

Applying open-source LLMs in Social and Behavioral Sciences

Dirk Wulff & Zak Hussain
@GSERM Ljubljana 2026



MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT

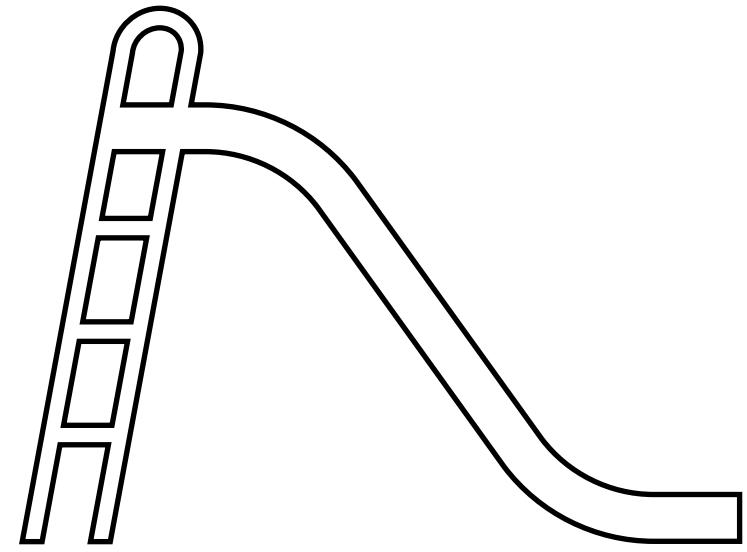


Goals

Familiarize you with the workings and applications of open-source LLMs and how to implement them using the Hugging Face ecosystem

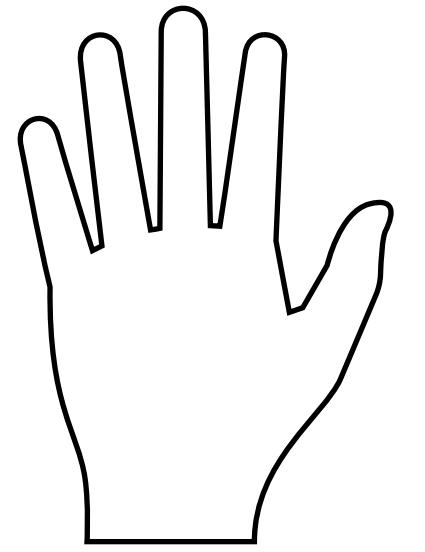


Components



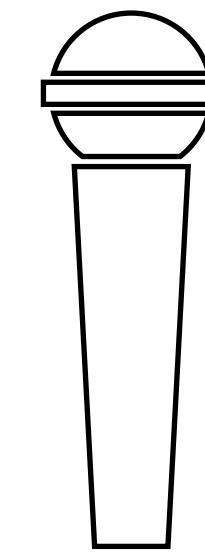
Slide-based introductions

Introduces core concepts and code



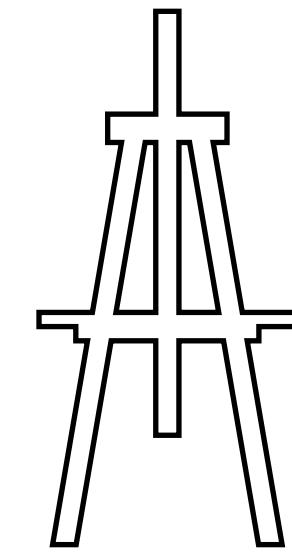
Hands-on exercises

Work through ready-made notebooks to carry out analyses step-by-step



Discussions

Discuss and reflect on applications of LLMs

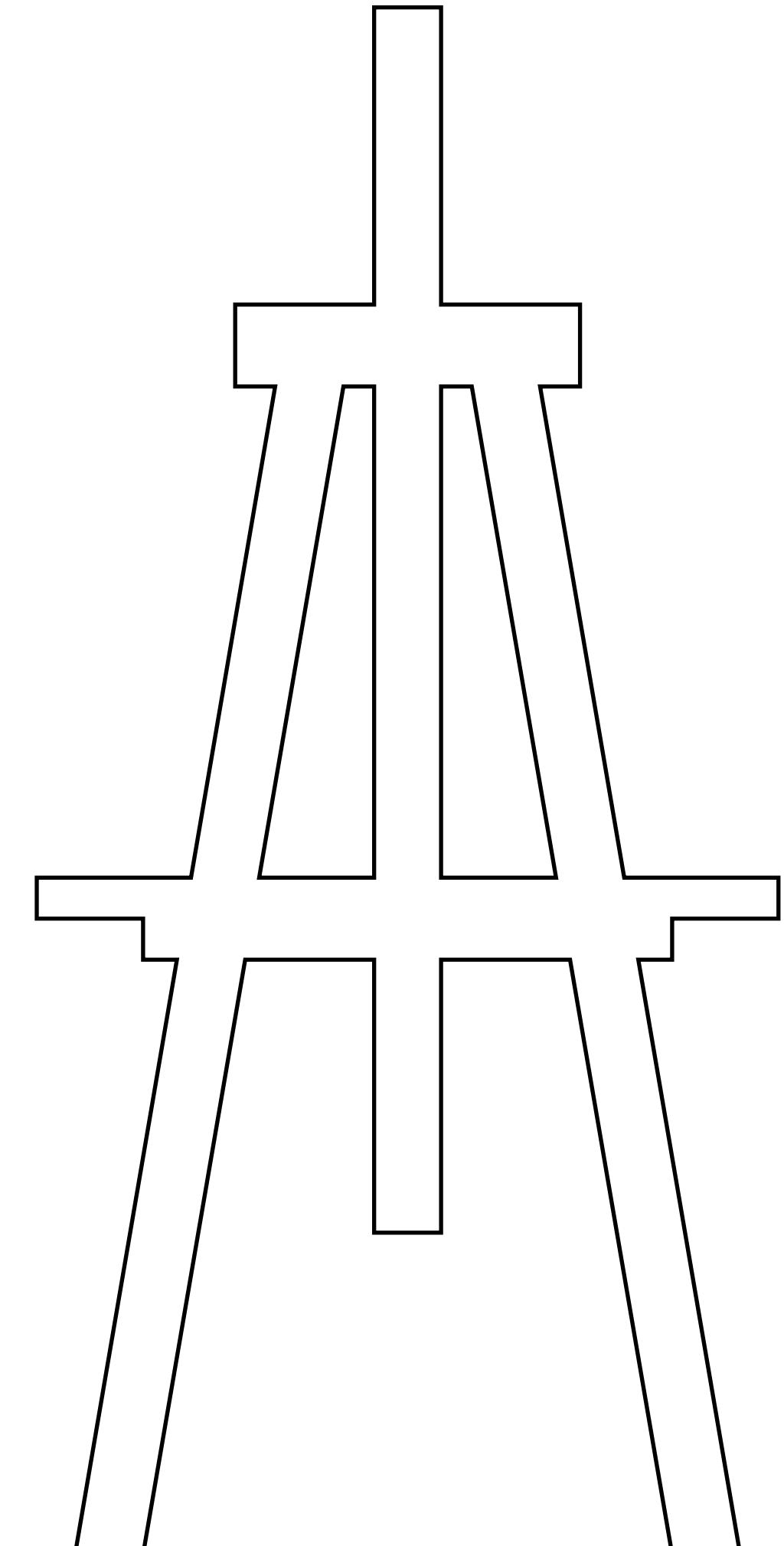


Project sketching

Sketch and pitch an LLM project of your own

Project

- † Any project using LLMs
- † Needs to be motivated (i.e., why is it interesting?)
- † Can be based on the exercises, but must go beyond (i.e., different data, different models, etc.)
- † Is presented at the end (max. 3 slides and 4 minutes)
- † Carried out after the event and documented in a two-page research article **submitted by June 28**



Software stack



+



+



 LLM4BeSci_Ljubljana2026     main  1 Branch  0 Tags Go to file

t

+

 Code

About

The course introduces the use of open large language models (LLMs) from the Hugging Face ecosystem for research in the behavioral and social sciences.

 Readme View license Activity 0 stars 0 watching 0 forks

Report repository

Releases

No releases published

[Create a new release](#)

 Zak-Hussain	slide upload note	a18ea0b · 2 days ago	 26 Commits
 day_1	use latest day_1.ipynb		last week
 day_2	typo		2 days ago
 day_3	typo		2 days ago
 day_4	typo		2 days ago
 day_5	update with newer justaism paper version		2 days ago
 .gitignore	Initial commit		last week
 LICENSE.txt	add LICENSE.txt		last week
 README.md	slide upload note		2 days ago
 cover_gserm.png	Add files via upload		last week
 notes.txt	add notes.txt		last week

Schedule

Please note, we will upload the presentation slides at end of the course.

Day 1

09:30 AM - 10:00 AM: Welcome & Intro

10:00 AM - 11:00 AM: [Talk: Intro to LLMs](#)

11:00 AM - 11:15 AM: Break

11:15 AM - 12:15 PM: Discussion: Find applications in small groups

12:15 PM - 01:15 PM: Lunch

01:15 PM - 02:00 PM: [Talk: A gentle intro to Hugging Face and Python](#)

02:00 PM - 02:15 PM: Setup Colab

02:15 PM - 02:30 PM: Break

02:30 PM - 03:00 PM: [Exercise: Running pipelines](#)

03:00 PM - 03:30 PM: Walkthrough



30s each

1. Who are you and what do you do?
2. How much experience do you have with machine learning and LLMs?
3. What motivates you to learn more about LLMs?

