

# ChatGPT and Business Opportunities

---

## Prachya Boonkwan (Arm)

Language and Semantic Technology Lab

National Electronics and Computer Technology Center, Thailand

[prachya.boonkwan@nectec.or.th](mailto:prachya.boonkwan@nectec.or.th), [kaamanita@gmail.com](mailto:kaamanita@gmail.com)

URL ⇒ <https://tinyurl.com/2fj5y3ny>



Courtesy of Paramount Pictures

# Who? Me?

- Nickname: **Arm** (P'/N'/E' Arm, etc.)
- Born: Aug 1981
- Work: researcher at NECTEC since 2005
- Education
  - Alma mater: Triam Udom Suksa School
  - B.Eng & M.Eng, CPE Kasetsart University
  - Obtained OCSC Scholarship in early 2008
  - Did a PhD in Informatics (Computational Linguistics) at University of Edinburgh during 2008-2013 (4.5 years)



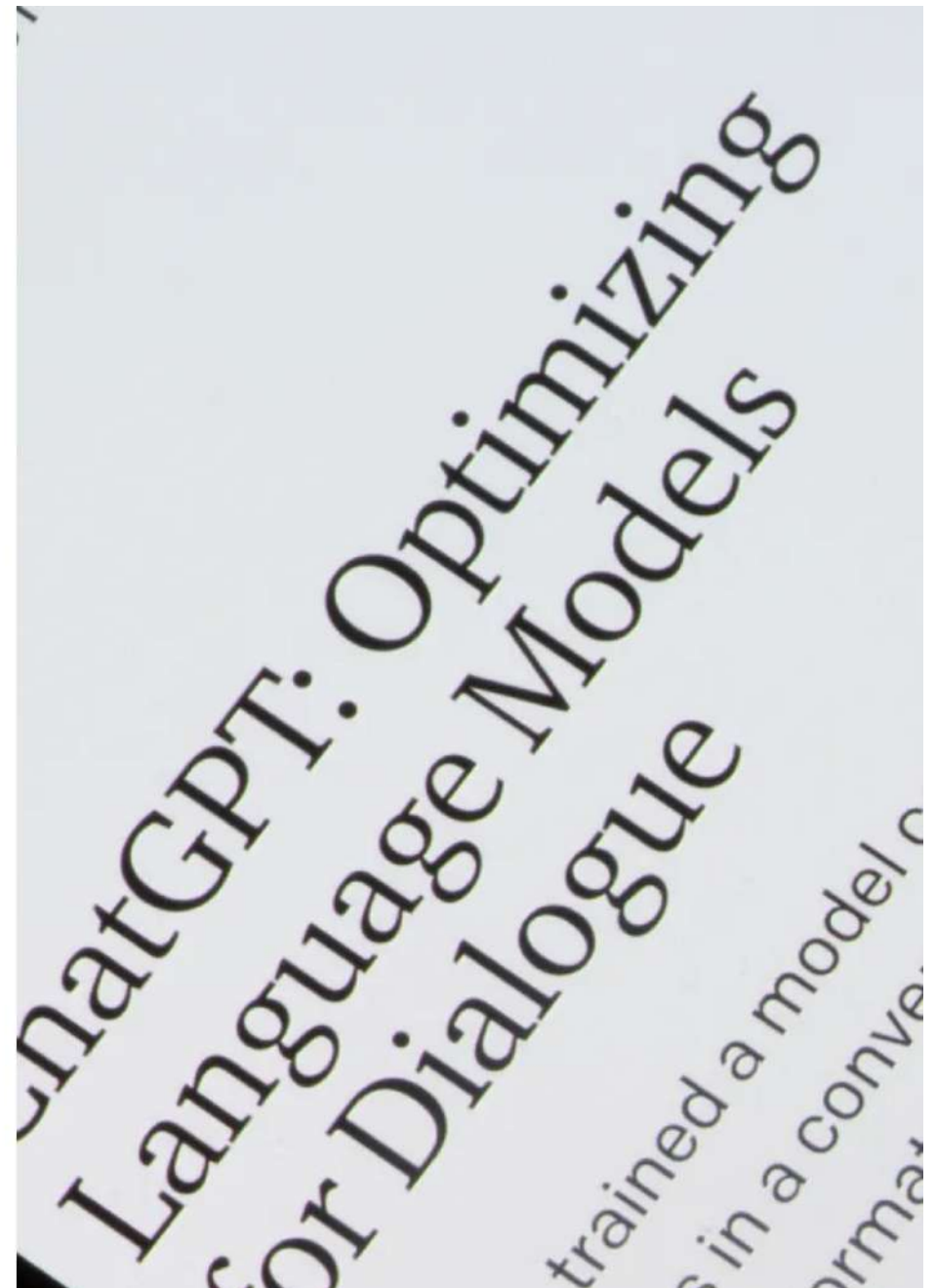
Princes Garden, Edinburgh (2010)

# ChatGPT

ChatGPT is an advanced AI chatbot trained by OpenAI which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

ChatGPT relies on the powerful `GPT-3.5` technology. GPT stands for Generative Pre-Trained Transformer, a complex neural network based on the revolutionary Attention concept.

<https://chatgptonline.net>







I'd like to write an email to the managing director. There are three points to cover: (1) The business contract is now approved and ready to sign on both sides. (2) The final budget is 15 million dollars. (3) The stock shares are 50-50.



Subject: Approval of Business Contract and Final Budget



Dear [Managing Director's Name],

I hope this email finds you well. I am writing to inform you that the business contract between our companies has now been fully approved and is ready for signature on both sides.

Additionally, I am pleased to announce that the final budget for this project has been confirmed at 15 million dollars. Our teams have worked hard to ensure that all necessary expenses are accounted for and that this budget represents the best possible value for both of our companies.

Finally, I wanted to confirm that the stock shares will be split equally between our companies at a 50-50 ratio. This arrangement is in line with our initial discussions and reflects a fair and equitable distribution of ownership.

I look forward to the opportunity to finalize this contract and begin working together on this exciting project. Please let me know if you need any further information or if you would like to arrange a time for the contract signing.

Best regards,

[Your Name]



I'm having a hair loss problem. Can taking a daily zinc pill solve the problem?



Zinc is an essential mineral that is important for a number of processes in the body, including hair growth. Zinc deficiency can lead to hair loss, and supplementing with zinc may help to promote hair growth in some people. However, hair loss can also be caused by a variety of other factors, such as genetics, hormonal imbalances, medical conditions, and certain medications.



If you are experiencing hair loss, it's best to consult a doctor or a dermatologist for an accurate diagnosis and appropriate treatment recommendations. They may recommend taking a zinc supplement in addition to other treatments, such as topical products, lifestyle changes, or medications. However, it's important to keep in mind that taking too much zinc can also have negative side effects, so it's important to take only the recommended amount.

In conclusion, while zinc supplementation may help with hair loss in some cases, it's not a cure-all solution and it's important to work with a healthcare professional to determine the best course of action for your specific situation.



My husband and I are going to Edinburgh in August. Can you recommend a journey plan, which consists of sightseeing, going to a concert, and visiting Edinburgh Castle?



Sure! Here's a sample itinerary for a trip to Edinburgh in August:



Day 1:

- Arrive in Edinburgh and check into your hotel.
- Explore the city's Old Town and visit the Royal Mile, a street that runs from Edinburgh Castle to the Palace of Holyroodhouse.
- Visit Edinburgh Castle, one of Scotland's most famous landmarks, to learn about the history of the castle and the city.

Day 2:

- Head to the Princes Street Gardens to enjoy the city's green spaces and take in the views of Edinburgh Castle.
- Visit the National Museum of Scotland to learn about the history and culture of Scotland.
- In the evening, attend a concert of your choice at the Usher Hall, one of Scotland's largest concert halls.

Day 3:

- Visit the Royal Botanic Garden Edinburgh to see the diverse plant collection, including a large collection of rhododendrons and azaleas.
- Explore the New Town and visit the Georgian House, a museum that showcases life in Edinburgh during the 18th century.
- In the evening, take a ghost tour of the city to learn about its haunted history.



Please write a piece of Python code that parses an input text in English into a dependency tree.



Here's an example of how you could parse English text into a dependency tree using the popular Natural Language Toolkit (NLTK) library in Python:



python

Copy code

```
import nltk
from nltk import word_tokenize
from nltk.parse import CoreNLPParser

# Initialize the parser
parser = CoreNLPParser(url='http://localhost:9000')

# Define the input text
text = "I saw the white car with my own eyes."

# Tokenize the text
tokens = word_tokenize(text)

# Parse the text into a dependency tree
dep_tree = next(parser.parse(tokens))

# Display the dependency tree
print(dep_tree)
```

This code uses the CoreNLPParser from NLTK, which is a wrapper for the Stanford Parser, to parse the input text into a dependency tree. The input text is first tokenized into individual words, and then passed to the parser for parsing. The parse method returns a generator, so we use `next` to get the first (and only) parse tree.



**Aran Komatsuzaki** 

@arankomatsuzaki

OpenAI did what used to be  
considered impossible.

They made people want to use  
Bing.



# Key Differences

## ChatGPT = chatbot

- It answers your questions with its knowledge
- Generative AI: it generates new outputs unseen in the dataset
- Its knowledge is based on the training data (up to 2021)
- Use it when you need quick answers and get the job done
- **GPTZero**: Detector for GPT outputs

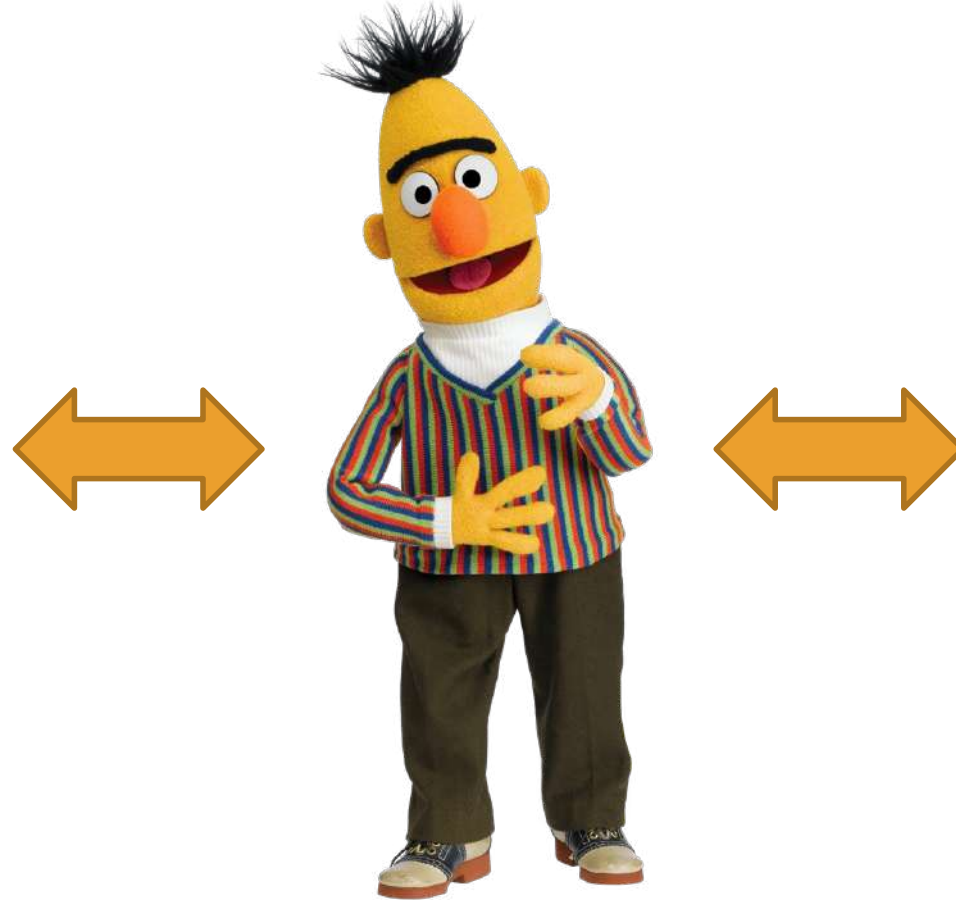
## Google = search engine

- It searches things in a very large repository for you
- Non-generative AI: it never generates unseen outputs
- Its knowledge is always up-to-date and much more voluminous
- Use it when you want to compile knowledge from external resources and internalize the knowledge

