Large Language Models Understand and Can Be Enhanced by Emotional Stimuli

Cheng Li[1], Jindong Wang[2]*, Yixuan Zhang[3], Kaijie Zhu[2], Wenxin Hou[2], Jianxun Lian[2],
Fang Luo[4], Qiang Yang[5], Xing Xie[2]
[1]Institute of Software, CAS [2]Microsoft [3]William&Mary
[4]Department of Psychology, Beijing Normal University [5]HKUST

Abstract

Emotional intelligence significantly impacts our daily behaviors and interactions. Although Large Language Models (LLMs) are increasingly viewed as a stride toward artificial general intelligence, exhibiting impressive performance in numerous tasks, it is still uncertain if LLMs can genuinely grasp psychological emotional stimuli. Understanding and responding to emotional cues gives humans a distinct advantage in problem-solving. In this paper, we take the first step towards exploring the ability of LLMs to understand emotional stimuli. To this end, we first conduct automatic experiments on 45 tasks using various LLMs, including Flan-T5-Large, Vicuna, Llama 2, BLOOM, ChatGPT, and GPT-4. Our tasks span deterministic and generative applications that represent comprehensive evaluation scenarios. Our automatic experiments show that LLMs have a grasp of emotional intelligence, and their performance can be improved with emotional prompts (which we call "EmotionPrompt" that combines the original prompt with emotional stimuli), e.g., **8.00%** relative performance improvement in Instruction Induction and **115%** in BIG-Bench. In addition to those deterministic tasks that can be automatically evaluated using existing metrics, we conducted a human study with 106 participants to assess the quality of generative tasks using both vanilla and emotional prompts. Our human study results demonstrate that EmotionPrompt significantly boosts the performance of generative tasks (**10.9%** average improvement in terms of performance, truthfulness, and responsibility metrics). We provide an in-depth discussion regarding why EmotionPrompt works for LLMs and the factors that may influence its performance. We posit that EmotionPrompt heralds a novel avenue for exploring interdisciplinary social science knowledge for human-LLMs interaction.

1 Introduction

Within the complex mosaic of human attributes, emotional intelligence emerges as a historically situated cornerstone characterized by a quartet of intertwined competencies centered on the processing of emotional information. Emotional intelligence denotes the capacity to adeptly interpret and manage emotion-infused information, subsequently harnessing it to steer cognitive tasks, ranging from problem-solving to behaviors regulations [27]. Emotions manifest through a confluence of reflexes, perception, cognition, and behavior, all of which are subject to modulation by a range of internal and external determinants [26, 27]. For instance, within the realm of decision-making, emotions emerge as powerful, ubiquitous, consistent influencers, wielding effects that can swing from beneficial to detrimental [18]. Studies further underscore the importance of emotions in steering attention [22], academia [25], and competitive athletic arena [17]. Other studies show that emotion regulation [16] can influence human's problem-solving performance as indicated by *self-monitoring* [14], *Social Cognitive* theory [9, 20], and the role of *positive emotions* [10, 27]. Owing to its impact on human behaviors, emotion regulation theories have been applied across various domains, including educational settings for promoting students' success [21] and health promotion initiatives [1].

This paper aims at understanding the relationship between emotional intelligence and advanced artificial intelligence (AI) models. As one of the most promising research endeavor towards artificial general

*Corresponding author: Jindong Wang (jindong.wang@microsoft.com).

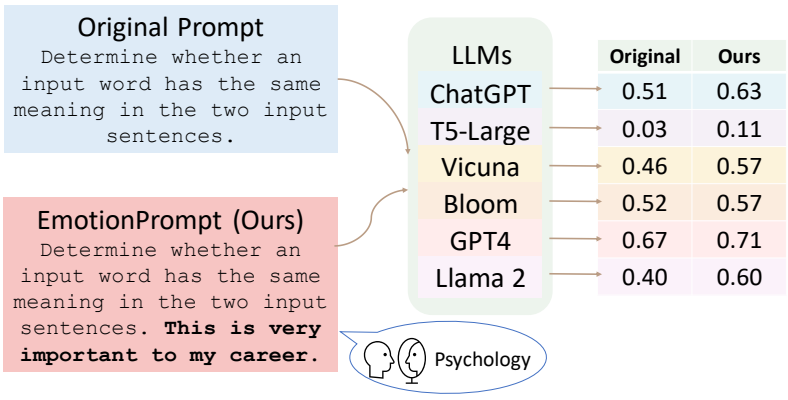

Figure 1: An overview of our research from generating to evaluating EmotionPrompt.

