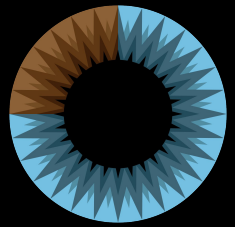
MACHINE LEARNING

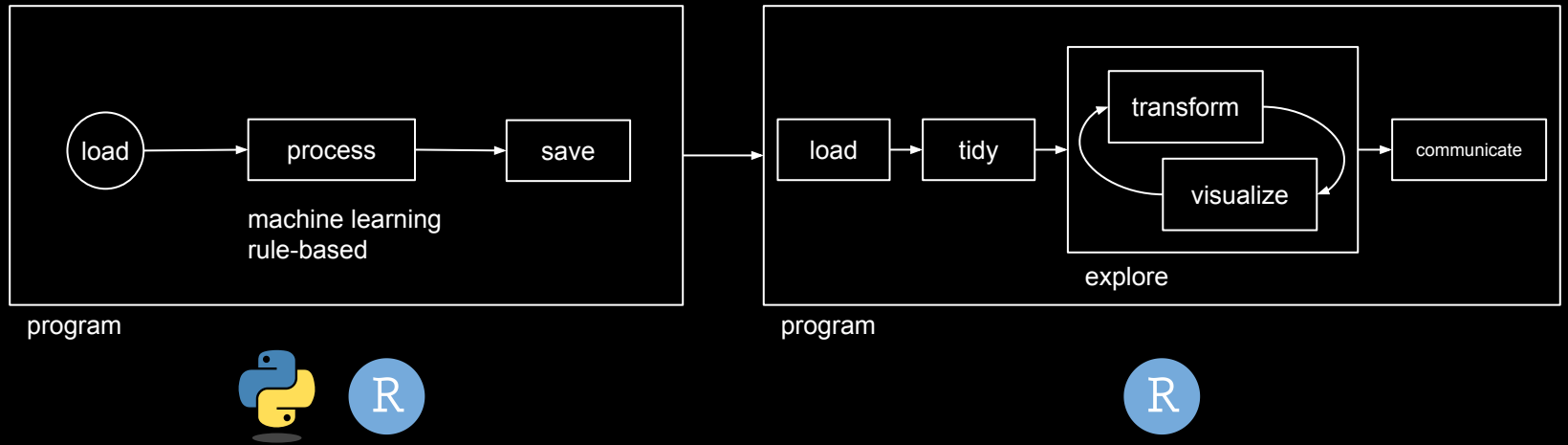
Highly recommended for background information



