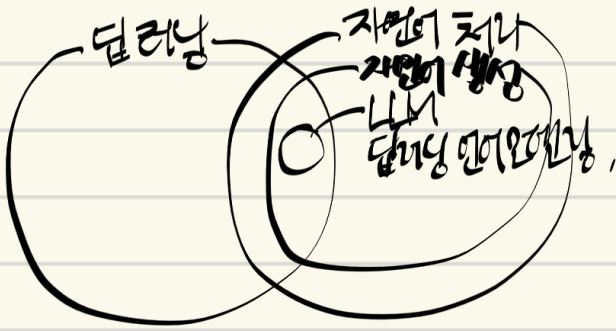


[LLM 지도]

① 데이터 & 언어 모델링

- LLM이 데이터를 몇 가지 처리해야 하는지 생각해볼 것



□	원스텝	2013
□	트랜스포머	2017
□	GPT-1	2018

- 데이터의 문제 해결 방법

① 원본 데이터 → ② 데이터셋 준비 → ③ 반복 학습 → 비정형 데이터도 준비해?

- vs 비정형 데이터

Who? → 데이터셋

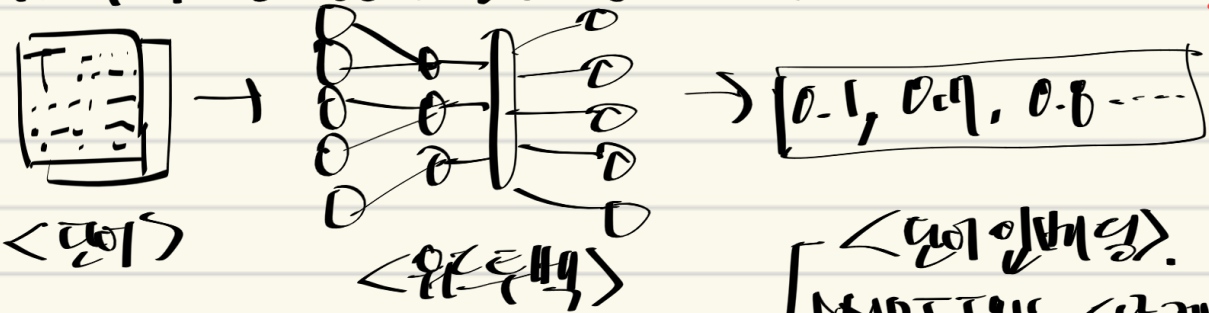
- 일베당

데이터를 그 의미를 알아 어떻게 숫자 값으로 표현 like MBTI
 What 이걸? → MBTI에 대한 특징을 나타낼 수 있는

그걸 "아, 나는 잘했어!" + 어떻게? MBTI에 대한
 ⇒ MBTI를 매 month 각 사람의 두의 결과가 적혀서 + 평가가 있다.
 데이터

III 워드 투 벡 (word vec)

워드 투 벡이란 어떤 문장을 일베당으로 변환할 수 있게 변환된 값을 '문장 일베당'



<문 일베당>
 MBTI를 어떻게 보면
 사람 성격에 따라 다르게 표현됨.
 또한, 두 사람의 벡터를 합쳐서
 장차 어떤 일을 하게 될지