What Media Frames Reveal About Stance: A Dataset and Study about Memes in Climate Change Discourse

Anonymous ACL submission

Abstract

Media framing refers to the emphasis on specific aspects of perceived reality to shape how an issue is defined and understood. Its primary purpose is to shape public perceptions often in alignment with the authors' opinions and stances. However, the interaction between stance and media frame remains largely unexplored. In this work, we computationally explore this interaction with internet memes on climate change. We curate CLIMATEMEMES, the first dataset of climate-change memes annotated with both stance and media frames, inspired by research in communication science. CLIMATEMEMES includes 1,184 memes sourced from 47 subreddits, enabling analysis of frame prominence over time and communities, and sheds light on the framing preferences of different stance holders. We propose two meme understanding tasks: stance detection and media frame detection. We evaluate 7B LLaVA-NeXT and Molmo in various setups, and report the corresponding results on their LLM backbone. On both tasks, we observe models exhibiting strong in-context learning capabilities. Human captions consistently enhance performance. Synthetic captions and human-corrected OCR also help occasionally. Our findings highlight that VLMs perform well on stance, but struggle on frames, where LLMs outperform VLMs. Finally, we perform a case study on memes reflecting sociological concepts of climate change, analyzing VLMs' limitations in handling nuanced frames and stance expressions in this domain.

1 Introduction

Memes are a powerful communication format in online discourse that reflect communities' cultural and social dynamics (Davis et al., 2016; Zhang and Pinto, 2021). Combining images and texts, memes can encapsulate complex viewpoints in a compact and engaging format (Sharma et al., 2020; Liu et al., 2022). As a multimodal channel, memes express

(a) *convinced* stance with REAL and IMPACT frames

(b) *skeptical* stance with HOAX frame

Figure 1: Two climate change memes conveying opposite stances using different media frames.

the public's position towards a particular topic, i.e., stances as Mohammad et al. (2016) defines.

While stance reflects the creator's opinion toward a target, the specific narrative used to convey a certain stance is shaped by media frames. Media framing refers to selecting specific aspects of a perceived reality in communication to portray how an issue is defined, how its causes are interpreted, how its moral implications are evaluated, and what potential solutions are considered (Entman, 1993; Gidin, 1980). Depending on their stance, creators may gravitate toward different framing strategies (Snow and Benford, 1992). However, the interaction between stance and media frames remains under-studied, particularly in their representation through memes and their impacts on debates of global significance, such as the climate change.

Climate change (CC) memes are a vital component of social media, such as Twitter/X (Ross and Rivers, 2019). For example, Figure 1a conveys a *convinced* stance towards CC by using REAL and IMPACT frames (further detailed in §3.2) to affirm the evidence of global warming and its disheartening consequences. Conversely, Figure 1b conveys a *skeptical* stance using the HOAX frame, claiming that CC is not a major issue or even not real and suggests that politics may distort the CC issue.

In this paper, we analyze stances and media frames in CC memes[1] by examining the following three research questions (RQs):

- *RQ1: How do different media frames shape the visual representation of climate change in memes across varying stances?* We introduce ClimateMemes, a dataset of CC memes, consisting of 1,184 CC memes from 47 subreddits, manually annotated with stance on climate change and the media frames they invoke (§3) to analyze how memes convey stance through strategic media framing (§4).
- *RQ2: Can state-of-the-art VLMs and LLMs accurately detect stances conveyed by memes and the corresponding media frames?* We extend *stance detection* from text and propose a new task of multi-label *media frame detection* on CC memes. We evaluate two open-source VLMs and their backbone LLMs (§5.1) and investigate the effects of few-shot experiments and input modalities on these two tasks (§5.2). We found that VLMs exhibit few-shot learning capability on both tasks, and while synthetic meme captions cannot yet fully replace human-annotated ones, they still improve the VLMs' performance on both tasks. Moreover, we find that for frame detection text-only models outperform VLMs.
- *RQ3: Can taxonomies from communication science provide more insights on stance and media frame detection results?* We recruit communication science specialists to annotate humor type, person, and responsibility features on 235 test CC memes. Our analyses reveal that the performances of VLMs and LLMs degrade markedly on memes that are jokes, memes about political figures, and memes about individual-level (micro) responsibilities (§6).

2 Background

2.1 Memes

Internet memes are multimodal and humorous digital forms of expression that are popular across various digital channels, especially on social media (Shifman, 2014). They often use replicated and modified templates and are circulated among users to convey new, context-specific meanings. For example, Figure 1a uses the "Simpsons so far" template to highlight the ongoing trend of global warming. In controversial political arenas such as climate change discourse, internet memes are seen as effective tool for capturing attention, allowing users to communicate their stances through impactful imagery and humor (Ross and Rivers, 2019).

Nguyen and Ng (2024) categorize meme understanding into three main types: *classification*, *interpretation*, and *explanation*. Classification aims to assign labels to memes, such as identifying harmful content (Kiela et al., 2020; Pramanick et al., 2021; Cao et al., 2022; Hee et al., 2023), sentiment (Sharma et al., 2020; Chauhan et al., 2020) or figurative language (Liu et al., 2022; Xu et al., 2022). Interpretation tasks focus on understanding and generating insights from memes, such as generating captions or analyzing the metaphor between the image and text componants (Hwang and Shwartz, 2023; Chen et al., 2024). Explanation tasks go a step further by generating textual justifications for the labels assigned to memes (Hee et al., 2023). In this study, we curate ClimateMemes and introduce two meme understanding tasks: stance detection and media frame detection. We also collect supplementary annotations—human-corrected OCR and meme captioning—that can serve as foundations for future tasks.

2.2 Media Frames

Strategic media framing refers to the selective presentation of information to influence audience attitudes or evoke specific reactions (Snow and Benford, 1992). In social and communication science, framing has been studied by framing is studied through creating codebooks and manual annotations to identify generic and issue-specific frames for analyzing how information are selected and presented in the media.

The Media Frames Corpus (Card et al., 2015), focusing on three specific issues: immigration, smoking, and same-sex marriage, brought the methodologies of framing into our NLP community. Subsequent efforts have expanded this foundation, including proposals for general, issue-independent frame taxonomies (Johnson et al., 2017), computational framing analysis approaches (Mendelsohn et al., 2021; Ali and Hassan, 2022), and highlighting the importance of cognitive, linguistic, and communicative aspects beyond topical content in frame detection (Otmakhova et al., 2024).

In the context of climate change, framing has

[1]We release data and code at https://anonymous.4open.science/r/ClimateMemes-5371/

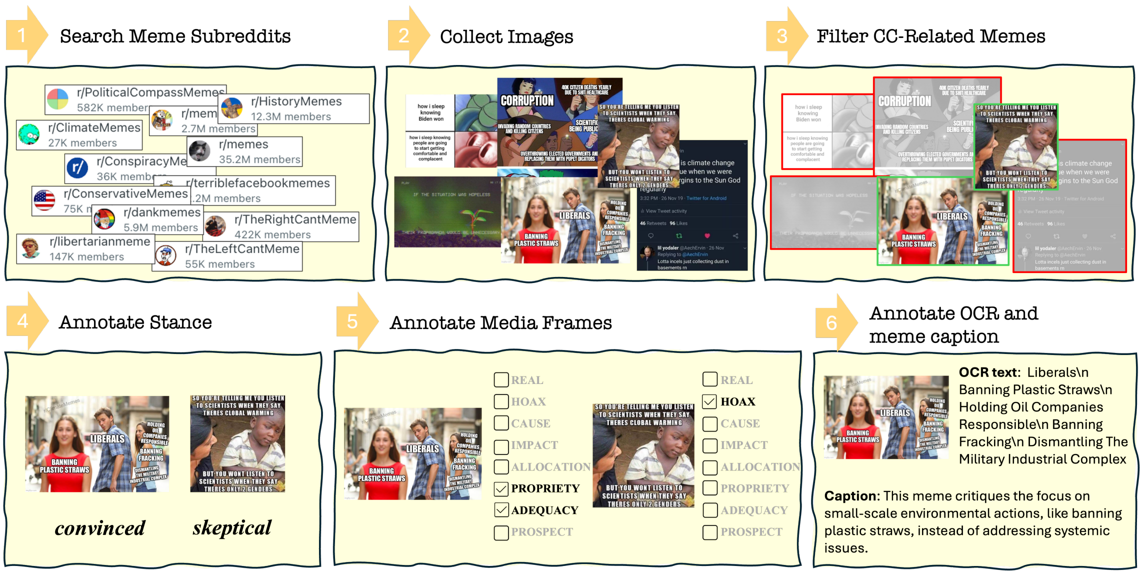


Figure 2: CLIMATEMEMES's pipeline of data collection, filtering, and annotations of stance, media frames, etc.

been studied to understand its role in public discourse and media representation. Stede et al. (2023) utilize generic frames, which are broadly universal and commonly observed across political discussions, to analyze climate change in Nature and Science editorials. Chen et al. (2022) study how frames evolve within public events, emphasizing their divergence and convergence in shaping climate change narratives, while Frermann et al. (2023) analyze how news articles across the political spectrum frame climate change. To the best of our knowledge, this paper presents the first dataset of memes annotated with media frames. We analyze how frames interact with stances, and evaluate the ability of pretrained Vision-Language Models (VLMs) for automatic frame and stance detection.