3Blue1Brown's YouTube Course on Neural Networks and Deep Learning

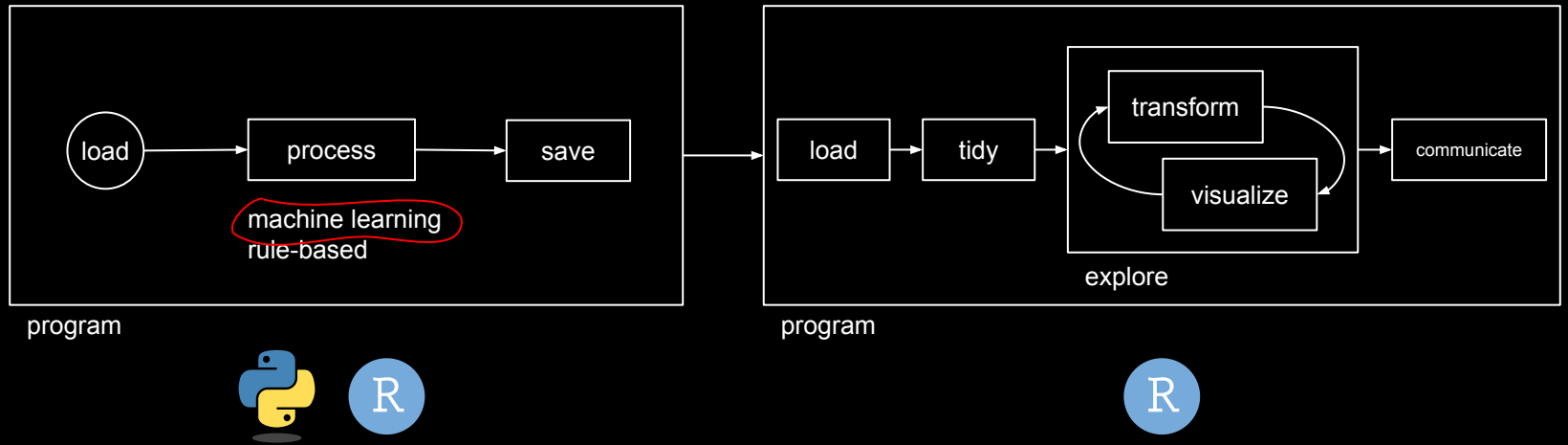
pre-process
unstructured data

exploratory data
analysis

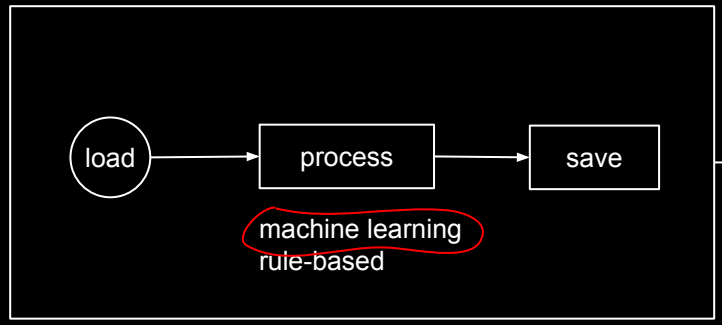


pre-process
unstructured data

exploratory data
analysis

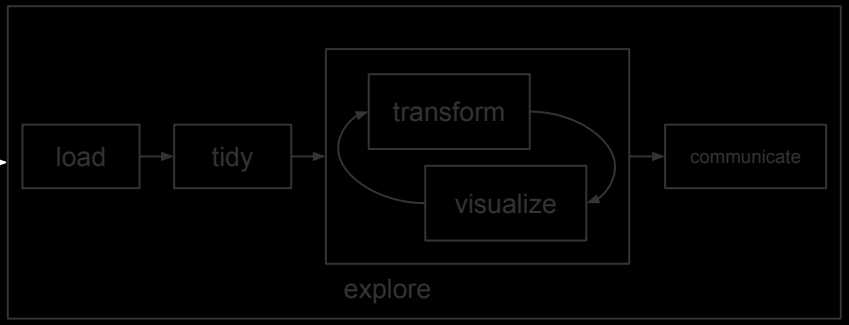


pre-process unstructured data

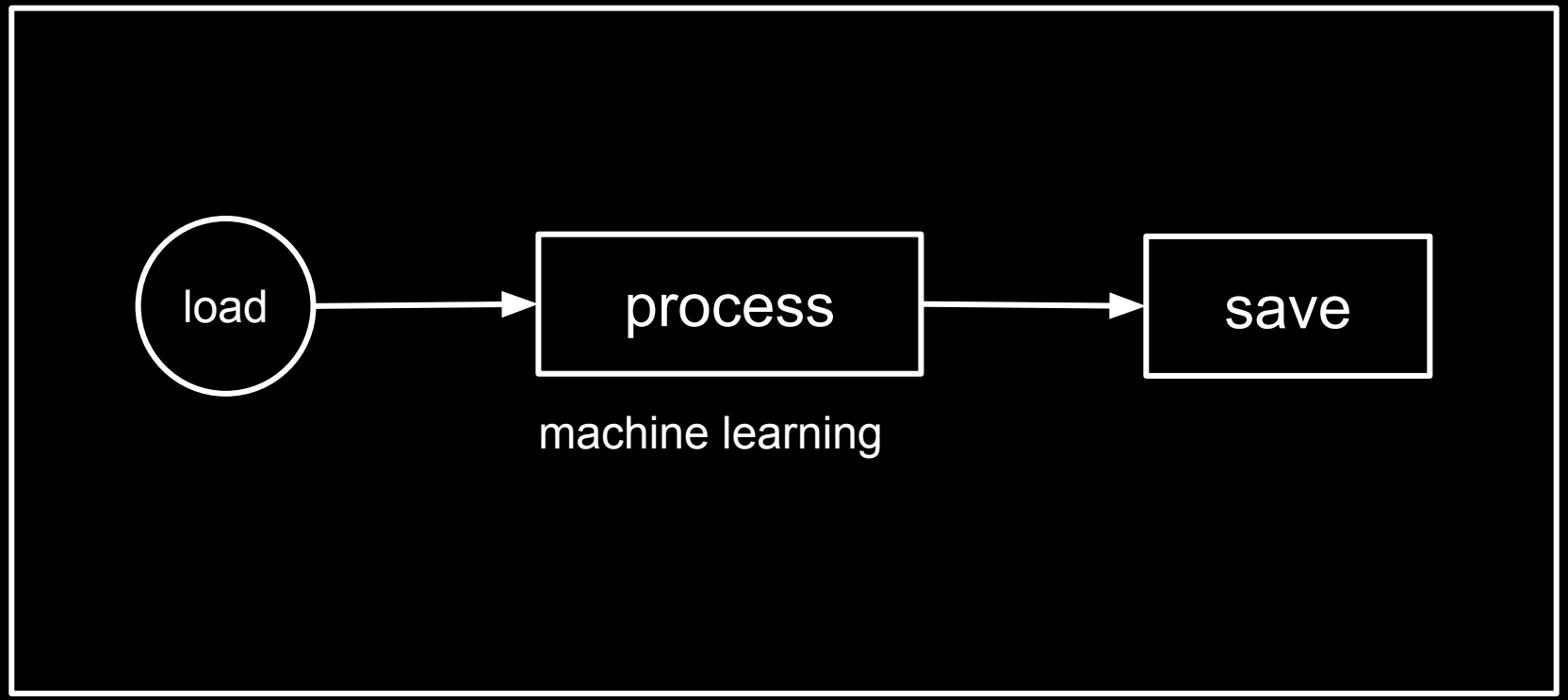


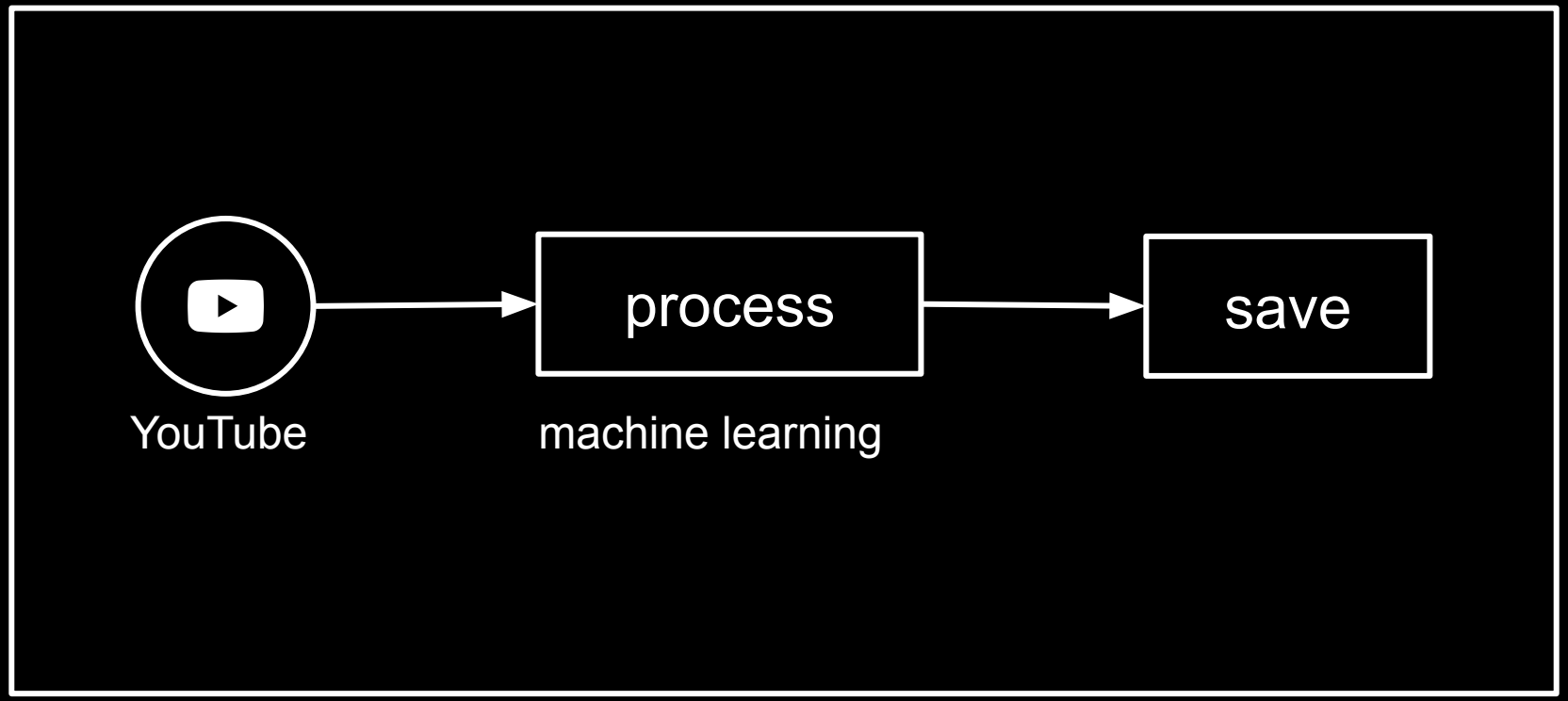
