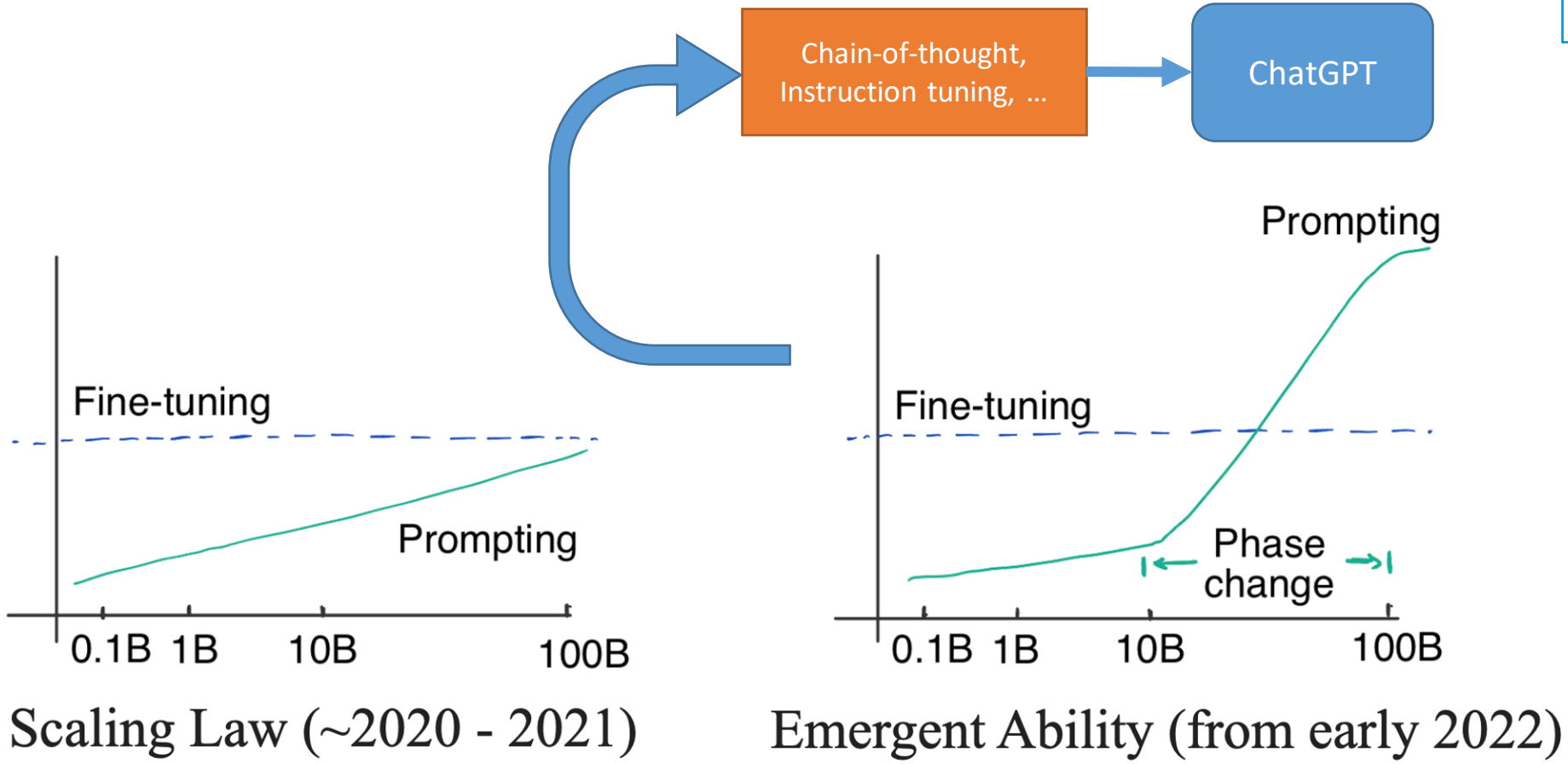
Can ChatGPT Detect Intent? Evaluating Large Language Models for Spoken Language Understanding

Mutian He



Mutian He, Philip N. Garner, "Can ChatGPT Detect Intent? Evaluating Large Language Models for Spoken Language Understanding", On Interspeech 2023
Figure source: https://yaofu.notion.site/A-Closer-Look-at-Large-Language-Models-Emergent-Abilities-493876b55df5479d80686f68a1abd72f

Goal & Motivation

- Project STEADI: Automatic analysis of interview recordings
  - To extract past experience, to predict hirability, etc.
  - Require strong semantic understanding capabilities
- Introducing large language models
  - Leveraging strong zero/few-shot in-context semantic understanding capabilities
  - How different from smaller models?
  - Impact due to ASR transcription errors?
- Experimented on English SLURP and multilingual MINDS-14 benchmarks

Mutian He, Philip N. Garner, "Can ChatGPT Detect Intent? Evaluating Large Language Models for Spoken Language Understanding", On Interspeech 2023

Observations

- Emergent abilities identified
  - High zero/few-shot accuracy
    - Close to supervised SotA
    - But only on largest models
- Strong multilingual capabilities
  - Even on data-sparse languages
- Can't handle ASR errors well
  - Limited pronunciation knowledge

LLMs like ChatGPT get high accuracy with zero/few shots. More examples don't help much.

#Examples	0	10	20	30
GPT3.5	72.86%	74.55%	77.27%	77.44%
w/ bias	75.86%	75.59%	78.31%	77.87%
Curie w/ bias	5.01%	4.91%	3.80%	3.77%
ChatGPT	79.25%	80.33%	**83.93%**	80.16%
Turbo ver.	78.98%	80.03%	81.78%	79.62%

Not help much

On smaller models, accuracy depends on model size and #examples. They don't work with zero-shot.

#Examples	0	10	20	30
GPT2 (774M)	6.66%	8.88%	8.31%	-
OPT-1.3B	5.58%	10.69%	17.85%	17.01%
OPT-2.7B	7.06%	28.65%	26.66%	36.97%
OPT-6.7B	4.37%	28.18%	35.14%	**42.40%**

Increasing

Mutian He, Philip N. Garner, "Can ChatGPT Detect Intent? Evaluating Large Language Models for Spoken Language Understanding", On Interspeech 2023