

GEN AI 인텐시브 과정

강사장철원

Section 0

코스소개

DAY1

DAY2

DAY3

DAY4

DAY5

DAY6

DAY7

DAY8

LLM
Basic
Concept

Transformers
paper
review

Transformers
LangChain
LangGraph

LLM
service
develop

Final Project

□ 트랜스포머 기반 LLM

GEN AI 인텐시브 과정

Section 1. LLM

Section 1-1. 트랜스포머 기반 LLM

트랜스포머 기반 LLM

1. **GPT**(Generative Pre-trained Transformer) - Open AI
2. **LLaMA**(Large Language Model Meta AI) - Meta
3. **Mistral** - Mistral AI
4. **BERT**(Bidirectional Encoder Representations from Transformers) - Google
5. **ELECTRA**(Efficiently Learning an Encoder that Classifies Token Replacements Accurately) - Google
6. **BART**(Bidirectional and Auto-Regressive Transformer) - Meta
7. **T5**(Text-To-Text Transfer Transformer) - Google

트랜스포머 기반 LLM

1. **GPT** - 생성형 언어 모델, 대규모 텍스트를 기반으로 문장을 예측하여 생성

2. **LLaMA** - 경량화된 LLM

3. **Mistral** - 효율적인 성능과 적은 파라미터 수를 강조한 모델

4. **BERT** - 문맥을 양방향으로 이해하는 모델, 주로 문장 이해 및 검색 최적화

5. **ELECTRA** - 적은 계산량으로 BERT보다 효율적으로 학습 가능

6. **BART** - 문맥 이해와 문장생성 두가지 역할 모두 가능

7. **T5** - 모든 NLP 테스트를 "텍스트 입력 -> 텍스트 출력" 방식으로 변환

디코더 기반 모델

인코더 기반 모델

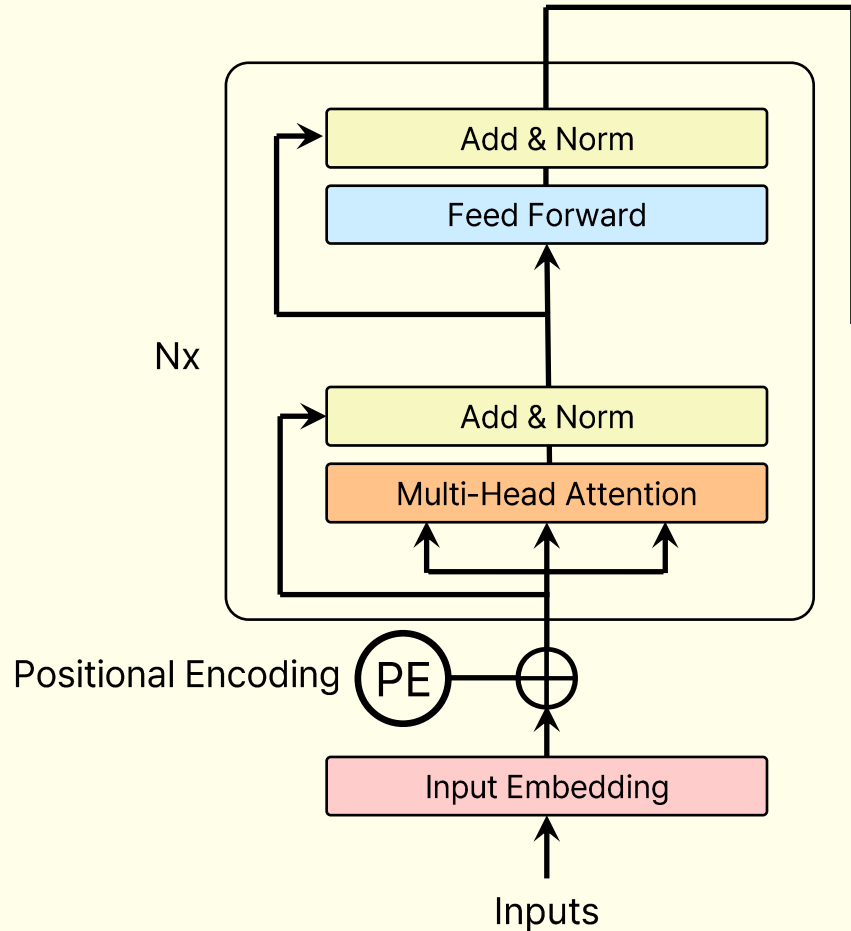
인코더-디코더 기반 모델

트랜스포머 기반 LLM

모델	출시일	개발사	주요특징
BERT	2018년 10월	Google	인코더 기반
T5	2019년 10월	Google	인코더-디코더 기반
BART	2019년 10월	Meta	인코더-디코더 기반
ELECTRA	2020년 3월	Google	인코더 기반
GPT3	2020년 6월	OpenAI	디코더 기반
LLaMA	2023년 2월	Meta	디코더 기반
Mistral	2023년 9월	Mistral AI	디코더 기반

