Name: Soham Pathak

Class: Final Year

Sec: B

PRN: 21070521079

Subject: Generative AI

Topic: Summary of Research Paper named as Attention Is All You Need

**Transformers: Redefining Machine Translation with Self-Attention Mechanisms**

The Transformer model, a ground-breaking design for sequence transduction problems, especially in machine translation, is introduced in the paper "Attention Is All You Need". The Transformer's primary innovation is its only dependence on self-attention processes, which enables parallelization and drastically cuts down on training durations when compared to conventional convolutional and recurrent neural networks (RNNs and CNNs). With the help of feed-forward neural networks and layers of self-attention, the encoder-decoder structure of the architecture allows the model to capture long-range dependencies in input data. The Transformer uses positional encodings to preserve sequence order, making sure the model comprehends the input's structure. As demonstrated by notable gains in BLEU scores, the Transformer's performance outperforms earlier models that relied on RNNs and CNNs, attaining state-of-the-art results on a variety of machine translation benchmarks. The architecture's easy scalability is another benefit that makes it appropriate for training on big datasets and utilizing the capabilities of contemporary technology. All things considered, the Transformer model is a major breakthrough in natural language processing, showing that attention mechanisms by themselves can accomplish better results in sequence transduction tasks and opening the door for additional study and applications in a variety of fields, such as text generation and summarization.