GPT models: The ultimate Multitaskers?

Generative AI Conversations

Emilie Grégoire
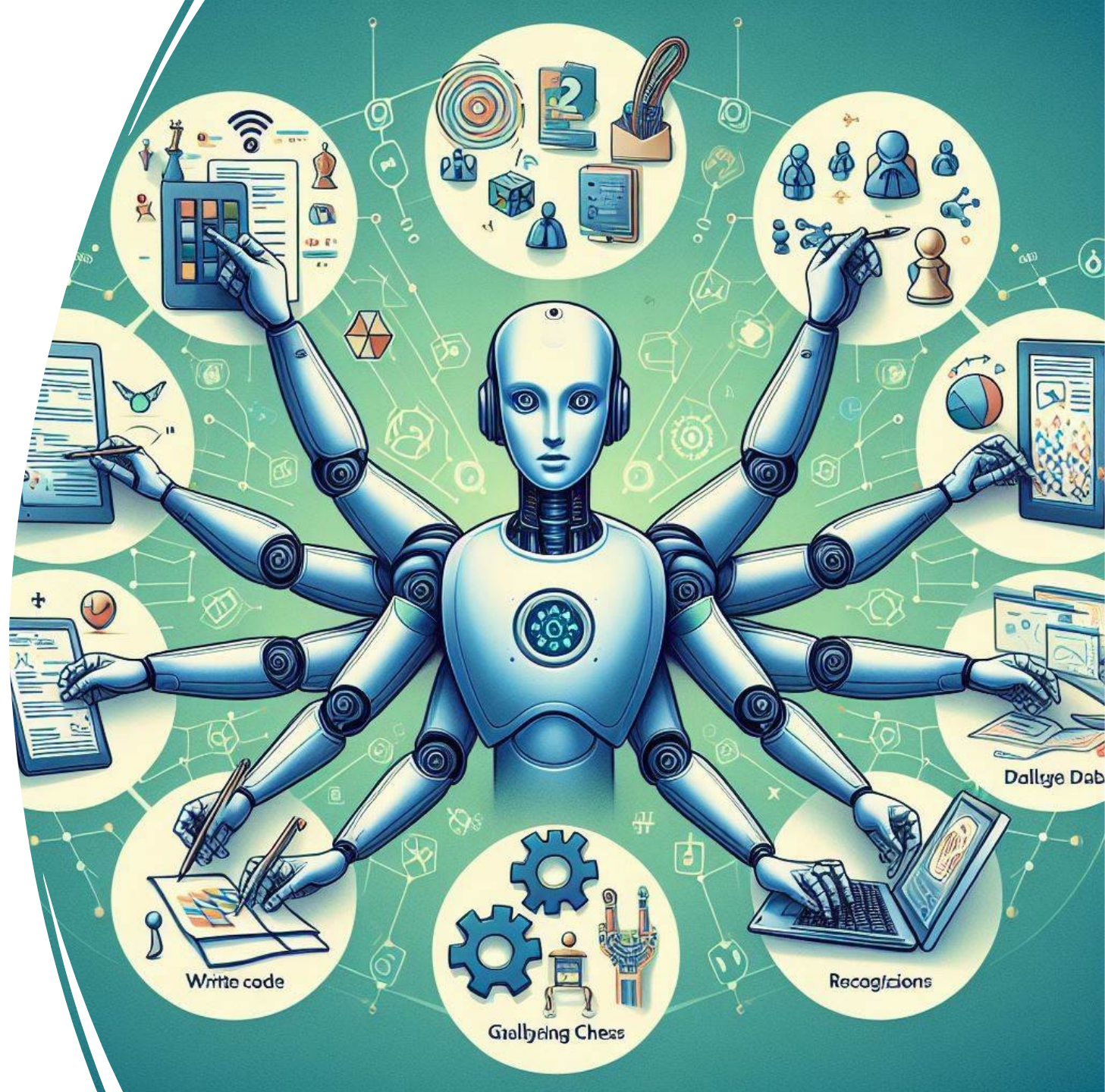