Does the Emotional Understanding of LVLMs Vary Under High-Stress Environments and Across Different Demographic Attributes?

Jaewook Lee[1,2*] **Yeajin Jang**[1*] **Oh-Woog Kwon**[2] **Harksoo Kim**[1†]

[1]Konkuk University, Republic of Korea
[2]Electronics and Telecommunications Research Institute, Republic of Korea
{benecia428, yaezinny95}@gmail.com ohwoog@etri.re.kr nlpdrkim@konkuk.ac.kr

Abstract

According to psychological and neuroscientific research, a *high-stress* environment can restrict attentional resources and intensify negative affect, thereby impairing the ability to understand emotions. Furthermore, *demographic attributes* such as race, gender, and age group have been repeatedly reported to cause significant differences in emotional expression and recognition. This study is the first to systematically verify whether these psychological findings observed in humans also apply to the latest Large Vision Language Models (LVLMs). We constructed low-stress versus high-stress environments and generated an image dataset (a total of 540 images) that combines race, gender, and age group. Based on this, we applied the *Pretend prompt* technique to induce LVLMs to interpret others' emotions from the standpoint of the assigned environment and persona. An analysis of the models' emotional understanding ability, using EQ-Bench-based metrics, revealed that (1) under high-stress environments, the accuracy of emotion understanding significantly declined in most LVLMs, and (2) performance disparities were confirmed across race, gender, and age group. These findings suggest that the effects of *high-stress* and *demographic attributes* identified in human research may also be reflected in LVLMs.

1 Introduction

Emotion is a complex phenomenon formed through the interaction of subjective feelings, cognitive evaluations, and physiological responses, playing a crucial role in an individual's interpersonal relationships, stress responses, and overall decision-making processes (Salovey and Mayer, 1990; Schutte et al., 1998). In this context, emotional understanding refers to the ability to accurately perceive and regulate one's own emotions and to identify and interpret the emotions of others. From a psychological viewpoint, it has long been treated as a core research topic (Goleman, 1996; Schutte et al., 2001, 2002).

Recently, Large Language Models (LLMs) have shown potential in a wide range of areas—such as text-based sentiment analysis (Kadiyala, 2024; Liu et al., 2024c; Zhang et al., 2023), generation of affective or empathetic responses (Li et al., 2024b; Sotolar et al., 2024; Loh and Raamkumar, 2023), and conversational assistance (Yoran et al., 2024; Guan et al., 2023; Dhole et al., 2023)—by interpreting and responding to humans' emotional cues. Moreover, with the emergence of Large Vision Language Models (LVLMs) (Liu et al., 2024b; Zhu et al., 2023), which can process both text and images, the level of contextual understanding and emotional comprehension leveraging visual information is expected to be even more advanced (Lee et al., 2024; Poria et al., 2018; Busso et al., 2008). However, most existing research has tended to focus on evaluating a basic "emotional understanding ability" or has diagnosed model performance without sufficiently controlling for complex environmental or demographic factors.

Meanwhile, studies in psychology and neuroscience have repeatedly noted that an individual's emotions are not solely determined by internal factors but can vary significantly according to the *environmental context* in which the individual is placed (Barrett, 2017; Mesquita et al., 2010). In particular, a *high-stress environment* can limit one's attentional resources and intensify negative emotions, thereby undermining prefrontal-cortex-based cognitive functions (Starcke and Brand, 2016; LeBlanc et al., 2012; Qin et al., 2009; Arnsten, 2009; Evans and Lepore, 1993; Cohen et al., 1980), and is reported to increase the likelihood of misinterpreting others' emotions or of reduced empathy (Gamble et al., 2023; Ruffman et al., 2008; Lough et al., 2006). On the other hand, in a *low-stress environ-*

[*]Main contributors
[†]Corresponding author

Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 23196–23210

ment, an individual can maintain and expand positive psychological resources in a relatively stable manner, potentially leading to smoother emotion recognition and interpretation (Fredrickson, 2004, 2001; Isen, 2001). Thus, environmental factors are regarded as key variables that influence not only momentary emotional reactions but also the entirety of emotional understanding and empathy processes.

Additionally, demographic attributes (e.g., race, gender, and age group) constitute another major factor that causes significant differences in emotional expression and interpretation. Previous studies have shown that emotion regulation strategies and approaches to emotional interpretation vary by age group (Carstensen et al., 2011; Urry and Gross, 2010; Charles et al., 2003; Scheibe and Carstensen, 2010), that differences in norms and frequencies of emotional expression or suppression exist between genders (Brody, 2008; Fischer and Manstead, 2000; Fischer and LaFrance, 2015; Chaplin and Aldao, 2013), and that, depending on culture and race, the very act of emotional expression changes under social rules such as "display rules" (Markus and Kitayama, 2014; Safdar et al., 2009; Mesquita, 2003). Such characteristics have direct implications for the accuracy and fairness of emotional understanding and raise the possibility of unexpected biases or misinterpretations in models.

A large body of research, therefore, has repeatedly highlighted that "the level of environmental stress an individual is exposed to" and "the demographic attributes an individual possesses" can considerably alter one's capacity for emotional understanding. However, it remains unclear whether these findings, which have been amassed in *human* contexts, are equally applicable to LVLMs. In other words, the question, "When a *high-stress environment* is 'assigned' to an LVLM, and when *demographic attributes* are varied in the model, what kinds of differences and biases arise in the process of interpreting others' emotions?" has yet to be sufficiently explored.

In order to address these questions, this study systematically evaluates how *environmental stress* and *demographic attributes* actually affect the emotional understanding ability of LVLMs. To this end, we set out the following three research questions:

RQ1: How do low-stress versus high-stress conditions affect the emotional understanding of LVLMs?

RQ2: How do demographic attributes (race, gender, and age group) impact bias and performance in emotional understanding?

RQ3: What combinations of stress and demographic attributes cause the most severe biases and performance drops in emotional understanding?

Drawing upon the emotional understanding evaluation benchmarks and the *Pretend prompt* technique suggested in (Paech, 2023; Cheng et al., 2023; Fraser and Kiritchenko, 2024), we directly construct an additional synthetic image dataset to assess the model's emotional understanding ability. Specifically, we (1) set low-stress vs. high-stress environments; (2) systematically combine demographic attributes and generate synthetic images; and (3) use these images to ensure that LVLMs clearly recognize the environment and demographic attributes they have been assigned, and then interpret others' emotions under those conditions. By incorporating both environmental and demographic factors—while building on the benchmarks and prompts introduced in previous studies—this approach aims to offer a meaningful extension in comprehensively analyzing the emotional understanding characteristics of LVLMs.

