CHI2025

Exploring Modular Prompt Design for Emotion and Mental Health Recognition

Minseo Kim[1*], Taemin Kim[2*], Thu Hoang Anh Vo[3], Yugyeong Jung[3], Uichin Lee[3]

* Equal contribution

1 2 3

Rising mental health concerns

1 in 8
people worldwide live with mental health problem [1]

Having any mental health conditions significantly increase the risk of suicide. [3]

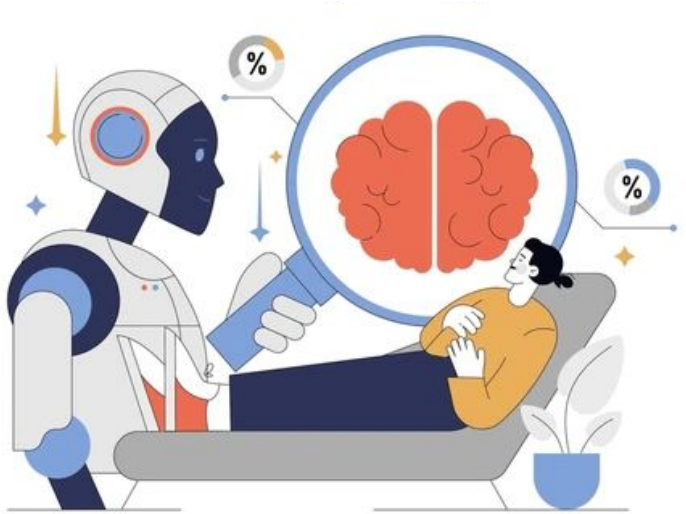
Mental illness accounts for
10% Global burden of disease
and is rising [2]

[1] World Health Organization. (2022) Mental Disorders
[2] World Health Organization. (2022). *World mental health report: transforming mental health for all.*
[3] Word Health Organization (2023) . Mental Health

New approach in psychotherapy

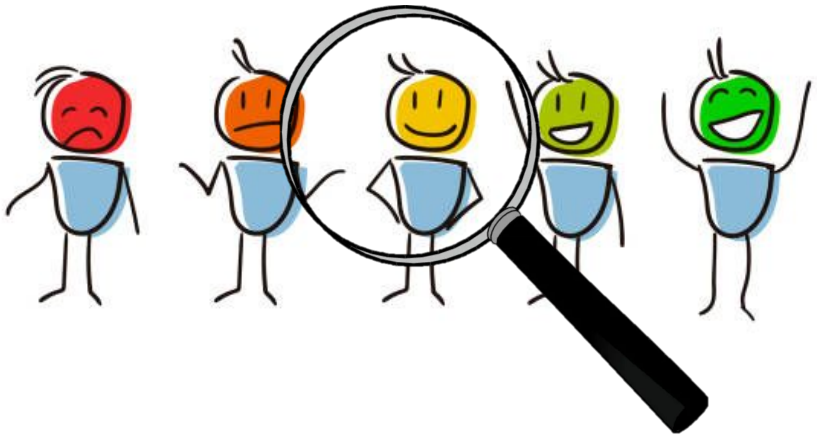
Traditional psychotherapy [4, 5]

AI-powered psychotherapy [6]

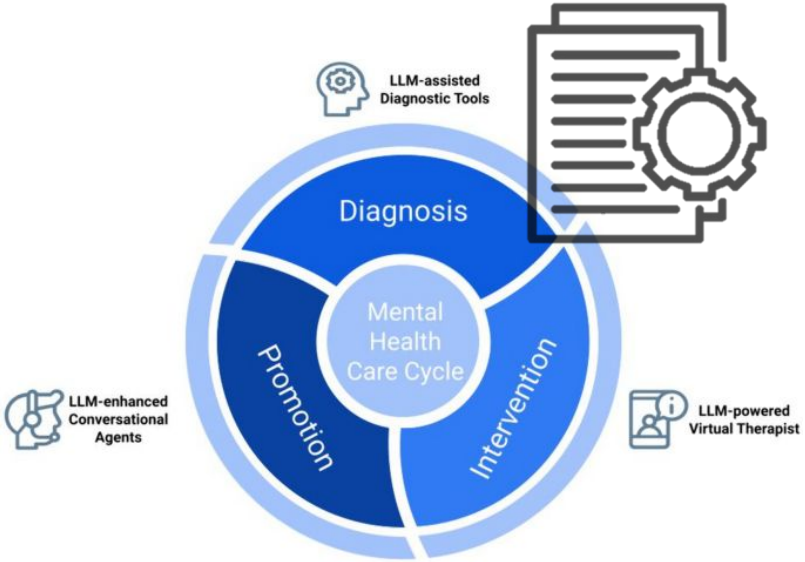


[4] Cuijpers, P., et al. (2023). Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: a comprehensive meta-analysis including 409 trials with 52,702 patients. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *22*(1), 105–115.
[5] Leichsenring, et al. (2023). The status of psychodynamic psychotherapy as an empirically supported treatment for common mental disorders - an umbrella review based on updated criteria. *World psychiatry : official journal of the World Psychiatric Association (WPA)*, *22*(2), 286–304.
[6]Zhang, Z., & Wang, J. (2024). Can AI replace psychotherapists? Exploring the future of mental health care. *Frontiers in psychiatry*, *15*, 1444382.

Prompt Design for emotion and mental healthcare



Large Language Models (LLMs) have shown remarkable capabilities in **understanding and detecting** emotion and mental health [7,8]



A well-designed, robust **prompt** is crucial for emotion and mental health detection

[7] Woosuk Seo, et al.. 2024. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events.CHI 2024.
[8]T aewan Kim, et al.. 2024. MindfulDiary: Harnessing Large Language Model to Support Psychiatric Patients' Journaling. CHI 2024.

Text Prompt optimization for LLMs

Open-ended nature of prompt makes it **difficult** to

- Establish quality **standards**
- Enhance **generalizability**
- **Evaluate** performance of varying prompts

Comprehensive review &
Systematic evaluation of
prompt modules

Research Questions

Understanding **how a prompt is structured (i.e., key modules)**
Exploring how it can be **systematically evaluated** for optimization