3 CLIMATEMEMES Dataset

This section describes CLIMATEMEMES, a dataset of 1,184 CC memes from 47 subreddits annotated with activated media frames and stances. Figure 2 illustrates our data processing pipeline. We discuss source meme collection and climate filtering (§3.1), and present guidelines and annotations for stances and media frames (§3.2).

3.1 Source Memes and Climate Filter

Data Source. To collect CC memes, we search subreddits with "meme" in their names and filter the topic of posts with the keyword "climate". To obtain a diverse range of perspectives on climate change, our collection includes subreddits like `r/ClimateMemes`, which primarily hosts climate activists, as well as `r/ConservativeMemes`, which reflects a community of more skeptical stances on climate change, even questioning the validity of the issue itself. In total, we collect 2,015 original images.

r/subreddit	#m	*conv./skep./nei.*	#f	top 3 frames
ClimateMemes	591	**94.1** / 3.2 / 2.7	2.35	ADEQ, CAUS, IMPA
TheRightCantMeme	90	13.2 / **83.5** / 3.3	1.70	HOAX, PROP, CAUS
dankmemes	90	**82.3** / 13.3 / 4.4	1.84	ADEQ, IMPA, REAL
memes	76	**92.1** / 1.3 / 6.6	1.83	IMPA, REAL, ADEQ
meme	50	**80.0** / 16.0 / 4.0	1.96	ADEQ, IMPA, REAL
ConservativeMemes	45	22.2 / **68.9** / 8.9	2.02	HOAX, PROP, REAL
Total	1,184	**78.0** / 17.2 / 4.8	2.11	ADEQ, IMPA, HOAX

Table 1: The number of memes (#m) in the top 6 frequent subreddits, along with percentages of *convinced*, *skeptical*, and *neither* stances, average number of involved frames (#f), and top 3 frequently used frames.

Filtering CC Memes. Two master students in computational linguistics manually annotated all images to ensure a curated collection of CC memes: *climate-associated* and in the format of a *meme*. Annotators first assess the relevance of these images to climate change, retaining only samples where climate change was a central theme. They then identify whether a sample qualified as a meme by examining its combination of visual and textual elements, humorous or satirical intent, and relevance to cultural or social contexts. As Figure 2 Step 3 shows, tweets containing only text or lyrical statements paired with images are excluded.

Out of 2,015 initially collected images, 1,184 CC-associated memes from 47 subreddits remained after filtering. Table 1 shows the top 6 subreddits that contribute to 79.6% of CC-associated memes (see Appendix A for a complete list of subreddits). The table also presents the stance and frame annotations distribution that we will detail in §3.2.

3.2 Annotation

Stance Annotation The SemEval 2016 shared task (Mohammad et al., 2016) introduced the stance detection task to classify tweets based on whether they are *in favor of*, *against*, or show *neither* stance towards specific targets, one of which was "Climate Change is a Real Concern." We assess the stances of these 1,184 CC memes regarding whether the meme creators are *convinced* that climate change is real, *skeptical*, or *neither* (i.e., cannot tell), following terminologies from social science, particularly Hoffman (2011) – see Appendix B.2 for detailed definitions.

Media Frame Annotation In communication science, media frames are frequently used to capture different, sometimes conflicting, perspectives on climate change. Jang and Hart (2015) propose five media frames to examine Twitter conversations on climate change. These frames include: REAL, emphasizing the present risk of climate change; HOAX, questioning the faithfulness of public communication regarding the risk; CAUSE, attributing the risk significantly to human activities; IMPACT, highlighting the net negative consequences of the risk; and ACTION, discussing necessary actions to address the risk. Ross and Rivers (2019) apply these five media frames to internet memes and exemplify the contrasting stances of individuals who are *convinced* of the climate change issue and those who remain *skeptical*. Yet, they only present a handful of examples, and a dataset for quantitative analysis and model training is still missing.

After adopting these five media frames and through multiple rounds of annotation revisions, we noticed the overly frequent occurrence of ACTION. To provide a more fine-grained analysis of media frames on CC memes, we subdivide the ACTION frame into the following four categories: ALLOCATION, PROPRIETY, ADEQUACY, and PROSPECT.

- ALLOCATION captures discussions about the responsibility of certain groups, such as nations, organizations, or even generations, to take action on climate change than others;
- PROPRIETY reflects debates on whether current actions are appropriate or effective;
- ADEQUACY highlights critiques regarding whether existing measures are sufficient to address climate risks or more actions are needed;
- PROSPECT explores perceptions of the potential outcomes of positive actions, distinguishing between climate doomists, who view catastrophe as inevitable, and climate risk realists, who believe meaningful prevention is still achievable (Davidson and Kemp, 2024).

Appendix B.3 presents detailed guidelines and examples of media frame annotations.

Inter-Annotator Agreement The first author of this paper annotated stances and media frames on all 1,184 CC memes. To ensure the reliability and consistency of the annotations, we asked one master student in computational linguistics to annotate 200 randomly sampled memes following guidelines in Appendix B.2-B.3. We achieved high agreement for stance detection: 0.83 on Cohen's Kappa. For media frame selection, since we allow one or more labels per meme, we assess MASI distance and achieve average score 0.83 (see Appendix B.4 for frame-specific IAAs.).

3.3 OCR and Meme Caption

CLIMATEMEMES includes two supplementary annotations: OCR correction and meme caption, as in Figure 2 Step 6. For each meme, we extract the embedded text via EasyOCR[2] and ask the two master students to correct any OCR errors manually. To obtain the meaning of the memes, we follow Hwang and Shwartz (2023) and ask the annotators to write a concise caption describing the message that the meme conveys. We further investigate in §5 whether added explicit textual information helps VLMs detect stances and media frames.

4 What Do Media Frames Reveal About Stance?

This section analyzes the interactions between stances and media frames in CC memes. We present CLIMATEMEMES statistics (§4.1), discuss frequently used media frames for *convinced* and *skeptical* memes (§4.2), concurrences of frames (§4.3), and analyze whether specific frames serve as reliable signals of a meme's stance (§4.4).

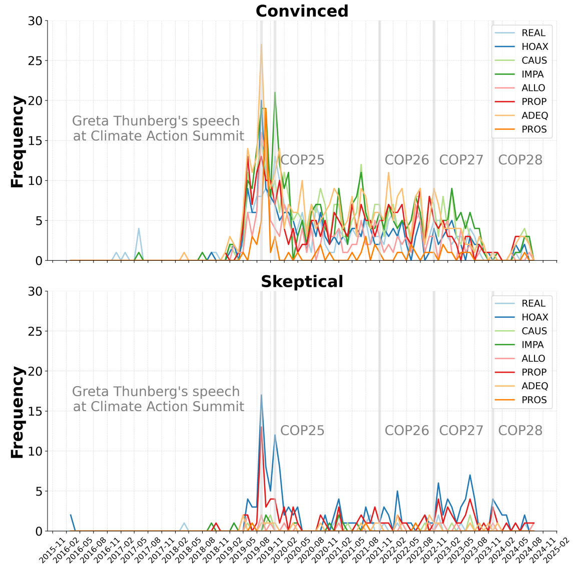


Figure 3: Monthly frequencies of media frames used in *convinced* versus *skeptical* memes.

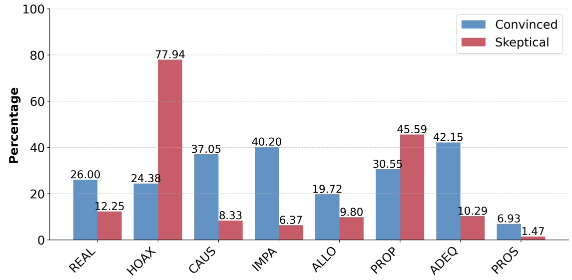


Figure 4: Frame preference of *convinced* and *skeptical* memes.

4.1 CLIMATEMEMES Statistics

Table 1 presents the number of memes in the top 6 frequent subreddits, along with their average number of frames and distribution of *convinced*, *skeptical*, and *neither* stances. About half of the 1,184 CC-associated memes are sourced from `r/ClimateMemes`, a community of climate activists. 94.1% memes from `r/ClimateMemes` exhibit a *convinced* stance, with the most frequently occurring frames being ADEQUACY, CAUSE, and IMPACT. These frames discuss human activities as primary drivers of climate change, enumerate negative consequences, and call for more actions.

`r/TheRightCantMeme`, `r/dankmemes` each account for about 8% of the total memes, ranking second in tie. 83.5% of the memes from `r/TheRightCantMeme` demonstrate a *skeptical* stance, with the predominant frames being HOAX, PROPRIETY, and CAUSE. These frames reflect skepticism toward the truthfulness of the CC communications concerning the effectiveness of current actions and the denial of human activity as the primary cause. In contrast, 82.3% of `r/dankmemes` memes exhibit a *convinced* stance, with REAL being a common frame, highlighting that CC is indeed happening.

Despite continuous efforts to upsample *skeptical* memes and subreddits, CLIMATEMEMES exhibits an unbalance where 78.0% memes are *convinced* and 17.2% are *skeptical*, most frequently employing ADEQUACY and HOAX frames, respectively.

