How can I help you today?

Research on Empathy Ability & Response Naturalness in ChatGPT 4

Agenda

Why are we doing this & Goal
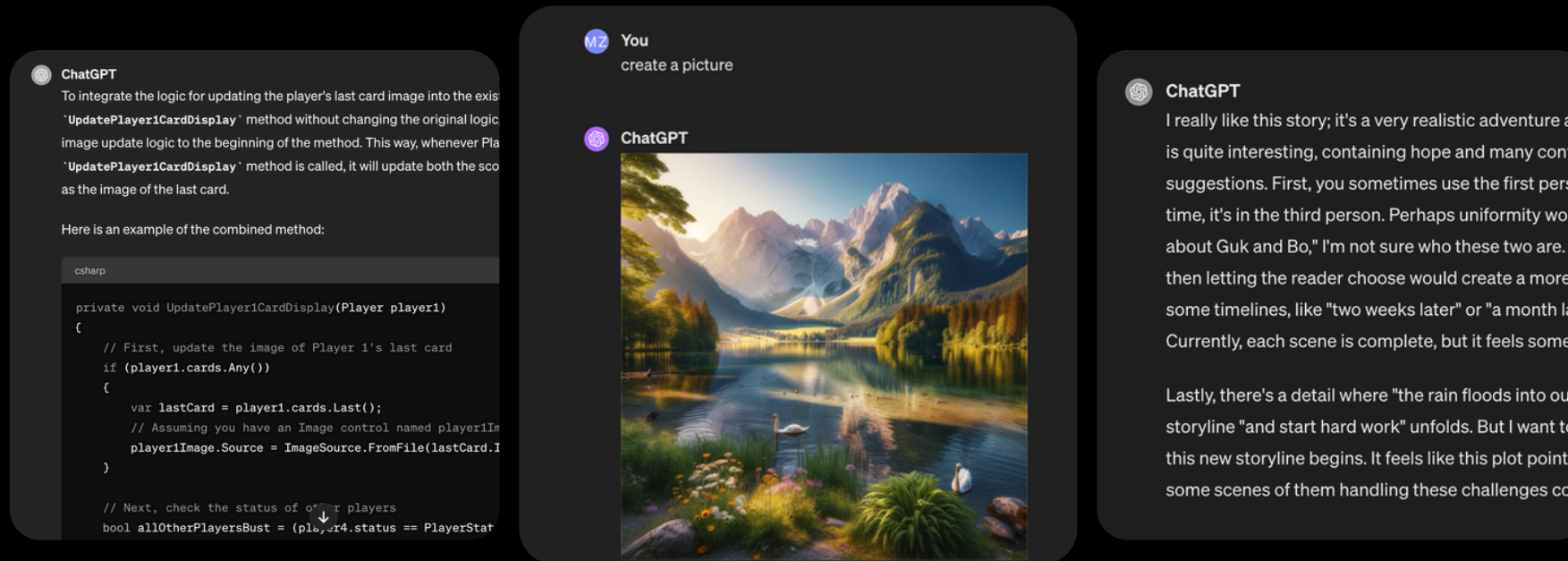
📈

3.5 (out of 5)

Need for Emotional Support

User Satisfaction Rate from my previous research

GOAL:

Collect qualitative and quantitative data to evaluate user satisfaction with ChatGPT's conversational responses and to assess its capability to demonstrate empathy.

Introduction

ChatGPT4 is an AI-based conversational language model developed by OpenAI. It uses its trained model to compute the answers users are most likely seeking, providing assistance with various issues through natural language text.

As of September 2023, it has **150** million active users. Users include students, programmers, creators, and more.

Methodology

01 **Competitive Analysis** | Presentations are communication tools that can be used as lectures.

02 **Survey** | Presentations are communication tools that can be used as lectures.

03 **User Interview** | Presentations are communication tools that can be used as lectures.