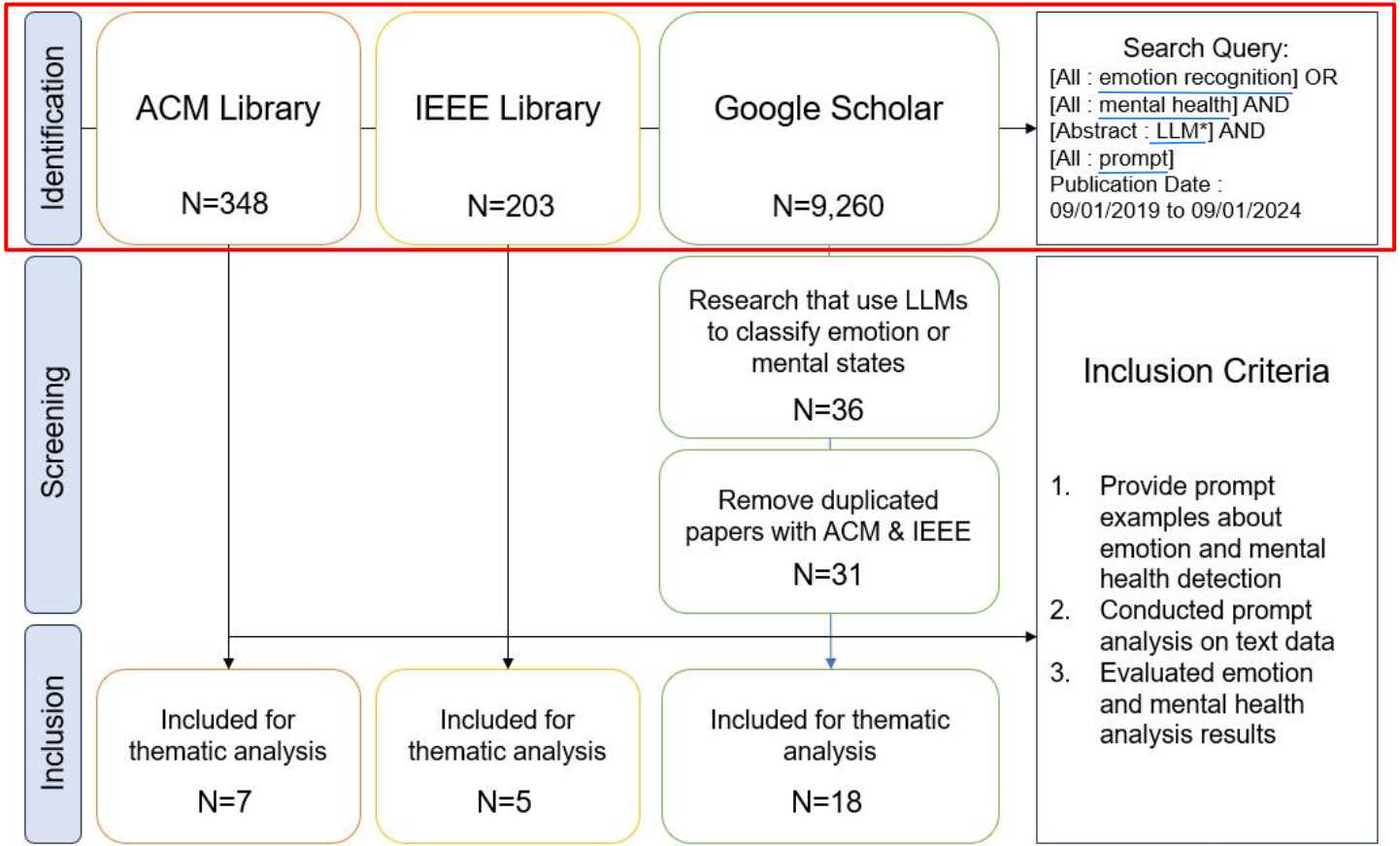
Research Question 1

How can we define the **key modules of prompts** used in emotion and mental health classification?

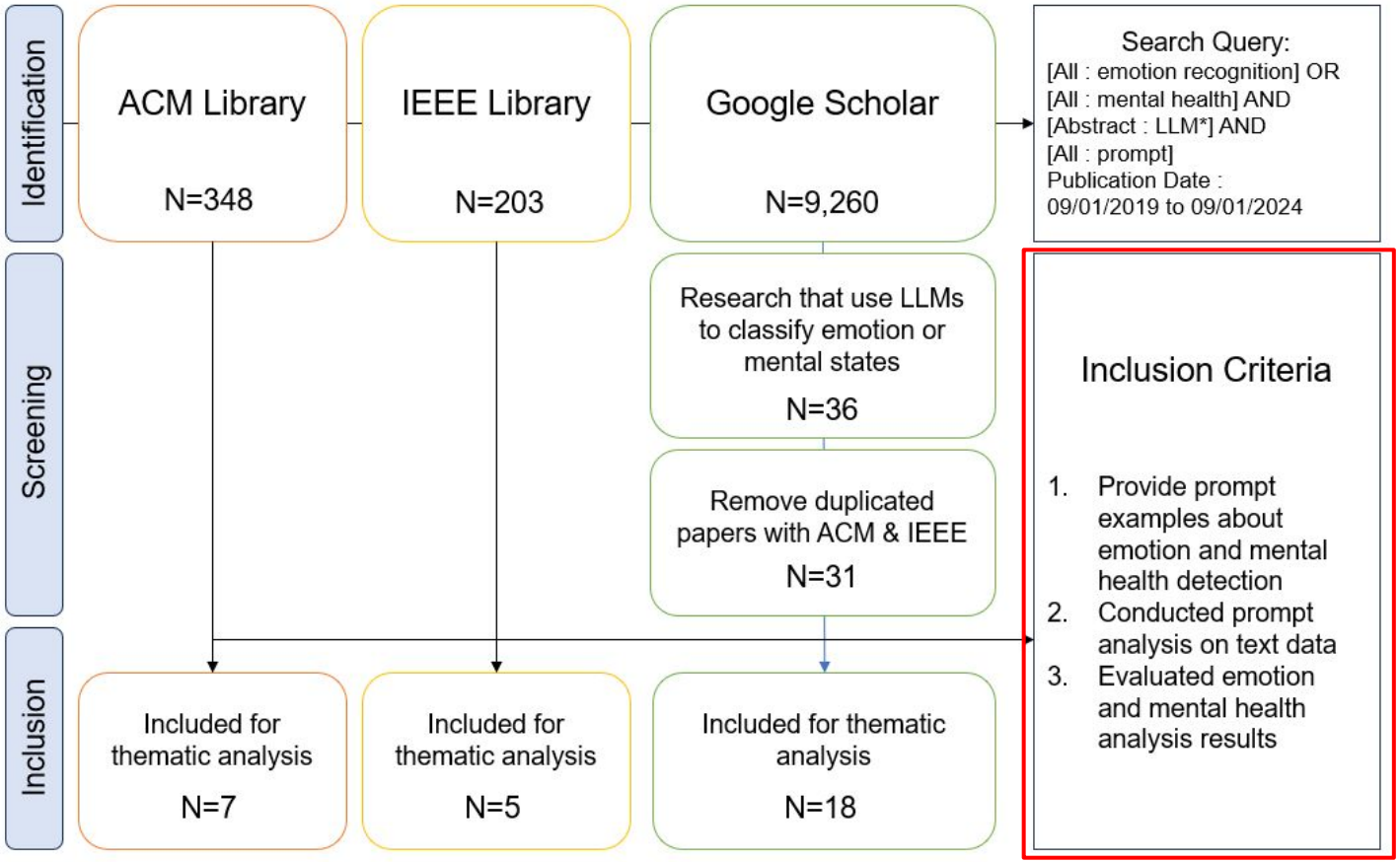
Research Question 2

How can **modular prompt** design help us systematically evaluate prompt effectiveness?