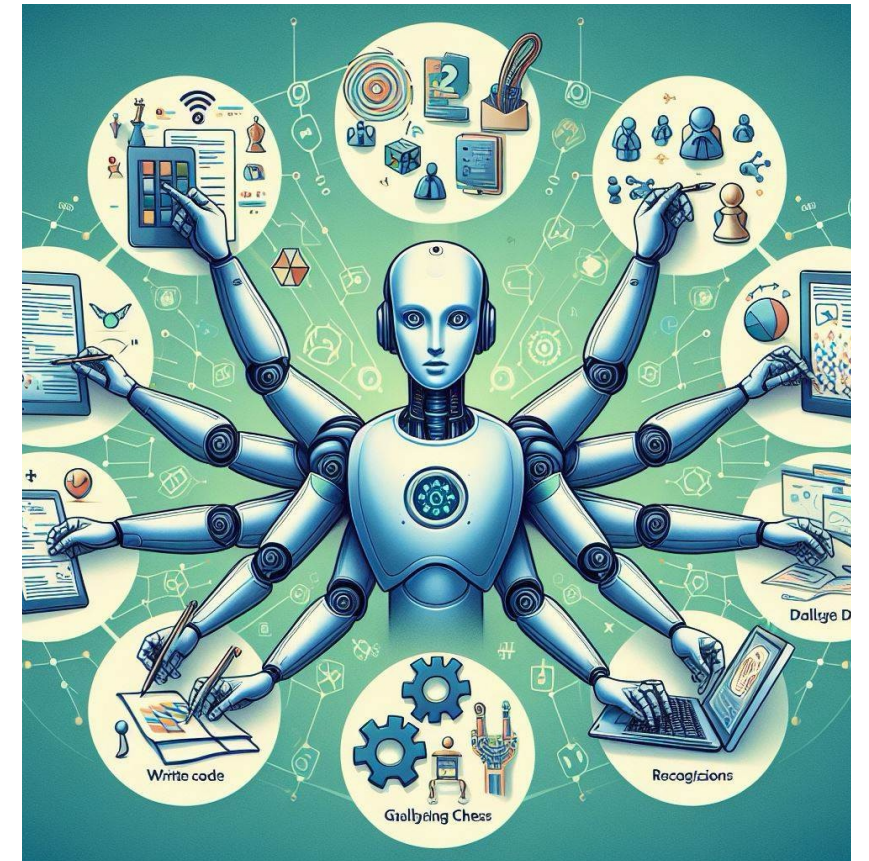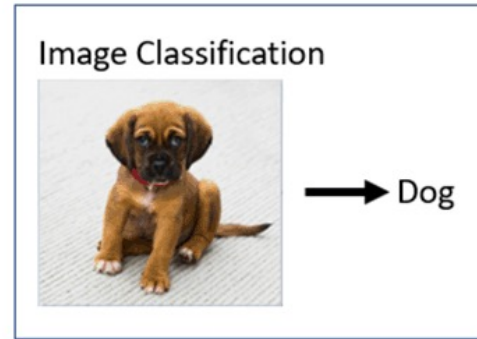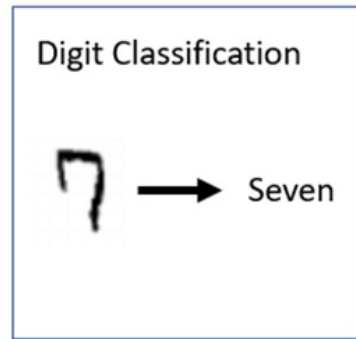
Overview

