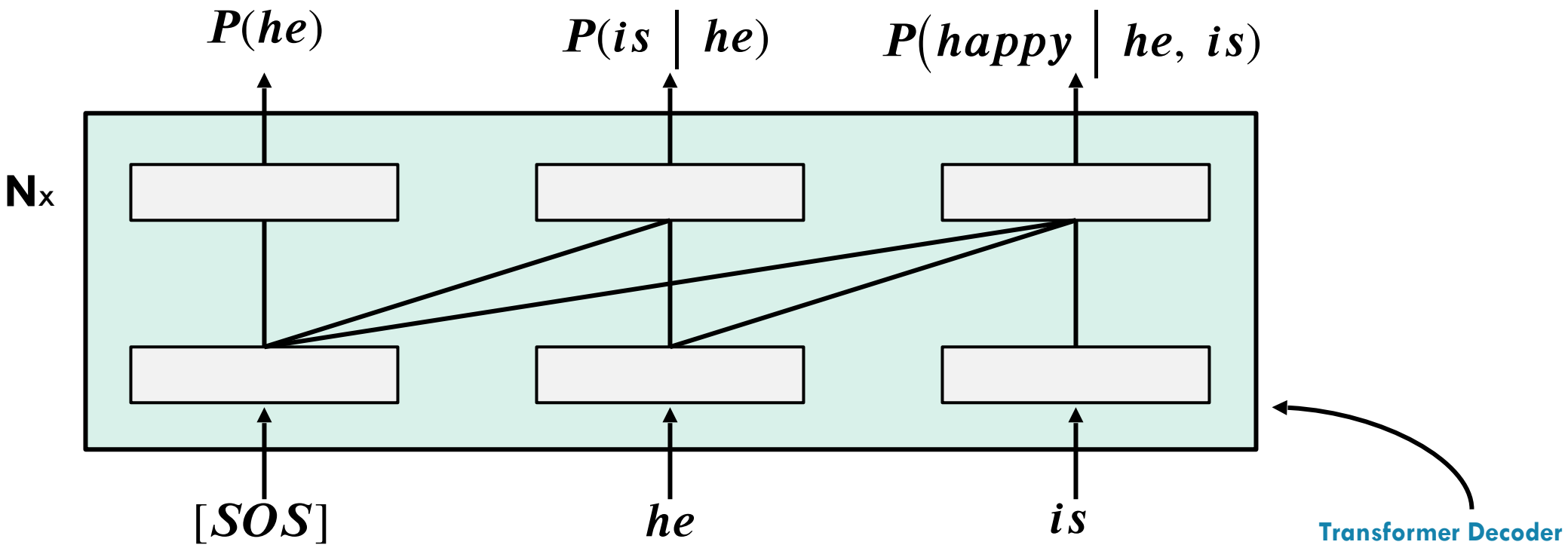

GPT & LLM

GPT

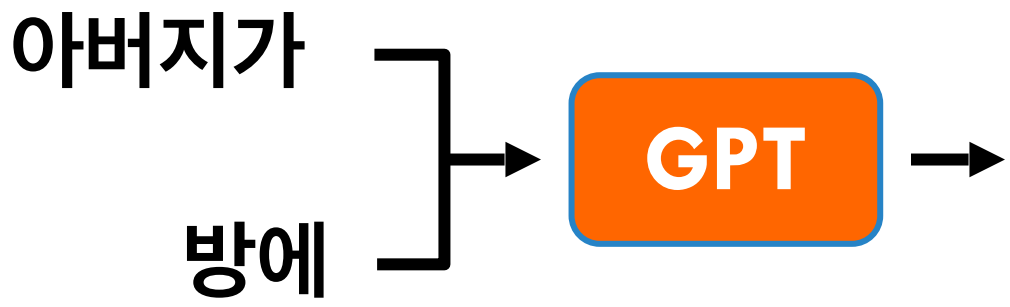
- GPT(Generative Pretrained Transformer)는 Transformer Decoder 기반의 Large Language Model

