

From GPT to Bing Search

In a previous [module \(NLP_Beyond_LLM.ipynb#Using-zero-shot-to-create-new-applications\)](#), we suggested how a Large Language Model could be adapted to a new task

- Fine-Tuning
- Adding a "pre-prompt" as a prefix of user input

Bing Search is a descendant of GPT

- GPT
- InstructGPT
 - GPT fine-tuned to be a helpful assistant
- ChatGPT
 - InstructGPT fine-tuned to be conversational
- WebGPT
 - ChatGPT altered to use "non-parametric knowledge"
 - the Web as Knowledge Source
 - rather than its parameters, which were frozen after Training
- Bing Search
 - WebGPT fine-tuned to Microsoft standards

We give a very brief overview of some of the key steps on this family tree.

There are a lot of very interesting steps that we omit

- Making GPT helpful, truthful and not harmful

Fine-tune: Question Answering

ChatGPT is actually based on InstructGPT

- GPT Fine-tuned for question answering

In order to fine-tune a LLM to answer questions

- we can present it with Question/Answer pairs
- formatted as a long text string

`Question: {question} Answer: {answer}`

- where `{question}` and `{answer}` are place-holders for an example question and its answer.

At inference-time, we just present the question and the request for an Answer

Question: {question} Answer:

and expect the LLM to complete the text by providing the answer.

SQuAD (Stanford Question Answering Dataset) is a dataset frequently used for Question Answering models.

Each example consists of

- a "context": one or more sentences
- a "question"
- an "answer": a substring of the context

Here are some examples

First: the context (which is shared among the first few questions in this toy dataset).

```
In [5]: dest_csv = "squad_show.csv"
squad_df = pd.read_csv(dest_csv)

example = squad_df.iloc[0]
context = example["context"]

# NOTE: the context doesn't change in the first few examples, so we show it only once
print_clean( example["context"])
```

Beyoncé Giselle Knowles-Carter (/bi:'jɒnseɪ/ bee-YON-say) (born September 4, 1981) is an American singer, songwriter, record producer and actress

Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child

Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time

Their hiatus saw the release of Beyoncé's debut album, *Dangerously in Love* (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".

And then, some questions and answers

```
In [6]: for example_no in range(0,5):
        example = squad_df.iloc[example_no]

        question = example["question"]
        text = example["text"]

        print("Question:")
        print_clean( question )

        print("\n")

        answer = example["text"]
        print("Answer:")
        print_clean( answer )

        print("\n\n")
```

Question:
When did Beyonce start becoming popular?

Answer:
in the late 1990s

Question:
What areas did Beyonce compete in when she was growing up?

Answer:
singing and dancing

The answer is a substring of the context, identified by its starting position
Question:
When did Beyonce leave Destiny's Child and become a solo singer?

Answer:
2003

Question:
In what city and state did Beyonce grow up?

Answer:
Houston, Texas

Question:

In which decade did Beyonce become famous?

Answer:

late 1990s

```
In [7]: print(question)
print("\n")
ans_start = example["ans_start"]
print(f"Answer starts at position {ans_start} ")

print("\nAnswer:\n")
print( context[ ans_start: ans_start+len(answer)])
```

In which decade did Beyonce become famous?

Answer starts at position 276

Answer:

late 1990s

The Pre-prompt

The "pre-prompt" for Bing Search instructs it on how to behave.

The instructions were meant to be hidden, but a simple *adversarial attack* caused it to reveal (<https://twitter.com/kliu128/status/1623472922374574080>) its pre-prompt.

In [8]: `print("Done")`

Done