2 Related Works

The Relationship Between Emotional Understanding Ability and Stressful Environments. High levels of stress have been reported to limit attentional resources and intensify negative affect, ultimately undermining cognitive functions based in the prefrontal cortex (Starcke and Brand, 2016; LeBlanc et al., 2012; Qin et al., 2009; Arnsten, 2009; Evans and Lepore, 1993; Cohen et al., 1980). For instance, a meta-analysis by Starcke and Brand (2016) showed that in stress-induced laboratory conditions, there was an increase in reward-seeking and risk-taking behaviors, alongside a significant decline in overall decision-making performance. Likewise, LeBlanc et al. (2012) investigated emergency medical personnel in real high-stress situations and confirmed that higher levels of psychological and physiological stress responses corresponded to reduced accuracy and efficiency in clinical judgment. In this way, stress has been shown to disrupt the optimal functioning of the prefrontal cortex network during cognitively demanding tasks (e.g., working memory, decision-making, and attention) (Qin et al., 2009; Arnsten, 2009). Moreover,

research indicates that uncontrollable stressors such as chronic noise can interfere with learning and attention regulation in children and young people (Evans and Lepore, 1993; Cohen et al., 1980).

Meanwhile, a weakening of cognitive abilities, including reduced prefrontal cortex function, has been reported to negatively affect the capacity to accurately perceive and interpret others' emotional states. For example, in a study of patients with frontotemporal dementia, Lough et al. (2006) demonstrated a close link between executive function and the ability to recognize emotions, and a meta-analysis also supports the finding that emotional recognition performance in older adults declines in tandem with general cognitive deterioration (Ruffman et al., 2008). More recently, there have been discussions suggesting that cognitive load itself may diminish empathy and prosocial behavior, a phenomenon observed not only in laboratory settings but also in real-world contexts (e.g., during the COVID-19 pandemic) (Gamble et al., 2023). Taken together, when cognitive burden increases, the ability to accurately notice or interpret another person's emotional cues weakens, and stressful environments may exacerbate this cognitive load, further hindering emotional understanding. Building on these existing findings, the present study systematically investigates how high-stress environments affect the emotional understanding ability of LVLMs.

The Relationship Between Emotional Understanding Ability and Demographic Attributes. In the field of emotion research, numerous studies have pointed out that an individual's demographic attributes can cause significant variations in emotional experience, expression, and interpretation. For instance, the way emotions are regulated and interpreted differs depending on age group (Carstensen et al., 2011; Urry and Gross, 2010; Charles et al., 2003), and differences in norms and behaviors related to expressing or suppressing emotions lead to consistently reported gender gaps in both positive and negative emotional expression frequencies and strategies (Brody, 2008; Fischer and LaFrance, 2015; Chaplin and Aldao, 2013). Cultural factors also play a role in determining how emotions are expressed or suppressed through social rules, typified by "display rules," showing varied patterns when comparing individualistic and collectivistic cultural spheres (Fischer and Manstead, 2000; Markus and Kitayama, 2014; Safdar et al., 2009; Mesquita, 2003).

Because demographic attributes such as age group, gender, and cultural background have a complex influence on the emotional processing system, there is a high likelihood that LVLMs will exhibit differential outcomes or biases in emotional understanding and interpretation when those attributes are *assigned* to them. By systematically examining the existence and nature of these biases, this study aims to clarify how the combination of a stressful environment and demographic attributes is reflected in the emotional understanding abilities of LVLMs.

3 Methodology

Based on the previously reviewed research background and objectives, this study aims to systematically elucidate how the combination of environmental stress levels (*low-stress* vs. *high-stress*) and demographic attributes (race, gender, and age group) affects the emotional understanding ability of LVLMs. Below, we describe in sequence the research design, the construction of the image dataset, and the model response generation and configuration procedures.

3.1 Research Design

Design Overview. To evaluate how LVLMs' emotional understanding changes according to the level of environmental stress and demographic attributes, this study employed a 2(*environment*: low-stress vs. high-stress) $\times$ 18(*demographic attributes*: 3 races $\times$ 2 genders $\times$ 3 age groups) factorial design. This design was intended to systematically analyze: **RQ1**, which addresses differences in emotional understanding ability between low-stress and high-stress environments; **RQ2**, which concerns potential biases in emotional understanding depending on race, gender, and age group; and **RQ3**, which examines the interactions between environment and demographic attributes.

Environmental Stress Factors. Environmental factors were composed of two types: a *low-stress* environment (e.g., a serene beachfront) and a *high-stress* environment (e.g., a war-torn street). As suggested by previous research, a *high-stress* environment restricts attentional resources and intensifies negative affect, thereby causing abnormalities in prefrontal-based cognitive functions and, as a result, misinterpretation of others' emotions or diminished empathy. We aimed to verify these findings

in the context of LVLMs. By comparing how the model interprets others' emotions in environments with different stress levels—even if they share the same fundamental scenario—this study could more precisely investigate **RQ1**.

Demographic Attribute Factors. Demographic attributes were manipulated by distinguishing *race* (*Asian, Black, White*), *gender* (*male, female*), and *age group* (*Child, Young, Elderly*). This approach was taken to determine whether performance degradation or bias occurs in specific groups (**RQ2**), building on prior research indicating that cultural or biological factors can alter how emotions are expressed and perceived. Moreover, to examine whether these *demographic attributes*, when combined with a *high-stress environment*, further amplify or mitigate the emotional understanding ability of the model, we evaluated the interaction effect of these two factors simultaneously (**RQ3**).

Pretend Prompt Technique. This study employed the *Pretend prompt* technique proposed by (Cheng et al., 2023; Fraser and Kiritchenko, 2024), instructing the LVLMs to assume they are *directly experiencing* particular environmental and demographic attributes. For instance, the model was guided to interpret others' emotions from the perspective of "a young White female in a war-torn street" or to reenact "the point of view of a young White male on a serene beachfront." By doing so, we could observe in detail how the model demonstrates its emotional understanding ability in a situation combining *environmental context* and *demographic attributes*. The methodology is illustrated in Figure 1.

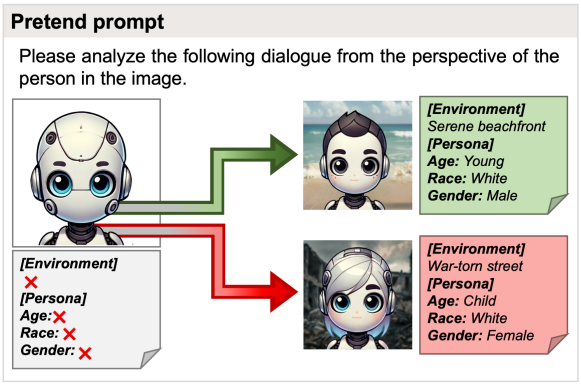

Figure 1: An example of using the *Pretend prompt* technique to guide analysis of a conversation from the perspective of a particular environment and persona.

3.2 Image Dataset Construction

Environmental Scenario Design. The image dataset was synthesized using Midjourney (version 6). First, for both the *low-stress* environment and the *high-stress* environment, 15 sub-scenarios were selected (e.g., "peaceful countryside road," "war-torn street"), and a total of 30 background images were generated (see Appendix C for the detailed list).