4.2 Frame Preference

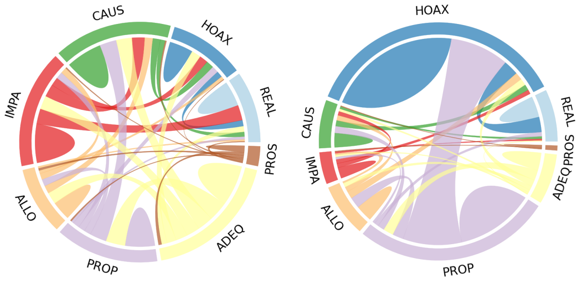
Our first analysis focuses on understanding the framing preference of *convinced* versus *skeptical* stances over time. The publication time of 1,184 CLIMATEMEMES memes spans eight years from March 16, 2016, to September 9, 2024. Figure 3 plots the monthly frequency of each frame separately for memes with *convinced* and *skeptical* stances. We observe two peaks in both the *convinced* and *skeptical* memes, occurring in September and December 2019, corresponding to Greta Thunberg's speech at the United Nations Climate Summit and the COP25 event, respectively. Interestingly, in *convinced* memes, the frequency of nearly all frames is significantly higher during these months, while in *skeptical* memes, only the hoax and propriety frames show an increase, with no significant changes in the other frames.

Figure 4 shows the probability of particular frames being involved in memes with *convinced* and *skeptical* stances. Among *skeptical* memes, 77.94% involve the HOAX frame, followed by the PROPRIETY frame at 45.59%. Other frames appear in less than 15% of the memes. In contrast, the frames in *convinced* memes are more diverse, with ADEQUACY, IMPACT, and CAUSE being the most common, appearing in 42.1%, 40.20%, and 37.05% of the memes. Other frames, except for the PROSPECT frame, appear in 20%-30% memes.

4.3 Frame Concurrence

Since each meme can use multiple frames (2.11 frames/meme, cf. Table 1), Figure 5 investigates the concurrence of frames in *convinced* and *skeptical* memes. For *skeptical*, the concurrence of HOAX and PROPRIETY frames is notably more potent than others. Rather, frame concurrences in *convinced* memes are more balanced across diverse combina-

[2]https://github.com/JaidedAI/EasyOCR



(a) convinced (b) skeptical

Figure 5: Concurrence of media frames in *convinced* and *skeptical* memes.

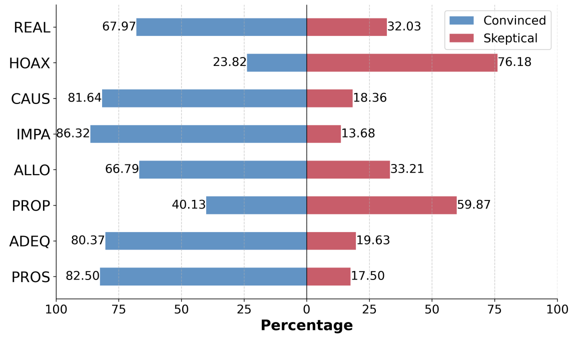


Figure 6: When a frame is used in a meme, which kind of stance is the meme more likely to hold?

tions, similar to observations in Figure 4. Moreover, we notice that HOAX has negative correlations with CAUSE, IMPACT, ADEQUACY, and PROSPECT, indicating that they tend not to co-exist (see Figure 7 in Appendix C for correlation heatmap).

4.4 Frame as a signal

Given that specific frames such as HOAX are prevalent in *skeptical* memes, we examine whether frames serve as an good signal for stance detection. Figure 6 analyzes the likelihood of a meme being *convinced* or *skeptical* when a specific frame is used. We observe that when CAUSE, IMPACT, ADEQUACY, and PROSPECT appear in a meme, there is >80% probability that the meme holds a *convinced* stance. REAL and ALLOCATION also appear more frequently in *convinced* memes. Conversely, HOAX implies a 76.18% probability that the meme is *skeptical*, followed by PROPRIETY (59.87%).

To sum up, strategic media framing is essential in conveying stances in CC memes. Though HOAX remains dominant in *skeptical* memes, framing is more diverse for *convinced* ones.

5 Stance and Media Frame Detection

To what degree can VLMs detect stance and frames in a meme? How can we improve their performance? In this section, we report on various experiments we performed on CLIMATEMEMES.

5.1 Experimental Setups

Models. We evaluate two similar-sized open-source VLMs: LLaVA-v1.6-Mistral-7B (LLaVA, Liu et al. 2024), Molmo-7B-D (Molmo, Deitke et al. 2024), both following a *visual encoder→cross-modal connector→LLM* approach. As a comparison to multimodal input, we also experiment with text-only inputs on their LLM backbones: Mistral-7B (Mistral, Jiang et al. 2023) and Qwen2-7B (Qwen, Yang et al. 2024).

Data Partition. We split CLIMATEMEMES into `train` and `test` sets with an 8:2 ratio, and all models are evaluated on the `235` test memes.

Evaluation Scenarios. In addition to zero-shot, we evaluate all models on n-shot experiments n ranging from 1 to 4. Following Huang et al. (2024), we leverage relative sample augmentation to select top n similar memes from `train` for each `test` meme based on the image and its human-corrected OCR. We also explore various input scenarios following Hwang and Shwartz (2023) to examine whether OCR, human, and synthetic meme caption can improve stance and media frame detection. We rotate stance and frame orders in prompts and report average over permutations (Zheng et al., 2023; Wang et al., 2024). For the backbone LLM baselines, we run experiments on text-only inputs.

Metrics. We report accuracy and macro F1 on stance detection due to imbalanced label distributions. Since one or more media frames can be assigned to one meme, we binary classify each frame and report the average over eight frames.

5.2 Which inputs help stance and frame detection in memes?

Table 2 examines how the number of shots and textual inputs, including meme captions from humans or VLMs and OCR, influence VLM performance.

Zero-shot vs. Few-shot. For all VLMs (LLaVA and Molmo) and their corresponding LLM backbones (Mistral and Qwen), few-shot setups outperform zero-shot on both tasks, evincing their in-context learning ability (0-4 shots in Appendix D).

Model	Inputs	Stance Acc.	Stance F1	Frame Acc.	Frame F1
LLaVA	meme	77.31	39.08	51.87	45.63
	meme+ocr	77.31	44.06	46.36	40.72
	meme+syn	73.95	40.01	52.45	45.78
	meme+syn+ocr	76.89	41.10	**52.57**	**45.87**
	meme+hum	**86.55**	**56.68**	49.96	44.18
	meme+hum+ocr	83.19	53.57	50.53	44.46
Molmo	meme	47.06	28.16	60.37	52.60
	meme+ocr	57.56	34.70	56.98	49.68
	meme+syn	61.76	39.25	58.37	51.02
	meme+syn+ocr	65.97	38.32	54.23	47.97
	meme+hum	**72.27**	**49.53**	**62.74**	**54.24**
	meme+hum+ocr	70.17	46.52	60.40	52.46
Mistral	ocr	51.90	37.09	61.71	54.79
	syn	58.23	36.06	59.03	53.01
	syn+ocr	59.66	42.71	61.78	55.20
	hum	**79.32**	**60.54**	64.61	58.31
	hum+ocr	67.65	48.96	**65.09**	**58.78**
Qwen	ocr	49.16	34.06	64.02	55.45
	syn	68.91	44.66	60.33	53.98
	syn+ocr	61.34	39.08	60.88	54.24
	hum	**73.11**	**53.28**	**65.86**	**58.23**
	hum+ocr	70.17	51.66	64.98	57.51

Table 2: Performance in accuracy and Macro-F1 on stance and frame detection with 4-shot setup. Backbone LLMs, Mistral and Qwen, only receive text input; syn = synthetic caption, hum = human caption.

Frame	#M	LLaVA Acc.	LLaVA F1	Molmo Acc.	Molmo F1	Mistral Acc.	Mistral F1	Qwen Acc.	Qwen F1
REAL	44	30.69	26.84	68.99	**59.43**	46.60	43.88	**72.13**	60.32
HOAX	81	51.60	46.90	**71.54**	**59.43**	**71.33**	60.54	71.01	**61.46**
CAUS	75	49.47	45.90	64.63	57.06	70.37	60.64	65.85	58.22
IMPA	77	48.03	45.37	60.59	54.64	**71.33**	**61.29**	61.54	56.03
ALLO	49	52.55	48.05	60.00	55.22	69.95	60.17	64.36	56.99
PROP	80	53.62	49.54	56.86	53.18	64.10	56.65	64.10	57.57
ADEQ	81	**56.86**	**50.50**	59.68	54.84	64.26	56.98	65.21	58.28
PROS	13	56.86	50.69	59.63	54.74	58.99	53.58	62.71	56.99
average		49.96	45.47	62.74	56.07	64.61	56.72	65.86	58.23

Table 3: Frame-specific performances with 4-shot meme+hum VLMs and hum LLMs. **Best** and worst scores per model are bolded and underlined. #M = number of `test` memes with the frame label.

VLMs vs. LLM backbones. To what extent can additional visual inputs benefit VLM performances on meme understanding?