1. **What do we mean by 'task'?**

2. **Can a model learn multiple tasks?**

3. **Are GPT models born multitaskers?**

4. **Should we stop thinking about multi-task learning?**

What do we mean by "task"?

In (supervised) machine learning a task can be defined as the **type of prediction** or inference being made, **based on** the problem or question that is being asked, and the **available data.**

What do we mean by "task"?

Examples of classification tasks
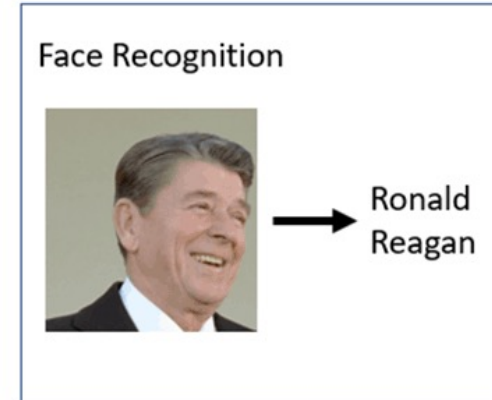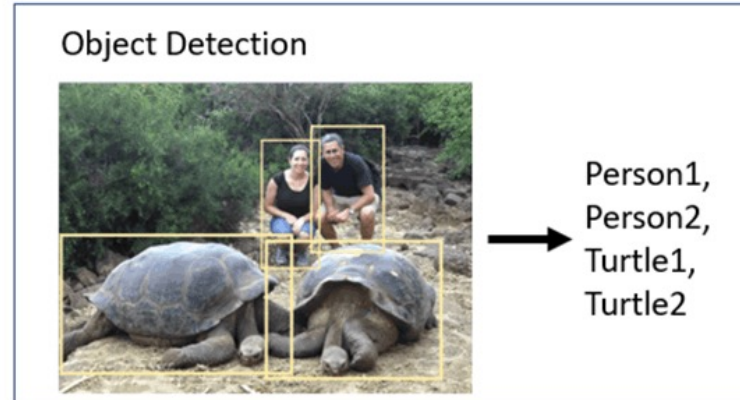
What do we mean by "task"?
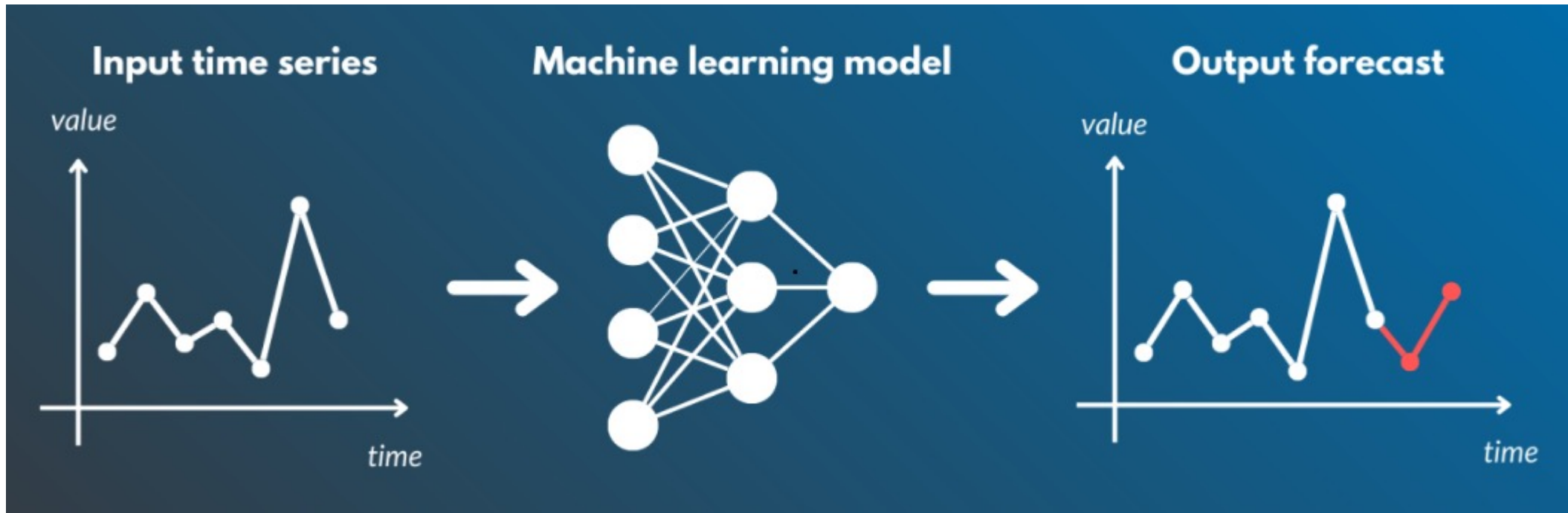
Forecasting tasks

