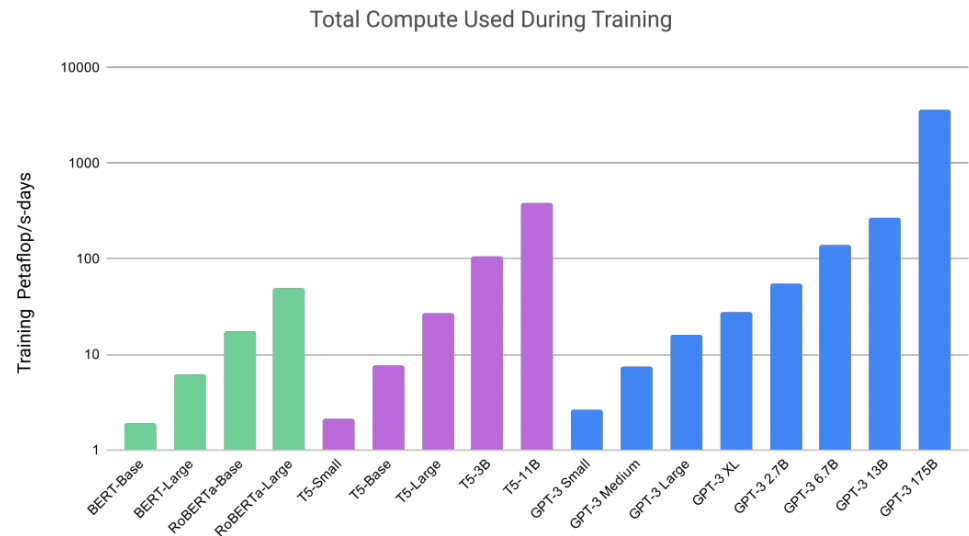
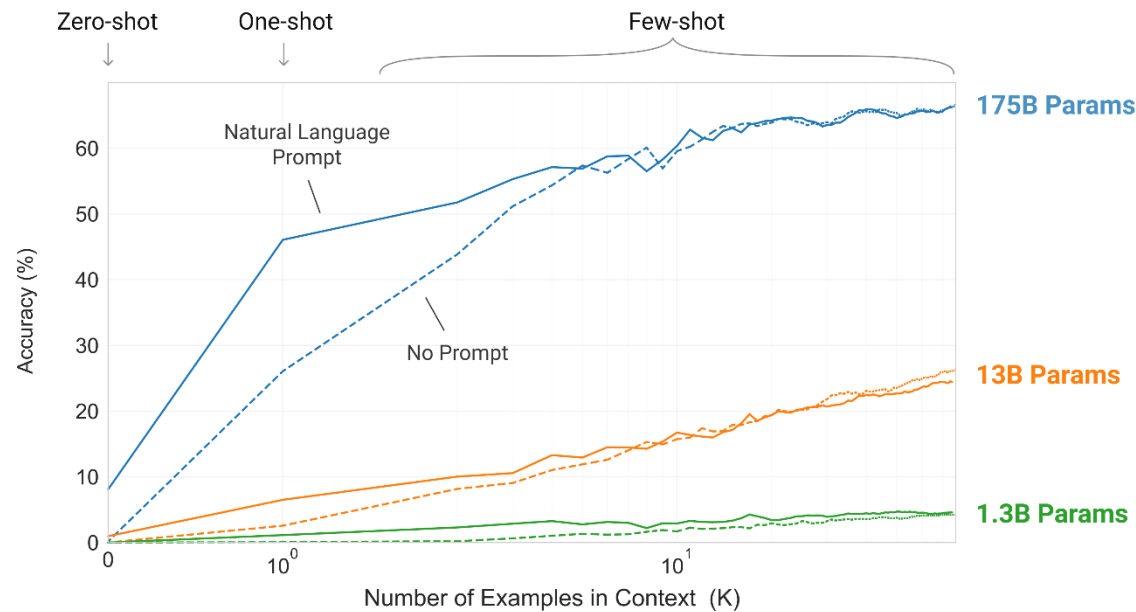


# Zero-Shot, One-Shot, Few-Shot Learning



**OpenAI 2020: Language Models are Few-Shot Learners**  
[arxiv.org/pdf/2005.14165.pdf](https://arxiv.org/pdf/2005.14165.pdf)