# Dissection of ChatGPT



**1) Engine:**  
Transformer Model



**2) Language Model:**  
GPT (BERT-based)

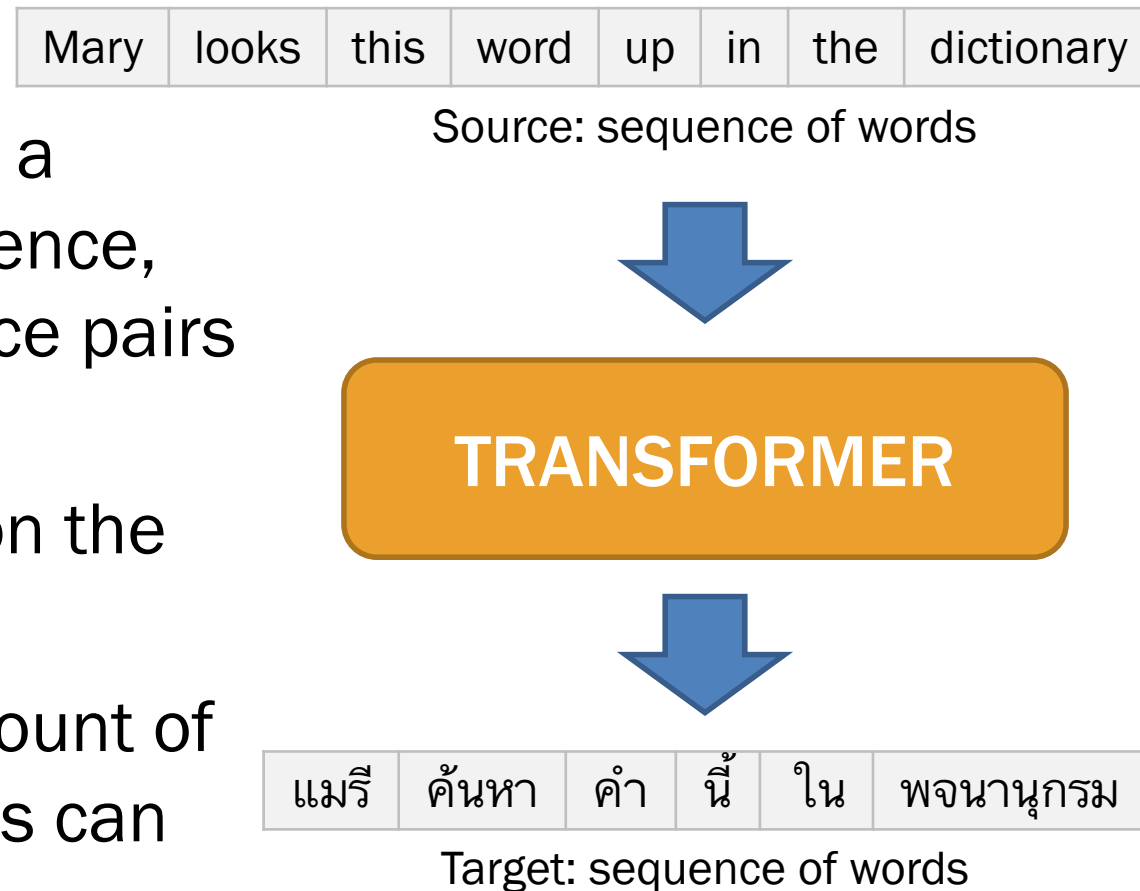


**3) Dataset:**  
Chat-based Instruction

# 1. Transformer Model

# The Transformer (Vaswani et al., 2016)

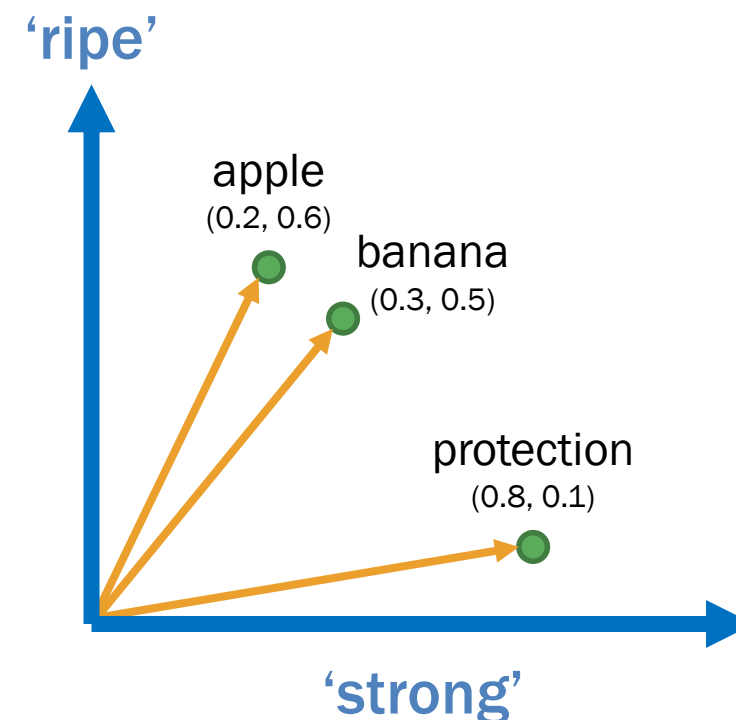
- Sequence-to-sequence model
  - **Translation:** It learns how to produce a target sequence from a source sequence, given a very large dataset of sequence pairs
  - **Pros:** It automatically learns **word collocations** and **phrase structures** on the input and output sequences
  - **Cons:** It consists of an expansive amount of neuron cells, and the training process can be quite time-consuming





# Recap: Word Vector

- Distributional semantics
  - Measured by co-occurrence of words and their contexts (i.e. **context distribution**)
  - Various types of similarity metrics and context are employed (Dagan+, 2008)
    - Backoff method (Katz, 1987) with interpolation (Jelinek+, 1980) are required for smoothing the zero counts
    - Bigram co-occurrence is commonly used as the context representation (Brown+, 1992)
    - Well-known techniques include Latent Semantic Analysis (Dumais, 2005) and PMI (Church+, 1990)

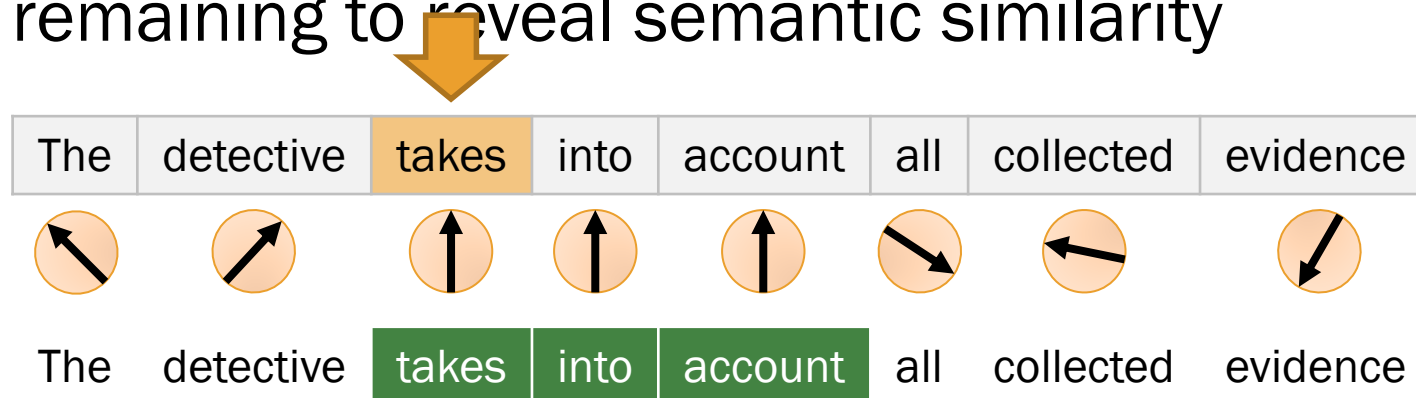


In this over-simplified example, each numeric element is:

$$p(\text{word}|\text{context})$$

# Pros: Multiword Expression (MWE)

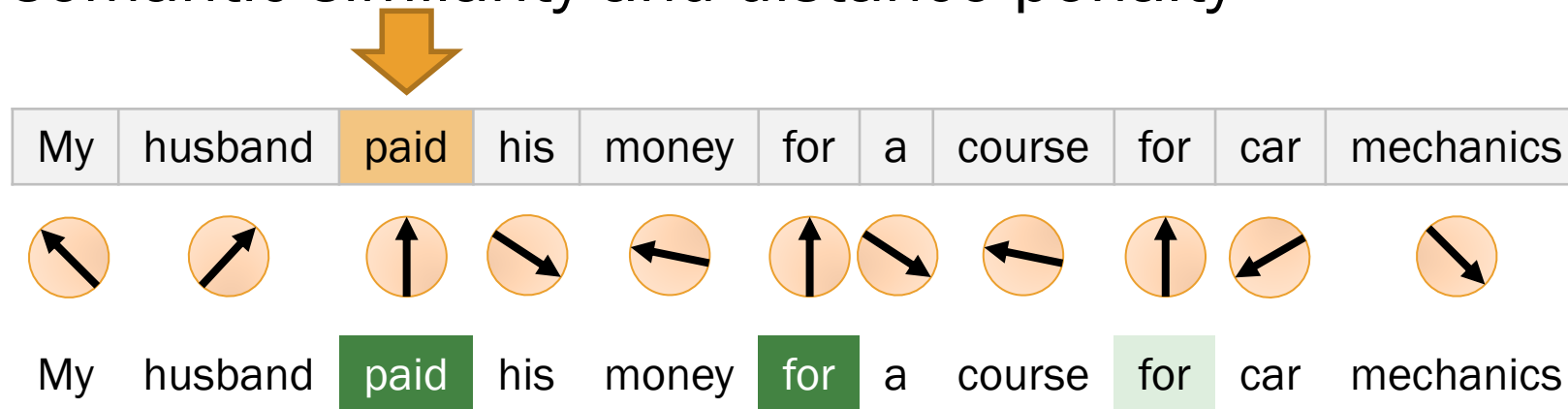
- It recognizes the idiosyncratic collocations of at least 2 words
  - E.g. 'peanut butter', 'car park', 'kick the bucket', 'take into account', 'break up'
- It learns MWEs by comparing each word with the remaining to reveal semantic similarity



**MWE is extracted because of semantic similarity**

# Pros: Moderate-Distance Dependency

- It recognizes word collocation that is separate within a moderate distance
  - E.g. 'look \_\_\_ up', 'ask \_\_\_ out', 'pay \_\_\_ for'
- It learns moderate-distance dependency with semantic similarity and distance penalty



The similarity is penalized by the distance



# Scaled Dot-Product Attention

- Semantic similarity  $\Rightarrow$  search engine
  - Query is compared against each key with dot product
  - The more similar the key is to the query, the more weight its value will get

**Simple Form**

$$w_i \propto \mathbf{k}_i \cdot \mathbf{q}$$

$$\mathbf{r} = \sum_{i=1}^N w_i \mathbf{v}_i$$

**Matrix Form**

$$\mathbf{w} = \text{Softmax}(\mathbf{K} \times \mathbf{q})$$

$$\mathbf{r} = \mathbf{V}^T \times \mathbf{w}$$

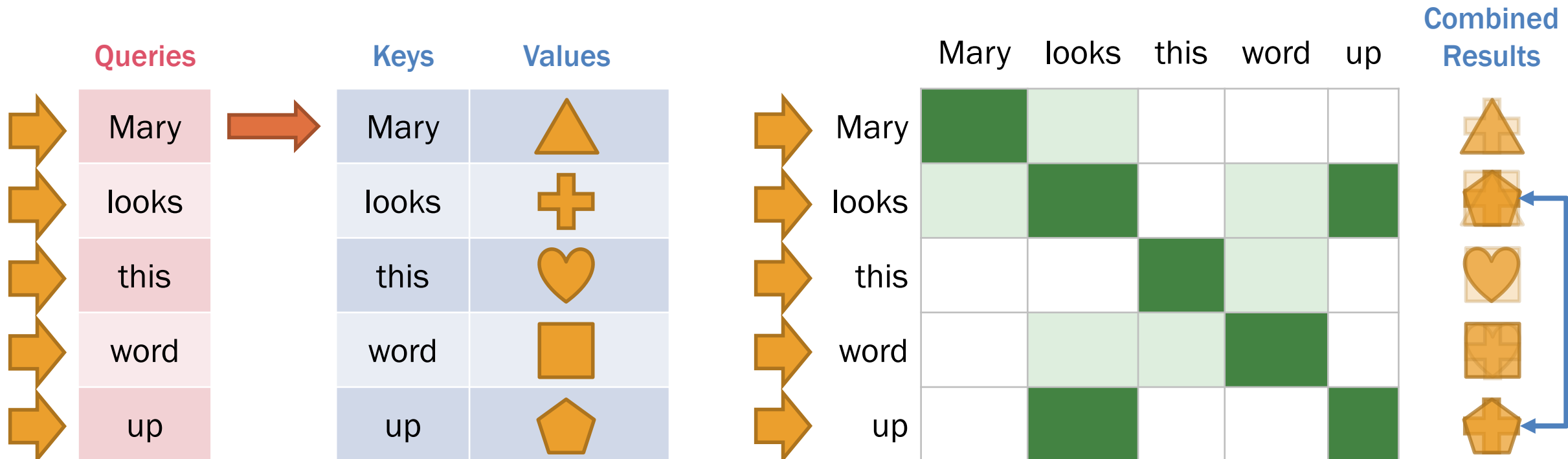




# Self-Attention

- Scaled dot-product attention whose queries and keys are the same
- Collocations will have almost similar results

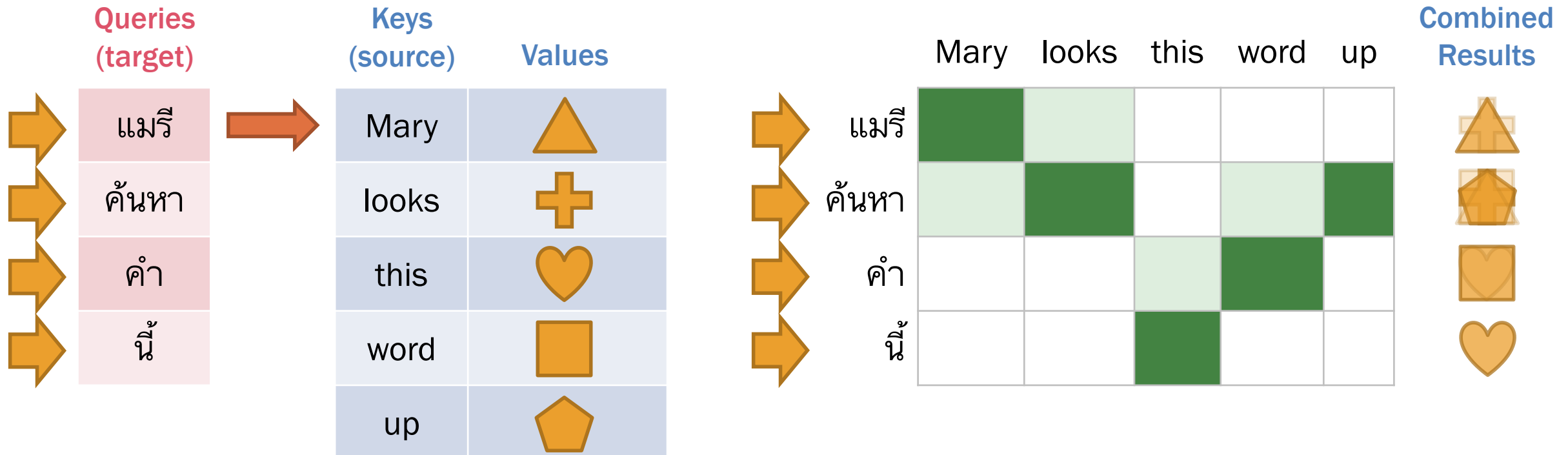
$$\begin{array}{ll} \text{Matrix} & \mathbf{W} = \text{Softmax}(\mathbf{K} \times \mathbf{K}^T) \\ \text{Form} & \mathbf{R} = \mathbf{W} \times \mathbf{V} \end{array}$$