What do we mean by "task"?
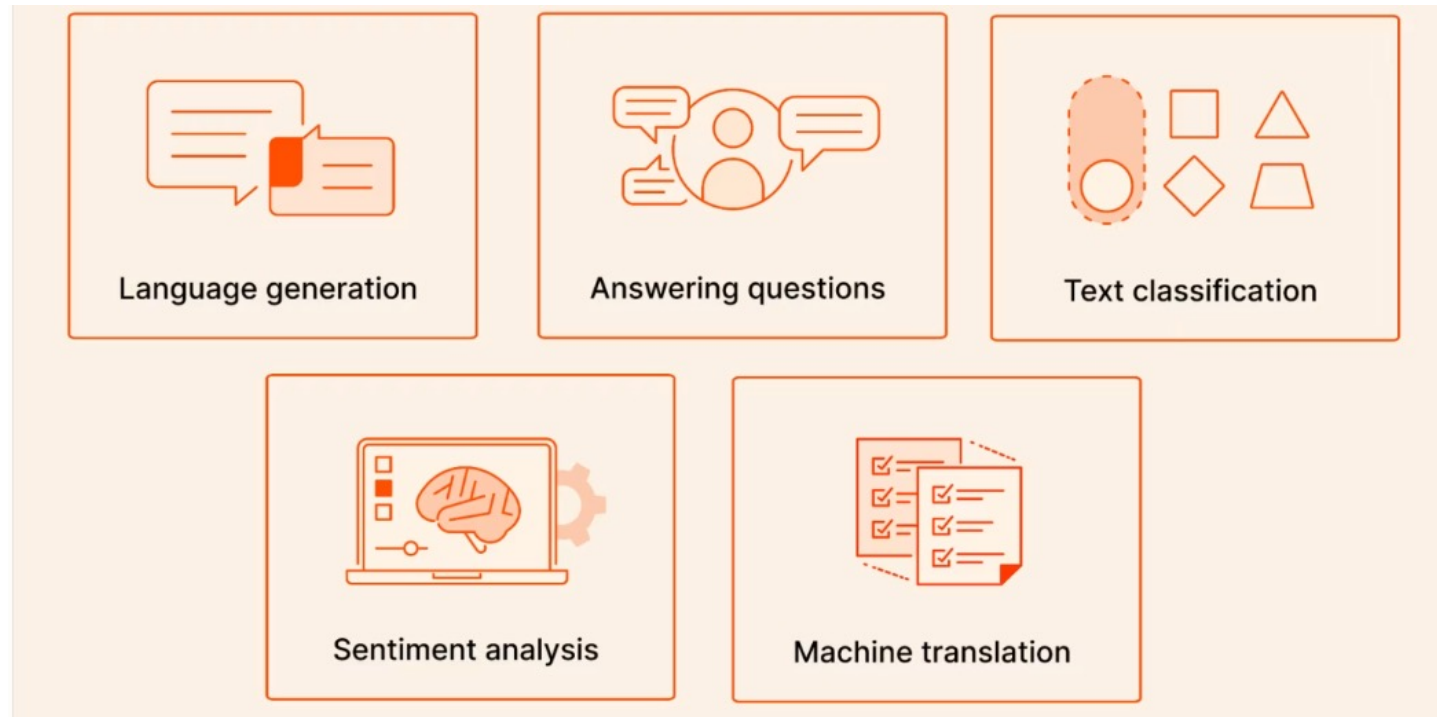
NLP tasks

What do we mean by "learning a task"?
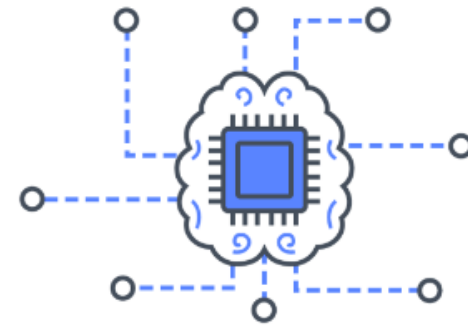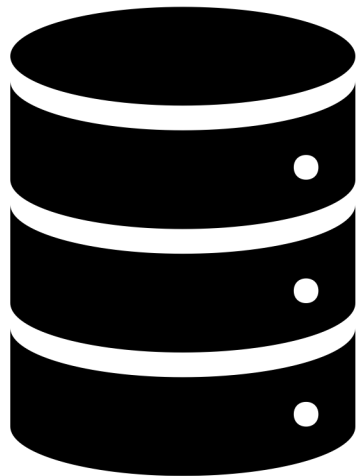
Standard supervised learning: 1 task = 1 model = 1 dataset

