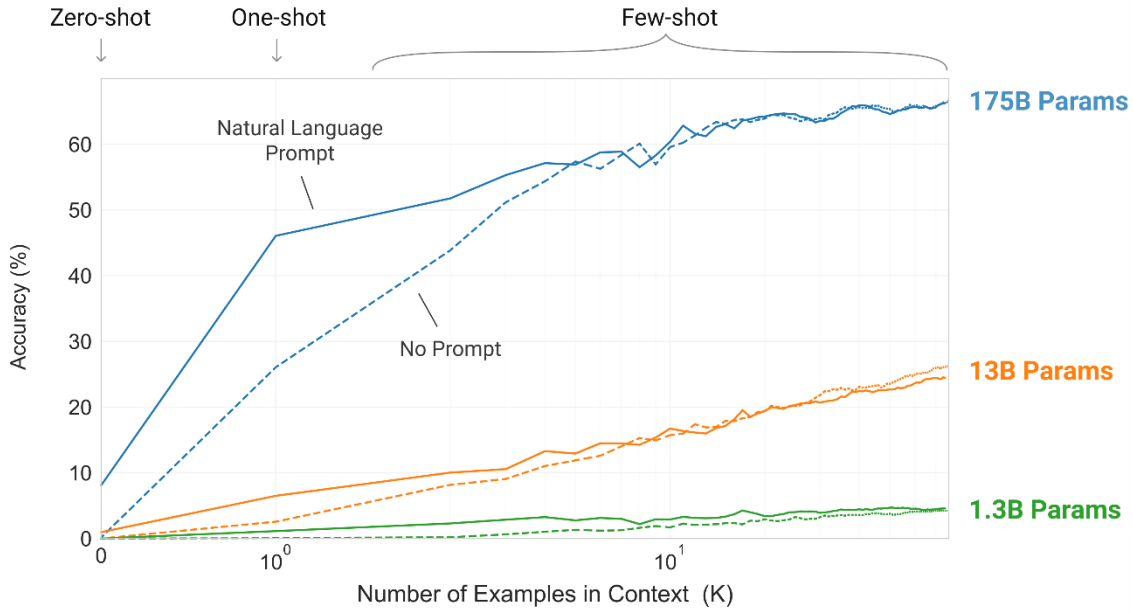
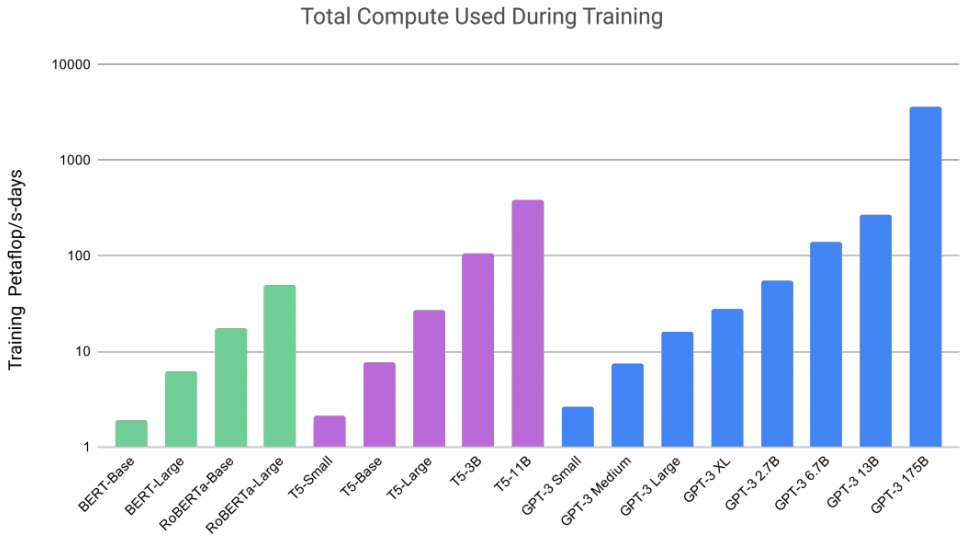
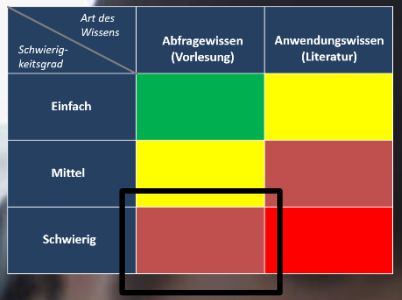
Zero-Shot, One-Shot, Few-Shot Learning





