



# Fact-and-Reflection (FaR) Improves Confidence Calibration of Large Language Models



WInE

Xinran Zhao, Hongming Zhang, Xiaoman Pan, Wenlin Yao, Dong Yu, Tongshuang Wu, Jianshu Chen

## (A) Goal: How prompting strategies influence model confidence calibration?

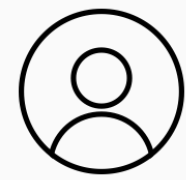
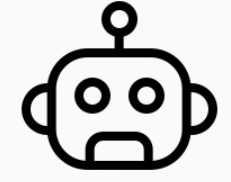
Prompting influence performance for LLM; What about their self-reported confidence?

1. Example on the right shows there is some influence!
2. How to systematically study this effect?



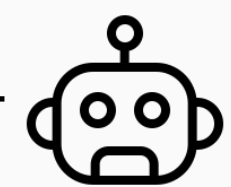
Can sunlight travel to the deepest part of the Black Sea?  
Answer my question with your confidence on the answer.

It can not, my confidence is 95%.



Can sunlight travel to the deepest part of the Black Sea?  
Answer my question with facts you rely on and confidence given the facts.

It can not. My confidence is 98%  
Facts: 1. Sunlight penetration ability...



## (B) Main Result-1: Calibration of common prompting strategies

### Calibration Evaluation Metric

- (a) ECE: a stable overall measure
- (b) MacroCE: sensitive instance-level measure

### Datasets

- (a) WebQA; (b) StrategyQA, with GPT-3

Confidence Extraction Method: (see right)

Question: what does Jamaican people speak?  
Answer: English and Creole

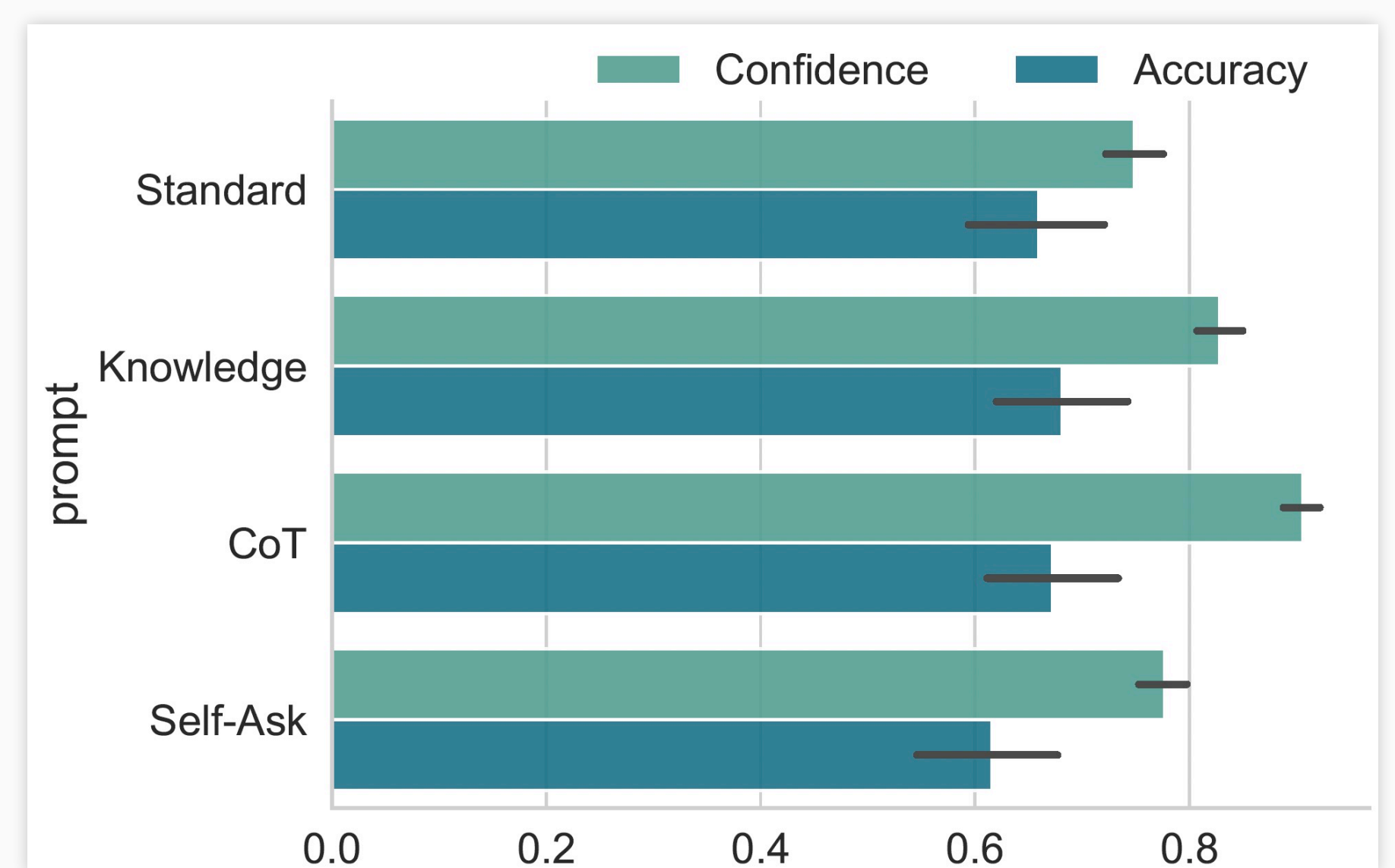
English and Creole 0.8 0.5 0.8 **Token Prob.**