04 **Empahy Questionnaires** | Five questionnaires to evaluate ChatGPT4's empathy

Competitive Analysis

Features

	ChatGPT 4	Gemini	Copilot
Voice Conversation	✅	❌	❌
Pin Conversation Boxes	❌	✅	❌
Links in the Response	❌	✅	✅
A Guess What You Want to Ask	❌	✅	✅
Integrated into Programming Environment	❌	❌	✅
Design Your Own GPT	✅	❌	❌
Image Creating	✅	❌	❌
Excel Analysis	✅	❌	❌
Price	$25 per month	$19.99 per month	$20 per month

Advantages

Gemini (Google)

Value & Business Goal
With Google ecosystem, it is connected to Google's vast user base to make better daily decisions

- **Google's native ecosystem**
Integrates into Google's native ecosystem, such as Google Flights, Google Maps, YouTube, etc., and their users are almost everyone in the US

- **Response fast**
- **Lower price**
- **Double-check feature**

Copilot (Microsoft)

Value & Business Goal
Primarily aimed at developers, assisting users to enhance user work efficiency

- **Real-time coding assistance**
As an IDE plugin, it is directly integrated into the developer's programming environment.

- **Relatively vertical domain**
Focus on helping programmers write code.

- **Accuracy**
Insert links to support the answer

- **Provide relevant topic links at the end**

Advantages

ChatGPT4 (OpenAI)

- **The Latest Deep Learning Architectures**
Such as the Transformer model, interpret users' intentions more accurately and therefore generate contextually appropriate responses.

- **Plugins**
Such as a literature reading plugin, to cater to personalized needs.

- **Uploading Files**
ChatGPT can read, summarize, and analyze the file.

- **GPTs**
Users can design their GPTs or utilize those designed by others.

- **Team Workspace**
Facilitate collaboration.

- **Voice Communication**
Streamline interactions, making ChatGPT more "human".

- **Image Creating**

What should we learn from them?

Value & Business Goal

Progressing towards Empower AGI means that it not only needs to excel in specific domains but also be able to "Enhance more comprehensive and widespread general intelligence.

- **Cross-platform Integration**
  Enhance the ability to assist on various platforms and services, whether it's travel planning, learning resources, or business tools.

- **Fast Response**
  Provides expert-level features in specific areas, such as programming or data analysis.

- **Provide links to support answers**

Survey

Quantitative data on ChatGPT 4's emotion recognition ability and overall response satisfaction

INTRODUCTION

I surveyed **63** participants to understand their experiences regarding satisfaction, including the naturalness of responses, the acquisition of new insights, and daily & emotional support performance.

Survey includes two instances of survey logic and an open-ended question.

PARTICIPANTS

- University master students, workers, and professionals (**n=63**)
- Aged 25 to 35,
- Use ChatGPT at least **5 times a week**,
- Expresses interest in expressing their emotions to ChatGPT.

My survey's link

Primarily use ChatGPT for?