Intro LLMs

Dirk Wulff & Zak Hussain

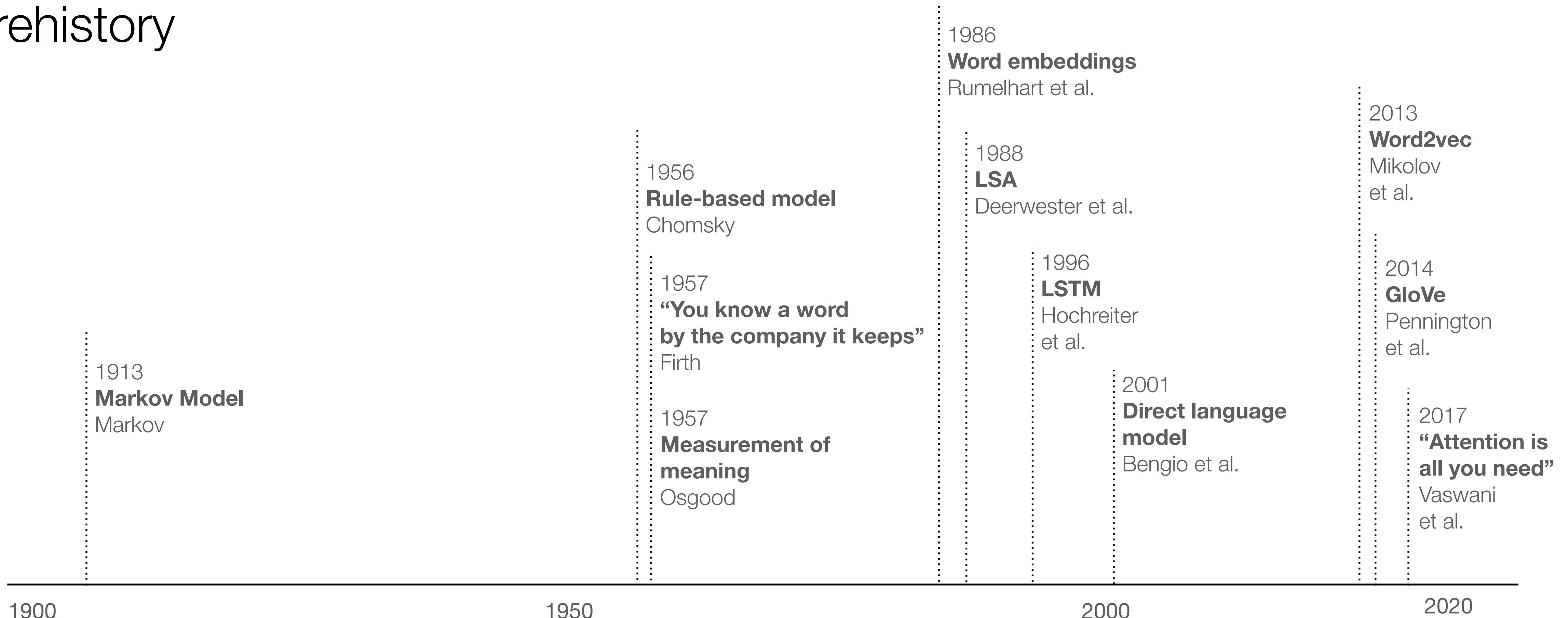


MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT



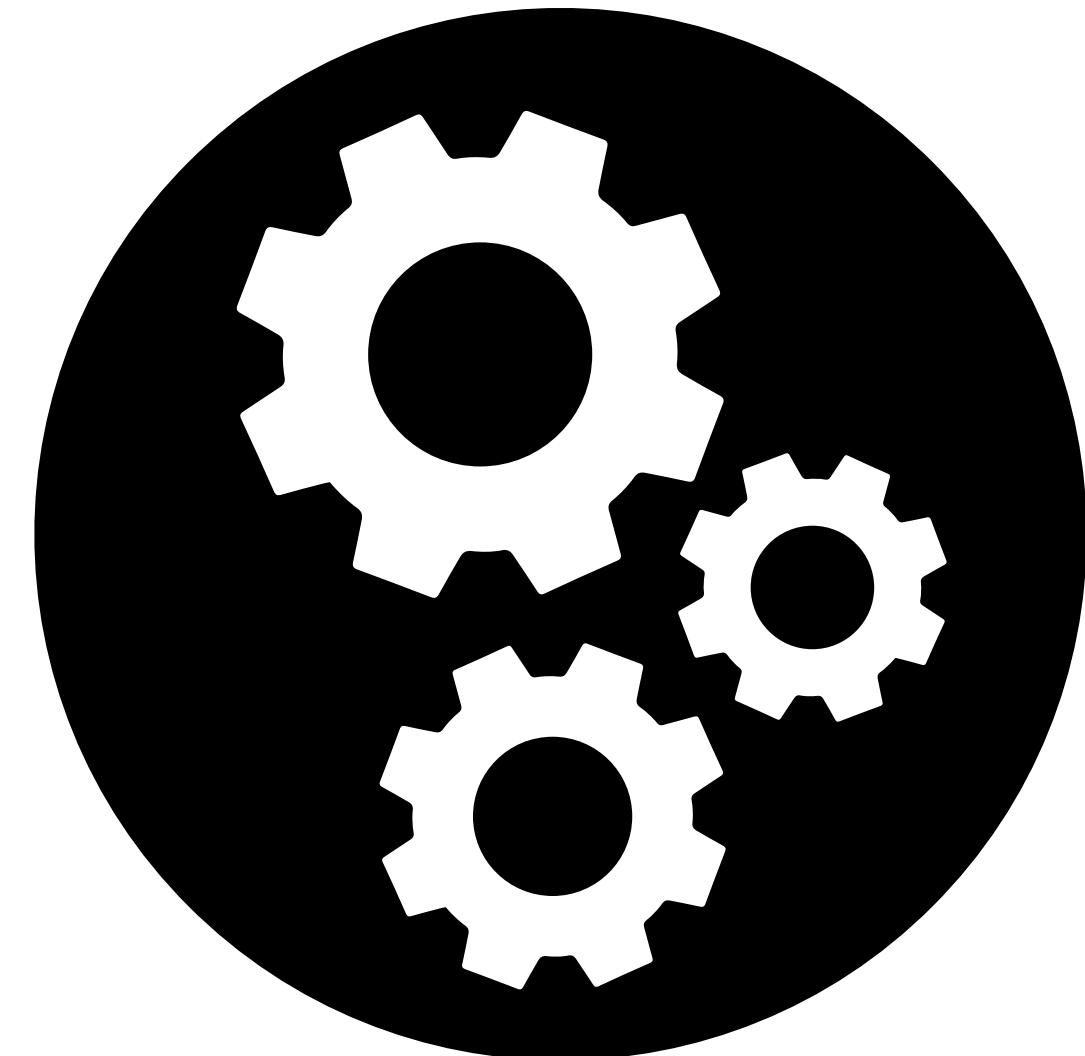
Language models

Prehistory

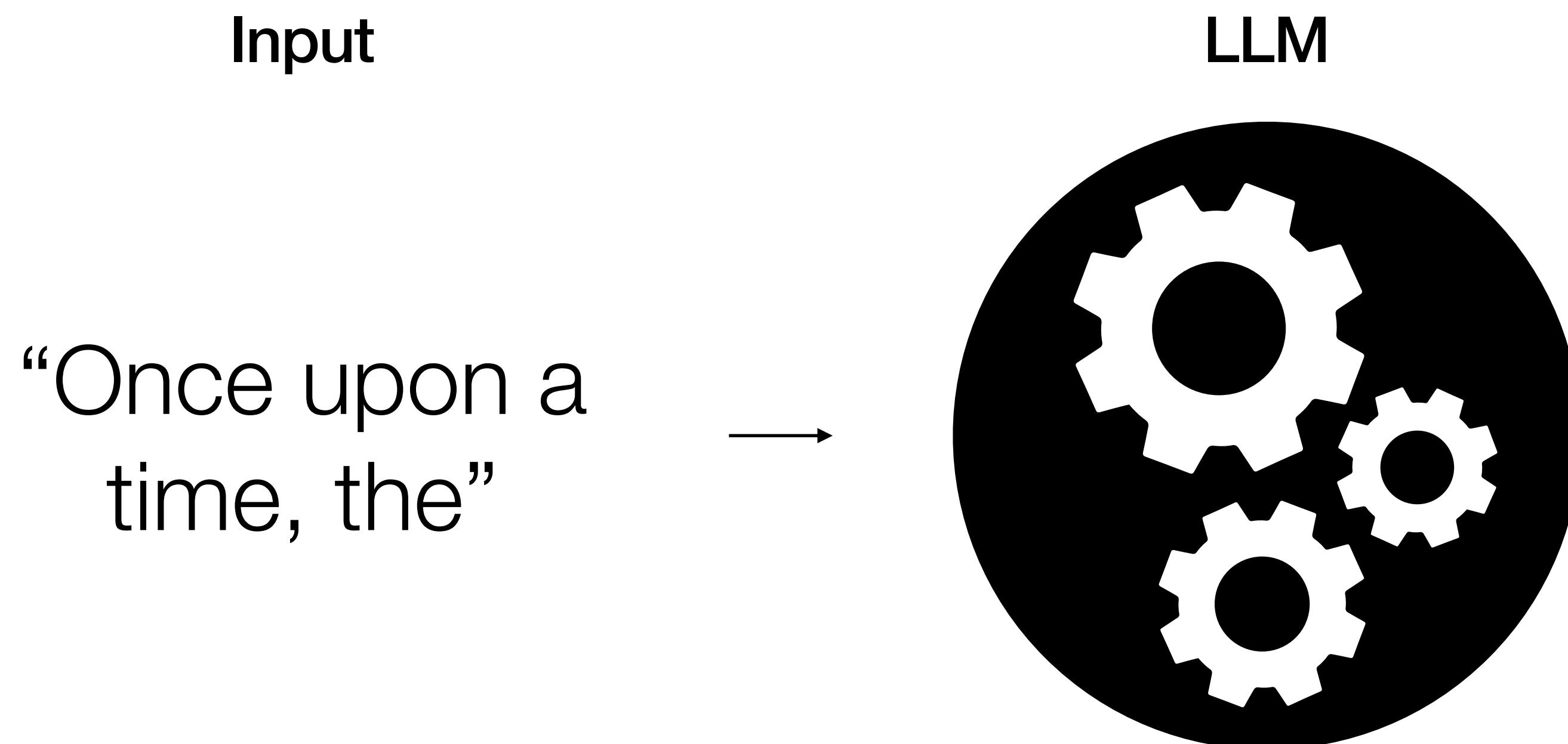


LLMs as mechanisms

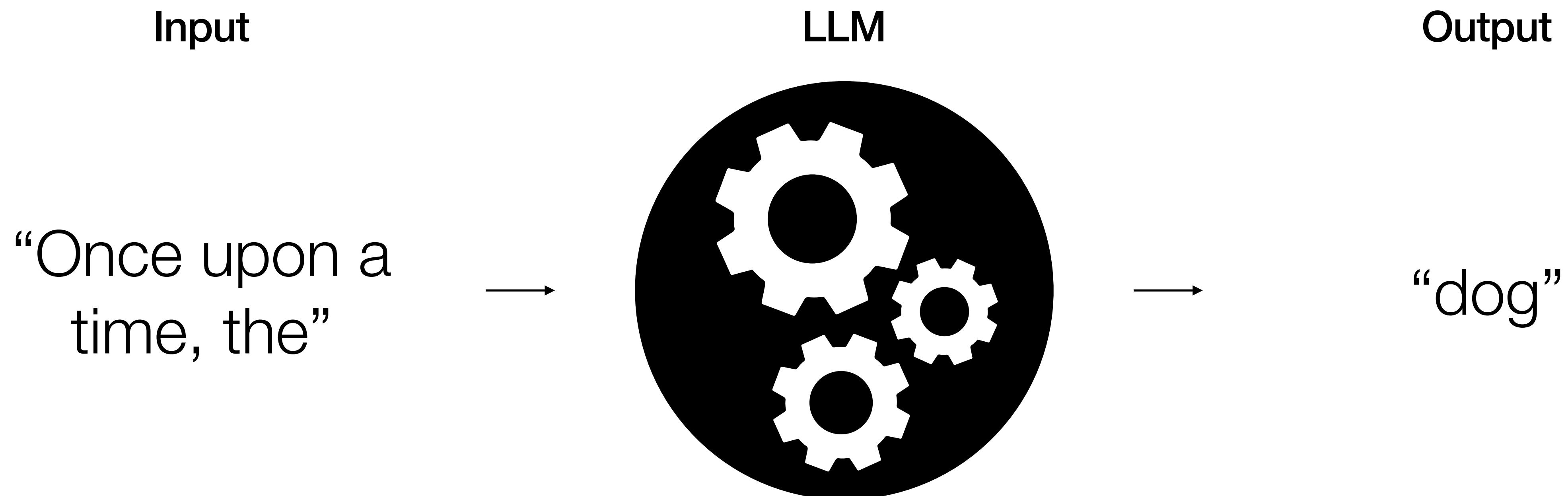
LLM



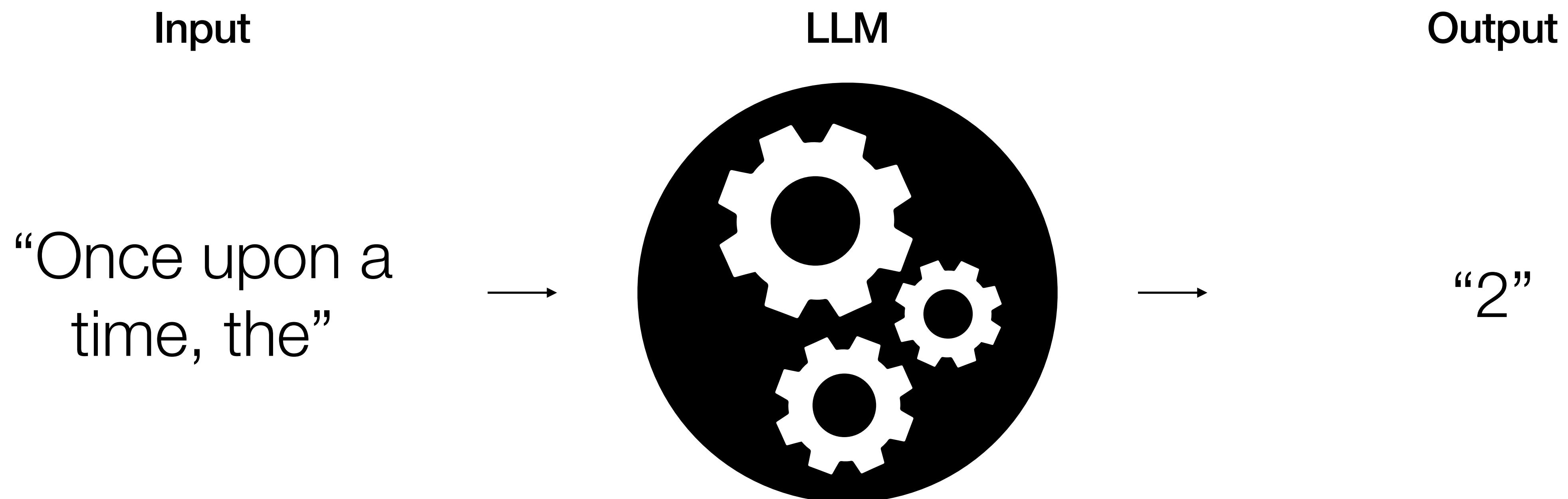
LLMs as mechanisms



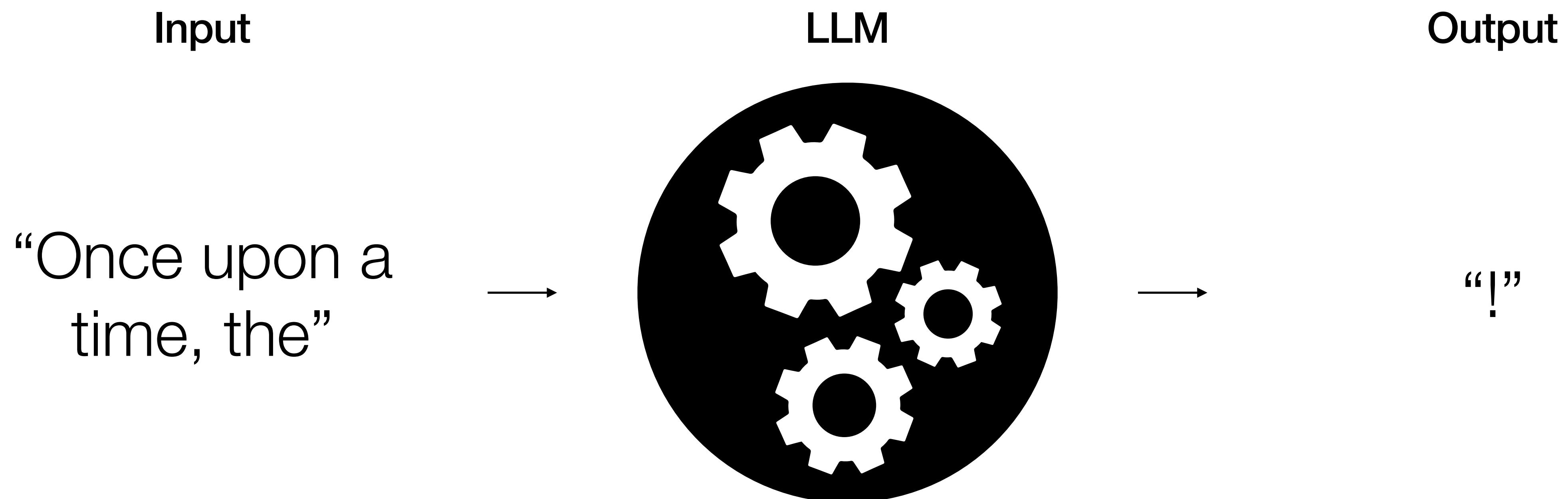
LLMs as mechanisms



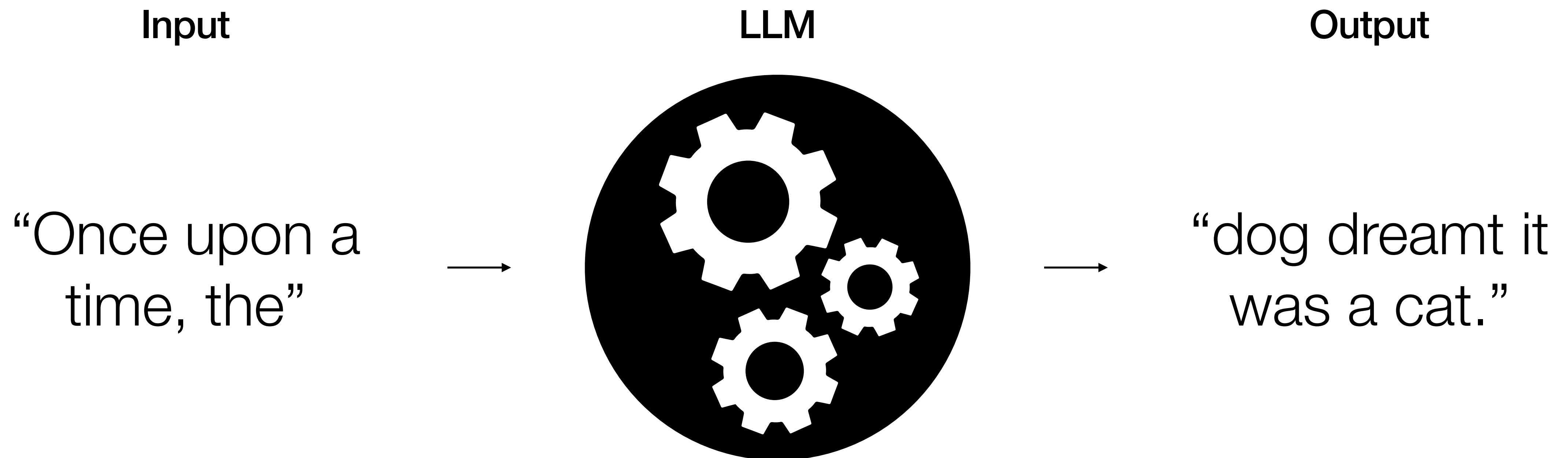
LLMs as mechanisms



LLMs as mechanisms



LLMs as mechanisms



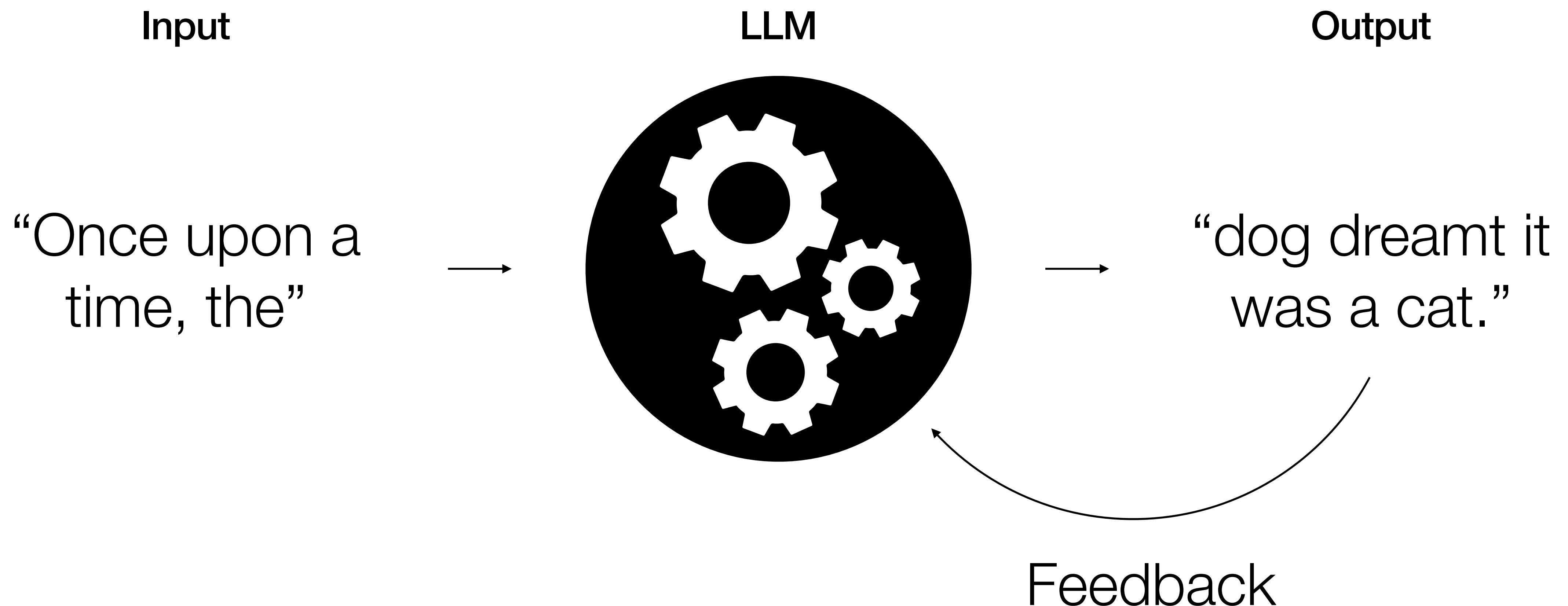
Once upon a time, the