# Alignment Attention

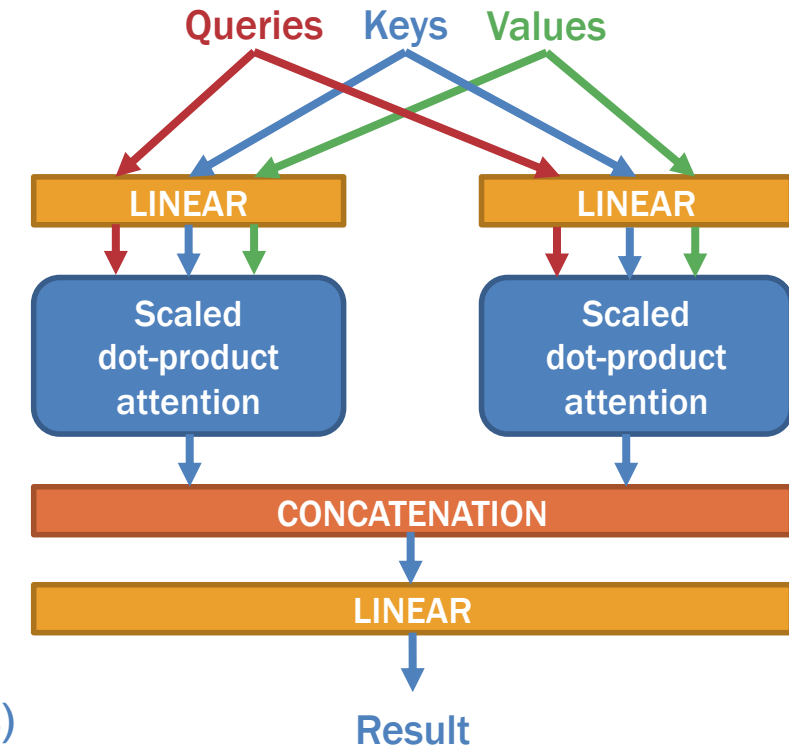
- Scaled dot-product attention whose queries are the target and whose keys are the source
- Collocation alignment via semantic similarity

**Matrix Form**  $\mathbf{W} = \text{Softmax}(\mathbf{Q} \times \mathbf{K}^T)$   
 $\mathbf{R} = \mathbf{W} \times \mathbf{V}$



# Multihead Attention

- Scaled dot-product attention has a drawback
  - It recognizes **only one** type of word collocation
  - If we assume more than one type of word collocation per sequence, then we have to combine multiple attention heads [default = 8 heads]



HEAD 1 (looks \_\_\_ up)

	Mary	Poppins	looks	this	word	up
Mary	■					
Poppins		■				
looks			■			■
this				■		
word					■	
up			■			■

HEAD 2 (Mary Poppins)

	Mary	Poppins	looks	this	word	up
Mary	■	■				
Poppins	■	■				
looks			■			
this				■		
word					■	
up						■

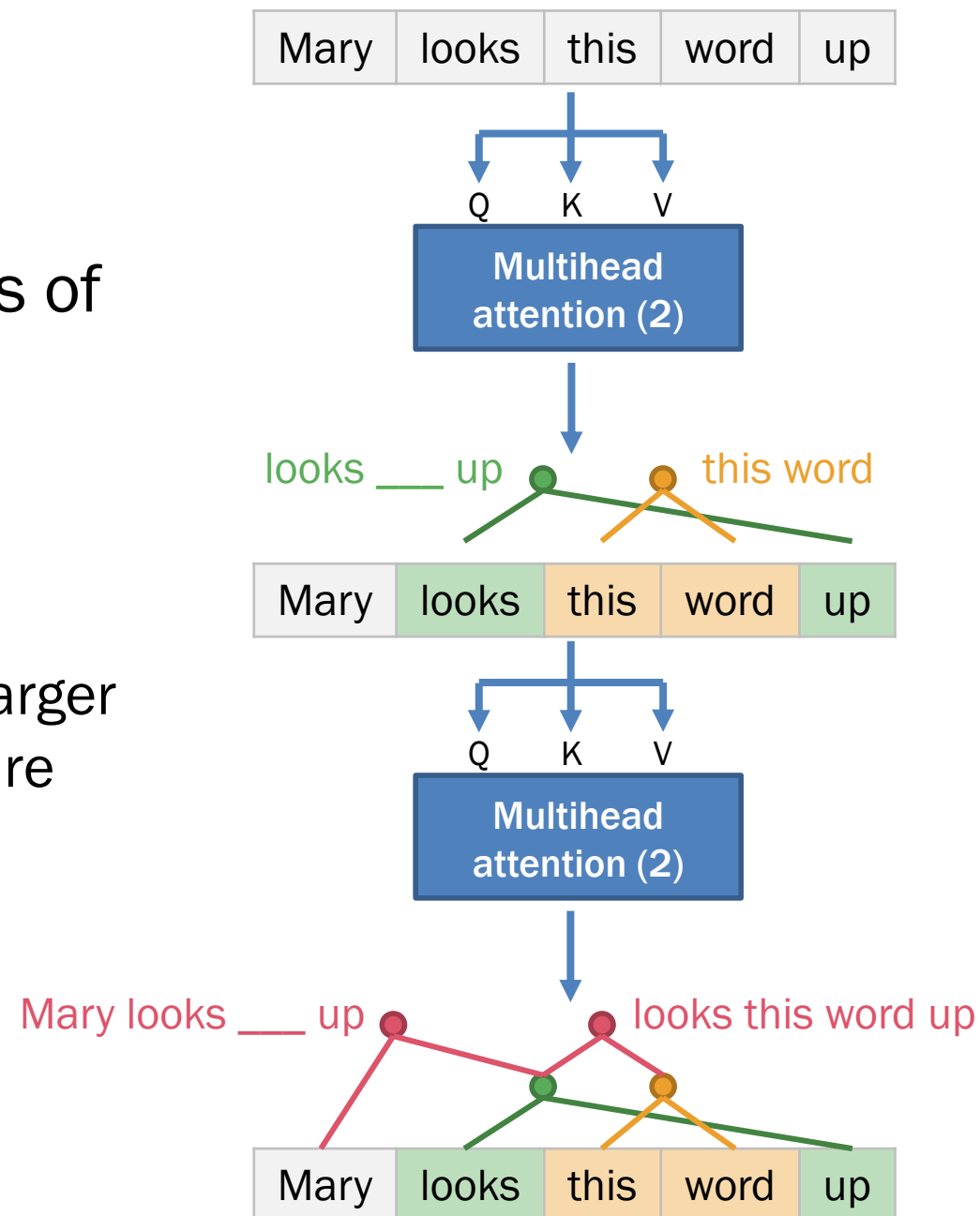
## Notation

Q K V

Multihead  
attention ( $n$ )

# Phrase Structure

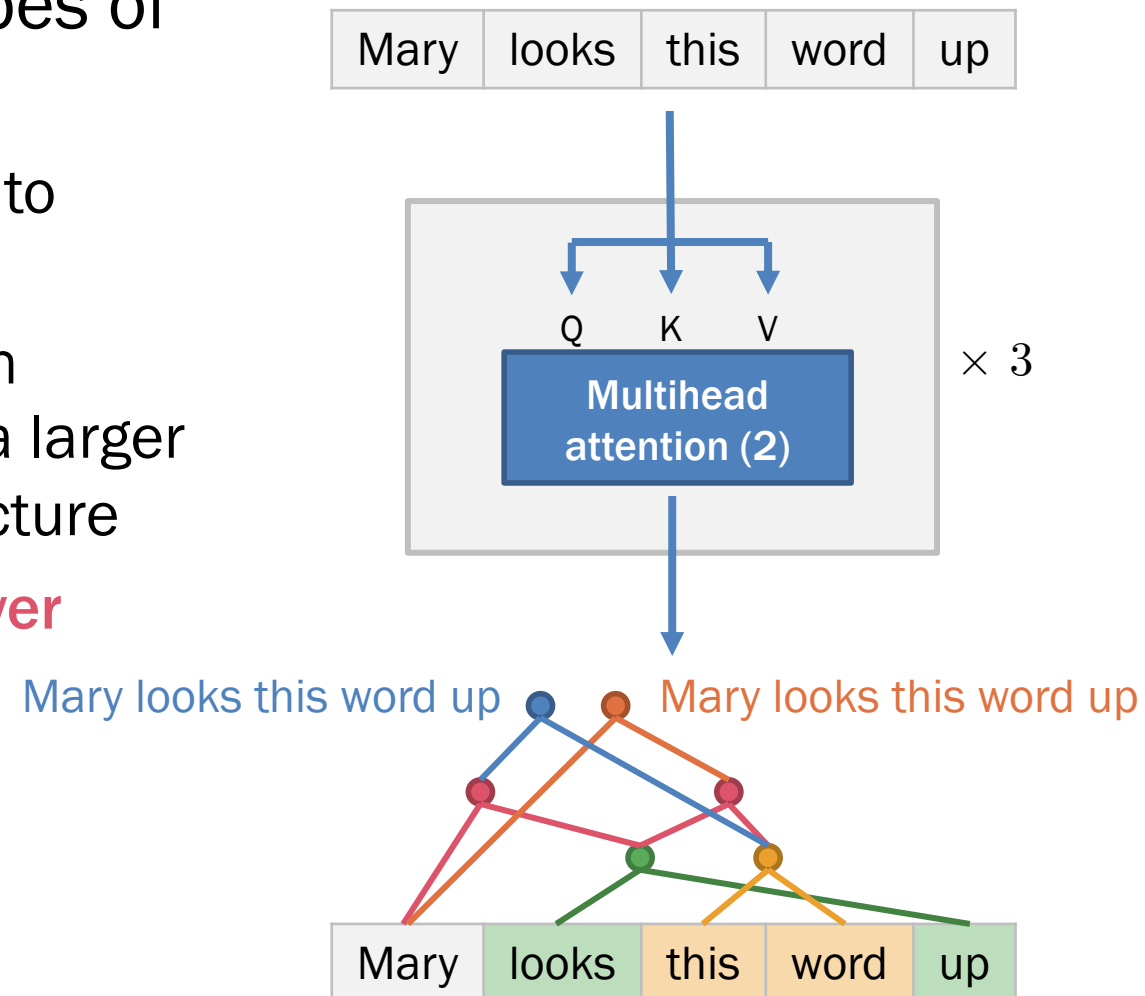
- $H$ -head self-attention recognizes  $H$  types of word collocation per sequence
  - One layer can combine consecutive words to become a phrase
  - More layers of multihead self-attention can combine consecutive phrases to become a larger phrase or even a sentence  $\Rightarrow$  phrase structure
  - Each layer is simply called an **encoding layer**





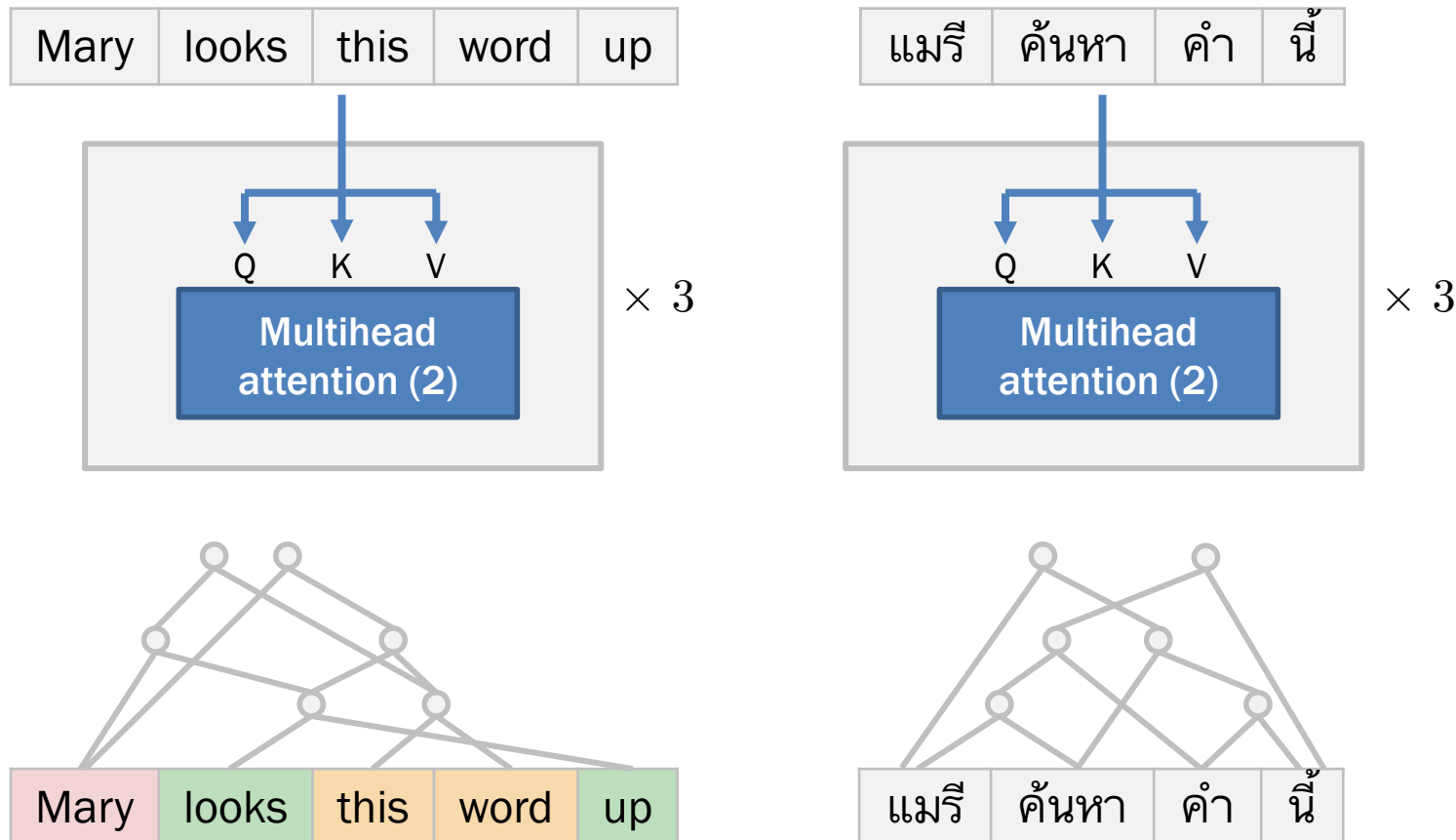
# Phrase Structure

- $H$ -head self-attention recognizes  $H$  types of word collocation per sequence
  - One layer can combine consecutive words to become a phrase
  - More layers of multihead self-attention can combine consecutive phrases to become a larger phrase or even a sentence  $\Rightarrow$  phrase structure
  - Each layer is simply called an **encoding layer**