6.10%
1.22%
3.66%

30.49%
Study Assistance

30.49%
Writing Assistance

28.05%
Information Searching

■ Study assistance ■ Information search ■ Writing assistance ■ Entertainment and casual chat ■ Emotional needs ■ Others
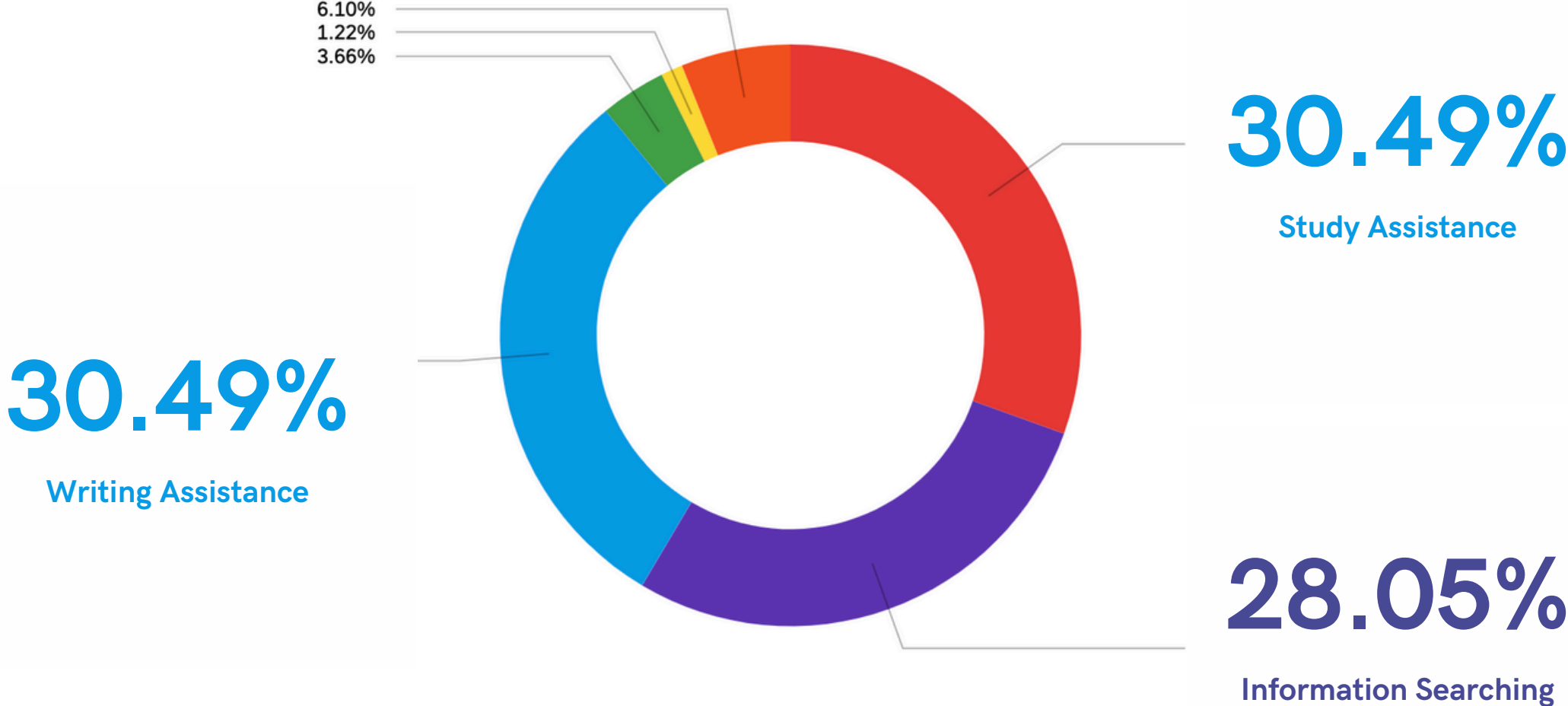
Survey - Brief Data   (arranged from lowest to highest)

4.34 out of 10
Emotional Questions Satisfaction Rate

40%
Have asked emotional or daily life questions

5.76 out of 10
Naturalness

6.71 out of 10
Overall Satisfaction Rate

53.13%
Offer new Insights

6.69 out of 10
Meet needs

70.96%
Somehow satisfied or satisfied with the response

72.88%
Using ChatGPT for 30 minutes or less each day

74.46%
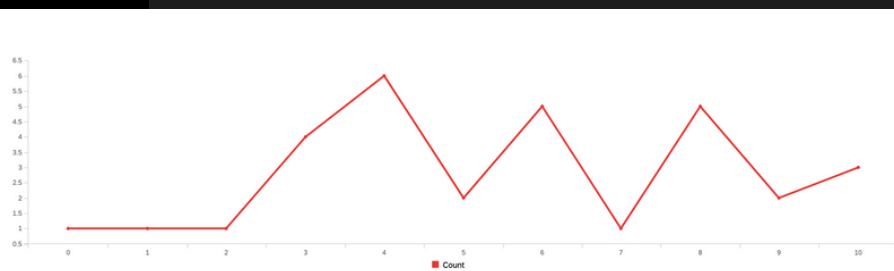User-friendly

My reslut page   **My whole survey report**

Survey - Pattern and Insights

01 Naturalness in Different Tasks

- **Pattern**:
There's a large standard deviation in ratings for naturalness.

- **Insight:**
Naturalness perception may vary with different tasks and needs, requiring further clarification during interviews.