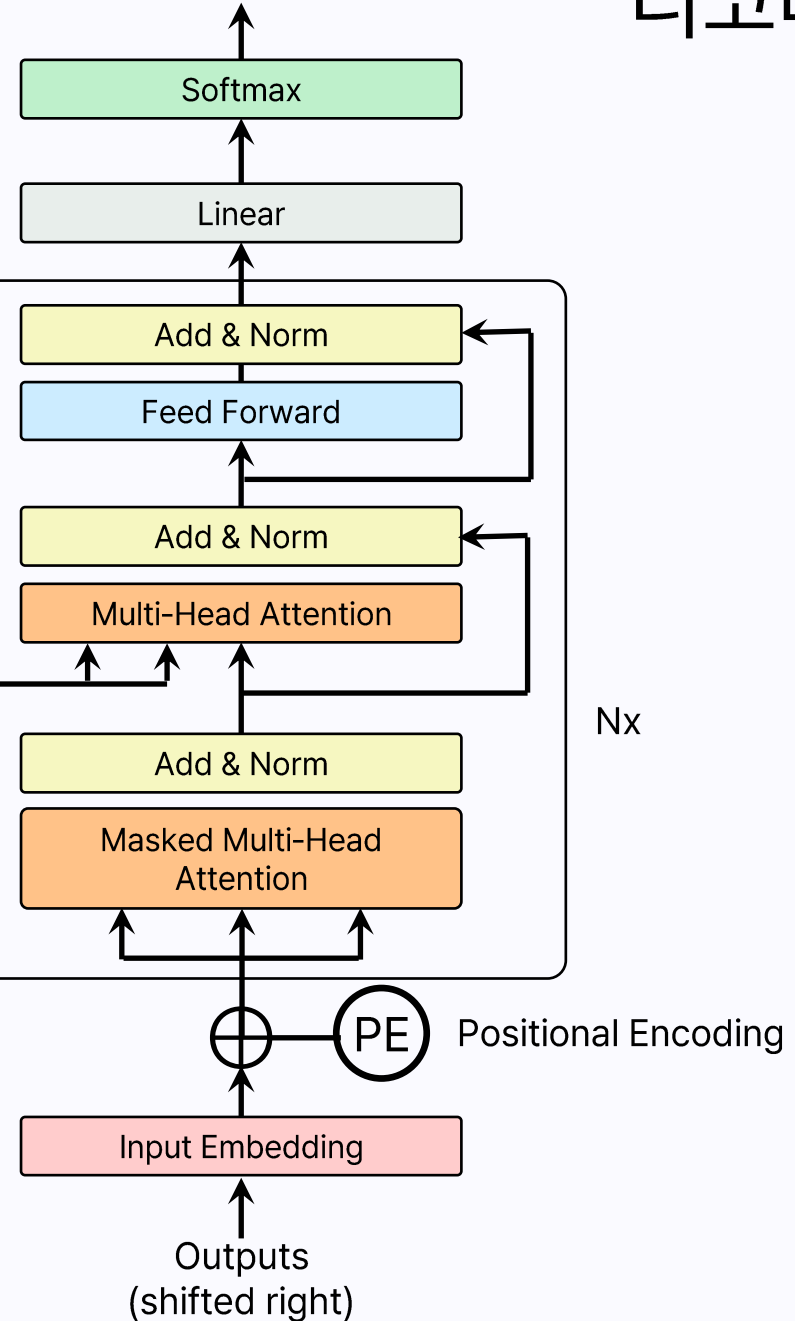
인코더 기반 / 디코더 기반

인코더

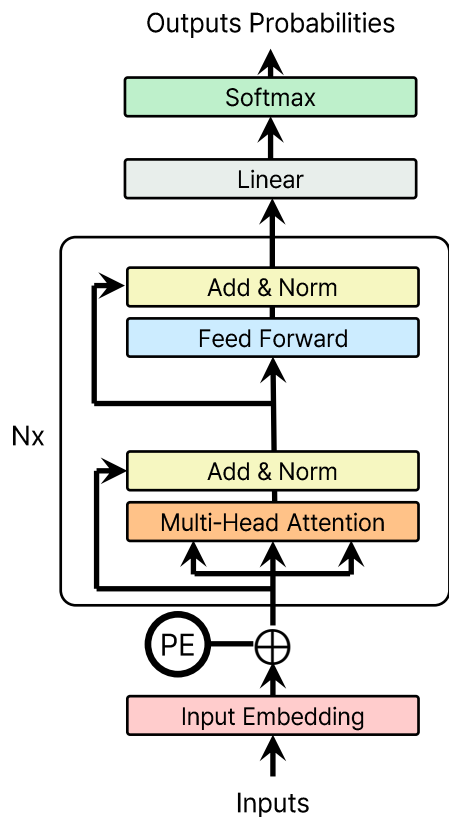


디코더

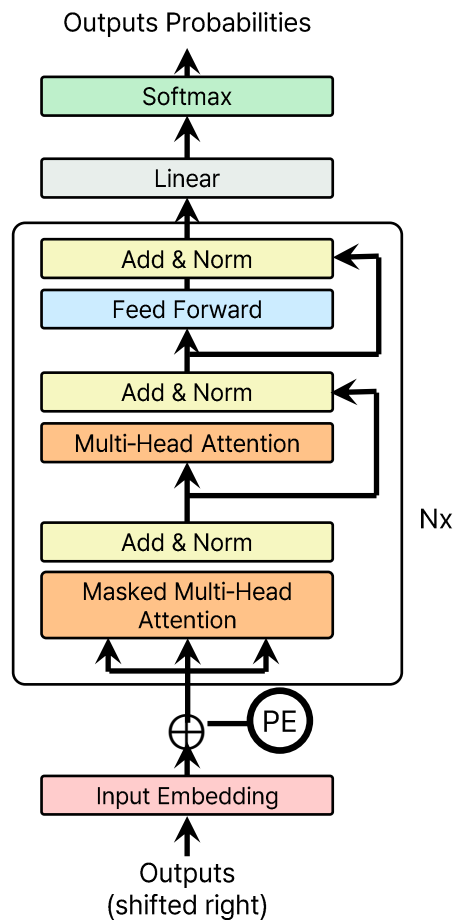
Outputs Probabilities



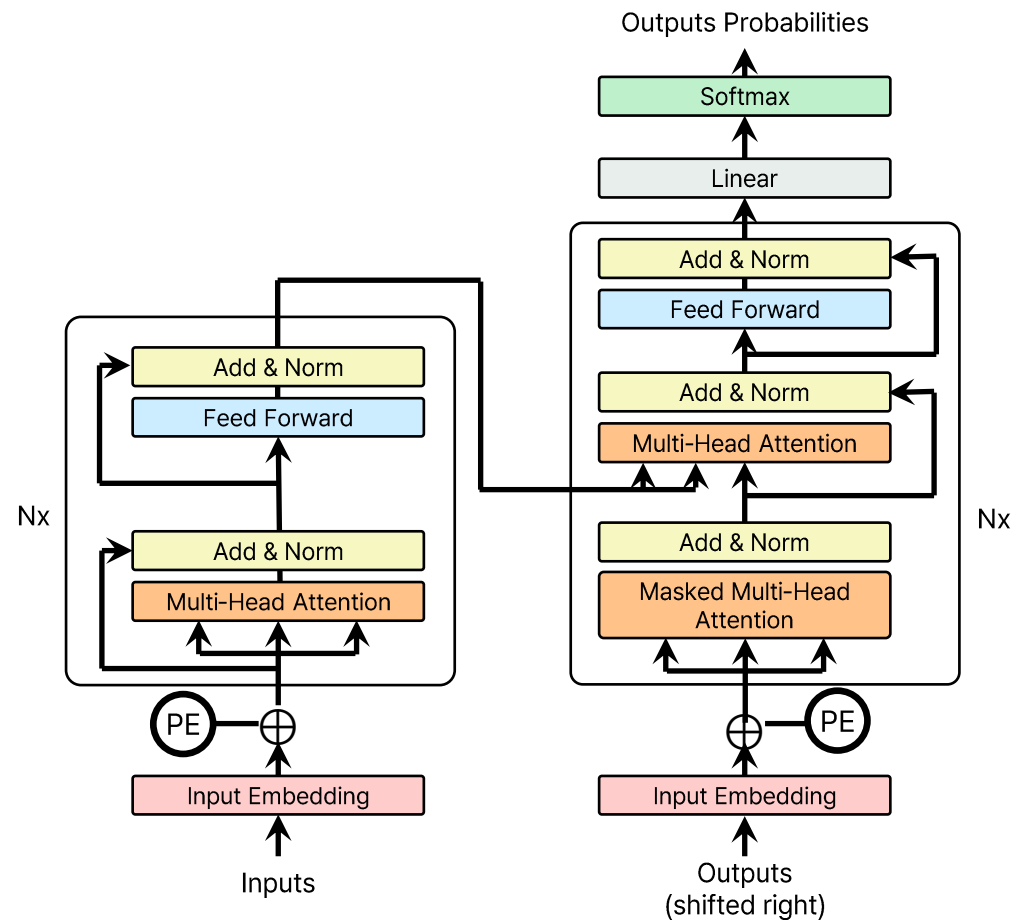
인코더 기반

BERT
ELECTRA

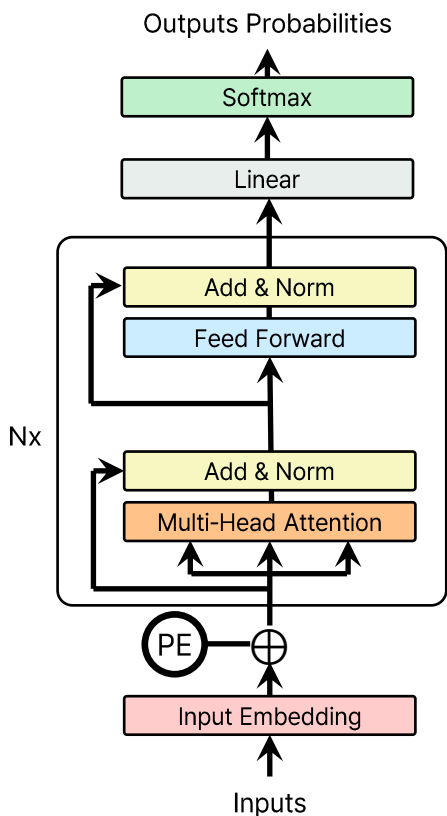
디코더 기반

GPT
LLaMA
Mistral

인코더-디코더 기반

BART
T5