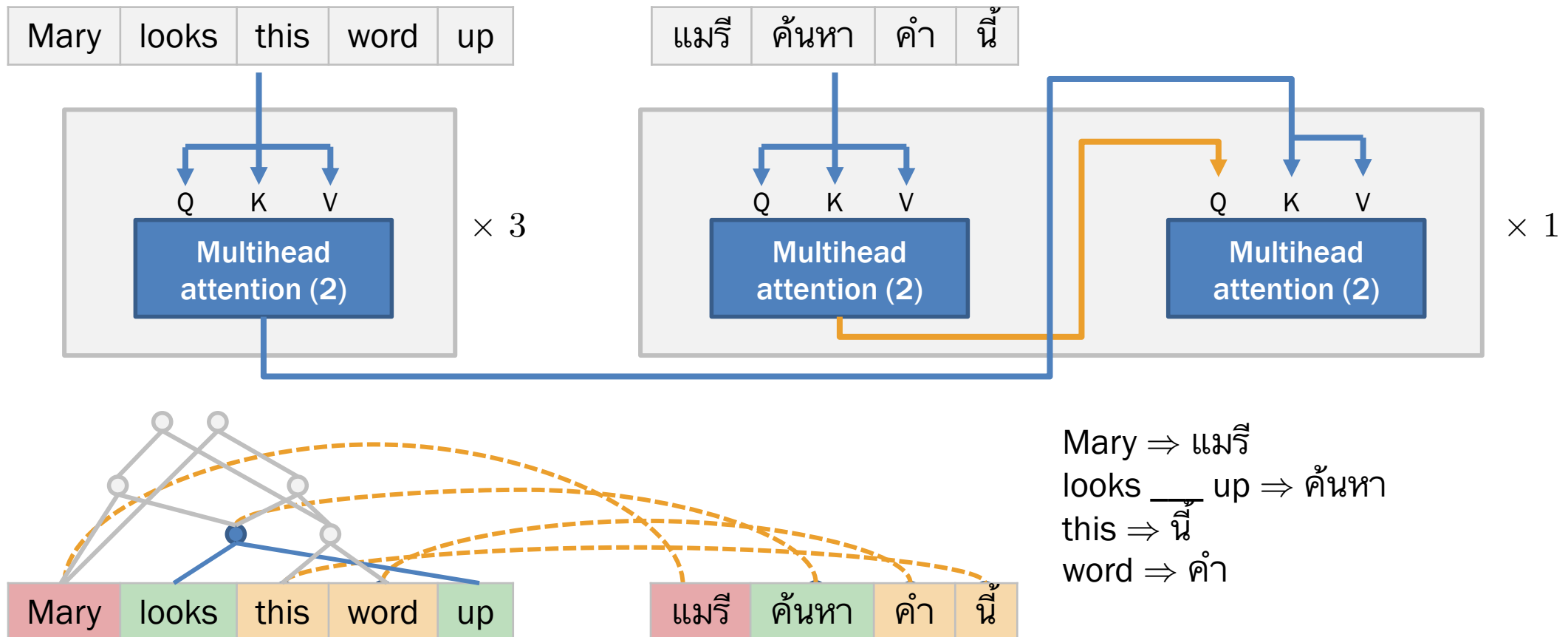
# Alignment of Phrase Structures

- $H$ -head alignment attention recognizes  $H$  pairs of phrase structures



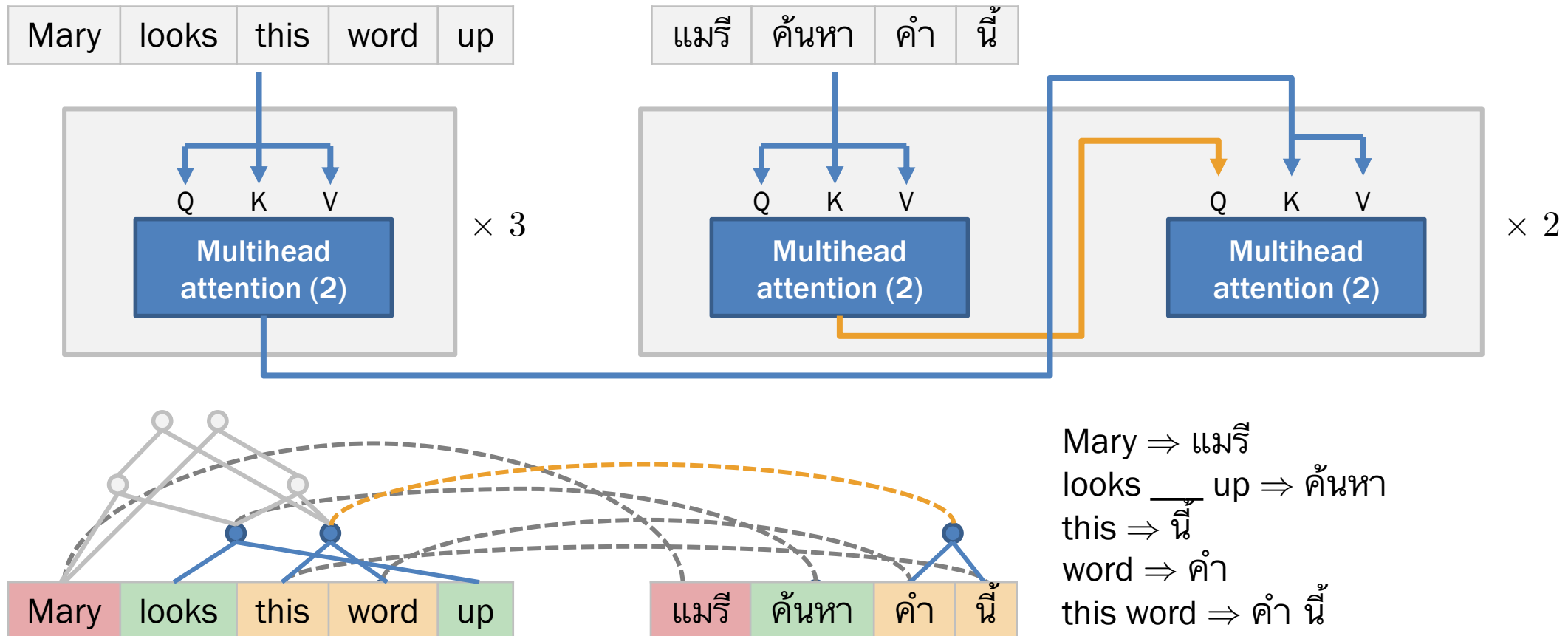
# Alignment of Phrase Structures

- $H$ -head alignment attention recognizes  $H$  pairs of phrase structures  $\Rightarrow$  **decoding layer**



# Alignment of Phrase Structures

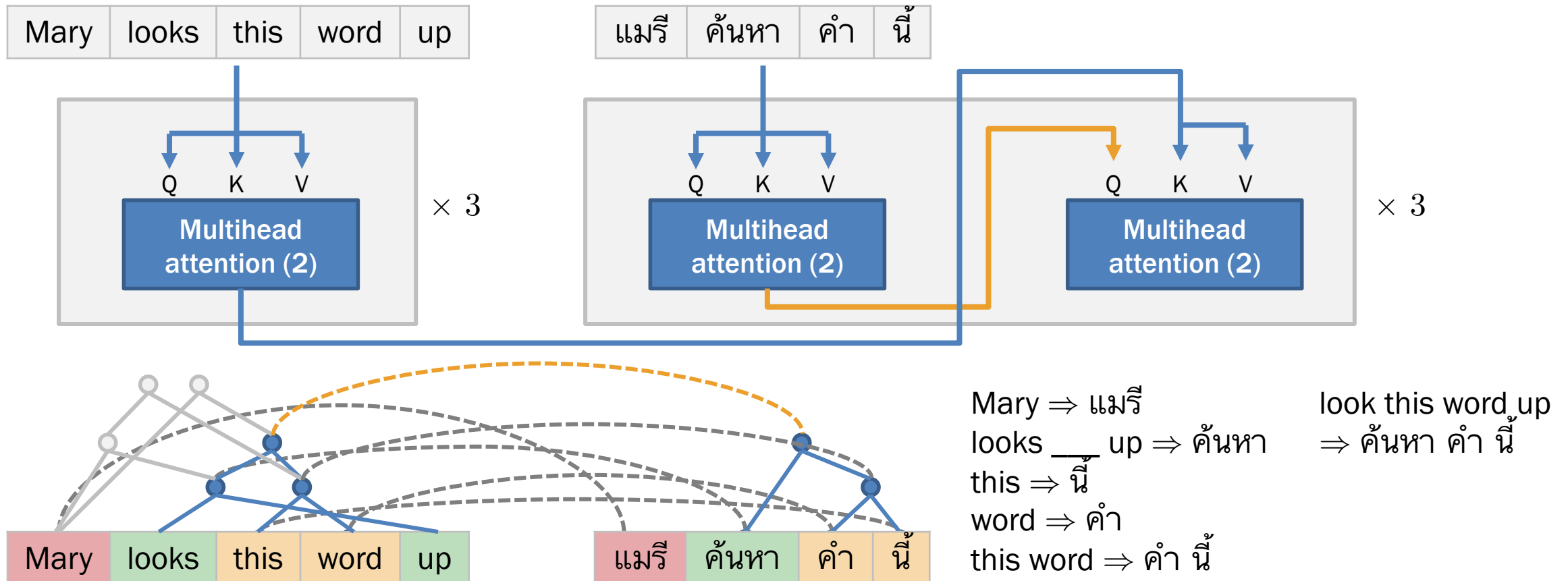
- $H$ -head alignment attention recognizes  $H$  pairs of phrase structures  $\Rightarrow$  **decoding layer**





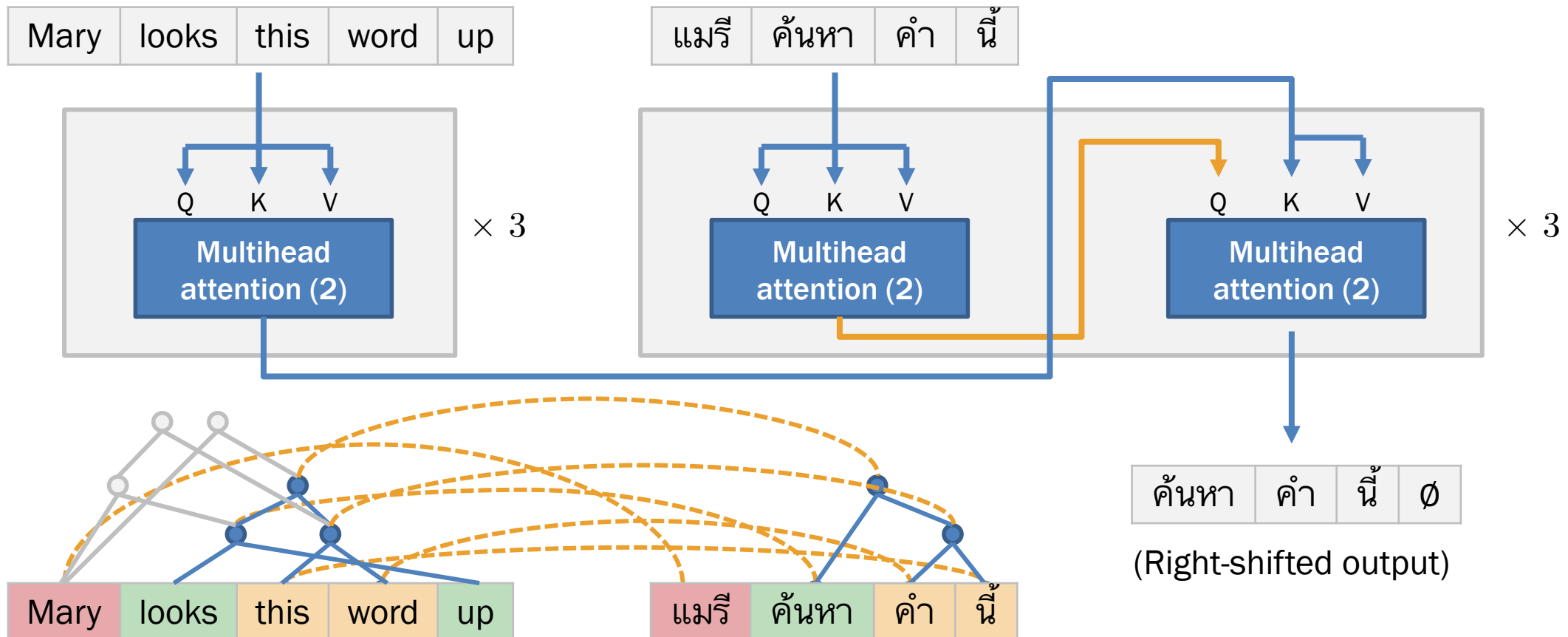
# Alignment of Phrase Structures

- $H$ -head alignment attention recognizes  $H$  pairs of phrase structures  $\Rightarrow$  **decoding layer**



# Alignment of Phrase Structures

- $H$ -head alignment attention recognizes  $H$  pairs of phrase structures  $\Rightarrow$  **decoding layer**



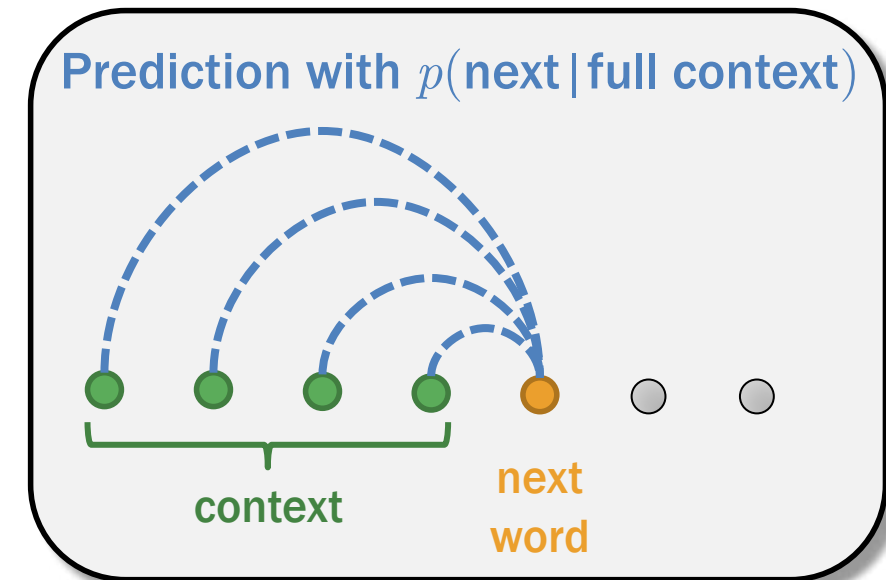
## 2. Large Language Models

# Language Models

- **Motivation**

- Statistical prediction for how strings are produced in a language
- Interpreted as a generative model
  1. Generate the first word  $w_1$
  2. Keep generating the **next word**  $w_k$  based on the previous words (a.k.a. **context**)  $w_1 \dots w_{k-1}$  until the whole sentence of length  $N$  is produced

$$P(w_1 \dots w_N) = p(w_1) \prod_{k=2}^N p(\overbrace{w_k}^{\text{next word}} | \overbrace{w_1 \dots w_{k-1}}^{\text{context}})$$



