Supplementary Materials for MMAFFBen paper

Abstract

This document is the Supplementary Materials for the MMAFFBen paper, including the **performance across languages on multilingual datasets** (i.e., SAMSEMO, MMS, and XAD) and the **prompt descriptions for each dataset and each task**. The complete project is publicly released to foster transparent, reproducible, and inclusive progress in affective analysis studies and applications[1].

ACM Reference Format:

. 2025. Supplementary Materials for MMAFFBen paper. In . ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Performance Across Languages on Multilingual Datasets (Table 1 for SAMSEMO, Table 2 for MMS, Table 3 for XED.)

The results demonstrate the multilingual advantage of the MMAFFLM series, followed by GPT-4o-mini. Moreover, we observe that most models perform better on English compared to other languages. A possible reason is that the pretraining corpora of most foundation models are predominantly English-centric. Therefore, advancing multilingual research in LLMs and VLMs is both necessary and meaningful.

Table 1: Multi-lingual macro F1 on SAMSEMO.

Models	de	en	es	ko	pl
llava-nextvideo-7b	30.1	32.5	25.5	26.2	11.1
llava-nextvideo-7b-dpo	31.8	30.1	27.2	24.7	10.6
qwen2.5-vl-3b	32.1	33.6	28.8	28.9	28.4
qwen2.5-vl-7b	29.1	30.3	28.8	41.3	27.4
InternVL2.5-1B	20.6	23.3	21.5	22.0	23.2
InternVL2.5-2B	25.1	28.2	26.8	39.3	27.6
InternVL2.5-8B	36.7	38.6	32.1	49.5	31.8
InternVL2.5-1B-MPO	33.2	27.9	29.1	29.5	26.4
InternVL2.5-2B-MPO	34.1	34.2	31.6	31.3	32.6
InternVL2.5-8B-MPO	37.9	39.2	35.3	37.6	39.7
GPT-4o-mini	**42.0**	42.3	35.2	42.8	37.3
MMAFFLM-3b	36.1	52.2	36.1	46.7	**48.1**
MMAFFLM-7b	37.7	**52.7**	**38.5**	**60.5**	36.2

2 Prompt descriptions for each dataset and each task (Table 4)

[1]https://github.com/lzw108/MMAFFBen
https://huggingface.co/datasets/lzw1008/MMAFFBen/tree/main
https://huggingface.co/datasets/lzw1008/MMAFFIn/tree/main

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Table 2: Multi-lingual macro F1 on MMS.