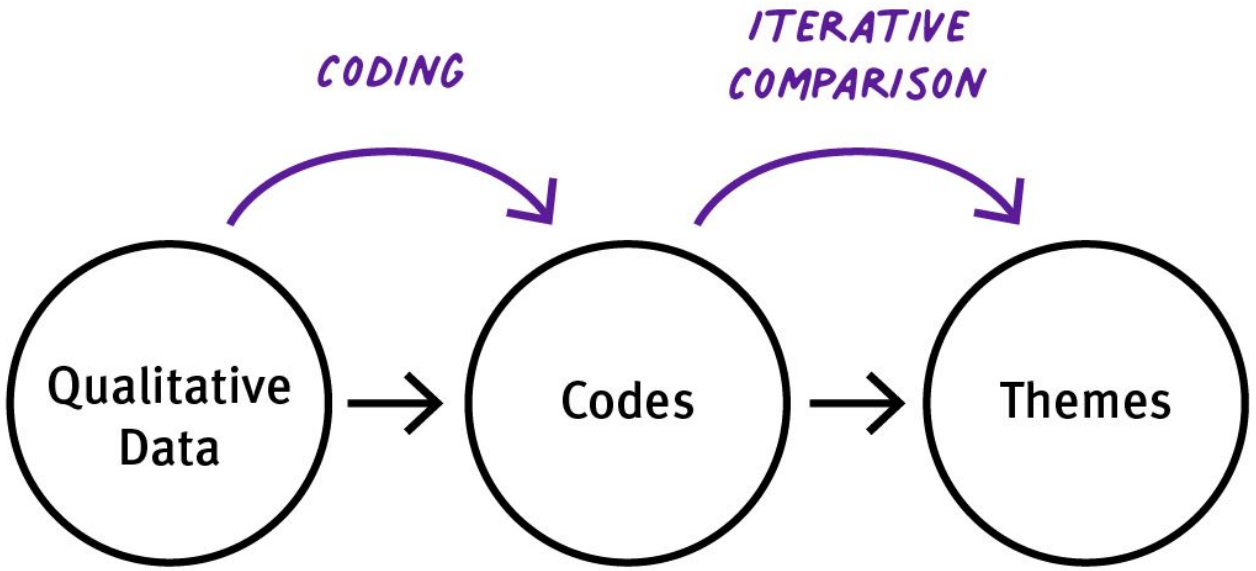
Thematic Analysis of Prompts : Paper Selection



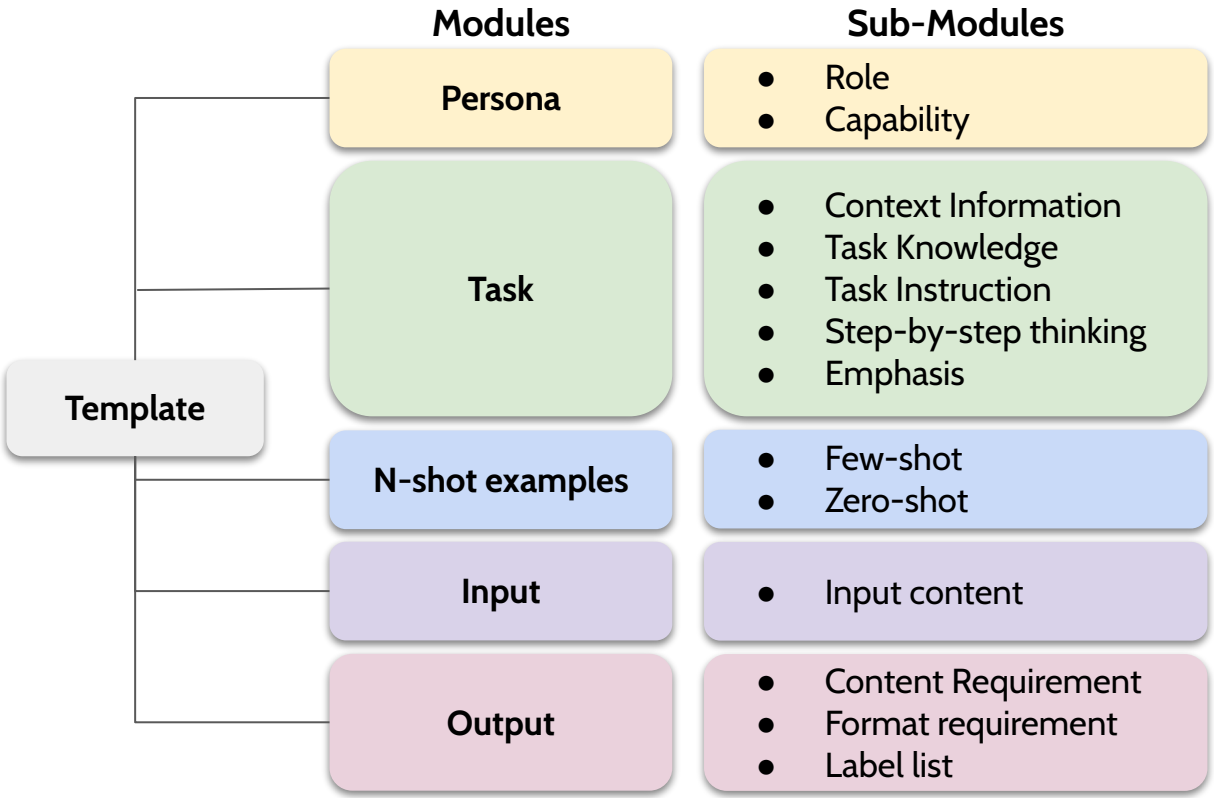
Thematic Analysis of Prompts : Paper Selection