OpenAI 2020: Language Models are Few-Shot Learners
arxiv.org/pdf/2005.14165.pdf

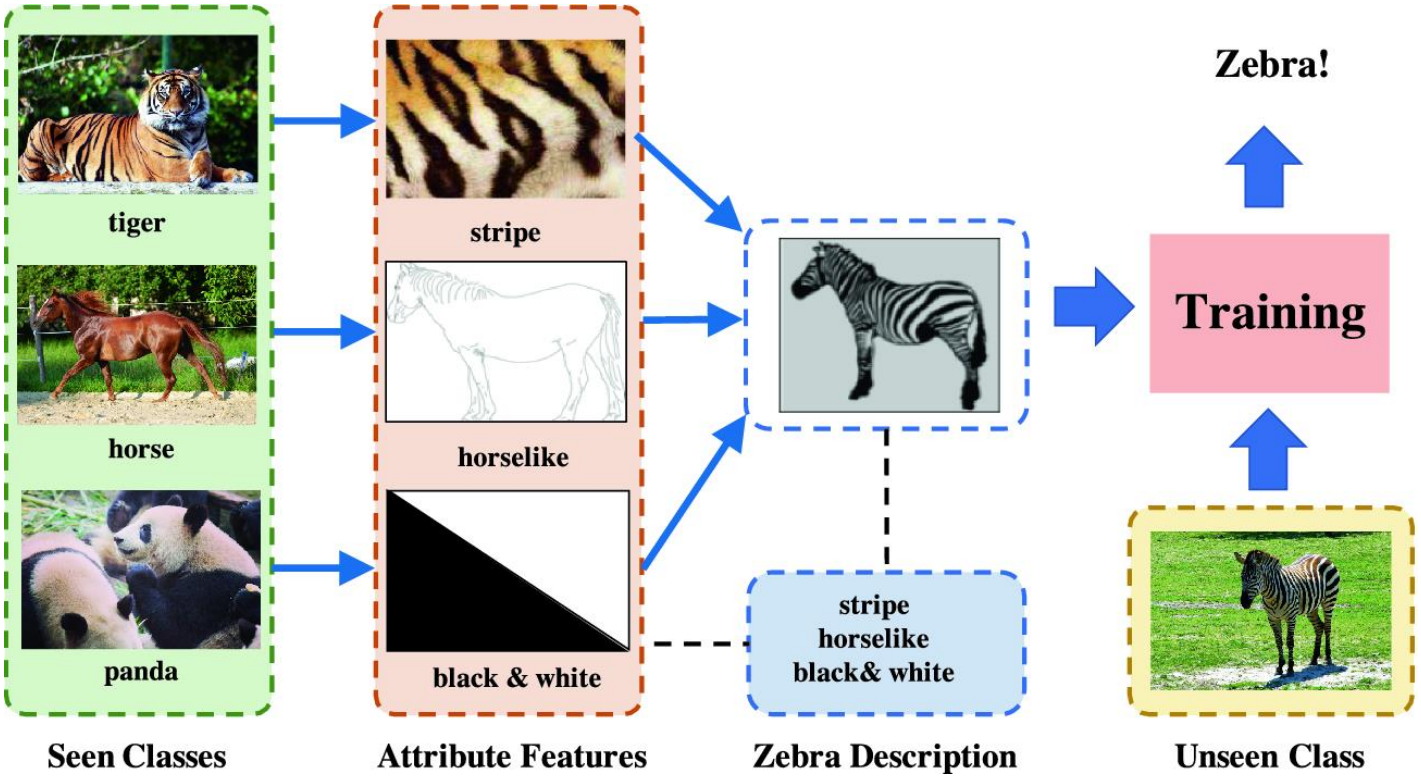
Which statement on GPT3 is FALSE?



a) **GPT-3** shows **characteristics** of **zero-shot learning**.
b) **GPT-3** was an **open-source project acquired by Microsoft**.
c) **BERT** and **GPT-3** are widely used **pretrained Transformers**.
d) In terms of **parameters**, **GPT-3** is **smaller than BERT**.
e) GPT-3 uses a **Context Window** of size 2048.