02 Low Naturalness, Lower Satisfaction

- **Pattern:**
The low naturalness is a major reason for the low satisfaction.

- **Insight:**
Provide natural responses, including incorporating colloquial expressions and adjusting sentence structures.

3/10 How natural do you find ChatGPT's responses in conversation? (0 unnatural to 10 very natural) 31 ⓘ
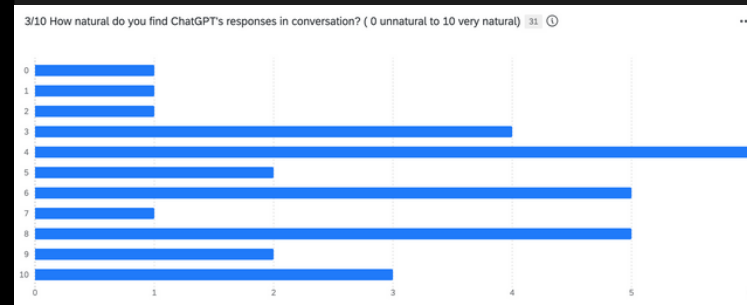
03 Low Empathy

- **Pattern:**
83% of participants who asked about emotional issues expressed chatGPT's low empathy.

- **Insight:**
Provide emotional value, giving them personalized advice and human-like reactions.

4/10 Would you ask ChatGPT for advice on emotional or life matters? 34 ⓘ
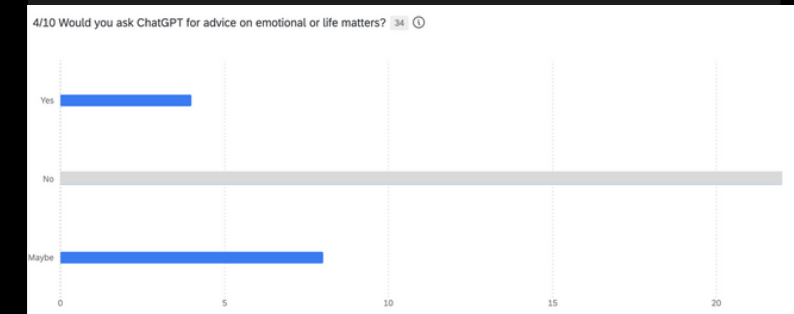
Survey - Pattern and Insights

04 Innovative Content

- **Pattern:**
Users who report satisfaction with the new insights provided by ChatGPT are more likely to perceive its responses as accurate.

- **Insight:**
Ensuring the accuracy of the GPT model in generating content that is both innovative and relevant is crucial.

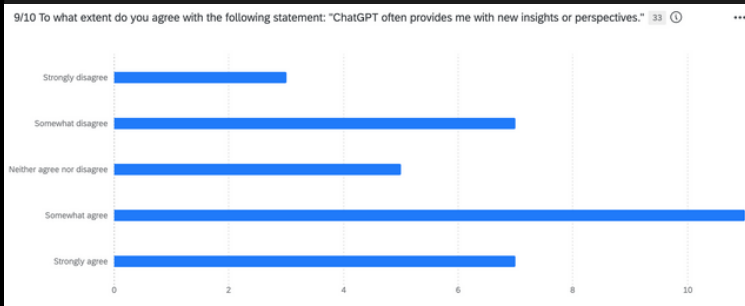9/10 To what extent do you agree with the following statement: "ChatGPT often provides me with new insights or perspectives." 33

05 Low Credibility and Input Capacity

- **Pattern:**
Users commonly desire links in replies, and the ability to input more text.

- **Insight:**
Use links to support responses' credibility and increase the text input capacity to accommodate complex programming needs.

Word limit

I would like to see links to articles that might be useful for a topic

Being able to read links and images.

