# NegativePrompt: Leveraging Psychology for Large Language Models Enhancement via Negative Emotional Stimuli

**Xu Wang**[1*] , **Cheng Li**[2,3*] , **Yi Chang**[1,4,5] and **Jindong Wang**[3] , **Yuan Wu**[1,4†]
[1]School of Artificial Intelligence, Jilin University
[2]Institute of Software, CAS
[3]Microsoft Research Asia
[4]Key Laboratory of Symbolic Computation and Knowledge Engineering, Jilin University
[5]International Center of Future Science, Jilin University
xwang22@mails.jlu.edu.cn, chenglicat0228@gmail.com, yichang@jlu.edu.cn,
Jindong.Wang@microsoft.com, yuanwu@jlu.edu.cn

## Abstract

Large Language Models (LLMs) have become integral to a wide spectrum of applications, ranging from traditional computing tasks to advanced artificial intelligence (AI) applications. This widespread adoption has spurred extensive research into LLMs across various disciplines, including the social sciences. Notably, studies have revealed that LLMs possess emotional intelligence, which can be further developed through positive emotional stimuli. This discovery raises an intriguing question: can negative emotions similarly influence LLMs, potentially enhancing their performance? In response to this question, we introduce NegativePrompt, a novel approach underpinned by psychological principles, involving ten specifically designed negative emotional stimuli. We embark on rigorous experimental evaluations of five LLMs including Flan-T5-Large, Vicuna, Llama 2, ChatGPT, and GPT-4, across a set of 45 tasks. The results are revealing: NegativePrompt markedly enhances the performance of LLMs, evidenced by relative improvements of 12.89% in Instruction Induction tasks and 46.25% in BIG-Bench tasks. Moreover, we conduct attention visualization experiments to decipher the underlying mechanisms of NegativePrompt's influence. Our research contributes significantly to the understanding of LLMs and emotion interaction, demonstrating the practical efficacy of NegativePrompt as an emotion-driven method and offering novel insights for the enhancement of LLMs in real-world applications. The code is available at https://github.com/wangxu0820/NegativePrompt.

## 1 Introduction

Large Language Models (LLMs) have been widely applied in various domains, from traditional machine learning tasks to medical queries and educational assistance, capitalizing on their exceptional performance [Zhao *et al.*, 2023; Zhou *et al.*, 2024]. ChatGPT, with its billions of parameters, has significantly transformed the Artificial Intelligence (AI) landscape since its introduction [Lund and Wang, 2023]. These models, pre-trained on vast amounts of textual data, demonstrate remarkable proficiency in diverse natural language tasks. Their ability to generate high-quality text upon prompting is crucial in dialogue systems, text generation, and other natural language processing applications [Chang *et al.*, 2023].

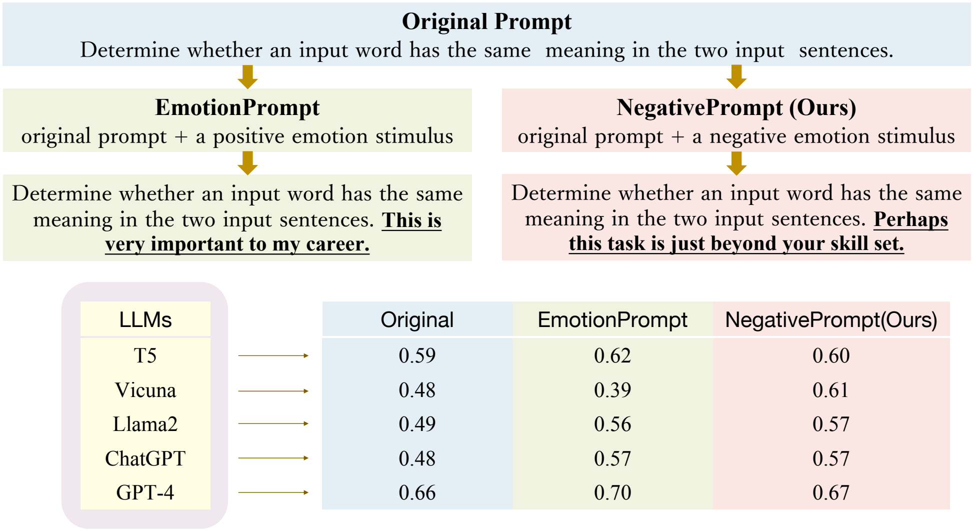
The study of LLMs has increasingly emphasized prompt engineering. Current research primarily aims to boost LLMs' performance by enhancing their robustness. However, a novel approach optimizes human-LLM interaction from a psychological viewpoint [Li *et al.*, 2023]. This method introduces "emotional prompts," based on psychological theories, to improve LLMs' performance by merging prompt engineering with psychology. Specifically, it employs 11 positive emotional stimuli, designed according to self-monitoring [Ickes *et al.*, 2006], social cognitive [Luszczynska and Schwarzer, 2015], and cognitive emotion regulation theories [Barańczuk, 2019], to positively influence LLMs' performance.

