

The background features a complex network of thin grey lines and dots, forming a web-like structure. Scattered throughout are various triangles of different sizes and orientations, some with solid black dots at their vertices. The overall aesthetic is minimalist and technological.

# AI 11기 신성윤

---

AI 심리상담  
챗봇

데이터 선정이유 및  
문제설정

01

모델 선택

02

데이터 전처리 설명

03

## TABLE OF CONTENTS

04

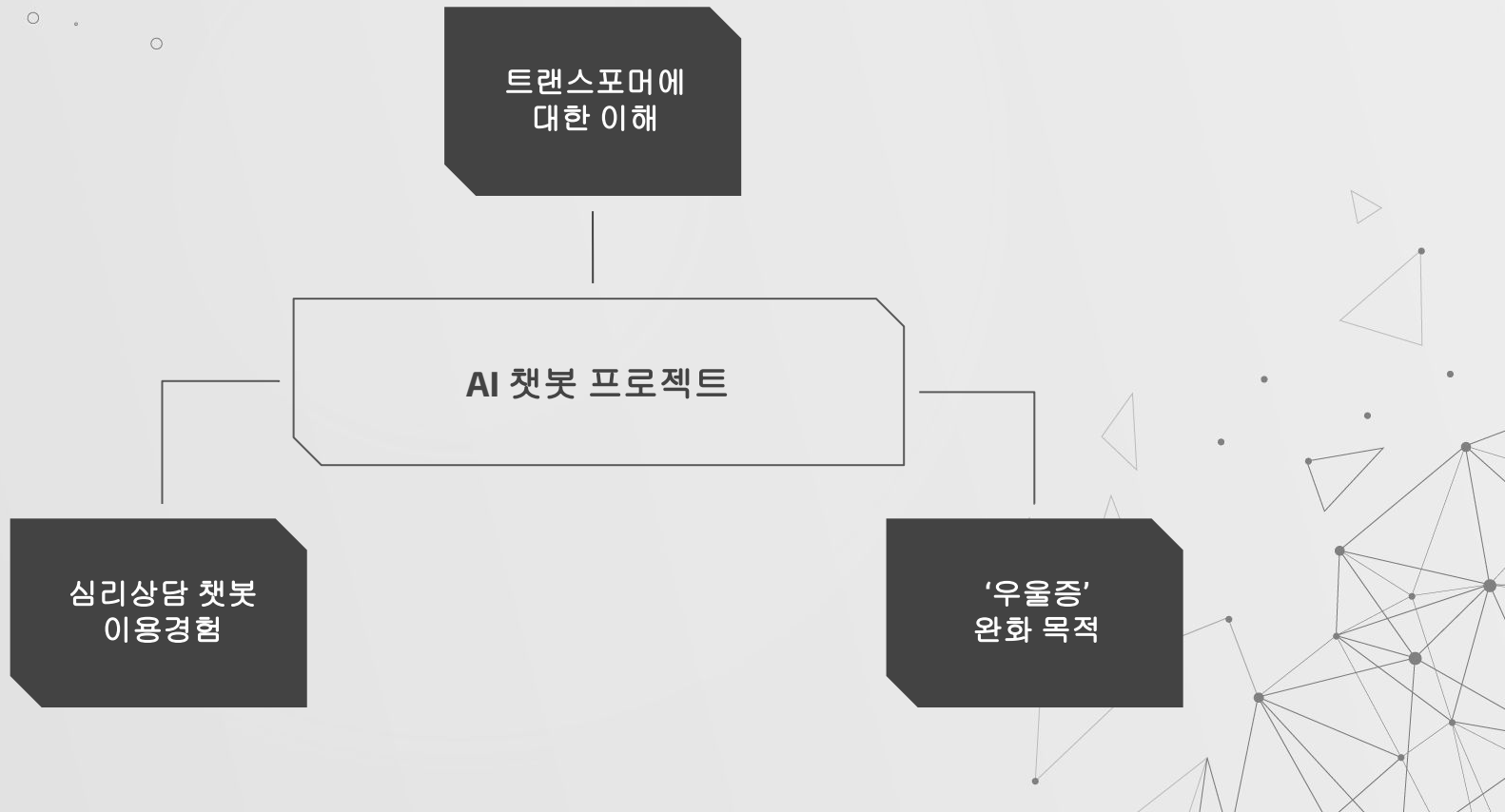
한계점 및 추후방안



# 01

## 데이터 선정이유 및 문제 설정

# 데이터 선정이유 및 문제설정





# 02

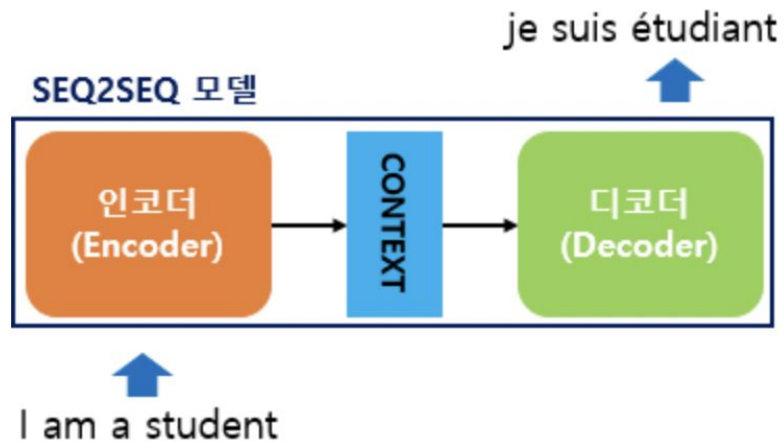
## 모델 선택

---

