

Transformer Architecture: Three Key Innovations

① Positional Encoding: Adds order information to all words at once. Parallelization leads to 100x speedup by using all GPU cores.

