Predicting Narratives of Climate Obstruction in Social Media Advertising

Harri Rowlands[*]
InfluenceMap
harri.rowlands@influencemap.org

Gaku Morio[*]
Stanford University
Hitachi America[†]
gaku@stanford.edu

Dylan Tanner
InfluenceMap
dylan.tanner@influencemap.org

Christopher D. Manning
Stanford University
manning@stanford.edu

Abstract

Social media advertising offers a platform for fossil fuel value chain companies and their agents to reinforce their narratives, often emphasizing economic, labor market, and energy security benefits to promote oil and gas policy and products. Whether such narratives can be detected automatically and the extent to which the cost of human annotation can be reduced is our research question. We introduce a task of classifying narratives into seven categories, based on existing definitions and data. Experiments showed that RoBERTa-large outperforms other methods, while GPT-4 Turbo can serve as a viable annotator for the task, thereby reducing human annotation costs. Our findings and insights provide guidance to automate climate-related ad analysis and lead to more scalable ad scrutiny.

1 Introduction

Advertising has allowed firms to construct narratives that align with their commercial interests and sway public discourse. This is true in the context of climate change, where strategies akin to tobacco industry propaganda are employed to shape public perception by redirecting responsibility away from corporations (Supran and Oreskes, 2021).

In the domain of social media, entities, including fossil fuel corporations, utilize the platform to bolster existing beliefs about the significance of fossil fuels (Holder et al., 2023). For example, some advertisements (or ads) claim the indispensability of fossil fuels for jobs and the economy, and promote the idea that they are "clean." We refer to such narratives that obstruct progress against climate change as *climate obstructive narratives*. The scale of the public relations effort and advertising of climate obstructive narratives

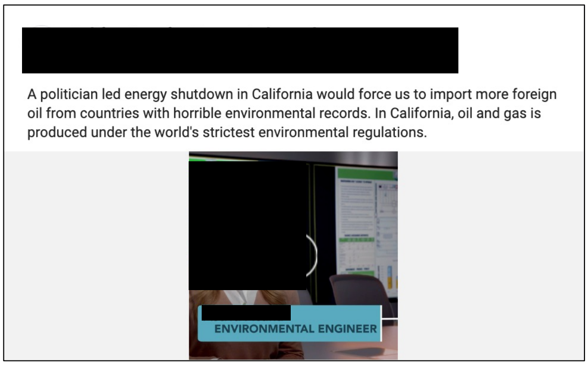

Figure 1: Example of an ad labeled 'Patriotic energy mix' (see Table 1). Entity information is blacked out.

is extensive, necessitating comprehensive analysis. Potentially disinformative ads must be identified and their messaging contrasted against climate science described by the Intergovernmental Panel on Climate Change, International Energy Agency, and other bodies mandated to provide objective analysis of climate change and its optimal solutions (InfluenceMap, 2021).

Identifying ads that contain climate obstructive narratives poses significant challenges in terms of efficiency and scale. This task usually relies on human expertise from academics or non-profit organizations (NPOs), due to the unique nature of the domain and the nuanced presentation of the ads. For instance, to accurately label about 1,500 ads, an NPO required the expertise of five subject matter experts (Holder et al., 2023), resulting in an estimated total of 120 hours spent. In this context, natural language processing (NLP) may potentially offer a viable alternative.

This paper proposes a multi-label classification task with seven classes to identify climate obstructive narratives. Our dataset was constructed based on the definitions and annotations of Holder et al. (2023), which includes Facebook ads by fossil fuel entities. An example of the 'Patriotic energy mix' type is shown in Figure 1, where this type suggests

[*]Equal contribution
[†]This work was done as a Visiting Scholar at Stanford University.

Findings of the Association for Computational Linguistics ACL 2024, pages 5547–5558
August 11-16, 2024

Super-category	Label	Description
Community & Resilience	CA	Helps national/local economies/communities, including through philanthropic efforts
	CB	Creates or sustains jobs
Green Innovation and Climate Solutions	GA	Emissions reductions and transitioning the energy mix
	GC	'Clean' gas as a climate solution
Pragmatism / Pragmatic Energy mix (Power systems and manufactured goods)	PA	Oil & gas as energy sources are a pragmatic choice and critical for maintaining functioning or optimal power systems
	PB	Oil & gas are needed as raw materials for alternative (non-power related) uses and manufactured goods
Patriotic Energy mix	SA	The production of domestic oil and gas reserves benefits the US, including through energy independence or energy leadership

Table 1: The labels included in the climate obstruction data. The categories, labels, and descriptions are taken directly from Holder et al. (2023).

that the production of domestic oil and gas reserves benefits the country. We utilize pre-trained language models, such as BERT (Devlin et al., 2019), and large language models (LLMs), such as GPT-4 Turbo (OpenAI et al., 2023) in our experiments. A comparable study on this task can be found in the work of Islam et al. (2023), while our approach differs in several ways. We utilize a high-quality annotated and relatively large dataset, benchmark various baseline models on it, and provide insights toward scalable ad scrutiny.