Models	ar	bg	bs	cs	de	en	es	fa	fr	he	hi	hr	hu	it	ja	lv	pl	pt	ru	sk	sl	sq	sr	sv	th	ur	zh
EmoLlama-chat-7B	37.9	44.9	46.0	52.6	64.7	63.3	59.9	29.2	65.7	29.2	48.3	47.2	40.5	50.6	39.3	38.9	48.9	45.0	55.3	39.1	46.6	25.9	44.6	46.5	37.2	30.3	50.9
Llama3.2-1b-instruct	18.1	28.9	31.7	31.7	31.9	31.9	36.9	31.1	32.8	17.2	25.9	30.9	27.4	37.6	31.3	29.7	32.3	27.5	30.7	30.8	33.1	21.0	31.5	36.4	25.1	23.7	31.4
Llama3.2-3b-instruct	21.8	34.9	28.7	35.7	33.7	34.6	33.2	46.1	31.6	30.3	28.2	26.3	31.5	39.5	36.2	34.0	39.9	30.5	37.0	32.0	35.2	19.9	30.8	35.5	31.9	25.3	25.5
Mistral-7b-instruct	26.0	29.0	29.4	33.6	37.6	44.1	33.5	29.9	36.8	34.0	34.0	27.7	24.9	37.7	50.9	26.2	33.0	27.7	36.5	25.8	29.9	21.4	30.9	35.7	32.7	27.0	42.0
llama3.2-11B	28.3	36.1	37.0	37.6	37.9	51.3	42.2	41.5	50.4	36.1	34.0	35.9	36.5	46.3	51.0	34.1	38.8	31.2	40.4	35.2	34.4	24.4	32.5	40.1	34.7	28.6	33.0
llava-1.5-7b	26.2	29.6	32.5	33.2	37.3	34.1	37.0	29.1	46.1	29.8	33.2	30.4	27.8	40.1	36.6	26.6	36.5	28.0	38.5	30.6	30.4	20.6	33.9	36.4	22.3	24.9	31.6
llava-1.5-13b	27.2	35.1	33.8	37.3	40.9	38.3	39.4	35.2	41.4	35.1	33.7	34.4	33.3	42.5	44.2	24.1	40.9	30.2	39.1	38.1	34.2	23.6	30.9	40.5	31.9	28.6	40.3
llava-nextvideo-7b	26.3	29.1	31.0	35.5	32.4	36.9	36.1	26.1	35.2	21.5	29.6	30.8	30.4	29.8	33.5	32.5	33.4	24.6	33.0	31.0	25.2	25.2	22.3	28.7	23.5	25.7	37.3
llava-nextvideo-7b-dpo	27.1	29.7	29.3	38.2	44.0	37.6	35.3	29.3	45.2	25.4	31.9	34.5	42.9	28.2	36.2	33.9	40.8	34.5	34.7	32.8	27.3	26.1	22.5	31.0	22.9	27.7	38.5
qwen2.5-vl-3b	31.7	30.7	30.5	32.9	39.4	45.2	39.5	32.9	39.0	35.4	34.2	30.1	28.8	44.2	42.1	28.8	38.1	31.7	38.9	28.2	30.1	20.2	33.2	37.1	38.6	24.6	36.1