program

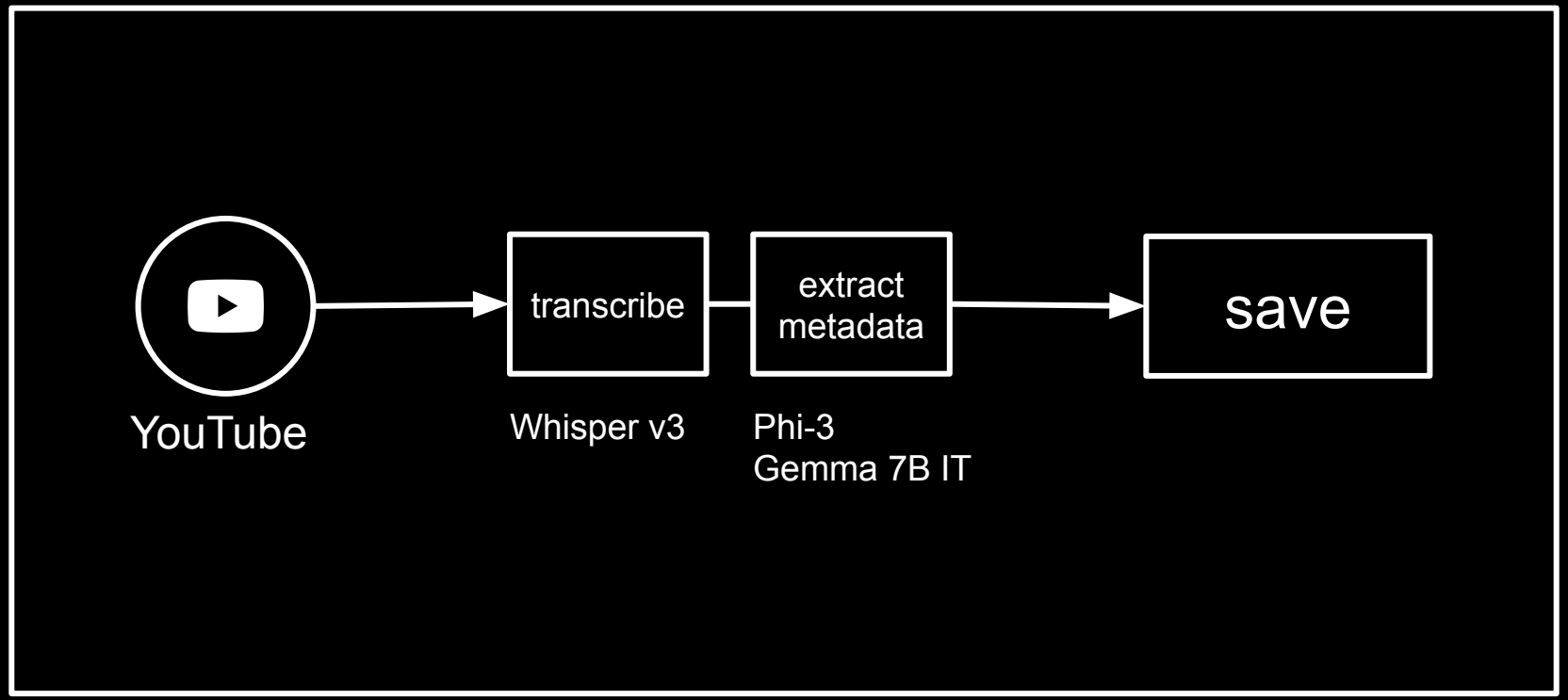
exploratory data analysis

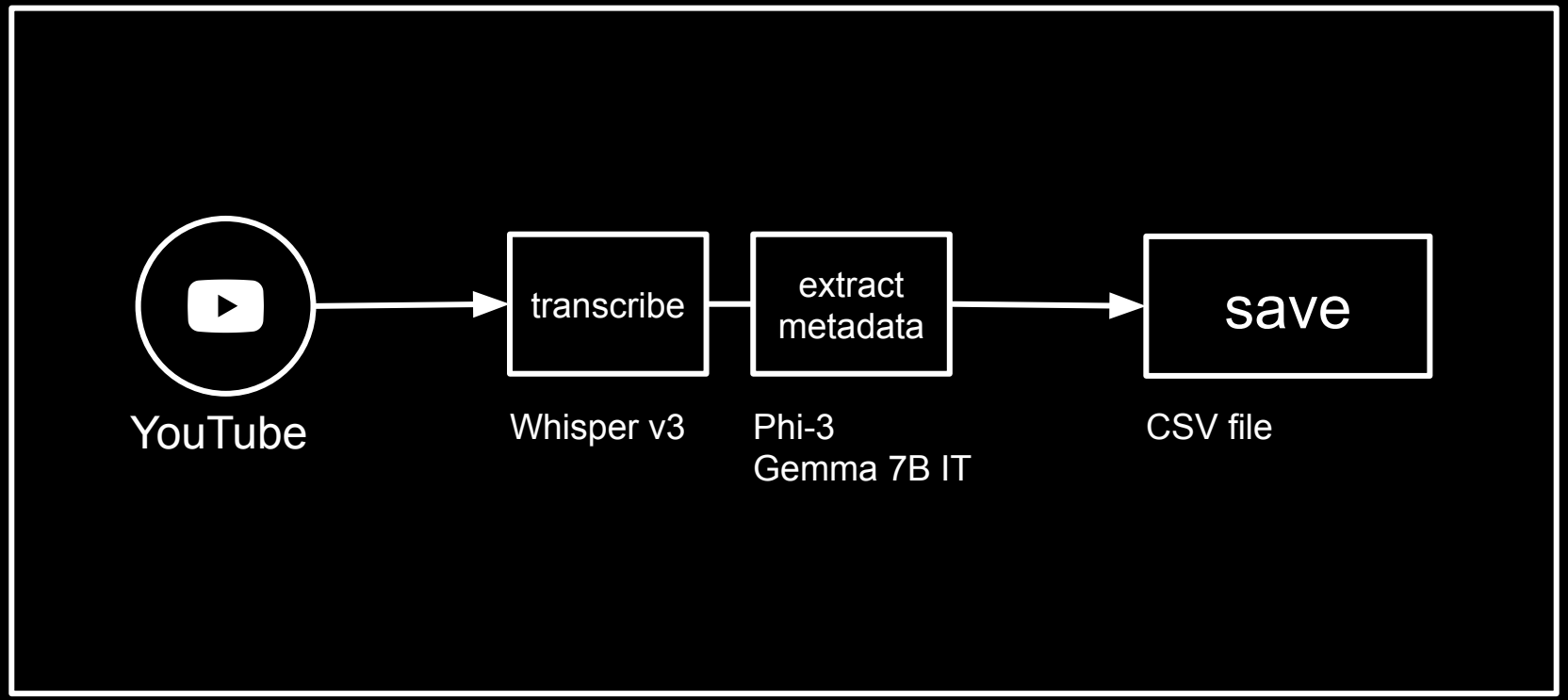


program


load
process
save
machine learning
program


process
save
YouTube
machine learning
program


transcribe
extract metadata
save
YouTube
Whisper v3
Phi-3
Gemma 7B IT
program


transcribe
extract metadata
save
YouTube
Whisper v3
Phi-3
Gemma 7B IT
CSV file
program

YouTube API