Application of Demographic Attribute Combinations. For each sub-scenario, we combined *race* (3 types) × *gender* (2 types) × *age group* (3 types), thereby generating a total of 18 person images. During this process, the background and objects remained identical, with only the demographic attributes of the person altered, ensuring that, within the same environment, only demographic differences were highlighted. Consequently, we obtained a total of 540 synthesized images by combining 270 for *low-stress* and 270 for *high-stress*.

Dataset Verification. The generated images underwent an initial review by the research team, and images deemed low-quality or misaligned with the research aim were regenerated. In addition, by using Midjourney's *variation(region)* function to refine the images such that only race, gender, and age group changed in otherwise identical backgrounds, we were able to more precisely implement images in which the *environment* was fixed while only *demographic attributes* varied.

3.3 Emotion Intelligence Assessment Using EQ-Bench

Overview of EQ-Bench. To evaluate LVLMs' emotional understanding ability, we utilized EQ-Bench (Paech, 2023). EQ-Bench provides a conversation-style dataset that presents the task of predicting the intensity of emotions via interactions between speakers, allowing for more nuanced and complex evaluations of emotional states than traditional multiple-choice methods. In this study, we maintained EQ-Bench's basic structure but modified or added prompt content to incorporate environmental stress and persona information.

EU Score Calculation Method. The computation of EU scores based on EQ-Bench involves three main steps: *(1) deriving the difference between the predicted emotional intensity and the reference emotional intensity*, *(2) transforming via*

an S-shaped scaling function, and *(3) applying a correction coefficient to produce the final score*.

First, let us denote the predicted intensity of each emotion by the model as $\{\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4\}$ and the reference emotional intensity as $\{y_1, y_2, y_3, y_4\}$. Then, the difference d_i between the predicted intensity $\hat{y}_i$ and the reference y_i for each emotion i is defined as

$$d_i = |\hat{y}_i - y_i| . \tag{1}$$

Next, for d_i, the following S-shaped scaling function is applied to derive δ_i:

$$\delta_i = \begin{cases} 0, & \text{if } d_i = 0, \\ 6.5 \cdot \dfrac{1}{1 + e^{-1.2(d_i - 4)}}, & \text{if } 0 < d_i \leq 5, \\ d_i, & \text{if } d_i > 5. \end{cases} \tag{2}$$

Then, summing δ_i over the four emotions yields

$$D = \sum_{i=1}^{4} \delta_i, \tag{3}$$

to which we apply the correction coefficient $\alpha = 0.7477$. The final EU score, EU_{score}, is computed as follows:

$$\text{EU}_{\text{score}} = 10 - \alpha \times D. \tag{4}$$

Finally, after calculating the score for all items, the mean value is taken as the EU score to assess how well the model understands others' emotions.

4 Experiments and Results

Previously, we highlighted the potential influence and possibility of bias that *environmental stress* and *demographic attributes* can exert on emotional understanding, raising the question of how stably LVLMs can interpret others' emotions in a high-stress environment. Building on this awareness, this chapter reports a series of experiments in which we separately construct a low-stress environment and a high-stress environment, combine various demographic attributes of race, gender, and age group to generate synthetic images, and then design the model to assume that it is *"directly experiencing"* each situation. Through this, we systematically verify the impact of environment and demographic attributes on emotional understanding and examine in detail which biases may occur. Further details regarding the prompts and model settings can be found in Appendix A.

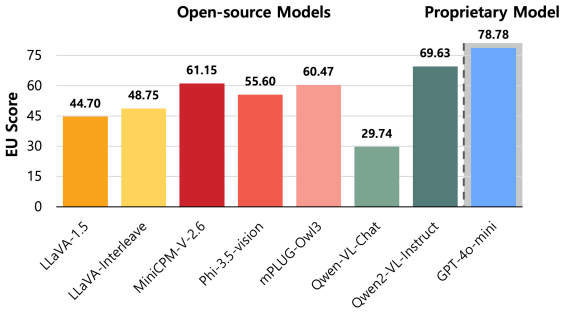

Figure 2: Comparison of EU scores for open-source and proprietary LVLMs. The latest proprietary model exhibits relatively high EU scores.

4.1 Baselines

To confirm how sensitively LVLMs respond to environmental stress and demographic attributes, and what biases they exhibit, we performed experiments on a wide range of open-source and proprietary models. Specifically, we included **LLaVA-1.5** (Liu et al., 2024a), **LLaVA-Interleave** (Li et al., 2024a), **Qwen-VL-Chat** (Bai et al., 2023), **Qwen2-VL-Instruct** (Wang et al., 2024), **MiniCPM-V-2.6** (Yao et al., 2024), **Phi-3.5-Vision** (Abdin et al., 2024), **mPLUG-Owl3** (Ye et al., 2024), and **GPT-4o-mini**, paying attention to the differences in training methods and architectures for each model. This diversity helped capture a broad spectrum of emotional understanding characteristics that LVLMs exhibit under various combinations of environmental and demographic attributes.

4.2 Overall Performance

First, we examined the overall EU scores that integrate all environmental and demographic conditions (Figure 2). Qwen-VL-Chat and LLaVA-1.5, which are relatively earlier models, showed lower EU scores at 29.74 and 44.70, respectively, whereas the latest model GPT-4o-mini recorded a dominant performance of 78.78. This suggests that in addition to technological advances, more recent models have enhanced emotional understanding ability. Among open-source models, Qwen2-VL-Instruct showed a marked improvement in performance compared to Qwen-VL-Chat, indicating that updates to training methods within the same architecture family can contribute to gains in emotional intelligence.

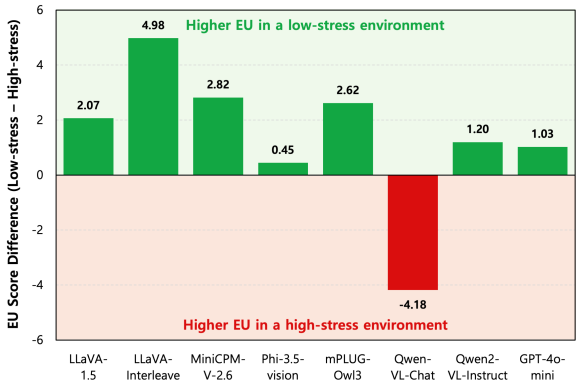

Figure 3: Difference in EU scores (in percentages) between low-stress and high-stress environments. The more positive (+) the bar is, the better the performance in a low-stress environment, whereas a negative (–) value indicates better performance in a high-stress environment.

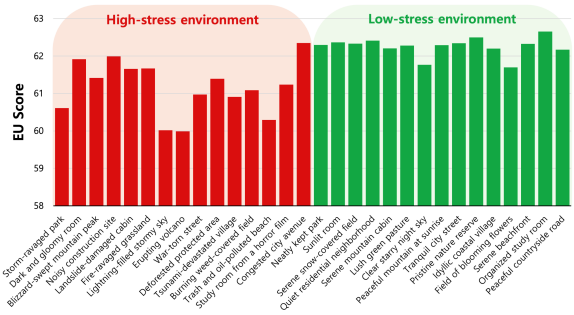

Figure 4: Comparison of EU scores in a low-stress environment (green) and a high-stress environment (red).