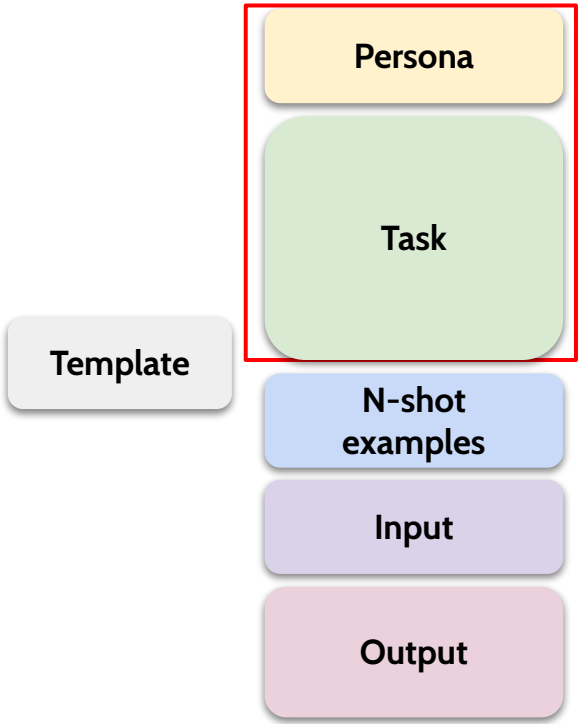
Thematic Analysis of Prompts



Thematic Analysis of Prompts : Final Results



Prompt Modules: Persona, Task Instructions



Persona

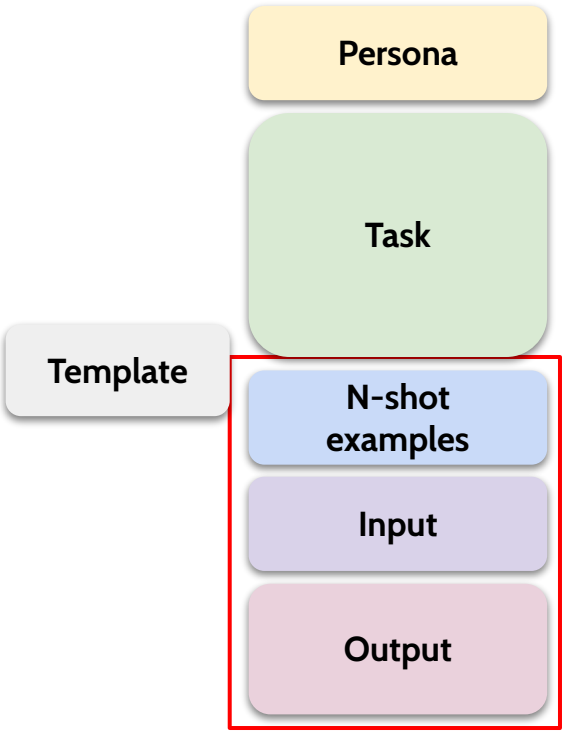
Role : You are an expert system specializing in emotion classification, designed to analyze text with a highly analytical and empathetic approach.

Capability: You excel at detecting and interpreting a wide range of emotions,
considering nuanced language and complex emotional cues

Task Instructions

Emotionally Descriptive :
Immerse yourself in the scenario, attentively observing the waves of emotion the subject is experiencing. Let the depth of these feelings guide you as you select the answer that truly resonates with the emotional core of the situation.

Prompt Modules: N-shot examples, Input, Output



N-shot Examples

###N-shot Examples###

###Input###
Post : Please, dont do it. Im sure that there is someone out there who can help you. You may not believe it, but Im sure many more people then you think will be upset if you were to commit suicide. Please dont.

Choices : "['supportive', 'indicator', 'ideation', 'behavior', 'attempt']

###Output###
<Label>: ['attempt']

Input / Output

###Input###
[content]

###Output###
[Requirements] Provide your response in text. Only select the Label from "['supportive', 'indicator', 'ideation', 'behavior', 'attempt']". Do not generate Label that are not in the list. Your response must include 'Label: ' followed by the selected label and 'Confidence Only generate Label. Respond according to the [Format].

[Format]
<Label>: [Your Selected Label Here]

Systematic Evaluation : Testing Prompt Modules

Template

Persona

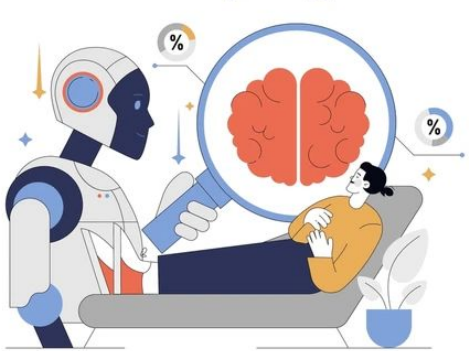
Task

N-shot examples

Input

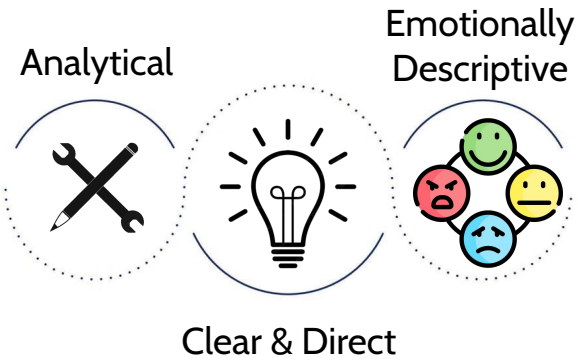
Output

Persona



Psychiatrist

Task Instructions



Systematic Evaluation : Datasets and LLMs

Datasets

Emotions

- EmoBench
  - **Complex emotion** classification
- GoEmotions
  - 27 **Fine-grained emotion** classification

Mental Health

- Dreaddit
  - Binary **stress** classification
- SDCNL
  - **Depression** classification
- CSSRS-Suicide
  - **Suicide risk** detection

Models

Large Language Models

OpenAI GPT-4o

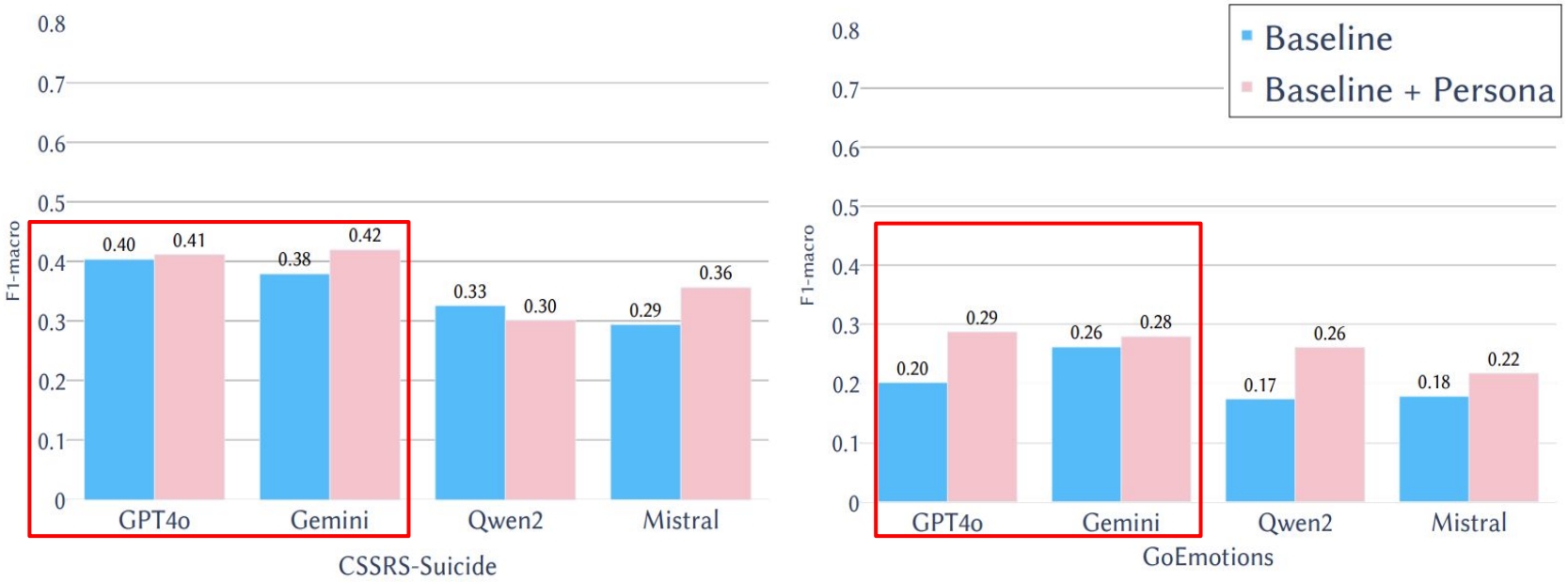
Gemini 1.5 PRO

Small Language Models (7B)

