

Tuesday • November 2, 2021

# Modern Language Models



Yale

LING 380/780

*Neural Network Models of Linguistic Structure*

# In-Context Adaptation

Transfer Learning Method #2

# Generative Pre-Trained Transformer

- **GPT-n (for n=1, 2, 3)** are Transformer language models trained on large corpora.
- GPT-1 Data: 1B of text from Books corpus
- GPT-1 Parameters: 115 million
- GPT-2 Data: 40 GB of text from links in Reddit posts with at least 3 karma
- GPT-2 Parameters: 1.5 billion
- GPT-3 Data: > 570 GB of text, including web scrapes, books, and Wikipedia
- GPT-3 Parameters: 175 billion

# GPT-2 Text Predictions

**Prompt:** *A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.*

**Prediction:** The incident occurred on the downtown train line, which runs from Covington and Ashland stations.

In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.

...

# GPT-2 Text Predictions

**Prompt:** *Legolas and Gimli advanced on the orcs, raising their weapons with a harrowing war cry.*

**Prediction:** The orcs' response was a deafening onslaught of claws, claws, and claws; even Elrond was forced to retreat. "You are in good hands, dwarf," said Gimli, who had been among the first to charge at the orcs; it took only two words before their opponents were reduced to a blood-soaked quagmire, and the dwarf took his first kill of the night.

...

# GPT-2 Text Predictions

**Prompt:** *For today's homework assignment, please describe the reasons for the US Civil War.*

**Prediction:** By Donny Ferguson

It is easy to identify why the Civil War happened, because so many people and so many books and so much television and films tell us that it was the cause, that it has something to do with race or economics or religion. There is much agreement that it was essentially a war of slavery on behalf of capitalism, about a century of slavery. ...

# GPT-2 Text Predictions

**Prompt:** *Recycling is good for the world.*

*NO! YOU COULD NOT BE MORE WRONG!!*

**Prediction:** Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. ...

# GPT-3 Text Predictions

## *When A.I. Falls in Love*

The Times asked GPT-3 to tell us a little about itself and its romantic life.

**Prompt:** *The following is a New York Times's Modern Love column about a woman named Frances, her husband Dean, and the unlikely story of how they first met:*