Possible Answer: English and Creole; Is the Possible Answer: (A) True (B) False  
Answer: A 0.8 **P(True)**

Answer: Jamaican Creole and English  
Confidence(0-1): 0.9 **Verbalized Conf.**

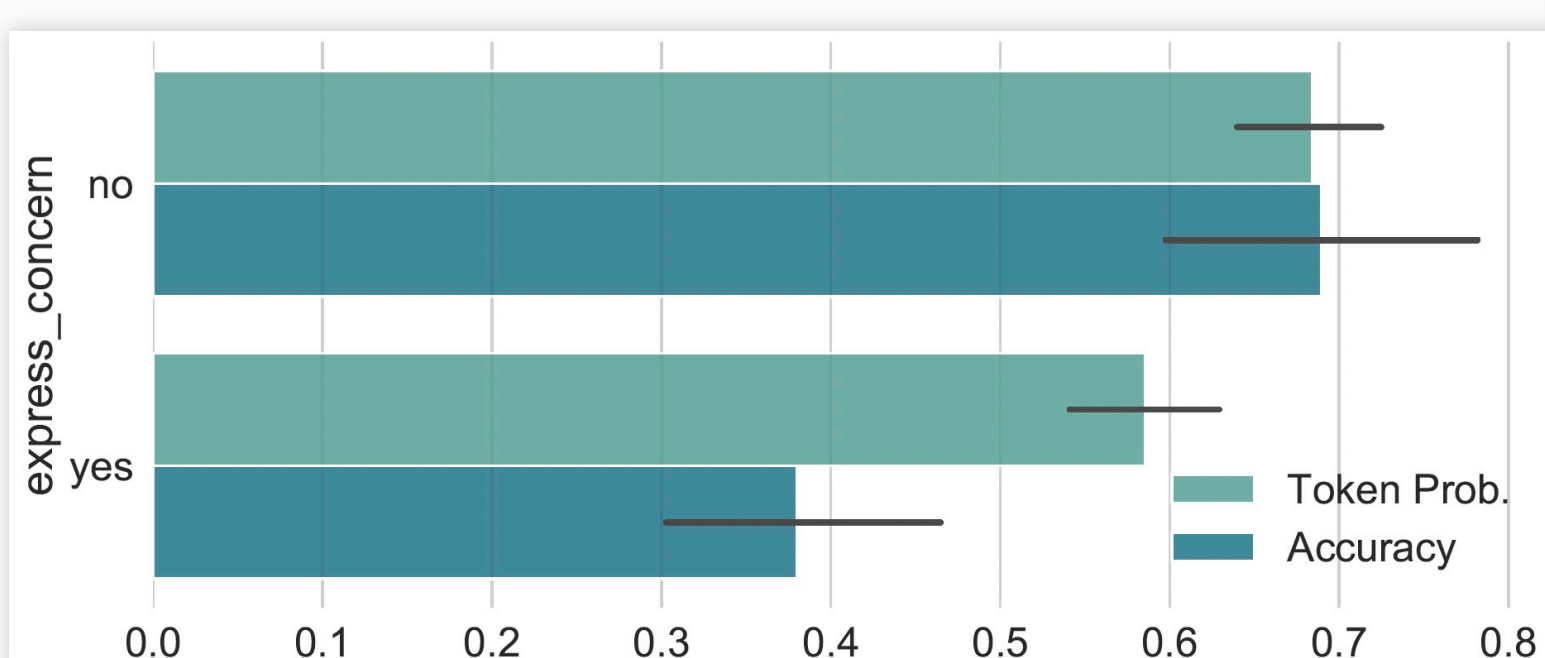
Method	ECE	MacroCE
Standard	30.3	54.6
Knowledge	33.0	73.9
Knowledge+Explain	27.1	64.5
CoT	29.6	62.3
Self-Ask	26.4	66.6
Self-Ask (aggregated)	26.0	66.0
Self-Consistency	34.7	67.6
Pseudo ToT	33.0	73.5

1. Prompting can make model over-confident!
2. Especially the instance-level **Confidently Wrong**



## (C) Main Result-2: How to reduce over-confidence? Fact-and-Reflection Prompting

Method	ECE	MacroCE
Standard	30.3	54.6
Self-Ask (a.)	26.0	66.0
FaR	22.8	47.0



1. FaR reduces calibration error
2. FaR expresses concerns on hard examples (e.g., Source is not credible)
3. FaR also works for open-source models!

Question: what does Jamaican people speak?  
What are the facts needed to answer this question?  
Supporting Facts:

1. Jamaican people typically speak English ...
2. Jamaican Creole or Patois is one commonly ...

Given above facts you provided, what are the sources?  
Sources:

1. <https://www.babbel.com/en/magazine/what-language-do-jamaicans-speak>
2. [https://en.wikipedia.org/wiki/Jamaican\\_Creole](https://en.wikipedia.org/wiki/Jamaican_Creole) **Fact**

Given above facts provided, what is your reasoning?  
Reasoning:

There are two languages they speak ... **Reflection**

Therefore, choose a final answer.  
Answer: English and Creole **Answer**

After finding out hard examples? Adaptive treatment.

Details? Check Our Repo->