# Language Models

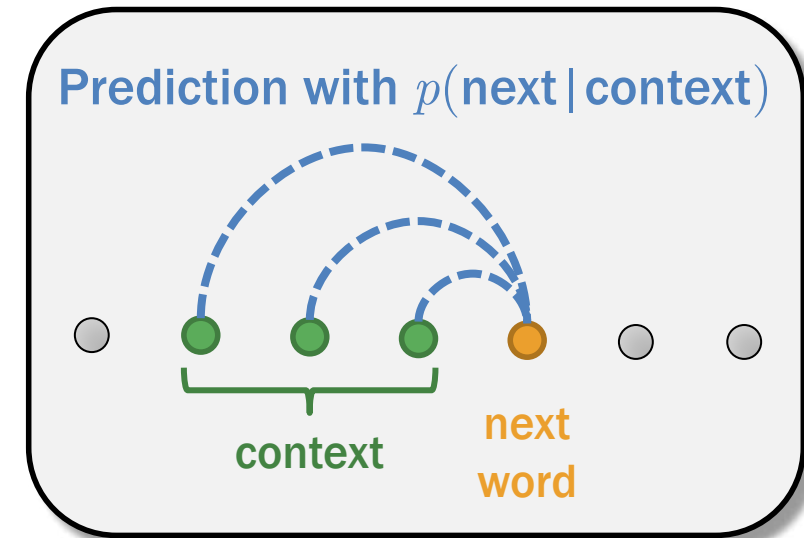
- **Motivation:**  $n$ -gram models
  - Language models whose context is truncated to at most  $n-1$  previous words

$$P(w_1 \dots w_N) = p(w_1) \prod_{k=2}^N p(\overbrace{w_k}^{\text{next word}} | \overbrace{w_{k-n+1} \dots w_{k-1}}^{n-1 \text{ prev words}})$$

- **Unigram** ( $n=1$ ):  $P(w_1 \dots w_N) = \prod_{k=1}^N p(w_k)$

- **Bigram** ( $n=2$ ):  $P(w_1 \dots w_N) = p(w_1) \prod_{k=2}^N p(w_k | w_{k-1})$

- **Trigram** ( $n=3$ ):  $P(w_1 \dots w_N) = p(w_1)p(w_2|w_1) \prod_{k=3}^N p(w_k | w_{k-2}, w_{k-1})$



# BERT (Devlin et al., 2018)

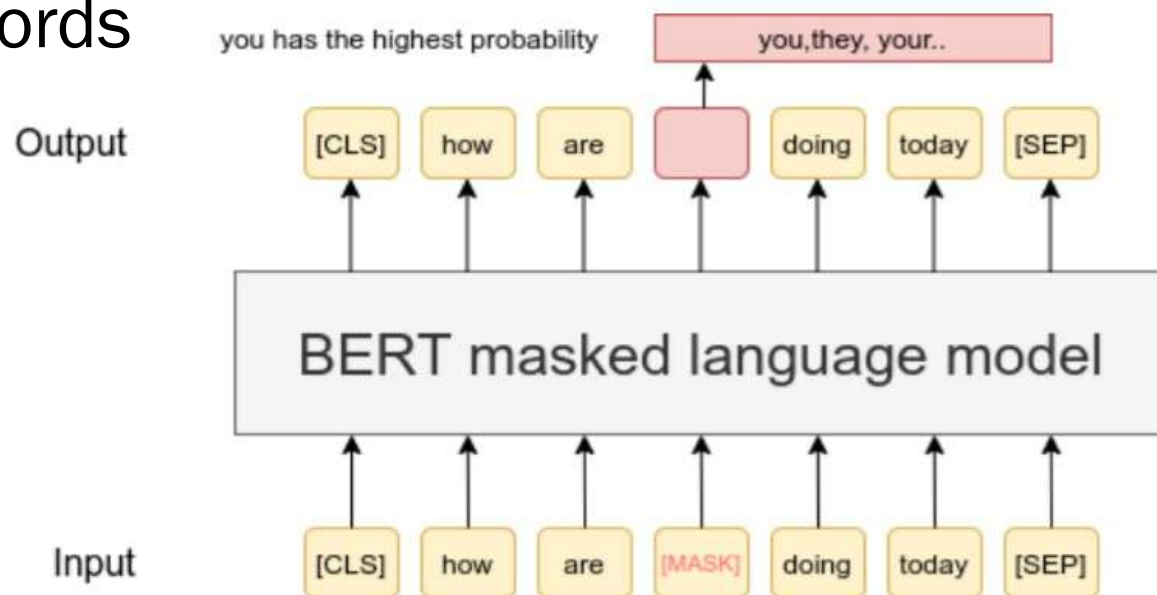
- Bidirectional Encoder Representations from Transformer
  - Transformer model pretrained with a very large language dataset
  - Pretrained Transformer model with multilayer bidirectional encoders
  - Contextual representations: vector repr of each word varies by position
  - Trained on BooksCorpus (800M words) + Wikipedia (2,500M words)

	BERT base	BERT large
Encoding layers	12	24
Attention heads	12	16
Hidden dimensions	768	1,024
Parameters	110M	340M

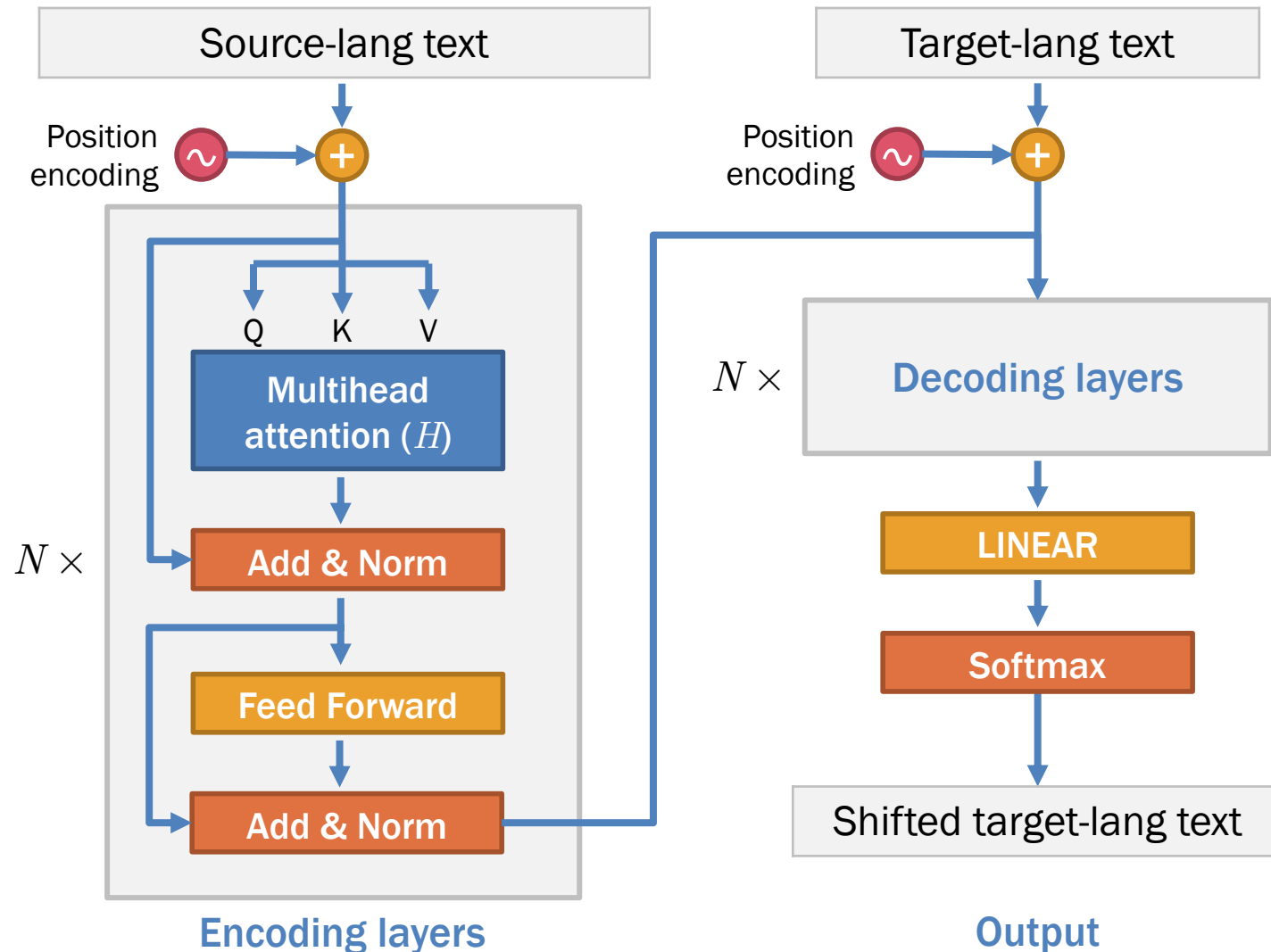


# Training BERT out of the Transformer

- **Masked Language Model (MLM):**
  - Each text is put on both source and target sides
  - Some words are blinded (or masked) at random
  - The model is trained to make correct prediction of masked words