Zero-Shot Learning (Kernidee)

In **Machine Learning**, **Zero-Shot Learning** is a **problem setup** where, **at test stage**, a learner **recognizes object from classes not previously seen at training stage**. This problem is widely studied in **computer vision**, **natural language processing** and **machine perception**.

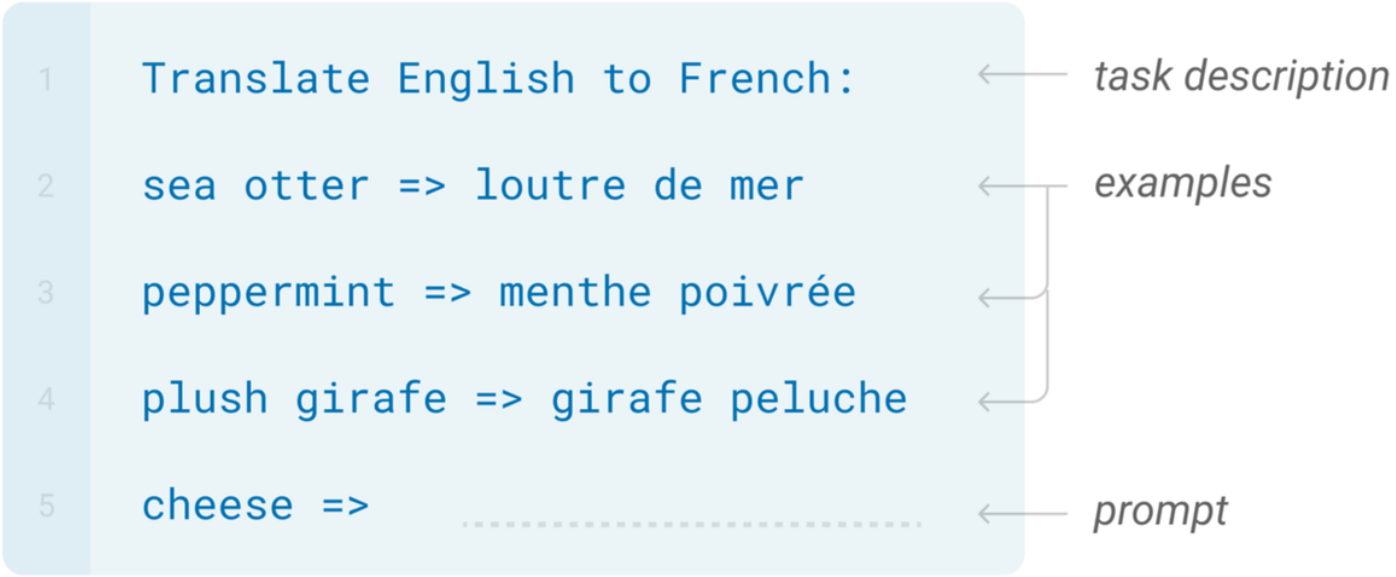


en.wikipedia.org/wiki/Zero-shot_learning
deepai.org/publication/zero-shot-learning-by-generating-pseudo-feature-representations

Zero-Shot, One-Shot, Few-Shot Learning (Kernidee)