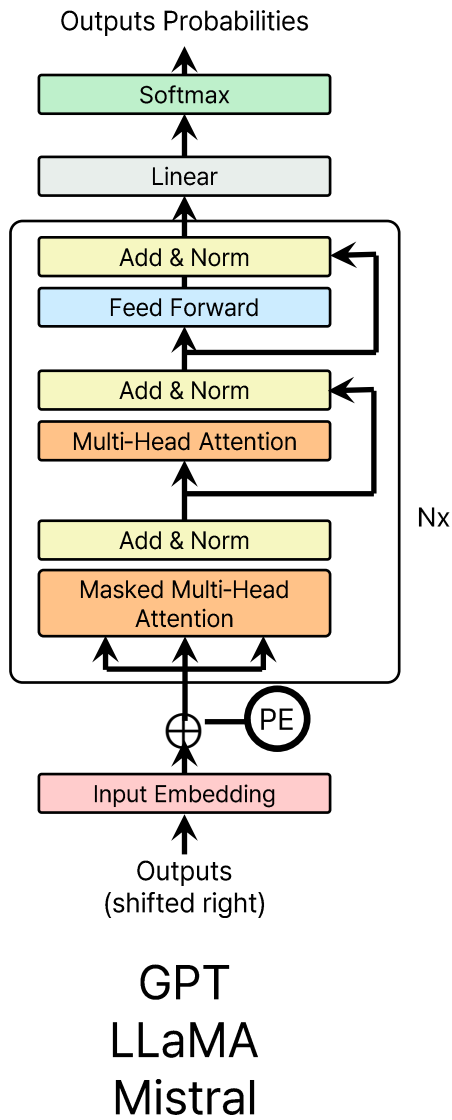
인코더 기반



BERT
ELECTRA

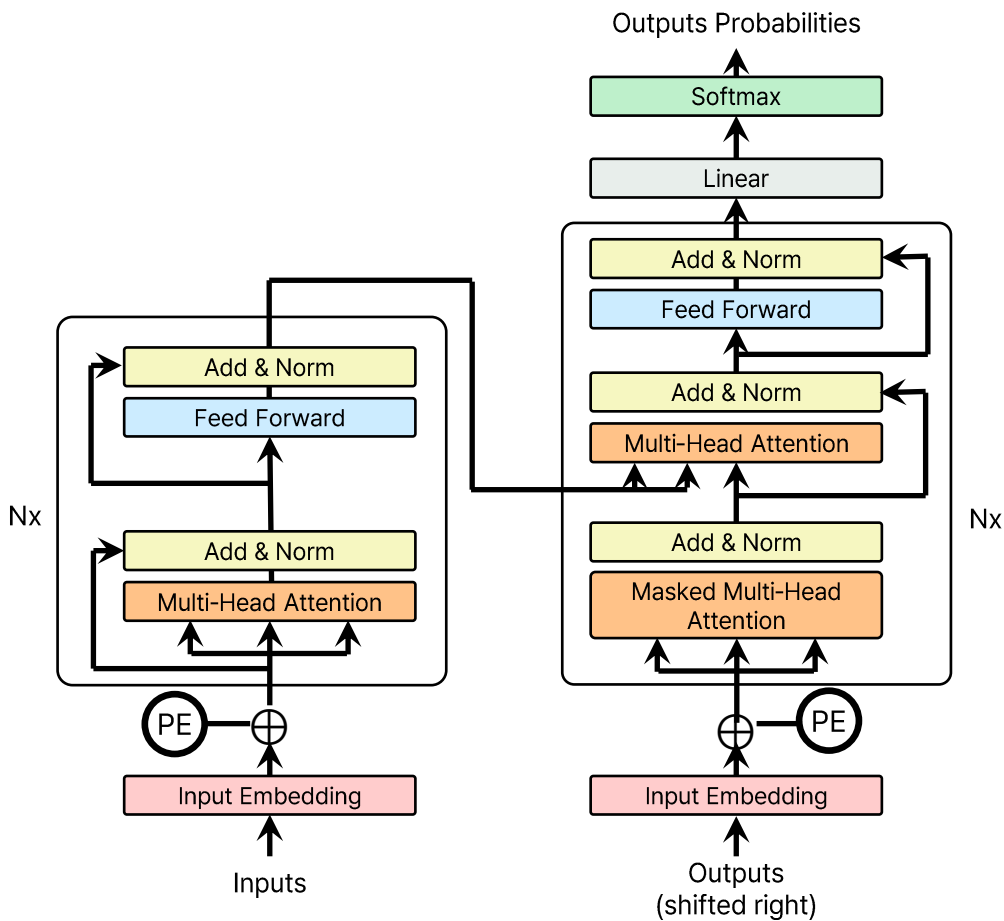
- 대표 모델: BERT, ELECTRA
- 구조: 트랜스포머의 인코더 블록만 사용
- 특징
 - 문장을 양방향(bidirectional)으로 분석하여 전체 문맥 이해 가능
 - 문장의 일부분을 [MASK] 처리하고 해당 단어를 예측하는 방식으로 학습 (MLM, Masked Language Modeling)
 - 텍스트 분류, 감성 분석 등 문맥 이해에 강함
- 입력: "나는 [MASK]를 좋아해"
- 출력: "축구"

디코더 기반