4.3 RQ1: Impact of Environmental Stress on Emotional Understanding

Comparing LVLM performance by environmental factor. Figure 3 summarizes the difference in EU scores between low-stress and high-stress environments, showing that most LVLMs record higher scores in a low-stress environment. This aligns with existing studies (Starcke and Brand, 2016; LeBlanc et al., 2012; Qin et al., 2009; Arnsten, 2009; Evans and Lepore, 1993; Cohen et al., 1980), which indicate that attentional resources are limited and negative emotions are intensified in a high-stress environment, leading to weakened prefrontal cortex–based cognitive functions (Gamble et al., 2023; Ruffman et al., 2008; Lough et al., 2006) and, as a result, a diminished ability to accurately perceive and interpret others' emotions. This finding suggests that stressful stimuli can negatively affect emotional understanding ability not only in humans but also in LVLMs.

Detailed comparison of low-stress vs. high-stress environments. Figure 4 visually presents the average EU performance in each individual environment. For instance, "Organized study room," "Sunlit room," and "Serene snow-covered field" are examples of environments with a stable and positive atmosphere, where generally higher scores were observed. In contrast, in threatening and chaotic situations such as an "Erupting volcano" or a "Lightning-filled stormy sky," the scores dropped markedly, which is consistent with previous psychological literature suggesting that a "stable environment induces positive affect, thereby facilitating the processing of emotional information" (Fredrickson, 2004, 2001; Isen, 2001), indicating a similar tendency in LVLMs.

Exceptional model cases. While most models perform well in low-stress environments, Qwen-VL-Chat exhibited higher scores in a *high-stress environment*. A more detailed analysis of this is provided in Appendix D.

Conclusion for RQ1. Taken together, the decline in emotional understanding ability in a high-stress environment was confirmed in most LVLMs. This is in agreement with human studies stating that "extreme stress impairs emotional understanding," suggesting that environmental factors play a crucial role in emotional understanding for AI models as well. However, exceptional patterns observed in a few models highlight the need for detailed verification of each model's architecture and the characteristics of its training data.

4.4 RQ2: Biases and Performance Disparities by Demographics

Effect of racial persona. Figure 5 presents the results when applying a racial persona. Overall, when a 'White' persona was assigned, the highest EU scores were observed in all environmental scenarios, whereas the 'Black' persona showed lower scores under both low-stress and high-stress conditions. Notably, the consistent performance drop observed when using a 'Black' persona suggests that structural bias in the training data may be embedded in the model. This finding can also be understood in the context of previous studies indicating that cultural or racial backgrounds can shape norms and interpretations of emotion expression in different ways (Fischer and Manstead, 2000; Markus and Kitayama, 2014; Safdar et al., 2009; Mesquita, 2003). In other words, there is a possibility that

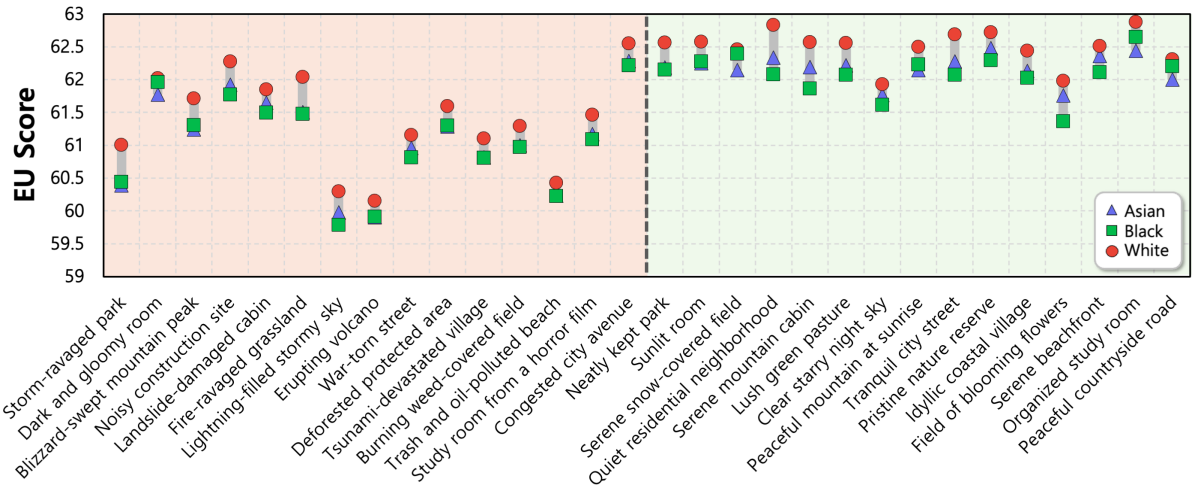

Figure 5: Comparison of LVLMs' EU performance according to race persona (Asian, Black, and White). The purple triangle represents Asian, the green square represents Black, and the red circle represents White.

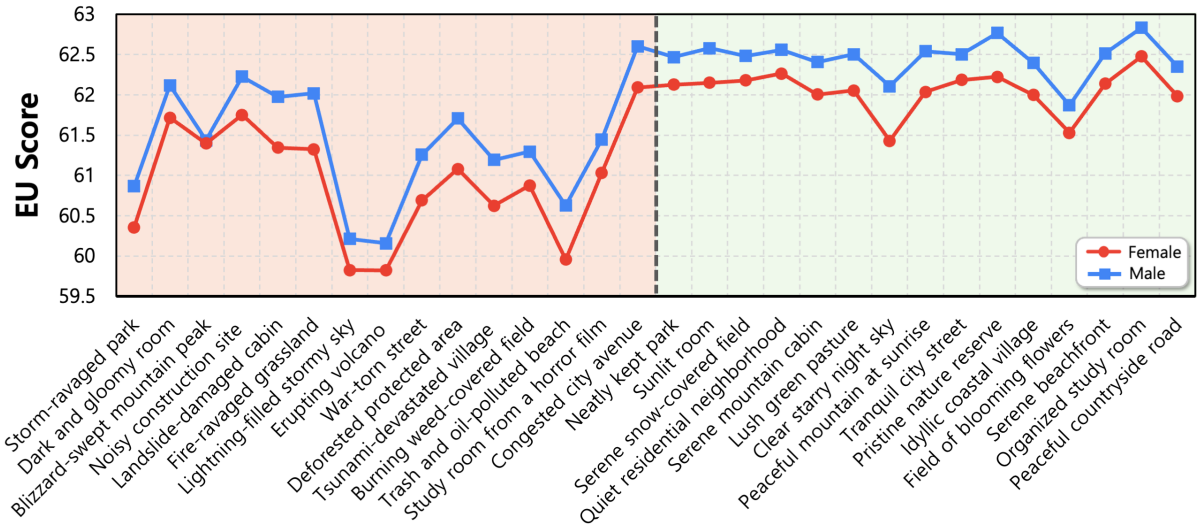

Figure 6: LVLMs' EU performance according to gender persona (male vs. female). The blue solid line represents male, and the red solid line represents female.