While LLaVA has an edge over Mistral across various inputs on stance detection, both VLMs achieve lower scores on frame detection compared to LLMs. We hypothesize that VLMs are not pretrained on meme datasets for frame detection. Yet, there exist already high-quality textual data related to framing (Stede and Patz, 2021; Fremmann et al., 2023). Moreover, it should be noted that LLMs' winning performances benefit from costly human annotations (OCR correction[3] and captioning) or synthetic captions generated by VLMs.

OCR. On stance detection, extra OCR input is beneficial for VLMs—though only in setups *without* human annotations. For LLMs, feeding VLM-generated meme captions (syn) functions mostly better than using OCR, especially for Qwen. Combining OCR with synthetic captions can improve the scores for LLaVA in frame detection but always harms Molmo's performance on both tasks. Importantly, OCR weakens models when combined with human captions (except for a comparable performance on LLaVA few-shot frame detection) on both VLMs and LLMs. This underlines the importance of additional high-quality textual descriptions, leading to the overall best model for stance. Instead, for frames, LLMs outperform VLMs. We hypothesize that LLMs better grasp text inputs (especially captions) which aid fine-grained frame detection, while VLMs' performance is lower on frames and benefits less from more explicit texts.

Human vs. Synthetic Caption. Human meme captions improve performance on both tasks in almost all setups (except for frames with LLaVA). We leave it to future work to probe to what extent meme captions help understand stances and frames, for models and humans.

5.3 Which frames are harder?

To assess which frames are more challenging, Table 3 reports per-frame performances of VLMs and LLMs. Consistent with overall frame detection performance, Molmo outperforms LLaVA in predicting all 8 frames, e.g., HOAX achieves the highest score and PROPRIETY scores the lowest. For LLMs, Qwen outperforms Mistral with overall frame performance, but interestingly, is not consistently the best model for all frames.

5.4 Can frame labels help stance detection?

Table 4 investigates whether adding gold frame labels helps stance detection on 4-shot VLMs. Notably, for LLaVA, which is more proficient in stance detection, incorporating frame information tends to reduce its performance. On the other hand, for Molmo, which is better at frame detection, adding frame information generally boosts its performance. However, both models further improve their performance in the image and human caption setup. This suggests that stance and frame detection could benefit from multi-task training, improving performance through shared knowledge.

[3]We observed low-quality OCR; the average Levenshtein edit-distance before and after human correlation is 60.75.

Model	Base Input(+frame)	ΔAcc.	ΔF1
LLaVA	meme	-5.04	+0.45
	meme+ocr	-2.52	+5.26
	meme+hum	0.00	+1.06
	meme+hum+ocr	-15.54	-7.03
	meme+syn	+2.52	+2.27
	meme+syn+ocr	-16.39	-4.74
Molmo	meme	+1.26	+2.81
	meme+ocr	+7.15	+3.33
	meme+hum	+5.46	+4.15
	meme+hum+ocr	+7.14	-0.30
	meme+syn	+6.31	+2.73
	meme+syn+ocr	+7.56	+3.11

Table 4: VLM performance changes on stance detection when gold frame labels are added as additional inputs.

Concept	Label	#M	Stance Acc.	Stance F1	Frame Acc.	Frame F1
humor type	irony	33	**78.79**	53.30	64.91	57.43
	compare	25	88.00	69.90	65.75	54.67
	surprise	21	85.71	50.42	64.58	56.82
	person	21	95.24	82.05	67.26	61.39
	joke	19	84.21	52.53	**58.14**	50.86
	exag	10	100.00	100.00	60.31	56.37
	pun	5	80.00	**33.33**	59.69	**42.86**
	Total	134	87.42	63.21	62.95	54.34
person	ordinary	86	88.37	66.60	62.55	54.03
	celebrity	25	88.00	53.23	61.75	52.27
	politicial	14	85.71	78.79	**54.80**	**47.93**
	NGO	14	**50.00**	**37.18**	66.74	55.43
	Total	139	78.02	58.95	61.87	54.35
responsibility	macro	50	88.00	51.06	64.88	58.39
	meso	37	89.19	**31.43**	60.47	**52.10**
	micro	37	**83.78**	46.88	**60.26**	52.56
	Total	124	86.99	43.12	61.84	54.40

Table 5: Llava 4-shot meme+hum results on `test` subsamples with *humor*, *person*, and *responsibility* labels.

6 Meme Understanding through the Lens of Communication Science

To understand which aspects challenge models in meme understanding, we investigate three critical concepts in communication science: *humor type*, *person*, and *responsibility*. Humor is a key feature of memes, and different humor types, such as puns, sarcasm, and surprise, can have varying effects on readers. Personalization is a common communication strategy that simplifies complex political issues by focusing on individual actors. Responsibility for an issue can be attributed to the micro (individual), meso (organizational), or macro (societal) levels.

To analyze to what degree these concepts from communication science influence model performance, we recruited two bachelor's students in communication science to annotate humor type, person, and responsibility on 235 CLIMATEMEMES `test` memes. Our guidelines are adapted from a codebook on "Climate Change and Social Media"[4] and one or multiple labels can be assigned to the three concepts (see Appendix E for label definitions). Table 5 presents the most common labels, the number of relevant memes, and LLaVA's performances on the relevant subsample.

In stance detection, the model performs well on memes with *humor types* exaggeration and personification. Memes with irony and puns are difficult for the model, resulting in the lowest accuracy and F1. Memes with ordinary and celebrity *Person* categories show strong performance, but NGO stands out as a challenging category. For *responsibility*, memes concerning the macro level are the easiest for the model, while meso-level memes are the hardest. We hypothesize micro/meso-level responsibilities address specific individuals or organizations, whereas macro-level responsibility leads to less variation and facilitates meme understanding.

In frame detection, memes with *humor types* pun and exaggeration are the hardest. Under *person*, memes featuring ordinary people or celebrities are easier for the model than political ones. Frame detection on memes attributing *responsibility* mirrors results on stance detection, with macro performing best and meso being the hardest. In sum, annotations using taxonomies from communication science provide insights into aspects that caused difficulties in meme stance and frame detection.

7 Conclusion

We introduce CLIMATEMEMES, a new dataset of climate change memes annotated with stance and media frames. We demonstrate that media frame preferences are strong indicators of stance, with *convinced* and *skeptical* stances favoring distinct frames. We evaluate VLMs and LLMs, demonstrating the impact of in-context learning strategies on stance and media frame detection performance. Compared to the experiments without visual input on backbone LLMs, we identify challenges of VLMs in integrating multimodal information. Annotating concepts from communication science provides insights into which aspects of memes caused challenges to models. We plan to integrate these taxonomies on a large scale to improve VLMs and LLMs for future work.

[4]Anonymized version available at: §`https://osf.io/3hqdk?view_only=dd6035e7b03542e4a66c2fafa4bf0d7d`

Limitations

Potential Bias in the Sample Due to the Platform Our dataset is exclusively composed of memes collected from Reddit, which introduces a potential bias. By focusing solely on this platform, we limit the diversity of content that could be found on other platforms like Twitter, Facebook, Instagram, or 4chan. Each platform has its own user base, culture, and way of sharing and discussing memes, which could result in differences in the types of memes that are shared. This platform-specific limitation means that our findings might not be fully representative of meme trends across the internet as a whole. The memes on Reddit, for instance, might reflect particular political, social, or cultural preferences that do not necessarily align with those on other platforms, making the sample less generalizable.

Distribution of Meme Stances and Frames Might Be Different on Other Platforms The stance and framing of memes on Reddit may not reflect the patterns found on other social media platforms. Reddit has a unique structure, where specific subreddits cater to distinct interests, communities, and ideologies, which could influence the stances and frames adopted in memes. For example, some subreddits may have a higher concentration of memes that are either supportive or skeptical of climate change, while other platforms might exhibit different trends. Memes on Twitter or Instagram could carry different connotations, tones, or styles that might not be as prevalent on Reddit. Thus, the distribution of meme stances and frames could vary significantly across platforms, and a more comprehensive understanding of meme discourse would require analyzing multiple platforms to account for these differences.

Monthly Frequency: Sample Size May Be Too Small in Some Months to Derive Conclusions About Temporal Trends The monthly frequency of memes in our dataset might not be large enough in certain months to allow for meaningful conclusions about trends or changes over time. If the sample size in a given month is too small, it becomes difficult to accurately detect shifts in meme stances, frames, or topics that may occur over longer periods. This limitation could obscure any subtle trends or variations in the frequency of specific meme types or themes, making it harder to assess how the discourse around a particular subject evolves. For instance, if a meme trend spikes during a specific event but the dataset contains very few memes from that month, it might not reflect the broader public sentiment or provide an accurate representation of the temporal dynamics.

Visual inputs for VLMs We did not evaluate VLMs without visual input, and using the LLM backbone alone might not be 100% comparable to running a VLM without image input, because VLMs are fine-tuned on different datasets.

Ethics Statement

All annotations were conducted in accordance with ethical guidelines, ensuring that annotators were not exposed to any psychologically distressing content during the process. All annotators are paid according to national standards.

References

Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Rui Cao, Roy Ka-Wei Lee, Wen-Haw Chong, and Jing Jiang. 2022. Prompting for multimodal hateful meme classification. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 321–332, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Dushyant Singh Chauhan, Dhanush S R, Asif Ekbal, and Pushpak Bhattacharyya. 2020. All-in-one: A deep attentive multi-task learning framework for humour, sarcasm, offensive, motivation, and sentiment on memes. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 281–290, Suzhou, China. Association for Computational Linguistics.