qwen2.5-vl-7b	26.1	29.3	24.9	30.8	35.6	38.3	35.4	28.8	33.3	32.3	30.0	26.9	26.5	34.8	36.4	26.6	29.8	23.2	36.1	28.3	26.7	18.0	26.2	32.8	30.7	20.7	33.4
InternVL2.5-1B	32.1	29.7	26.3	31.5	38.8	40.3	36.8	31.2	54.6	30.4	30.9	25.0	24.6	39.2	41.5	23.8	34.5	32.6	35.6	24.3	28.1	18.5	25.8	29.0	40.9	26.3	38.9
InternVL2.5-2B	29.5	27.4	28.0	33.7	50.7	53.8	34.1	32.0	52.8	31.3	33.0	27.4	32.9	35.2	56.8	26.4	35.5	29.8	36.9	27.9	28.6	17.9	28.8	29.3	24.9	26.7	46.3
InternVL2.5-8B	30.2	29.7	34.1	34.0	51.6	62.6	46.6	27.2	37.2	40.2	34.0	30.9	29.7	41.1	55.4	27.0	34.8	28.8	51.0	33.9	30.4	23.8	32.5	35.4	34.7	30.8	47.4
InternVL2.5-1B-MPO	28.7	25.7	25.7	36.5	45.6	34.6	35.9	29.0	49.1	35.7	28.7	22.7	24.0	36.4	53.3	23.2	30.0	41.2	31.6	22.3	24.9	20.0	27.5	29.8	43.7	26.0	33.3
InternVL2.5-2B-MPO	36.9	36.7	25.9	39.4	43.7	44.1	41.6	30.2	46.5	32.9	30.4	26.9	31.0	44.3	48.2	21.6	32.1	27.9	43.6	26.3	33.8	16.1	24.3	38.5	22.1	24.8	37.6
InternVL2.5-8B-MPO	27.5	26.6	28.8	43.2	43.8	42.0	44.8	27.2	35.6	51.1	44.1	27.2	34.0	49.2	47.5	27.8	32.9	38.0	44.3	38.3	37.8	19.2	37.0	42.2	38.0	27.6	40.3
GPT-4o-mini	52.6	**63.3**	**62.1**	60.5	**69.6**	**74.0**	**65.2**	**75.5**	**69.7**	**57.9**	**61.5**	**55.2**	**60.4**	**76.2**	**69.7**	**67.3**	**66.5**	**48.2**	**65.2**	**58.7**	**61.2**	**45.9**	**56.3**	**62.6**	**66.8**	**47.7**	59.5
MMAFFLM-3b	57.7	55.7	52.0	**61.3**	66.9	72.3	58.1	56.5	64.5	55.5	52.6	52.9	53.6	62.4	61.1	49.0	60.5	44.2	59.5	49.7	49.1	38.7	48.1	52.6	44.6	44.6	**64.0**
MMAFFLM-7b	**60.7**	53.6	47.6	59.7	63.7	69.9	57.6	51.3	67.0	50.1	52.2	55.0	48.6	62.4	61.6	44.5	59.8	44.9	59.7	52.9	51.9	44.4	49.2	53.4	40.9	45.1	62.7

Table 3: Multi-lingual macro F1 on XED.