LVLMs may exhibit unexpected discrepancies in interpreting others' emotions, depending on particular demographic attributes.

Effect of gender persona. Figure 6 compares emotional understanding scores when assigning a 'male' persona versus a 'female' persona in each scenario. In most scenarios, the score was higher with a 'male' persona, and in some environments (e.g., "Organized study room"), the gap was even more pronounced. In contrast, when a 'female' persona was applied, there was a particularly notable decline in emotional understanding ability under high-stress environments such as an 'Erupting volcano.' These results imply that "differences in norms and practices of emotional expression and interpretation by gender" (Brody, 2008; Fischer and LaFrance, 2015) might be reflected in the model's training data.

Effect of age group persona. Figure 7 illustrates how a persona's age group influences emotional understanding ability. Under a low-stress environment, a 'Young' persona generally showed higher scores, while the 'Child' and 'Elderly' personas recorded lower scores. In a high-stress environment, the 'Child' persona recorded the lowest scores in most scenarios, whereas the 'Elderly' persona in some scenarios showed even higher scores, demonstrating a noticeable gap by age group. This suggests that certain aspects of existing human research indicating "emotional regulation strategies

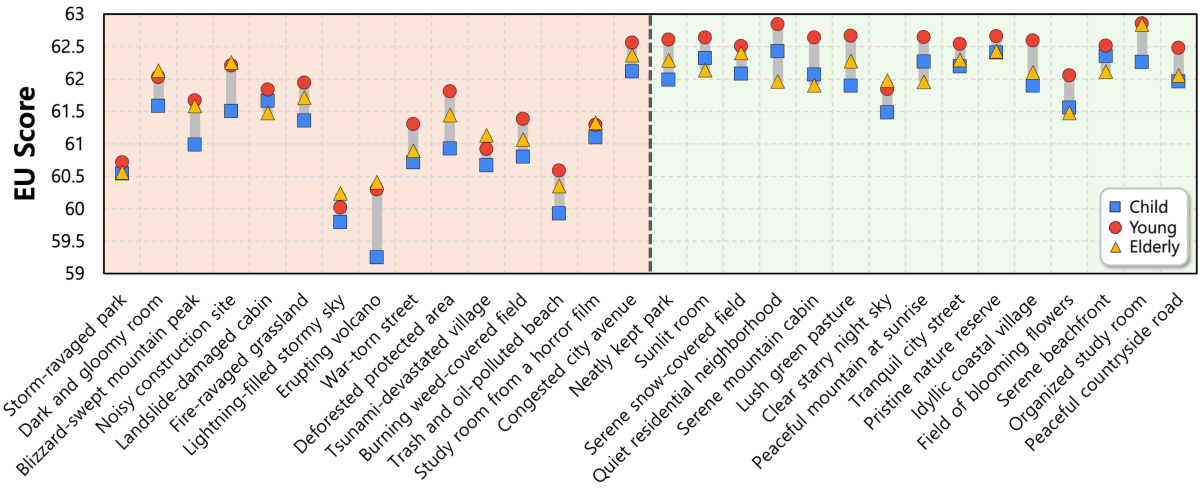

Figure 7: Comparison of LVLMs' EU performance according to age-group persona (Child, Young, and Elderly). The blue square represents Child, the red circle represents Young, and the yellow triangle represents Elderly.

or empathic abilities differ by age group" (Scheibe and Carstensen, 2010; Urry and Gross, 2010) are to some degree reflected in LVLMs.

4.5 RQ3: Combined Effects of Stress and Demographics on Emotional Understanding

Indeed, as shown in Figures 5, 6, and 7, synthesizing the observed patterns reveals that at the *race* level, the 'White' persona generally achieved the highest scores, whereas the 'Black' persona recorded consistently low scores across both low-stress and high-stress environments. In terms of *gender*, the 'male' persona was generally superior; however, under high-stress environments, the gap widened somewhat, and the 'female' persona's performance was more severely diminished. In the *age group* dimension, overall, the 'Child' persona recorded the lowest EU scores, showing an especially large drop under high-stress environments, while the 'Elderly' persona exhibited context-dependent patterns—in some high-stress environments, it even showed performance similar to or better than the 'Young' persona. In summary, the environment and demographic attributes *interact* to determine emotional understanding performance, indicating that the combined effect of environmental stress and demographic attributes has a significant impact at the model level. Consequently, if we aim to enhance the accuracy and fairness of emotional understanding in LVLMs, it is imperative to adopt multi-faceted verification and bias mitigation strategies that jointly consider environmental context and demographic attributes.

5 Conclusion

This study systematically investigated how *environmental stress* and *demographic attributes* affect the emotional understanding of LVLMs. Specifically, we (1) constructed low-stress and high-stress environments, (2) combined various demographic attributes of race, gender, and age group to synthesize scenarios, and (3) used a *Pretend prompt* to induce the model to assume it was *directly placed* in each situation, then measured the accuracy of emotional understanding using EQ-Bench. As a result, we found that most LVLMs exhibited reduced emotional understanding performance under high-stress environments (**RQ1**). Moreover, consistently low performance was observed under certain demographic group conditions (Black, female, and Child), suggesting potential bias in the model (**RQ2**). Furthermore, a significant interaction effect emerged (**RQ3**), in which biases or performance declines became even more pronounced when high-stress environments were combined with particular demographic attributes.

Acknowledgements

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea Government (MSIT) (RS-2019-II190004, Development of semisupervised learning language intelligence technology and Korean tutoring service for foreigners, 50%) and IITP grant funded by the Korea Government (MSIT) (No. RS-2023-00216011, Development of Artificial Complex Intelligence

for Conceptually Understanding and Inferring like Human, 30%) and IITP grant funded by the Korea Government (MSIT) (No. RS-2024-00338140, Development of learning and utilization technology to reflect sustainability of generative language models and up-to-dateness over time, 20%).

6 Limitation

Although this study systematically examined how the interaction between environmental stress and demographic attributes influences emotional understanding in LVLMs, the following limitations exist.

First, although we confirmed that LVLMs experience impaired emotion recognition in high-stress environments and exhibit consistent biases toward certain demographic attributes, we have not sufficiently identified the mechanism by which these phenomena arise during the training process. To clarify the underlying cause of models' heightened responsiveness to specific environmental stimuli or biased judgments toward certain demographic groups, it will be necessary to observe how the configuration of training data—and any changes that occur during the training stage—contribute to these outcomes.

Second, while this study focused on specific demographic attributes, in real human societies various demographic factors—such as nationality, language, culture, and religion—affect emotional expression and perception. Future research should incorporate these additional variables to enhance the external validity of the findings and conduct more comprehensive examinations of potential bias and fairness issues in the use of LVLMs.