GPT

$$P(he\ is\ happy) = P(he) \bullet P(is \mid he) \bullet P(happy \mid he,\ is)$$



GPT



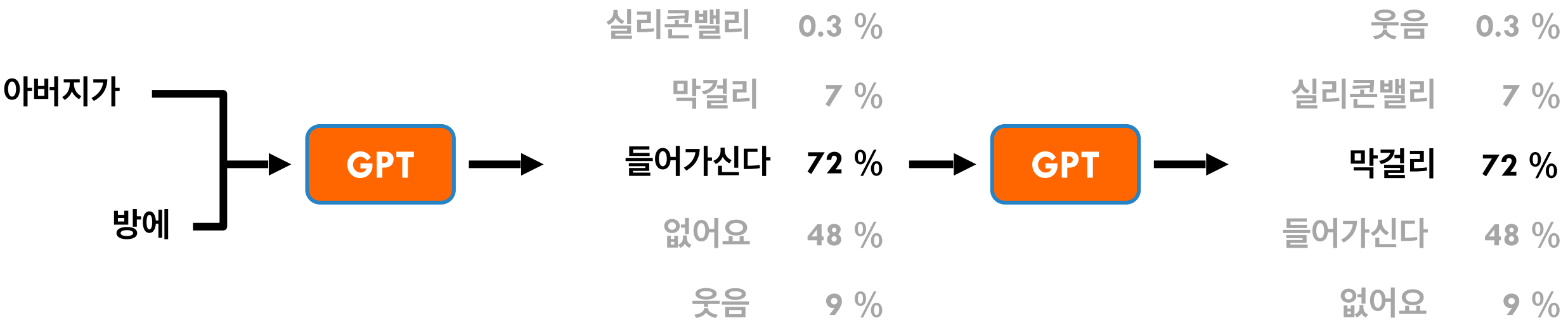
실리콘밸리 0.3 %

막걸리 7 %

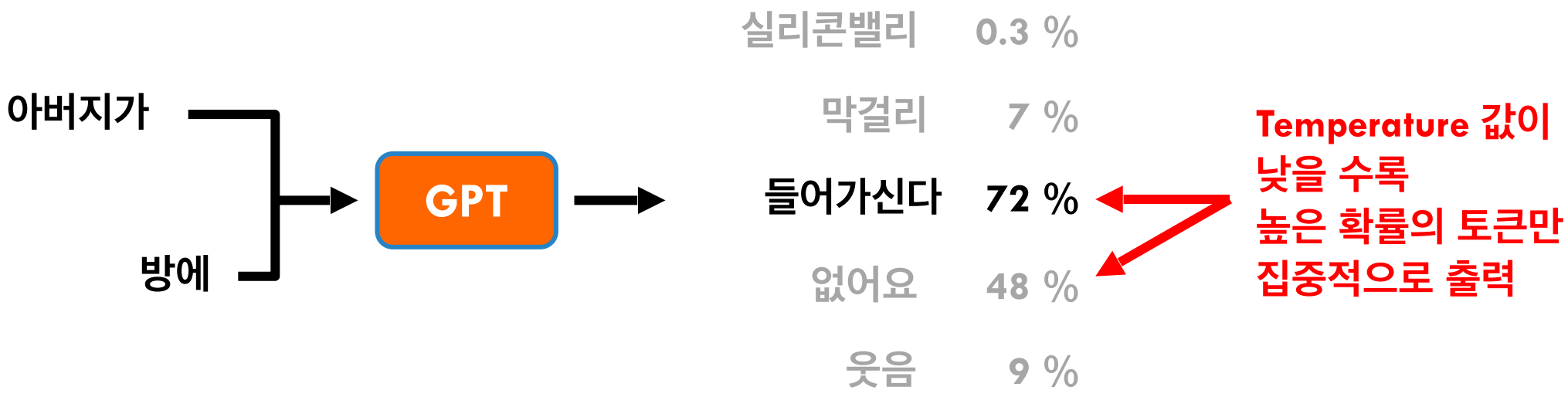
없어요 48 %

쿠당 9 %

GPT