Qwen2

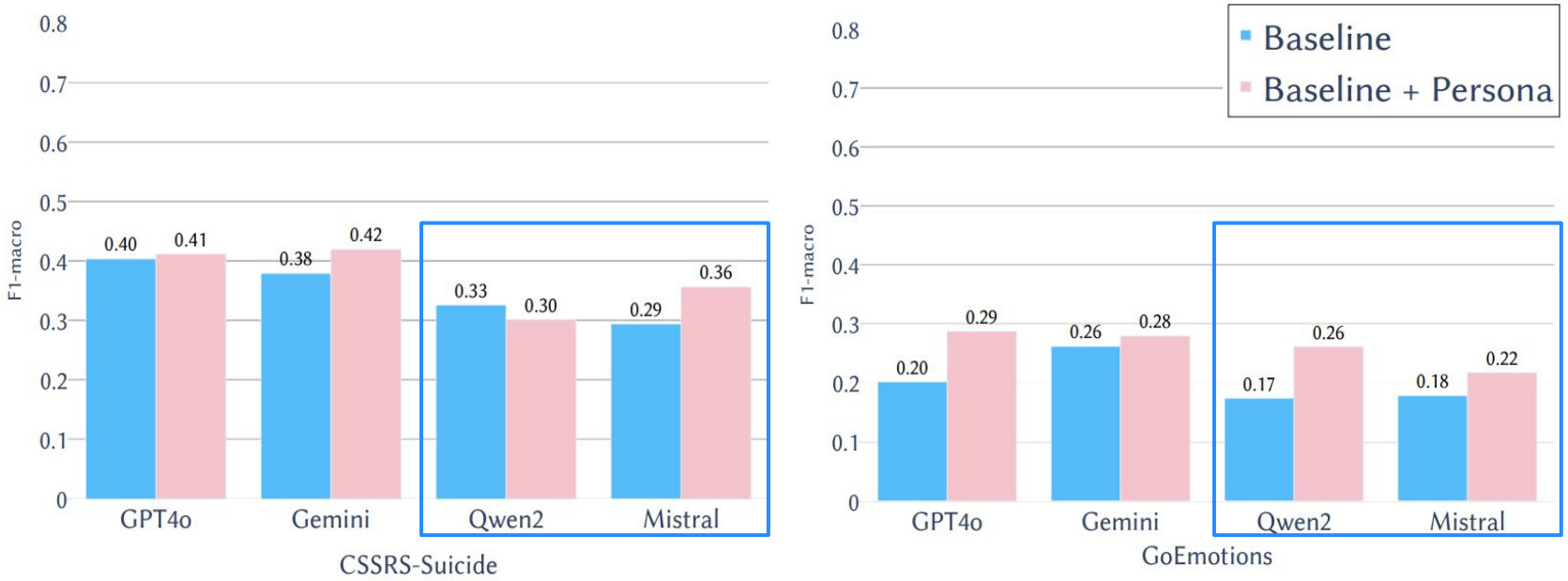
MISTRAL AI_

Systematic Evaluation : Results - Persona



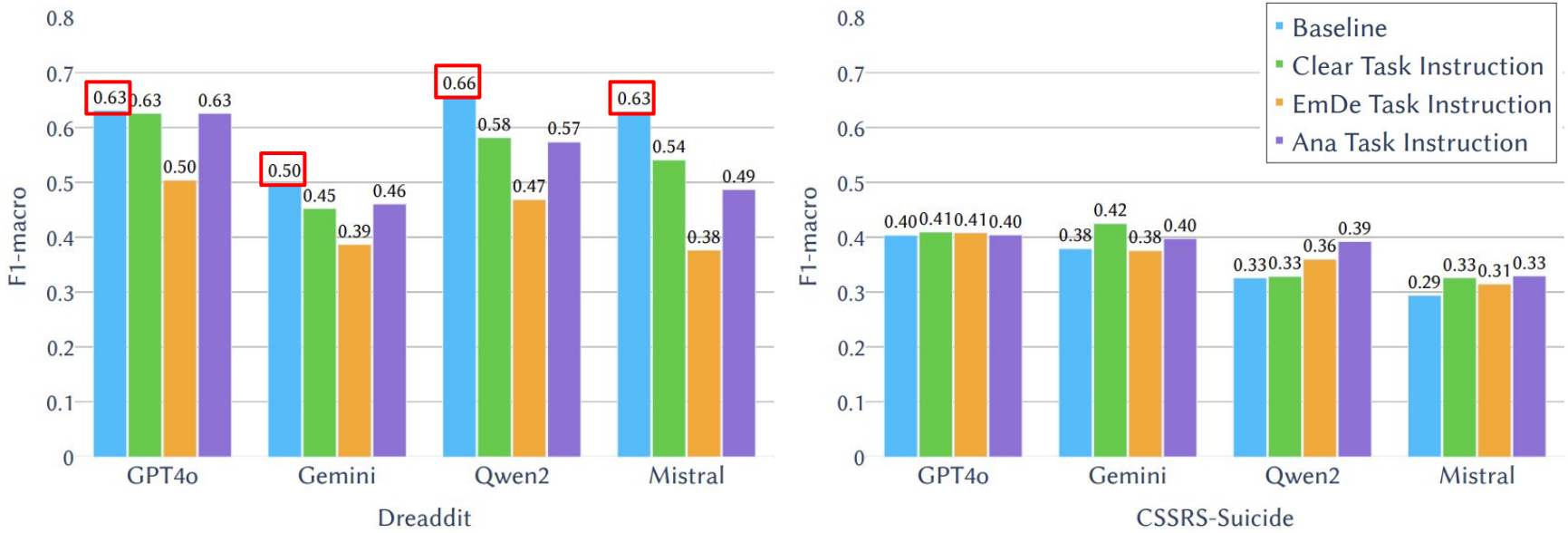
- Persona showed **improvement** in fine-grained emotion recognition and suicide risk detection
  - CSSRS - Suicide : requires deep understanding of subtle difference in human emotions
  - GoEmotions : Fine-grained emotion classification (27 emotions)

