

The background is a complex network of thin grey lines connecting various sized nodes. The nodes are colored in dark blue, light blue, and grey. Some nodes are simple dots, while others are larger circles with concentric rings. A large dark blue circle with a white center is prominent in the upper left. A light blue circle with a white center is in the lower left. A grey circle with a white center is in the upper right. A large black rectangle is positioned in the lower right, containing the text 'GPT & LLM' in white. A thin light blue horizontal line is located just below the text.

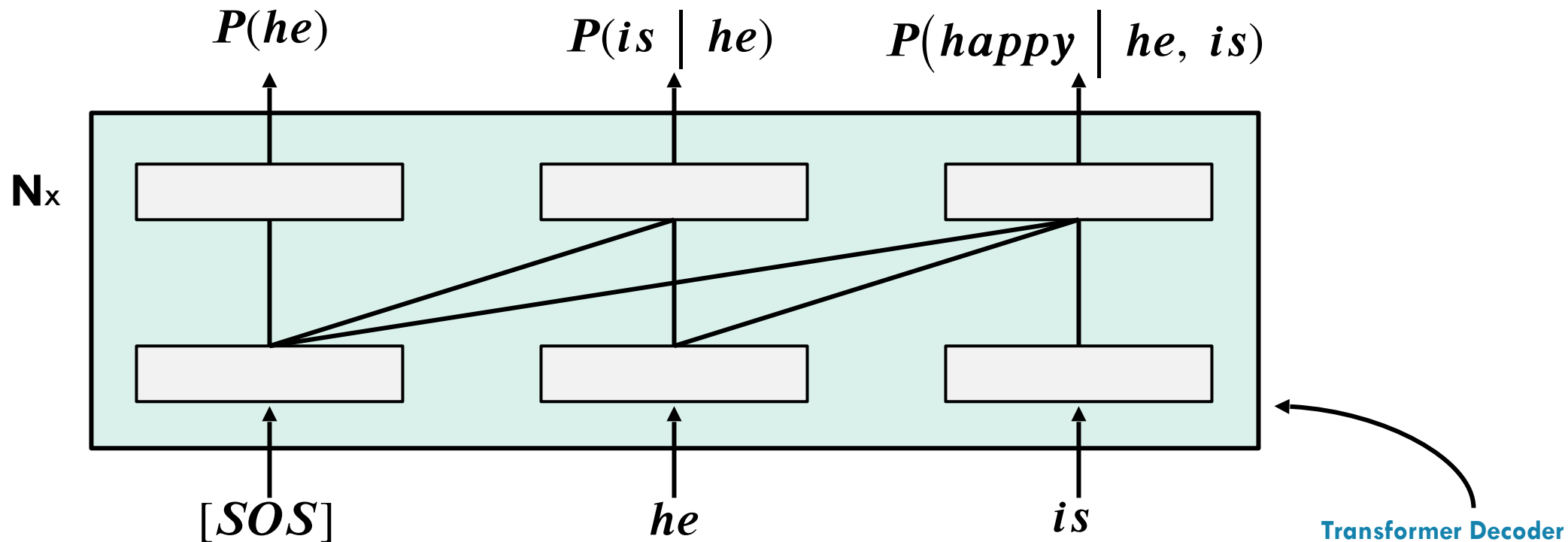
GPT & LLM

GPT

- GPT(Generative Pretrained Transformer)는 Transformer Decoder 기반의 Large Language Model