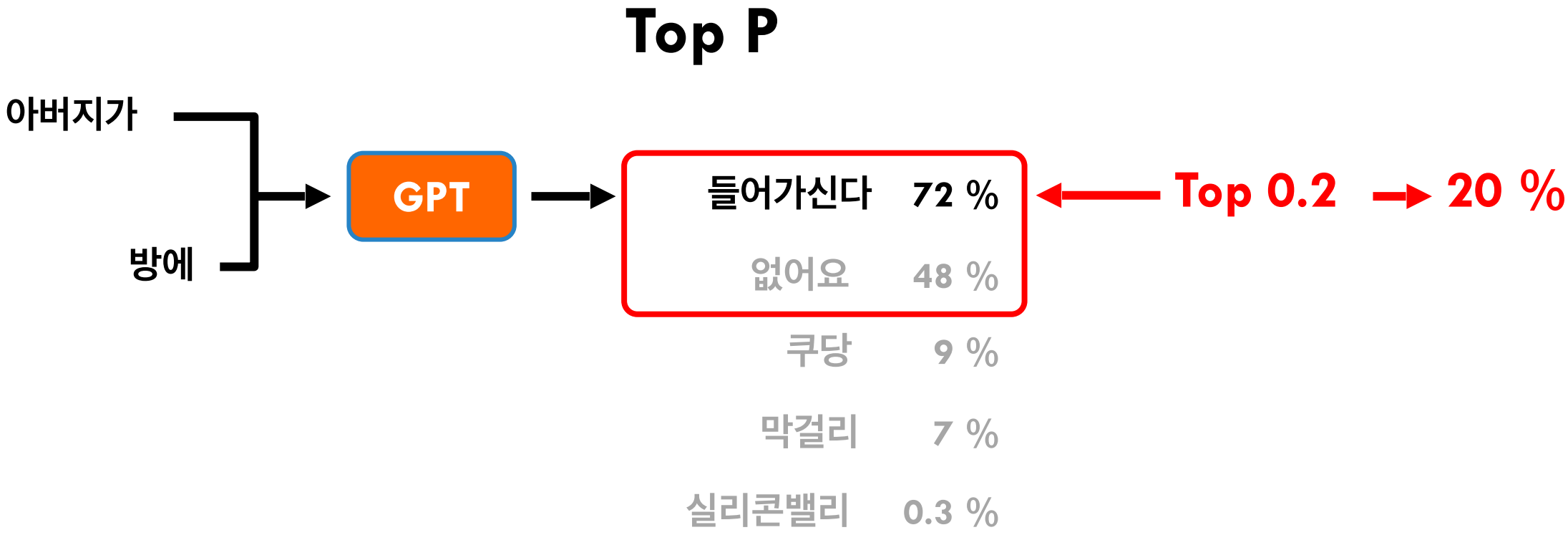

아버지가
방에
GPT
실리콘밸리 0.3 %
막걸리 7 %
들어가신다 72 %
없어요 48 %
웃음 9 %
GPT
웃음 0.3 %
실리콘밸리 7 %
막걸리 72 %
들어가신다 48 %
없어요 9 %

GPT


아버지가
방에
GPT
실리콘밸리 0.3 %
막걸리 7 %
들어가신다 72 %
없어요 48 %
웃음 9 %
Temperature 값이
낮을 수록
높은 확률의 토큰만
집중적으로 출력