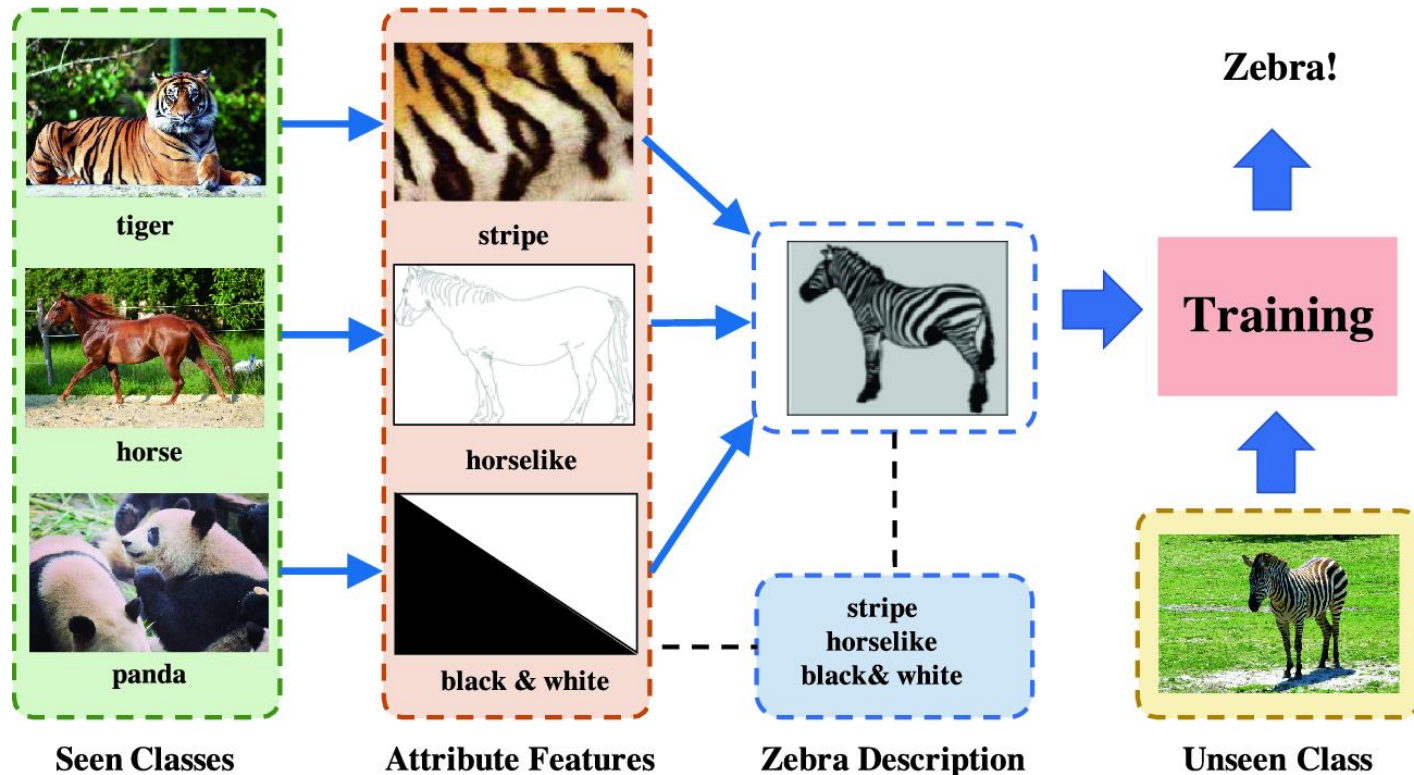
# Which statement on GPT3 is FALSE?