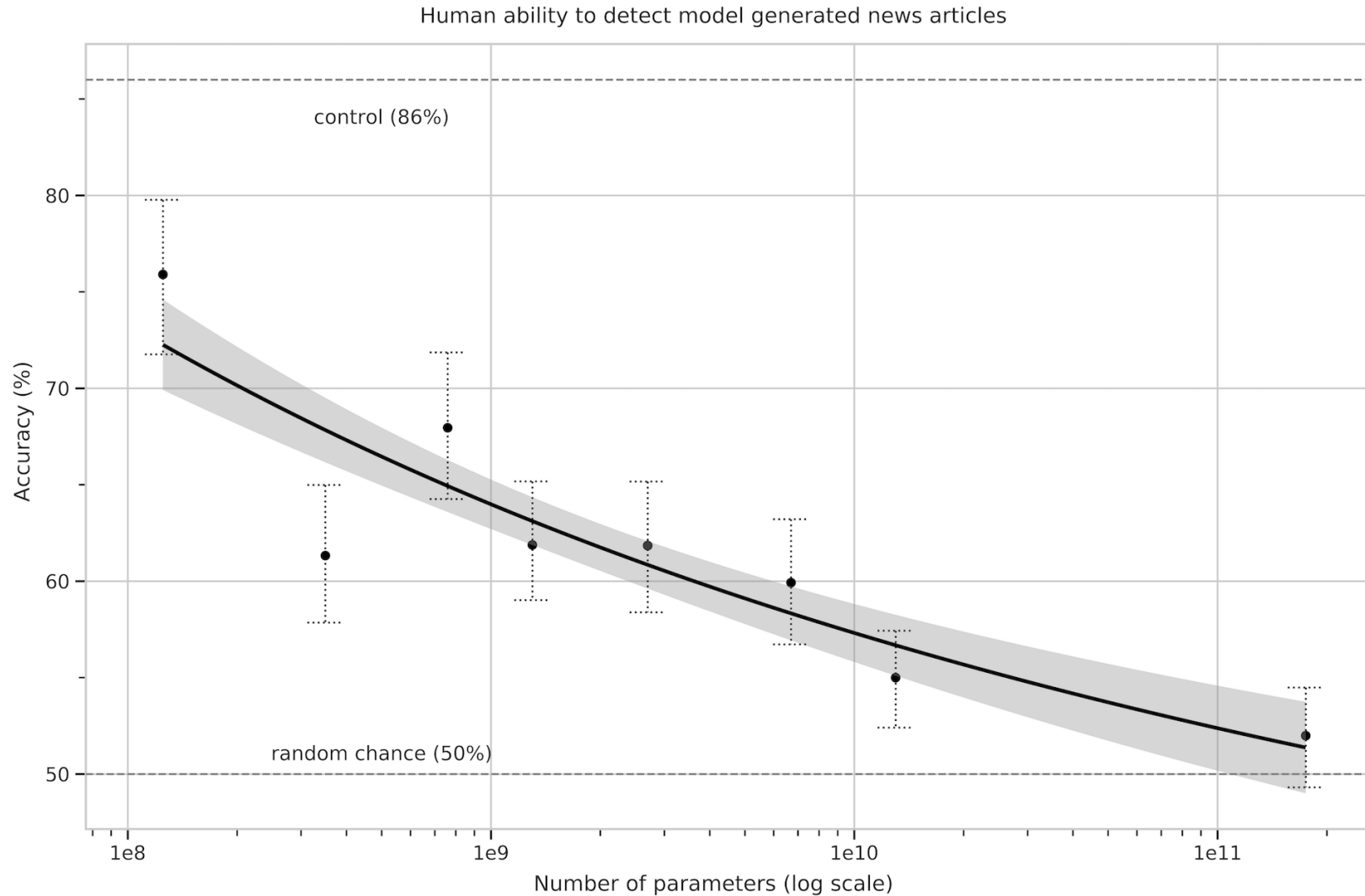
**Prediction:** After my fiancé died, my mother told me to “get out there again.” She wanted me to go to a singles bar. I told her I’d rather go to the dentist.

“Just once,” she said. “Just to see what it’s like.”

One day, early last year, I found myself driving to a singles bar in winter snow. I sat in my car for 15 minutes, then drove away. The next day, I went back and sat in my car for another 15 minutes. I did this for a couple of weeks, until I finally mustered up the nerve to walk in.



# Text Prediction and Model Size



# Summarization with GPT-2

**Prompt:** *Amina Ali Qassim is sitting with her youngest grandchild on her lap, wiping away tears with her headscarf. Only a few months old, this is the baby girl whose ears she desperately tried to cover the night the aerial bombardment started. She lay awake, she says, in a village mosque on the Yemeni island of Birim, counting explosions as the baby cried. ...* **TL;DR:**

**Prediction:** Yemen is in the middle of a civil war. Saudi Arabia is leading the coalition bombing campaign. It's been bombing Yemen for more than two months now.

# Causal Language Modeling



WHEN YOU TRAIN PREDICTIVE MODELS  
ON INPUT FROM YOUR USERS, IT CAN  
LEAK INFORMATION IN UNEXPECTED WAYS.

# Causal Language Modeling

- Text corpora contain a lot of information.
- **Idea:** Extract information from language model predictions.