GPT

GPT

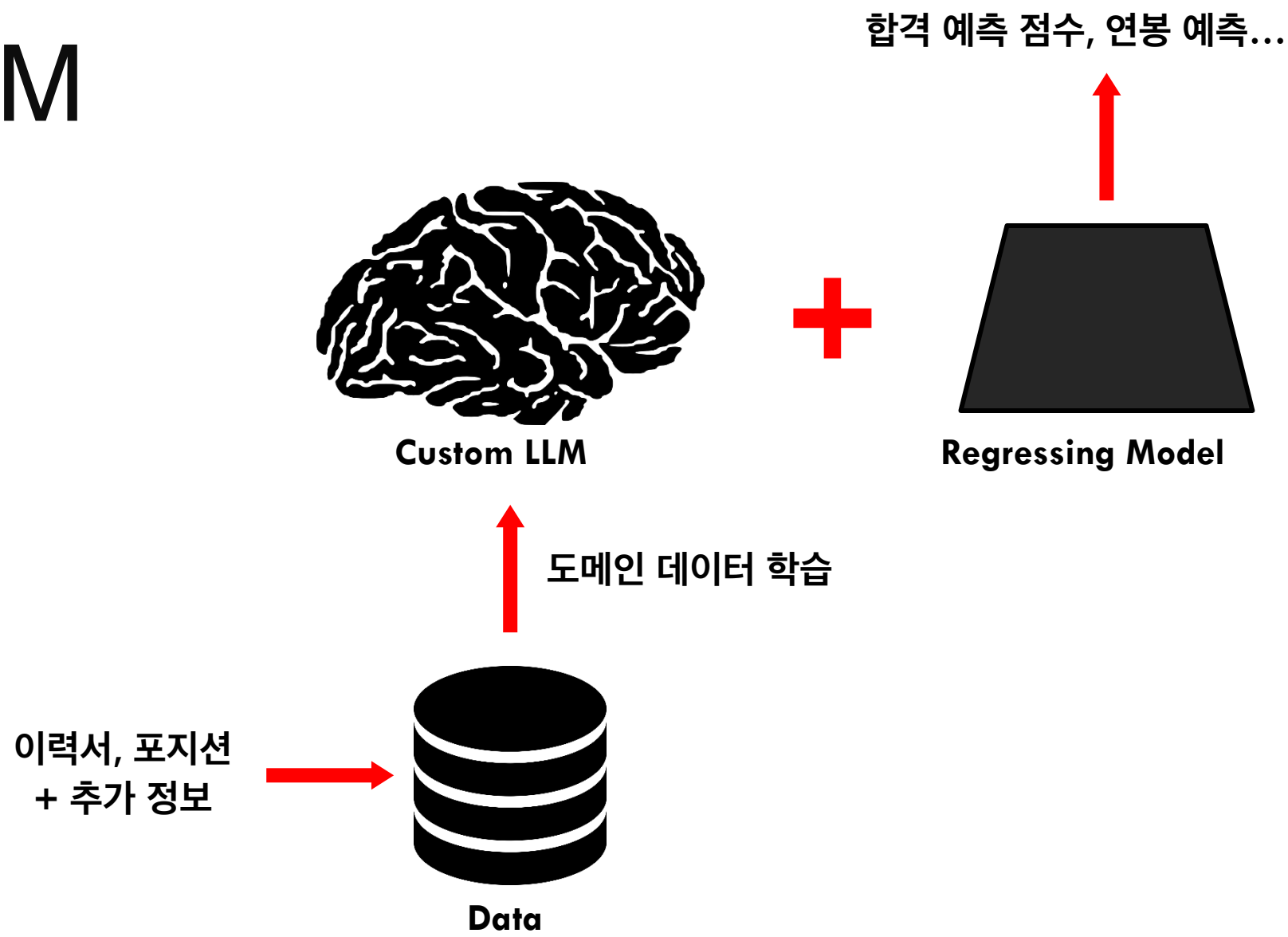


LLM

- Large Language Model은 P(text)를 계산 할 수 있는 큰 모델

  현재까지는 일반적으로 Transformer

LLM


합격 예측 점수, 연봉 예측…
Custom LLM
Regressing Model
도메인 데이터 학습
이력서, 포지션
+ 추가 정보
Data

LLM

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

LLM

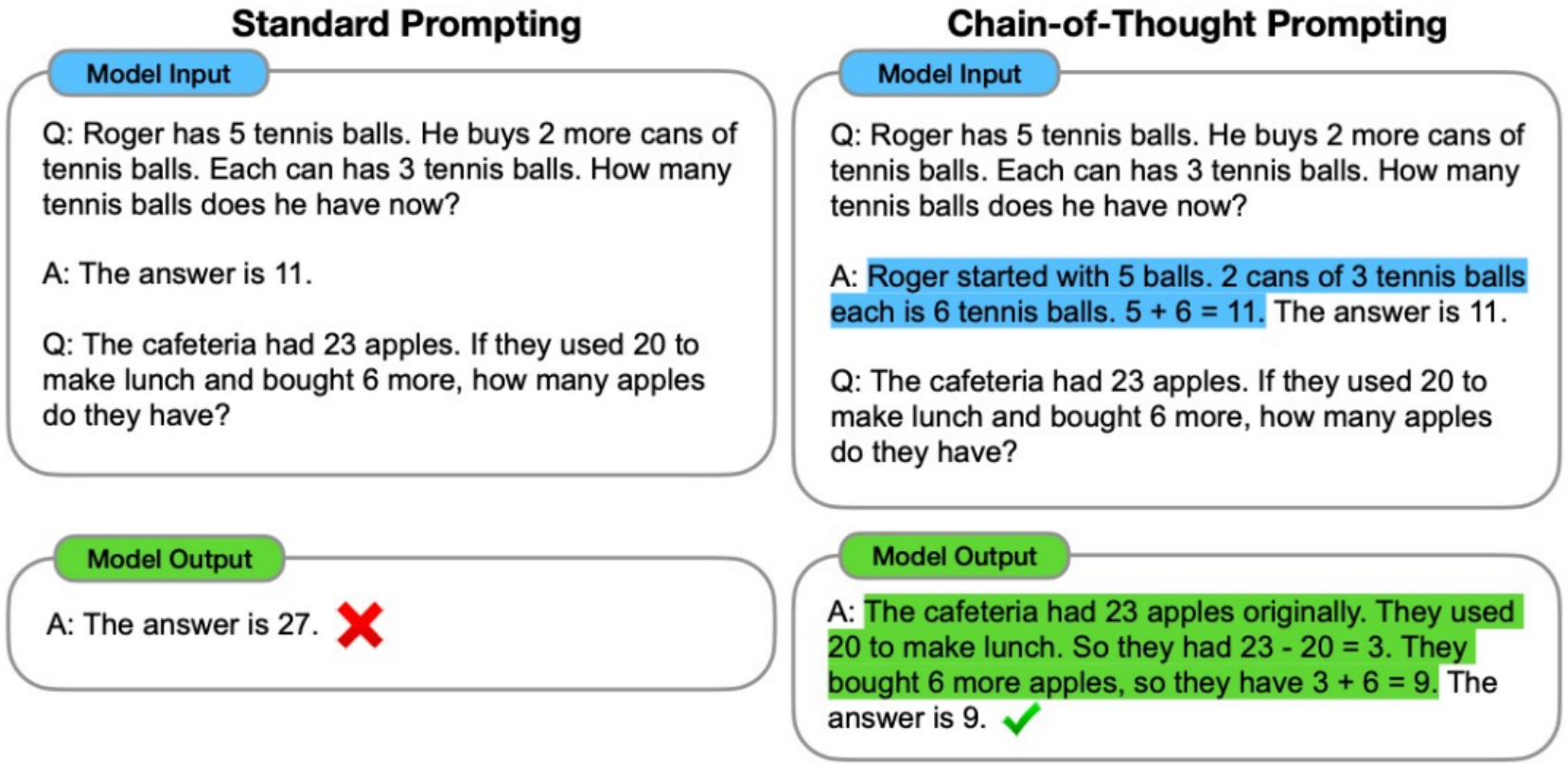
