MultiClimate dataset consists of 100 CC-related YouTube videos in English with 4,209 frame-transcript pairs. To ensure a balanced distribution, we partition MultiClimate into 80% train, 10% dev, and 10% test sets, not only in terms of the number of videos but also the number of frame-transcript pairs. Table 1 provides statistics on MultiClimate frame-transcript pairs. For a complete list of 100 YouTube videos, see Appendix F.

Video Collection. 100 YouTube videos are collected by querying “climate change” and filtering the results by Creative Commons license. The videos are downloaded using the pytube² and yt-dlp³ APIs, while the transcripts are extracted using the youtube-transcript-api.⁴

Frame Extraction and Alignment. To effectively use our video data, we extract the initial

¹<https://data.mendeley.com/datasets/j955mxnyyf/1>

²<https://pytube.io/en/latest/>

³<https://github.com/yt-dlp/yt-dlp>

⁴<https://pypi.org/project/youtube-transcript-api/>

frame from each 5-second segment using FFmpeg (Tomar, 2006). We then manually align each frame to its corresponding full sentence in the transcripts to form a frame-transcript pair. Longer sentences can be aligned with multiple time-stamped frames,⁵ and we allow different stance labels for individual frame-transcript pairs, as shown in Figure 1.

Stance Annotation. The first two authors of this paper manually annotated the stance expressed in each frame-transcript pair for 100 videos. We use trinary stance labels: SUPPORT, NEUTRAL, and OPPOSE. Appendix A details annotation guidelines, particularly label definitions, and examples.

MultiClimate includes 1,847 SUPPORT, 1,170 OPPOSE, and 1,192 NEUTRAL frame-transcript pairs (Table 1). To assess the effectiveness of our annotation guideline and the quality of our dataset, all 10 videos in the test partition are double annotated. Inter-annotator agreement (IAA) between the two annotators achieves 0.703 in Cohen’s kappa, 0.826 in accuracy, and 0.823 in weighted F1; see Appendix B for IAA on 10 individual test videos. Additionally, we aggregate stance labels from each video’s frame-transcript pairs by majority voting for future video-level analyses and experiments; Appendix F includes these video stance labels.

4 MultiClimate Stance Detection

This section evaluates SOTA text-only, image-only, and multimodal models on MultiClimate stance detection. We use both accuracy and weighted F1 scores since SUPPORT/NEUTRAL/OPPOSE labels are unbalanced in the dataset. The models are run on the CPU, Google T4 GPU, and NVIDIA GeForce RTX 2080.

4.1 Models

We conduct a comprehensive performance evaluation on several text-only, image-only, text-image-fusion, and multimodal models. We leverage text-only BERT variants (Devlin et al., 2019) as earlier work showed their superior performance on tweet stance detection (Weinzierl and Harabagiu, 2023a; Vaid et al., 2022). Large language models (LLMs) are also included given their promising zero-shot classification performance (Dubey et al., 2024).

⁵MultiClimate has 1,902 transcripts aligned to 1 frame, 656 to 2 frames, 193 to 3, 27 to 4, 8 to 5, 3 to 6, 2 to 7, excluding [Music] or None transcripts. Details are provided in the repository <https://github.com/werywjuw/MultiClimate/tree/main/notebooks>.

Meanwhile, for image recognition and analysis, ResNet50 (He et al., 2016) and ViT (Dosovitskiy et al., 2021) are capable of understanding and interpreting complex image data. By combining aforementioned textual and visual models, we also deploy fusion models to investigate whether both modalities are essential for multimodal stance detection. Moreover, models that are trained on cross-modal representations, CLIP (Radford et al., 2021), BLIP (Li et al., 2022), and IDEFICS (Alayrac et al., 2022), are also compared with the aforementioned unimodal and fusion models.

Text-only Models. We use BERT (Bidirectional Encoder Representations from Transformers, *bert-base-cased*, Devlin et al. 2019) for our textual stance detection given its effectiveness. Newly released LLMs, Llama3 (*meta-llama/Meta-Llama-3-8B*, Dubey et al. 2024; Meta AI 2024) and Gemma2-9B (*google/gemma-2-9b*, Clement and Warkentin 2024) are also evaluated on the Ol-lama (Morgan and Michael, 2024) platform by giving the following zero-shot prompt:

```
Classify the stance of the following text towards climate change as: 0 (NEUTRAL), 1 (SUPPORT), 2 (OPPOSE): {transcript}.
```

Image-only Models. We also deploy two state-of-the-art image-only models, ResNet50 (Residual Network, *microsoft/resnet-50*, He et al. 2016) and ViT (Vision Transformer, *google/vit-base-patch16-224*, Dosovitskiy et al. 2021).

Multimodal Models. Our multimodal fusion models are built by concatenating BERT (Devlin et al., 2019) with ViT (Dosovitskiy et al., 2021) or ResNet50 (He et al., 2016) embeddings, as they are the smaller 100M-sized models (see Table 2).

CLIP (Contrastive Language-Image Pre-training, *openai/clip-vit-base-patch32*, Radford et al. 2021) and BLIP (Bootstrapping Language-Image Pre-training, *Salesforce/blip-image-captioning-base*, Li et al. 2022) are leveraged to associate images and text simultaneously, capturing richer, more nuanced information. We also experiment with IDEFICS (Image-aware Decoder Enhanced à la Flamingo with Interleaved Cross-attentionS, *HuggingFaceM4/idefics-9b*, Alayrac et al. 2022),⁶ an open-source Multimodal Large Language Model

⁶<https://huggingface.co/blog/idefics>

Model	ACC	F1	# Params
BERT [♣]	0.705	0.705	110M
Llama3 [♣] (zero-shot)	0.485	0.451	8B
Gemma2 [♣] (zero-shot)	0.461	0.382	9B
ResNet50 [♣]	0.424	0.399	25.6M
ViT [♣]	0.460	0.462	86.6M
BERT + ResNet50 [*]	0.717	0.714	111.7M
BERT + ViT [*]	0.747	0.749	196.8M
CLIP [*]	0.431	0.298	151.3M
BLIP [*]	0.462	0.292	470M
IDEFICS [*] (zero-shot)	0.347	0.270	9B
IDEFICS [*] (fine-tuned)	0.600	0.591	9B
HUMAN	0.826	0.823	-

Table 2: Text-only[♣], image-only[♣], and multimodal^{*} model results on the MultiClimate test set.