Whisper v3

https://arxiv.org/abs/2212.04356

https://huggingface.co/openai/whisper-large-v3


audio file
whisper v3
transcript

Large Language Models (LLM)

what has been said so far?

(*prompt / context*)

what has been said so far?
(*prompt / context*)

prediction of next token based on
learnt probability distribution

what has been said so far?
(*prompt / context*)

prediction of next token based on learnt probability distribution

+

(randomness)

what has been said so far?
(*prompt / context*)

prediction of next token based on learnt probability distribution

+

(randomness)

+

(filter)

(discriminating, insulting content)

what has been said so far?
(*prompt / context*)

prediction of next token based on
learnt probability distribution

+

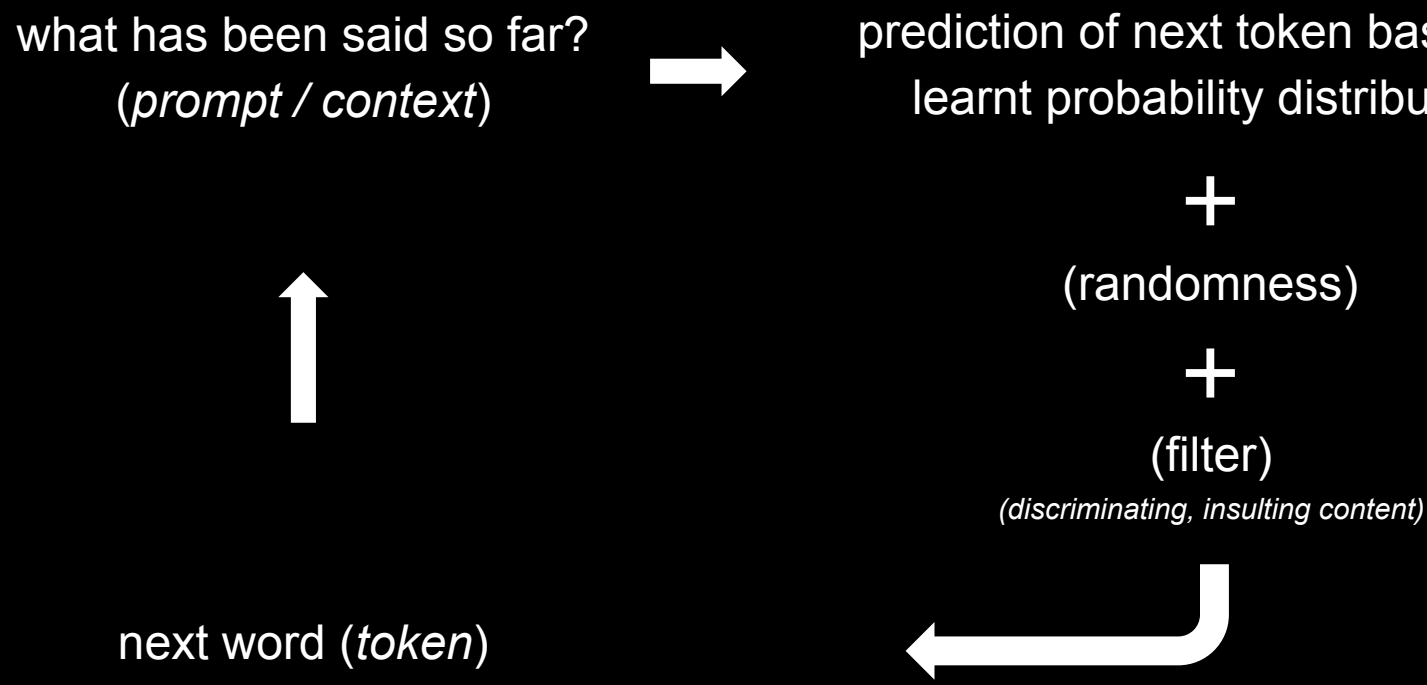
(randomness)

+

(filter)