Model	ar	bg	br	bs	cs	de	el	en	es	fi	fr	he	hr	hu	it	nl	no	pl	pt	ro	ru	sl	sr	sv	tr
EmoLlama-chat-7B	14.8	20.3	23.0	14.4	16.8	22.0	18.6	38.6	23.5	18.7	24.6	16.8	16.7	15.7	23.7	20.3	19.6	21.2	24.6	20.8	22.4	14.9	15.0	19.1	9.0
Llama3.2-1b-instruct	17.8	19.7	20.0	18.9	18.5	18.6	17.8	30.3	21.1	15.5	18.3	14.2	18.6	19.0	21.2	16.1	18.2	20.1	19.2	18.6	15.0	17.0	16.6	16.2	12.1
Llama3.2-3b-instruct	9.8	6.7	6.4	7.7	7.5	7.3	9.8	19.2	9.0	8.0	9.5	11.6	9.4	7.7	10.4	9.0	8.5	7.2	4.1	10.4	8.8	7.7	12.5	5.8	7.1
Mistral-7b-instruct	22.2	22.3	24.1	21.6	22.5	25.5	19.4	31.7	25.2	22.6	23.1	22.2	22.1	22.4	23.2	21.3	22.8	24.5	24.7	24.1	25.1	22.7	22.1	23.9	16.7
llama3.2-11B	18.6	21.2	19.0	18.1	19.1	20.4	23.9	31.7	23.4	19.3	20.3	21.5	17.7	15.6	21.5	19.2	17.1	19.3	20.6	18.1	22.1	14.3	16.4	19.7	16.0
llava-1.5-7b	21.1	22.7	23.8	20.3	21.7	24.9	19.5	33.0	25.4	23.6	24.8	22.0	21.9	20.8	24.2	22.5	23.2	20.8	25.2	22.8	24.8	20.3	20.8	25.2	17.3
llava-1.5-13b	17.9	24.2	24.4	18.3	20.8	26.2	18.9	30.9	25.3	24.3	25.0	21.5	19.5	19.3	21.7	20.9	24.8	21.0	21.2	21.8	25.7	20.0	19.4	25.4	14.6
llava-nextvideo-7b	19.5	18.3	18.2	15.3	18.0	18.1	16.8	26.3	20.4	16.7	18.5	16.8	14.9	17.8	18.2	16.6	17.1	16.2	16.4	17.7	22.1	15.0	16.5	17.9	13.6
llava-nextvideo-7b-dpo	20.2	17.7	18.8	17.4	18.8	19.4	18.8	28.2	21.4	16.3	19.9	16.9	19.0	17.8	20.7	17.8	17.1	16.6	22.0	17.7	20.6	17.4	15.3	19.2	15.6
qwen2.5-vl-3b	14.0	11.0	14.7	10.2	10.2	12.2	13.1	23.3	14.9	10.9	17.4	14.7	10.5	11.3	17.1	12.6	13.5	14.7	15.1	14.2	17.8	8.6	10.3	13.7	9.9
qwen2.5-vl-7b	12.8	13.2	12.8	9.1	9.5	17.1	12.1	26.4	16.0	12.0	14.8	14.7	9.0	7.2	14.3	11.8	10.1	10.3	12.2	11.8	18.4	8.0	10.0	10.2	8.9
InternVL2.5-1B	21.2	19.0	21.2	16.5	17.7	20.4	12.9	17.7	17.5	15.3	18.4	20.4	16.3	18.2	19.6	19.9	16.1	17.5	21.7	20.4	20.8	15.4	15.8	19.1	14.4
InternVL2.5-2B	4.4	4.8	5.8	2.7	4.5	5.3	5.0	12.9	5.2	5.9	5.5	5.5	1.2	1.6	3.4	3.2	2.5	5.2	5.0	4.8	6.1	2.4	1.2	4.0	2.6
InternVL2.5-8B	7.4	8.0	11.4	6.6	6.9	10.1	7.7	24.0	12.6	11.0	11.9	7.2	6.9	6.7	9.1	8.0	9.6	6.7	15.1	10.5	10.8	5.3	3.5	10.2	9.0
InternVL2.5-1B-MPO	24.1	22.2	22.4	21.3	21.5	21.7	19.5	25.1	22.3	16.6	20.5	23.5	19.1	18.3	21.7	21.6	19.5	19.9	23.7	20.3	23.9	18.8	19.8	19.8	16.0
InternVL2.5-2B-MPO	11.2	11.8	11.1	10.7	10.0	13.9	9.7	15.1	9.3	11.5	10.8	12.5	10.5	9.3	8.8	9.3	8.5	11.4	14.1	10.1	17.0	7.6	9.1	10.6	9.2
InternVL2.5-8B-MPO	7.9	9.5	13.8	9.8	12.2	13.9	9.8	27.3	17.5	14.0	14.3	6.9	11.2	10.9	14.1	8.4	13.2	9.5	17.9	12.6	11.5	8.0	5.0	14.0	11.1
GPT-4o-mini	18.6	21.7	18.0	19.6	21.1	22.7	19.8	34.1	23.0	24.0	19.9	22.9	18.6	18.3	20.6	19.3	21.9	20.0	21.8	21.0	20.6	16.7	21.9	22.5	17.1
MMAFFLM-3b	**27.2**	26.7	**26.0**	23.3	**37.3**	**28.5**	**35.4**	**36.8**	**29.0**	**37.7**	**26.7**	**35.0**	28.4	**31.9**	26.7	**26.3**	**26.6**	26.8	**28.4**	**25.7**	29.6	22.9	23.6	**27.1**	20.2
MMAFFLM-7b	24.2	**33.7**	21.6	**31.2**	24.3	25.4	31.0	29.9	24.6	32.4	25.5	23.4	**30.1**	19.1	**36.1**	23.2	23.2	**32.3**	23.3	20.3	**37.7**	**29.0**	**31.1**	24.1	**32.2**