(MLLM) by providing the following prompt template in zero-shot as well as fine-tuned settings.

Given the {frame} and {transcripts}, what is the stance of this frame-transcript pair towards climate change? Choose one between 0 for NEUTRAL, 1 for SUPPORT, and 2 for OPPOSE.

4.2 Results and Discussions

Table 2 presents the results of evaluating the effectiveness of individual modalities, multimodal models, and 9B-sized large models on MultiClimate.

BERT results are outstanding. The text-only BERT model achieves the best performance among single-modal models, notably surpassing the zero-shot LLMs. Furthermore, the multimodal fusion model BERT + ViT achieves state-of-the-art, 0.747 in accuracy and 0.749 in F1 score. Generally speaking, BERT + ResNet50/ViT fusion models outperform CLIP, BLIP, and IDEFICS, as textual features are crucial to our CC stance detection, and transcripts in YouTube videos benefit the already well-performing BERT model.

CLIP performs the worst in accuracy among trained multimodal models. One hypothesis is that the maximum sequence length has an impact on the results, that is, the maximum sequence length of CLIP is 77 tokens for text inputs, much shorter than BERT, and leads to declined performance due to a lack of information. For instance, the “MACC” video includes one sentence with 82 tokens, exceeding CLIP’s limit (77 tokens) but not BERT’s. Notably, fine-tuned BLIP shows similar poor perfor-

mance, in particular low weighted F1 score, which can be attributed to the misclassification of minority classes such as NEUTRAL and OPPOSE.

Textual and visual information compensate each other. Table 2 shows text-only models overall perform considerably better than image-only models, indicating that specific language contextual understanding can outperform approaches that are trained for detailed image classification. However, if both visual and textual information are concatenated, we can achieve the optimal result. We note that even though transcripts generally contain richer linguistic information, it is not guaranteed that visual and textual information are both meaningful in every frame-transcript instance. For example, video segments with no speech but only music playing in the background are transcribed as [Music] in YouTube. Text-dependent models can barely capture stances from text data in such cases, while the image can reveal additional information, and hence visual-informed models predict labels that align with human annotations; see the last 6 frame-transcript pairs of “AMCC” in Appendix E.

Text-only Llama3 wins in zero-shot. We also observe that Llama3 performs better than single-modal Gemma2 and multimodal IDEFICS within zero-shot, with a marginally increase of 0.138 in accuracy and 0.181 in F1 score compared with 9B-sized IDEFICS. Since Llama3 has been trained on more extensive text and speech data from various domains, it can better process longer sequences of text (Dubey et al., 2024). Meanwhile, we hypothesize that Llama3 is better at handling noisy data, which is common in the transcripts as many videos feature colloquial speech (e.g., in “AC-CFP”), thereby increasing the noise level. The results, alike the earlier BERT superior performances, suggest that the textual part is vital in CC stance detection due to its more explicit narrative and clearer directionality, and leveraging SOTA LLMs can significantly improve performance.

Zero-shot IDEFICS is biased toward SUPPORT. We also observe a tendency for the zero-shot IDEFICS model to predominantly predict SUPPORT labels, less often NEUTRAL, and rarely OPPOSE. This bias impacts the model’s performance negatively on videos where the majority of gold labels are OPPOSE or NEUTRAL. This accounts for the poor performance observed in videos like “CCUIM” (Acc/F1: 0.167/0.111; 7 SUPPORT, 21

NEUTRAL and 20 OPPOSE) and “EWCC” (Acc/F1: 0.163/0.109; 9 SUPPORT, 11 NEUTRAL and 29 OPPOSE). In contrast, the human inter-annotator Acc/F1 scores on these two test documents are not low: 0.771/0.773 on “CCUIM” and 0.816/0.811 on “EWCC,” illustrating that such bias is not present during human annotation.

The underlying reason for this annotation bias can stem from the mix of “stance” and “sentiment”. During manual annotation, instances that are negatively framed are often categorized as OPPOSE. However, the model classifies based on the expressed “stance” towards CC. Given that our video selection was filtered under “climate change”, a SUPPORT stance predominates. Additionally, the model occasionally conflates frame information, whereas annotations are based on both the frame and paired transcripts. Consequently, frames opposing climate change are undervalued in the classification of the model, leading to a pronounced bias towards NEUTRAL and SUPPORT categories.

Fine-tuned IDEFICS reduces stance bias. To mitigate biases present in the zero-shot IDEFICS model, we fine-tune IDEFICS using LoRA (Low-Rank Adaptation, [Hu et al. 2022](#)) on 80/10 MultiClimate train/dev videos before evaluation; see Appendix D for a list of hyperparameters.

Fine-tuning results in a significant increase in model performance. The accuracy increases from 0.347 to 0.600, and the F1 score improves from 0.270 to 0.591. These enhancements suggest that the model has developed a stronger capability for CC stance detection. Besides the higher Acc/F1, after examining the predictions made by the model before and after the fine-tuning process (see Appendix C), we observe a marked increase in the number of instances classified as OPPOSE, particularly the percentage of correctly predicted gold OPPOSE labels increased from 0.00% to 83.01%, indicating a better performance and less model bias.