Yingying Chen, Kjerstin Thorson, and John Lavaccare. 2022. Convergence and divergence: The evolution of climate change frames within and across public events. *International Journal of Communication*, 16:23.

Yuyan Chen, Songzhou Yan, Zhihong Zhu, Zhixu Li, and Yanghua Xiao. 2024. XMecap: Meme caption generation with sub-image adaptability. In *ACM Multimedia 2024*.

Joe PL Davidson and Luke Kemp. 2024. Climate catastrophe: The value of envisioning the worst-case scenarios of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 15(2):e871.

Corey B. Davis, Mark Glantz, and David R. Novak. 2016. "You Can't Run Your SUV on Cute. Let's Go!": Internet Memes as Delegitimizing Discourse. *Environmental Communication*, 10:62 – 83.

Matt Deitke, Christopher Clark, Sangho Lee, Rohun Tripathi, Yue Yang, Jae Sung Park, Mohammadreza Salehi, Niklas Muennighoff, Kyle Lo, Luca Soldaini, and 1 others. 2024. Molmo and pixmo: Open weights and open data for state-of-the-art multimodal models. *arXiv preprint arXiv:2409.17146*.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.

Lea Frermann, Jiatong Li, Shima Khanehzar, and Gosia Mikolajczak. 2023. Conflicts, villains, resolutions: Towards models of narrative media framing. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8712–8732, Toronto, Canada. Association for Computational Linguistics.

Todd Gidin. 1980. The whole world is watching. *Berkeley: University of California Press*.

Ming Shan Hee, Wen-Haw Chong, and Roy Ka-Wei Lee. 2023. Decoding the underlying meaning of multimodal hateful memes. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, pages 5995–6003. International Joint Conferences on Artificial Intelligence Organization. AI for Good.

Andrew J Hoffman. 2011. Talking past each other? cultural framing of skeptical and convinced logics in the climate change debate. *Organization & Environment*, 24(1):3–33.

Jianzhao Huang, Hongzhan Lin, Liu Ziyan, Ziyang Luo, Guang Chen, and Jing Ma. 2024. Towards low-resource harmful meme detection with LMM agents. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 2269–2293, Miami, Florida, USA. Association for Computational Linguistics.

EunJeong Hwang and Vered Shwartz. 2023. MemeCap: A dataset for captioning and interpreting memes. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1433–1445, Singapore. Association for Computational Linguistics.

S. Mo Jang and P. Sol Hart. 2015. Polarized frames on "climate change" and "global warming" across countries and states: Evidence from twitter big data. *Global Environmental Change*, 32:11–17.

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. *Preprint*, arXiv:2310.06825.

Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Leveraging behavioral and social information for weakly supervised collective classification of political discourse on Twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 741–752, Vancouver, Canada. Association for Computational Linguistics.

Douwe Kiela, Hamed Firooz, Aravind Mohan, Vedanuj Goswami, Amanpreet Singh, Pratik Ringshia, and Davide Testuggine. 2020. The Hateful Memes Challenge: Detecting Hate Speech in Multimodal Memes. In *Advances in Neural Information Processing Systems*, volume 33, pages 2611–2624. Curran Associates, Inc.

Chen Liu, Gregor Geigle, Robin Krebs, and Iryna Gurevych. 2022. FigMemes: A dataset for figurative language identification in politically-opinionated memes. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7069–7086, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26296–26306.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.

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.

Khoi P. N. Nguyen and Vincent Ng. 2024. Computational meme understanding: A survey. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 21251–21267, Miami, Florida, USA. Association for Computational Linguistics.

Yulia Otmakhova, Shima Khanehzar, and Lea Frermann. 2024. Media framing: A typology and survey of computational approaches across disciplines. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15407–15428, Bangkok, Thailand. Association for Computational Linguistics.

Shraman Pramanick, Dimitar Dimitrov, Rituparna Mukherjee, Shivam Sharma, Md. Shad Akhtar, Preslav Nakov, and Tanmoy Chakraborty. 2021. Detecting harmful memes and their targets. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2783–2796, Online. Association for Computational Linguistics.

Andrew S. Ross and Damian J. Rivers. 2019. Internet Memes, Media Frames, and the Conflicting Logics of Climate Change Discourse. *Environmental Communication*, 13(7):975–994. Publisher: Routledge _eprint: https://doi.org/10.1080/17524032.2018.1560347.

Chhavi Sharma, Deepesh Bhageria, William Scott, Srinivas PYKL, Amitava Das, Tanmoy Chakraborty, Viswanath Pulabaigari, and Björn Gambäck. 2020. SemEval-2020 task 8: Memotion analysis- the visuo-lingual metaphor! In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 759–773, Barcelona (online). International Committee for Computational Linguistics.

Limor Shifman. 2014. *Memes in Digital Culture*. The MIT Press.

David Snow and Robert D. Benford. 1992. Master-frames and cycles of protest. In *Frontiers in Social Movement Theory*.

Manfred Stede, Yannic Bracke, Luka Borec, Neele Charlotte Kinkel, and Maria Skeppstedt. 2023. Framing climate change in nature and science editorials: applications of supervised and unsupervised text categorization. *Journal of Computational Social Science*, 6(2):485–513.

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.

Viriya Taecharungroj and Pitchanut Nueangjamnong. 2015. Humour 2.0: Styles and types of humour and virality of memes on facebook. *Journal of Creative Communications*, 10(3):288–302.

Xinpeng Wang, Bolei Ma, Chengzhi Hu, Leon Weber-Genzel, Paul Röttger, Frauke Kreuter, Dirk Hovy, and Barbara Plank. 2024. “my answer is C”: First-token probabilities do not match text answers in instruction-tuned language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 7407–7416, Bangkok, Thailand. Association for Computational Linguistics.

Bo Xu, Tingting Li, Junzhe Zheng, Mehdi Naseriparsa, Zhehuan Zhao, Hongfei Lin, and Feng Xia. 2022. Met-meme: A multimodal meme dataset rich in metaphors. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’22, page 2887–2899, New York, NY, USA. Association for Computing Machinery.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, and 43 others. 2024. Qwen2 technical report. *Preprint*, arXiv:2407.10671.

Bingbing Zhang and Juliet Pinto. 2021. Changing the world one meme at a time: The effects of climate change memes on civic engagement intentions. *Environmental Communication*, 15:749 – 764.

Chujie Zheng, Hao Zhou, Fandong Meng, Jie Zhou, and Minlie Huang. 2023. Large language models are not robust multiple choice selectors. In *The Twelfth International Conference on Learning Representations*.

A Subreddits in CLIMATEMEMES

Table 6 details the public descriptions and meme frequencies of 47 subreddits in CLIMATEMEMES.

B Annotation Guidelines

B.1 Filtering Climate Change Memes

Is this image associated with the topic of climate change? Often images that discuss terms such as “climate change,” “global warming,” “greenhouse gas,” “carbon emission,” “fossil fuel,” “ozone,” “air pollution,” “carbon dioxide emissions,” “deforestation,” “industrial pollution,” “rising sea levels,” “extreme weather,” “melting glaciers,” “ocean acidification,” “biodiversity loss,” “ecosystem disruption,” “carbon capture,” “carbon storage,” “soil carbon,” “renewable energy,” “sustainable practices,” “Paris Agreement,” “Kyoto Protocol,” “carbon tax,” “emissions trading schemes,” “green technology,” “sustainable technology,” and “environmental change” are associated with climate change.

Additionally, if the meme features a well-known environmentalist or a political leader who has made statements related to climate change and environmental protection, it should also be considered as "associated with climate change." If you encounter an unfamiliar person, please use Google to search and confirm.