Schwierigkeitsgrad \ Art des Wissens	Abfragewissen (Vorlesung)	Anwendungswissen (Literatur)
Einfach		
Mittel		
Schwierig		

- 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)

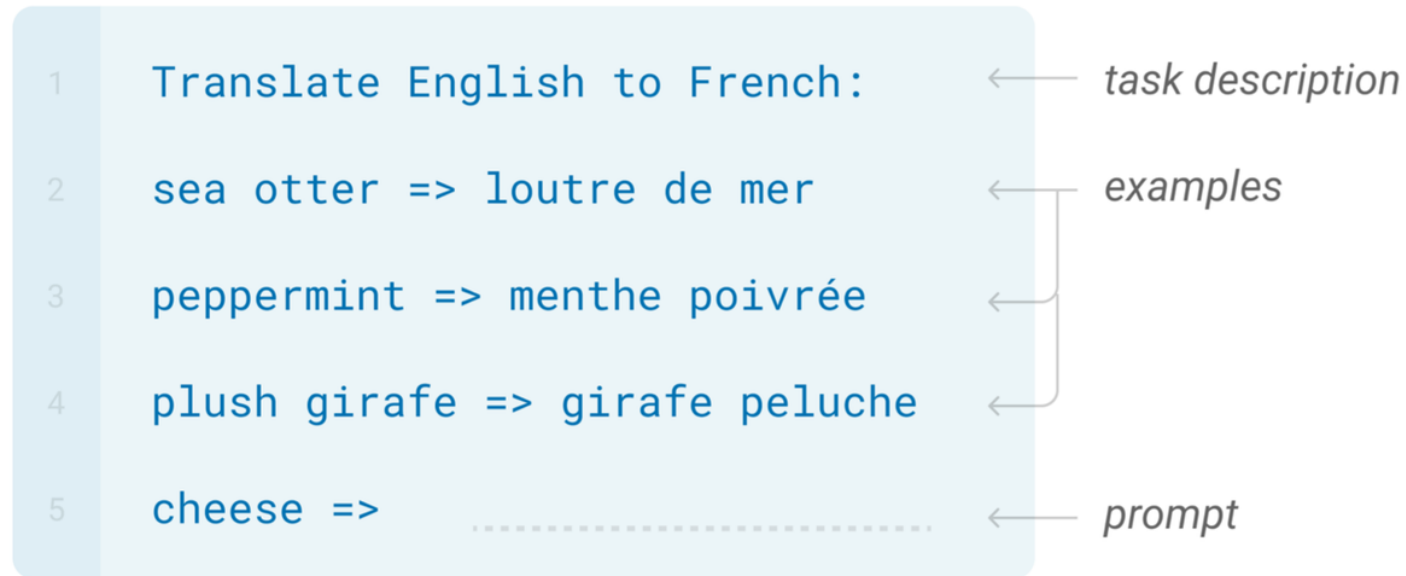
In **Machine Learning**, **Zero-Shot Learning** is a **problem setup** where, at **test stage**, a learner recognizes object from classes not previously seen at **training stage**. This problem is widely studied in **computer vision**, **natural language processing** and **machine perception**.



# 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.



# GPT-3 Transformer, OpenAI 2020

## Language Models are Few-Shot Learners

Tom B. Brown\*

Jared Kaplan†

Amanda Askell

Rewon Child

Christopher Hesse

Benjamin Chess

Sam McCandlish

Benjamin Mann\*

Prafulla Dhariwal

Sandhini Agarwal

Aditya Ramesh

Mark Chen

Alec Radford

Arvind Neelakantan

Ariel Herber

Daniel M. Zou

Eric Sigler

Jack Clark

OpenAI

Abstract

Recent work has demonstrated substantial gains on many NLP tasks using a single large language model trained on a large corpus of text followed by fine-tuning on a task-specific dataset. In this architecture, this method still requires task-specific fine-tuning on thousands of examples. By contrast, humans can generalize from a few examples or from simple instructions – sometimes even reaching competitive performance. Here we show that scaling up language models to billions of parameters enables few-shot performance, sometimes even reaching competitive performance with zero-shot and one-shot approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, and find that it can match or exceed the performance of models trained specifically on downstream tasks. We show that GPT-3 can learn to perform a wide variety of tasks from a few examples, and that it can be used to generate human-like text, code, and even to solve complex reasoning tasks. Our results suggest that large language models trained on a large corpus of text can be used as a general-purpose tool for many NLP tasks, and that they can be used to learn from a few examples, much like humans.