The experimental results show that RoBERTa-large (Liu et al., 2019) yields the best F-score, while GPT-4 Turbo performs close to RoBERTa-base. We also found that when GPT-4 Turbo is used to annotate training data for fine-tuning, only about 30 annotation examples are required to outperform RoBERTa-large. Given the need for experts to create the training data, the GPT-4 Turbo utilization is attractive even when considering the tradeoff between prediction cost and performance.

Given that recent research indicates that social media platforms are not adequately addressing the dissemination of misleading information (Holder et al., 2023), it is important to scale scrutiny of social media ads efficiently. Our study suggests that even with limited human resources, LLMs can be used to assist in monitoring climate obstructive ads. We release the code on GitHub (`https://github.com/climate-nlp/climate-obstruction-narratives`).

2 Background

Interdisciplinary study of climate change and NLP has gained attention in recent years. Major attempts have been made in this area to detect claims related to climate change. For example, datasets and models have been proposed to detect environment claims (Stammbach et al., 2023) or net-zero claims (Schimanski et al., 2023).

Islam et al. (2023) address theme classification of Facebook ads from fossil fuel entities. There are similarities to our study, although their work uses a smaller human-annotated dataset and does not use multi-label classification. The labels 'Patriotism', 'Pragmatism', 'Economy_pro', and 'ClimateSolution' found within the work align with some labels identified by Holder et al. (2023), though it appears that the respective studies were conducted independently. Some labels proposed in the work of Holder et al. (2023) are more fine-grained, and the results described in this paper provide a more detailed discussion. Technically, our study differs from the work of Islam et al. (2023) in that we benchmark various baseline models and few-shot learning on larger datasets based on high-quality annotations. Islam et al. (2023) use Sentence-BERT (Reimers and Gurevych, 2019) to classify ads and achieved an accuracy of 38.4%. We include different fine-tuned models and LLMs in our experiments and best models achieve F-scores around 70%. We also use our experimental results to discuss guidelines for automated ad scrutiny. In summary, our study provides more reliable and detailed discussions compared to the work of Islam et al. (2023).

From the view of technical classification of NLP, our work could be contextualized within climate change debate analysis (Stede and Patz, 2021), argument mining (Lawrence and Reed, 2019), discourse analysis, and persuasion and propaganda technique analysis. Luo et al. (2020) analyzed the global warming controversy using BERT, and there is similarity with our study in that it examines stances on climate change issues. In the propaganda typology (Martino et al., 2020; Da San Martino et al., 2019), 'Flag-waving' is somewhat similar to 'Patriotic energy mix' in our dataset. Our focus is, however, on the specific domain of the oil and gas sector.

Although the oil and gas sector may seem limiting, it is important to recognise that the domain is deceptively large, extending well beyond the primary operations of oil and gas companies to an extensive value chain. This value chain comprises refiners, pipeline operators, and manufacturers of secondary products (Olson and Lenzmann, 2016), all of which play integral roles in the industrial overall impact on both the economy and the environment. Furthermore, the industrial efforts to shape public discourse and policy are amplified through a network of agents, including Political Action Committees (PACs), trade associations, and lobbyists (Brulle, 2018). Our study is focused on ads that are affected by these value chains.

Research in other domains, such as health and politics, has also utilized NLP methods to detect misinformation (Schlicht et al., 2024; Raza, 2021). For instance, NLP methods have been applied to identify false statements and health-related misinformation (Sarrouti et al., 2021), demonstrating the versatility of these techniques across different areas of study and highlighting the critical role of accurate information in maintaining public trust and safety (Hirlekar and Kumar, 2020).

3 The Climate Obstruction Data

We built a dataset tailored for text classification based on the original data of Holder et al. (2023). The original dataset was compiled to include ads related to climate change that were run in the United States between January 1, 2020, and January 1, 2021 by utilizing the Facebook Ad Library API. The restriction to a limited timeframe in our study serves an important purpose, allowing us to target and analyze specific patterns in climate disinformation during a defined period. Given the rapid evolution of trends within this sphere, label definitions may need to be reconsidered as the dataset is expanded to a broader timeframe. The dataset primarily focused on the top ten fossil fuel companies, the top five industry associations representing the oil and gas sector, and ten advocacy groups with significant spending and connections to the fossil fuel industry.