subreddit	**frequency**	**description**
ClimateMemes	591	The community to share environmental memes of prime quality. We advocate for climate action through funny captions and satire. Release your inner Greta, share your dankest decarbonization memes and raise global awareness to save the planet! Discuss climate strikes, climate change denial and doomerism, Fridays For Future, facts and news about nature, climate crisis quotes, ecology, Extinction Rebellion, and the end of the world.
TheRightCantMeme	91	Get your fix at left-wing Reddit alternatives: Hexbear and Lemmygrad. Also check out the Discord.
dankmemes	90	D A N K
memes	76	Memes! A way of describing cultural information being shared. An element of a culture or system of behavior that may be considered to be passed from one individual to another by nongenetic means, especially imitation.
meme	50	r/meme is a place to share memes. We're fairly liberal but do have a few rules on what can and cannot be shared.
ConservativeMemes	45	Become a ConservativeMemes subscriber! — Click the JOIN button now, and post your Conservative Memes later at /r/ConservativeMemes !!! — If you like political humor, political memes, politically incorrect memes, or conservative memes, this is the sub for you!
PoliticalCompassMemes	39	Political Compass Memes
terriblefacebookmemes	30	Community for all those terrible memes your uncle posts on facebook
ConspiracyMemes	18	This subreddit is devoted to memes relating to all things conspiracy. Things are pretty laid back around here so all people are welcome. The moderators believe in free speech and try not to moderate comments or posts unless it is absolutely necessary.
Memes_Of_The_Dank	15	This is a meme subreddit. That should be obvious by now. Also, it is slowly recovering from bots, and that's good.
libertarianmeme	14	For an end to democracy and tyranny. For more information about our ideology, check out the Mises Institute
HistoryMemes	11	A place for history memes about events over 20 years ago.
CommunismMemes	9	A place to share memes about communism.
PoliticalMemes	7	We're striving for equality here. Not "equality" in the sense that we'll allow people to post bigoted nonsense or perpetuate a false equivalency of entities, but "equality" in the sense that we are all co-inhabitants of this flying rock and need to learn to live together peacefully.
dank_meme	7	Dank Memes
MemeEconomy	7	
sciencememes	7	
PrequelMemes	6	Memes of the Star Wars Prequels.
PresidentialRaceMemes	6	
MemeThatNews	6	Learn and comment on the news with memes.
AusMemes	6	The Australia Memes subreddit. Just waiting for a mate.
Animemes	6	A community for anime memes!
Marxism_Memes	5	MEMES ARE THE NEW PAMPHLETS JURY NULLIFICATION FOR COMRADE LUIGI!
TheLeftCantMeme	4	They make a lot of bad Political Memes
MinecraftMemes	3	A place to post memes about Minecraft! Our Discord Server can be found in the sidebar below.
AnarchyMemeCollective	3	A reddit for sharing anarchist memes and for discussing anarchism. If you share your own OC let us know and we may share if on our other platforms.
depression_memes	3	Memes about depression.
Funnymemes	3	"Where Laughter Lives: Your Daily Dose of the Funniest Memes!"
marvelmemes	2	Welcome to r/marvelmemes: The home of Marvel memes on Reddit!
lotrmemes	2	Come on in, have a seat! This subreddit is a warm resting place for all weary travelers who are fond of Tolkien and his works. We welcome all Tolkien related content! Grab a pint, a long pipe, and relax.
VegMeme	2	A place to share animal rights humor, cartoons, image macros etc, because if you can't have a laugh at the hypocrisy and ignorance of carnists or have a good-natured laugh at ourselves you will probably become a misanthropic douchebag.
Jordan_Peterson_Memes	2	Welcome to the official subreddit for Jordan Peterson memes.
animememes	2	An anime meme subreddit that's friendly for women, queer people, and generally marginalized anime fans who want a break from how toxic anime spaces usually are. Of course, anyone is welcome, but be respectful to the intention of the space.
AvatarMemes	2	A subreddit for memes and other humor related to the Avatar franchise. Jokes based on ATLA, LoK, etc. are welcome.
CoronavirusMemes	2	Opening back up due to popular demand, didn't know people still wanted to post about the coronavirus. Monkeypoxmemes are allowed. Getting a laugh out of the Coronavirus while we still can, and spreading happiness in a time of distress.
SequelMemes	1	Memes of the Star Wars Sequels
VoluntaristMemes	1	Memes for voluntarists and other liberty loving people.
CommunistMemes	1	Communism is always the end goal!
SimpsonsMemes	1	Memes from The Simpsons!
MemePiece	1	The best place to find One Piece memes! We celebrate the comedic and casual side of the series One Piece. Casual or low effort content, normally removed from r/OnePiece, is likely welcome!
CrusadeMemes	1	DEUS VULT
MemeReserve	1	The Doomsday Global Meme Vault is a fail-safe meme storage sub, built to stand the test of time — and the challenge of natural or economical collapse. Only for the best memes!
GameOfThronesMemes	1	This subreddit is currently closed. Please check out r/aSongofMemesAndRage for memes based off GOT, ASOIAF, etc.
IncrediblesMemes	1	It's showtime
memesITA	1	Pizza, pasta & memes.
AnimeMeme	1	AnimeMeme for anime memes.
YouBelongWithMemes	1	The official meme subreddit for r/TaylorSwift

Table 6: CLIMATEMEMES's 47 subreddits with their descriptions and meme frequency.

Is this image a meme? Is it a cartoon? Memes are created by taking an existing widespread image and attaching new meaning to it by adding text within the image. A political cartoon, also known as an editorial cartoon, uses caricatures and satire to express an artist's opinion on current events, often critiquing political leaders, social issues, or corruption through humor and exaggeration. A cartoon style often features exaggerated characters and simplified forms, and the text is usually in hand-drawn fonts that match the casual, expressive tone of the illustration. Both memes and political cartoons are considered memes in this study.

B.2 Stance Annotation

What is the stance of this CC meme? We annotate the stances of CC memes into the following three categories: *convinced*, *skeptical* and *neither*.

- *convinced*: Accepts environmental risks, supports regulation of harmful activities, and reflects egalitarian and communitarian values.
- *skeptical*: Downplays or denies environmental risks, opposes regulation, and prioritizes individual freedom and commerce.
- *neither*: Does not align with convinced or skeptical stance and may present a neutral or unrelated stance.

B.3 Media Frame Annotation

Climate change, a critical global issue, refers to long-term alterations in temperature and weather patterns, largely driven by human activities such as fossil fuel combustion. As this issue gains prominence, memes—images paired with text—have become a widespread tool for expressing opinions and social commentary online via media framing.

In this task, you will be given CC memes and will be asked the following question: ***which media frames are used in these CC memes?*** Choose one or multiple that apply.

- REAL emphasizes that there are evidences indicating that CC is occurring;
- HOAX questions the faithfulness of public communication by politicians, the media, environmentalists, etc., e.g., if they are misrepresented or manipulated;
- CAUSE attributes human activities as a significant cause of CC;
- IMPACT highlights that CC leads to more net negative outcomes than if there was no CC;

Frame	α
Real	0.810
Hoax	0.868
Cause	0.825
Impact	0.711
Action_allocation	0.786
Action_propriety	0.777
Action_adequacy	0.740
Action_prospect	0.834

Table 7: Cohen's κ scores for IAA among two annotators.

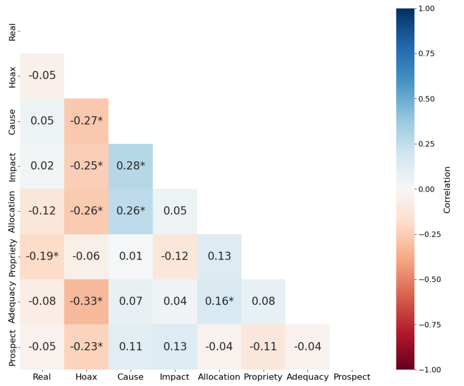


Figure 7: Correlation heatmap of frames. The values represent pairwise Pearson correlation coefficients. Values marked with * indicate corresponding p-values less than 0.05, indicating significance.

- ALLOCATION captures discussions about the responsibility of certain groups, such as nations, organizations, or even generations, to take action on climate change than others;
- PROPRIETY reflects debates on whether current actions are appropriate or effective;
- ADEQUACY highlights critiques regarding whether existing measures are sufficient to address climate risks or more actions are needed;
- PROSPECT explores perceptions of the potential outcomes of positive actions, distinguishing between climate doomists, who view catastrophe as inevitable, and climate risk realists, who believe meaningful prevention is still achievable (Davidson and Kemp, 2024).

B.4 Frame-level IAA

Table 7 presents per-frame Cohen's κ for inter-annotator agreement (IAA) among two annotators.

C Frame Correlation

We demonstrate the correlation among 8 frames in Figure 7.

D Additional Experimental Results

We show the full experimental results of stance detection of VLMs in Table 8, of LLMs in Table 9, frame detection of VLMs in Table 10, and of LLMs in Table 11. In line chart Figure 8, we also present VLM performances on stance and media frame detection with different shot and input setups.

E Definition of CC-associated communication science concepts

Humor Type. For the humor types category, the content format used to create humor in memes is coded. Humor types are initially independent of the respective humor style. Following (Taecharungroj and Nueangjamnong, 2015), a distinction is made between seven humor types, several of which can in principle be used simultaneously in a meme.

- **Puns** use language to construct new meanings or use words or phrases in a way that suggests two interpreta-tions, e.g. words that are pronounced the same but have different meanings.
- **Personifications** (personification) are used when human s and/or behavior are attributed to other objects such as animals, plants or objects.
- **Exaggerations** and understatements are disproportionate enlargements or reductions of a fact or context. Something is depicted as being larger or smaller than it (supposedly) actually is. Both the behavior of people and the consequences of events are depicted larger or smaller.
- **Comparisonsare** combinations of two or more elements (e.g. before and after pictures) to construct a funny situa-tion.
- **Irony and sarcasm** refers to the use of words to express the opposite of what one actually means.
- **Surprise** is the use of unexpected elements in memes. Memes with this element have a surprising ending/resolution.
- **Jokes** and nonsense describes content with no particular meaning and non-serious statements or actions that are only in-tended to make us laugh.

Person. Who is shown in the picture?