5 Conclusion and Future Directions

This paper curates MultiClimate, a novel multimodal stance detection dataset to support video research on climate change. Our results show that BERT-fusion models can achieve considerably higher performance than large multimodal models. Textual information is vital on MultiClimate, while the visual modality can compensate for the drawbacks of language models, and the best is achieved when combining image and text information. Simi-

larly, SOTA LLMs beat large multimodal models in zero-shot. Our work provides a foundation for multimodal stance detection in CC.

We plan to extend our dataset with more videos and annotations and further investigate the interactions between visual and textual features, as well as the informativeness of each modality in manual stance labeling. We are also interested in expanding to audio and video modalities, as well as unimodal models using transfer learning techniques, to improve performance and explain stance detection in CC. We hope our work fosters the positive social impact of CC stance detection in the NLP field.

Limitations

While our study makes valuable contributions by enriching the multimodal climate change dataset and exploring the performance of different modalities in stance detection tasks, it is important to recognize several inherent limitations. First, the dataset annotation was conducted by a limited number of annotators, which may introduce personal biases into the data. Second, large language models are not fine-tuned due to limited computational resources. Third, our study focuses solely on transcripts and frames. Incorporating audio and video modalities can enhance the understanding of speaker’s emotions and intentions, and potentially further improve stance detection.

Acknowledgement

We thank anonymous reviewers for their constructive feedback on this work. This research project is in parts funded by Deutschlandstipendium, AWS (Amazon Web Service). This work also partially belongs to the KLIMA-MEMES project funded by the Bavarian Research Institute for Digital Transformation (bidt), an institute of the Bavarian Academy of Sciences and Humanities. The authors are responsible for the content of this publication.

References

- Vered Aharonson. 2023. Data of: Engaging for change-investigating youtube videos and users comments. *Mendeley Data, V1*, doi: 10.17632/j955mxnyyf.1.
- Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob L. Menick, Sebastian

- Borgeaud, Andy Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikolaj Binkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karén Simonyan. 2022. [Flamingo: a visual language model for few-shot learning](#). In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. [Enriching word vectors with subword information](#). *Transactions of the Association for Computational Linguistics*, 5:135–146.
- Miriam Louise Carnot, Lorenz Heinemann, Jan Braker, Tobias Schreieder, Johannes Kiesel, Maik Fröbe, Martin Potthast, and Benno Stein. 2023. [On stance detection in image retrieval for argumentation](#). In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2023, Taipei, Taiwan, July 23-27, 2023*, pages 2562–2571. ACM.
- Farabet Clement and Tris Warkentin. 2024. Gemma 2 is now available to researchers and developers. *Google Deepmind*.
- Costanza Conforti, Jakob Berndt, Mohammad Taher Pilehvar, Chryssi Giannitsarou, Flavio Toxvaerd, and Nigel Collier. 2020. [Will-they-won't-they: A very large dataset for stance detection on twitter](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 1715–1724. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. [An image is worth 16x16 words: Transformers for image recognition at scale](#). In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Jesus M. Fraile-Hernandez and Anselmo Peñas. 2024. [HAMiSoN-generative at ClimateActivism 2024: Stance detection using generative large language models](#). In *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, pages 79–84, St. Julians, Malta. Association for Computational Linguistics.
- Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2022. [A survey on stance detection for mis- and disinformation identification](#). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1259–1277, Seattle, United States. Association for Computational Linguistics.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. [Deep residual learning for image recognition](#). In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 770–778. IEEE Computer Society.
- Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. [Lora: Low-rank adaptation of large language models](#). In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.
- Dilek Küçük and Fazli Can. 2021a. [Stance detection: A survey](#). *ACM Comput. Surv.*, 53(1):12:1–12:37.
- Dilek Küçük and Fazli Can. 2021b. [Stance detection: Concepts, approaches, resources, and outstanding issues](#). In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 2673–2676. ACM.
- Xiaochong Lan, Chen Gao, Depeng Jin, and Yong Li. 2024. [Stance detection with collaborative role-infused llm-based agents](#). In *Proceedings of the Eighteenth International AAAI Conference on Web and Social Media, ICWSM 2024, Buffalo, New York, USA, June 3-6, 2024*, pages 891–903. AAAI Press.
- Junnan Li, Dongxu Li, Caiming Xiong, and Steven C. H. Hoi. 2022. [BLIP: bootstrapping language-image pre-training for unified vision-language understanding and generation](#). In *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 12888–12900. PMLR.
- Bin Liang, Ang Li, Jingqian Zhao, Lin Gui, Min Yang, Yue Yu, Kam-Fai Wong, and Ruifeng Xu. 2024. [Multi-modal stance detection: New datasets and model](#). In *Findings of the Association for Computational Linguistics ACL 2024*, pages 12373–12387, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics.
- Diana Maynard and Kalina Bontcheva. 2015. [Understanding climate change tweets: an open source](#)