GPT

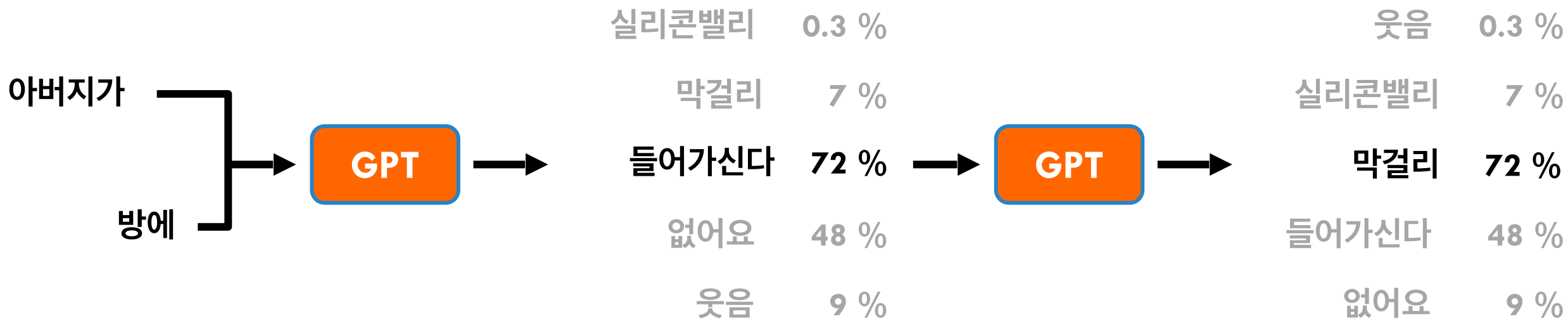
$$P(\textit{he is happy}) = P(\textit{he}) \cdot P(\textit{is} \mid \textit{he}) \cdot P(\textit{happy} \mid \textit{he}, \textit{is})$$