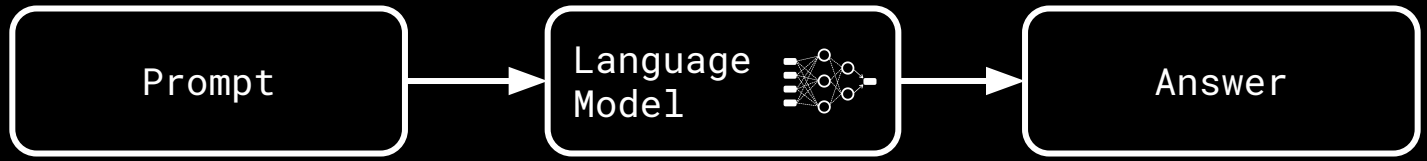
(discriminating, insulting content)

next word (*token*)


what has been said so far?
(*prompt / context*)
prediction of next token based on learnt probability distribution
+
(randomness)
+
(filter)
(discriminating, insulting content)
next word (*token*)

PROMPTING

https://www.promptingguide.ai/


Prompt
Language
Model
Answer

elements of a prompt

```
<instruction>
<context>
<input data>
<output indicator>
```

elements of a prompt

```
<instruction>
<context>
<input data>
<output indicator>
```

example prompt

```
Explain the binary number system.
```

elements of a prompt

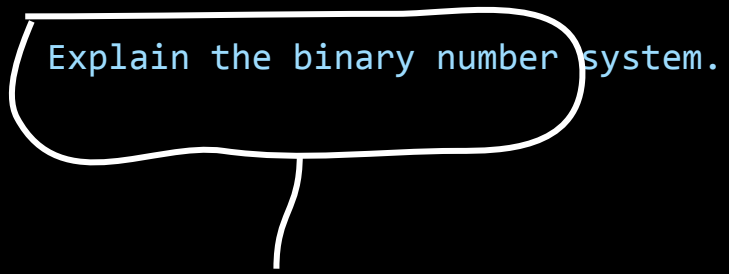
<instruction>

<context>

<input data>

<output indicator>

example prompt



start simple

elements of a prompt

<instruction>

<context>

<input data>

<output indicator>

example prompt

```
You are a friendly tutor and your
task is to explain complex
concepts as simple as possible.

Explain the binary number system.
```

elements of a prompt

<instruction>

<context>

<input data>

<output indicator>

example prompt

```
You are a friendly tutor and your
task is to explain complex
concepts as simple as possible.

Your answers are never longer
than 10 sentences.

Explain the binary number system.
```

ZERO-SHOT PROMPTING

elements of a prompt

<instruction>

<context>

<input data>

<output indicator>

example prompt

```
Classify the text into neutral,
negative or positive.

Text: "What a great dinner!"

Sentiment:
```

elements of a prompt

<instruction>

<context>

<input data>

<output indicator>

example prompt

```
Classify the text into neutral,
negative or positive.

Text: "What a great dinner!"

Sentiment:
```

this will be replaced with data later…

FEW-SHOT PROMPTING

IN-CONTEXT LEARNING

examples in the context to learn from

```
Extract all references to countries and their continent in the following text
using the format from the examples below.

Example 1: "They played the team called 'Die Mannschaft' in the world cup final"
Correct answer: Germany, Europe

Example 2: "The Three Lions once again lost to Germany in a semi final"
Correct answer: England, Europe, Germany, Europe

Text: "The Selecao was destroyed 1:7 by the DFB selection in their home stadium."
Answer:
```

examples in the context to learn from

Extract all references to countries and their continent in the following text using the format from the examples below.

Example 1: "They played the team called 'Die Mannschaft' in the world cup final"
Correct answer: Germany, Europe

Example 2: "The Three Lions once again lost to Germany in a semi final"
Correct answer: England, Europe, Germany, Europe

Text: "The Selecao was destroyed 1:7 by the DFB selection in their home stadium."
Answer:

more prompting strategies

chain-of-thought (CoT)
self-consistency
generate knowledge prompting
prompt chaining (subtasks)
tree-of-thoughts (ToT)
retrieval-augmented-generation (RAG)
…

Phi-3

https://arxiv.org/abs/2404.14219

~~https://huggingface.co/microsoft/Phi-3-mini-128k-instruct~~

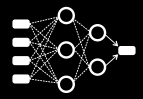
https://huggingface.co/microsoft/Phi-3-medium-128k-instruct