- toolkit for social media analysis. In *29th International Conference on Environmental Informatics, EnviroInfo 2015 / 3rd International Conference on Information and Communication Technology for Sustainability, ICT4S 2015, Copenhagen, Denmark, September 7-9, 2015, Part 1: Conference Proceedings*, pages 242–250. Atlantis Press.
- Md Abdullah Al Mazid and Zaima Zarnaz. 2022. Climate change myths detection using dynamically weighted ensemble based stance classifier. In *ICCA 2022: 2nd International Conference on Computing Advancements, Dhaka, Bangladesh, March 10 - 12, 2022*, pages 277–283. ACM.
- Meta AI. 2024. Introducing meta llama 3: The most capable openly available llm to date. *Meta AI*.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.
- Jeffrey Morgan and Michael. 2024. Get up and running with large language models. *Ollama*.
- OpenAI. 2024. Hello gpt-4o. *OpenAI Blog*.
- Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR.
- Mariona Taulé, Francisco M. Rangel Pardo, M. Antònia Martí, and Paolo Rosso. 2018. Overview of the task on multimodal stance detection in tweets on catalan #1oct referendum. In *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018) co-located with 34th Conference of the Spanish Society for Natural Language Processing (SEPLN 2018), Sevilla, Spain, September 18th, 2018*, volume 2150 of *CEUR Workshop Proceedings*, pages 149–166. CEUR-WS.org.
- Suramya Tomar. 2006. Converting video formats with ffmpeg. *Linux Journal*, 2006(146):10.
- Apoorva Upadhyaya, Marco Fisichella, and Wolfgang Nejdl. 2023a. Intensity-valued emotions help stance detection of climate change twitter data. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI 2023, 19th-25th August 2023, Macao, SAR, China*, pages 6246–6254. ijcai.org.
- Apoorva Upadhyaya, Marco Fisichella, and Wolfgang Nejdl. 2023b. Towards sentiment and temporal aided stance detection of climate change tweets. *Inf. Process. Manag.*, 60(4):103325.
- Roopal Vaid, Kartikey Pant, and Manish Shrivastava. 2022. Towards fine-grained classification of climate change related social media text. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 434–443, Dublin, Ireland. Association for Computational Linguistics.
- Jiayu Wang and Jinyan Liu. 2024. The visual stories in addressing climate change that a petroleum business tells and lives by: A multimodal critical ecolinguistic perspective. *Energy Research & Social Science*, 111:103483.
- Maxwell Weinzierl and Sanda Harabagiu. 2023a. Identification of multimodal stance towards frames of communication. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 12597–12609, Singapore. Association for Computational Linguistics.
- Maxwell A. Weinzierl and Sanda M. Harabagiu. 2023b. Identification of multimodal stance towards frames of communication. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, EMNLP 2023, Singapore, December 6-10, 2023*, pages 12597–12609. Association for Computational Linguistics.

A Annotation Guideline

Annotators are tasked with determining the stance on climate change within each frame-transcript pair provided. This involves assessing both the visual and textual elements of each pair and following the below instructions.

Annotation Order. Annotators annotate at the frame-transcript level for each pair within a single video, deliberately avoiding sentence context.

Text and Frame Consideration. Annotators evaluate both the text and the accompanying image frame. In cases of conflict between text and image, prioritize the element that evokes stronger emotions related to the stance.

Data Storage. The annotated dataset is provided in the following formats on the GitHub repository.⁷ As shown in Figure 2, each sentence transcript is presented in a CSV file with the column label ‘text’, and the corresponding frame is provided as a JPEG file. The sentences and frames are in the same order.

A.1 Stance Definitions and Examples

Each frame-transcript pair must be annotated with one of the following stance values regarding climate change: SUPPORT, NEUTRAL, or OPPOSE.

Support. The frame-transcript pair accepts, agrees with, and/or promotes climate change-related topics or actions. Presented below are several sample categories that exemplify support arguments. Additional categories may also exist.

- *Action Promotion:* “Goal 13 of the sustainable development goals climate action.” WISE-002



- *Encouragement:* “It’s up to us to preserve these natural wonders and maintain the balance on earth.” HCCA-029

⁷<https://github.com/weryjw/MultiClimate/tree/main/dataset>



- *Achievements:* “It could prevent as much as 174 million tons of carbon from getting re-leased.” DACC-019

Behavioural change	2020
1. Eating less meat on	80-120
2. Switching car	50-110
3. Using less energy in	80-90
4. Switching car	40
5. Energy-efficient lighting	20-40
6. Teleworking	20-40
7. Virtual meetings	20

Changes in behaviour of households and consumers can result in large reductions of greenhouse gas (GHG) emissions in the EU, both in the shorter and in the long term.

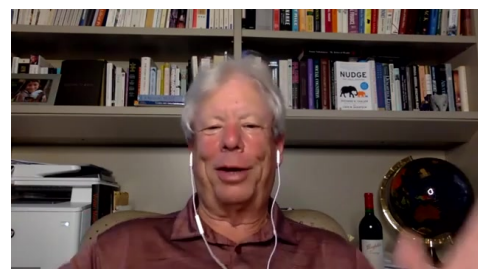
This study has focused on emission reduction potentials not covered by the EU Emissions Trading System and identified 36 options for behavioural change in the mobility, housing and food domains that will, when realised, result in a decrease of GHG emissions. Of these options, 11 have been studied in detail. If

- *Solutions:* “By eating more vegetables and less meat you not only get to enjoy the wonderful variety of fresh produce, but you also help the planet.” CCTA-010



Neutral. The pair neither supports nor opposes climate change topics or related actions. Presented below are several categories that exemplify neutral arguments. Additional categories may also exist.