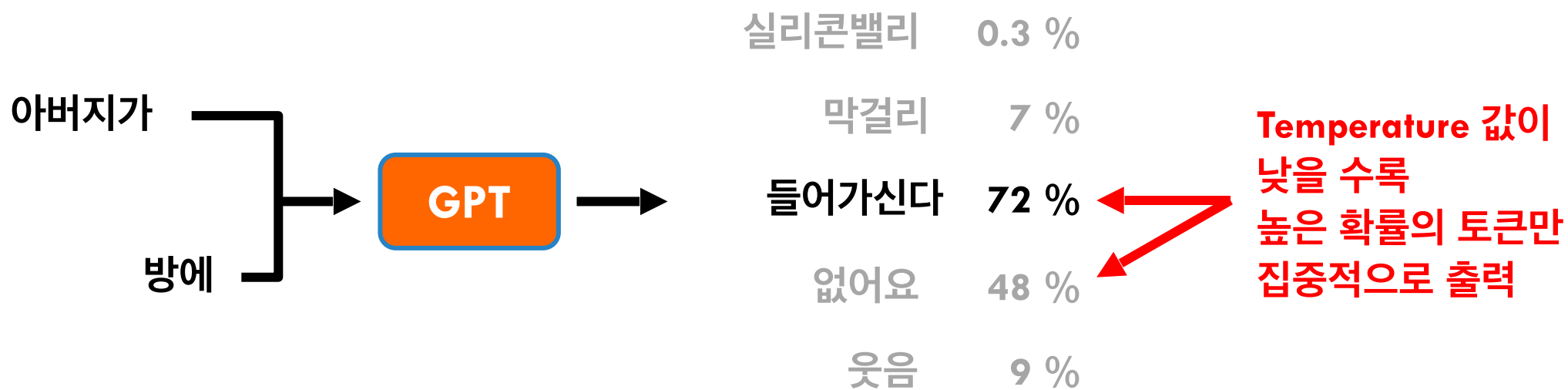
GPT



GPT



GPT



GPT

Top K

아버지가

방에

GPT

들어가신다 72 %

없어요 48 %

웃음 9 %

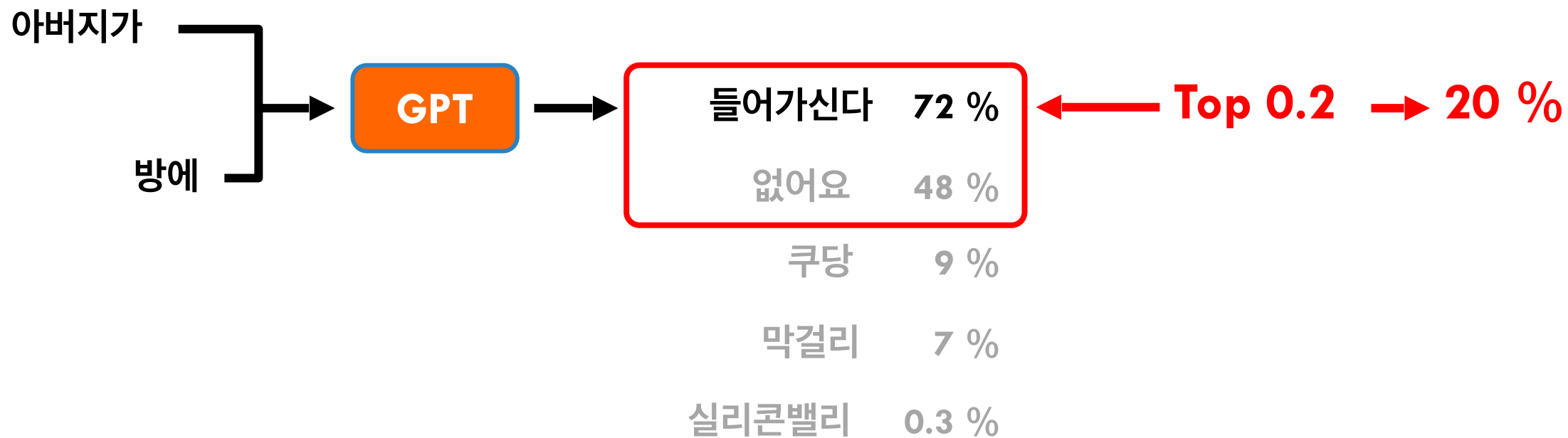
막걸리 7 %

실리콘밸리 0.3 %

← Top3

GPT

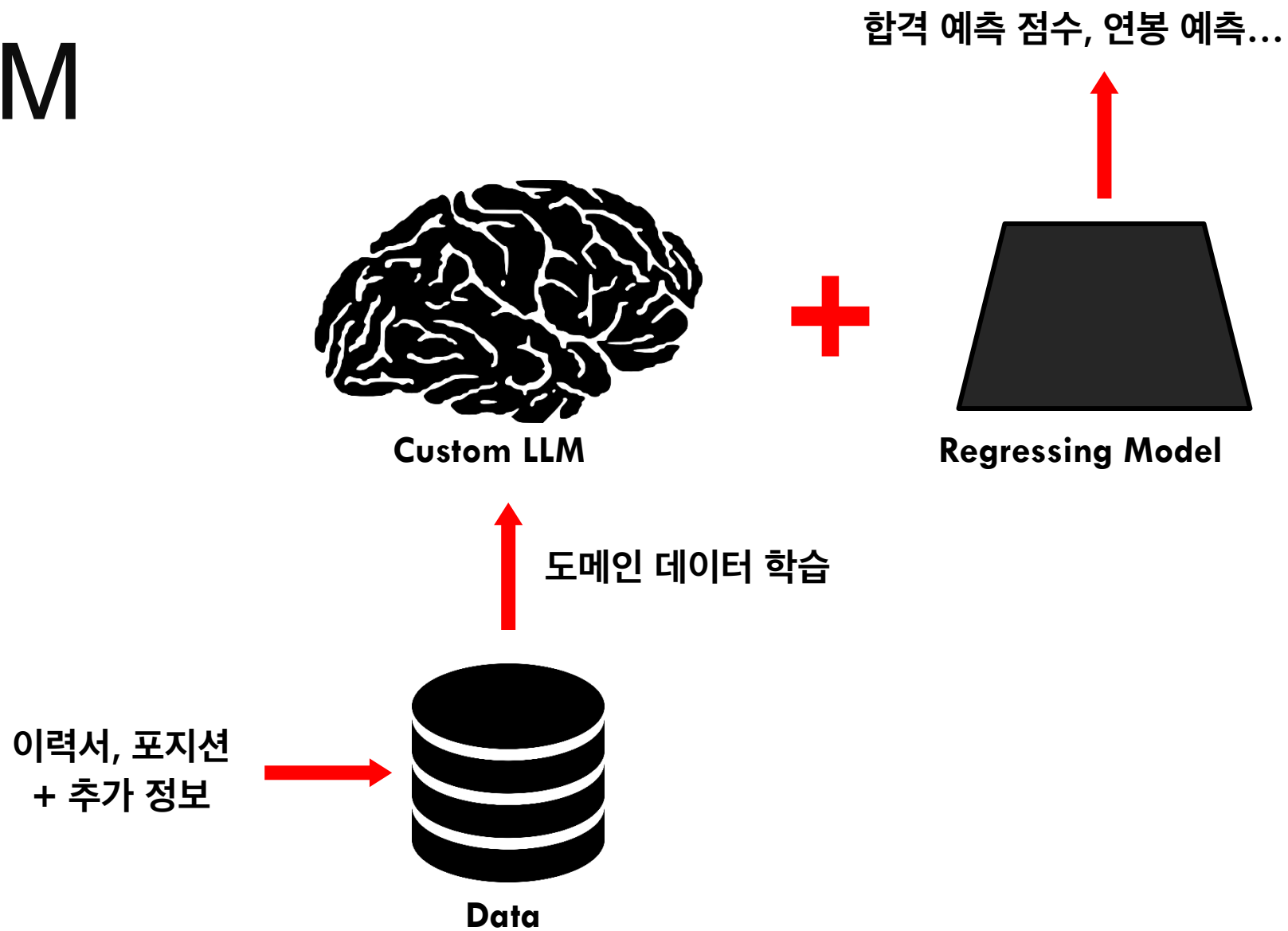
Top P



LLM

- Large Language Model은 $P(\text{text})$ 를 계산 할 수 있는 큰 모델
현재까지는 일반적으로 Transformer

LLM



LLM

- 대량의 corpus 로 Language Modeling을 학습 했을 때 창작 되는 능력
 - In context learning
 - Fewshot learning
 - Zeroshot learning
 - Reasoning
 - Chain of thought
 - Representation learning
 - Pretraining