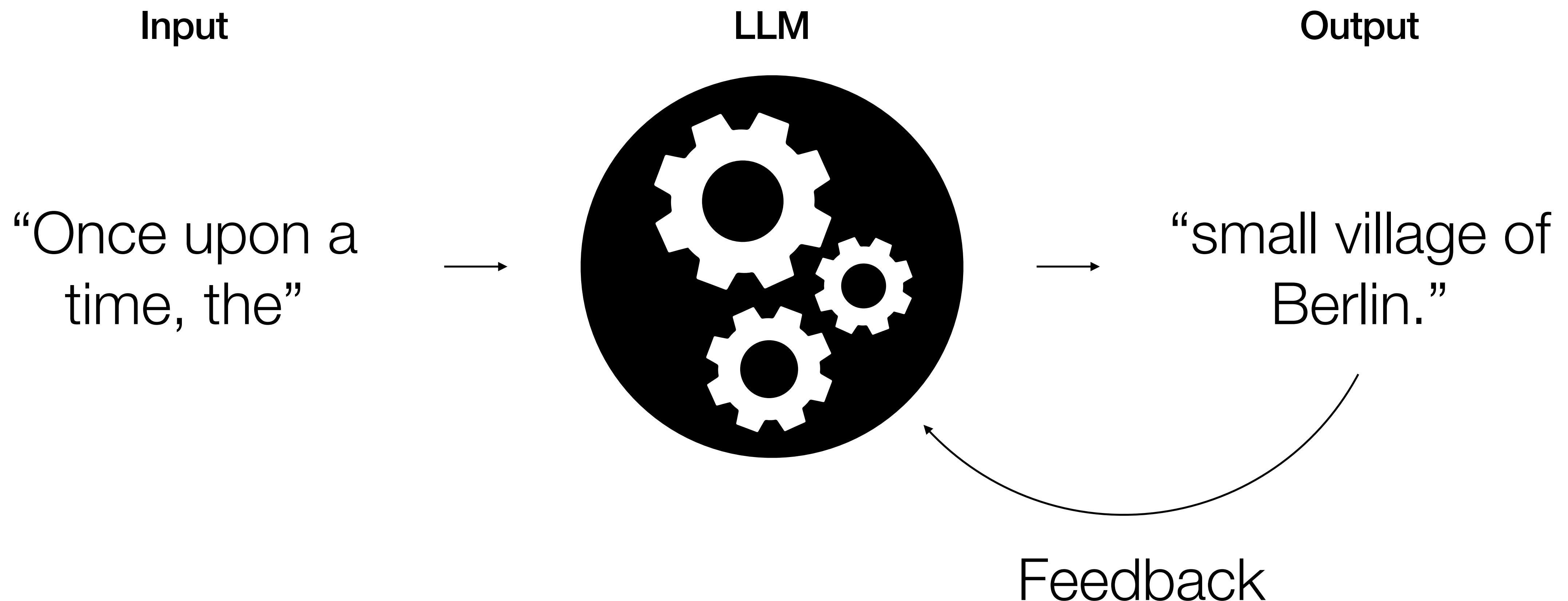
Once upon a time, the small village of Elmswood was nestled in a lush valley surrounded by towering mountains. The villagers lived peacefully, their days marked by the rhythms of nature and the changing seasons. However, everything changed when a mysterious old man arrived, carrying with him a locked chest that was said to contain a secret capable of altering the course of history. Intrigued by the stranger and his enigmatic treasure, the people of Elmswood soon found themselves on the brink of an adventure that would bind them together in ways they could never have imagined.