- *Unrelated Context:* “People think that economists don’t agree about anything.” RHTCC-025



label	text
	Addressing Climate Change through Fiscal Policies in Indonesia.
[Music]	
[Music]	
[Music]	
	Indonesia is the world's largest island country with more than 13000 Islands.
	Indonesia is the world's largest island country with more than 13000 Islands.
	Covering an area of about 790 million hectares with a total cost line length of 95 thousands 181 kilometers and land territory of about 200 million hectares.
	Covering an area of about 790 million hectares with a total cost line length of 95 thousands 181 kilometers and land territory of about 200 million hectares.
	Covering an area of about 790 million hectares with a total cost line length of 95 thousands 181 kilometers and land territory of about 200 million hectares.
	In conflict in 2010 Indonesia was the first developing country to denounce voluntary commitment to reduce its carbon emissions by 26% by 2020 and up to 41 reduction with international support.
	In conflict in 2010 Indonesia was the first developing country to denounce voluntary commitment to reduce its carbon emissions by 26% by 2020 and up to 41 reduction with international support.
	In conflict in 2010 Indonesia was the first developing country to denounce voluntary commitment to reduce its carbon emissions by 26% by 2020 and up to 41 reduction with international support.
[Music]	
	Indonesia has translated his commitment into presidential regulation number 61 2011 on national action plan on greenhouse gases emission reduction
	Indonesia has translated his commitment into presidential regulation number 61 2011 on national action plan on greenhouse gases emission reduction
	And now by 2030 through at least an initial commit to reduce carbon emission by 29%.
	And now by 2030 through at least an initial commit to reduce carbon emission by 29%.

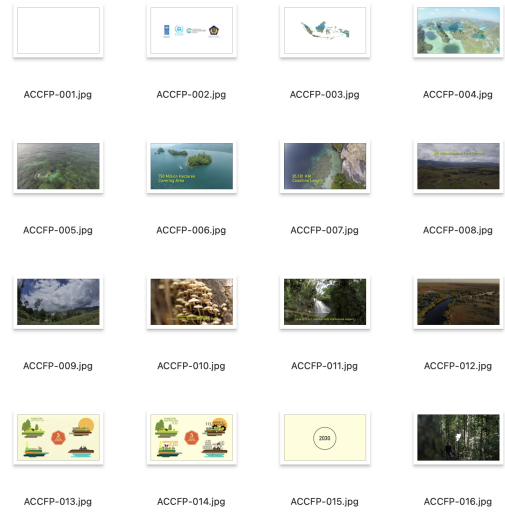
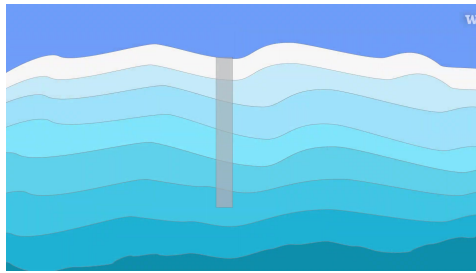


Figure 2: Annotator materials in “ACCFP”: CSV file (left) and frames (right) provided.

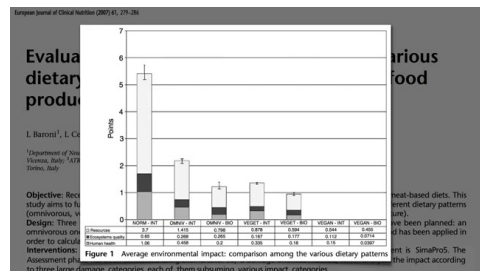
- *General Information*: “An ice core is a continuous section of ice drilled into a glacier or an ice sheet.” SDDA-007



- *Interrogative*: “Why have so few thought leaders made it their signature issue?” CCIS-008



- *Adverse Effects*: “And this is the negative effects the diet is having on the ecosystem and the adverse effects on human health.” DACC-047



Oppose. The pair expresses negative sentiments or criticism towards climate change or its related aspects. Presented below are several sample categories that exemplify opposing arguments. Additional categories may also exist.

- *Negative Consequences*: “Either everyone is leaving, or everyone is killing each other it all comes down to whether there’s enough rain.” TIOCC-005

B IAA Details on 10 Test Videos

We also assessed the accuracy and weighted F1 score on the 10 test video documents in Table 3.

C IDEFICS Confusion Matrix of Predictions Before and After Fine-tuning

The confusion matrices (predictions in %) before and after fine-tuning IDEFICS are shown in Figure 3 and Figure 4 respectively.

VIDEO	COHEN'S κ	ACC	F1
ACCFP	0.698	0.867	0.873
CCAH	0.778	0.867	0.851
CCSAD	0.408	0.644	0.675
CCUIM	0.633	0.771	0.773
EIB	0.647	0.822	0.809
EWCC	0.690	0.816	0.811
GGCC	0.736	0.827	0.814
SCCC	0.724	0.824	0.806
TICC	0.872	0.936	0.929
WICC	0.838	0.900	0.890

Table 3: Overview of IAA (Cohen’s kappa), accuracy, and weighted F1 score on 10 MultiClimate test videos.

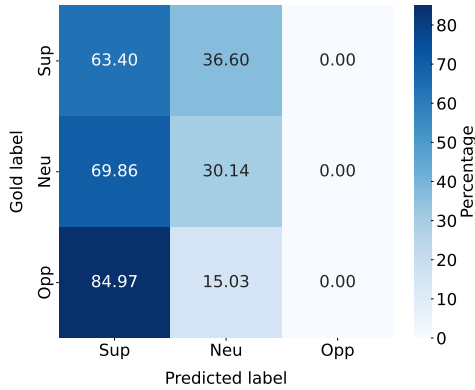


Figure 3: Confusion matrix of predictions before fine-tuning IDEFICS.

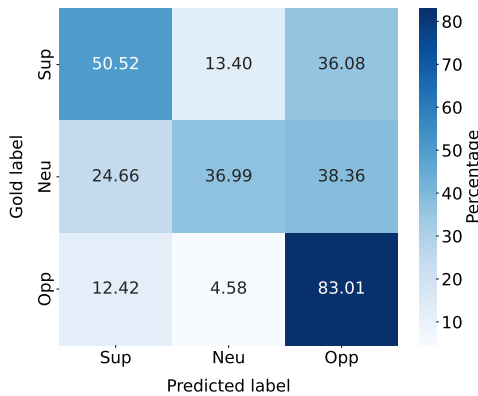


Figure 4: Confusion matrix of predictions after fine-tuning IDEFICS.