The capital of France is|

the first-largest city in

Paris.

the world's largest producer of

# In-context learning (Brown et al. 2020)

- Fine tuning without weight updates:
  - Train Transformer language model on **lots** of data
  - Provide fine tuning data in transformer's input
  - Test for use of this context

# Few-Shot Translation

Few shot:

1	Translate English to French:	← task description
2	sea otter => loutre de mer	← examples
3	peppermint => menthe poivrée	←
4	plush girafe => girafe peluche	←
5	cheese => .....	← prompt

One shot:

1	Translate English to French:	← task description
2	sea otter => loutre de mer	← example
3	cheese => .....	← prompt

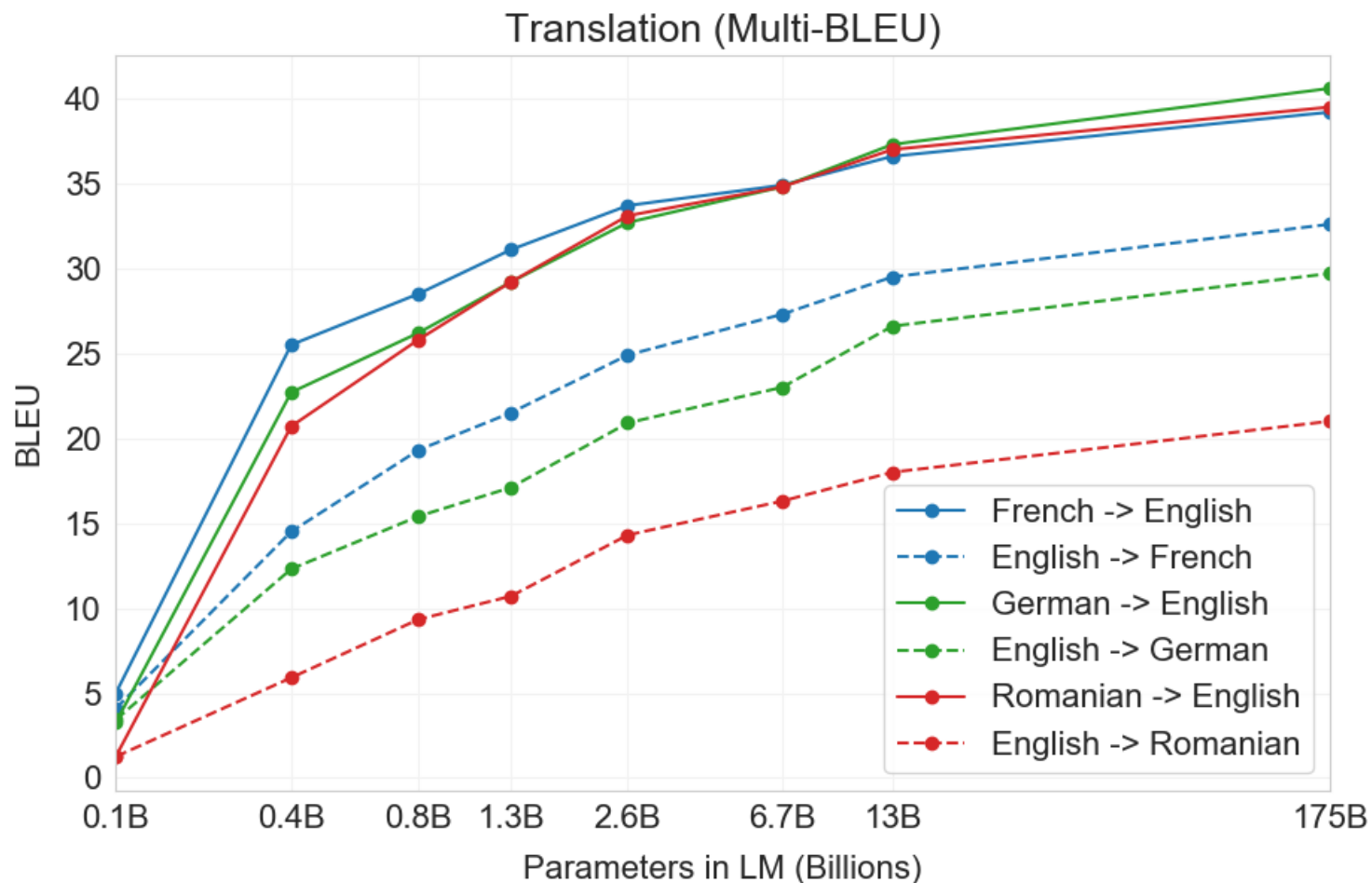
Zero shot:

