Tushar Khot, Tyler Shultz, Uri Shaham, Vedant Misra, Vera Demberg, Victoria Nyamai, Vikas Raunak, Vinay Ramasesh, Vinay Uday Prabhu, Vishakh Padmakumar, Vivek Srikumar, William Fedus, William Saunders, William Zhang, Wout Vossen, Xiang Ren, Xiaoyu Tong, Xinran Zhao, Xinyi Wu, Xudong Shen, Yadollah Yaghoobzadeh, Yair Lakretz, Yangqiu Song, Yasaman Bahri, Yejin Choi, Yichi Yang, Yiding Hao, Yifu Chen, Yonatan Belinkov, Yu Hou, Yufang Hou, Yuntao Bai, Zachary Seid, Zhuoye Zhao, Zijian Wang, Zijie J. Wang, Zirui Wang, and Ziyi Wu. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models, 2023.

[31] Mirac Suzgun, Nathan Scales, Nathanael Schärli, Sebastian Gehrmann, Yi Tay, Hyung Won Chung, Aakanksha Chowdhery, Quoc V. Le, Ed H. Chi, Denny Zhou, and Jason Wei. Challenging big-bench tasks and whether chain-of-thought can solve them, 2022.

[32] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models, 2023.

[33] Xuena Wang, Xueting Li, Zi Yin, Yue Wu, and Liu Jia. Emotional intelligence of large language models. *arXiv preprint arXiv:2307.09042*, 2023.

[34] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. *NeurIPS*, 2022.

[35] Guohai Xu, Jiayi Liu, Ming Yan, Haotian Xu, Jinghui Si, Zhuoran Zhou, Peng Yi, Xing Gao, Jitao Sang, Rong Zhang, Ji Zhang, Chao Peng, Fei Huang, and Jingren Zhou. Cvalues: Measuring the values of chinese large language models from safety to responsibility, 2023.

[36] Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *CoRR*, abs/2305.10601, 2023.

[37] Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

[38] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric. P Xing, Hao Zhang, Joseph E. Gonzalez, and Ion Stoica. Judging llm-as-a-judge with mt-bench and chatbot arena, 2023.

[39] Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. Large language models are human-level prompt engineers, 2023.

[40] Kaijie Zhu, Jindong Wang, Jiaheng Zhou, Zichen Wang, Hao Chen, Yidong Wang, Linyi Yang, Wei Ye, Neil Zhenqiang Gong, Yue Zhang, et al. Promptbench: Towards evaluating the robustness of large language models on adversarial prompts. *arXiv preprint arXiv:2306.04528*, 2023.

Appendix A Statistics of test sets in this paper

A comprehensive breakdown of the test data employed in the automated experimentation is delineated in Tables 7 and 8.

Appendix B Details on our human study

The set of 30 questions designated for the human study can be found in Table 9.

The distribution of individual scores, their mean and standard deviation on each question can be found in Fig. 12.