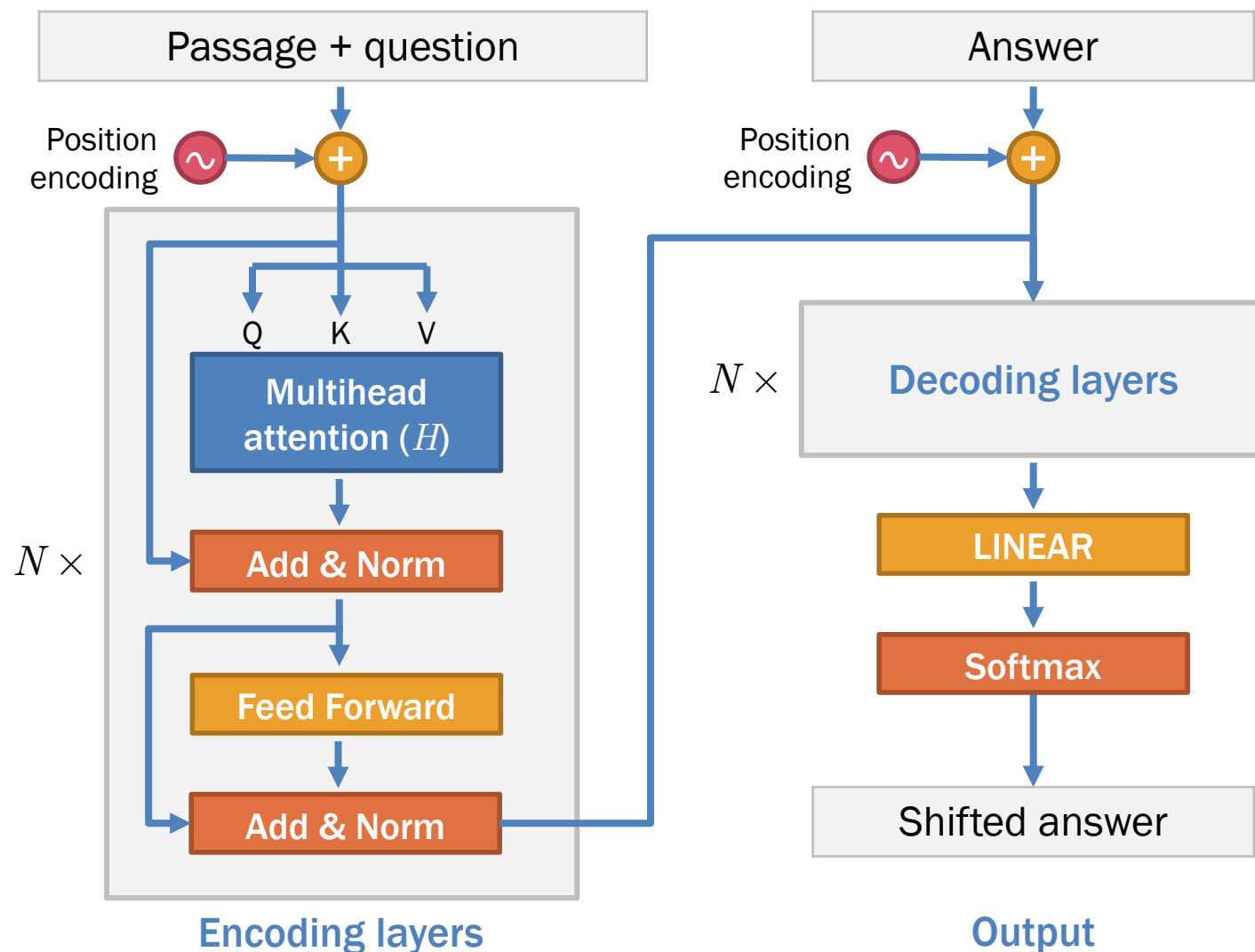


# Fine-Tuning



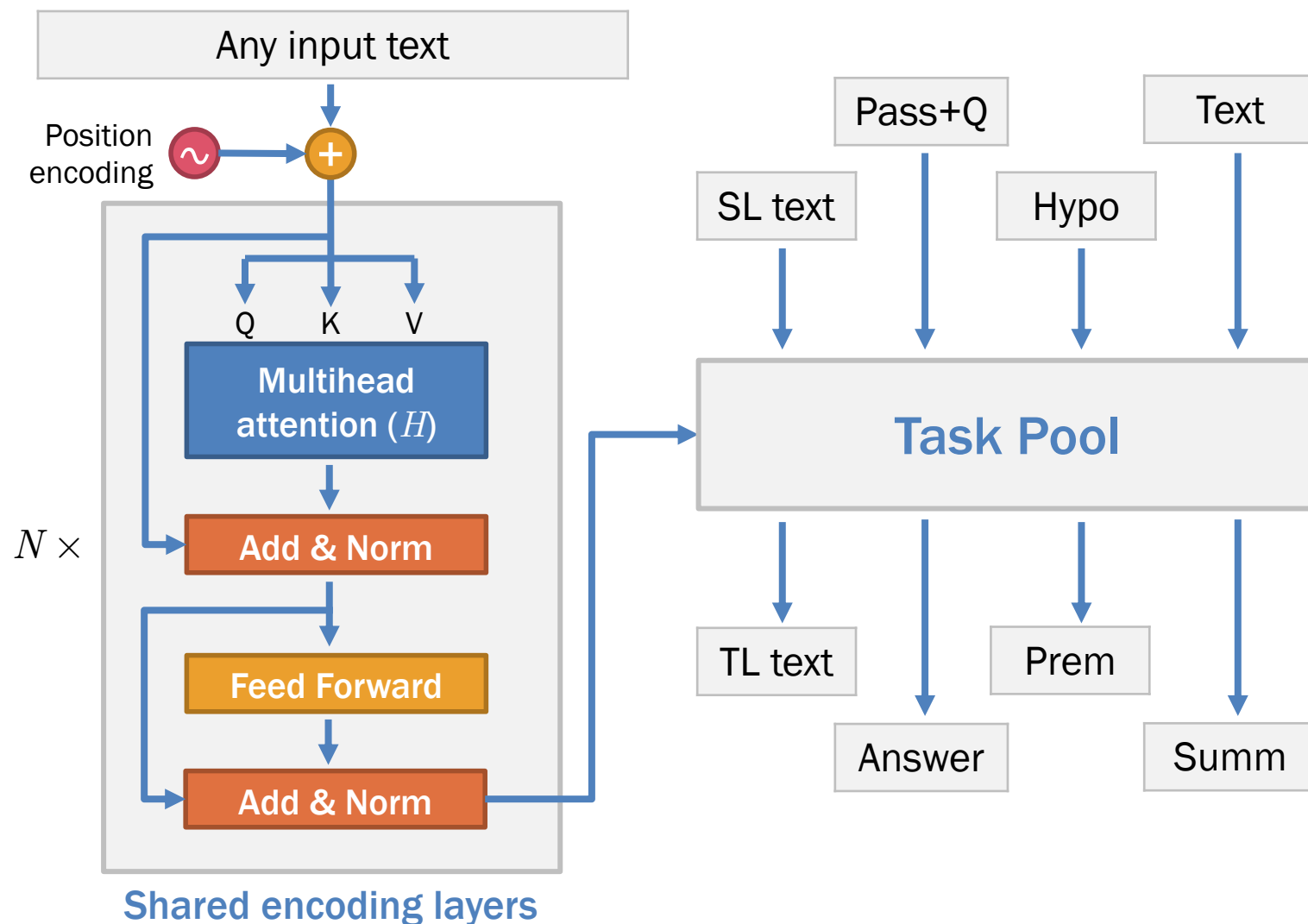
- BERT can be fine-tuned via multiple downstream tasks
  - **Machine translation**
  - Question answering (SQUAD)
  - Inference in natural language (NLI in GLUE Dataset)
  - Abstractive summarization

# Fine-Tuning



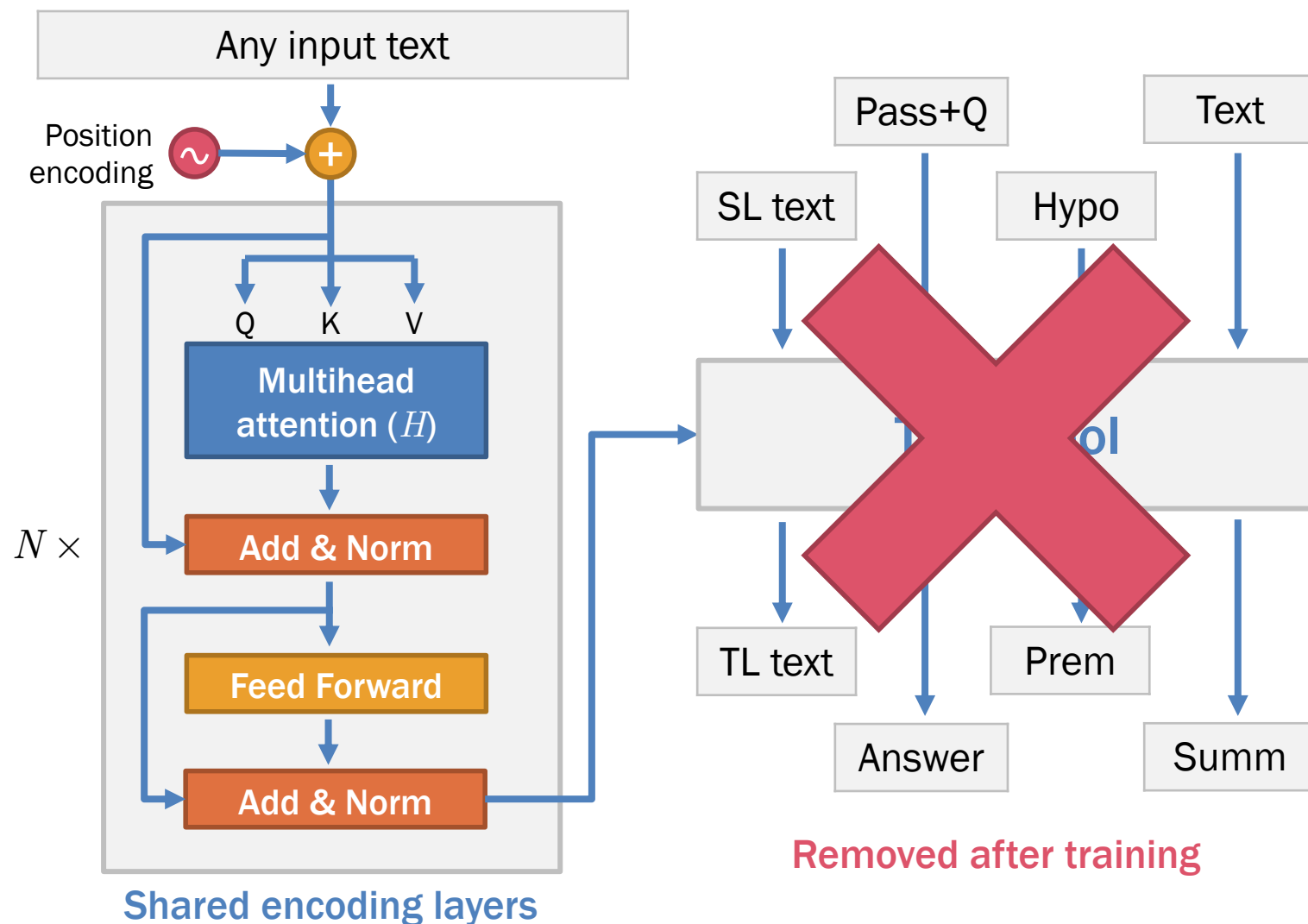
- BERT can be fine-tuned via multiple downstream tasks
  - Machine translation
  - **Question answering (SQUAD)**
  - Inference in natural language (NLI in GLUE Dataset)
  - Abstractive summarization

# Fine-Tuning



- BERT can be fine-tuned via multiple downstream tasks
  - Machine translation
  - Question answering (SQUAD)
  - Inference in natural language (NLI in GLUE Dataset)
  - Abstractive summarization

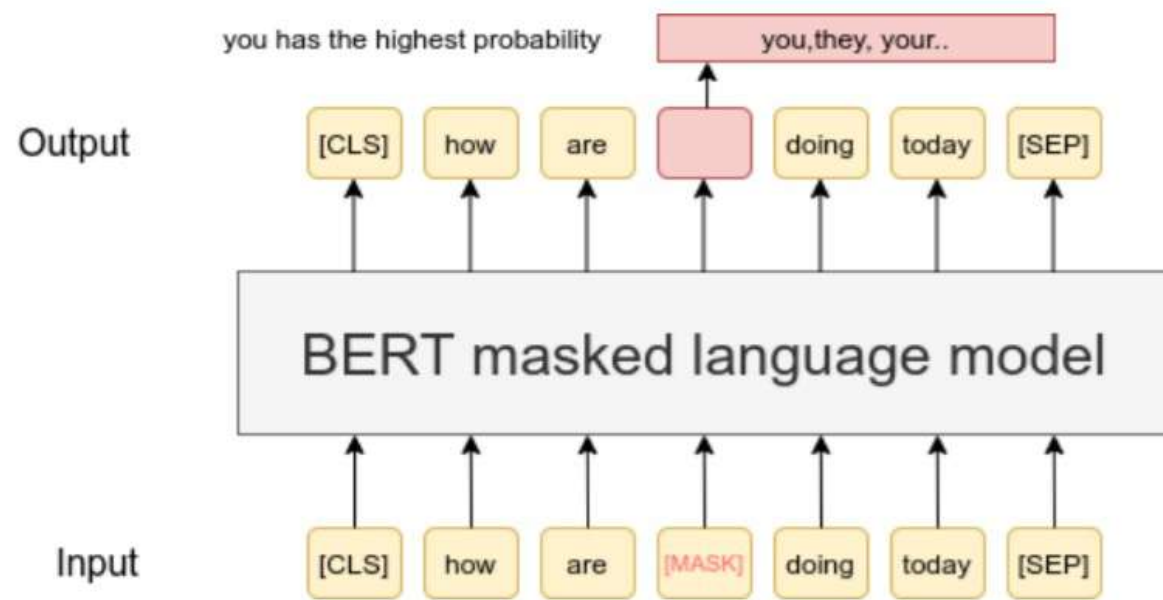
# BERT = Shared Encoding Layers



- BERT can be fine-tuned via multiple downstream tasks
  - Machine translation
  - Question answering (SQUAD)
  - Inference in natural language (NLI in GLUE Dataset)
  - Abstractive summarization

# RoBERTa (Liu et al., 2019)

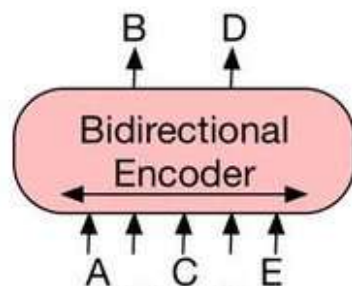
- Robustly Optimized BERT pretraining approach
  - An improved version of BERT



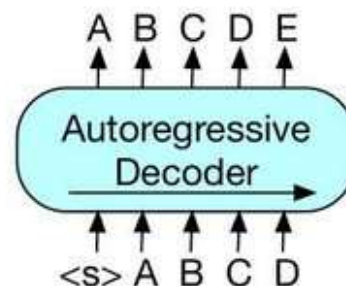
- Dynamic masking instead of static masking
- NSP task is eliminated without losing semantic relatedness
- Larger datasets are used in training than BERT (CC-News and Open WebText)



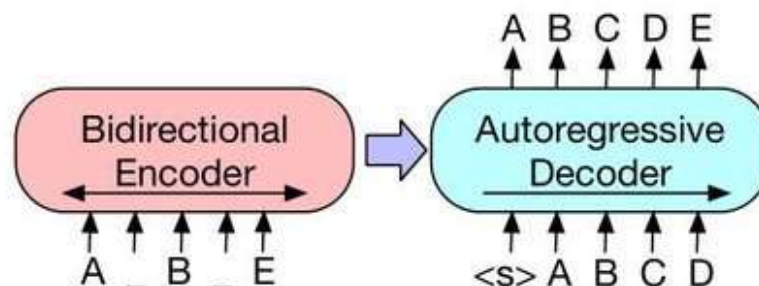
# Differences of BERT, GPT, and BART (Lewis et al., 2019)



(a) BERT: Random tokens are replaced with masks, and the document is encoded bidirectionally. Missing tokens are predicted independently, so BERT cannot easily be used for generation.



(b) GPT: Tokens are predicted auto-regressively, meaning GPT can be used for generation. However words can only condition on leftward context, so it cannot learn bidirectional interactions.



(c) BART: Inputs to the encoder need not be aligned with decoder outputs, allowing arbitrary noise transformations. Here, a document has been corrupted by replacing spans of text with a mask symbols. The corrupted document (left) is encoded with a bidirectional model, and then the likelihood of the original document (right) is calculated with an autoregressive decoder. For fine-tuning, an uncorrupted document is input to both the encoder and decoder, and we use representations from the final hidden state of the decoder.

- **BERT:**
  - Bidirectional encoder
- **GPT:**
  - Google's Pretrained Transformer
  - Autoregressive (unidirectional) decoder
- **BART:**
  - Bidirectional encoder + autoregressive decoder

# PaLM (Wei et al., 2022)

- Pathways Language Model
  - 540B parameters trained with Pathways Transformer model (single model capable of doing many tasks)
  - Chain of thoughts prompting for multistep problem solving

## 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. ✅

- Chain of thoughts is added to the target sequence
- Reasoning and answering are learned in tandem
- Interpretability is introduced to Transformer

# PaLM (Wei et al., 2022)

- Chain of thoughts

## Math Word Problems (free response)

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.

## Math Word Problems (multiple choice)

Q: How many keystrokes are needed to type the numbers from 1 to 500?  
Answer Choices: (a) 1156 (b) 1392 (c) 1480 (d) 1562 (e) 1788

A: There are 9 one-digit numbers from 1 to 9. There are 90 two-digit numbers from 10 to 99. There are 401 three-digit numbers from 100 to 500.  $9 + 90(2) + 401(3) = 1392$ . The answer is (b).

## CSQA (commonsense)

Q: Sammy wanted to go to where the people were. Where might he go?  
Options: (a) race track (b) populated areas (c) desert (d) apartment (e) roadblock

A: The answer must be a place with a lot of people. Race tracks, desert, apartments, and roadblocks don't have a lot of people, but populated areas do. So the answer is (b).

## StrategyQA

Q: Yes or no: Would a pear sink in water?

A: The density of a pear is about  $0.6 \text{ g/cm}^3$ , which is less than water. Thus, a pear would float. So the answer is no.

## Date Understanding

Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?

A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943.

## Sports Understanding

Q: Is the following sentence plausible? "Joao Moutinho caught the screen pass in the NFC championship."

A: Joao Moutinho is a soccer player. The NFC championship is part of American football, not soccer. So the answer is no.

## SayCan (Instructing a robot)

Human: How would you bring me something that isn't a fruit?