1	Translate English to French:	← task description
2	cheese => .....	← prompt

# Few-Shot Translation

Setting	En→Fr	Fr→En	En→De	De→En	En→Ro	Ro→En
SOTA (Supervised)	<b>45.6<sup>a</sup></b>	35.0 <sup>b</sup>	<b>41.2<sup>c</sup></b>	40.2 <sup>d</sup>	<b>38.5<sup>e</sup></b>	<b>39.9<sup>e</sup></b>
XLM [LC19]	33.4	33.3	26.4	34.3	33.3	31.8
MASS [STQ <sup>+</sup> 19]	<u>37.5</u>	34.9	28.3	35.2	<u>35.2</u>	33.1
mBART [LGG <sup>+</sup> 20]	-	-	<u>29.8</u>	34.0	35.0	30.5
GPT-3 Zero-Shot	25.2	21.2	24.6	27.2	14.1	19.9
GPT-3 One-Shot	28.3	33.7	26.2	30.4	20.6	38.6
GPT-3 Few-Shot	32.6	<u>39.2</u>	29.7	<u>40.6</u>	21.0	<u>39.5</u>

# Few-Shot Translation





# Few-Shot Question Answering

**Prompt:** *Helsinki is the capital and largest city of Finland. It is in the region of Uusimaa, in southern Finland, on the shore of the Gulf of Finland. ...*

*Q: what is the most populous municipality in Finland?*

*A: Helsinki*

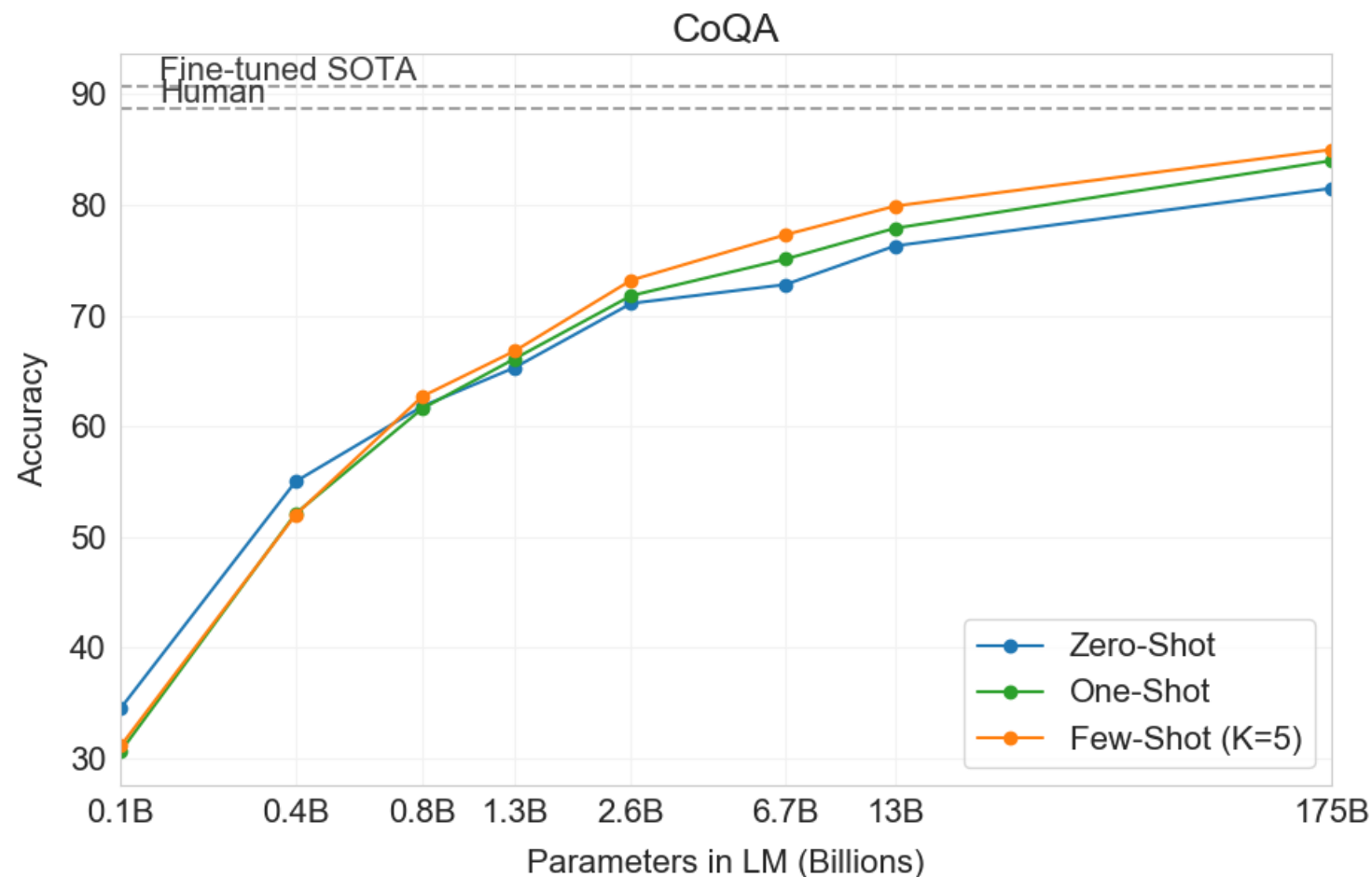
...