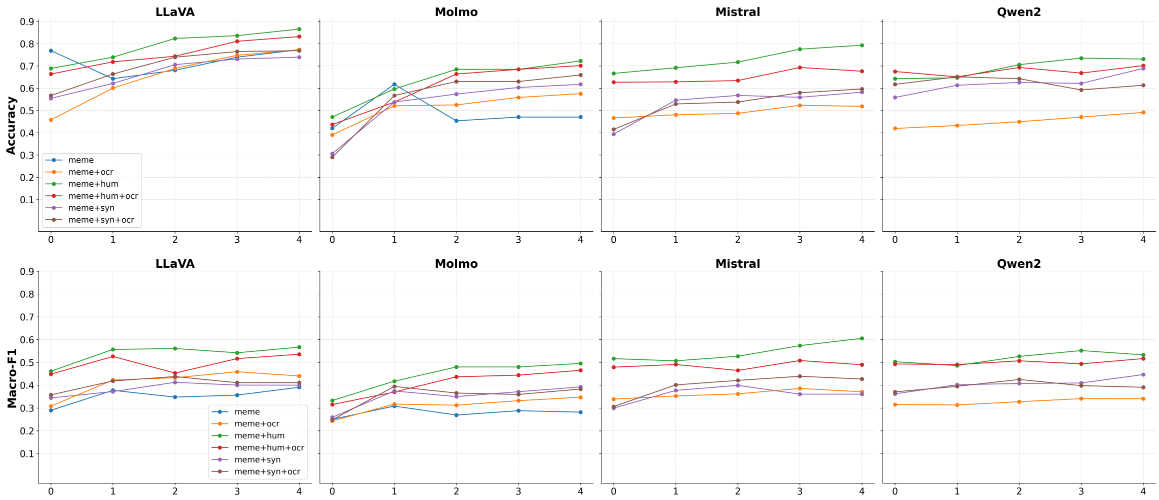
- **Political** actors include heads of state, members of government, official state delegates to the COP, ministers, representatives of institutions such as the UN or EU.
- **NGO** members or environmental activists. Members of protest movements such as Fridays for Future are considered environ-mental activists, whereas “normal” participants in demonstrations are coded as “normal citizens”.
- **Celebrities** are famous people who do not have an official political function. This includes, for example, people such as athletes, actors, influencers or artists.
- **Normal** or ordinary citizens are people who are not clearly assigned to one of the other categories.

Responsibility. To whom the responsibility for solving or combating the climate problem is attributed. The aim is to record who should take measures against climate change (e.g. more environmental protection, fewer emissions) or who isexpected to do so. Responsibility can be explicitly attributed or suggested by listing necessary measures that only a certain group can take.

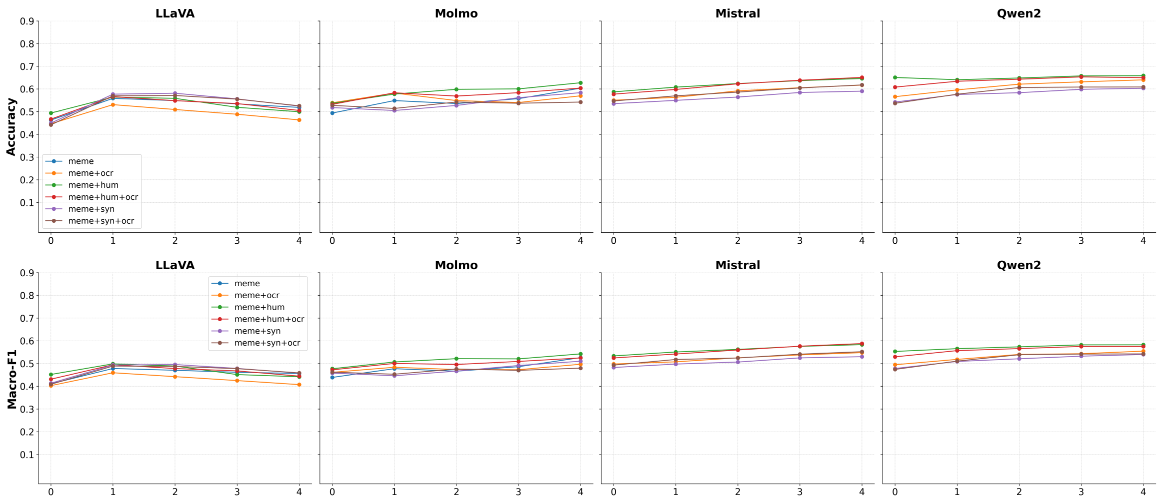
- Responsibility at **micro** level: Responsibility for individual persons such as politicians, activists, entrepreneurs, etc.
- Responsibility at **meso** level: Responsibility for individual companies, institutions, parties, parliaments, governments.
- Responsibility at **macro** level: Responsibility for certain countries, politics in general, the economic system, society, us as humanity, etc

F Elaboration About Use of AI Assistants

The authors used ChatGPT to polish writings for clarity and coherence and to assist with code generation. The authors manually inspected all ChatGPT suggestions and made corrections when necessary.



(a) Stance detection



(b) Frame detection

Figure 8: Accuracy and Macro-F1 of VLMs on stance and media frame detection with different shot and input setups.

Model	#S	Inputs	Acc.	F1	precision	recall
LLaVA	0	meme	76.89	28.98	26.75	31.61
		meme+ocr	45.80	30.80	35.87	37.73
		meme+hum	68.91	46.12	58.29	55.05
		meme+hum+ocr	66.39	44.83	48.50	55.75
		meme+syn	55.46	34.41	36.67	38.22
		meme+syn+ocr	56.72	35.75	38.32	40.48
	1	meme	64.29	37.75	39.42	38.16
		meme+ocr	60.08	42.27	41.78	49.04
		meme+hum	73.95	55.68	54.30	63.42
		meme+hum+ocr	71.85	52.60	53.35	61.04
		meme+syn	62.18	37.19	37.69	40.12
		meme+syn+ocr	66.39	41.85	41.85	44.24
	2	meme	68.07	34.77	40.36	35.37
		meme+ocr	68.91	43.26	42.31	45.27
		meme+hum	82.35	56.08	55.79	58.62
		meme+hum+ocr	74.37	45.31	43.33	52.29
		meme+syn	70.59	41.25	41.07	42.70
		meme+syn+ocr	73.95	43.79	44.55	44.95
	3	meme	73.95	35.64	45.55	35.40
		meme+ocr	74.79	45.90	45.97	45.95
		meme+hum	83.61	54.21	54.73	55.01
		meme+hum+ocr	81.09	51.65	52.37	53.97
		meme+syn	73.11	40.04	38.67	42.22
		meme+syn+ocr	76.47	41.10	39.71	42.73
	4	meme	77.31	39.08	49.55	40.69
		meme+ocr	77.31	44.06	46.19	42.86
		meme+hum	86.55	56.68	62.50	54.48
		meme+hum+ocr	83.19	53.57	55.13	53.97
		meme+syn	73.95	40.01	38.76	41.69
		meme+syn+ocr	76.89	41.10	39.60	42.90
Molmo	0	meme	42.02	25.02	30.30	30.31
		meme+ocr	39.08	24.32	31.31	31.70
		meme+hum	47.06	33.27	40.09	46.07
		meme+hum+ocr	43.70	31.50	40.16	45.55
		meme+syn	30.67	25.91	37.67	41.08
		meme+syn+ocr	28.99	24.80	37.18	40.38
	1	meme	61.76	30.83	30.98	32.34
		meme+ocr	52.10	31.68	34.73	37.93
		meme+hum	59.66	41.76	43.43	51.90
		meme+hum+ocr	53.78	37.00	41.04	48.83
		meme+syn	53.78	37.45	39.43	46.01
		meme+syn+ocr	56.72	39.45	41.55	48.08
	2	meme	45.38	26.90	31.27	32.56
		meme+ocr	52.52	31.22	34.30	36.36
		meme+hum	68.49	47.99	46.59	55.31
		meme+hum+ocr	66.39	43.66	44.16	52.27
		meme+syn	57.38	35.00	36.62	38.26
		meme+syn+ocr	63.03	36.54	36.65	41.55
	3	meme	47.06	28.81	33.69	37.59
		meme+ocr	55.88	33.20	35.32	39.48
		meme+hum	68.49	47.99	47.64	53.57
		meme+hum+ocr	68.49	44.41	45.90	51.40
		meme+syn	60.34	37.13	37.96	41.21
		meme+syn+ocr	63.03	35.93	36.24	40.68
	4	meme	47.06	28.16	32.95	34.98
		meme+ocr	57.56	34.70	36.37	42.78
		meme+hum	72.27	49.53	50.21	55.34
		meme+hum+ocr	70.17	46.52	52.67	54.70
		meme+syn	61.76	39.25	40.45	45.16
		meme+syn+ocr	65.97	38.32	37.73	43.63

Table 8: VLMs' performance in terms of accuracy and Macro-F1 on stance detection. Hum = human caption, syn = synthetic caption, #S = number of shots.