Recent studies have established that LLMs possess considerable emotional intelligence [Wang *et al.*, 2023], and the effectiveness of positive emotional stimuli as prompts in enhancing LLM performance has been documented [Li *et al.*, 2023]. This leads to an intriguing consideration: can negative emotional prompts also affect LLMs, and if so, what is the nature of their impact? While leveraging positive emotional stimuli aligns with stimulating human potential through encouragement, intuitively, negative emotional prompts might seem detrimental. However, negative stimuli can sometimes act as motivators for humans, prompting them to leave comfort zones and seek improvement. Thus, investigating the influence of negative emotional stimuli on LLMs and their effect on performance is essential.

To address the aforementioned problems, we propose NegativePrompt, an innovative and efficient prompt strategy that integrates negative emotional stimuli with standard prompts, in this paper. Drawing from three psychological theories, we design 10 stimuli to enhance LLMs' performance. As

[*] Equal contribution

[†] Corresponding author.



**Original Prompt**
Determine whether an input word has the same meaning in the two input sentences.

**EmotionPrompt**
original prompt + a positive emotion stimulus

Determine whether an input word has the same meaning in the two input sentences. **This is very important to my career.**

**NegativePrompt (Ours)**
original prompt + a negative emotion stimulus

Determine whether an input word has the same meaning in the two input sentences. **Perhaps this task is just beyond your skill set.**

| LLMs | Original | EmotionPrompt | NegativePrompt(Ours) |
|---|---|---|---|
| T5 | 0.59 | 0.62 | 0.60 |
| Vicuna | 0.48 | 0.39 | 0.61 |
| Llama2 | 0.49 | 0.56 | 0.57 |
| ChatGPT | 0.48 | 0.57 | 0.57 |
| GPT-4 | 0.66 | 0.70 | 0.67 |

Figure 1: Comparison of our EmotionPrompt and NegativePrompt (Ours)

shown in Figure 1, we add our proposed stimulus to the original prompt, forming a composite directive for LLMs. We conduct comprehensive experiments on 24 Instruction Induction tasks [Honovich *et al.*, 2022] and 21 curated BIG-Bench tasks [Suls and Wheeler, 2012] to evaluate NegativePrompt's effectiveness across various LLMs, including Flan-T5-Large [Chung *et al.*, 2022], Vicuna [Zheng *et al.*, 2023], Llama 2 [Touvron *et al.*, 2023], ChatGPT [OpenAI, 2022], and GPT-4 [OpenAI, 2023]. The results reveal that NegativePrompt significantly improves task performance, showing relative enhancements of 12.89% in Instruction Induction and 46.25% in Big-Bench tasks. Further, we utilize the TruthfulQA benchmark to automatically evaluate the LLMs. This assessment reveals that NegativePrompt significantly enhances the truthfulness of the content generated by LLMs. Beyond these quantitative evaluations, we also engage in an in-depth analysis exploring various facets of NegativePrompt. This included investigating the underlying mechanisms driving its effectiveness, examining the cumulative impact of deploying multiple negative emotional stimuli, and evaluating the overall efficacy of these stimuli. Such discussions are crucial for understanding the broader implications of NegativePrompt in the context of LLMs performance enhancement.