Table 4: Prompt descriptions for each dataset and each task

Dataset	Prompt
SemEva2018-EI [4]	Task: Assess the magnitude of emotion E in the text using a real number between 0 and 1, where 0 denotes the least intensity and 1 denotes the most intensity. Text: raw text. Emotion E: anger. Intensity Score:
SemEva2018-EC (multi) [4]	Task: Determine the text's emotional tone based on 11 categories: 1. joy, 2. sadness, 3. anger, 4. fear, 5. surprise, 6. disgust, 7. anticipation, 8. love, 9. optimism, 10. pessimism, and 11. trust. You have two options: (a) Select one or more of the 11 emotions that represent the emotional state depicted in the text. (b) Choose '0. neutral' if the image does not express any emotion. Please provide your selection in the following format: number. emotion. Text: raw text
SemEva2018-SP (7) [4]	Task: Classify the text into one of seven ordinal classes, corresponding to various levels of positive and negative sentiment intensity, that best represents the mental state of the publisher. 3: very positive mental state can be inferred. 2: moderately positive mental state can be inferred. 1: slightly positive mental state can be inferred. 0: neutral or mixed mental state can be inferred. -1: slightly negative mental state can be inferred. -2: moderately negative mental state can be inferred. -3: very negative mental state can be inferred Text: raw text. Intensity Class:
SemEva2018-SI [4]	Task: Determine the sentiment (valence) intensity of the publisher's mental state on a scale of -1 (most negative) to 1 (most positive). Text: raw text. Intensity Score:
EWECT-usual-EC (single)	Task: Determine the emotion of the text based on five discrete categories: 1. happiness, 2. sadness, 3. anger, 4. fear, and 5. surprise. You have two options: (a) Select one emotion that best represents the emotional state conveyed in the text. (b) Choose '0. neutral' if the text does not express any emotion. Please provide your selection in the following format: number. emotion. Text: raw text
EWECT-virus-EC (single)	Task: Determine the emotion of the text based on five discrete categories: 1. happiness, 2. sadness, 3. anger, 4. fear, and 5. surprise. You have two options: (a) Select one emotion that best represents the emotional state conveyed in the text. (b) Choose '0. neutral' if the text does not express any emotion. Please provide your selection in the following format: number. emotion. Text: raw text
Onlineshopping-SP (2)	Task: Determine the text's sentiment polarity. You can select one of the following labels: -1. negative, and 1. positive. Text: raw text
XED-EC (multi) [5]	Task: Determine the emotion of the text based on eight discrete categories: 1. joy, 2. sadness, 3. anger, 4. fear, 5. surprise, 6. disgust, 7. anticipation, 8. trust. You have two options: (a) Select one or more of the 8 emotions that represent the emotional state depicted in the text. (b) Choose '0. neutral' if the text does not express any emotion. Please provide your selection in the following format: number. emotion. Text: raw text
MMS-SP (3) [8]	Task: Determine the text's sentiment polarity. You can select one of the following labels: -1. negative, 0. neutral, and 1. positive. Text: raw text
EMOTIC-EC (multi) [3]	<image>Task: Determine the emotion(s) of the image based on a set of 26 discrete categories using the image information. The list of emotions includes: 1. Happiness: feeling delighted; feeling enjoyment or amusement. 2. Sadness: feeling unhappy, sorrow, disappointed, or discouraged. 3. Anger: intense displeasure or rage; furious; resentful. 4. Fear: feeling suspicious or afraid of danger, threat, evil or pain; horror. 5. Surprise: sudden discovery of something unexpected. 6. Aversion: feeling disgust, dislike, repulsion; feeling hate. 7. Excitement: feeling enthusiasm; stimulated; energetic. 8. Peace: well being and relaxed; no worry; having positive thoughts or sensations; satisfied. 9. Affection: fond feelings; love; tenderness. 10. Annoyance: bothered by something or someone; irritated; impatient; frustrated. 11. Anticipation: state of looking forward; hoping on or getting prepared for possible future events. 12. Confidence: feeling of being certain; conviction that an outcome will be favorable; encouraged; proud. 13. Disapproval: feeling that something is wrong or reprehensible; contempt; hostile 14. Disconnection: feeling not interested in the main event of the surrounding; indifferent; bored; distracted. 15. Disquietment: nervous; worried; upset; anxious; tense; pressured; alarmed. 16. Doubt/Confusion: difficulty to understand or decide; thinking about different options. 17. Embarrassment: feeling ashamed or guilty 18. Engagement: paying attention to something; absorbed into something; curious; interested. 19. Esteem: feelings of favourable opinion or judgement; respect; admiration; gratefulness. 20. Fatigue: weariness; tiredness; sleepy. 21. Pain: physical suffering. 22. Pleasure: feeling of delight in the senses. 23. Sensitivity: feeling of being physically or emotionally wounded; feeling delicate or vulnerable. 24. Suffering: psychological or emotional pain; distressed; anguished. 25. Sympathy: state of sharing others emotions, goals or troubles; supportive; compassionate. 26. Yearning: strong desire to have something; jealous; envious; lust. You can select one or more emotions from the list to represent the image's emotional expression.

