POST 18 - GEN AI

UNDERSTANDING THE CONCEPT OF "MULTI-HEAD ATTENTION" IN GENERATIVE AI & LLM

Generative AI Deep Dives,
Key concepts for Transformers - Part 5

GENERATIVE AI FOR ALL

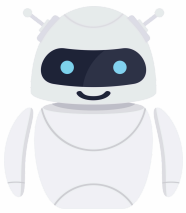
DINESH LAL
(DIRECTOR, DATA SCIENCE)

DINESHLAL

WHAT IS COVERED IN THIS DOCUMENT?

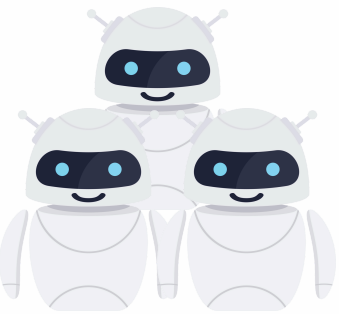
- This document explains the topic **Multi-Head Attention**
- First we will transition from self attention to Multi-Head Attention
- Document takes 3 different example of English sentences, and explains the concept with details
- It covers key process, by breaking down one of the example with approach
- Finally definition and summary are shared

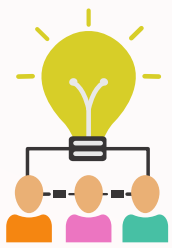
SELF ATTENTION TO MULTI-HEAD ATTENTION



- **Self Attention** which we covered in previous post is the building block of Multi Head Attention
- Self Attention Allows the Model to analyze the surrounding words to go beyond the meaning of individual words

- **Multi-head attention** takes this a step further.
- It allows to attend to the sentence in multiple ways simultaneously,
  - like focusing on grammar,
  - factual relationships,
  - or stylistic flow, all at once.





EXPLANATION WITH EXAMPLE

1

Consider this sentence

"The chef baked a delicious cake for the birthday party"

- **Head 1 (Focuses on Agent-Object Relationship):** This head prioritizes the connection between the person performing an action and the object affected by the action.
  - It would assign high attention scores to "chef" and "baked," "baked" and "cake," etc.

EXPLANATION WITH EXAMPLE

1

Consider this sentence

"The chef baked a delicious cake for the birthday party"

- **Head 2 (Focuses on Occasion and Purpose):** This head concentrates on words that indicate the context or reason for the action.
  - It would give high attention to "baked" and "cake," "cake" and "birthday party," etc.

EXPLANATION WITH EXAMPLE

1

Consider this sentence

"The chef baked a delicious cake for the birthday party"

- **Head 3 (Focuses on Adjective-Noun Relationship):** This head focuses on how adjectives describe nouns.
  - It would assign high attention scores to "baked" and "delicious," "cake" and "delicious," etc.

EXPLANATION WITH EXAMPLE

2

Consider this sentence

"The programmer struggled to debug the code"

- **Head 1 (Focuses on Agent-Action Relationship):** This head prioritizes the connection between who performs the action and the action itself.
  - It would assign high attention scores to "programmer" and "struggled," "programmer" and "debug," etc.

EXAMPLE

EXPLANATION WITH EXAMPLE

2

Consider this sentence

"The programmer struggled to debug the code"

- **Head 2 (Focuses on Difficulty):** This head concentrates on words that indicate the level of difficulty or challenge.
  - It would assign high attention to "struggled" and "debug," "struggled" and "difficult" (if present in the context), etc.

EXPLANATION WITH EXAMPLE

2

Consider this sentence

"The programmer struggled to debug the code"

- **Head 3 (Focuses on Problem-Solution):** This head focuses on the relationship between the problem and the attempted solution.
  - It would consider the connection between "debug" and "code," highlighting the issue the programmer is trying to fix.

EXAMPLE

EXPLANATION WITH EXAMPLE

3

Consider this sentence

"The rain in Spain falls mainly on the plain"

- **Head 1 (Focuses on Subject-Verb Agreement):** This head prioritizes the grammatical agreement between subject and verb.
  - It would give high attention scores to "rain" and "falls," "Spain" and "falls," etc.

EXPLANATION WITH EXAMPLE

3

Consider this sentence

"The rain in Spain falls mainly on the plain"

- **Head 2 (Focuses on Rhyme and Meter):** This head concentrates on the poetic aspects of the sentence, looking for patterns in sounds and rhythm.
  - It would assign high attention to "rain" and "Spain," "mainly" and "plain," etc., due to the rhyming words.

EXAMPLE

EXPLANATION WITH EXAMPLE

3

Consider this sentence

"The rain in Spain falls mainly on the plain"

- **Head 3 (Focuses on Location):** This head focuses on the geographical relationship between the elements.
  - It would give high attention to "rain" and "Spain," "in Spain" and "plain," highlighting the location where the rain falls.

EXAMPLE

BREAKING DOWN THE PROCESS

1. **Embedding:** Each word is converted into a vector representing its meaning.
2. **Head Splitting:** These embeddings are projected into separate queries, keys, and values for each of the three heads.
3. **Independent Attention:** Each head calculates attention scores based on its focus:
   - Head 1: "rain" (query) with "falls" (key) gets a high score, ensuring subject-verb agreement.
   - Head 2: "rain" (query) with "Spain" (key) gets a high score, identifying the rhyming pattern.
   - Head 3: "rain" (query) with "in Spain" (key) gets a high score, focusing on the location of the rain.

BREAKING DOWN THE PROCESS

4 **Head Outputs:** Each head generates a context vector for each word, indicating relevant information based on its focus (grammar, rhyme, or location).

5 **Combining Heads:** The outputs from all three heads (context vectors for each word) are concatenated.

6 **Final Output:** This combined output is processed by another layer to create the final representation of the sentence, incorporating information from all three attention perspectives.

DEFINITION

- Multi-head attention is a mechanism within Transformer models that allows the model to jointly attend to information from different representation subspaces at different positions.
- It does this by performing the attention function in parallel multiple times, with each "head" focusing on different parts of the input.
- This enables the model to capture various aspects of the input data, such as different types of relationships or features.

DEFINITION

- Imagine you're reading a complicated paragraph in a book. Multi-head attention is like having multiple super-powered highlighters that can focus on different things at the same time. Here's the idea:
- Regular Highlighter: A normal highlighter marks one thing, like the main ideas.
- Multi-Head Highlighter: This special highlighter has several tips, each focusing on a different aspect of the text.

Thank You

SPECIAL THANKS TO CHATGPT, OPEN AI, COPILOT, GEMINI FOR THE SUPPORT ON CONTENT

DINESHLAL