What do we mean by "learning a task"?
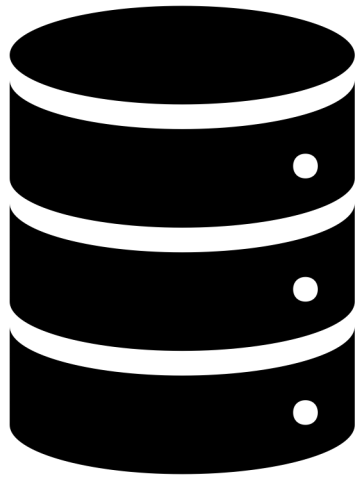
Supervised learning
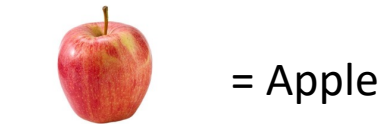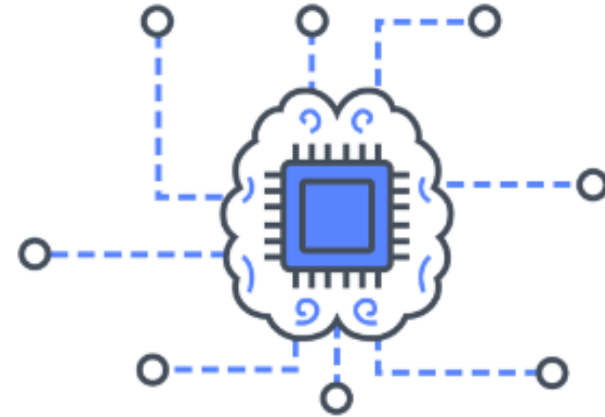
= Apple

= Orange

= Banana

What do we mean by "learning a task"?

Unlabeled unseen data
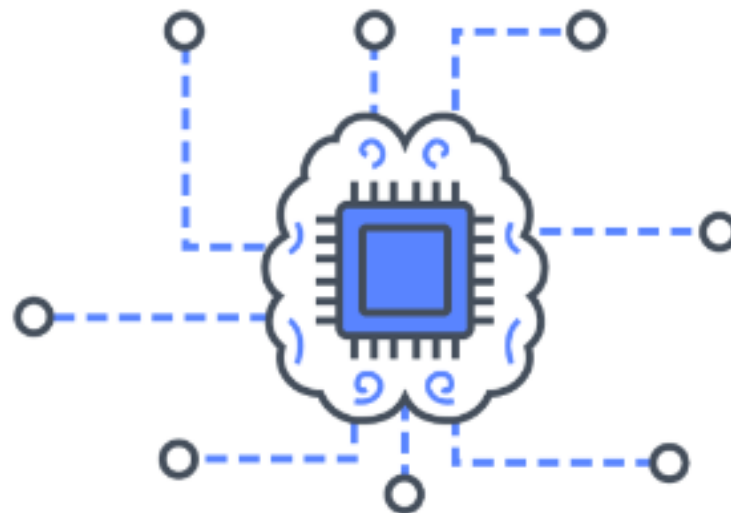
"Orange"

Can a model learn more than one task?

Can a model learn more than one task?

Simultaneously?

Can a model learn more than one task?

Simultaneously?

Three Single Task Models

Name **Maturity** **Quality**

"apple" "unripe" "medium"

One Multi-Task Model

Name **Quality**

"apple" **Maturity** "medium"

"unripe"

Which tasks should be learned together?

Can a model learn more than one task?

Simultaneously?

Three Single Task Models

Name "apple" **Maturity** "unripe" **Quality** "medium"

One Multi-Task Model

Name "apple" **Maturity** "unripe" **Quality** "medium"

Learning tasks simultaneously might also hurt the performance!

$$\mathcal{L}_{MTL}^{(t)} = \sum_{i=1}^{N_T} w_i \mathcal{L}_i(\theta_s^{(t)}, \theta_i^{(t)})$$

Can a model learn more than one task?

Sequentially?

Data Task 1 > Data Task 2

Name

Quality

Fine-tuned for task 2

Can a model learn more than one task?

Sequentially?

Data Task 1 > Data Task 2

Task 1

Model1 → Head → Predictions1   **Name**

Knowledge transfer

How much of the original model should be fine-tuned?

Task 2

Model1 → New Head → Predictions2   **Quality**

Fine-tuned for task 2

Are GPT models born multitaskers?

GPT can do more than "only" next-word prediction (language modeling)!

Text generation

Question answering

Translation

Summarization

Reading comprehension

Are GPT models born multitaskers?

GPT can do more than "only" next-word prediction (language modeling)!