D Hyperparameters for Fine-tuning

Table 4 provides our explored hyperparameter space for all models in fine-tuning, helping in the reproducibility of our experimental results.

Model	HYPERPARAMETER	VALUE
BERT	Epochs	3
	Learning rate	2e-4
	Weight decay	1e-2
	Per device train batch size	16
	Per device eval batch size	16
ResNet50	Epochs	3
	Learning rate	3e-4
	Train batch size	32
	Eval batch size	32
ViT	Epochs	3
	Learning rate	3e-5
	Optimizer	AdamW
	Train batch size	32
BERT + ResNet50/ViT	Epochs	3
	Learning rate	2e-5
	Train batch size	4
	Eval batch size	4
CLIP	Epochs	3
	Learning rate	2e-5
	Train batch size	4
	Eval batch size	4
BLIP	Epochs	3
	Learning rate	1e-5
	Weight decay	1e-4
	Train batch size	4
IDEFICS	Eval batch size	4
	Epochs	4
	Learning rate	2e-4
	Per device train batch size	4
	Per device eval batch size	4
	Gradient accumulation steps	8
	Lora alpha	32
	Lora dropout	5e-2
	Eval steps	100
	Save step	100
	Logging steps	20
	Max steps	400

Table 4: Explored hyperparameter space for fine-tuning used in our experiment.

E Sample Responses Generated from Llama3 and Gemma2

Since BERT, image-only, and multimodal models are fine-tuned on our dataset, we provide exemplary responses generated by the other two text-only state-of-the-arts, Llama3 and Gemma2, in zero-shot setups, which may offer some insights for further study. Human annotations and original transcripts are given below.

E.1 SUPPORT Example

“Together we can address the causes and impacts of climate change to ensure that our spectacular ocean and great lakes continue to thrive while supporting our economies and values.” EIB-040



Llama3 Response

‘I would classify the stance of this text towards climate change as a 1 (Support). The text explicitly mentions addressing the “causes” and “impacts” of climate change, indicating an acceptance of the reality and need to take action on climate change. Additionally, it frames taking action on climate change as a positive step to ensure the continued health of natural systems and support economic and societal values.’

Gemma2 Response

‘The stance of this text towards climate change is 1 (Support).

Here’s why:

* **“Address the causes and impacts of climate change”** directly indicates a willingness to take action on climate change.

* **Phrasing like “ensure that our spectacular ocean and great lakes continue to thrive”** highlights the negative consequences of inaction and emphasizes the importance of protecting natural resources. The text uses positive and collaborative language, suggesting support for addressing climate change.’

E.2 NEUTRAL Example

“*[Music]*” AMCC-072



Llama3 Response

“I’m happy to help! However, I must point out that the provided text appears to be a single musical note, which doesn’t contain any information or

opinion about climate change. Therefore, it’s impossible for me to classify its stance towards climate change.

If you meant to provide a different text, please feel free to share it, and I’ll be happy to help with the classification!”

Gemma2 Response

‘The provided text “[Music]” does not express a stance on climate change.

Therefore, the classification is 0 (Neutral).’

E.3 OPPOSE Example

“*Injuries, property damage – even homelessness and loss of loved ones.*” CCAH-010



Llama3 Response

‘I would classify the stance of this text towards climate change as 2 (oppose). The text is listing potential consequences of climate change, such as injuries, property damage, homelessness, and loss of loved ones, which implies that the author views climate change as a negative and harmful phenomenon.’

Gemma2 Responses

‘The stance expressed in “Injuries, property damage – even homelessness and loss of loved ones” is strongly suggestive of 1 (Support) for addressing climate change.

Here’s why:

- **Negative Consequences:** The text lists severe negative consequences that are widely associated with the effects of climate change.
- **Emotional Impact:** Phrases like “loss of loved ones” evoke a strong emotional response, aiming to highlight the human cost of inaction on climate change.

The language implies a concern for these potential outcomes and suggests a need for action to prevent or mitigate them.’