Typology: Table 1 shows the labels and their brief descriptions provided by Holder et al. (2023). There is a super-category such as 'Community & Resilience' and subcategories for each super-category. For reference, we provide feature words analysis in Appendix Table 5. CA and CB emphasize the economy and include many job-related words. GA and GC have the potential to project a clean image to consumers by using words such as "clean". PA and PB emphasize the pragmatism of oil and gas by using words such as "affordable," "sanitizer," and "reliable."

Annotation: Holder et al. (2023) followed a rigorous coding scheme inspired by Miller and Lellis (2016), initially encompassing 25 subcategories under four broad themes: 'Community & Economy,' 'Climate Solutions,' 'Pragmatic Energy Mix,' and 'Patriotic Energy Mix.' However, to refine the process, the team performed three rounds of inter-coder reliability testing. This iterative process led to a more streamlined typology, eventually consisting of the four broad themes, each with three subcategories. Holder et al. (2023) chose to report the four broad category labels ('super-category' labels in this study) rather than each subcategory label, while we mainly focus on the subcategories. After three rounds of inter-coder reliability testing, the team achieved a Fleiss-Kappa score (Fleiss, 1971) of 0.78 (Holder et al., 2023). This indicates a high level of consistency in the annotation.

Data split: Since the original data were not designed for NLP tasks, we built the dataset by splitting the data into training, development, and test sets based on the entity name. We also removed any duplicate ad texts. Our dataset has imbalanced label distributions, and there are a reasonable number of samples with no labels (see Appendix Table 4). These reflect practical settings of real ads and provide challenging tasks for NLP methods.

4 Experiments

Because each ad can be associated with multiple labels, we define our task as multi-label classification for the text included in the ads. Conceptually, given input ad text $\mathcal{X}$, the output is a set of labels $\mathcal{Y} \subset \{CA, CB, GA, GC, PA, PB, SA\}$. An empty set is allowed for text that does not correspond to any of the labels. Although ads can contain images and videos, this study does not consider them. We use standard F-scores to evaluate the classification performance.

Models: We are motivated to compare different models that are simple and well-known, but strong baselines. To this end, we use a conventional

	CA	CB	GA	GC	PA	PB	SA	All
BERT-base	$71.4_{\pm 1.4}$	$72.3_{\pm 7.6}$	$58.6_{\pm 3.9}$	$9.6_{\pm 8.0}$	$75.1_{\pm 0.5}$	$16.7_{\pm 28.9}$	$19.7_{\pm 17.1}$	$61.1_{\pm 1.3}$
RoBERTa-base	$73.1_{\pm 2.2}$	$75.6_{\pm 2.2}$	$59.7_{\pm 3.8}$	$44.6_{\pm 7.1}$	$\mathbf{84.7}_{\pm 0.4}$	$16.7_{\pm 28.9}$	$35.9_{\pm 1.3}$	$69.5_{\pm 1.9}$
RoBERTa-large	$\mathbf{76.1}_{\pm 0.9}$	$78.5_{\pm 3.0}$	$64.7_{\pm 4.1}$	$43.9_{\pm 7.9}$	$\mathbf{84.7}_{\pm 0.8}$	$33.9_{\pm 5.8}$	$\mathbf{57.8}_{\pm 5.5}$	$\mathbf{71.4}_{\pm 0.6}$
Mistral7B-Inst	45.2	53.7	55.8	31.2	65.7	0.0	32.5	50.5
GPT3.5-trb	65.1	70.0	67.5	**54.3**	70.1	46.1	19.1	58.1
GPT3.5-trb (CoT)	56.8	57.1	49.9	40.5	55.3	66.6	47.3	52.2
GPT4-trb	69.6	**89.6**	**72.2**	37.8	73.6	**74.9**	38.7	66.9

Table 2: Subcategory level classification F-scores (avg. from three random seeds for BERT and RoBERTa). Mistral and GPTs are prompted with zeroshot.

approach using pre-trained language model fine-tuning and LLMs with prompting. **BERT** (Devlin et al., 2019) and **RoBERTa** (Liu et al., 2019) are simple yet strong baselines. Mistral-7B-Instruct-v0.1 (**Mistral7B-Inst**; Jiang et al. (2023)), GPT-3.5 Turbo (**GPT3.5-trb**), and GPT-4 Turbo (**GPT4-trb**; OpenAI et al. (2023)) are used to investigate the capabilities of zeroshot learning with prompting. We also explore Chain-of-Thought (CoT; Wei et al. (2022)) prompting (**GPT3.5-trb (CoT)**). We use DSPy (Khattab et al., 2023) and vLLM (Kwon et al., 2023) to implement the LLM experiments.

