

# Llama 2 Fine-Tuning 실습 - KorQuad 데이터셋 프로젝트의 목적

---

에이아이스쿨(AISchool) 대표  
양진호 (솔라리스)

<http://aischool.ai>

<http://solarisailab.com>

---

## 프로젝트의 목적

- 프로젝트의 목적 : 특정 좁은 분야(narrow) 의 지식에 있어서는 현존 최강의 LLM(GTP-4)보다 강력한 Fine-Tuning 된 LLM을 만들어보자!

# 전이 학습(Transfer Learning)

- 전이 학습(Transfer Learning) 또는 Fine-Tuning이라고 부르는 기법은 이미 학습된 Neural Networks의 파라미터를 새로운 Task에 맞게 다시 미세조정(Fine-Tuning)하는 것을 의미합니다.
- 컴퓨터 비전 문제 영역에서는 ImageNet 등의 데이터셋에 미리 Pre-Training 시키고 이 파라미터들을 내가 풀고자하는 문제에 맞게 Fine-Tuning하는 과정이 광범위하게 사용되고 있었습니다.
- 최근에는 BERT, GPT 같은 대규모 자연어 처리 모델이 등장하면서 자연어 처리 문제 영역에서도 전이 학습의 개념이 광범위하게 사용되고 있습니다.

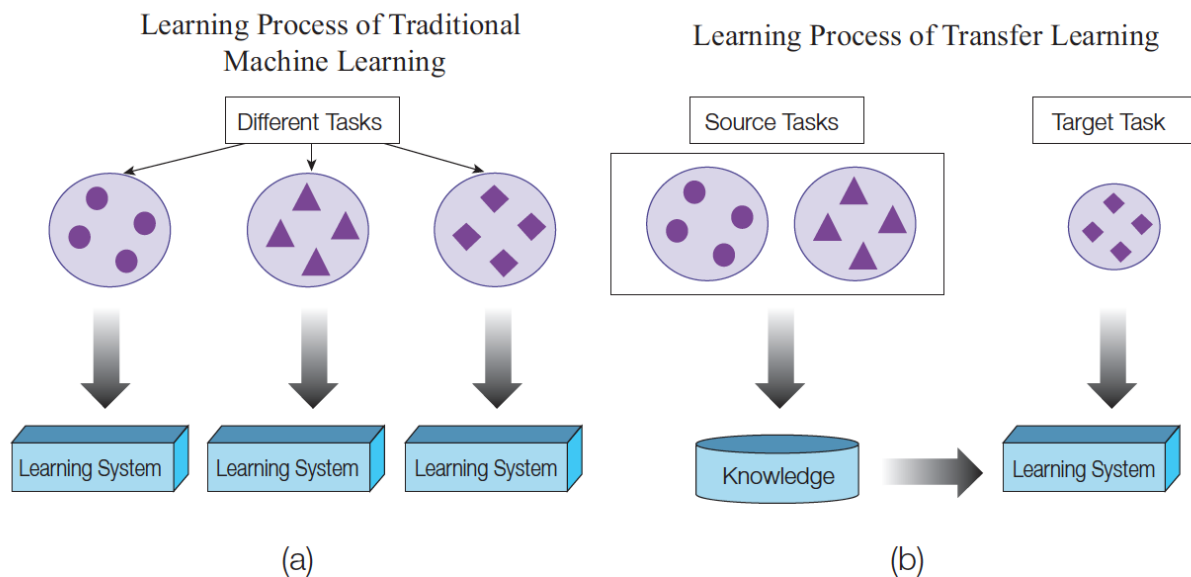


그림 6-5 | 전이 학습의 예

# Fine-Tuning Overview

- 대량의 corpus 데이터셋으로 Pre-training -> 목적에 맞게 Fine-Tuning

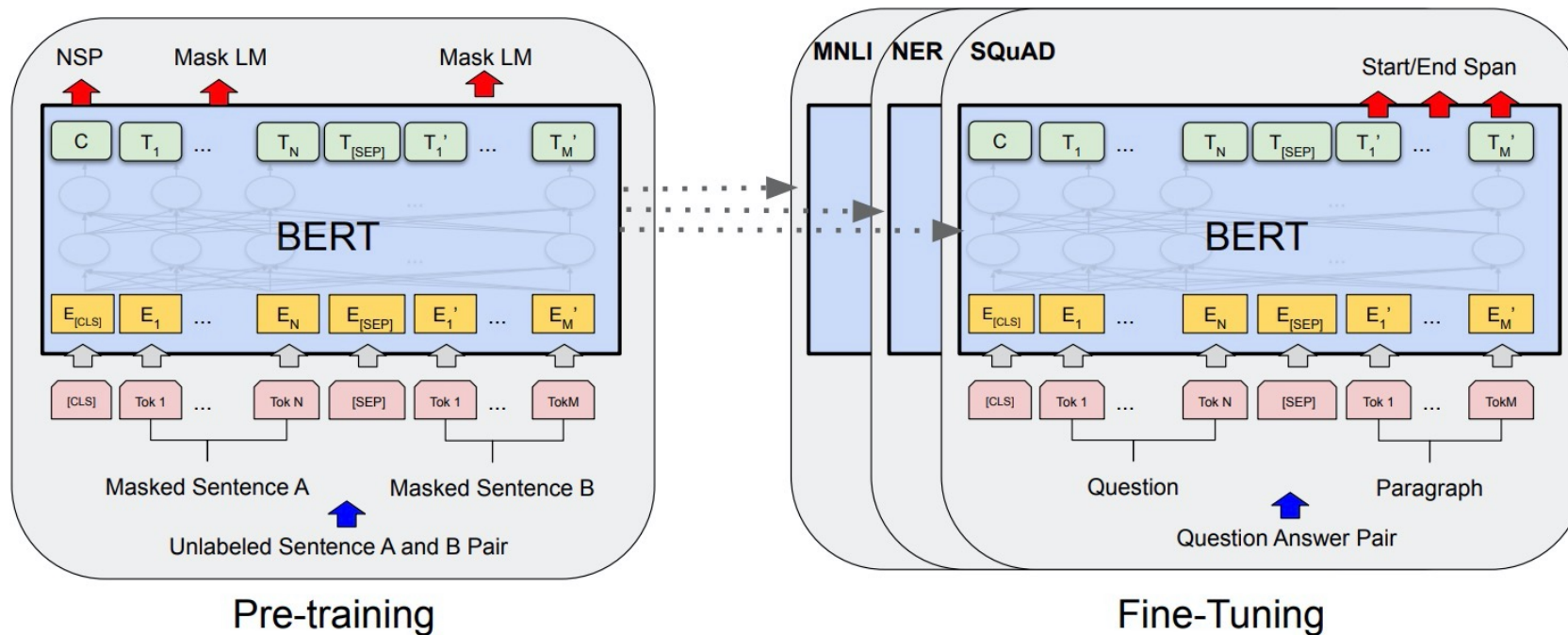


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. [CLS] is a special symbol added in front of every input example, and [SEP] is a special separator token (e.g. separating questions/answers).

---

## KorQuad v1.0 Dataset Example 1

- Question : 1989년 6월 30일 평양축전에 대표로 파견 된 인물은?
- Answer : 임수경

---

## KorQuad v1.0 Dataset Example 2

- **Question** : 1989년 2월 15일 여의도 농민 폭력 시위를 주도한 혐의로 지명수배된 사람의 이름은?
- **Answer** : 임종석