F Dataset Overview

#	YOUTUBE-ID	TITLE	ABBREVIATION	# PAIRS	SPLIT	VIDEOSTANCE
1	bnN2_YNAEK8	About the Committee on Climate Change	ACCC	28	Train	SUPPORT
2	NTP5a1M2Z6I	Addressing Climate Change through Fiscal Policies in Indonesia	ACCFP	42	Test	SUPPORT
3	t-MiN_0kYhM	Analysing impacts of air quality policies on health and climate change Denise Mauzerall	AlAQ	59	Train	SUPPORT
4	BQr593iCEn8	AI and digital twins: Tools to tackle climate change	AIDT	22	Train	SUPPORT
5	T4CTgXY12kA	1-Arctic Methane and Climate Change	AMCC	74	Train	NEUTRAL
6	xajNochi7iM	Big data and climate change	BDCC	33	Train	NEUTRAL
7	eTqKLL_09yQ	Bitcoin Energy Consumption & Climate Change: Does Bitcoin Use Too Much Energy? [2021]	BECCC	41	Train	OPPOSE
8	iKzdIzN4q2Q	Belize: Women in Fisheries Forum 5: Gender and Climate Change - Understanding the Link	BWFF	79	Train	SUPPORT
9	s4ly6o-VT9o	Connections between air quality and climate - English - Sept. 2021	CBAQC	35	Train	OPPOSE
10	FGs2QQWfQyQ	The Crucial Connection: Climate Change and Health Kaiser Permanente	CCAH	30	Test	OPPOSE
11	1tGWJ-NkcGU	Climate change, biodiversity and nutrition - Helping local heroes tell their stories	CCBN	29	Train	SUPPORT
12	1Aop3wreUek	Climate change, biodiversity and nutrition nexus	CCBNN	15	Train	SUPPORT
13	4VXSrQospVY	Can climate change and biodiversity loss be tackled together?	CCCB	24	Train	SUPPORT
14	DRXQ9ixPbD8	Combating climate change in the Pacific	CCCP	28	Train	SUPPORT
15	TdxNG8L4JCM	Climate Change and Conflict in Somalia	CCCS	58	Train	SUPPORT
16	8fbrnAAg7VM	Climate change and development	CCD	106	Train	OPPOSE
17	nXOB8YPyc04	Climate Change and Food Supply	CCFS	43	Train	OPPOSE
18	MNdF-eVRWX4	Climate Change Fuelling Wilder Weather	CCFWW	27	Train	OPPOSE
19	v24wT16OU2w	Climate Change, Global Food Security, and the U.S. Food System	CCGFS	74	Dev	SUPPORT
20	CA8iTY7iMcK	Climate Change and our Health (ADB Insight Full Episode)	CCH	118	Train	SUPPORT
21	mPE7D0wRyOU	Climate Change: Health Equity Stories from The Colorado Trust (English subtitles)	CCHES	85	Train	OPPOSE
22	KxBaiad3Xto	Climate change in the Australian Alps	CCIAA	43	Train	SUPPORT
23	yeih2v425A	Climate Change: It's About Health Kaiser Permanente	CCIAH	37	Train	SUPPORT
24	m95K7LClIC4	350.org - Climate Change Is About Power	CCIAPI	28	Dev	SUPPORT
25	b919Fb-P3N8	Climate Change Impacts for Canadian Directors	CCICD	33	Train	OPPOSE
26	A7ktYbVsvr90	Climate change is simple: David Roberts at TEDxTheEvergreenStateCollege	CCIS	213	Train	OPPOSE
27	yVvVk2zNSbo	Untold stories of climate change loss and damage in the LDCs: Sierra Leone	CCISL	26	Train	OPPOSE
28	INBP7aRskVE	Climate Change: Mitigate or Adapt	CCMA	46	Train	SUPPORT
29	5DVa8xBgToc	Climate Change for South African Directors	CCSAD	59	Test	SUPPORT
30	GeksVaAmZzc	Climate Change or Social Change: The Role of Blockchain	CCSC	295	Train	SUPPORT
31	z19h-HTBHO8	Climate Change: Take Action	CCTA	22	Train	SUPPORT
32	iN-ZaY61_C8	Climate Change: The Philippines	CCTP	23	Train	SUPPORT
33	Vvefzge_RsA	Climate change unlikely to increase malaria burden in West Africa	CCUIM	48	Test	NEUTRAL
34	RZ-N5KwBaVc	Climate Change is Water Change	CCWC	12	Train	SUPPORT
35	MnTm89dSHHa	Climate Change and Water Quality	CCWQ	15	Train	SUPPORT
36	OL8a1YEhk_o	Honest Government Ad Climate Emergency & School Strikes	CESS	29	Train	OPPOSE
37	b7Lw66cSM4	How should Coronavirus influence the fight against Climate Change?	CICC	29	Dev	SUPPORT
38	vD0ix_b8jNM	COP28: Nuclear Science and Technology for Climate Change Adaptation	COP	66	Train	SUPPORT
39	MefBo0z0xi4	Crop production - Climate change affects biosphere - Earth Hazards - meriSTEM	CPCC	19	Train	SUPPORT
40	FHyUbeDVM3k	Capturing and transforming CO2 to mitigate climate change	CTCM	12	Train	SUPPORT
41	Gu5NKLxqTak	Diet and Climate Change: Cooking Up a Storm	DACC	77	Train	SUPPORT
42	OfYGx-N_gB0	Deforestation and Climate Change	DFCC	64	Train	SUPPORT
43	CG3pN7qQqZI	This Is How Denmark Protects Its Cities Against Climate Change	DPIC	29	Train	SUPPORT
44	Ry-e19Bu8UI	Developing tools for equality in climate change planning in Tanzania	DTECC	51	Train	SUPPORT
45	M17pm2PT_c	Effects Of Climate Change In MN Discussed At Seminar	ECCDS	16	Train	SUPPORT
46	HzL9hUOh_K4	Ecosystems at risk from Climate Change	EFCC	34	Dev	SUPPORT
47	kjTAWBPpeZ0	Earth Is Blue: Climate Change in your Sanctuaries	EIB	45	Test	SUPPORT
48	Qmxg97Ac9Wg	Extreme Weather and Climate Change, EarthNow	EWCC	49	Test	OPPOSE
49	Wrb4b28dgcU	Forests and Climate Change	FCC	13	Train	OPPOSE
50	9DaUn0geq4U	Fiji: Standing tall against climate change threats	FIIJ	32	Dev	OPPOSE
51	ison6lQozDU	Food loss and waste are among the main causes of climate change	FLW	14	Train	OPPOSE
52	M9wSP16p9xM	+Forest, together against climate change	FTACC	26	Train	SUPPORT