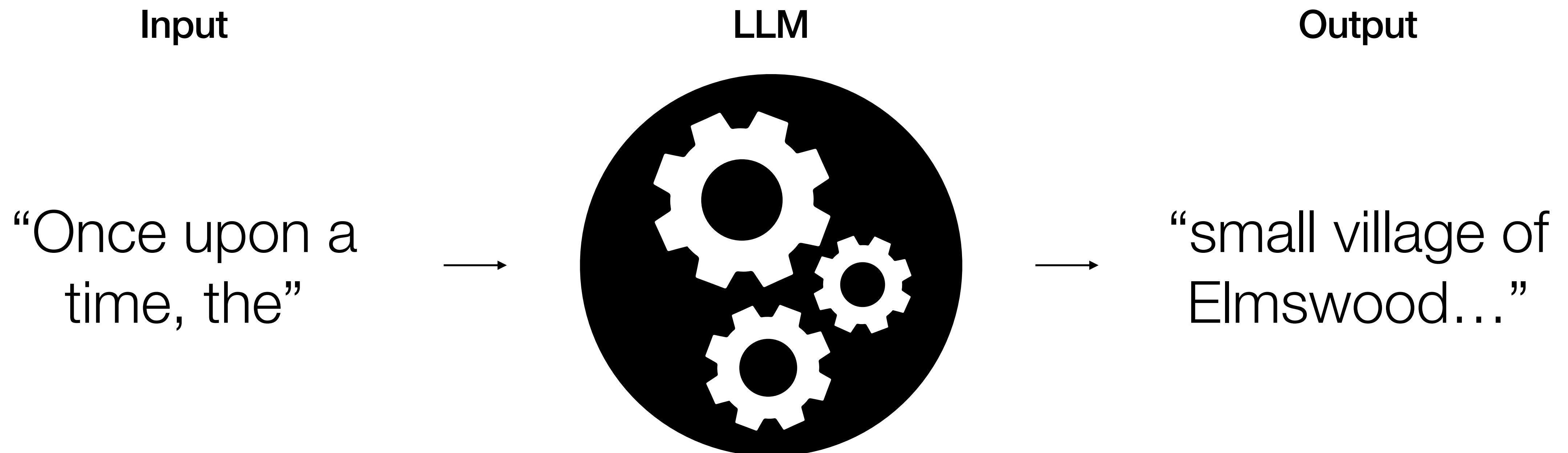
LLMs as **trained** mechanisms



LLMs as **trained** mechanisms



LLMs as **trained** mechanisms



Phi-3-mini-4K-Instruct and Phi-3-mini-128K-Instruct were trained over 7 days on 3.3T tokens using 512 H100-80G GPUs for each model. They followed advanced fine-tuning techniques to align with human preferences and safety standards.

The pre-training process followed two distinct and consecutive stages:

- In the first stage, the models were primarily exposed to a vast collection of web sources. This data helped the models develop general knowledge and language comprehension.
- In the second stage, the models were fine-tuned with a more rigorously selected subset of web data from the first phase, combined with additional synthetic data, to improve their logical reasoning and specialized abilities.

After these 2 stages, the models underwent additional training, which included supervised instruction fine-tuning and preference tuning, to enhance their stability and security.

The training dataset, made of 3.3 trillion tokens, is a meticulously curated mix of quality-filtered public documents, select educational materials, code, and newly generated synthetic data generated by LLMs. Specifically, the team filtered the web data to encompass the appropriate degree of knowledge and retained a greater number of web pages that may enhance the models' reasoning abilities. Instead of indiscriminately feeding vast amounts of data into the training model, the emphasis was placed on enhancing its reasoning capabilities, rather than one that merely has a vast repository of information.



Phi-3

LLM training = Pretraining + Fine-tuning

Trillions of tokens
Millions of power
consumption
Uses masked/next token
prediction

Hand-selected/crafted texts
Quality input-output pairs
Human feedback
Verifiable outcomes

Masked/next token prediction

"Once upon a time" is a [stock phrase](#) used to introduce a narrative of past events, typically in [fairy tales](#) and folk tales. It has been used in some form since at least 1380 (according to the [Oxford English Dictionary](#)) in [storytelling](#) in the [English language](#) and has started many narratives since 1600. These stories sometimes end with "and they all lived [happily ever after](#)", or, originally, "happily until their deaths".

The phrase is common in [fairy tales](#) for younger children. It was used in the original translations of the stories of [Charles Perrault](#) as a translation for the [French](#) "*il était une fois*", of [Hans Christian Andersen](#) as a translation for the [Danish](#) "*der var engang*" (literally "there was once"), the [Brothers Grimm](#) as a translation for the [German](#) "*es war einmal*" (literally "it was once") and [Joseph Jacobs](#) in [English](#) translations and fairy tales.

In *More English Fairy Tales*, Joseph Jacobs notes that:

"The opening formulae are varied enough, but none of them has much play of fancy. 'Once upon a time and a very good time it was, though it wasn't in my time nor in your time nor in any one else's time.' is effective enough for a fairy epoch, and is common, according to Mayhew (*London Labour*, III), among tramps."^[1]

https://en.wikipedia.org/wiki/Once_upon_a_time

LLM



Masked/next token prediction

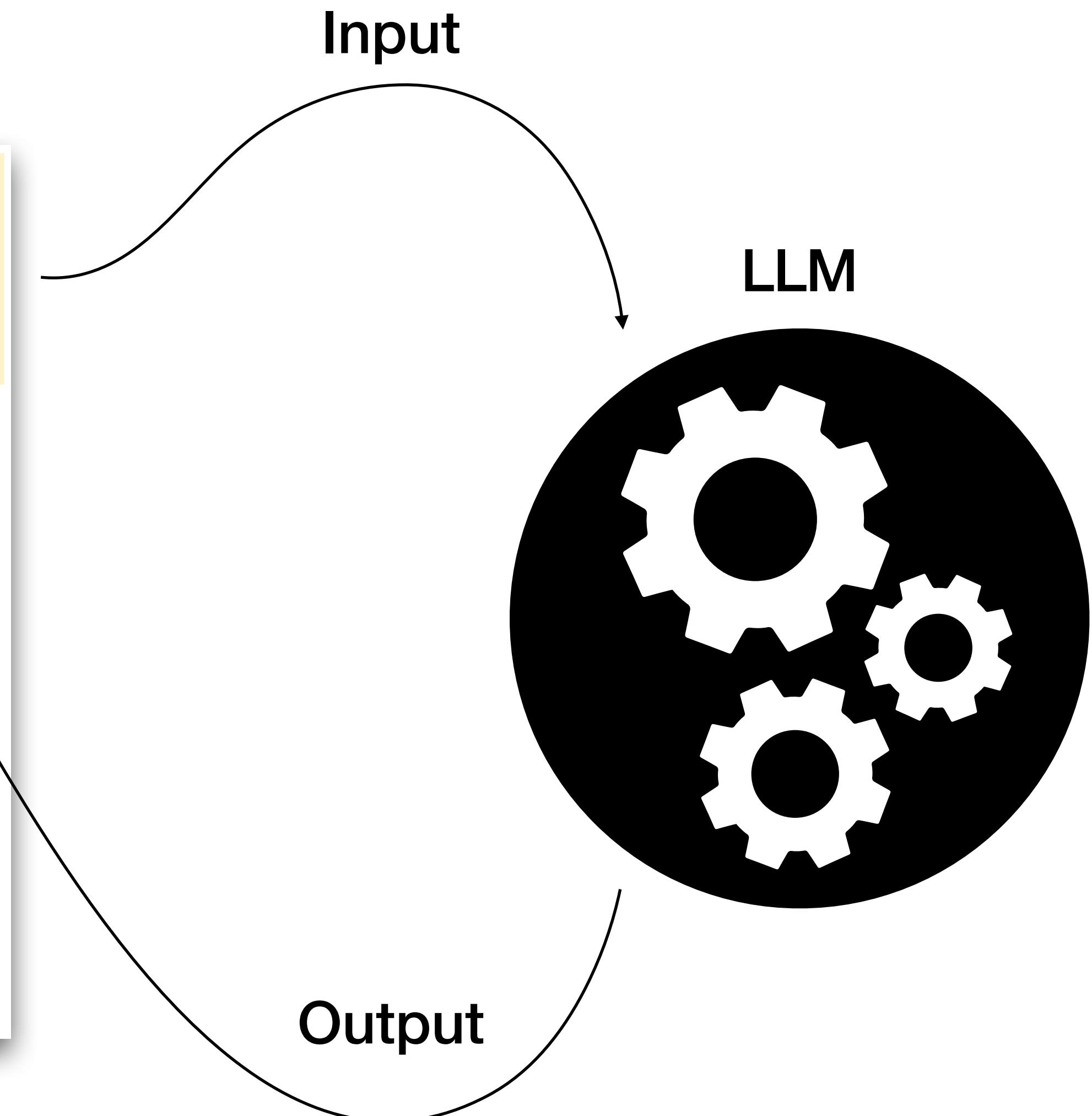
"Once upon a time" is a stock phrase used to introduce a narrative of past events, typically in fairy tales and folk tales. It has been used in some form since at least 1380 (according to the *Oxford English Dictionary*) in [REDACTED] in the English language and has started many narratives since 1600. These stories sometimes end with "and they all lived happily ever after", or, originally, "happily until their deaths".

The phrase is common in fairy tales for younger children. It was used in the original translations of the stories of Charles Perrault as a translation for the French "il était une fois", of Hans Christian Andersen as a translation for the Danish "der var engang" (literally "there was once"), the Brothers Grimm as a translation for the German "es war einmal" (literally "it was once") and Joseph Jacobs in English translations and fairy tales.

In *More English Fairy Tales*, Joseph Jacobs notes that:

"The opening formulae are varied enough, but none of them has much play of fancy. 'Once upon a time and a very good time it was, though it wasn't in my time nor in your time nor in any one else's time.' is effective enough for a fairy epoch, and is common, according to Mayhew (*London Labour*, III), among tramps."^[1]

https://en.wikipedia.org/wiki/Once_upon_a_time



Masked/next token prediction

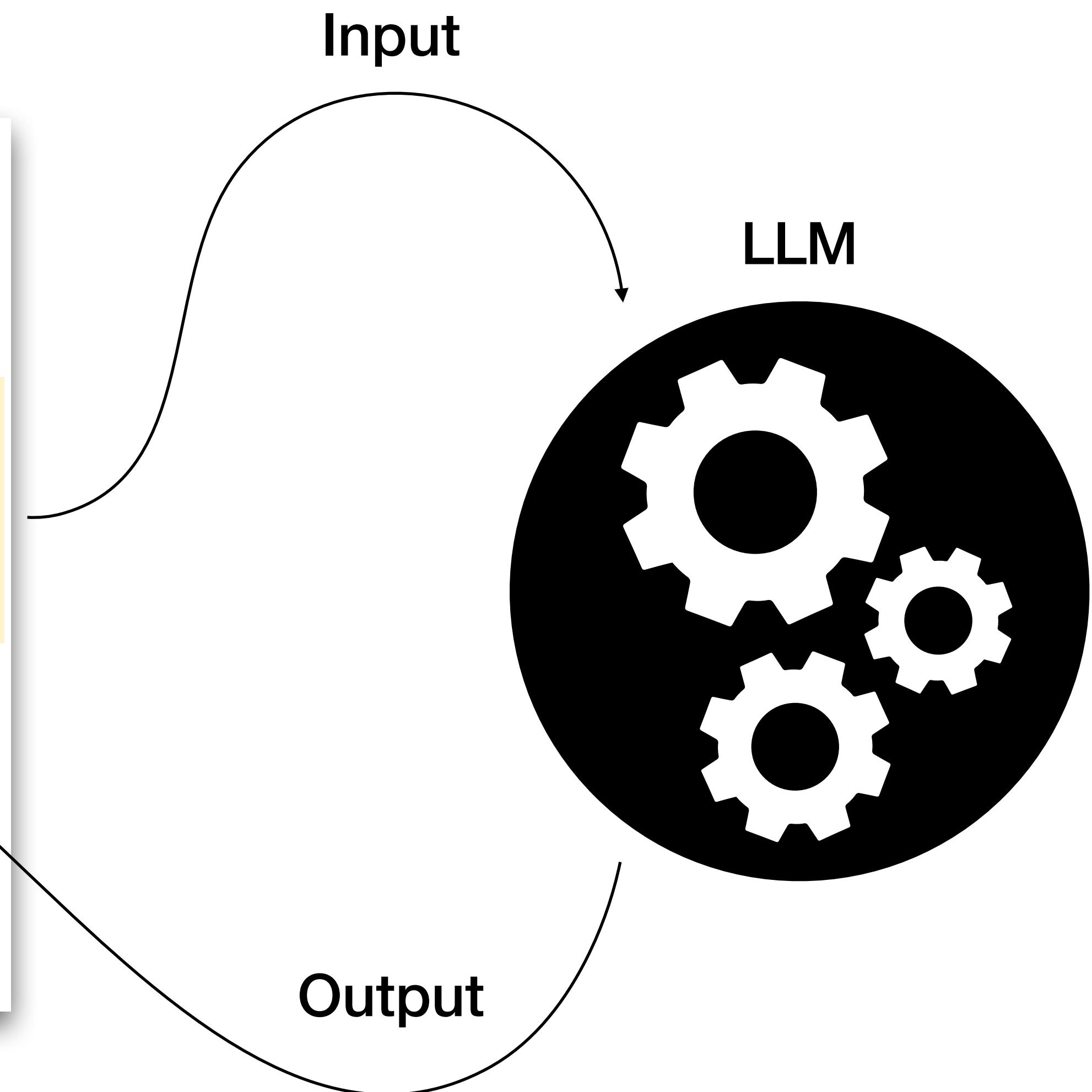
"Once upon a time" is a [stock phrase](#) used to introduce a narrative of past events, typically in [fairy tales](#) and folk tales. It has been used in some form since at least 1380 (according to the [Oxford English Dictionary](#)) in [storytelling](#) in the [English language](#) and has started many narratives since 1600. These stories sometimes end with "and they all lived [happily ever after](#)", or, originally, "happily until their deaths".

