

068 트랜스포머 아키텍처

Transformer Architecture

문맥을 고려해 단어 간 관계를 학습하는 AI 모델 구조

- 텍스트의 순서보다 단어 사이의 상호 연관성을 중심으로 정보를 처리하는 신경망 구조
- AI가 언어의 맥락을 이해하고 자연스러운 문장을 생성하도록 하는 LLM의 핵심 기반 기술

트랜스포머 아키텍처 개요

트랜스포머 아키텍처는 2017년 구글 연구진이 제안한 언어 처리용 신경망 구조로, AI가 문장을 더 깊이 이해하도록 만든 핵심 기술입니다. 기존의 순환신경망(RNN)은 단어를 순서대로 분석해 긴 문장을 처리하기 어렵고, 앞부분의 정보가 뒤로 갈수록 사라지는 문제가 있었습니다. 트랜스포머는 이러한 한계를 극복하기 위해 문장 전체를 한 번에 살피며 단어 간의 관계를 파악합니다. 즉, “문장 안의 모든 단어가 서로에게 주의를 기울인다”는 원리에 따라, 각 단어가 다른 단어와 어떤 의미적 연관을 가지는지를 계산합니다. 이를 통해 모델은 단어의 순서뿐 아니라 문맥 전체를 이해할 수 있어, 번역·요약·질의응답 등 복잡한 언어 작업에서 뛰어난 성능을 보입니다.

트랜스포머 아키텍처의 구성

트랜스포머는 입력을 해석하는 인코더와 출력을 생성하는 디코더로 구성됩니다. 인코더는 문장의 각 단어를 숫자 형태로 변환하고, 문장 내 다른 단어와의 관계를 계산합니다. 이때 핵심이 되는 어텐션(attention) 구조는 모든 단어가 서로를 참조하며 문맥적 중요도를 스스로 조정하도록 합니다. 예를 들어 “그녀는 사과를 먹었다”라는 문장에서 모델은 ‘그녀’와 ‘먹었다’의 관계를 인식해 주어와 동사의 연결을 이해합니다. 디코더는 이러한 정보로 다음 단어를 예측하거나 문장을 완성합니다. 트랜스포머는 계산을 동시에 수행하는 병렬 구조를 사용해 처리 속도를 크게 높였으며, 긴 문맥을 안정적으로 다루는 데 유리합니다. 이 구조 덕분에 대규모 데이터 학습이 가능해졌고, 인간 언어의 복잡한 의미 패턴을 정교하게 포착할 수 있게 되었습니다.

트랜스포머 아키텍처의 의의

트랜스포머의 등장은 AI 언어 이해 능력을 한 단계 끌어올린 혁신으로 평가됩니다. 이 구조를 기반으로 GPT, BERT, T5 등 다양한 언어모델이 등장하며, 챗봇·번역기·요약 도구 등 실제 서비스에 폭넓게 활용되고 있습니다. 더 나아가 시각·음성·텍스트를 함께 처리하는 멀티모달 AI로 확장되며, AI가 복합적 정보를 다루는 기반이 되었습니다. 그러나 성능이 높을수록 막대한 연산 자원과 전력 소모가 요구되고, 문맥을 단순한 통계 패턴으로 해석해 부정확한 답변을 내놓는 한계도 있습니다. 이를 개선하기 위해 모델 경량화, 장기 문맥 처리, 해석 가능한 AI 연구가 활발히 이루어지고 있습니다. 트랜스포머는 여전히 현대 AI의 표준 구조로 자리하며, AI 발전의 방향을 결정짓는 기술로 평가받고 있습니다.