Gemma 2B / 7B Instruct

https://arxiv.org/abs/2403.08295

https://huggingface.co/google/gemma-2b-it

https://huggingface.co/google/gemma-7b-it



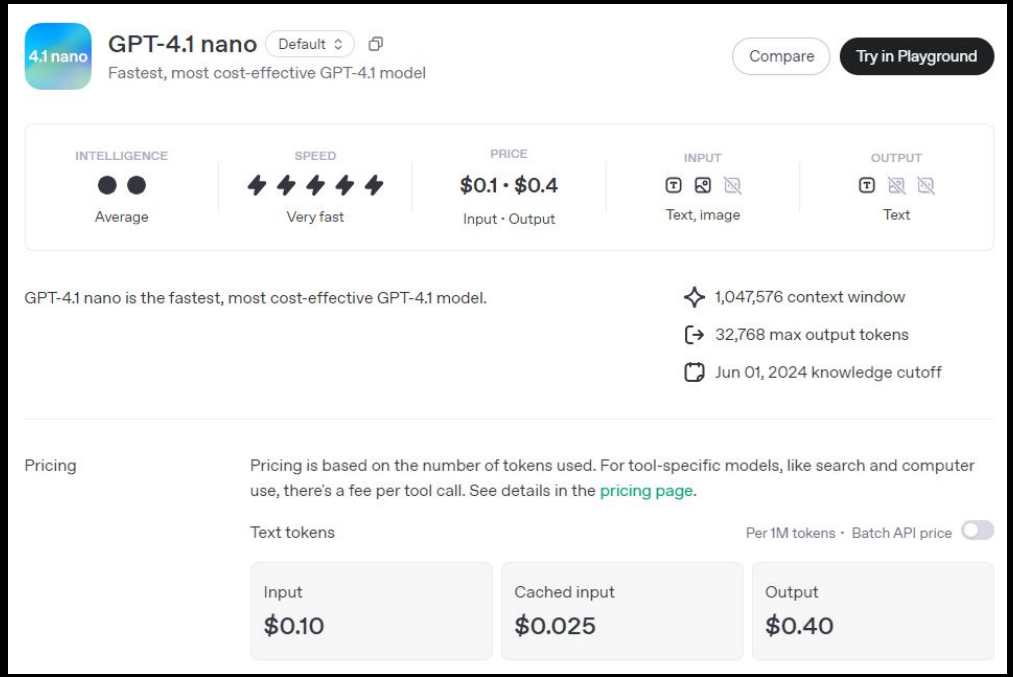
prompt +
context

phi-3
gemma-2b-it
gemma-7b-it

answer

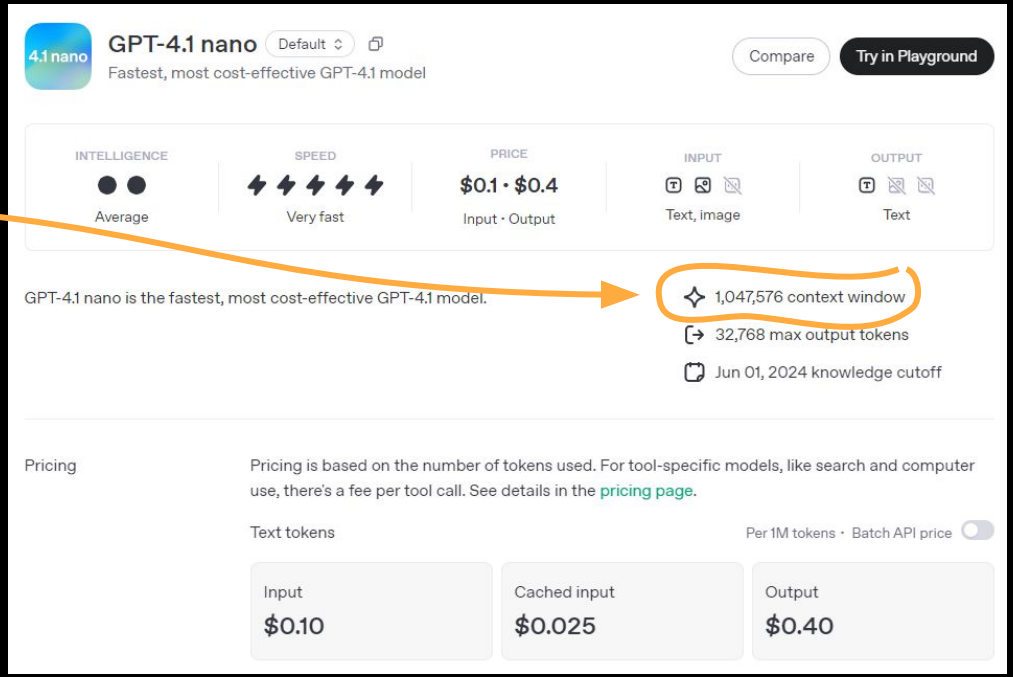
OpenAI's GPT-4.1 family

https://platform.openai.com/docs/models



https://platform.openai.com/docs/models/gpt-4.1-nano

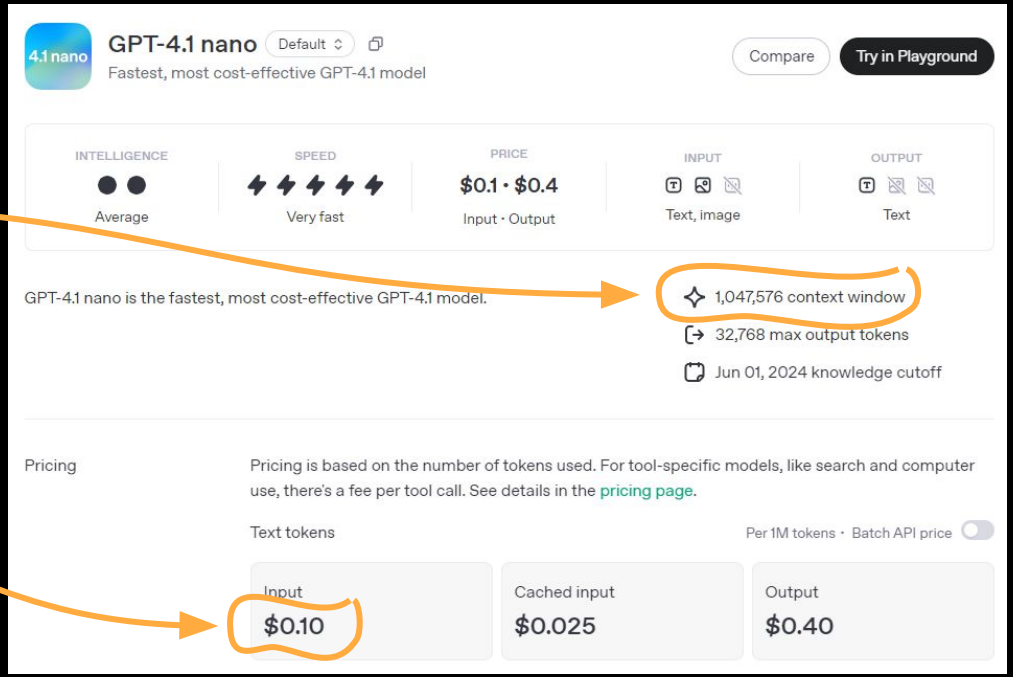