The phrase is common in [fairy tales](#) for younger children. It was used in the original translations of the stories of [Charles Perrault](#) as a translation for the [French](#) "*il était une fois*", of [Hans Christian Andersen](#) as a translation for the [Danish](#) "*der var engang*" (literally "there was once"), the [Brothers Grimm](#) as a translation for the [German](#) "*es war einmal*" (literally "it was once") and [Joseph Jacobs](#) in [English](#) translations and fairy tales.

In *More English Fairy Tales*, Joseph Jacobs notes that:

"The opening formulae are varied enough, but none of them has much play of fancy. 'Once upon a time and a very good time it was, though it wasn't in my time nor in your time nor in any one else's time.' is effective enough for a fairy epoch, and is common, according to Mayhew (*London Labour, III*), among tramps."^[1]

https://en.wikipedia.org/wiki/Once_upon_a_time



Masked/next token prediction

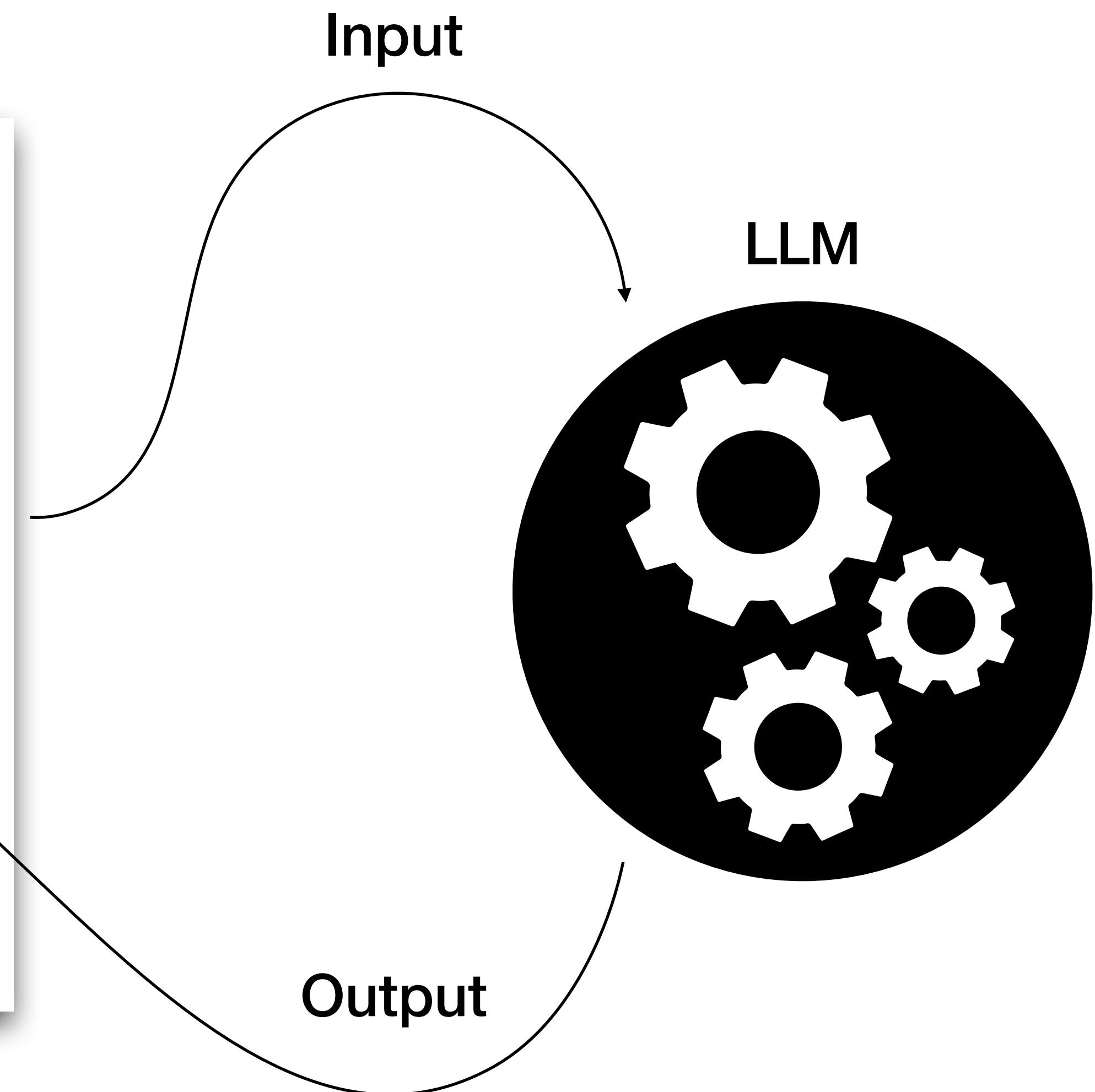
"Once upon a time" is a [stock phrase](#) used to introduce a narrative of past events, typically in [fairy tales](#) and folk tales. It has been used in some form since at least 1380 (according to the [Oxford English Dictionary](#)) in [storytelling](#) in the [English language](#) and has started many narratives since 1600. These stories sometimes end with "and they all lived [happily ever after](#)", or, originally, "happily until their deaths".

The phrase is common in [fairy tales](#) for younger children. It was used in the original translations of the stories of [Charles Perrault](#) as a translation for the [French](#) "*il était une fois*", of [Hans Christian Andersen](#) as a translation for the [Danish](#) "*der var engang*" (literally "there was once"), the [Brothers Grimm](#) as a translation for the [German](#) "*es war einmal*" (literally "it was once") and [Joseph Jacobs](#) in [English](#) translations and fairy tales.

In *More English Fairy Tales*, Joseph Jacobs notes that:

"The opening formulae are varied enough, but none of them has much play of fancy. 'Once upon a time and a very good time it was, though it wasn't in my time nor in your time nor in any one else's time.' is effective enough for a fairy epoch, and is common, according to Mayhew (*London Labour, III*), among tramps."^[1]

https://en.wikipedia.org/wiki/Once_upon_a_time



Masked/next token prediction

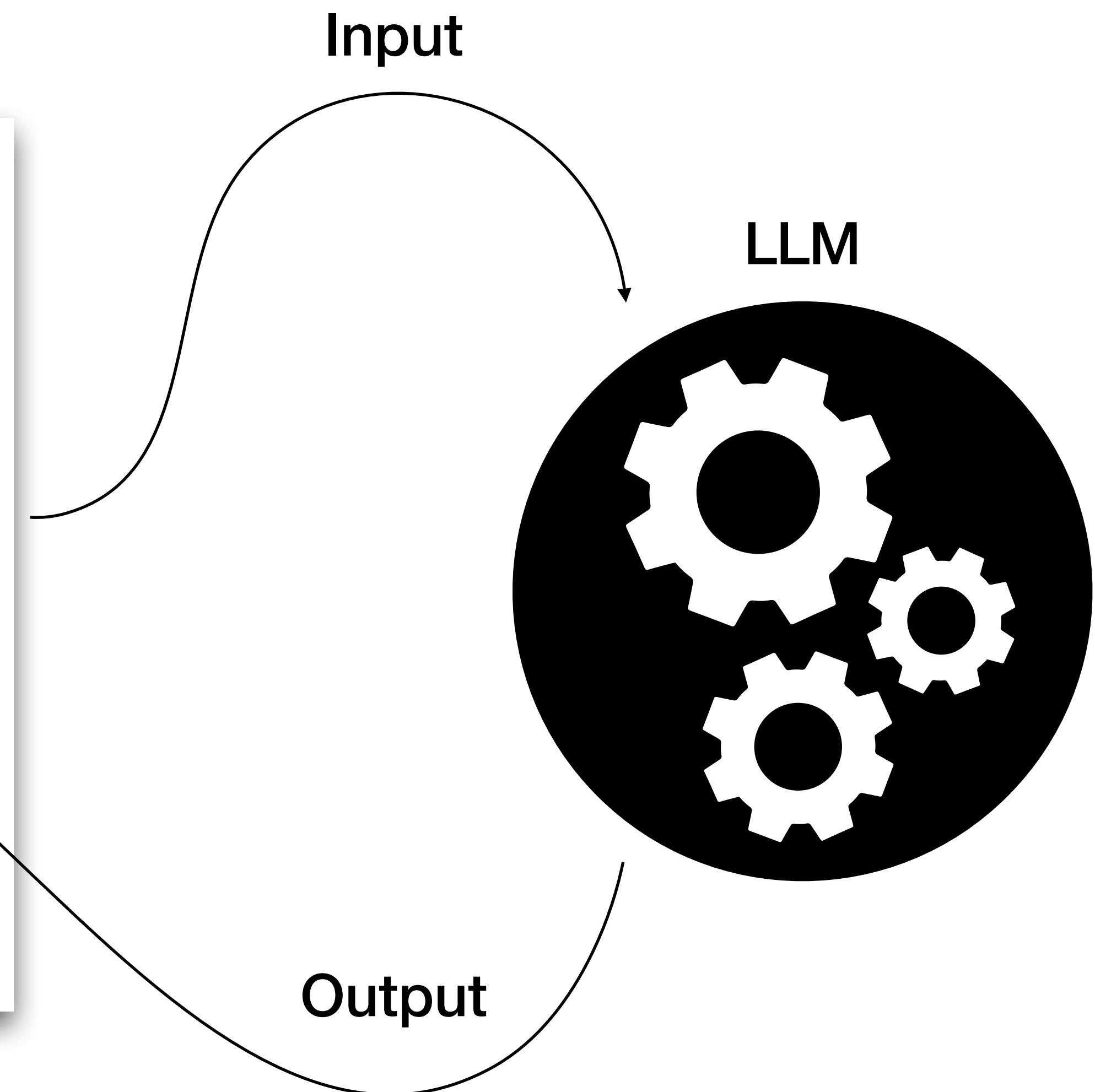
"Once upon a time" is a [stock phrase](#) used to introduce a narrative of past events, typically in [fairy tales](#) and folk tales. It has been used in some form since at least 1380 (according to the [Oxford English Dictionary](#)) in [storytelling](#) in the [English language](#) and has started many narratives since 1600. These stories sometimes end with "and they all lived [happily ever after](#)", or, originally, "happily until their deaths".

The phrase is common in [fairy tales](#) for younger children. It was used in the original translations of the stories of [Charles Perrault](#) as a translation for the [French](#) "*il était une fois*", of [Hans Christian Andersen](#) as a translation for the [Danish](#) "*der var engang*" (literally "there was once"), the [Brothers Grimm](#)  a translation for the [German](#) "*es war einmal*" (literally "it was once") and [Joseph Jacobs](#) in [English](#) translations and fairy tales.