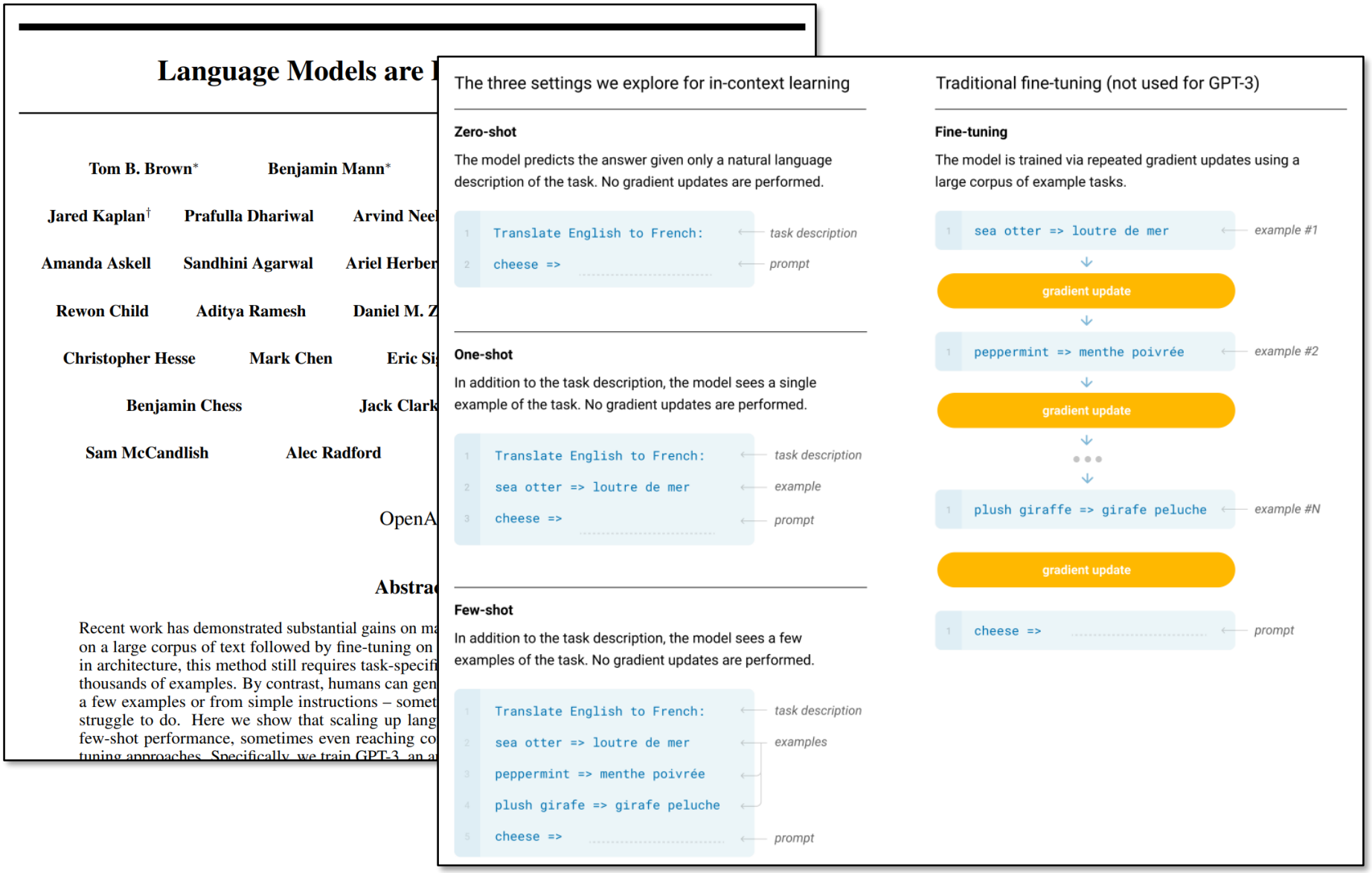
Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



OpenAI 2020: Language Models are Few-Shot Learners
arxiv.org/pdf/2005.14165.pdf

GPT-3 Transformer, OpenAI 2020



Language Models are

Tom B. Brown* Benjamin Mann*

Jared Kaplan† Prafulla Dhariwal Arvind Nee

Amanda Askell Sandhini Agarwal Ariel Herber

Rewon Child Aditya Ramesh Daniel M. Z

Christopher Hesse Mark Chen Eric Si

Benjamin Chess Jack Clark

Sam McCandlish Alec Radford

OpenA

Abstra

Recent work has demonstrated substantial gains on ma
on a large corpus of text followed by fine-tuning on
in architecture, this method still requires task-specifi
thousands of examples. By contrast, humans can gen
a few examples or from simple instructions – somet
struggle to do. Here we show that scaling up lang
few-shot performance, sometimes even reaching co
tuning approaches. Specifically, we train GPT-3, an a

The three settings we explore for in-context learning

Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

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

One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

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

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

```
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
```

Traditional fine-tuning (not used for GPT-3)

Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.

```
1 sea otter => loutre de mer          ← example #1
```

gradient update

```
1 peppermint => menthe poivrée        ← example #2
```

gradient update

```
1 plush giraffe => girafe peluche     ← example #N
```

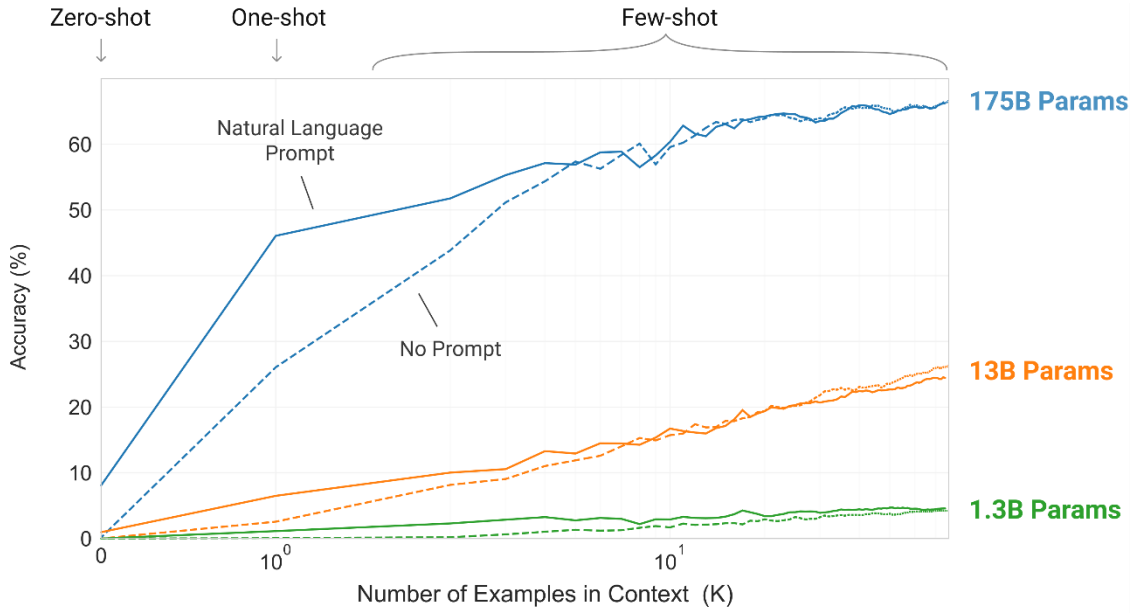
gradient update

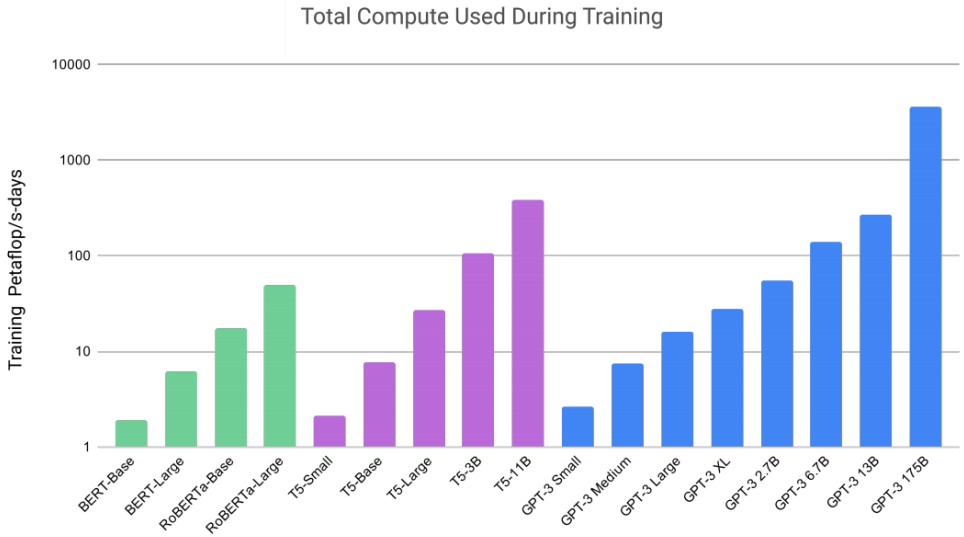
```
1 cheese =>                           ← prompt
```

OpenAI 2020: Language Models are Few-Shot Learners

arxiv.org/pdf/2005.14165.pdf

Zero-Shot, One-Shot, Few-Shot Learning





OpenAI 2020: Language Models are Few-Shot Learners
arxiv.org/pdf/2005.14165.pdf