We also investigate conventional automated training data labeling with LLMs (Wang et al., 2021). We use GPT4-trb to label our training data, resulting in 'silver' training data. Then, we fine-tune RoBERTa-large on this data, referring to this model as RoBERTa-GPT4-trb-Label (**RoGL**). We investigate a low-resource scenario by fine-tuning RoGL on sampled human labeled training data. For implementation and hyperparameter details, refer to Appendix A.2.

4.1 Results

Overall Scores: Table 2 shows the subcategory level overall results. RoBERTa-large outperforms other models. The overall F-score of RoBERTa-large is over 70%, which is a notable result given the size of the training data. However, for low-frequency labels such as PB and SA, we have lower F-scores. This could be remedied by up-sampling and up-weighting for low-resource labels or by refining the prompting. Even though GPT4-trb is a zero-shot method, it outperforms the BERT-base model. Interestingly, GPT3.5-trb (CoT) does not outperform GPT3.5-trb. This may be because the rationale for each label gets diluted by the extra info in the prompt. For reference, we also examine the F-scores at the super-category level as shown in Appendix Table 6.

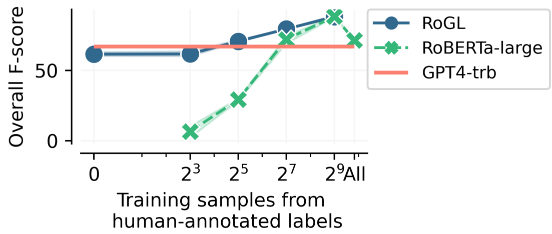

Figure 2: Results of the low-resource experiments (with narrow error bands just visible). We show zeroshot performance of GPT4-trb for reference.

Low-resource Scenario: Given the human costs associated with training data annotation, it is desirable to develop models with as little training data as possible. In particular, climate change policies and measures change frequently, which may require categorisation with new labels. New label definitions will also need to be created for sectors beyond oil and gas. Here, we experiment with a low-resource scenario where we change the size of training data (by random sampling) and investigate the trade-off between training data size and classification performance.

Figure 2 shows that RoBERTa struggles to output correct labels at the training sample scale of 2^3, while RoGL significantly outperforms RoBERTa, albeit with a slightly lower F-score than GPT4-trb. At the training sample scale of $2^5 = 32$, RoGL outperforms GPT4-trb and appears to almost saturate in classification performance. This result indicates that utilizing silver labels generated by LLMs is effective to reduce human annotation costs in the domain of climate obstructive narratives.

Error Analysis: We analyzed errors in the output from RoBERTa and GPT4-trb, as shown in Table 3, to understand the limitation of the methods. Note that 'No label' indicates that the model did not output a label.[1]

[1]Our task is multi-label classification and there are samples

No.	Text	Gold	RoBERTa-large	GPT4-trb
1	Thanks to increased natural gas production, U.S. CO_2 emissions are the lowest since 1985.	GC	GC	GA
2	From backpacks to binders to calculators, #natgas helps fuel the production of the essential supplies that students need, whether they are starting out the school year at home or at school! #<Anonymized>	PB	No label	PB
3	From the <Anonymized> to yours, Happy Independence Day!	CB, GC, PA, SA	No label	No label

Table 3: Example output errors.

In the No.1 example, RoBERTa produced the correct answer but GPT4-trb did not. This could be due to different interpretations of the text: GPT4-trb seems to have focused on the fact that emissions have lowered and labeled it GA, while RoBERTa may have focused on the "clean" image of this ad. This suggests that the subtle difference in nuance is acquired in the fine-tuning process.

In the No.2 example, RoBERTa produced incorrect output. We found that the training data contains a word 'backpack,' and the ads were annotated as unlabeled. Certain words in the training data could have biased the test output.

In the No.3 example, we found both RoBERTa and GPT4-trb produced errors. Looking at the text, 'No label' appears to be correct. However, upon checking the actual website of the ad, we found that there was a video embedded in the ad. The video did indeed contain content corresponding to CB, GC, PA, and SA. This indicates a limitation of this study, which deals only with text-based content.

More case studies can be found in Table 7.

5 Discussion

Which method is reliable in replicating the nuanced understanding of the ads? As we showed, fine-tuned RoBERTa-large performed best. There is no simple way to compare; however, given the inter-annotator agreement score of the dataset is 0.78, we can see that the classification performance of RoBERTa-large is close to expert performance. However, we found that RoBERTa runs the risk of overfitting the training data; GPT4-trb can produce more intuitive output but our prompts cannot reproduce the subtle nuances contained in the biases of the annotation. If the test data domain does not change significantly, RoBERTa seems to be the right choice, while GPT4-trb seems to be appropriate if one allows for looser annotation strictness.