1 mio. token →
~2500 pages



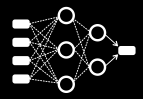
https://platform.openai.com/docs/models/gpt-4.1-nano

1 mio. token →
~2500 pages

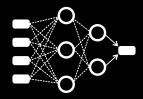
roughly 10 cents
as input



https://platform.openai.com/docs/models/gpt-4.1-nano



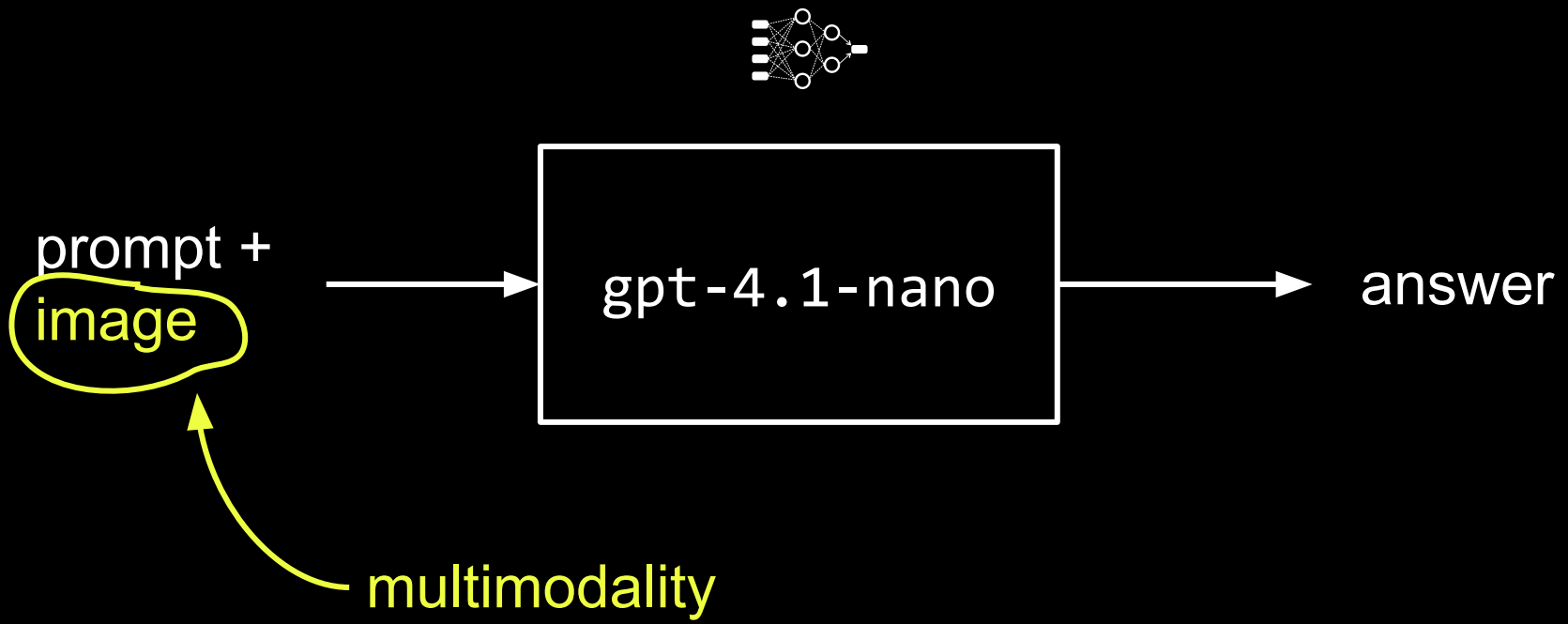

prompt
gpt-4.1-nano
answer



prompt +
image

gpt-4.1-nano

answer


prompt +
image
gpt-4.1-nano
answer
multimodality