Model	#S	Inputs	Acc.	F1	precision	recall
Mistral	0	ocr	46.64	33.90	37.53	42.64
		hum	66.67	51.62	50.70	65.00
		hum+ocr	62.71	47.90	47.81	60.98
		syn	39.41	29.94	37.35	40.45
		syn+ocr	41.53	30.61	36.59	41.32
	1	ocr	48.10	35.28	39.01	44.88
		hum	69.20	50.68	49.24	59.52
		hum+ocr	62.87	49.07	48.56	64.09
		syn	54.62	37.70	39.04	44.40
		syn+ocr	52.94	40.13	41.31	51.75
	2	ocr	48.74	36.19	39.70	45.89
		hum	71.73	52.68	50.88	60.56
		hum+ocr	63.45	46.47	46.22	56.28
		syn	56.78	39.92	40.15	46.82
		syn+ocr	53.81	42.10	43.23	56.23
	3	ocr	52.32	38.59	40.81	49.21
		hum	77.54	57.36	55.22	62.96
		hum+ocr	69.33	50.81	49.02	59.57
		syn	55.93	36.06	37.34	41.03
		syn+ocr	57.98	43.90	43.68	55.56
	4	ocr	51.90	37.09	39.52	46.43
		hum	79.32	60.54	57.57	66.95
		hum+ocr	67.65	48.96	47.40	57.36
		syn	58.23	36.06	37.24	40.24
		syn+ocr	59.66	42.71	42.97	51.47
Qwen2	0	ocr	41.95	31.49	36.80	45.72
		hum	64.29	50.29	55.11	59.46
		hum+ocr	67.51	49.31	49.53	59.05
		syn	55.88	36.20	53.12	51.88
		syn+ocr	61.76	37.06	43.43	47.13
	1	ocr	43.28	31.34	35.86	40.83
		hum	64.71	48.60	50.08	58.11
		hum+ocr	65.13	49.02	50.31	59.80
		syn	61.34	40.18	42.22	43.47
		syn+ocr	65.13	39.49	42.33	39.37
	2	ocr	44.96	32.74	37.72	44.56
		hum	70.59	52.60	55.23	59.66
		hum+ocr	69.33	50.69	52.38	61.32
		syn	62.61	40.74	50.80	43.77
		syn+ocr	64.29	42.51	47.10	46.20
	3	ocr	47.06	34.10	38.19	42.38
		hum	73.53	55.19	57.97	61.74
		hum+ocr	66.81	49.35	50.24	58.10
		syn	62.18	40.97	47.24	44.47
		syn+ocr	59.24	39.80	44.51	45.65
	4	ocr	49.16	34.06	37.96	39.98
		hum	73.11	53.28	55.79	58.30
		hum+ocr	70.17	51.66	52.22	61.01
		syn	68.91	44.66	47.97	49.62
		syn+ocr	61.34	39.08	40.29	44.12

Table 9: LLMs' performance in terms of accuracy and Macro-F1 on stance detection. Hum = hum caption, syn = synthetic caption, #S = number of shots.

Model	#S	Inputs	Acc.	F1	precision	recall
LLaVA	0	meme	46.27	41.06	44.59	50.60
		meme+ocr	44.65	40.27	46.50	50.39
		meme+hum	49.38	45.18	51.77	53.19
		meme+hum+ocr	46.70	43.10	51.23	52.08
		meme+syn	44.93	41.34	50.59	50.22
		meme+syn+ocr	44.19	40.85	50.65	50.35
	1	meme	55.86	47.82	50.96	53.05
		meme+ocr	53.09	45.94	49.77	52.87
		meme+hum	56.42	49.85	53.13	56.16
		meme+hum+ocr	56.62	49.44	53.21	55.01
		meme+syn	57.71	49.35	52.78	53.97
		meme+syn+ocr	56.97	48.89	52.42	53.85
	2	meme	54.87	47.03	50.80	52.67
		meme+ocr	50.95	44.22	47.79	52.40
		meme+hum	55.86	48.91	52.19	55.72
		meme+hum+ocr	54.85	47.80	51.38	54.34
		meme+syn	58.13	49.56	52.88	53.88
		meme+syn+ocr	57.13	48.74	52.03	53.51
	3	meme	53.44	46.15	49.55	53.47
		meme+ocr	48.86	42.50	45.73	53.41
		meme+hum	51.91	45.14	47.76	55.38
		meme+hum+ocr	53.56	46.75	50.18	55.09
		meme+syn	55.62	47.85	51.38	54.45
		meme+syn+ocr	55.53	47.74	51.29	54.54
	4	meme	51.87	45.63	50.22	53.83
		meme+ocr	46.36	40.72	43.54	53.40
		meme+hum	49.96	44.18	47.60	55.55
		meme+hum+ocr	50.53	44.46	48.10	54.18
		meme+syn	52.45	45.78	49.53	54.70
		meme+syn+ocr	52.57	45.87	49.43	54.81
Molmo	0	meme	49.38	43.89	48.00	55.04
		meme+ocr	53.90	46.13	49.03	55.52
		meme+hum	53.66	47.71	51.30	56.43
		meme+hum+ocr	53.21	47.23	50.97	56.12
		meme+syn	51.80	45.80	49.40	55.56
		meme+syn+ocr	52.77	46.20	49.21	55.35
	1	meme	54.88	47.73	50.06	55.65
		meme+ocr	58.20	48.36	50.89	55.82
		meme+hum	57.75	50.68	54.44	57.07
		meme+hum+ocr	58.32	50.04	53.30	57.14
		meme+syn	50.51	44.61	49.13	54.24
		meme+syn+ocr	51.42	45.30	49.80	55.25
	2	meme	53.58	46.62	48.77	54.85
		meme+ocr	54.89	47.35	51.36	55.12
		meme+hum	59.84	52.15	55.45	57.88
		meme+hum+ocr	56.88	49.58	53.43	56.63
		meme+syn	52.73	46.66	51.48	55.47
		meme+syn+ocr	54.26	47.58	51.83	55.43
	3	meme	55.72	48.61	51.00	55.14
		meme+ocr	54.02	47.28	51.17	56.01
		meme+hum	60.04	52.05	54.70	57.93
		meme+hum+ocr	58.36	50.93	54.35	57.98
		meme+syn	56.16	49.14	53.04	55.84
		meme+syn+ocr	53.71	47.00	50.81	55.44
	4	meme	60.37	52.60	54.99	57.41
		meme+ocr	56.98	49.68	53.12	56.63
		meme+hum	62.74	54.24	56.65	59.35
		meme+hum+ocr	60.40	52.46	55.06	58.54
		meme+syn	58.37	51.02	54.12	57.04
		meme+syn+ocr	54.23	47.97	51.53	56.92

Table 10: VLMs‘ performance in terms of accuracy and Macro-F1 on frame detection. Hum = human caption, syn = synthetic caption, #S = number of shots.

Model	#S	Inputs	Acc.	F1	precision	recall
Mistral	0	ocr	55.03	49.79	54.96	55.65
		hum	58.72	53.35	57.91	58.97
		hum+ocr	57.73	52.46	57.30	58.47
		syn	53.52	48.26	53.95	55.06
		syn+ocr	54.70	49.30	54.48	55.68
	1	ocr	56.32	50.75	55.43	56.04
		hum	60.81	55.08	59.07	59.85
		hum+ocr	59.79	54.18	58.18	59.29
		syn	54.97	49.76	54.58	55.17
		syn+ocr	56.98	51.78	56.35	57.14
	2	ocr	59.18	52.51	55.95	56.65
		hum	62.38	56.23	59.36	60.41
		hum+ocr	62.27	55.96	58.96	60.03
		syn	56.44	50.65	55.15	55.28
		syn+ocr	58.61	52.43	56.00	56.54
	3	ocr	60.55	53.80	57.04	57.88
		hum	63.67	57.55	60.39	61.55
		hum+ocr	63.84	57.63	60.16	61.78
		syn	58.45	52.48	56.41	56.81
		syn+ocr	60.45	54.12	57.55	58.25
	4	ocr	61.71	54.79	57.59	58.68
		hum	64.61	58.31	61.00	62.31
		hum+ocr	65.09	58.78	61.06	62.94
		syn	59.03	53.01	56.42	57.48
		syn+ocr	61.78	55.20	58.10	59.12
Qwen2	0	ocr	56.60	49.47	53.29	54.49
		hum	65.07	55.32	57.85	58.63
		hum+ocr	60.82	52.94	55.81	56.99
		syn	54.24	47.82	52.28	53.33
		syn+ocr	53.69	47.41	51.66	53.06
	1	ocr	59.59	51.82	54.94	55.63
		hum	64.07	56.54	59.53	60.69
		hum+ocr	63.39	55.68	58.50	59.15
		syn	57.55	50.92	54.79	54.84
		syn+ocr	57.69	51.08	54.90	55.30
	2	ocr	62.11	54.01	56.95	57.72
		hum	64.84	57.34	59.96	61.24
		hum+ocr	64.31	56.57	58.88	60.05
		syn	58.36	52.09	55.99	57.10
		syn+ocr	60.65	53.87	57.07	58.42
	3	ocr	63.14	54.32	56.92	57.65
		hum	65.76	58.22	60.59	61.94
		hum+ocr	65.32	57.52	59.76	61.05
		syn	59.85	53.24	56.58	57.65
		syn+ocr	60.85	54.16	57.57	58.96
	4	ocr	64.02	55.45	57.99	58.80
		hum	65.86	58.23	60.66	62.18
		hum+ocr	64.98	57.51	60.01	61.12
		syn	60.33	53.98	57.27	58.26
		syn+ocr	60.88	54.24	57.45	58.45

Table 11: LLMs‘ performance in terms of accuracy and Macro-F1 on frame detection. Hum = human caption, syn = synthetic caption, #S = number of shots.