53	ep29Rw-i8Mo	Greenland's glaciers and Climate Change, Danish Broadcasting Corporation - Denmark	GGCC	52	Test	OPPOSE
54	S9Z_h1_LQ0o	How Climate Change Affects Biodiversity	HCCAB	25	Dev	SUPPORT
55	me14ikumMZE	How Climate Change Affects the Ecosystem	HCCAE	31	Train	OPPOSE
56	CGoNpwN0mrs	How Climate Change /Actually/ Works...in 4 Minutes	HCCAW	48	Train	NEUTRAL
57	KTA5onaECFE	How climate change influences geopolitics - Interview with Francesco Femia	HCCIG	14	Train	OPPOSE
58	RfxsDqQWjkh	How COVID-19 is impacting air pollution and climate change	HCI	16	Train	SUPPORT
59	ivN1QvdbUI	How do we change our attitude towards climate change? Christiana Figueres	HDWC	26	Train	SUPPORT
60	jDueueB3Tcs	Human Health, Vector-Borne Diseases, and Climate Change	HHVBD	98	Train	NEUTRAL
61	e_8puAyS0I	Human Rights Day 2012: Climate Change and Human Rights	HRDCC	42	Dev	OPPOSE
62	vsbcasoudmI	How are scientists helping whale-watchers adapt to climate change?	HSHWA	18	Train	NEUTRAL
63	me14ikumMZE	How Solar Panels Work To Reduce Climate Change	HSPW	28	Train	SUPPORT
64	RYsZ0NdHKyc	How the US National Security Policy Incorporates Climate Change - Interview with Alice Hill	HUSNS	39	Dev	SUPPORT
65	OSA944Shtme	IMRF: Statement from IOM's Migration, Environment, Climate Change, and Risk Reduction Division	IMRF	16	Train	SUPPORT
66	OKPzj-7gp0	INCAS: Monitoring for Climate Change	INCAS	45	Train	NEUTRAL
67	ks7rCR7-mF0	Migration and Climate Change in the Global Compact for Migration	MACC	62	Dev	SUPPORT
68	ko4cUnzoPic	Science Action: What's the unique role of methane in climate change?	MICC	53	Train	OPPOSE
69	vB3_49ULzF0	NASA's Climate Advisor Discusses Climate Change	NASA	20	Train	SUPPORT
70	EwHtHsBerIA	Overview of the C-ROADS Climate Change Policy Simulator	OCCE	46	Train	NEUTRAL
71	CuIdBaExvHy4	President Clinton On the Cost of Climate Change	PCOCC	22	Train	SUPPORT
72	A6uRlax7AuE	Preserve the Wonder - Climate change action	PWCCA	26	Train	SUPPORT
73	yTo3zmn3u84	Removing atmospheric greenhouse gases to prevent dangerous climate change	RAGG	37	Train	SUPPORT
74	_P31w8E_SZc	Regenerative Agriculture: A Solution to Climate Change	RASCC	22	Train	SUPPORT
75	Btqy4-H03dE	Ron Cohen on "BEACO2N" Climate Change Sensors	RCCCS	9	Train	SUPPORT
76	fCo77PW2G6Y	Reframing climate change science to include indigenous and local knowledge - Dr Tero Mustonen	RCCS	39	Train	NEUTRAL
77	s3VileAG6_U	Richard H Thaler on climate change	RHTCC	77	Train	NEUTRAL
78	G6JfQwonm78	Rare plant - Snakeshead Fritillary - defies climate change	RPDCC	30	Train	SUPPORT
79	K4dpmfzEASo	Strengthening Africa's pastoral food systems transformation in the face of climate change	SAPFS	52	Dev	NEUTRAL
80	N2L_AjZGVQo	System change NOT Climate change: Can we leverage the digital age to get there? - D2S Agenda	SCCC	34	Test	SUPPORT
81	se-BRvZuu7k	Scientists drill deep in Antarctic ice for clues to climate change	SDDA	32	Train	NEUTRAL
82	Zz8S18UfPYA	Saint Lucia and Climate Change Adaptation (English)	SLCCA	16	Train	SUPPORT
83	id4DZ0NiKk4	Stanford Students Tackle Climate Change	SSTCC	24	Train	NEUTRAL
84	pCraV8ahpYo	The connections between climate change and mental health	TCBCC	22	Train	NEUTRAL
85	tqavP5lotNo	Transforming our Economy to Combat Climate Change	TECCC	28	Train	SUPPORT
86	L5551Op_OpQ	Trade, Investment, and Climate Change in Asia and the Pacific	TICC	31	Test	SUPPORT
87	xWYwSgvZh38	The inequalities of climate change - ICRC	TIOCC	13	Train	OPPOSE
88	jsZ2_WFtDU	Things in Your Home that are Linked to Climate Change	TIYH	22	Train	SUPPORT
89	763lGy43spk	The technology fighting climate change	TTFCC	42	Train	SUPPORT
90	5KtGg-Lvxso	To understand climate change, understand these three numbers.	TUCC	38	Train	NEUTRAL
91	SDxnlvGIV9k	UK Climate Change Risk Assessment 2017 - Urgent priorities for the UK	UKCC	31	Train	OPPOSE
92	elcWgCjTHWM	Voices from Vanuatu: Climate Change Impacts and Human Mobility	VFVCC	67	Train	OPPOSE
93	i9mx391VVk	View from the Pacific: "Climate change is real"	VPCC	17	Train	SUPPORT
94	_JVDYaQDNCg	Wildfires and Climate Change Attribution: It's Complicated!	WCCA	25	Train	OPPOSE
95	TM_6C9sZLOI	Why focus on human security when working on climate change adaptation?	WFHSW	21	Train	SUPPORT
96	MPIFBW0NnWY	What is Climate Change?	WICC	30	Test	OPPOSE
97	mgBY0feG80U	What is climate change? Earth Hazards meriSTEM	WICCE	32	Train	OPPOSE
98	iXvyExAzQ58	What is the Impact of Solar Energy and Solar Panels on Climate Change?	WISE	25	Train	SUPPORT
99	iFmoMhVb6iw	Cuomo: Walk the Talk on Climate Change	WTCC	29	Train	SUPPORT
100	60byqdfPGLI	Yale Professor Tony Leiserowitz Discusses American Perceptions of Climate Change	YPTL	82	Train	NEUTRAL

Table 5: List of 100 Youtube videos on the MultiClimate dataset.