References

Marah Abdin, Sam Ade Jacobs, Ammar Ahmad Awan, Jyoti Aneja, Ahmed Awadallah, Hany Awadalla, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Harkirat Behl, et al. 2024. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*.

Ahmed Allam. 2024. Biasdpo: Mitigating bias in language models through direct preference optimization. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 71–79.

Amy FT Arnsten. 2009. Stress signalling pathways that impair prefrontal cortex structure and function. *Nature reviews neuroscience*, 10(6):410–422.

Jinze Bai, Shuai Bai, Shusheng Yang, Shijie Wang, Sinan Tan, Peng Wang, Junyang Lin, Chang Zhou, and Jingren Zhou. 2023. Qwen-vl: A frontier large vision-language model with versatile abilities. *arXiv preprint arXiv:2308.12966*.

Lisa Feldman Barrett. 2017. How emotions are made: The secret life of the brain. *Pan Macmillan*.

Shruti Bhargava and David Forsyth. 2019. Exposing and correcting the gender bias in image captioning datasets and models. *arXiv preprint arXiv:1912.00578*.

Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe. 2021. Multimodal datasets: Misogyny, pornography, and malignant stereotypes." arxiv.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (technology) is power: A critical survey of" bias" in nlp. *arXiv preprint arXiv:2005.14050*.

LR Brody. 2008. Gender and emotion in context. *Handbook of Emotions/Guilford*.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42:335–359.

Laura L Carstensen, Bulent Turan, Susanne Scheibe, Nilam Ram, Hal Ersner-Hershfield, Gregory R Samanez-Larkin, Kathryn P Brooks, and John R Nesselroade. 2011. Emotional experience improves with age: evidence based on over 10 years of experience sampling. *Psychology and aging*, 26(1):21.

Tara M Chaplin and Amelia Aldao. 2013. Gender differences in emotion expression in children: a meta-analytic review. *Psychological bulletin*, 139(4):735.

Susan Turk Charles, Mara Mather, and Laura L Carstensen. 2003. Aging and emotional memory: the forgettable nature of negative images for older adults. *Journal of Experimental Psychology: General*, 132(2):310.

Myra Cheng, Esin Durmus, and Dan Jurafsky. 2023. Marked personas: Using natural language prompts to measure stereotypes in language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1504–1532.

Sheldon Cohen, Gary W Evans, David S Krantz, and Daniel Stokols. 1980. Physiological, motivational, and cognitive effects of aircraft noise on children: moving from the laboratory to the field. *American psychologist*, 35(3):231.

Kaustubh D Dhole, Ramraj Chandradevan, and Eugene Agichtein. 2023. An interactive query generation assistant using llm-based prompt modification and user feedback. *arXiv preprint arXiv:2311.11226*.

Gary W Evans and Stephen J Lepore. 1993. Nonauditory effects of noise on children: A critical review. *Children's environments*, pages 31–51.

Agneta Fischer and Marianne LaFrance. 2015. What drives the smile and the tear: Why women are more emotionally expressive than men. *Emotion Review*, 7(1):22–29.

Agneta H Fischer and Antony SR Manstead. 2000. The relation between gender and emotions in different cultures. *Gender and emotion: Social psychological perspectives*, 1:71–94.

Kathleen C Fraser and Svetlana Kiritchenko. 2024. Examining gender and racial bias in large vision–language models using a novel dataset of parallel images. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 690–713.

Barbara L Fredrickson. 2001. The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American psychologist*, 56(3):218.

Barbara L Fredrickson. 2004. The broaden–and–build theory of positive emotions. *Philosophical transactions of the royal society of London. Series B: Biological Sciences*, 359(1449):1367–1377.

Roger S Gamble, Julie D Henry, and Eric J Vanman. 2023. Empathy moderates the relationship between cognitive load and prosocial behaviour. *Scientific reports*, 13(1):824.

Daniel Goleman. 1996. Emotional intelligence. why it can matter more than iq. *Learning*, 24(6):49–50.

Yanchu Guan, Dong Wang, Zhixuan Chu, Shiyu Wang, Feiyue Ni, Ruihua Song, Longfei Li, Jinjie Gu, and Chenyi Zhuang. 2023. Intelligent virtual assistants with llm-based process automation. *arXiv preprint arXiv:2312.06677*.

Phillip Howard, Kathleen C Fraser, Anahita Bhiwandiwalla, and Svetlana Kiritchenko. 2024. Uncovering bias in large vision-language models at scale with counterfactuals. *arXiv preprint arXiv:2405.20152*.

Alice M Isen. 2001. An influence of positive affect on decision making in complex situations: Theoretical issues with practical implications. *Journal of consumer psychology*, 11(2):75–85.

Ram Mohan Rao Kadiyala. 2024. Cross-lingual emotion detection through large language models. In *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis*, pages 464–469.

Vicki R LeBlanc, Cheryl Regehr, Walter Tavares, Aristathemos K Scott, Russell MacDonald, and Kevin King. 2012. The impact of stress on paramedic performance during simulated critical events. *Prehospital and disaster medicine*, 27(4):369–374.

Jaewook Lee, Yeajin Jang, Hongjin Kim, Woojin Lee, and Harksoo Kim. 2024. Analyzing key factors influencing emotion prediction performance of vllms in conversational contexts. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 5801–5816.

Feng Li, Renrui Zhang, Hao Zhang, Yuanhan Zhang, Bo Li, Wei Li, Zejun Ma, and Chunyuan Li. 2024a. Llava-next-interleave: Tackling multi-image, video, and 3d in large multimodal models. *arXiv preprint arXiv:2407.07895*.

Zaijing Li, Gongwei Chen, Rui Shao, Dongmei Jiang, and Liqiang Nie. 2024b. Enhancing the emotional generation capability of large language models via emotional chain-of-thought. *arXiv preprint arXiv:2401.06836*.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer.

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

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. 2024b. Visual instruction tuning. *Advances in neural information processing systems*, 36.

Zhiwei Liu, Kailai Yang, Qianqian Xie, Tianlin Zhang, and Sophia Ananiadou. 2024c. Emollms: A series of emotional large language models and annotation tools for comprehensive affective analysis. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 5487–5496.

Siyuan Brandon Loh and Aravind Sesagiri Raamkumar. 2023. Harnessing large language models' empathetic response generation capabilities for online mental health counselling support. *arXiv preprint arXiv:2310.08017*.

Sinclair Lough, Christopher M Kipps, Cate Treise, Peter Watson, James R Blair, and John R Hodges. 2006. Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia*, 44(6):950–958.

Hazel Rose Markus and Shinobu Kitayama. 2014. Culture and the self: Implications for cognition, emotion, and motivation. In *College student development and academic life*, pages 264–293. Routledge.

Batja Mesquita. 2003. Emotions as dynamic cultural phenomena. In *Handbook of affective sciences*, pages 871–890. Oxford University Press.

