

Transformers_and_Language_Models_Basics

Introduction aux Transformers

Architecture créée par Vaswani et al. en 2017. Révolutionne le traitement séquentiel par attention multi-têtes.

Fonctionnement du Transformer

- Encodeur : analyse de la séquence
- Décodeur : génération d'une nouvelle séquence

Attention mechanism détaillé

Chaque mot se connecte aux autres mots du contexte par pondération de pertinence (self-attention).

Applications modernes

ChatGPT (OpenAI), PaLM (Google DeepMind), LLaMA (Meta), Claude (Anthropic).

Comparaison RNN vs LSTM vs Transformer

Transformer gère mieux la parallélisation et les longues séquences que les RNN et LSTM classiques.