- 대표 모델: GPT, LLaMA, Mistral
- 구조: 트랜스포머의 디코더 블록만 사용
- 특징
 - 문장을 단방향(left to right)으로 처리하여 이전 단어들만 참고 가능
 - 과거 정보만 사용하여 다음 단어를 한 번에 하나씩 예측
(Masked Self-Attention을 사용하여 미래 단어를 참조하지 않도록 제한)
 - 문장 생성 능력이 뛰어나며, 챗봇, 텍스트 요약, 번역 등의 작업에 강함
- 입력: "나는 축구를"
- 출력: "좋아해"

인코더 - 디코더 기반



BART
T5

- 대표 모델: BART, T5
- 구조: 트랜스포머의 인코더 + 디코더 모두 사용
- 특징
 - 문맥을 이해하면서도 새로운 문장 생성 가능
- 인코더와 디코더 모두 사용하면 좋지만 효율이 떨어지는 경향이 있어 최근에는 인코더와 디코더 중 한 쪽만 사용하는 방법을 주로 사용함

언제 어떤 모델을 사용해야할까?

인코더 기반

- 문맥을 깊이 이해해야하는 작업
 - 문장 분류, 감성 분석 등
 - 구글 검색 엔진, 추천 시스템 등

디코더 기반

- 자연스럽게 문장을 생성해야하는 작업
 - 챗봇, AI 글쓰기 도구, 번역 등

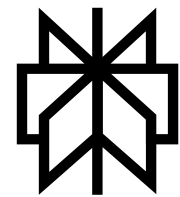
다양한 트랜스포머 기반 서비스



ChatGPT



Claude



perplexity

Gemini

감사합니다.

Q & A