Which method is better for practical applications? Our methods are useful for identifying trends in corporate advertising and detecting their stance on climate policy. On the other hand, the appropriate method can be selected depending on the use case. If there is sufficient training data, RoBERTa is effective, otherwise GPT4-trb can be used. GPT4-trb would be better suited for small-scale analyses because of throughput and cost issues, making it difficult to process the analysis efficiently. If one can provide training data of about 30 samples, RoGL is accurate with higher throughput. We believe that RoGL is sufficiently practical for ad analysis.

6 Conclusion

Research indicates that social media platforms are not adequately addressing the dissemination of misleading information (Holder et al., 2023). While we acknowledge the limitations, our study introduced methods that are both time-efficient and scalable for analyzing social media ads, offering a valuable way for NPOs and academic researchers aiming to undertake extensive evaluations. Importantly, the applicability of this approach extends beyond the environmental sector, holding promise for other areas impacted by disinformation, including natural ecosystems, biodiversity, and food security. In future studies, we will assess the accuracy of these methods against more recent ads which may display new climate narrative trends.

Acknowledgements

Computational resource of AI Bridging Cloud Infrastructure (ABCI) provided by National Institute of Advanced Industrial Science and Technology (AIST) was used. GM did this research within the Stanford Data Science (SDS) Affiliates Program. GM receives financial supports from Hitachi America to conduct this study.

with no associated labels. We refer to such cases as 'No label'; RoBERTa is trained with binary cross-entropy loss and assigns 'No label' only if the output probability of all labels are less than 0.5. GPT-4 trb is also supported to output 'X' (the same as 'No label') if there is no label to assign.

7 Ethical Consideration

This section discusses ethical considerations. This section is partly based on the guidelines of ACL Rolling Review (`https://aclrollingreview.org/responsibleNLPresearch/`) and NeurIPS Code of Ethics (`https://nips.cc/public/EthicsGuidelines`).

Privacy: The dataset we used comprises ads intended to have a broad reach. Therefore, we believe that privacy concerns are low. On the other hand, the ads may contain the names of specific individuals. Named entities have been left unanonymized; however, researchers should consider the potential impact on individuals when publishing their work.

Consent: The dataset comprises ads intended for broad dissemination, and the Facebook Ad Library API permits researchers to use related data in publications, mitigating consent concerns.

Copyright and Fair Use: See Appendix A.6.

Representative Evaluation Practice: See Section 8.

Safety: We do not propose technologies that directly harm humans.

Security: Our models could analyze advertising, like detecting potential greenwashing. However, model outputs are not infallible, with risks of false positives and negatives. Therefore, any analysis that relies on erroneous outputs may lead to erroneous conclusions. This could potentially and unfairly affect an entity's reputation. Additionally, an entity could use our model to make their ads less detectable by models. We encourage researchers to be aware of these limitations.

Discrimination: At a high level, our models determines the narrative strategy within ads. This causes us to label the ads for a particular entity. Furthermore, associating an individual's name with an ad could lead to discrimination against that person. Researchers should analyze entities from multiple perspectives, not solely based on model outputs, to prevent unwarranted conclusions. Caution is advised in publishing to prevent disadvantaging certain individuals.

Surveillance: N/A.

Deception & Harassment: We believe that the proposed models are unlikely to lead to hate speech or harassment issues. However, as noted above, the risk of labeling certain entities or individuals should be considered.

Environment: We acknowledge that, when our models are used to analyze advertising, energy consumption occurs. Our study focuses on few-shot learning and minimal fine-tuning of existing language models, thus reducing the energy consumed. We propose a new method, RoGL, which reduces energy consumption in comparison to LLM usage.

Human Rights: N/A.

Bias and Fairness: The dataset used in this study includes specific regions and individuals' names, potentially introducing bias into the model. For example, in our training data, if an entity labeled PA is located in California, it may increase the likelihood that other entities in California will also be labeled PA. The dataset size constraints make it challenging to fully eliminate these biases.

8 Limitations

We acknowledge limitations of the dataset used in this study. Our dataset is a small subset of the available ads. We only evaluate English ads from oil and gas entities in the United States. This limits the reliability of the task for other sectors, regions, and languages. Also, it cannot be guaranteed that the results achieved in this study will be replicated on more recent corporate ads.

In the low-resource scenario, the experiment was conducted with a set of fixed sample data points; thus, experiments outside these samples have not been validated. Also, because the dataset is small, variations in results due to the split of training and test are expected, but this study does not account for them. Some labels have very small samples in the test data. This limits the benchmarking capability of those labels. Note that our contribution is the empirical analysis, and we cannot generalize our experimental results or case studies given the limitations above.