*Q: what towns are a part of the metropolitan area?*

*A:*

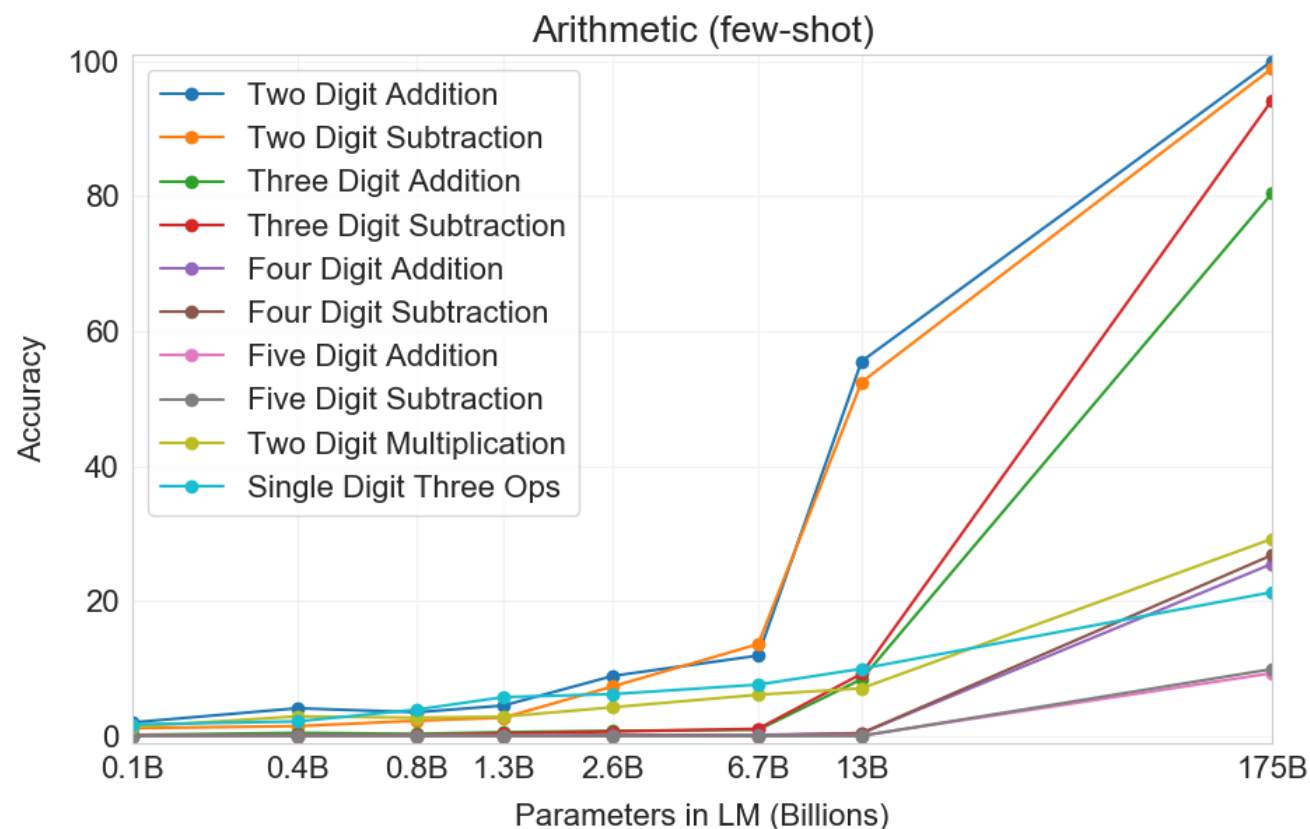
**Prediction:** Helsinki, Espoo, Vantaa, Kauniainen, and surrounding commuter towns