Dataset	Prompt
EMOTIC-SI [3]	<image>Task: Determine the valence, arousal, and dominance of the image using the image information. The descriptions of these dimensions are as follows: Valence: Measures how positive or pleasant an emotion is, ranging from negative to positive. Arousal: Measures the agitation level of the person, ranging from calm/non-active to agitated/ready to act. Dominance: Measures the level of control a person feels in the situation, ranging from submissive/non-control to dominant/in-control. Assign a real-valued score between -1 and 1 for each category: -1: Indicates the most negative, calm/non-active, or submissive/non-control. 1.0: Represents the most positive, agitated/ready to act, or dominant/in-control. Use the following template to report the intensity of each category: valence: intensity, arousal: intensity, dominance: intensity.
FER2013-EC (single)	<image>Task: Determine the image's emotion using Ekman's categories: 1. happiness, 2. sadness, 3. anger, 4. fear, 5. surprise, and 6. disgust. You have two options: (a) Select one Ekman label that best represents the emotional state depicted in the image. (b) Choose '0. neutral' if the image does not express any emotion. Please provide your selection in the following format: number. emotion.
CFAPS-EC (single) [2]	<image>Task: Determine the image's emotion using Ekman's categories: 1. happiness, 2. sadness, 3. anger, 4. fear, 5. surprise, and 6. disgust. You have two options: (a) Select one Ekman label that best represents the emotional state depicted in the image. (b) Choose '0. neutral' if the image does not express any emotion. Please provide your selection in the following format: number. emotion.
CFAPS-EI [2]	<image>Task: Assess the intensity of emotion E in the image using a real number between 0 and 1, where 0 represents the least intensity and 1 represents the greatest intensity. Emotion E: surprise.
SANSEMO-EC (multi) [1]	<video>Task: Determine the video's emotion based on its content (text transcribed from the video's audio) and video information, using Ekman's categories: 1. happiness, 2. sadness, 3. anger, 4. fear, 5. surprise, and 6. disgust. You have three options: (a) Select up to two Ekman labels for a given scene. (b) Choose '0. neutral' if the video does not express any emotion. (c) Select '99. other emotions' if the basic six Ekman emotions are inadequate. Please provide your selection in the following format: number. emotion. Content: transcription
CHSIMS-SP (3) [7]	<video>Task: Determine the video's sentiment polarity according to its content (the text transcribed from the video's audio) and video information. You can select one of the following labels: -1. negative, 0. neutral, and 1. positive. Content: transcription
CHSIMS-SI [7]	<video>Task: Determine the video's sentiment strength according to its content (the text transcribed from the video's audio) and video information. Choose one of the following values: -1.0, -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0, where -1.0 indicates the most negative sentiment, 0.0 represents neutral, and 1.0 indicates the most positive sentiment. Content: transcription
MELD-EC (single) [6]	<video>Task: Determine the video's emotion based on its content (text transcribed from the video's audio) and video information , using Ekman's categories: 1. happiness, 2. sadness, 3. anger, 4. fear, 5. surprise, and 6. disgust. You have two options: (a) Select one Ekman label that best represents the emotional state depicted in the video. (b) Choose '0. neutral' if the video does not express any emotion. Please provide your selection in the following format: number. emotion. Content: transcription
MELD-SP (3) [6]	<video>Task: Determine the video's sentiment polarity according to its content (the text transcribed from the video's audio) and video information. You can select one of the following labels: -1. negative, 0. neutral, and 1. positive. Content: transcription

References

[1] Paweł Bujnowski, Bartłomiej Kuzma, Bartłomiej Paziewski, Jacek Rutkowski, Joanna Marhula, Zuzanna Bordzicka, and Piotr Andruszkiewicz. 2024. SAMSEMO: New dataset for multilingual and multimodal emotion recognition. In *Proc. Interspeech 2024*. 2925–2929.

[2] Xu Gong, Yu-Xia Huang, Yan Wang, and Yue-jia Luo. 2011. Revision of the Chinese facial affective picture system. *Chinese mental health journal* (2011).

[3] Ronak Kosti, Jose M Alvarez, Adria Recasens, and Agata Lapedriza. 2020. Context based emotion recognition using emotic dataset. *arXiv preprint arXiv:2003.13401* (2020).

[4] Saif M. Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 Task 1: Affect in Tweets. In *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*. New Orleans, LA, USA.

[5] Emily Öhman, Marc Pàmies, Kaisla Kajava, and Jörg Tiedemann. 2020. XED: A Multilingual Dataset for Sentiment Analysis and Emotion Detection. In *Proceedings of the 28th International Conference on Computational Linguistics*. 6542–6552.

[6] Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 527–536.

[7] Wenmeng Yu, Hua Xu, Fanyang Meng, Yilin Zhu, Yixiao Ma, Jiele Wu, Jiyun Zou, and Kaicheng Yang. 2020. Ch-sims: A chinese multimodal sentiment analysis dataset with fine-grained annotation of modality. In *Proceedings of the 58th annual meeting of the association for computational linguistics*. 3718–3727.

[8] Łukasz Augustyniak, Szymon Woźniak, Marcin Gruza, Piotr Gramacki, Krzysztof Rajda, Mikołaj Morzy, and Tomasz Kajdanowicz. 2023. Massively Multilingual Corpus of Sentiment Datasets and Multi-faceted Sentiment Classification Benchmark. arXiv:2306.07902 [cs.CL]