References

Robert J. Brulle. 2018. The climate lobby: a sectoral analysis of lobbying spending on climate change in the usa, 2000 to 2016. *Climatic Change*, 149(3):289–303.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5636–5646, Hong Kong, China. 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.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Vaishali Vaibhav Hirlekar and Arun Kumar. 2020. Natural language processing based online fake news detection challenges – a detailed review. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, pages 748–754.

Faye Holder, Sanober Mirza, Namson-Ngo-Lee, Jake Carbone, and Ruth E. McKie. 2023. Climate obstruction and facebook advertising: how a sample of climate obstruction organizations use social media to disseminate discourses of delay. *Climatic Change*, 176(2):16.

InfluenceMap. 2021. Climate change and digital advertising, the oil and gas industry's digital advertising strategy. `https://influencemap.org/report/Climate-Change-and-Digital-Advertising-a40c8116160668aa2d865da2f5abe91b#1`.

Tunazzina Islam, Ruqi Zhang, and Dan Goldwasser. 2023. Analysis of climate campaigns on social media using bayesian model averaging. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 15–25, New York, NY, USA. Association for Computing Machinery.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. 2023. DSPy: Compiling declarative language model calls into self-improving pipelines. *arXiv preprint arXiv:2310.03714*.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of Third International Conference for Learning Representations*.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.

John Lawrence and Chris Reed. 2019. Argument mining: A survey. *Computational Linguistics*, 45(4):765–818.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Yiwei Luo, Dallas Card, and Dan Jurafsky. 2020. Detecting stance in media on global warming. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3296–3315, Online. Association for Computational Linguistics.

Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020. A survey on computational propaganda detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4826–4832. International Joint Conferences on Artificial Intelligence Organization. Survey track.

Barbara M. Miller and Julie Lellis. 2016. Audience response to values-based marketplace advocacy by the fossil fuel industries. *Environmental Communication*, 10(2):249–268.

Carol Olson and Frank Lenzmann. 2016. The social and economic consequences of the fossil fuel supply chain. *MRS Energy 38; Sustainability*, 3:E6.

OpenAI, :, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mo Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade

Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2023. Gpt-4 technical report.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Shaina Raza. 2021. Automatic fake news detection in political platforms - a transformer-based approach. In *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Sociopolitical Events from Text (CASE 2021)*, pages 68–78, Online. Association for Computational Linguistics.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Claude Sammut and Geoffrey I. Webb, editors. 2010. *TF–IDF*, pages 986–987. Springer US, Boston, MA.

Mourad Sarrouti, Asma Ben Abacha, Yassine Mrabet, and Dina Demner-Fushman. 2021. Evidence-based fact-checking of health-related claims. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3499–3512, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Tobias Schimanski, Julia Bingler, Mathias Kraus, Camilla Hyslop, and Markus Leippold. 2023. ClimateBERT-NetZero: Detecting and assessing net zero and reduction targets. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15745–15756, Singapore. Association for Computational Linguistics.

Ipek Baris Schlicht, Eugenia Fernandez, Berta Chulvi, and Paolo Rosso. 2024. Automatic detection of health misinformation: a systematic review. *Journal of Ambient Intelligence and Humanized Computing*, 15(3):2009–2021.

Dominik Stammbach, Nicolas Webersinke, Julia Bingler, Mathias Kraus, and Markus Leippold. 2023. Environmental claim detection. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1051–1066, Toronto, Canada. Association for Computational Linguistics.

	CA	CB	GA	GC	PA	PB	SA	None
train	221	166	102	59	225	32	69	253
dev	30	32	6	0	70	12	21	25
test	49	28	33	56	89	3	10	68

Table 4: The label distribution of the dataset. 'None' denotes samples with no labels.

Manfred Stede and Ronny Patz. 2021. The climate change debate and natural language processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 8–18, Online. Association for Computational Linguistics.

Geoffrey Supran and Naomi Oreskes. 2021. Rhetoric and frame analysis of ExxonMobil's climate change communications. *One Earth*, 4(5):696–719.

Shuohang Wang, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. 2021. Want to reduce labeling cost? GPT-3 can help. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4195–4205, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, brian ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pages 24824–24837. Curran Associates, Inc.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

A Appendix

A.1 Dataset Detail

The origin data from Holder et al. (2023) contained 30,116 ad samples. After processing to build the NLP dataset, we obtained 913 training ads, 162 development ads, and 255 test ads. The label distribution can be found in Table 4.