### 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 ← examples

4 plush girafe => girafe peluche ← examples

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

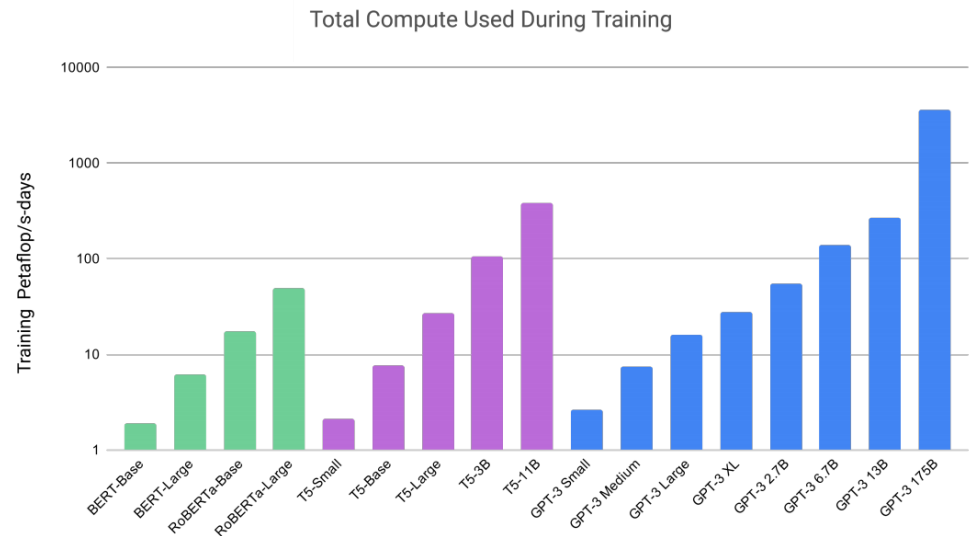
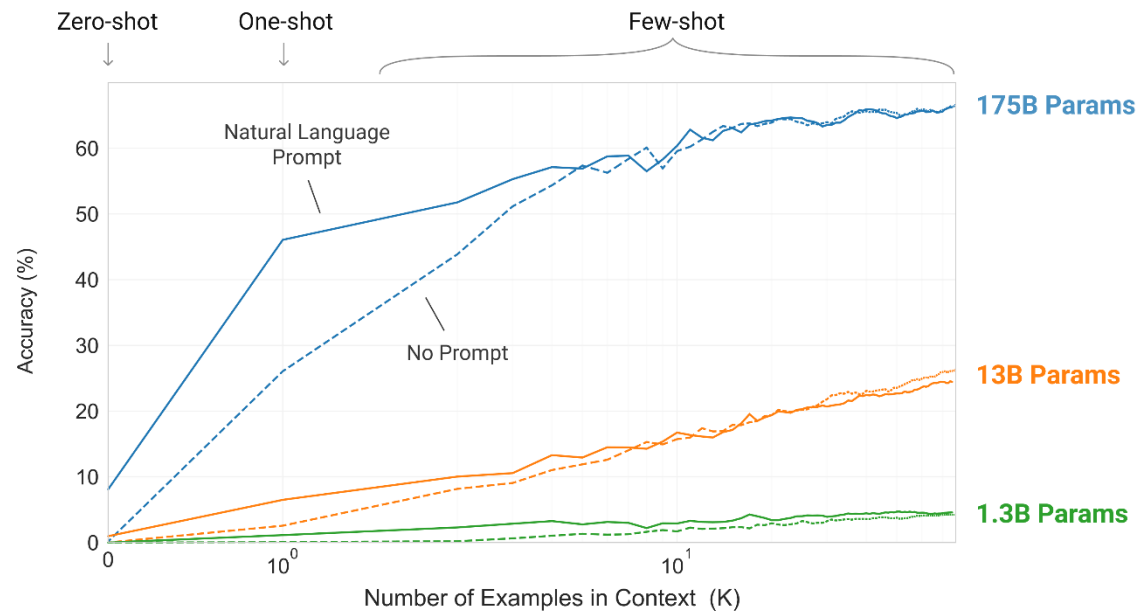
...

1 plush giraffe => girafe peluche ← example #N

gradient update

1 cheese => ..... ← prompt

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



**OpenAI 2020: Language Models are Few-Shot Learners**  
[arxiv.org/pdf/2005.14165.pdf](https://arxiv.org/pdf/2005.14165.pdf)