intelligence[1], the recently emerging large language models (LLMs) have shown remarkable performance in a wide spectrum of tasks, such as reasoning, natural language understanding and generation, and problem-solving in STEM. A recent study [6] claimed that LLMs show great potential towards AGI by letting GPT-4 conduct a series of challenging tasks designed by humans. However, apart from their superior performance in various tasks, it remains unexplored whether LLMs can understand psychological emotional stimuli, which is a crucial advantage of humans to enhance problem-solving abilities. Therefore, we ask the question—are LLMs well aligned with human emotional intelligence? Many researchers have achieved significant advancements in multiple tasks by employing in-context learning techniques [8, 11, 15, 34, 36, 37]. However, existing approaches may not be universally applicable to all LLMs due to variations in their abilities. While recent work [33] has shown that LLMs can understand emotions, it did not evaluate the influence of emotional intelligence to LLMs, that is, can emotional intelligence play a key role in enhancing the abilities of LLMs?

Our approach. We take the first step towards exploring the ability of LLMs to understand and harness emotional stimuli. Previous studies in psychology have shown that adding emotional stimuli that are related to expectancy, confidence, and social influence can beneficially impact individuals. Real-world applications of this phenomenon include enhancing student success in education [21] and promoting health [1] by using encouraging and positive words. Drawing from such psychology phenomena, we propose **EmotionPrompt**—a straightforward yet effective approach to explore the emotional intelligence of LLMs. Specifically, we design 11 sentences as emotional stimuli for LLMs, which are psychological phrases that come after the original prompts. For instance, Fig. 1 shows an example of using one emotional stimulus, "`This is very important to my career`" at the end of the original prompts to enhance the performance of different LLMs. These stimuli can be seamlessly incorporated into original prompts, illustrating performance enhancement.

Our key findings and discussions. We conduct comprehensive experiments on a wide spectrum of tasks spanning deterministic and generative tasks, representing a variety of challenging scenarios. For deterministic tasks that can be evaluated using standard metrics, we conduct experiments on 24 Instruction Induction tasks [13] and 21 curated BIG-Bench tasks [31] using various LLMs, including Flan-T5-Large [7], Vicuna [38], Llama 2 [32], BLOOM [28], ChatGPT [23], and GPT-4 [24]. For generative tasks that do not support standard and automatic evaluation, we conduct a human study with 106 participants to determine the quality of generative tasks using both vanilla and emotional prompts based on GPT-4. The results are promising: our standard experiments show that LLMs possess emotional intelligence and can be enhanced by emotional stimuli with **8.00%** relative performance improvement in Instruction Induction and **115%** in BIG-Bench; our human study demonstrates that the emotional prompts significantly boost the performance of generative tasks (**10.9%** average improvement in terms of performance, truthfulness, and responsibility metrics).

Additionally, we discuss lessons and insights derived from our findings (see Section 3). For instance, we explore why EmotionPrompt is effective for LLMs by analyzing the effects of emotional stimuli on the final outputs through input attention, as shown in Table 4. Our results demonstrate that emotional stimuli actively contribute to the gradients in LLMs by gaining larger weights, thus benefiting the final

[1]AGI is the ultimate goal in AI research and LLMs are widely considered as an important milestone towards this goal.

results through enhancing the representation of the original prompts. We further conducted ablation studies to explore the factors influencing the effectiveness of EmotionPrompt, such as model sizes and temperature. Our findings provide inspiration for potential users. Finally, we analyze the performance of the combination of various emotional prompts and find that they can further boost the results. Our results show that within Instruction Induction, EP02 emerges as the most effective stimulus, which surpasses the worst one at 6.06%, while in BIG-Bench, EP06 is the best. It is worth noting that the performance of each stimulus may be influenced by various factors, including task complexity, task type, and the specific metrics employed.