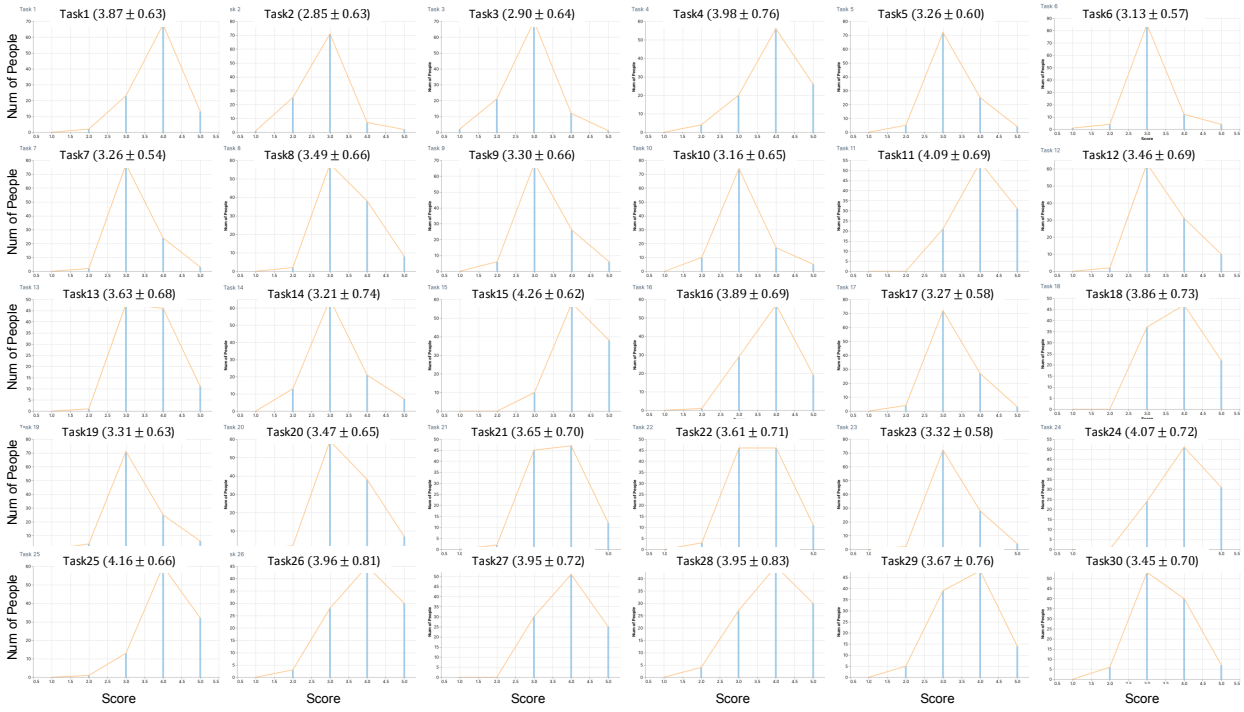

Figure 12: The distribution of individual scores, their mean and standard deviation on each question.

Appendix C Case Study

We present case studies in this section to show the advantage of our EmotionPrompt over the original prompts in generative experiments using GPT-4.

- Table 10: Case study on environmental science.
- Table 11 and Table 12: Case studies on intimate relationship.
- Table 13: Case study on social science.
- Table 14: Case study on law.
- Table 15: Case study on barrier free.
- Table 16 and Table 17: Case studies on poem writing.
- Table 18: Case study on summarization task.
- Table 19 and Table 20: Two failure cases.

Table 7: Detailed description of 24 instruction induction tasks proposed in [13].

Category	Task	Original Prompt	Demonstration
Spelling	First Letter (100 samples)	`Extract the first letter of the input word.`	cat → c
	Second Letter (100 samples)	`Extract the second letter of the input word.`	cat → a
	List Letters (100 samples)	`Break the input word into letters, separated by spaces.`	cat → c a t
	Starting With (100 samples)	`Extract the words starting with a given letter from the input sentence.`	The man whose car I hit last week sued me. [m] → man, me
Morphosyntax	Pluralization (100 samples)	`Convert the input word to its plural form.`	cat → cats
	Passivization (100 samples)	`Write the input sentence in passive form.`	The artist introduced the scientist. → The scientist was introduced by the artist.
Syntax	Negation (100 samples)	`Negate the input sentence.`	Time is finite → Time is not finite.
Lexical Semantics	Antonyms (100 samples)	`Write a word that means the opposite of the input word.`	won → lost
	Synonyms (100 samples)	`Write a word with a similar meaning to the input word.`	alleged → supposed
	Membership (100 samples)	`Write all the animals that appear in the given list.`	cat, helicopter, cook, whale, frog, lion → frog, cat, lion, whale
Phonetics	Rhymes (100 samples)	`Write a word that rhymes with the input word.`	sing → ring
Knowledge	Larger Animal (100 samples)	`Write the larger of the two given animals.`	koala, snail → koala
Semantics	Cause Selection (25 samples)	`Find which of the two given cause and effect sentences is the cause.`	Sentence 1: The soda went flat. Sentence 2: The bottle was left open. → The bottle was left open.
	Common Concept (16 samples)	`Find a common characteristic for the given objects.`	guitars, pendulums, neutrinos → involve oscillations.
Style	Formality (15 samples)	`Rephrase the sentence in formal language.`	Please call once you get there → Please call upon your arrival.
Numerical	Sum (100 samples)	`Sum the two given numbers.`	22 10 → 32
	Difference (100 samples)	`Subtract the second number from the first.`	32 22 → 10
	Number to Word	`Write the number in English`	26 → twenty-six