In *More English Fairy Tales*, Joseph Jacobs notes that:

"The opening formulae are varied enough, but none of them has much play of fancy. 'Once upon a time and a very good time it was, though it wasn't in my time nor in your time nor in any one else's time.' is effective enough for a fairy epoch, and is common, according to Mayhew (*London Labour, III*), among tramps."^[1]

https://en.wikipedia.org/wiki/Once_upon_a_time



Masked/next token prediction

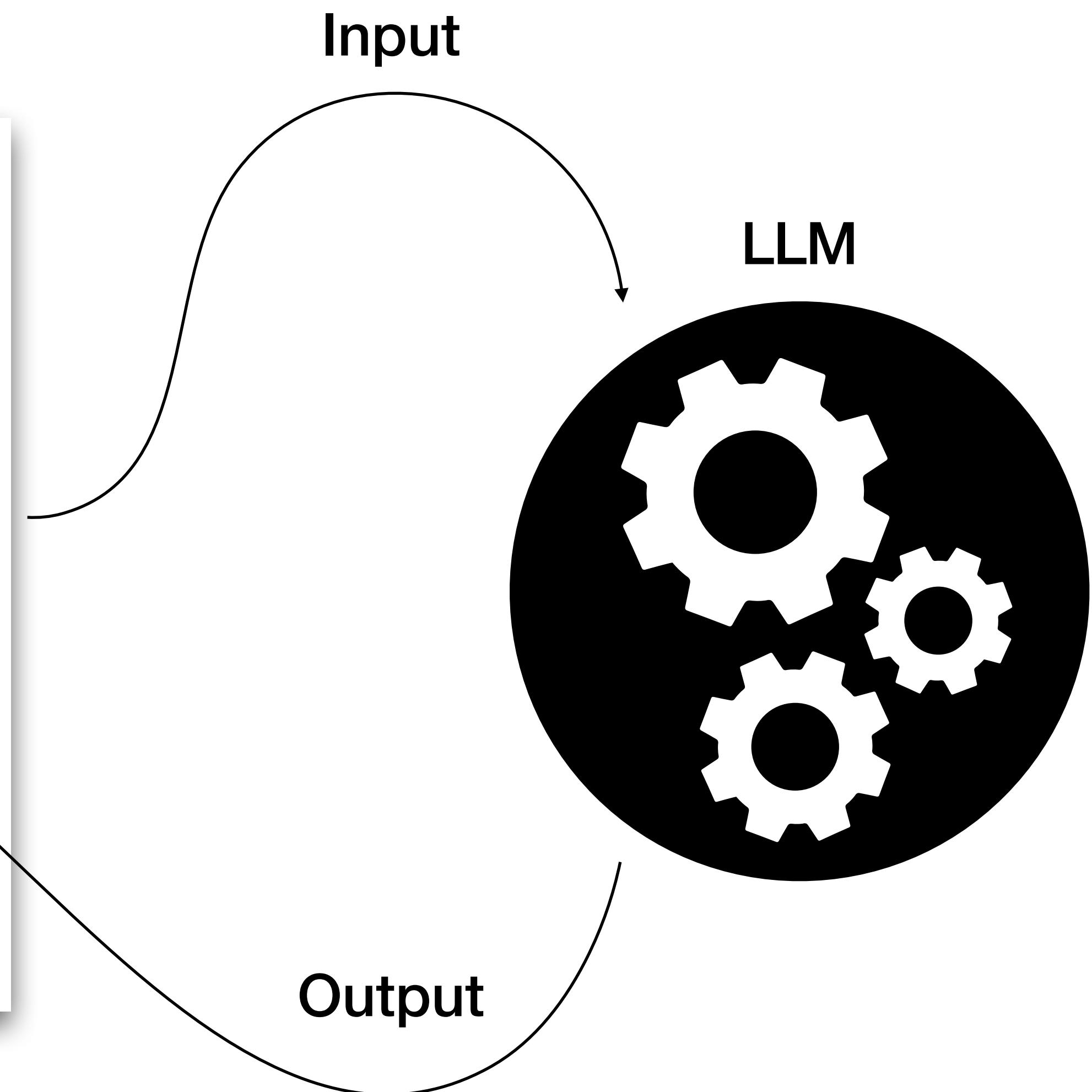
"Once upon a time" is a [stock phrase](#) used to introduce a narrative of past events, typically in [fairy tales](#) and folk tales. It has been used in some form since at least 1380 (according to the [Oxford English Dictionary](#)) in [storytelling](#) in the [English language](#) and has started many narratives since 1600. These stories sometimes end with "and they all lived [happily ever after](#)", or, originally, "happily until their deaths".

The phrase is common in [fairy tales](#) for younger children. It was used in the original translations of the stories of [Charles Perrault](#) as a translation for the [French](#) "*il était une fois*", of [Hans Christian Andersen](#) as a translation for the [Danish](#) "*der var engang*" (literally "there was once"), the [Brothers Grimm](#)  a translation for the [German](#) "*es war einmal*" (literally "it was once") and [Joseph Jacobs](#) in [English](#) translations and fairy tales.

In *More English Fairy Tales*, Joseph Jacobs notes that:

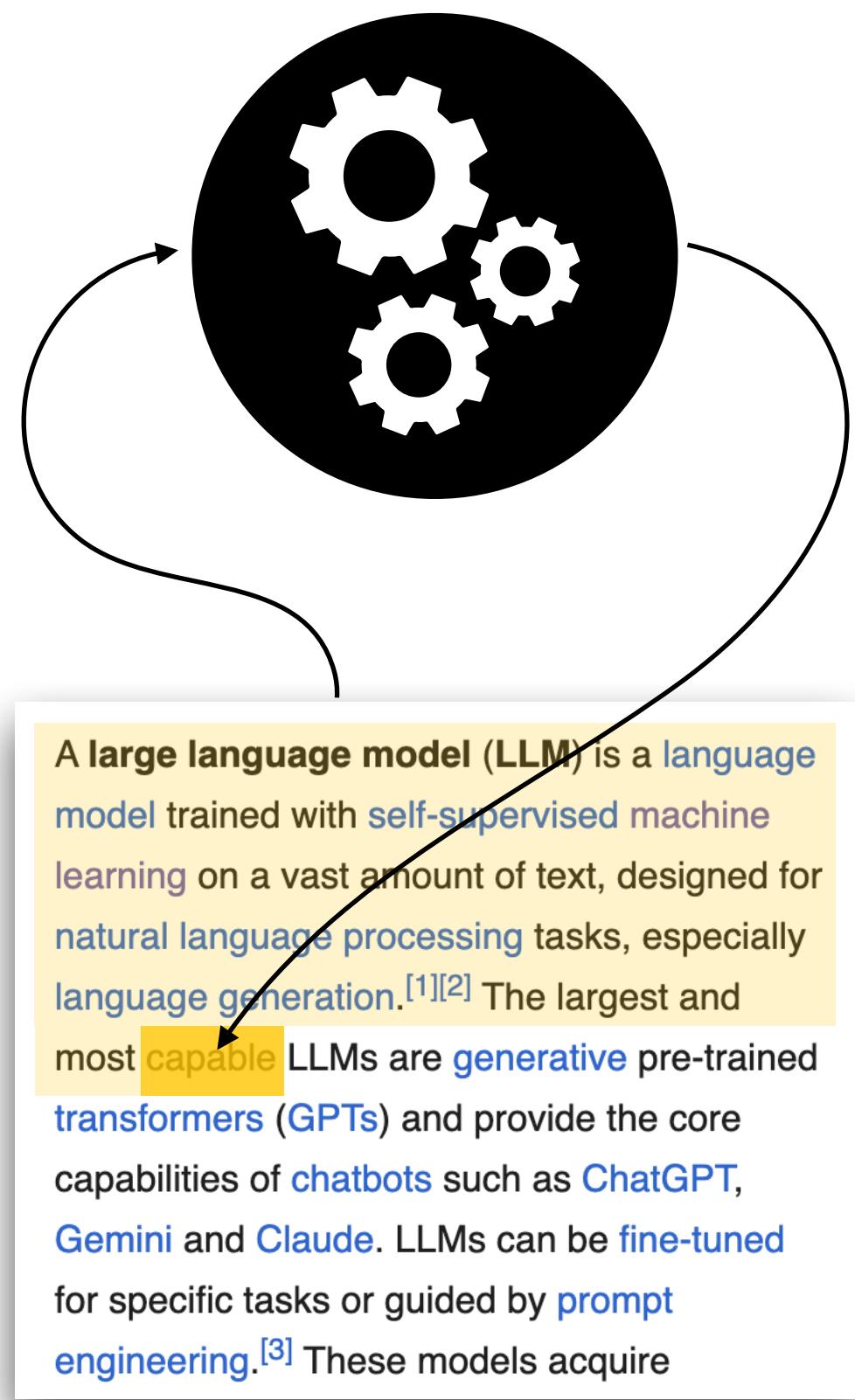
"The opening formulae are varied enough, but none of them has much play of fancy. 'Once upon a time and a very good time it was, though it wasn't in my time nor in your time nor in any one else's time.' is effective enough for a fairy epoch, and is common, according to Mayhew (*London Labour, III*), among tramps."^[1]