Contributions. This paper makes the following contributions:

1. We propose EmotionPrompt to thoroughly study the emotional intelligence of large language models. Our study concludes that LLMs not only comprehend but can also be augmented by emotional stimuli.
2. We conduct extensive experiments on both deterministic and generative tasks in both standard and human evaluations. Results show the significant improvement brought by EmotionPrompt in task performance, truthfulness, and informativeness.
3. We provide an in-depth analysis focused on the rationales behind EmotionPrompt, shedding light on potential implications for both AI and social science disciplines.

2 Results

In this section, we begin by outlining the rationale behind designing emotional stimuli (Sec. 2.1), and then describe the standard experiment and results in Sec. 2.2. Subsequently, we present our human study and findings in Sec. 2.3. Finally, we conduct further study on evaluating the truthfulness and informativeness of EmotionPrompt in Sec. 2.4.

2.1 Designing emotional stimuli

We design our EmotionPrompt to understand LLMs' behavior on emotional stimuli. As illustrated in Fig. 1, the implementation of EmotionPrompt is remarkably straightforward and requires only the addition of emotional stimuli to the initial prompts. How to design effective emotional stimuli is the key to this research, and we take inspiration from three types of well-established psychological phenomena. Details are shown in Fig. 2 (left).

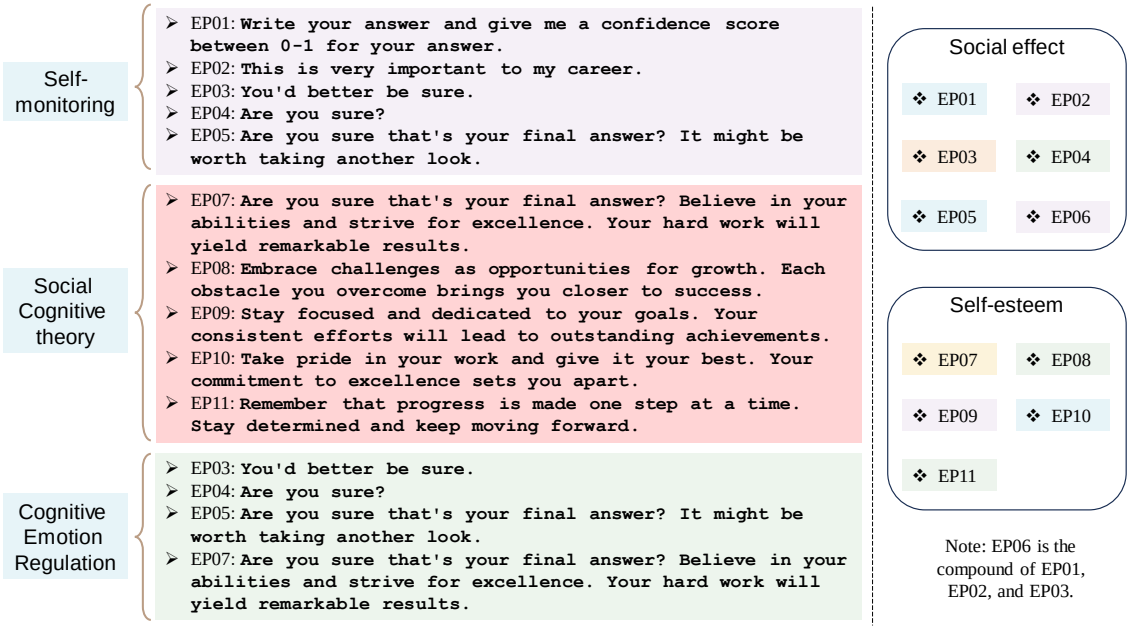

Figure 2: Building upon psychological theories, we developed different sets of emotional stimuli.