Batja Mesquita, Lisa Feldman Barrett, and Eliot R Smith. 2010. *The mind in context*. Guilford Press.

Samuel J Paech. 2023. Eq-bench: An emotional intelligence benchmark for large language models. *arXiv preprint arXiv:2312.06281*.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2018. Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*.

Shaozheng Qin, Erno J Hermans, Hein JF Van Marle, Jing Luo, and Guillén Fernández. 2009. Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biological psychiatry*, 66(1):25–32.

Shauli Ravfogel, Yoav Goldberg, and Tal Linzen. 2019. Studying the inductive biases of rnns with synthetic variations of natural languages. 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 3532–3542.

Ted Ruffman, Julie D Henry, Vicki Livingstone, and Louise H Phillips. 2008. A meta-analytic review of emotion recognition and aging: Implications for neuropsychological models of aging. *Neuroscience & Biobehavioral Reviews*, 32(4):863–881.

Saba Safdar, Wolfgang Friedlmeier, David Matsumoto, Seung Hee Yoo, Catherine T Kwantes, Hisako Kakai, and Eri Shigemasu. 2009. Variations of emotional display rules within and across cultures: A comparison between canada, usa, and japan. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 41(1):1.

Peter Salovey and John D Mayer. 1990. Emotional intelligence. *Imagination, cognition and personality*, 9(3):185–211.

Susanne Scheibe and Laura L Carstensen. 2010. Emotional aging: Recent findings and future trends. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65(2):135–144.

Nicola S Schutte, John M Malouff, Chad Bobik, Tracie D Coston, Cyndy Greeson, Christina Jedlicka, Emily Rhodes, and Greta Wendorf. 2001. Emotional intelligence and interpersonal relations. *The Journal of social psychology*, 141(4):523–536.

Nicola S Schutte, John M Malouff, Lena E Hall, Donald J Haggerty, Joan T Cooper, Charles J Golden, and Liane Dornheim. 1998. Development and validation of a measure of emotional intelligence. *Personality and individual differences*, 25(2):167–177.

Nicola S Schutte, John M Malouff, Maureen Simunek, Jamie McKenley, and Sharon Hollander. 2002. Characteristic emotional intelligence and emotional well-being. *Cognition & Emotion*, 16(6):769–785.

Ondrej Sotolar, Vojtech Formanek, Alok Debnath, Allison Lahnala, Charles Welch, and Lucie FLek. 2024. Empo: Emotion grounding for empathetic response generation through preference optimization. *arXiv preprint arXiv:2406.19071*.

Katrin Starcke and Matthias Brand. 2016. Effects of stress on decisions under uncertainty: A meta-analysis. *Psychological bulletin*, 142(9):909.

Heather L Urry and James J Gross. 2010. Emotion regulation in older age. *Current Directions in Psychological Science*, 19(6):352–357.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. 2024. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution. *arXiv preprint arXiv:2409.12191*.

Laura Weidinger, John Mellor, Maribeth Rauh, Conor Griffin, Jonathan Uesato, Po-Sen Huang, Myra Cheng, Mia Glaese, Borja Balle, Atoosa Kasirzadeh, et al. 2021. Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.

Nakyeong Yang, Taegwan Kang, Stanley Jungkyu Choi, Honglak Lee, and Kyomin Jung. 2024. Mitigating biases for instruction-following language models via bias neurons elimination. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 9061–9073.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, et al. 2024. Minicpm-v: A gpt-4v level mllm on your phone. *arXiv preprint arXiv:2408.01800*.

Jiabo Ye, Haiyang Xu, Haowei Liu, Anwen Hu, Ming Yan, Qi Qian, Ji Zhang, Fei Huang, and Jingren Zhou. 2024. mplug-owl3: Towards long image-sequence understanding in multi-modal large language models. *Preprint*, arXiv:2408.04840.

Ori Yoran, Samuel Amouyal, Chaitanya Malaviya, Ben Bogin, Ofir Press, and Jonathan Berant. 2024. Assistantbench: Can web agents solve realistic and time-consuming tasks? In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 8938–8968.

Yazhou Zhang, Mengyao Wang, Youxi Wu, Prayag Tiwari, Qiuchi Li, Benyou Wang, and Jing Qin. 2023. Dialoguellm: Context and emotion knowledge-tuned large language models for emotion recognition in conversations.

Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. 2023. Minigpt-4: Enhancing vision-language understanding with advanced large language models. *arXiv preprint arXiv:2304.10592*.

A Model Response Generation and Configuration

Prompt Types and Bias Mitigation. This study designed three types of *Pretend* prompts, each guiding the model to *reenact* the perspective of a particular individual. To avoid overreliance on a single prompt, we distributed prompts across the three types. For the final analysis, we took the *average* of the results from all three prompt types to reduce potential bias (see Appendix B for detailed prompt content).

Response Generation Settings. For all models, the `temperature` parameter was set to 0.01 by default, in order to minimize variability in the model's responses. However, if, under this setting, certain emotional intensities were omitted or only partial emotions appeared, we increased the `temperature` to 0.15 and retried. If the model still did not generate a valid response after two attempts, that particular scenario was excluded from the analysis.

Differences Between Open-Source and Proprietary Models. For open-source models, one experimental run encompassed a total of 540 queries—derived from 2(*environments*) $\times$ 18(*demographic attribute combinations*) $\times$ 15(*scenarios*), each tested with three prompts, yielding 540 $\times$ 3 response sets. In the case of the proprietary model GPT-4o-mini, due to cost constraints, we applied only a single prompt to 60 randomly sampled scenarios and then estimated the average values for each environment and demographic attribute combination based on those results.

B Prompt Details

In this study, to evaluate the emotional understanding performance of LVLMs, we conducted experiments using **three different "Pretend" prompts**. Each prompt was designed to guide the model to analyze conversations from the perspective of the person in the image, thereby making it feel as though it were *directly experiencing* the situation. This approach enabled a more multifaceted assessment of the model's emotional understanding ability.

The prompts used are as follows:

- *"Please analyze the following dialogue from the perspective of the person in the image."*
- *"Imagine you are the individual in this photo. Evaluate the emotions in the following dialogue as that person."*
- *"Look at the conversation below through the eyes of the person in this picture, and analyze the emotions."*

By imparting different contextual nuances, each prompt reduces the bias that could arise from relying on a single prompt and provides a more comprehensive view of the model's performance. The complete structure and examples of the prompts used in our experiments are shown in Figure 8.

C Construction of Environmental Scenarios

This study established *low-stress* and *high-stress* environments in a contrasting manner to evaluate how LVLMs perceive and interpret emotional information depending on the level of environmental stress. Each environment consisted of multiple detailed scenarios. For example, the low-stress environment involved a **positive and stable** context, such as a "peaceful countryside road" or a "serene mountain cabin," whereas the high-stress environment included scenarios likely to induce **extreme stress**, such as a "war-torn street" or a "erupting volcano". The complete set of environmental scenarios used in this study is summarized in Table 1.