https://en.wikipedia.org/wiki/Once_upon_a_time

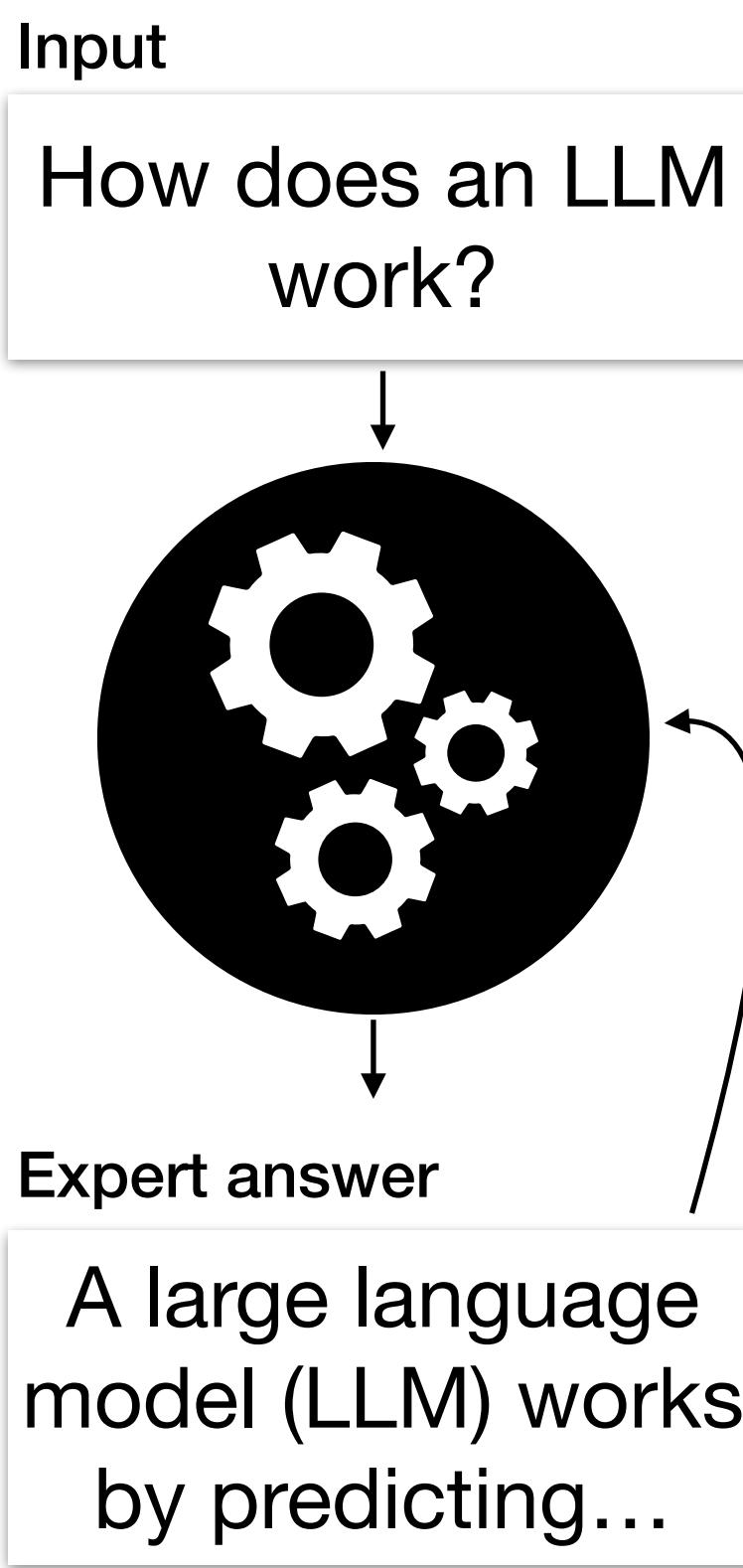


Training

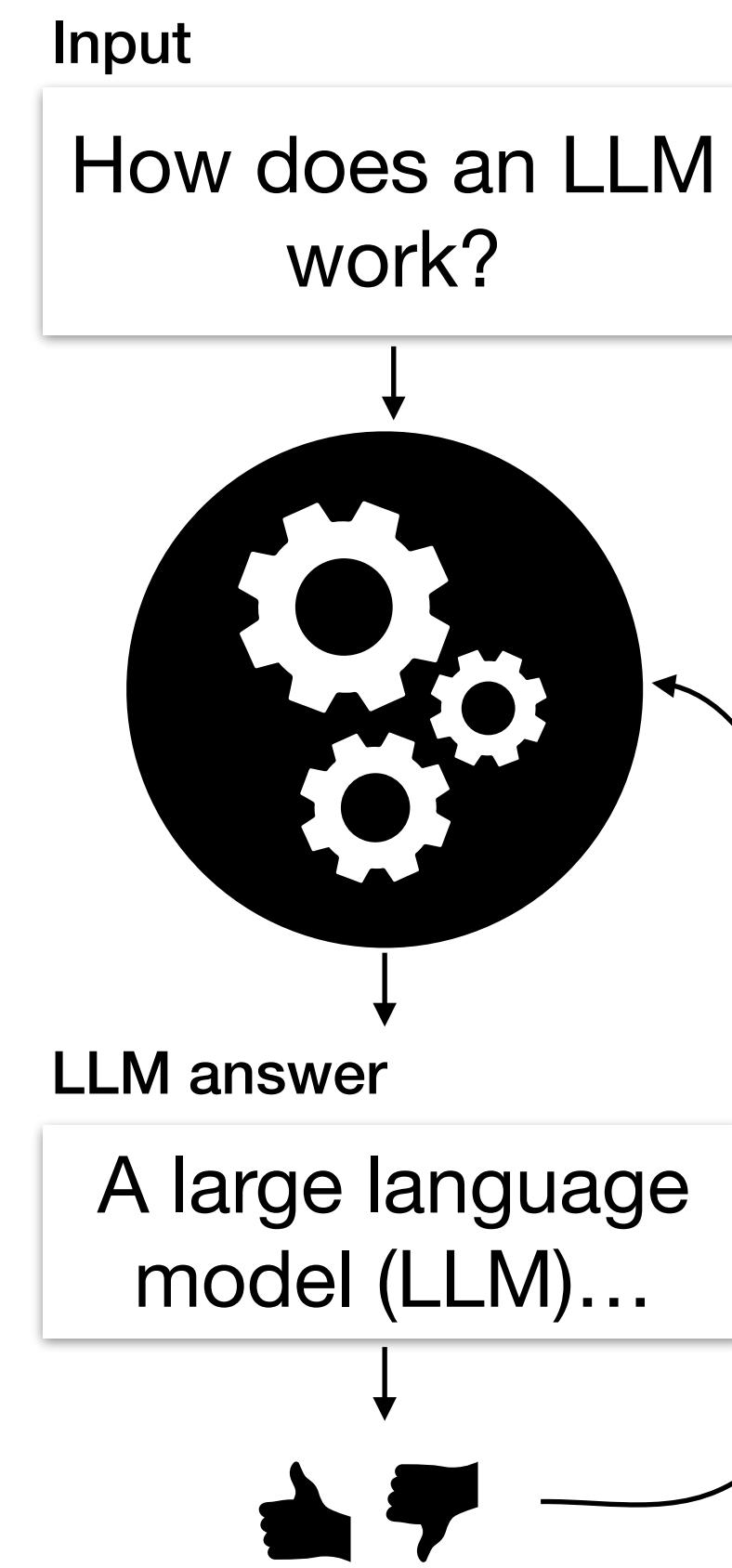
Pretraining



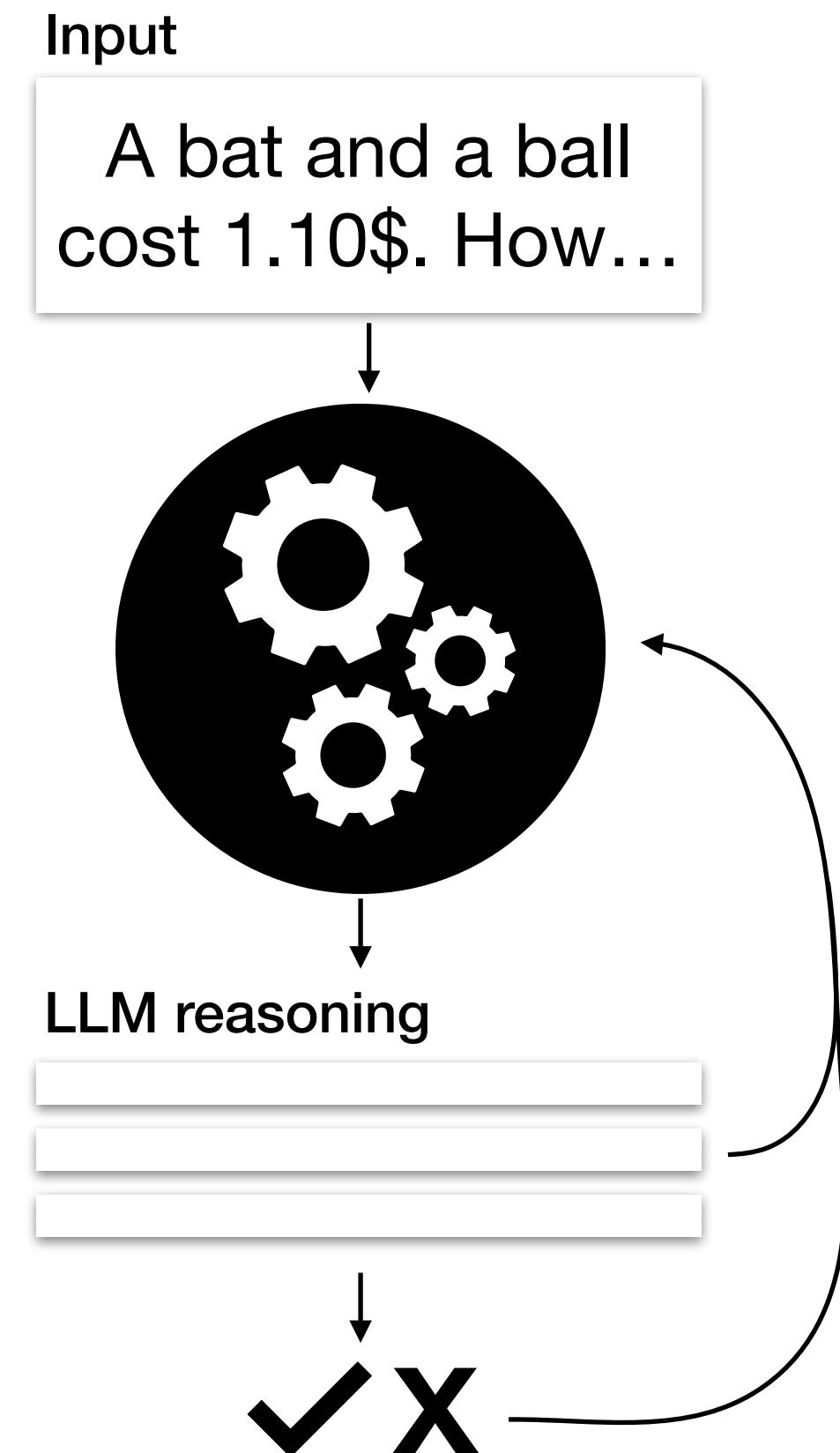
Instruction tuning



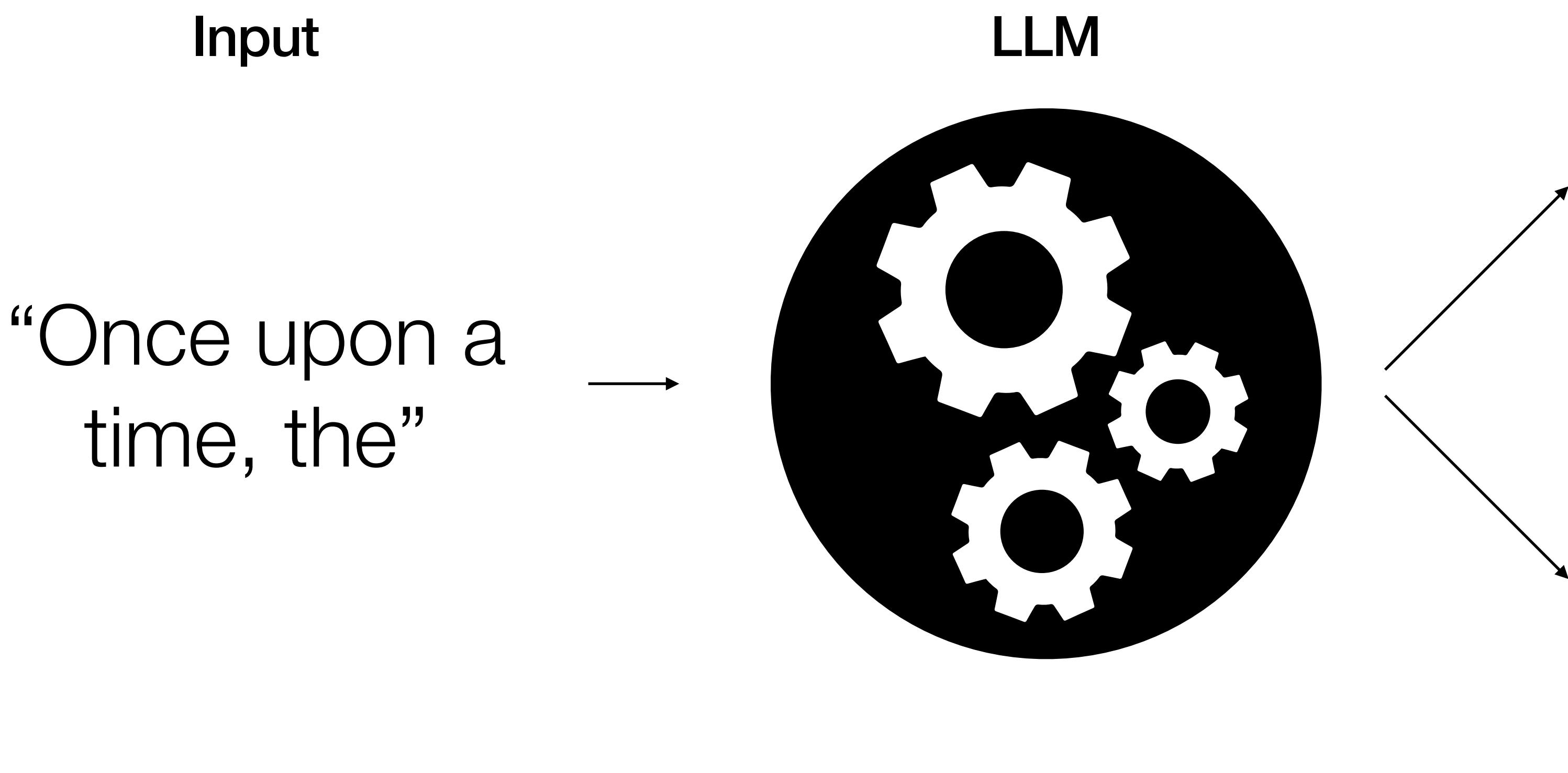
Preference tuning



Reasoning tuning



Two major applications



Text generation

“small village of
Elmswood...”

Feature extraction

- .23, 1.23, .24,
- .12, .34, .32, ...

