ANU-SEARCH THE INTERNATIONAL JOURNAL OF PHARMA AND ALLIED SCIENCES

Research article

(*[ISSN: XXXX-XXXX] Journal Homepage:https://anusearch-ijpas.com/*)

Artificial Intelligence / Machine Learning / Natural Language Processing

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, ?ukasz Kaiser, Illia Polosukhin

ARTICLE INFO

Received On: 29-07-25

Modified On: 29-07-25

Published On: 29-07-25

DOI: XXXX-XXXX

ABSTRACT

This paper introduces the **Transformer**, a novel neural network architecture that relies entirely on **self-attention** mechanisms, dispensing with recurrence and convolutions entirely. The model achieves significantly better results in translation tasks and is much more parallelizable, leading to reduced training time.

KEYWORDS

Proposed the Transformer architecture. Introduced the concept of multi-head self-attention. Outperformed previous models (like RNNs and LSTMs) on translation benchmarks. Enabled the development of later large-scale models like BERT, GPT, and T5.

ANU-SEARCH THE INTERNATIONAL JOURNAL OF PHARMA AND ALLIED SCIENCES

INTRODUCTION:

Positional Encoding: Adds position information to input embeddings. Scaled Dot-Product Attention: Core mechanism to focus on relevant words. Multi-Head Attention: Allows the model to jointly attend to information from different representation subspaces. Feedforward Networks: Used after attention layers. Layer Normalization and Residual Connections: Improve training stability and convergence.

MATERIALS AND METHODS:

This paper laid the foundation for modern NLP models. The Transformer architecture is the backbone of ChatGPT, BERT, T5, and many other state-of-the-art models.

RESULTS:

This paper introduces the Transformer, a novel neural network architecture that relies entirely on self-attention mechanisms, dispensing with recurrence and convolutions entirely. The model achieves significantly better results in translation tasks and is much more parallelizable, leading to reduced training time.

DISCUSSION:

Artificial Intelligence / Machine Learning / Natural Language Processing

CONCLUSION:

Positional Encoding: Adds position information to input embeddings. Scaled Dot-Product Attention: Core mechanism to focus on relevant words. Multi-Head Attention: Allows the model to jointly attend to information from different representation subspaces. Feedforward Networks: Used after attention layers. Layer Normalization and Residual Connections: Improve training stability and convergence.

ANU-SEARCH THE INTERNATIONAL JOURNAL OF PHARMA AND ALLIED SCIENCES

ABBREVIATIONS:

Positional Encoding: Adds position information to input embeddings. Scaled Dot-Product Attention: Core mechanism to focus on relevant words. Multi-Head Attention: Allows the model to jointly attend to information from different representation subspaces. Feedforward Networks: Used after attention layers. Layer Normalization and Residual Connections: Improve training stability and convergence.

FUNDING:

This paper introduces the Transformer, a novel neural network architecture that relies entirely on self-attention mechanisms, dispensing with recurrence and convolutions entirely. The model achieves significantly better results in translation tasks and is much more parallelizable, leading to reduced training time.

CONFLICT OF INTERESTS:

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, ?ukasz Kaiser, Illia Polosukhin

ACKNOWLEDGMENT:

DECLARATIONS:

ANU-SEARCH THE INTERNATIONAL JOURNAL OF PHARMA AND ALLIED SCIENCES

REFERENCES

1. This paper introduces the Transformer, a novel neural network architecture that relies entirely on self-attention mechanisms, dispensing with recurrence and convolutions entirely. The model achieves significantly better results in translation tasks and is much more parallelizable, leading to reduced training time.

HOW TO CITE: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, ?ukasz Kaiser, Illia Polosukhin, *Artificial Intelligence / Machine Learning / Natural Language Processing*, Int. J. of Pharm. Sci., 2025, Vol 3, Issue 1, 1803–1809.
doi: XXXX-XXXX