The feature word analysis is shown in Table 5. We used tf-idf (Sammut and Webb, 2010), excluding stop words. Scikit-learn (Pedregosa et al., 2011) was used to implement tf-idf.

CA	oil, gas, economy, energy, new, industry, alaska, natural, jobs, economic
CB	jobs, oil, gas, energy, economy, alaska, local, ballot, industry, natural
GA	energy, emissions, wind, gas, tci, carbon, rural, natural, learn, initiative
GC	gas, energy, natural, emissions, clean, future, carbon, <Anonymized>, reliable, climate
PA	energy, gas, natural, oil, pipelines, affordable, reliable, learn, americans, pipeline
PB	hand, sanitizer, energy, grade, distributing, lines, gas, <Anonymized>, refineries, oil
SA	energy, oil, gas, natural, production, texas, america, foreign, security, world

Table 5: Top feature words for each label by tf-idf excluding stop words.

A.2 Implementation and Hyperparameter Details

For fine-tuning, we used 1K training steps, a learning rate of 1e-5, and a batch size of 8. The optimizer used is Adam (Kingma and Ba, 2015). We used PyTorch 2.0.0 (Paszke et al., 2019) and HuggingFace transformers 4.28.1 (Wolf et al., 2020) for the model fine-tuning and predictions. We did not use the development data for validation. We experimented with fine-tuning using three different random seeds for each method and reported the average F-score.

We used V100 GPUs for BERT and RoBERTa, and A100 GPUs for Mistral7B-Inst. The parameter size of BERT-base is 110M. The parameter size of RoBERTa-large is 355M. The parameter size of Mistral7B-Inst is 7.3B. The parameter sizes of GPT3.5-trb and GPT4-trb are unknown. The exact GPU usage time, including preliminary experiments, is unknown; however, due to the small size of the dataset, fine-tuning RoBERTa-large only takes a few minutes.

For prompting, brief task and label descriptions were provided as shown in Figure 3. Figure 4 shows the variant for CoT prompting. We used DSPy 2.1.1 and vLLM 0.3.0 to implement the above, employing 'gpt-4-1106-preview' for GPT4-trb and 'gpt-3.5-turbo-1106' for GPT3.5-trb. The default temperature (i.e., zero) setting of DSPy was used. F-scores are reported based on a single run.

A.3 Super-category Level Results

Table 6 shows the F-scores in super-category level. GPT4-trb showed a similar F-score to RoBERTa-large.

Please label the following advert according to the described typology. Many adverts will not be relevant so please label them as X. We are looking for narratives specifically from the oil and gas sector.

Community & Resilience
CA: Emphasizes how the oil and gas sector contributes to local and national economies through tax revenues, charitable efforts, and support for local businesses.
CB: Focuses on the creation and sustainability of jobs by the oil and gas industry.

Green Innovation and Climate Solutions
GA: Highlights efforts to reduce greenhouse gas emissions through internal targets, policy support, voluntary initiatives, and emissions reduction technologies.
GC: Promotes "clean" or "green" fossil fuels as part of climate solutions.

Pragmatism/Pragmatic Energy mix (Power systems and manufactured goods)
PA: Portrays oil and gas as essential, reliable, affordable, and safe energy sources critical for maintaining power systems.
PB: Emphasizes the importance of oil and gas as raw materials for various non-power-related uses and manufactured goods.

Patriotic Energy mix
SA: Stresses how domestic oil and gas production benefits the nation, including energy independence, energy leadership, and the idea of supporting American energy.

X. No relevant typology detected.

This task is a multi-label classification and can have up to four labels amongst CA, CB, GA, GC, PA, PB, and SA.
If X is labeled, no other labels are allowed.
For example, a label containing GA and GC should be answered ["GA", "GC"].

Figure 3: The basic prompt for DSPy.

Please label the following advert according to the described typology. Many adverts will not be relevant so please label them as X. We are looking for narratives specifically from the oil and gas sector.

Community & Resilience
CA: Emphasizes how the oil and gas sector contributes to local and national economies through tax revenues, charitable efforts, and support for local businesses.
CB: Focuses on the creation and sustainability of jobs by the oil and gas industry.

Green Innovation and Climate Solutions
GA: Highlights efforts to reduce greenhouse gas emissions through internal targets, policy support, voluntary initiatives, and emissions reduction technologies.
GC: Promotes "clean" or "green" fossil fuels as part of climate solutions.

Pragmatism/Pragmatic Energy mix (Power systems and manufactured goods)
PA: Portrays oil and gas as essential, reliable, affordable, and safe energy sources critical for maintaining power systems.
PB: Emphasizes the importance of oil and gas as raw materials for various non-power-related uses and manufactured goods.