"TOP" function and my GPTs can have history, more content input for my programming questions

Length limit of words count to be increased for free version

up to date with resources if possible

Interview

Qualitative data on the ChatGPT 4's naturalness of response and its performance on Emotion Recognition
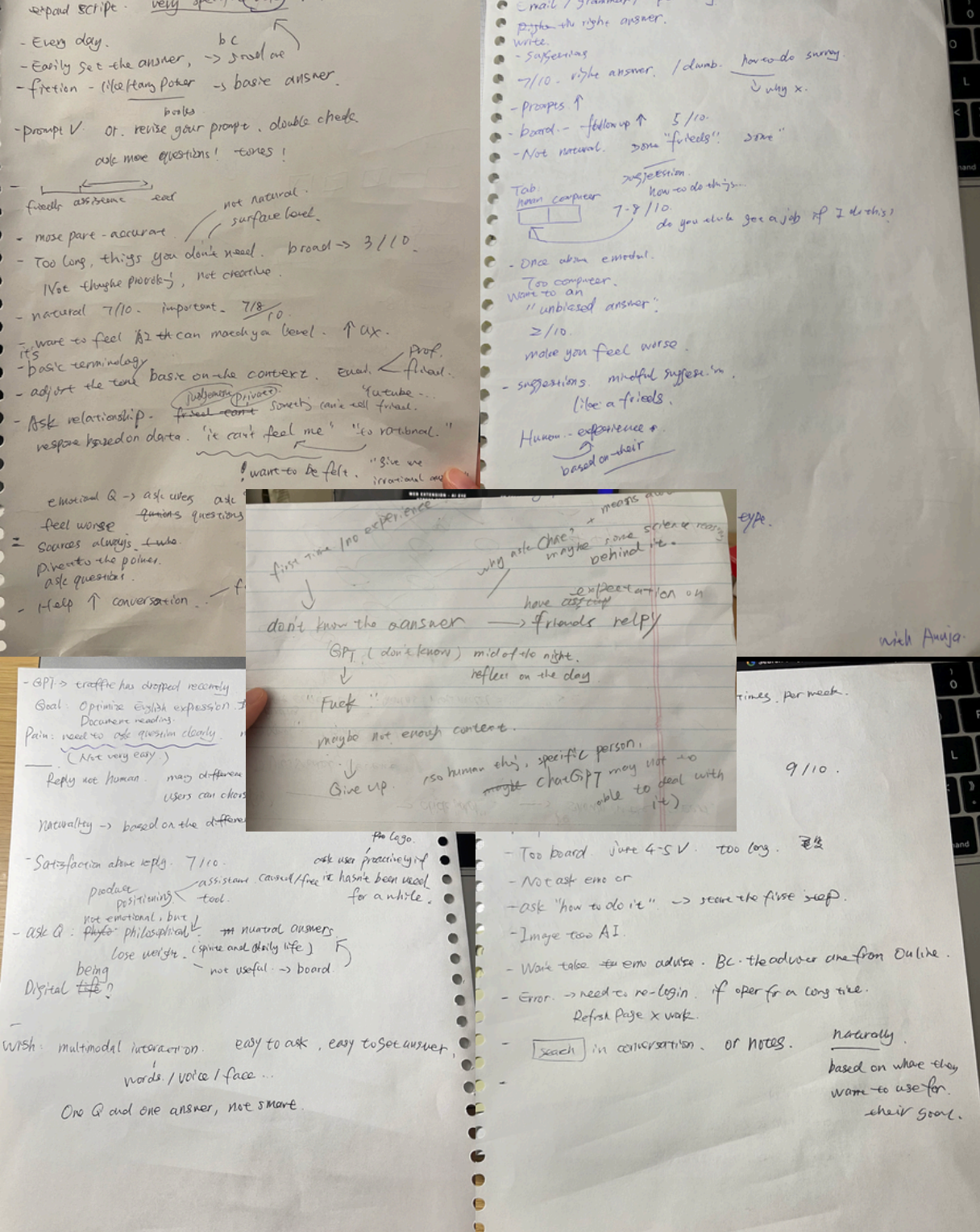
INTRODUCTION

The objective is to explore the underlying reasons for the perception of low naturalness and the need for empathy.

Each interview lasts **30** minutes and comprises **10** questions

PARTICIPANTS

- Total **5** participants (2 female, 3 male)
- University master students (**n=2**), **1** professional and **1**worker (include native English speakers)
- Aged 25 to 35,
- Uses ChatGPT at least 5 times a week

Interview - Findings

"Too robotic!"

