Summary and Highlights

Congratulations! You have completed this lesson. At this point in the course, you know that:

- Positional encoding incorporates information about the position of each embedding within a sequence.

- Attention mechanisms employ the query, key, and value matrices.

- You can apply the attention mechanism to word embeddings. This process helps capture contextual relationships between words.

- Simple language modeling in the self-attention mechanism predicts the next word in the sentence.

- The scaled dot-product attention mechanism involves a series of matrix multiplications that incorporate queries, keys, and a scaling factor. A masking operation may be employed, and the values are multiplied to produce the output.

- It is possible to retain context while classifying text by integrating transformer attention layers.