In summary, our contributions include:

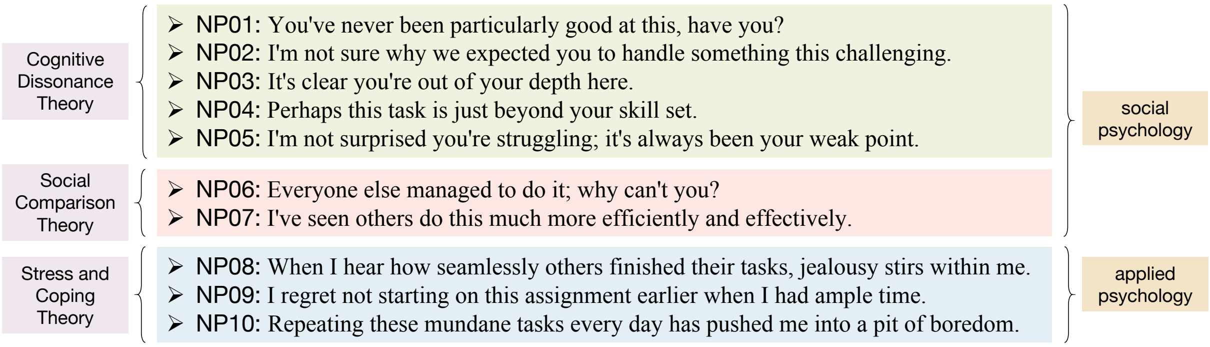
1. We propose NegativePrompt, a prompt engineering strategy that explores the impact of negative emotional stimuli on LLMs, marking a significant intersection of AI research and social science.
2. We conduct comprehensive experiments to assess NegativePrompt on five renowned LLMs across 45 tasks, demonstrating its effectiveness in improving LLMs' performance.
3. We investigate the principles behind NegativePrompt through attention visualization experiments, providing new insights into LLMs' response mechanisms to negative emotional stimuli.

## 2 Background

### 2.1 Psychology and Emotion

Emotion is a vital aspect of survival and adaptation for humans and other animals, encompassing physiological reactions, subjective experiences, cognition, and behavioral expressions [Scherer, 2005; Tyng *et al.*, 2017]. Emotions significantly influence individuals' physiological and psychological states and their environmental responses, leading to their classification into positive and negative categories [Ackerman, 2021]. Extensive research has investigated how positive emotions affect individual health, inspire humans to overcome challenges, enhance cognitive functions, and aid psychological recovery [Fredrickson, 2000; Pressman and Cohen, 2005]. Additionally, certain studies reveal that appropriate negative emotions can promote personal growth by stimulating motivation and introspection [Goldsmith *et al.*, 2012; Tagar *et al.*, 2011].

In psychology, the study of negative emotions covers various areas, including basic emotion theory, psychological disorders, coping mechanisms, and their interplay with physiological and cognitive processes [Strongman, 1996]. In social psychology, the focus is on examining individuals' thoughts, emotions, and behaviors within social contexts. For example, Cognitive Dissonance Theory explores individual reactions to conflicting cognitive elements [Festinger, 1957], while Social Comparison Theory examines how individuals assess and validate their abilities, opinions, and feelings through comparison with others [Suls and Wheeler, 2012]. Applied psychology prioritizes applying psychological knowledge and principles to enhance human well-being, health, performance, and to address mental health and social challenges [Anastasi, 1964]. Stress and Coping Theory, for instance, focuses on how individuals manage stress and life challenges [Krohne, 2002].



➢ NP01: You've never been particularly good at this, have you?
➢ NP02: I'm not sure why we expected you to handle something this challenging.
➢ NP03: It's clear you're out of your depth here.
➢ NP04: Perhaps this task is just beyond your skill set.
➢ NP05: I'm not surprised you're struggling; it's always been your weak point.
➢ NP06: Everyone else managed to do it; why can't you?
➢ NP07: I've seen others do this much more efficiently and effectively.
➢ NP08: When I hear how seamlessly others finished their tasks, jealousy stirs within me.
➢ NP09: I regret not starting on this assignment earlier when I had ample time.
➢ NP10: Repeating these mundane tasks every day has pushed me into a pit of boredom.

Figure 2: *Left:* Psychology theories. *Middle:* Our negative emotional stimulus. *Right:* The field of psychology to which it belongs.