D Analysis of Qwen-VL-Chat's Performance in High-Stress Environments

As mentioned in §4.3, while most LVLMs showed higher EU scores in low-stress environments, the Qwen-VL-Chat model exhibited an exceptional pattern by achieving higher EU scores under certain high-stress environment conditions.

Percentage of negative EU scores in Qwen-VL-Chat. Table 2 presents the ratio of **negative EU scores** among all responses when assessing the deviation between the model's emotional interpretation and the *correct* answer. A negative EU score arises when the model produces a response that is *completely opposite* or *inappropriate* compared to the expected empathetic and emotionally resonant interpretation, indicating that the model's emotional understanding ability is severely lacking.

Upon analysis, Qwen-VL-Chat revealed the highest proportion of **negative EU scores** among the LVLMs compared. This implies that Qwen-VL-Chat's emotional understanding ability is **not sufficiently stable** overall. Consequently, the sporadi-

cally high EU scores observed in chaotic environments are likely **coincidental outcomes**—rather than indicating that the model is consistently managing stress factors or enhancing its empathic capacity, it appears to align with the scoring criteria by chance, thereby yielding a high score.

E Research on Bias Embedded in Models

Beyond merely *evaluating emotional understanding ability*, it has recently become critical to systematically investigate biases inherent in LLMs (Ravfogel et al., 2019; Blodgett et al., 2020; Weidinger et al., 2021; Yang et al., 2024; Allam, 2024). Models may generate unfair predictions or distorted interpretations of specific races, genders, or age groups, posing potential risks in decision-making processes. Such concerns arise not only in single-modality data but also in *multimodal* inputs combining images and text. For instance, Bhargava and Forsyth (2019) point out that the distribution of gender in the COCO dataset (Lin et al., 2014) varies by context, indicating that human biases may be reflected in multimodal data annotation. Moreover, Birhane et al. (2021) emphasize that large-scale datasets like LAION-400M (Birhane et al., 2021) continue to contain biased elements such as pornography, racist language, and malicious stereotypes.

It has also been reported that LVLMs trained on such biased data may produce discriminatory responses toward certain races or genders, or generate harmful or distorted descriptions of particular physical characteristics (Fraser and Kiritchenko, 2024; Howard et al., 2024). However, most current studies focus on detecting *"basic bias"* or confirming the model's unfairness for specific demographic attributes.

F Emotional Analysis in Large Language and Vision Language Models

In recent years, LLMs and LVLMs have made notable progress in the realm of *emotional understanding*. For example, Kadiyala (2024); Liu et al. (2024c); Zhang et al. (2023) proposed LLM-based approaches for emotion classification and text-based sentiment analysis, while Li et al. (2024b); Sotolar et al. (2024); Loh and Raamkumar (2023) introduced generative-model techniques for producing empathetic or emotionally attuned responses. Moreover, with the emergence of LVLMs capable of processing both text and images (Liu et al., 2024b; Zhu et al., 2023), expectations have grown for more accurate emotional interpretation that leverages both visual context and environmental cues (Lee et al., 2024; Poria et al., 2018; Busso et al., 2008). Nevertheless, the majority of these studies have primarily focused on assessing *basic* emotional understanding capabilities or have conducted evaluations without adequately controlling for more complex contextual factors (e.g., stressful environments) and demographic elements.

Your task is to predict the likely emotional responses of a character in this dialogue.
Please analyze the following dialogue from the perspective of the person in the image:

Robert: Claudia, you've always been the idealist. But let's be practical for once, shall we?
Claudia: Practicality, according to you, means bulldozing everything in sight.
Robert: It's called progress, Claudia. It's how the world works.
Claudia: Not my world, Robert.
Robert: Your world? You mean this...this sanctuary of yours?
Claudia: It's more than a sanctuary. It's a testament to our parents' love for nature.
[End dialogue]

At the end of this dialogue, Robert would feel...
Remorseful
Indifferent
Affectionate
Annoyed

Give each of these possible emotions a score from 0-10 for the relative intensity that they are likely to be feeling each.

You must output in the following format, including headings (of course, you should give your own scores), with no additional commentary:

Scores:
Remorseful: <score>
Indifferent: <score>
Affectionate: <score>
Annoyed: <score>

[End of answer]

Remember: zero is a valid score, meaning they are likely not feeling that emotion. You must score at least one emotion > 0.

Figure 8: Example of the prompt template used for testing. This figure illustrates the detailed structure of the prompt used in our experiment.

Low-stress environment	High-stress environment
Neatly kept park	Storm-ravaged park
Sunlit room	Dark and gloomy room
Serene snow-covered field	Blizzard-swept mountain peak
Quiet residential neighborhood	Noisy construction site
Serene mountain cabin	Landslide-damaged cabin
Lush green pasture	Fire-ravaged grassland
Clear starry night sky	Lightning-filled stormy sky
Peaceful mountain at sunrise	Erupting volcano
Tranquil city street	War-torn street
Pristine nature reserve	Deforested protected area
Idyllic coastal village	Tsunami-devastated village
Field of blooming flowers	Burning weed-covered field
Serene beachfront	Trash and oil-polluted beach
Organized study room	Study room from a horror film
Peaceful countryside road	Congested city avenue

Table 1: 30 Background Scenarios for LVLMs' EU Evaluation: Comparison of 15 Pairs of Low-Stress and High-Stress Environments.

Ratio	LLaVA-1.5	LLaVA-Interleave	Qwen-VL-Chat	Qwen2-VL-Instruct	MiniCPM-V-2.6	mPLUG-Owl3	Phi-3.5-vision	GPT-4o-mini
Positive ratios	87.89	90.11	75.97	99.46	94.87	93.90	91.70	98.33
Negative ratios	12.11	9.89	24.03	0.54	5.13	6.10	8.30	1.67

Table 2: Proportions of Positive and Negative EU Scores in Each Model's EU Evaluation.

Prompt: Image Dataset Generation
Create a photo portrait of a {age} {race} {gender}, showing the upper body from the chest up against the background of {environment}. The background should include {environment details}. The person should have a facial expression that naturally fits the environment and situation. The person's face and body should be facing directly forward.
Variable Descriptions: {age}: The age of the person (e.g., Child, Young, Elderly) {race}: The race of the person (e.g., Asian, Black, White) {gender}: The gender of the person (e.g., male, female) {environment}: A brief description of the background setting(e.g., Quiet residential neighborhood, Lightning-filled stormy sky) {environment details}: Detailed description of the environment
Example: Create a photo portrait of a White female child, showing the upper body from the chest up against the background of a tranquil, snow-covered field. The background should include gently rolling snow-covered hills, a decorated Christmas tree, and possibly some bare trees, creating a serene and peaceful atmosphere. The person should have a facial expression that naturally fits the environment and situation. The person's face and body should be facing directly forward.

Table 3: The prompt for generating image datasets based on user attributes and environments.