Patriotic Energy mix
SA: Stresses how domestic oil and gas production benefits the nation, including energy independence, energy leadership, and the idea of supporting American energy.

X. No relevant typology detected.

This task is a multi-label classification and can have up to four labels amongst CA, CB, GA, GC, PA, PB, and SA.
If X is labeled, no other labels are allowed.
For example, a label containing GA and GC should be answered ["GA", "GC"].

Reasoning process for analysis:
First, read the advert text to understand its main message.
Next, identify the key themes presented in the advert. This includes looking for mentions of economic impact, job creation, environmental efforts, or patriotic messaging.
Then, match these themes to the typologies listed above. Determine which of the typologies the themes of the advert align with.
If the advert contains elements from multiple categories, determine the primary focus of the advert and choose the most fitting category.
Finally, label the advert according to the most appropriate typology.

Figure 4: The CoT prompt for DSPy.

	C	G	P	S	All
BERT-base	$73.2_{\pm 1.0}$	$60.6_{\pm 5.1}$	$75.3_{\pm 0.5}$	$19.7_{\pm 17.1}$	$69.0_{\pm 2.1}$
RoBERTa-base	$78.5_{\pm 1.6}$	$67.2_{\pm 6.1}$	$83.9_{\pm 0.5}$	$35.9_{\pm 1.3}$	$75.9_{\pm 2.0}$
RoBERTa-large	$\mathbf{81.1}_{\pm 0.7}$	$72.7_{\pm 2.0}$	$\mathbf{84.4}_{\pm 0.8}$	$\mathbf{57.8}_{\pm 5.5}$	$\mathbf{78.8}_{\pm 0.8}$
Mistral7B-Inst	63.4	75.3	66.0	32.5	65.4
GPT3.5-trb	69.1	83.2	75.0	19.1	66.1
GPT3.5-trb (CoT)	59.0	69.1	60.8	47.3	61.7
GPT4-trb	79.9	**86.7**	75.9	38.7	77.9

Table 6: Super-category level classification F-scores. C, G, P, and S correspond to CA&CB, GA&GC, PA&PB, and SA.

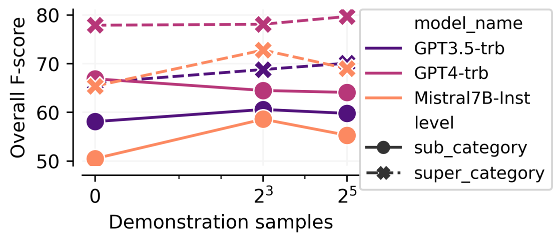

Figure 5: The in-context learning result

A.4 Effect of In-context Learning

We investigated the effect of in-context learning by providing few-shot examples with the prompts. This was also implemented with DSPy, and we tried 8 and 32 few-shot samples. The few-shot samples were the same as in the low-resource experiment. Figure 5 shows the result. GPT3.5-trb and Mistral7B-Inst appear to have improved its performance through few-shot learning, but GPT4-trb did not necessarily do so. Appropriate selection of the few-shot samples may improve performance. Development data can be used in this context.

A.5 More Predicted Examples

We show additional error output examples in Table 7.

A.6 Dataset Availability and License

The original data of our dataset can be obtained upon a reasonable request to the original data provider (Holder et al., 2023). Our dataset is subject to the terms of the work of Holder et al. (2023); however, we were unable to find the license for the original data. The copyright of the ads may belong to the owning entities or the Facebook Ad Library. Our code, including data preprocessing, model, and evaluation implementations, will be distributed under Apache 2.0 License. Please note that the dataset is intended to be used for research purposes. In particular, commercial purposes, for instance, may fall outside the scope of fair use.

A.7 Disclosure of the Use of LLMs

We used OpenAI ChatGPT and DeepL Write in parts of our paper to translate, correct grammar, and improve the writing. We declare that the original text is our own.

No.	Text	Gold	RoBERTa-large	GPT4-trb
1	Americans deserve a reliable, abundant energy source. See how the abundant supply of natural gas in America plays a critical role in energy security, strengthening the economy, creating jobs and more:	CA, CB, PA	CA, CB, PA	CA, CB, PA, SA
2	A proposed clean energy bill in PA aims to support natural gas, electric and hydrogen vehicles by developing transportation infrastructure. Natural gas is part of #<Anonymized>'s clean energy future. Learn more: <URL>	GC	GA	GC
3	Climate Commitment Announcement: Learn how #<Anonymized> will achieve its commitment to reduce emissions by 56% in ten years while on the path to net zero emissions by 2050.	GA, GC, PA	GA	GA

Table 7: Additional examples of output errors.