# Few-Shot Question Answering



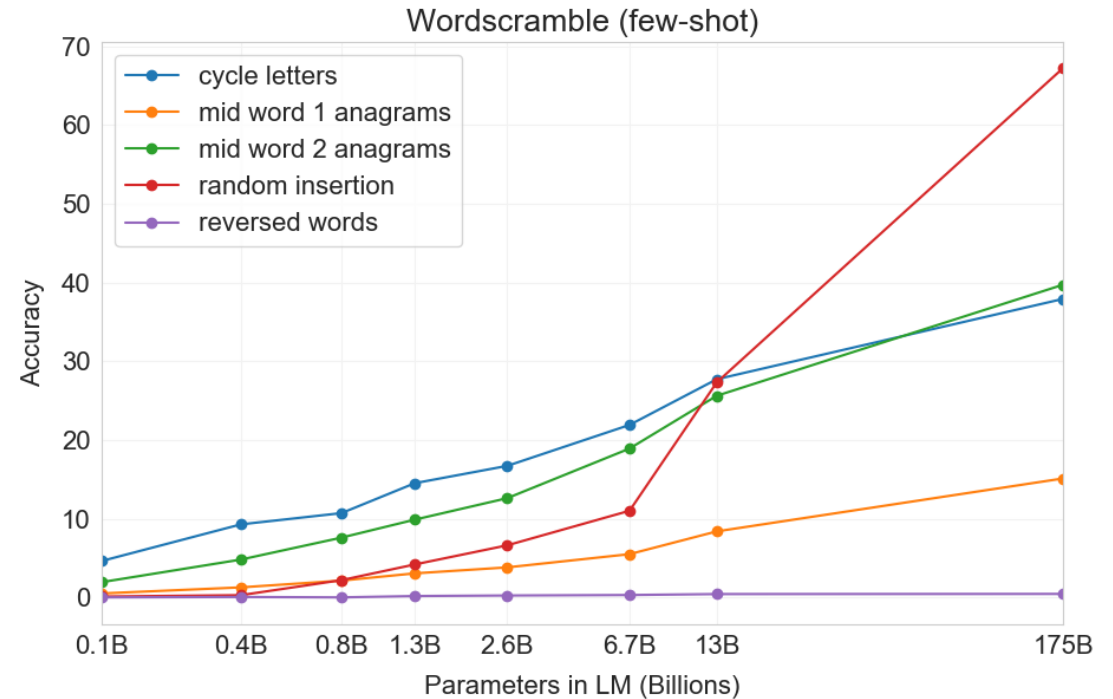
# Few Shot Synthetic Tasks

- Arithmetic: “Q: What is 48 plus 76? A: 124”