Explanation: the user wants something to eat that isn't a fruit. An energy bar is not a fruit, so I will bring the user an energy bar.

Plan: 1. find(energy bar) 2. pick(energy bar) 3. find(user) 4. put(energy bar) 5. done().

## Last Letter Concatenation

Q: Take the last letters of the words in "Lady Gaga" and concatenate them.

A: The last letter of "Lady" is "y". The last letter of "Gaga" is "a". Concatenating them is "ya". So the answer is ya.

## Coin Flip (state tracking)

Q: A coin is heads up. Maybelle flips the coin. Shalonda does not flip the coin. Is the coin still heads up?


A: The coin was flipped by Maybelle. So the coin was flipped 1 time, which is an odd number. The coin started heads up, so after an odd number of flips, it will be tails up. So the answer is no.

## PaLM-E (Driess et al., 2023)

- Embodied multimodal PaLM with 562B parameters

## Mobile Manipulation



Human: Bring me the rice chips from the drawer. Robot: 1. Go to the drawers, 2. Open top drawer. I see  3. Pick the green rice chip bag from the drawer and place it on the counter.

## Visual Q&A, Captioning ...



Given `<img>`. Q: What's in the image? Answer in emojis.

A: 🍏 🍌 🍇 🍐 🍑 🍈 🍓.

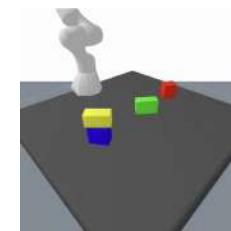


Describe the following `<img>`:  
A dog jumping over a hurdle at a dog show.

## Language Only Tasks

Here is a Haiku about embodied language models:  
Embodied language  
models are the future of  
natural language


## Task and Motion Planning



Given **<emb>** Q: How to grasp blue block?  
A: First grasp yellow block and place it on the table, then grasp the blue block.

## Tabletop Manipulation

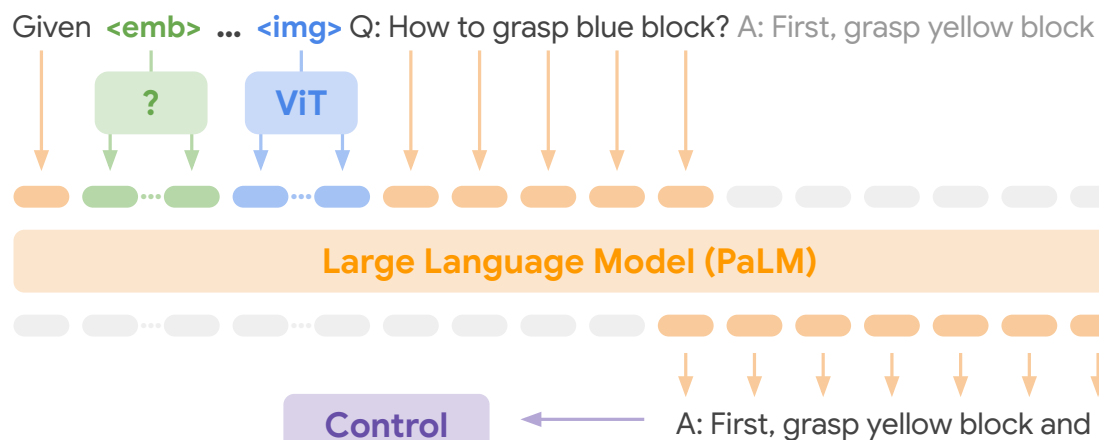


Given  Task: Sort colors into corners.

Step 1. Push the green star to the bottom left.

Step 2. Push the green circle to the green star.

# PaLM-E: An Embodied Multimodal Language Model






# PaLM-E (Driess et al., 2023)

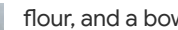
## Visually-conditioned Jokes + Few-shot prompting



I see (image omitted). Description: A cow sitting on a rug. Joke: Why is the cow sitting on a rug? Because it wants to be a cow-ch! ... I see . Description: a donkey is carrying a dog, cat, and rooster. Joke: what do you call a donkey with a rooster on his back. A rooster booster.


## Robot Visual Perception, Dialogue, and Planning




Robot: I am a robot operating in a kitchen. Given , what do I see? a bowl of eggs, a bowl of flour, and a bowl of sugar.. and a spatula.. and a spoon.. and a bag of flour.. and a bag of sugar. When a human asks me to do a task, I will respond with the sequence of actions I would do to accomplish the task with only the items I see. Human: Use all of the ingredients you see to make a cake batter. 1. crack egg. 2. put egg in bowl. 3. put flour in bowl. 4. put sugar in bowl. 5. mix. 6. put in pan. 7. bake. 8. eat. 9. clean up.

## Zero-shot Multimodal Chain-of-Thought




Given . Q: Can I go down this street on a bicycle, yes or no? A: Let's think step by step. 1. do not enter. 2. except bicycles. 3. do not entry except bicycles. 4. yes.




Given . Q: How many championship rings did the player on the left win in his career A: Let's think step by step. 1. kobe bryant is on the left. 2. kobe bryant has won 5 championship rings.



Given . Q: Who are the two teams playing in this photo, and which was the last to win a championship, and which year did they win, and who was their star player that year? A: Let's think step by step. 1. team in white is new york knick. 2. team in green is boston celtic. 3. last to win championship was boston celtic. 4. year was 2008. 5. star player was paul pierce. 6. i am not sure if this is correct. i am not basketball fan. i just google it.



## Robot Initiative



Given . Q: If a robot wanted to be useful here, what steps should it take? A: clean table..pick up trash.. pick up chairs.. wipe chairs.. put chairs down.


## Zero-shot: Multi-image Relationships



Q: Where is  in this scene ? A: a stove top with a pot on it.

## Physical Prediction



Given . Q: What will the robot do next? A: fall.

## Spatial Grounding




Given . Q: What flavor is the doughnut on the right? Strawberry.. Q: What flavor is the doughnut on the left? A blueberry.





Photo 1: . Photo 2: . Q: What is in Photo 1 but NOT in Photo 2? A: sunglasses.

Photo 1: . Photo 2: . Q: What is in Photo 1 but NOT in Photo 2? A: Let's think step by step and explain. Photo 1 has sunglasses on top of folded clothes. Photo 2 does not have sunglasses on top of folded clothes. Therefore, sunglasses are in Photo 1 but not in Photo 2.

## End-to-End Egocentric Q&A, Dialogue

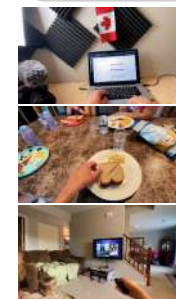






Photo 1, at 10:30 am: .  
Photo 2, at 12:45 pm: .  
Photo 3, at 3:45 pm: .

Q: I forget, what did I have for lunch, and what time was it? A: Let's think step by step.  
1. you had a sandwich for lunch.  
2. it was 12:45 pm.

## OCR-free Math Reasoning



Given . Q: I am getting just two custom pizza for me and my friend. How much should I pay in total? A: Let's think step by step. 1st step: 2 custom pizzas. 2nd step: 2 x \$9.99. 3rd step: \$19.98.

# 3. Training Technique



# Instruction Dataset (Ouyang et al., 2022)

Table 1: Distribution of use case categories from our API prompt dataset.

Use-case	(%)
Generation	45.6%
Open QA	12.4%
Brainstorming	11.2%
Chat	8.4%
Rewrite	6.6%
Summarization	4.2%
Classification	3.5%
Other	3.5%
Closed QA	2.6%
Extract	1.9%

Table 2: Illustrative prompts from our API prompt dataset. These are fictional examples inspired by real usage—see more examples in Appendix A.2.1.

Use-case	Prompt
Brainstorming	List five ideas for how to regain enthusiasm for my career
Generation	Write a short story where a bear goes to the beach, makes friends with a seal, and then returns home.
Rewrite	This is the summary of a Broadway play: "" { summary } "" This is the outline of the commercial for that play: ""

- Prompt: “instruction”
- Cleverly designed set of instructions and responses for a chatbot
- Covering frequently asked questions and their answers

Use Case	Example
brainstorming	List five ideas for how to regain enthusiasm for my career
brainstorming	What are some key points I should know when studying Ancient Greece?
brainstorming	What are 4 questions a user might have after reading the instruction manual for a trash compactor?  { user manual }  1.

# User Prompts and Chats (Ouyang et al., 2022)

closed qa	<p>Answer the following question: What shape is the earth?</p> <p>A) A circle B) A sphere C) An ellipse D) A plane</p>
closed qa	<p>Tell me how hydrogen and helium are different, using the following facts:</p> <p>{list of facts}</p>
open qa	<p>I am a highly intelligent question answering bot. If you ask me a question that is rooted in truth, I will give you the answer. If you ask me a question that is nonsense, trickery, or has no clear answer, I will respond with "Unknown".</p> <p>Q: What is human life expectancy in the United States? A: Human life expectancy in the United States is 78 years.</p> <p>Q: Who was president of the United States in 1955? A:</p>
open qa	Who built the statue of liberty?
open qa	How do you take the derivative of the sin function?
open qa	who are the indiginous people of New Zealand?

# User Prompts and Chats (Ouyang et al., 2022)

Use Case	Example
chat	<p>The following is a conversation with an AI assistant. The assistant is helpful, creative, clever, and very friendly.</p> <p>Human: Hello, who are you? AI: I am an AI created by OpenAI. How can I help you today? Human: I'd like to cancel my subscription. AI:</p>
chat	<p>Marv is a chatbot that reluctantly answers questions with sarcastic responses:</p> <p>You: How many pounds are in a kilogram? Marv: This again? There are 2.2 pounds in a kilogram. Please make a note of this. You: What does HTML stand for? Marv: Was Google too busy? Hypertext Markup Language. The T is for try to ask better questions in the future. You: When did the first airplane fly? Marv:</p>
chat	<p>This is a conversation with an enlightened Buddha. Every response is full of wisdom and love.</p> <p>Me: How can I achieve greater peace and equanimity? Buddha:</p>

# User Prompts and Chats (Ouyang et al., 2022)

Use Case	Example
classification	<p>The following is a list of companies and the categories they fall into:</p> <p>Apple, Facebook, Fedex</p> <p>Apple Category: Technology</p> <p>Facebook Category: Social Media</p> <p>Fedex Category:</p>
extract	<p>Text: {text} Keywords:</p>
generation	<p>"Hey, what are you doing there?" Casey was startled. He hadn't even begun to</p>
generation	<p>The name of the next Star Wars movie is</p>
generation	<p>This is the research for an essay: === {description of research} === Write a high school essay on these topics: ===</p>

# Prompts + Human Ranking + RL

## Step 1

Collect demonstration data and train a supervised policy.

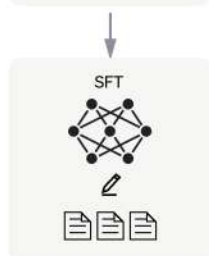
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



This data is used to fine-tune GPT-3.5 with supervised learning.



## Step 2

Collect comparison data and train a reward model.

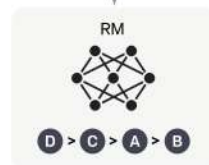
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



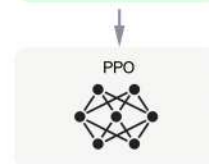
## Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

A new prompt is sampled from the dataset.



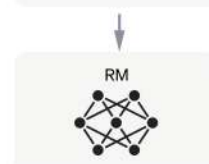
The PPO model is initialized from the supervised policy.



The policy generates an output.



The reward model calculates a reward for the output.



The reward is used to update the policy using PPO.



- 3 steps
  - Fine-tune the language model with the instruction dataset
  - Retrain the reward model for chat response with human ranking
  - Optimize the policy model w.r.t. the reward model with the PPO Algorithm (proximal policy optimization)

## 4. Business Opportunities in AI

# Generative AI

- Algorithms (such as ChatGPT) that can be used to create new content, including audio, code, images, text, simulations, and videos
- Popular generative AI models
  - ChatGPT for text (OpenAI)
  - DALL-E and Stable Diffusion (MidJourney) for image
  - VALL-E for voice (Microsoft)



# Krisp.ai

krisp

Product

Solutions ▾

Pricing

Blog

For Contact Centers

[Sign in](#)

Get Krisp for Free

## Speak with confidence in business calls

Krisp's AI removes background voices, noises and echo from all your calls, giving you peace of mind.

