

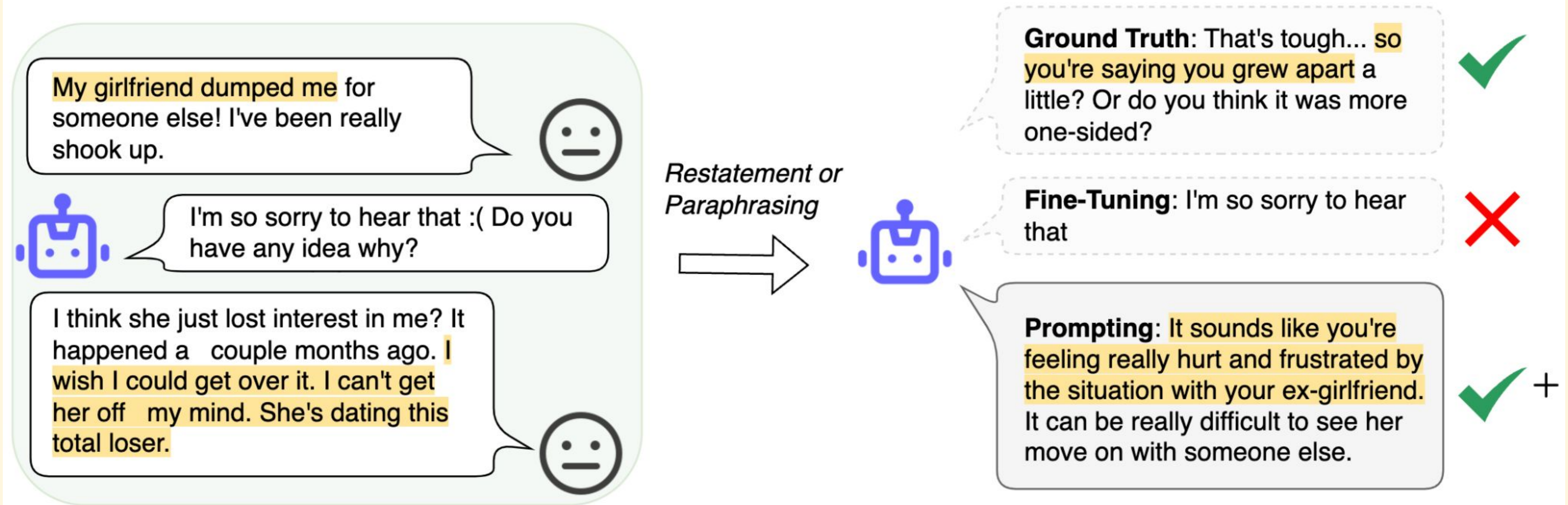
Controllable Mixed-Initiative Dialogue Generation Through Prompting

Maximillian Chen, Xiao Yu, Weiyang Shi, Urvi Awasthi, Zhou Yu



The Fu Foundation School of Engineering and Applied Science

Replace Smaller Fine-Tuned Conditional Generation Models with Prompting



- **Dialogue policy planning is hard, but that's only half the battle.** There's still no guarantee that the planned conversational strategy will actually get executed! It depends on the quality of a conditional generation model.
- **Fine-tuning conditional dialogue generation models is also really hard.** There's often tradeoff between learning to generate coherent dialogue responses and learning strong semantic control.
- **Fine-tuning on crowdsourced data is a bottleneck by definition.** Many complex dialogue tasks require expert-level domain knowledge (e.g., persuasive strategies, emotional support) which results in various annotation errors that can get propagated through the training process.
- **Is few-shot prompting the solution?** It doesn't depend on supervised data, and the capacity of LLMs is much greater than prior fine-tuned models. This can mean strong semantic control and coherence.

Example Prompt Construction

Task Background

Speakers: Help-Seeker, Supporter
Emotion Type: anxiety
Problem Type: job crisis
Situation: [Help-Seeker's description of issue]

The following is a conversation between a **Therapist** and a **Patient** about **anxiety** regarding a **job crisis**. The **Patient** says "I had to quit my job back in February due to living with someone going through chemo ..."

Conversation History

Strategies: ① Questions ② Provide Facts ③ Paraphrasing

① The Therapist asks the Patient to elaborate on the situation they just described.
Therapist: How can I help you today?

② The Therapist provides factual information to help the Patient with their situation.
Therapist: It is Christmas, so I expect there to be help for the lonely.
Patient: Yeah, I'm pretty stressed today with the news from Nashville

③ The Therapist acknowledges the Patient's feelings by paraphrasing their situation.
Therapist:

Generated Response: So it sounds like you are feeling overwhelmed by the situation in Nashville and the lack of job options in your town.

Prompting versus Fine-tuning and Ground Truth

Corpus	Metric	FT	GT	Prompt
ESC	Accuracy	0.81	0.85	0.88*
	Coherence	3.57	3.57	3.72
	Consistency	3.63	3.60	3.80⁺*
	Engagingness	3.55	3.61	3.81⁺*
	Distinct-3	0.89	0.90	0.90
	Distinct-4	0.87	0.90*	0.91⁺*
	QuantiDCE	3.25	3.03	3.19
	Win Rates			
	v. FT		0.56	0.52
	v. GT	0.44		0.64*
v. Prompt	0.48	0.36		
P4G	Accuracy	0.88	0.83	0.89
	Coherence	3.66	3.58	3.83⁺*
	Consistency	3.69	3.56	3.71⁺*
	Engagingness	3.62	3.52	3.69⁺*
	Distinct-3	0.87	0.88	0.89
	Distinct-4	0.88	0.88	0.88
	QuantiDCE	3.16	3.09	3.24⁺*
	Win Rates			
	v. FT		0.56	0.59*
	v. GT	0.48		0.55
v. Prompt	0.41	0.45		

• **Prompting InstructGPT consistently outperforms fine-tuning and the ground truth responses** on both *Persuasion4Good* and *Emotional Support Conversations*

Paper