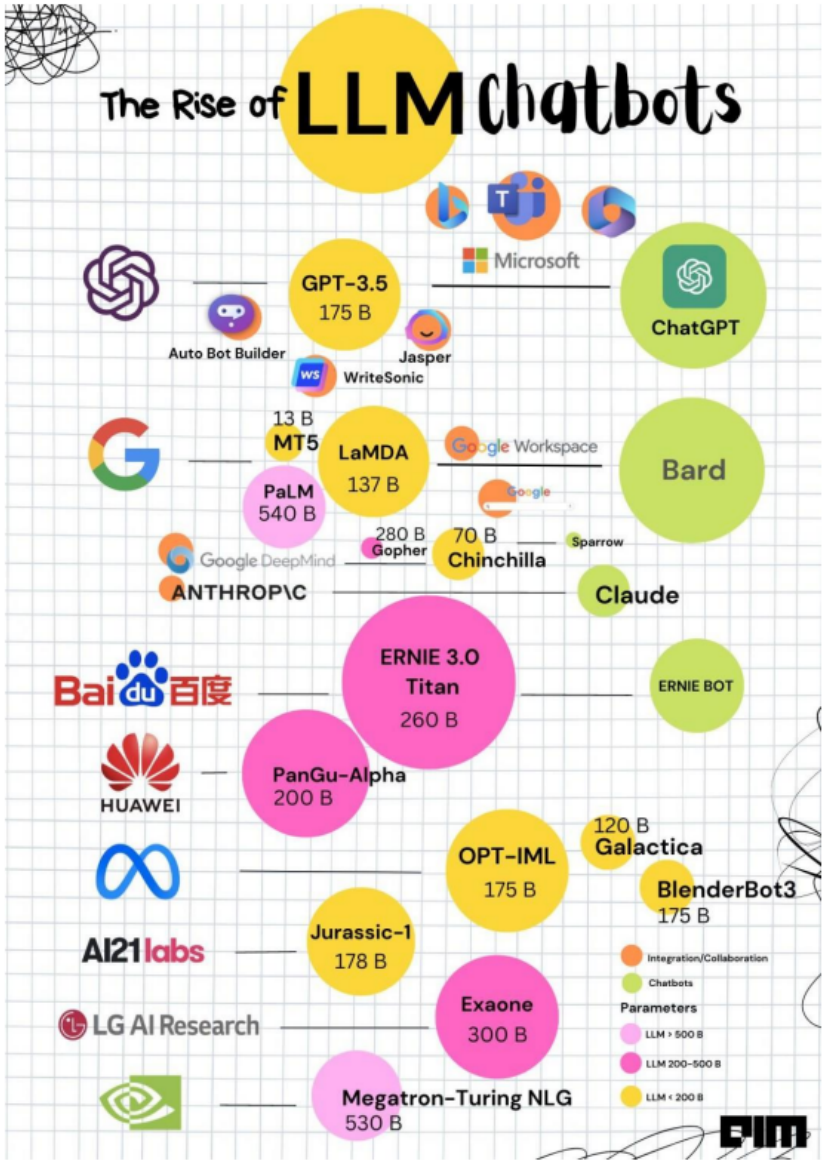


Figure 1: Chain-of-thought prompting enables large language models to tackle complex arithmetic, commonsense, and symbolic reasoning tasks. Chain-of-thought reasoning processes are highlighted.

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