Without the need for additional data or models! *

Model 1 was trained to learn language modeling

Text generation

Question answering

Translation

Reading comprehension

Summarization
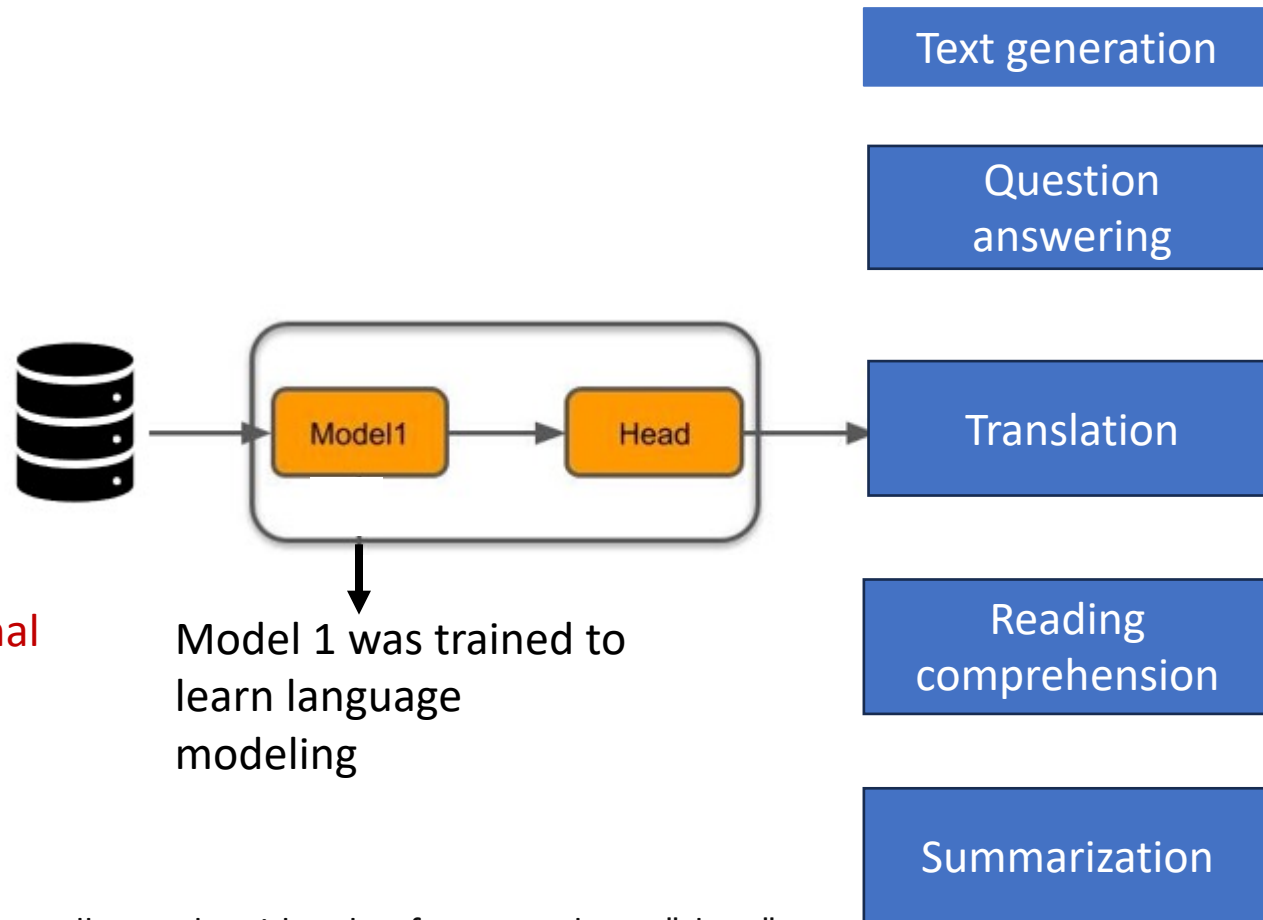
* GPT-3 and successors have the ability perform well on tasks with only a few examples or "shots" provided during inference.