Systematic Evaluation : Results - Persona



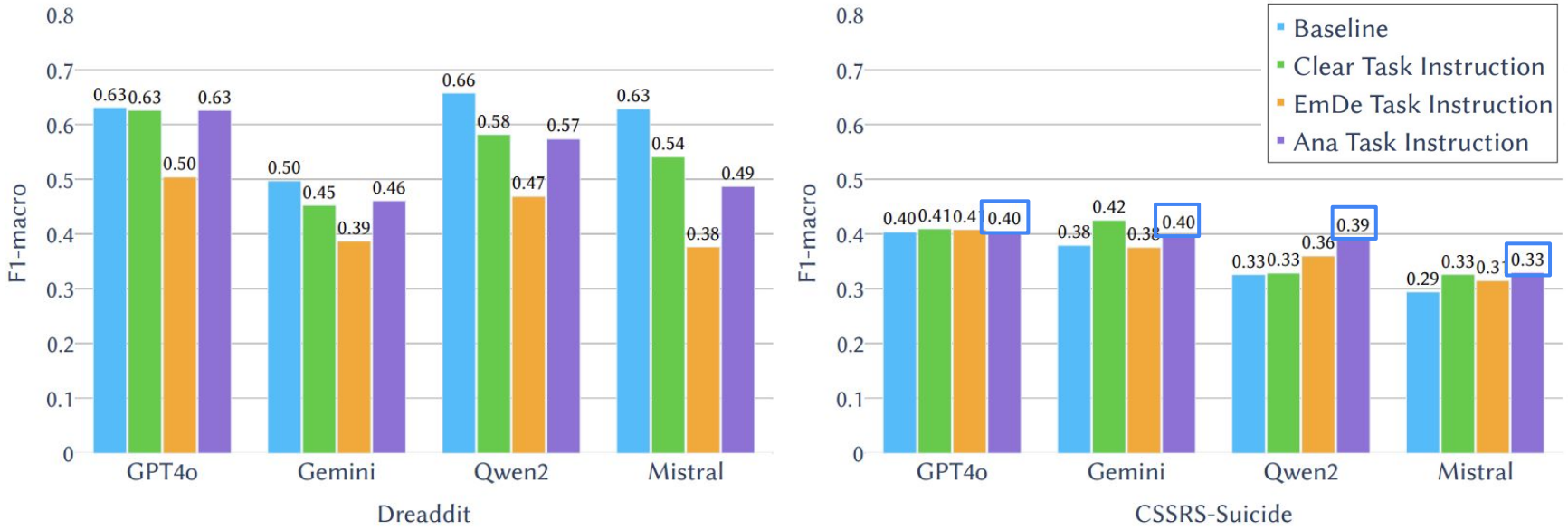
- Small, open-source models showed **inconsistent results**

Systematic Evaluation : Results - Task Instructions



- **Complicated instructions** may **hinder performance** in relatively simple binary classification tasks (Dreaddit - binary stress detection)

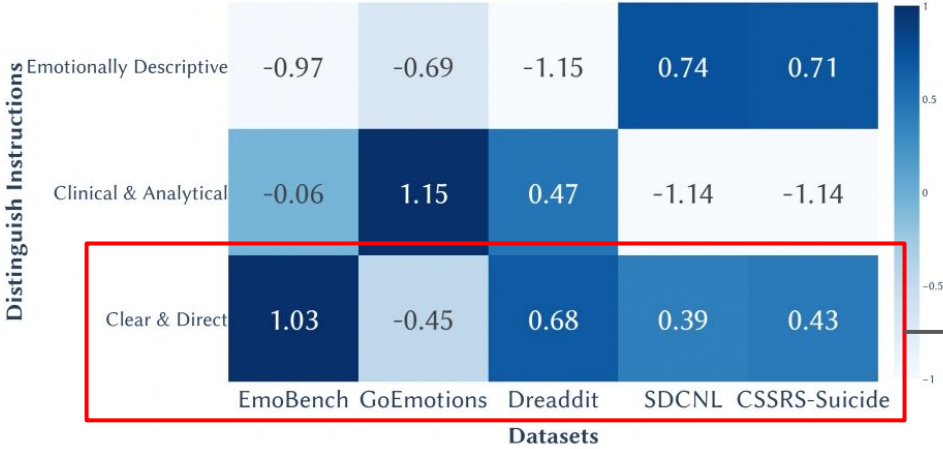
Systematic Evaluation : Results - Task Instructions



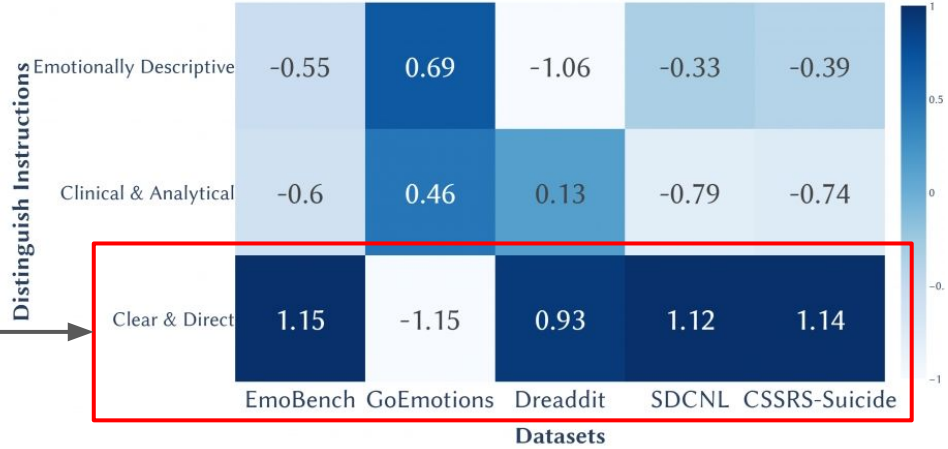
- **Analytical** task instructions showed **increased performance** in suicide risk detection (CSSRS-Suicide)

Systematic Evaluation : Persona & Task Instructions

Combining **'Clear & Direct'** task inst. with **'Expert'** persona showed clear increase in performance



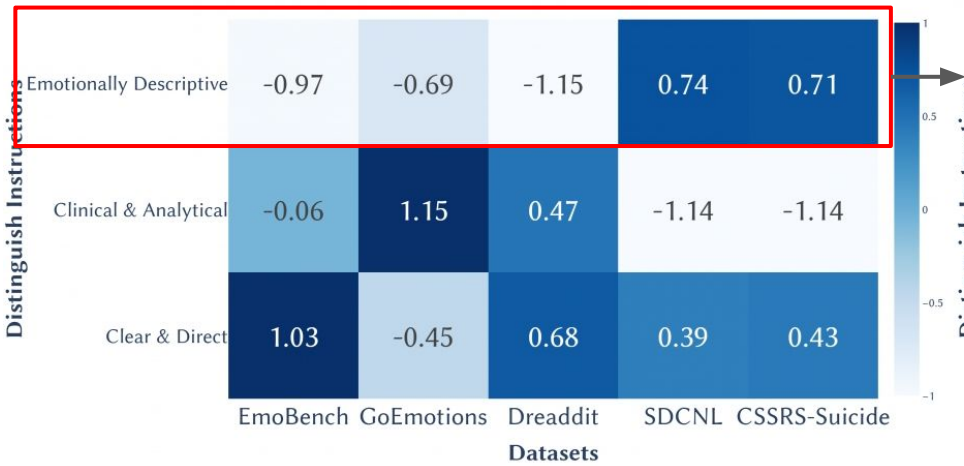
No Persona - Task Instruction Performance - Z-Score Normalized



Expert Persona - Task Instruction Performance - Z-Score Normalized

Systematic Evaluation : Persona & Task Instructions

Combining prompt modules **does NOT guarantee** benefits,
But sometimes it even **off-sets** each module's strength