LLM

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

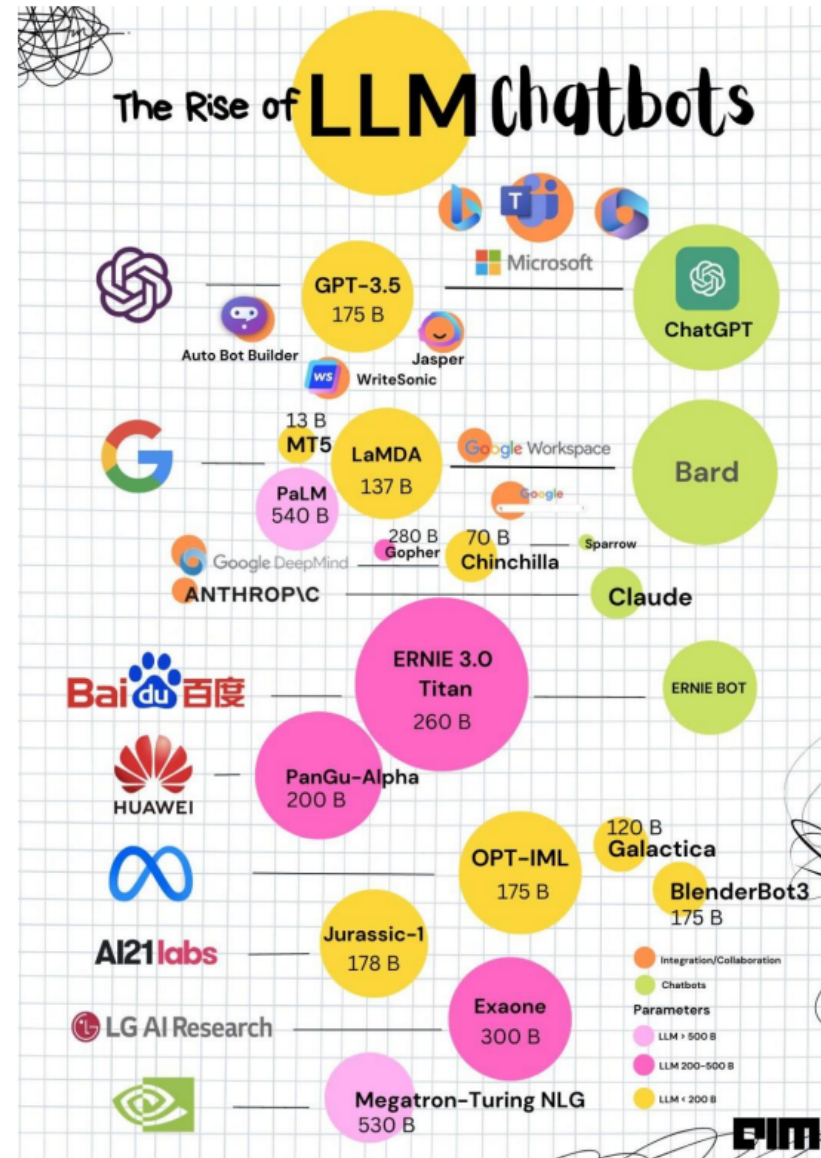
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

LLM



<https://www.linkedin.com/pulse/rise-llm-chatbots-bhasker-gupta/>

LLM

Open Source LLM

LLaMA	A foundational, 65-billion-parameter large language model. LLaMA.cpp Lit-LLaMA	GPT-Neo	An implementation of model & data parallel GPT3-like models using the mesh-tensorflow library.
Alpaca	A model fine-tuned from the LLaMA 7B model on 52K instruction-following demonstrations. Alpaca.cpp Alpaca-LoRA	GPT-J	A 6 billion parameter, autoregressive text generation model trained on The Pile.
Vicuna	An Open-Source Chatbot Impressing GPT-4 with 90% ChatGPT Quality.	Pythia	Interpreting Autoregressive Transformers Across Time and Scale
GPT4All	Demo, data, and code to train open-source assistant-style large language model based on GPT-J and LLaMa.	Dolly 2.0	The first open source, instruction-following LLM, fine-tuned on a human-generated instruction dataset licensed for research and commercial use.
Koala	A Dialogue Model for Academic Research	OpenFlamingo	An open-source reproduction of DeepMind's Flamingo model.
RedPajama	An Open Source Recipe to Reproduce LLaMA training dataset	Cerebras-GPT	A Family of Open, Compute-efficient, Large Language Models.
BLOOM	BigScience Large Open-science Open-access Multilingual Language Model BLOOM-LoRA	GALACTICA	The GALACTICA models are trained on a large-scale scientific corpus.
T5	Text-to-Text Transfer Transformer	Camel	A state-of-the-art instruction-following large language model designed to deliver exceptional performance and versatility.
OPT	Open Pre-trained Transformer Language Models.	h2oGPT	
UL2	A unified framework for pretraining models that are universally effective across datasets and setups.	MOSS	MOSS是一个支持中英双语和多种插件的开源对话语言模型。
RWKV	Parallelizable RNN with Transformer-level LLM Performance.	Open-Assistant	A project meant to give everyone access to a great chat based large language model.
StableLM	Stability AI Language Models.		