Are GPT models born multitaskers?
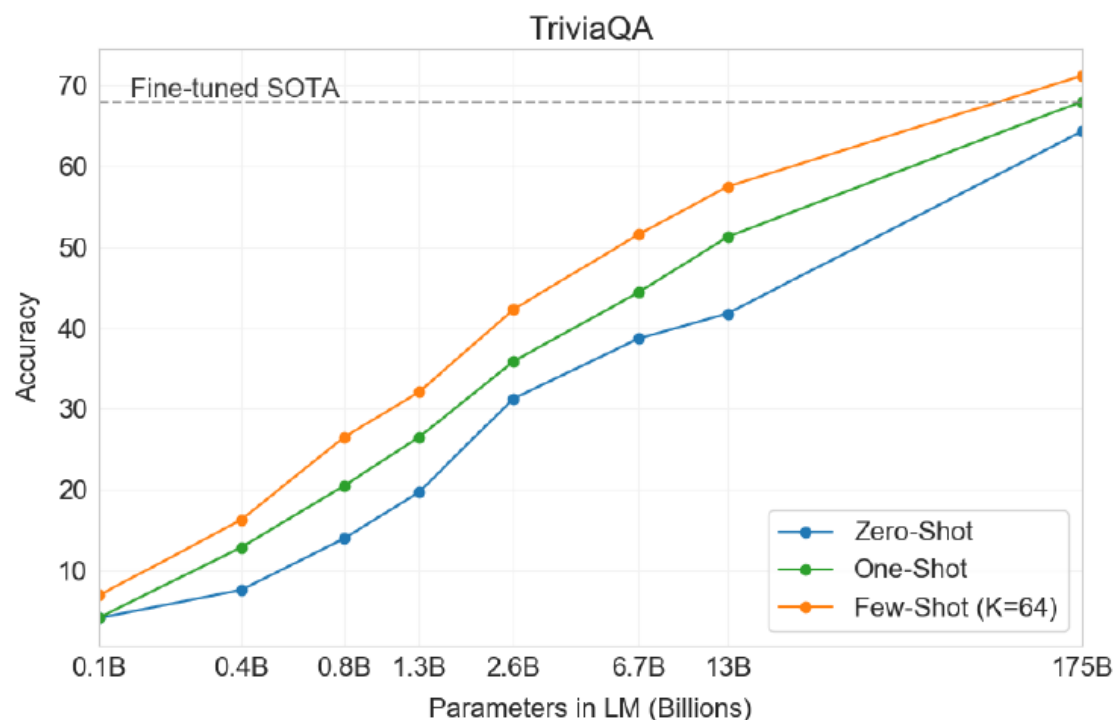

TriviaQA

Figure 3.3: On TriviaQA GPT3's performance grows smoothly with model size, suggesting that language models continue to absorb knowledge as their capacity increases. One-shot and few-shot performance make significant gains over zero-shot behavior, matching and exceeding the performance of the SOTA fine-tuned open-domain model, RAG [LPP+20]

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Are GPT models born multitaskers?