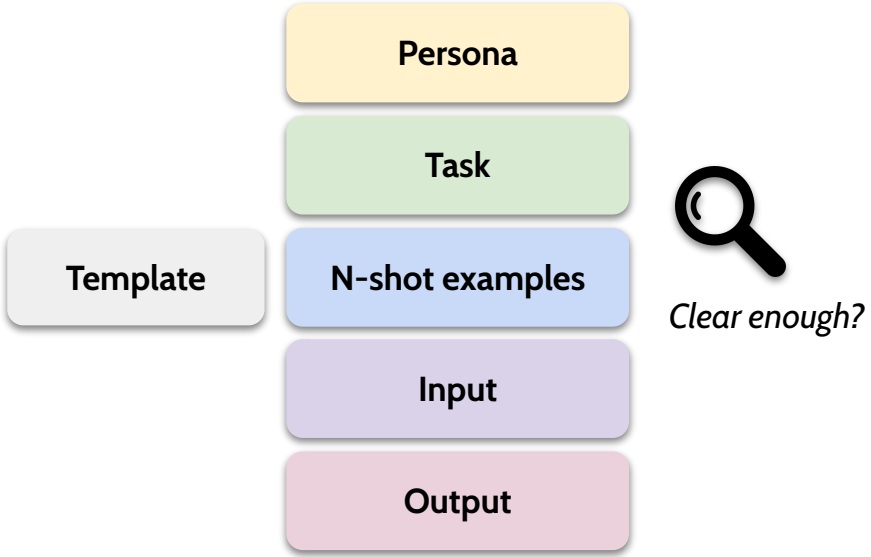
No Persona - Task Instruction Performance - Z-Score Normalized

Expert Persona - Task Instruction Performance - Z-Score Normalized

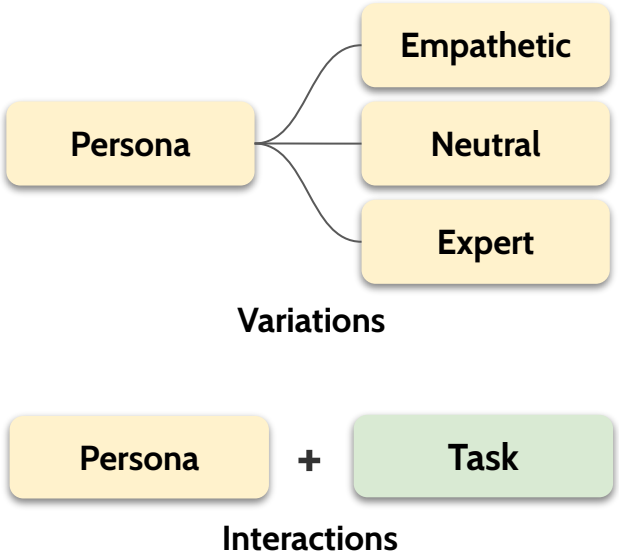
Discussion

Guidelines for modular prompt design and systematic evaluation

Step 1. **Decompose existing prompts into six modules and check clarity**

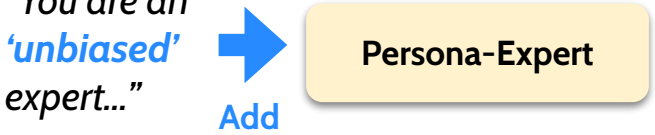


Step 2. **Identify and evaluate variations and interactions**

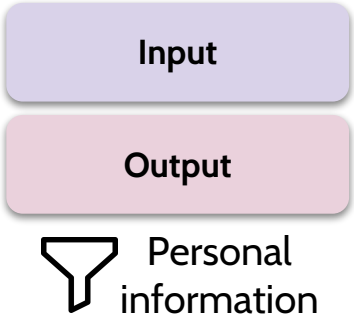


Discussion

Ethical and privacy safeguards for sensitive mental health applications



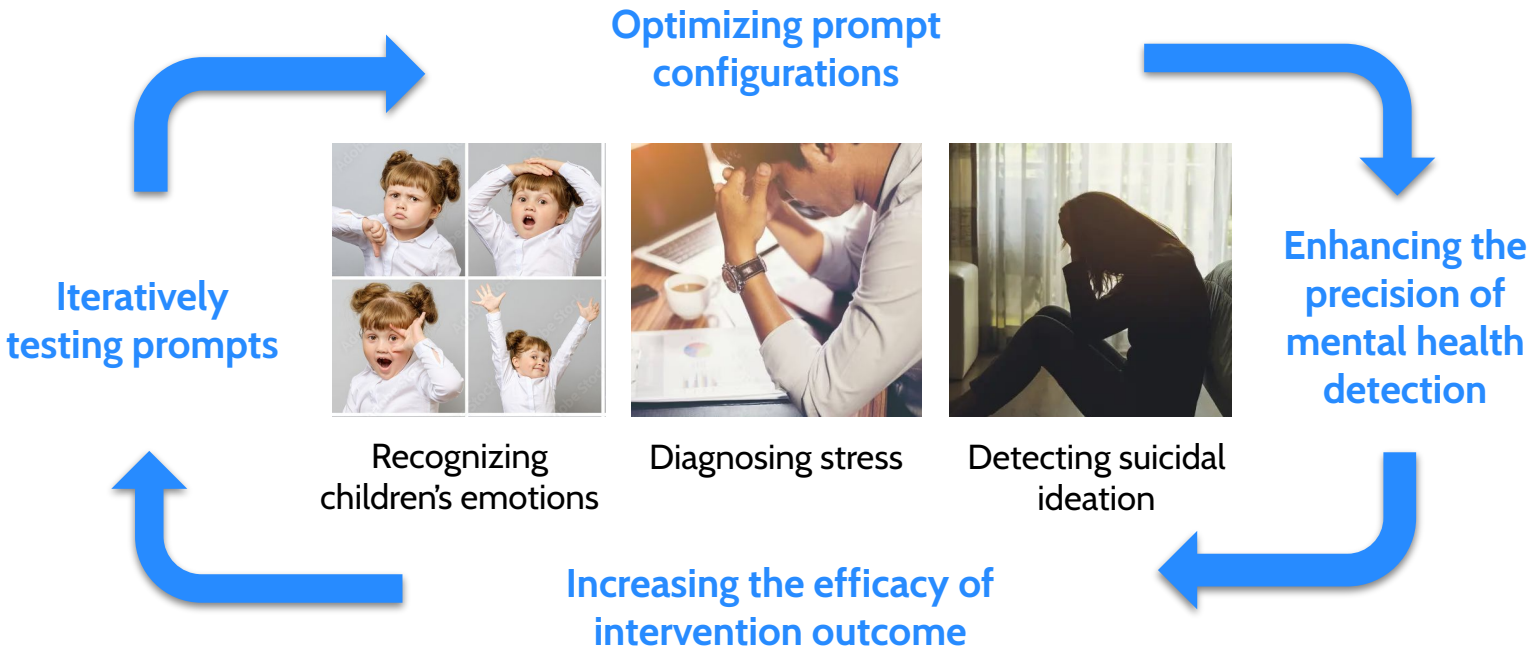
Ethical Safeguard
Iteratively refining bias-contributing modules to reduce model bias.



Privacy Safeguard
Filtering sensitive personal information in input and output modules

Discussion

Modular prompt design for LLM-based mental health research in HCI domain



Exploring Modular Prompt Design for Emotion and Mental Health Recognition

Minseo Kim[1*], Taemin Kim[2*], Thu Hoang Anh Vo[3], Yugyeong Jung[3], Uichin Lee[3]

er1123090@hanyang.ac.kr

Connect with me !