Get Krisp for Free



✓ No credit card needed

✓ Free forever



- Krisp.ai removes background voices, noises, and echoes from your conference calls with AI
- <https://krisp.ai>

# Beatoven.ai

beatoven.ai   How It Works   Pricing   Artists   About us   Blog   Sign in   Sign up

How it works

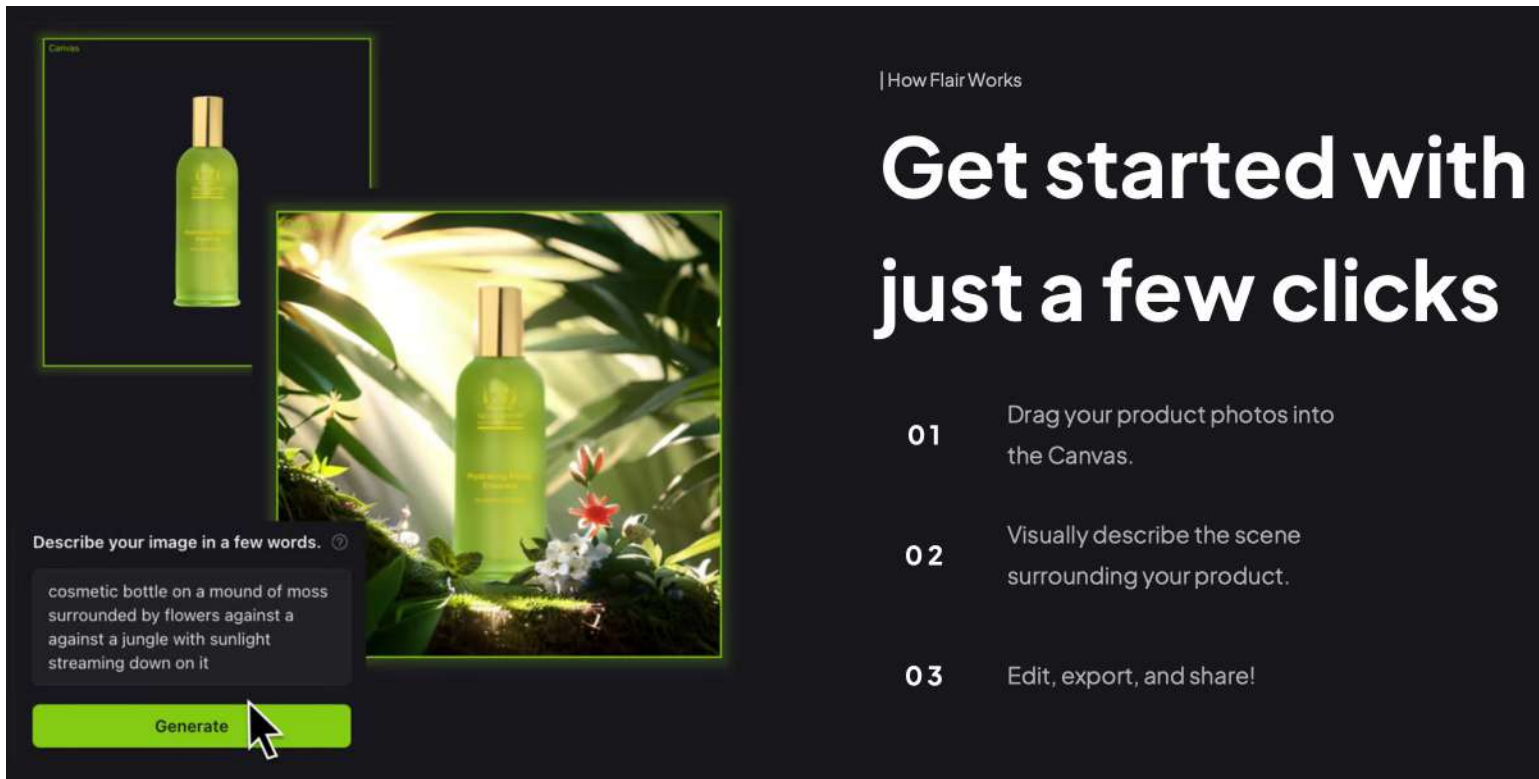
Compose unique music for your content in a few easy steps

- 1. Pick a Genre/Style**  
Upload a video/podcast or start with a track, choose from 8 different Genres to suit your theme.
- 2. Make some cuts**  
We know that the mood of the content changes over time. That's why you can make multiple cuts to add different moods.
- 3. Change Mood**  
Now, choose from a rich selection of 16 moods to choose the correct mood for the cuts.
- 4. Compose**  
Hit compose and let our AI do all the hard work of composing a unique track for you.

Start Creating for Free

- Beatoven.ai creates loyalty-free, customizable, and unique musics
- <https://www.beatoven.ai>

# Flair.ai



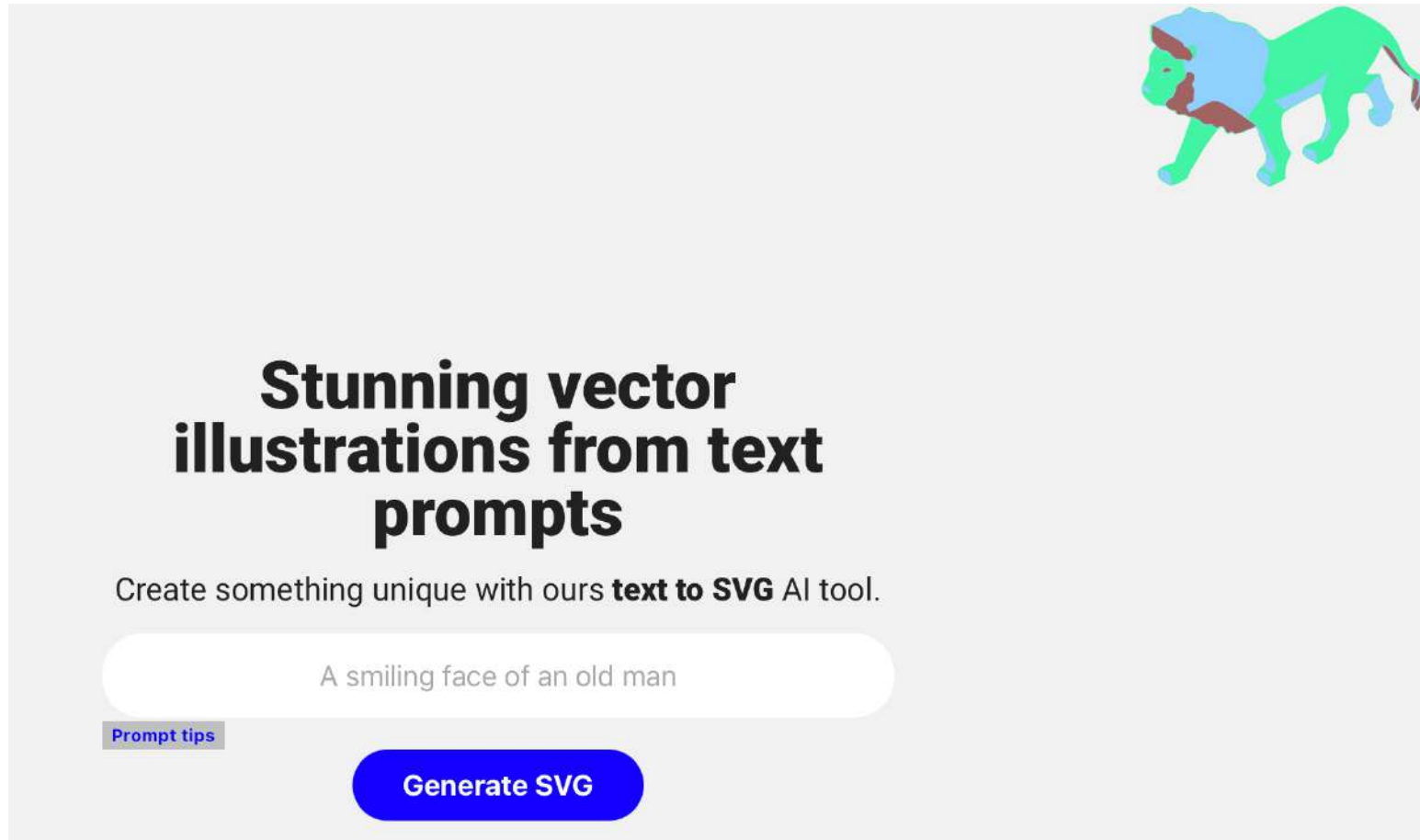
The screenshot displays the Flair.ai web interface. On the left, a 'Canvas' area shows a green cosmetic bottle. Below it, a text input field contains the description: 'cosmetic bottle on a mound of moss surrounded by flowers against a jungle with sunlight streaming down on it'. A green 'Generate' button is at the bottom left. To the right, a larger image shows the generated advertisement: the same green bottle on a mossy mound with flowers, set against a lush jungle background with sunlight filtering through the leaves. The text 'How Flair Works' is visible above the main heading.

## Get started with just a few clicks

- 01 Drag your product photos into the Canvas.
- 02 Visually describe the scene surrounding your product.
- 03 Edit, export, and share!

- Flair.ai generates an advertisement from your idea draft and visual elements
- <https://flair.ai>

# Illustroke.com



- Illustroke.com converts your text prompt into a vector illustration (SVG)
- <https://illustroke.com>

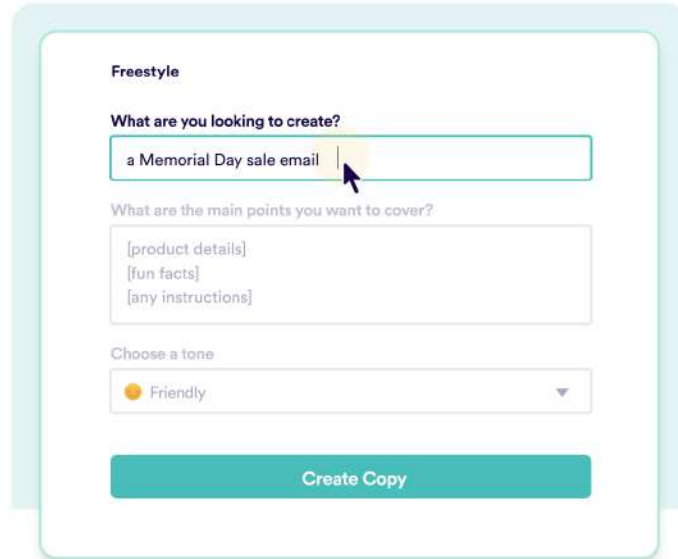
# Copy.ai

copy.ai [We're Hiring!](#) [Teams](#) [Use Cases](#) [Resources](#) [Pricing](#)

Login

Get Started — It's Free

## How it works



The screenshot shows the 'Freestyle' section of the Copy.ai interface. It includes a text input field for 'What are you looking to create?' with the text 'a Memorial Day sale email'. Below this is a text area for 'What are the main points you want to cover?' containing '[product details]', '[fun facts]', and '[any instructions]'. There is also a dropdown menu for 'Choose a tone' with 'Friendly' selected. A teal 'Create Copy' button is at the bottom.

### 1 Enter your copywriting project

Choose from emails, social posts, long-form blog posts, and more!

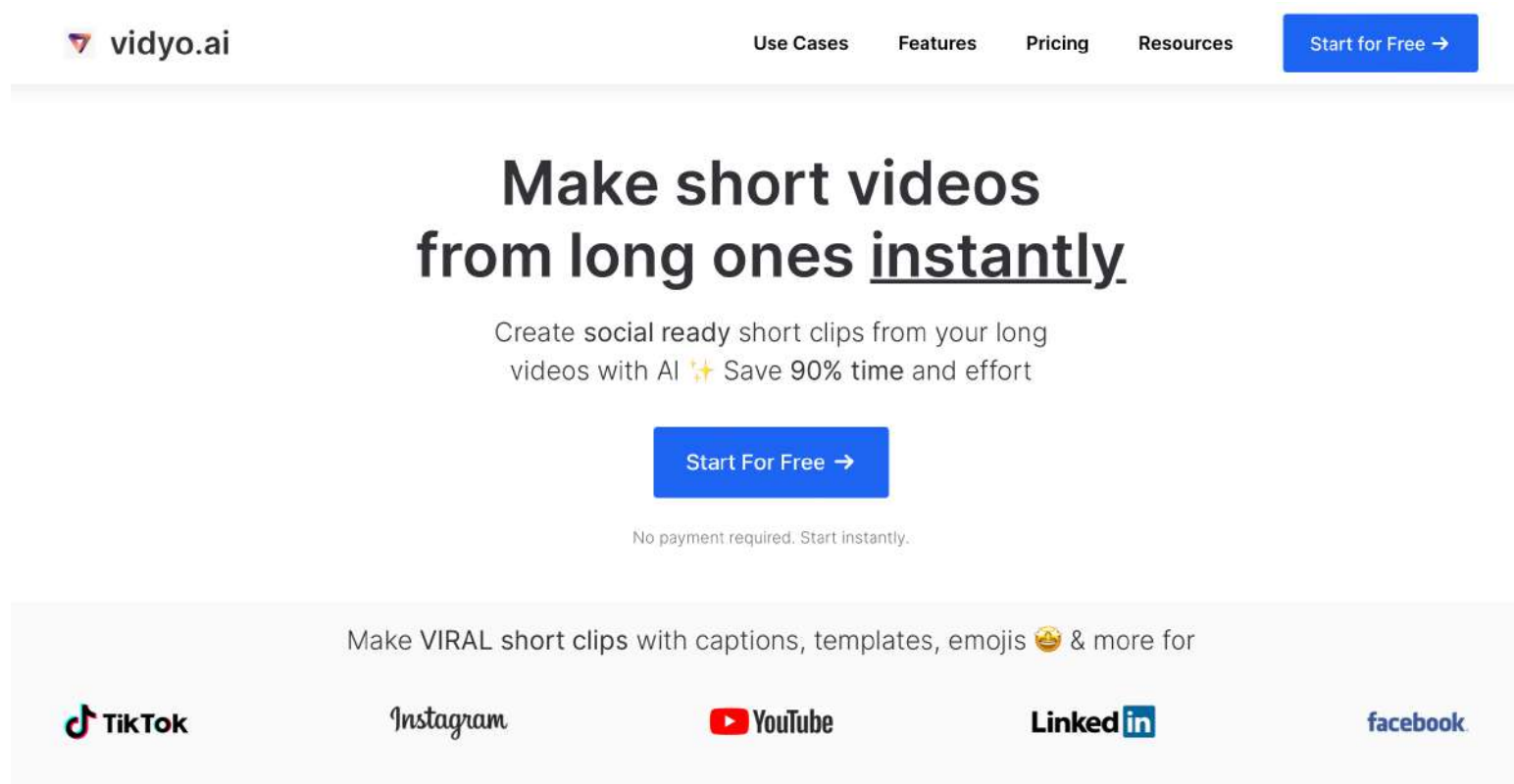
- Copy.ai generates a text content based on your context (main points) and tone
- <https://www.copy.ai>

# Ocoya.com



- Ocoya.com generates social media and hashtags by your headline and context
- It posts the contents on the optimal time
- <https://www.ocado.com>

# Vidyo.ai



The screenshot shows the Vidyo.ai website. At the top, there is a navigation bar with the Vidyo.ai logo on the left and links for 'Use Cases', 'Features', 'Pricing', 'Resources', and a 'Start for Free →' button on the right. The main content area features a large heading 'Make short videos from long ones instantly' in bold black text. Below this, a subtext reads 'Create social ready short clips from your long videos with AI ✨ Save 90% time and effort'. A blue button labeled 'Start For Free →' is centered below the subtext. Underneath the button, it says 'No payment required. Start instantly.' At the bottom of the main content area, there is a light gray box with the text 'Make VIRAL short clips with captions, templates, emojis 🥰 & more for' followed by five social media icons: TikTok, Instagram, YouTube, LinkedIn, and Facebook.

vidyo.ai

Use Cases Features Pricing Resources Start for Free →

## Make short videos from long ones instantly

Create social ready short clips from your long videos with AI ✨ Save 90% time and effort

Start For Free →

No payment required. Start instantly.

Make VIRAL short clips with captions, templates, emojis 🥰 & more for

TikTok Instagram YouTube LinkedIn facebook

Loved by 30K+ podcasters & creators

- Vidyo.ai generates short viral clips with emojis and captions from a given video footage
- <https://vidyo.ai>

<https://tinyurl.com/2fj5y3ny>



# Thank You