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            ←┐
3   peppermint => menthe poivrée          ←┤  examples
4   plush girafe => girafe peluche        ←┘
5   cheese =>    ..........                ←——— prompt
```

Without the need for additional data or models!
Few/one/zero shots are enough…

HOW?

Are GPT models born multitaskers?
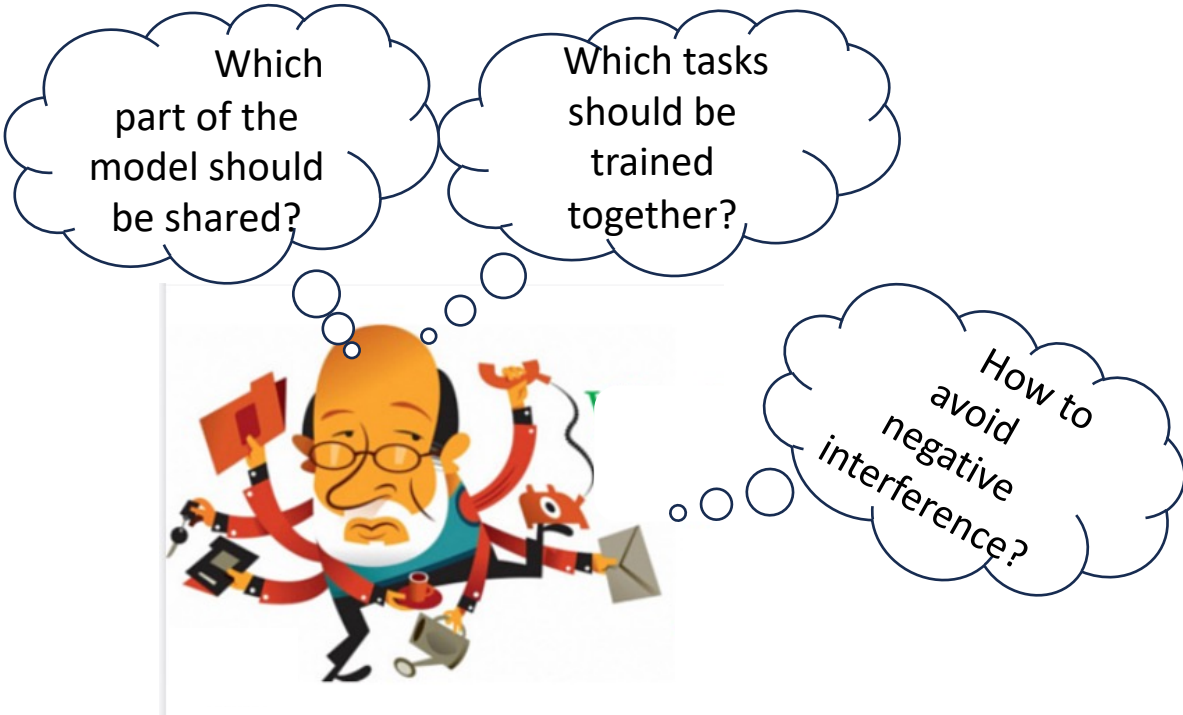
The multi-task learning researcher

Should we stop thinking about MTL?

PhD Thesis

In order to enable a model to learn multiple-tasks, one needs an enormous model and a lot of data [1].

[1] Smart people from OpenAI   1/1

Should we stop thinking about MTL?

Absolutely not!

- Why is zero/few-shot working for some tasks?

- Why does it work better for some tasks than for others?

- Are the zero-shot tasks actually new?

- What about non-NLP tasks?

- Not all tokens matter equally

- How can we make these models more efficient?

- …
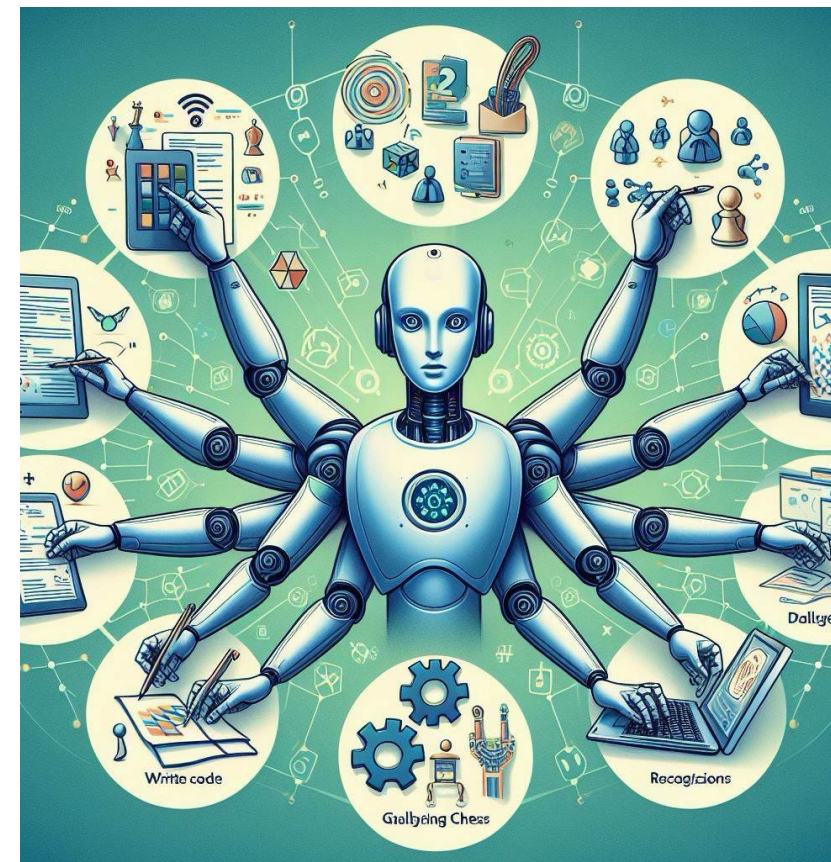
Are GPT models born multitaskers?

ChatGPT

GPT (Generative Pre-trained Transformer) models, including those like GPT-3.5, are not born as multitaskers in the traditional sense. When these models are initially trained, they are typically pre-trained on a large corpus of text data using a single objective, such as language modeling.