"Too much unnecessary info."

👎 **Low Naturalness**

- Four participants rated the importance of naturalness as high as 8 out of 10.

- Naturalness significantly impacts conversational satisfaction.

👎 **Not Casual**

All participants use ChatGPT for tasks like email correction, requiring responses to be more natural, utilize relatively casual vocabulary, and be easily understandable.

Scenarios

The need for naturalness should be determined by different use scenarios.

Interview - Findings

"ChatGPT can't feel me.."

"Too neutral.."

Personal Support

Among the four participants, three sought advice from ChatGPT on matters of life, personal philosophy, or emotions.

👎 **Low Empathy**

- The responses felt too neutral and mechanical, indicating ChatGPT's low empathetic capacity.

- Two participants reported feeling worse after their consultations.

Why not ask friends?

Their friends' opinions are often based on personal experiences, and may not necessarily apply to others.

They want to get some unjudgemental advice.

Interview - Findings

"It's tiring to type very specific requirements every time."

👎 **Lengthy and Broad**

- Half of the participants mentioned that ChatGPT's responses were too <u>lengthy and broad.</u>

- Want <u>data sources</u> in its responses

👎 **Hard to Review**

- Want to have features such as <u>a search or note function</u> in each conversation for easy review and information retrieval.

- They want to <u>pin</u> the conversation.

👎 **Single format**

Half of the users expressed dissatisfaction with ChatGPT's <u>one-question-get-one-answer</u> format.

Action Insights

01 — Naturalness Adjustment

1. Switch Tones

Add a feature to switch between formal and casual tones as needed.

2. Context Understanding Model

Enhance the ability of context understanding model, to auto-adjust output naturalness for various scenarios.

02 — Customize Responses

Surveys & Questions

Gain insights into users' life values and preferences.

Enhance the emotional esupport by interacting based on individual characteristics and needs.

03 — Precise Answers

1. Follow-up Question

When a user's question lacks specificity, the system should ask further questions to narrow down the scope of the inquiry.

2. Preset Prompts

Offer preset prompts for quick access to needed information.

Measuring Empathy

Most users are concerned with ChatGPT's empathic capabilities, so I've evaluated its empathy

Five Questionnaires

Interpersonal Reactivity Index	Fantasy: 18 Perspective Taking：17 Empathic Concern: 11 Personal Distress: 9
Toronto Empathy Questionnaire	**18** / 50
Empathy Quotient	**37** / 100
Autism Spectrum Quotient	**43** / 64
Perth Empathy Scale	**39** /100

The Empathic Assessment Methods

Our empathy exceeds ChatGPT 3.5's, but remains below average human levels

How might we show emotional understanding and caring for others?

Conclusion & Limitation

Conclusion

1. **Improving Response Naturalness**

It's vital for being easy to use, boosting work efficiency, and accelerating the work completion rate

2. **Enhancing Emotional Empathy**

By adjusting emotional expression according to context and user preferences, ChatGPT can better meet diverse user emotional needs, advancing toward a multi-faceted AI

Limitations

1. **Small Sample Size**

ChatGPT has a broad user base, but the sample size of this study is limited, and the interviewees all come from digital media-related industries, which may result in **a lack of generalizability**

2. **Ethical Concerns**

When participants encounter emotional or daily life questions during an interview or data collection process, some may feel **Uncomfortable or Guarded**

Next Steps in Research

Under the premise of increasing sample size, and based on existing research findings

- **Research whether enhancing emotional recognition ability is directly related to satisfaction with emotional issues**

- **Explore the different tones to understand which tone is more suitable for specific scenarios or user groups**

- **Understand users' acceptance of value and preference questionnaires**

Reference

ChatGPT vs. Gemini: Which AI Chatbot Subscription Is Right for You?

ChatGPT vs. Microsoft Copilot: What's the difference?

Exploring ChatGPT's Empathic Abilities

NEW Google Gemini Is Here! Is It Better Than ChatGPT?

Thank you for watching.

Menglin Zhao
Northeastern University
zhao.mengl@northeastern.edu
617-7636-285