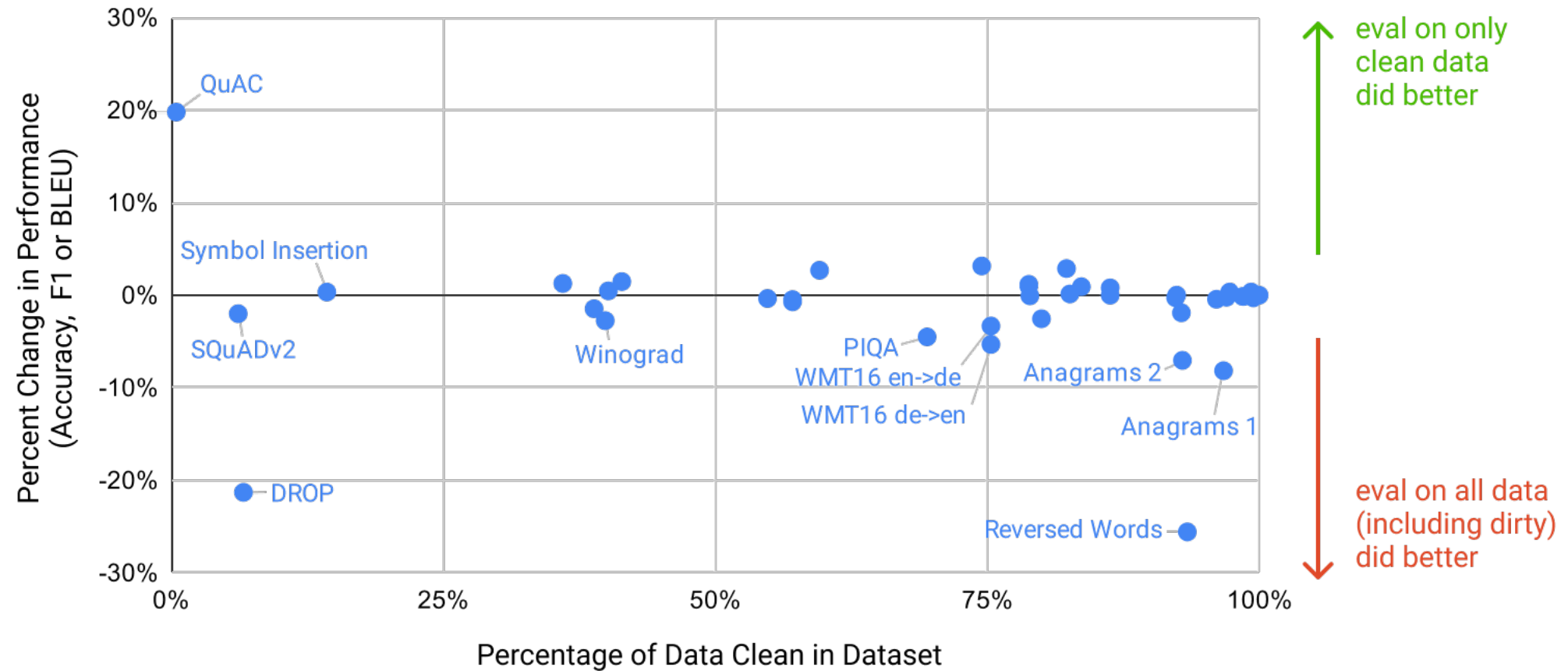


# Few Shot Synthetic Tasks

- Word Manipulation:
  - Cycle Letters
  - A1: Anagrams of all but first and last letters
  - A2: Anagrams of all but first and last two letters
  - Random Insertion
  - Reversed Words



# Contamination Analysis



# Few Shot Synthetic Tasks

- Word Manipulation:
  - Cycle Letters
  - A1: Anagrams of all but first and last letters
  - A2: Anagrams of all but first and last two letters
  - Random Insertion
  - Reversed Words

Setting	CL	A1	A2	RI	RW
GPT-3 Zero-shot	3.66	2.28	8.91	8.26	0.09
GPT-3 One-shot	21.7	8.62	25.9	45.4	0.48
GPT-3 Few-shot	37.9	15.1	39.7	67.2	0.44