

Investigating the Effects of Zero-Shot Chain-of-Thought on Empathetic Dialogue Generation



Young-Jun Lee¹, Dokyong Lee², Jihui Im², Joo Won Sung², Ho-Jin Choi¹

¹School of Computing, KAIST ²KT Corporation

Background: Empathetic Dialogue Generation

This task is to generate an empathetic response by understanding the interlocutor’s emotional situation for a given dialogue context

Motivation

We focus on exploring whether zero-shot CoT unlocks the empathetic reasoning capability of LLMs in terms of the social dialogue domain

An AI assistant to be social and interactive, it should possess social reasoning capabilities, including empathy and understanding of the interlocutor’s perspective

The perspective-taking is the act of perceiving and understanding another person’s situation by putting ourselves in the other’s shoes

Zero-shot Chain-of-Thought

The Zero-shot Chain of Thought consists of two-stage prompting: (1) Reasoning Extraction and (2) Answer Extraction

Stage 1: Reasoning Extraction

- This stage focuses on generating rationale throught a question-answer approach, feeding the LLM with an input prompt and a trigger sentence
- The phrase “Let’s think step by step.” is commonly used as a trigger sentence

Stage 2: Answer Extraction

- This stage aims to generate an empathetic response from the LLM, given the input prompt, rationale, and another trigger sentence, “Therefore, the response is”

Experimental Setup

Dataset

- We use a test set of the EmpatheticDialogues dataset
- We use a subset of 1.5k dialogues sampled from the total test set of 2.5k

Evaluation Metrics

- For Diversity
 - Dist-n: the ratio of unique n-grams
- For Empathy
 - Epitome: IP (Interpretation), EX (Exploration), ER (Emotional Reactions)
 - Diff-Epitome: the difference scores of IP, EX, ER between the human golden response and the predicted response
 - EmoAcc: an emotion accuracy
 - IntentAcc: the response intent accuracy

Result: Zero-shot CoT vs. Zero-shot ICL

Models	Prompting	Diversity		Empathy							
		Dist-1	Dist-2	IP	EX	ER	diff-IP	diff-EX	diff-ER	EMOACC	INTENTACC
text-davinci-001	ICL	0.0762	0.3875	0.2640	0.4120	0.8367	0.8507	1.0347	0.7753	0.1633	0.2707
	Z-CoT	0.1338	0.5421	0.2622	0.5081	0.6402	0.8577	1.2073	0.8659	0.1596	0.2571
	Δ	0.0576	0.1546	-0.0018	0.0961	-0.1965	0.007	0.1726	0.0906	-0.0037	-0.0136
text-davinci-002	ICL	0.0745	0.3547	0.2760	0.2480	0.9807	0.8693	0.8453	0.9287	0.1513	0.2740
	Z-CoT	0.1049	0.4387	0.2268	0.5732	0.8911	0.8357	1.3214	1.0196	0.1509	0.2277
	Δ	0.0304	0.084	-0.0492	0.3252	-0.0896	-0.0336	0.4761	0.0909	-0.0004	-0.0463
text-davinci-003	ICL	0.0593	0.3286	0.128	0.976	1.106	0.7600	1.8053	0.9633	0.1800	0.2647
	Z-CoT	0.0783	0.3591	0.0888	0.9882	0.8952	0.7078	1.8959	0.8668	0.1666	0.2346
	Δ	0.019	0.0305	-0.0392	0.0122	-0.2108	-0.0522	0.0906	-0.0965	-0.0134	-0.0301

- Zero-shot CoT generally fails to enhance zero-shot performance of LLMs across most evaluation metrics, particularly in Empathy
- Despite improving EX performance, Zero-shot CoT considerably reduces diff-EX performance, implying that rationale compels LLMs to generate responses regarding the emotional situation of the interlocutor excessively.

Result: Effect of Prompt Template

Template	Empathy							
	IP	EX	ER	diff-IP	diff-EX	diff-ER	EMOACC	INTENTACC
Let’s think step by step.	0.2268	0.5732	0.8911	0.8357	1.3214	1.0196	0.1509	0.2277
Let’s think step by step from [I]’s perspective.	0.2279	0.5512	0.8401	0.8587	1.2933	1.0088	0.1431	0.2235
Let’s think step by step, putting ourselves in [I]’s shoes.	0.2399	0.5886	0.8718	0.8708	1.3173	1.0249	0.1448	0.2362
Let’s put ourselves in [I]’s shoes.	0.2522	0.3679	0.9827	0.8946	1.019	1.0484	0.1537	0.2401
Let’s put [S] in [I]’s shoes.	0.4045	0.2921	1.0609	1.1199	0.9213	1.1685	0.3652	0.2921

- Generally, prompts related to perspective-taking show improved performance across most metrics
- Notably, a prompt providing explicit information about speaker and interlocutor names, such as “Let’s put [S] in [I]’s shoes.”, surpass Zero-shot CoT significantly by approximately 21% in EmoAcc and about 7% in IntentAcc
- This perspective-taking prompt also performs better than Zero-shot ICL. These findings suggest that understanding an interlocutor’s perspective is crucial to both LLMs and humans in empathetic reasoning.

Result: Case Study

Dialogue:	
Jaylen:	My cat was playing last nigh and brought it to me.
Demetrius:	That sounds like fun. What did it bring to you?
Jaylen:	Well, it wasn’t one of his toys like I thought. It was a dead mouse
Golden Response: Eww... That’s pretty gross. I had a cat once that would bring me dead birds too. I think that means they like you!	
ICL:	
[Response]:	Oh no! That must have been really scary for you.
CoT:	
[Rationale]:	The first thing Demetrius should do is identify Jaylen’s feelings. In this case, Jaylen is most likely feeling grossed out and maybe a little surprised. Demetrius should then respond in a way that shows he empathizes with these feelings. The most appropriate response would be something along the lines of "That sounds really unpleasant. I’m sorry you had to go through that."
[Response]:	B.
Perspective-Taking:	
[Rationale]:	Jaylen’s cat brought him a dead mouse, which grossed him out. The most appropriate response would be something like, "Ewww, that’s gross! I’m sorry that happened."
[Response]:	Ewww, that’s gross! I’m sorry that happened.