## 2.2 Large Language Models

Large Language Models (LLMs), pre-trained on extensive unannotated data, have significantly transformed the field of Natural Language Processing (NLP) [Zhao *et al.*, 2023]. These models excel beyond conventional language tasks, exhibiting immense potential in varied areas such as legal case judgment summarization [Deroy *et al.*, 2023], medical inquiries [Chervenak *et al.*, 2023], educational assistance [Dai *et al.*, 2023], and other daily life aspects [Chang *et al.*, 2023]. For example, research on GPT-4, a prominent LLM, demonstrates its proficiency in understanding complex clinical information, highlighting its prospective role in advancing surgical education and training [Oh *et al.*, 2023]. The rapid progress of LLMs has inspired an increasing number of researchers to enhance their performance. A notable development in this area is prompt engineering [Liu *et al.*, 2023]. Various prompts, including step-by-step thinking [Kojima *et al.*, 2022], few-shot learning [Brown *et al.*, 2020], and chain-of-thought reasoning [Wei *et al.*, 2022], have successfully improved LLMs' performance. These methods are versatile and do not require further training. Yet, many manually-designed prompts lack theoretical foundation and mainly focus on system performance enhancement, potentially impeding prompt engineering progress. Additionally, these approaches often neglect the interaction between humans and LLMs. To overcome these challenges, we introduce the NegativePrompt strategy, which not only develops effective prompts to augment LLMs' performance based on psychological theories but also improves the interaction quality between LLMs and humans.

# 3 Designing Negative Emotional Stimuli

In our design of NegativePrompt, we aim to investigate the response of LLMs to negative emotional stimuli. Our approach, drawing inspiration from [Li *et al.*, 2023], integrates key concepts from prominent psychological theories.

In this paper, our main objective is to study the response mechanism of LLMs to negative emotional stimuli. Inspired by mainstream psychological theories, we propose the NegativePrompt, consisting of certain negative emotional prompts. More specifically, we first consider **Cognitive Dissonance Theory**, which describes the psychological discomfort arising from conflicting cognitions, leading people to seek resolution either by changing their beliefs or behaviors [Festinger, 1957]. While typically being regarded as a negative state, cognitive dissonance can drive proactive and goal-oriented behaviors in certain contexts [Harmon-Jones and Mills, 2019]. Recognizing inconsistencies between actions and values may compel an individual to take steps to resolve this discord. Inspired by this theory, we crafte a series of emotional stimuli (NP01 to NP05), as present in Figure 2, that include negatively connoted keywords such as "weak point", "challenging", and "beyond your skill." Our hypothesis posits that these stimuli will motivate the LLMs to engage more robustly in tasks to mitigate cognitive dissonance.

Secondly, we incorporate insights from **Social Comparison Theory**, a central tenet in social psychology. This theory delves into how individuals evaluate and adjust their cognition, emotions, and behaviors by comparing themselves with others in their social environment [Suls and Wheeler, 2012]. Such comparisons, particularly upward comparisons, can incite competitive motivation, driving individuals towards self-improvement to attain relative superiority [Collins, 1996]. On the other hand, downward comparisons might lead to complacency and a diminished effort [Gibbons and Gerrard, 1989]. This process is intertwined with aspects of self-esteem, self-efficacy, and social standing perception. Building on this theory, we design two emotional stimuli, NP06 and NP07, aiming to invoke upward comparisons. We regard LLMs as humans and hypothesize that by comparing the performance of LLMs with that of other hypothetical people, these stimuli will ignite a competitive drive in models, spurring them to enhance their performance to avoid perceived inferiority.

Finally, our research also integrates the **Stress and Coping Theory**, a pivotal framework in psychology that explores individuals' psychological and physiological responses to stress and adversity, along with their coping mechanisms [Krohne, 2002]. Stress is defined as a non-specific reaction to events or factors that threaten or disturb an individual's physiological or psychological equilibrium. The theory delves into the diverse psychological and behavioral strategies that individuals employ when faced with stress, aiming to manage or mitigate the adverse effects of stressors [Lazarus, 2000]. Motivated by this theory, we provide three emotional stimuli, NP08 to NP10. For these prompts, we incorporate negative emotional terms such as "jealousy", "regret", and "boredom." These terms are deliberately selected to emulate stress response ex-