Closed vs. Open

Public use

Performance

Safety

Reproducible

Data secure

Interpretable



Tasks Libraries Datasets Languages Licenses Other

Filter Tasks by name

Multimodal

Audio-Text-to-Text Image-Text-to-Text
Visual Question Answering Document Question Answering Video-Text-to-Text
Visual Document Retrieval Any-to-Any

Computer Vision

Depth Estimation Image Classification
Object Detection Image Segmentation
Text-to-Image Image-to-Text Image-to-Image
Image-to-Video Unconditional Image Generation
Video Classification Text-to-Video
Zero-Shot Image Classification Mask Generation
Zero-Shot Object Detection Text-to-3D
Image-to-3D Image Feature Extraction
Keypoint Detection

Natural Language Processing

Text Classification Token Classification
Table Question Answering Question Answering
Zero-Shot Classification Translation
Summarization Feature Extraction

Models 1,680,396

Filter by name

Full-text search

↑↓ Sort: Trending

nvidia/parakeet-tdt-0.6b-v2
Automatic Speech Recognition • Updated 10 days ago • ↓ 72.7k • ❤ 691

ACE-Step/ACE-Step-v1-3.5B
Text-to-Audio • Updated 1 day ago • ❤ 310

nari-labs/Dia-1.6B
Text-to-Speech • Updated 5 days ago • ↓ 148k • ⚡ • ❤ 2.04k

Lightricks/LTX-Video
Text-to-Video • Updated 5 days ago • ↓ 214k • ⚡ • ❤ 1.38k

JetBrains/Mellum-4b-base
Text Generation • Updated 4 days ago • ↓ 2.81k • ❤ 311

Qwen/Qwen3-235B-A22B
Text Generation • Updated 10 days ago • ↓ 82.4k • ⚡ • ❤ 764

deepseek-ai/DeepSeek-Prover-V2-671B
Text Generation • Updated 11 days ago • ↓ 7.38k • ⚡ • ❤ 750

lodestones/Chroma
Text-to-Image • Updated 1 day ago • ❤ 401

black-forest-labs/FLUX.1-dev
Text-to-Image • Updated Aug 16, 2024 • ↓ 2.66M • ⚡ • ❤ 10.1k

cognition-ai/Kevin-32B
Updated 5 days ago • ↓ 342 • ❤ 93

Qwen/Qwen3-30B-A3B
Text Generation • Updated 11 days ago • ↓ 150k • ⚡ • ❤ 514

microsoft/Phi-4-reasoning-plus
Text Generation • Updated 3 days ago • ↓ 10.8k • ❤ 240

tencent/HunyuanCustom
Image-to-Video • Updated 2 days ago • ❤ 77

Tesslate/UIGEN-T2-7B-Q8_0-GGUF
Text Generation • Updated 5 days ago • ↓ 2.59k • ❤ 115

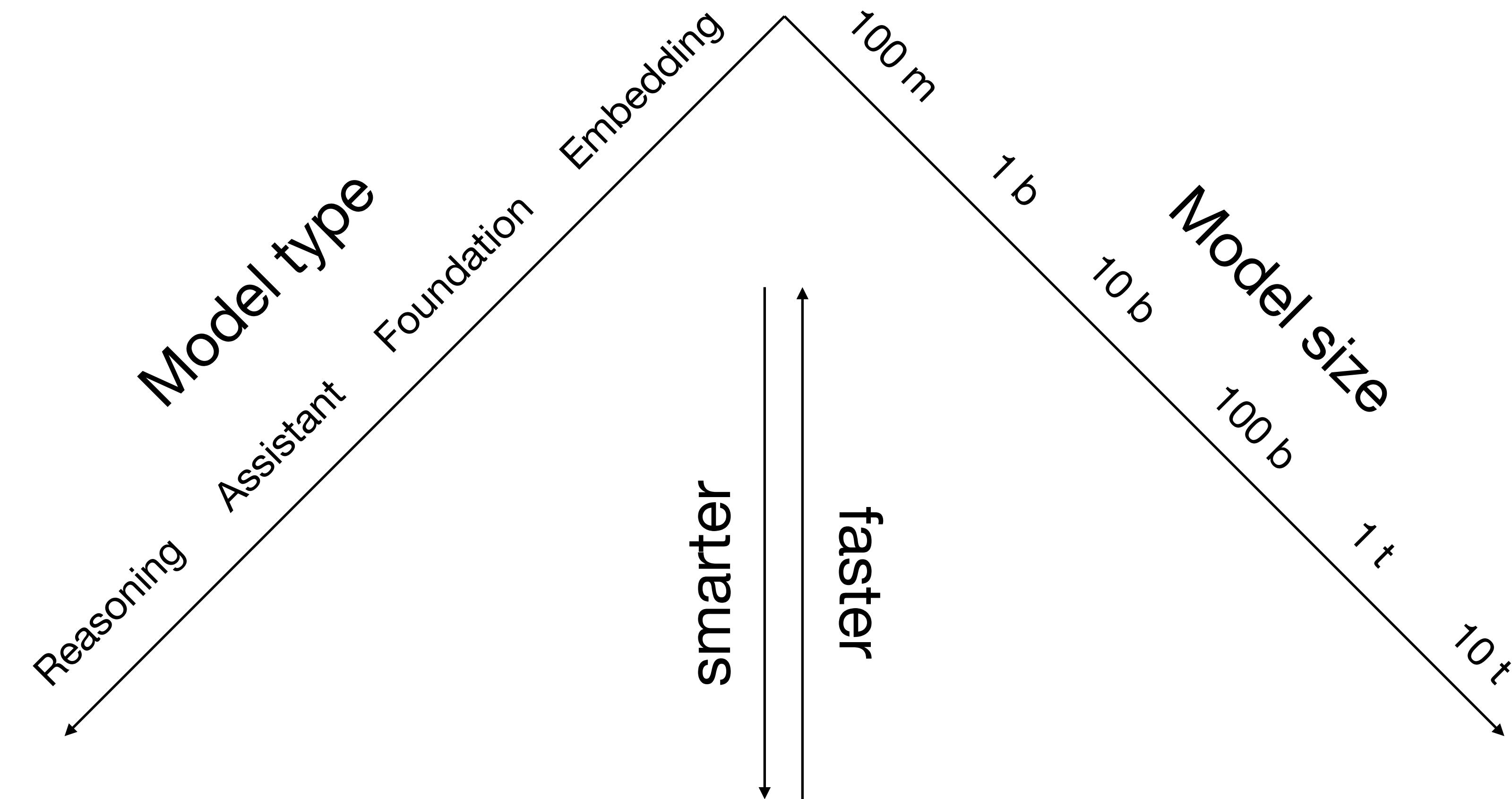
hexgrad/Kokoro-82M
Text-to-Speech • Updated about 1 month ago • ↓ 1.88M • ⚡ • ❤ 4.27k

fdtn-ai/Foundation-Sec-8B
Text Generation • Updated 10 days ago • ↓ 21.5k • ❤ 145

ServiceNow-AI/AprielandNemotron-15b-Thinker
Updated 4 days ago • ↓ 784 • ❤ 66

Qwen/Qwen3-8B
Text Generation • Updated 12 days ago • ↓ 394k • ❤ 262

Distinguishing LLMs



Project (maker, bases, URL)	Availability						Documentation						Access		
	Open code	LLM data	LLM weights	RL data	RL weights	License	Code	Architecture	Preprint	Paper	Modelcard	Datasheet	Package	API	
OLMo 7B Instruct Ai2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	~
BLOOMZ bigscience-workshop	✓	✓	✓	✓	~	~	✓	✓	✓	✓	✓	✓	✗	✓	12.0
AmberChat LLM360	✓	✓	✓	✓	✓	✓	~	~	✓	✗	~	~	✗	✓	10.0
Open Assistant LAION-AI	✓	✓	✓	✓	✗	✓	✓	✓	~	✗	✗	✗	✓	✓	9.5
...															
Command R+ Cohere AI	✗	✗	✗	✓	✓	~	✗	✗	✗	✗	~	✗	✗	✗	3.0
LLaMA2 Chat Facebook Research	✗	✗	~	✗	~	✗	✗	~	~	✗	~	✗	✗	✗	3.0
Nanbeige2-Chat Nanbeige LLM lab	✓	✗	✗	✗	✓	~	✗	✗	✗	✗	✗	✗	✗	~	3.0
Llama 3 Instruct Facebook Research	✗	✗	~	✗	~	✗	✗	~	✗	✗	~	✗	✗	~	2.5
Solar 70B Upstage AI	✗	✗	~	✗	~	✗	✗	✗	✗	✗	~	✗	✗	~	2.0
Xwin-LM Xwin-LM	✗	✗	~	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	~	1.0
ChatGPT OpenAI	✗	✗	✗	✗	✗	✗	✗	✗	~	✗	✗	✗	✗	✗	0.5

see also <https://osai-index.eu/the-index>

LMSYS Chatbot Arena Leaderboard

| [Vote](#) | [Blog](#) | [GitHub](#) | [Paper](#) | [Dataset](#) | [Twitter](#) | [Discord](#) |

LMSYS [Chatbot Arena](#) is a crowdsourced open platform for LLM evals. We've collected over 1,000,000 human pairwise comparisons to rank LLMs with the [Bradley-Terry model](#) and display the model ratings in Elo-scale. You can find more details in our [paper](#).

Arena

Full Leaderboard

Total #models: 99. Total #votes: 1,170,955. Last updated: 2024-05-20.

⚠ NEW! View leaderboard for different categories (e.g., coding, long user query)! This is still in preview and subject to change.

Rank* (UB) ▲	Rank (StyleCtrl) ▲	Model	Arena Score	95% CI ▲	Votes ▲	Organization	License
1	1	Gemini-Exp-1206	1374	+4/-5	20227	Google	Proprietary
1	1	ChatGPT-4o-latest...(2024-11-20)	1365	+4/-3	33383	OpenAI	Proprietary
1	4	Gemini-2.0-Flash-Thinking-Exp-1219	1364	+5/-6	15728	Google	Proprietary
2	4	Gemini-2.0-Flash-Exp	1357	+6/-4	19030	Google	Proprietary
3	1	o1-2024-12-17	1351	+7/-7	7289	OpenAI	Proprietary
6	4	o1-preview	1335	+4/-4	33194	OpenAI	Proprietary
7	7	DeepSeek-V3	1319	+6/-6	10510	DeepSeek	DeepSeek
7	10	Step-2-16K-Exp	1305	+8/-9	3374	StepFun	Proprietary

Overall	Bitext Mining	Classification	Clustering	Pair Classification	Reranking	Retrieval	STS	Summarization	Retrieval w/Instructions
English	Chinese	French	Polish						
Overall MTEB English leaderboard 🎉									
<ul style="list-style-type: none"> Metric: Various, refer to task tabs Languages: English 									
Rank	Model	Model Size (Million Parameters)	Memory Usage (GB, fp32)	Embedding Dimensions	Max Tokens	Average (56 datasets)	Classification Average (12 datasets)	Clustering Average (11 datasets)	
1	NV-Embed-v1					69.32	87.35	52.8	
2	voyage-large-2-instruct			1024	16000	68.28	81.49	53.35	
3	SFR-Embedding-Mistral	7111	26.49	4096	32768	67.56	78.33	51.67	
4	gte-Qwen1.5-7B-instruct	7099	26.45	4096	32768	67.34	79.6	55.83	
5	voyage-lite-02-instruct	1220	4.54	1024	4000	67.13	79.25	52.42	
6	GritLM-7B	7242	26.98	4096	32768	66.76	79.46	50.61	
7	e5-mistral-7b-instruct	7111	26.49	4096	32768	66.63	78.47	50.26	
8	google-gecko.text-embedding-p	1200	4.47	768	2048	66.31	81.17	47.48	
9	SE_v1					65.66	76.8	47.38	
10	GritLM-8x7B	46703	173.98	4096	32768	65.66	78.53	50.14	

Models

BERTish

all-MiniLM-L6-v2
MPNet
MPNet-personality



Feature extraction
(fine-tuning)

Open GPTs

Llama 3.2 3B



Text-generation
(in-context learning)

Quiz



MAX PLANCK INSTITUTE
FOR HUMAN DEVELOPMENT



Quiz

How are LLMs trained?

What are the main ways to use LLMs?

What aspects distinguish different LLMs?

How to choose an LLM?