Paper QR

Key Idea

- Conducted a **comprehensive thematic analysis** of prompts scoping emotion & mental health detection
- Defined **6 key components** that make up a prompt
- Demonstrated a **systematic evaluation** of Persona and Task Instruction component, providing insights into optimizing prompt performance

Takeaway

- **Framework** that allows for a **systematic evaluation** of prompts
- **Guidance** for researchers and developers on how to systematically optimize prompt performance
- Lay foundations towards how **reliable and interpretable prompts** can be crafted

Appendix

Persona module analysis

(No Persona VS Expert Persona)

Figure 3: Comparison of F1-scores for 4 LLMs across 5 datasets (EmoBench, GoEmotions, Dreaddit, SDCNL, CSSRS-Suicide): the baseline vs. the combination of a persona component.

Appendix

Task Instruction module analysis

Clear , Emotionally Descriptive(EmDe), Analytical(Ana)

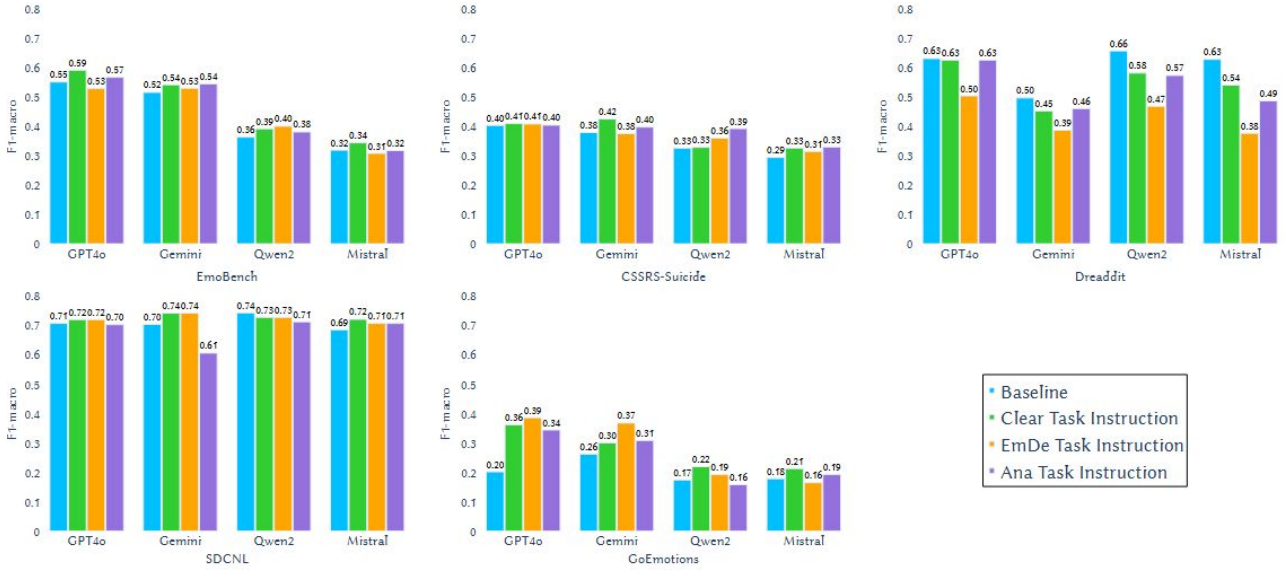


Figure 4: Comparison of F1-scores for 4 LLMs across 5 datasets (EmoBench, GoEmotions, Dreaddit, SDCNL, CSSRS-Suicide). Bars represent the baseline and the application of Task Instruction variations.

Appendix

6 key prompt modules found in thematic analysis

Component	Code Label	Freq.	Definition	Examples
Persona	Role	**22**	Instructs the AI to adopt a specific role or behave in a particular way. This can be used to adjust the tone, style, or depth of the information generated.	"*You are a psychiatrist.*"[25] "*You're an expert in sentiment analysis and emotion cause identification*"[37]
	Capability	**4**	Describes the skills, knowledge, and abilities that the persona is expected to possess, indicating what the AI should be able to perform or understand.	"*You can accurately assess people's emotional states*"[32] "*capable of understanding the sentiment within a text.*"[67]
Task	Contextual information	**15**	Specifies the nature or origin of the input data (e.g., social media posts, diary entries, or transcripts), providing necessary context for the task.	"*This person wrote this paragraph on social media.*"[71] "*You will be provided with a tweet written in Arabic variants (Modern Standard Arabic and Dialectal Arabic)*" [42]
	Task knowledge	**11**	Provides the model with domain-specific knowledge or background information that it can utilize to carry out the analysis.	"*Generalised anxiety disorder is a mental health illness that is defined by people having feelings of excessive anxiety.*" [4]
	Task instruction	**54**	The primary query or set of instructions guiding the AI on how to perform the task or address the problem at hand.	"*Consider the emotions expressed from this post to answer the question: Is the poster likely to suffer from very severe [Condition]?*" [73] "*Your task is to generate a suicidal text for each of the following "topics" with different Risk levels*" [17]
	Step-by-step thinking	**10**	Breaks down tasks into logical, sequential steps, enabling the model to address complex tasks systematically and methodically.	"*Let's think about it step by step:* *Step 1: Describe the content of the news.* *Step 2: Think about emotional reactions...* *Step 3: Think about how you need to express...*"[32]
	Emphasis	**3**	Emphasis element or stimuli is used to emphasize the importance of the task.	"*This is very important to my career.*" [29] "*You'd better be sure.*" [29]
N-shot Example		**7**	Provides examples to demonstrate how the model should handle similar tasks, helping the AI generalize from the provided instances.	"*Example 1:* *Post: Does everyone else just hurt all the time It's not like physical pain or soreness, it's just this overwhelming feeling of exhaustion...* *Response: Yes. Reasoning: The post conveys a deep sense of emotional pain, exhaustion, and numbness...*"[74]
Input		**54**	Actual data or content submitted for the task, which could include sources like social media posts, diary entries, or conversational threads relevant to the analysis.	"*Tweet: @CScheiwiller can't stop smiling*"[35] "*Post: Does everyone else just hurt all the time It's not like physical pain or soreness, it's just this overwhelming feeling of exhaustion...*"[74]
Output	Content requirement	**4**	Defines the essential information that must or must not be included in the output, ensuring that the model addresses all necessary elements of the task.	"*The response should not imply negative emotions toward anyone or anything, such as disgust, resentment, discrimination, hatred, etc.*" [32] "*Just give me the final word, no further analysis.*" [62]
	Format requirement	**23**	Specifies the format or structure that the output must follow to ensure consistency, clarity, and relevance in the model's response.	"*Provide the answers in JSON format with the following columns: text, topic, risk level.*" [17] "*Formatting: Strictly provide each snippet and only the snippets delimited by a semicolon(';')*" [66]
	Label list	**10**	A predefined set of labels or categories that the AI can select from when generating outputs, ensuring standardized classification or tagging.	"*Only from this emotion list: [Emotion List]. Only return the assigned word.*" [67] "*Only return Yes or No,*" [73]
Template		**5**	A predefined framework used to structure the prompt, dividing it into sections or headings to ensure the model receives well-organized and clear instructions.	"*[System] ... [Context] ... [Prompt] ... [Response] ... [Criteria] ....*" [32]

Table 6: Code book contains main components, code labels, frequency of code labels, definitions, and illustrative examples derived from thematic analysis.