## 모델 선택

### 트랜스포머 (Transformer)

기존의 Seq2Seq 모델에서  
병렬 연산을 수행할 수 있게  
개선된 모델.  
성능 또한 우수함.



기존 Seq - to Seq 모델



# 03

## 모델 학습 진행

---

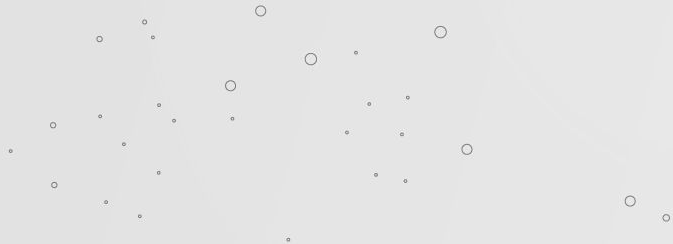
## 데이터 출처

**<https://aihub.or.kr/opendata/keti-data/recognition-language/KETI-02-006>** (AI HUB  
정신상담데이터)

**[https://github.com/songys/Chatbot\\_data](https://github.com/songys/Chatbot_data)**(개  
인깃협 '송영숙'님 자료)

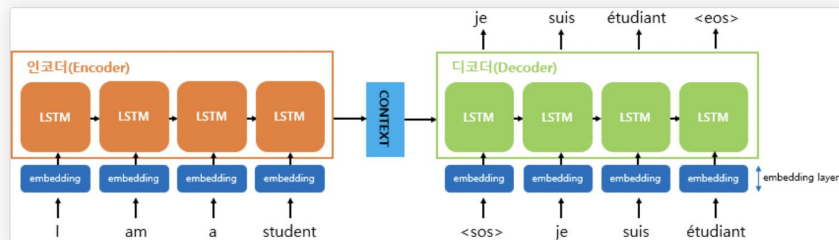


# 코랩 실행



# 교사 강요 ( Teaching forcing)

이제 단어를 입력으로 제공하여 학습을 시작합니다.



출처: 딥러닝을 이용한 자연어처리

○ 디코더의 학습과정에서 예측 데이터가 아닌 실제 데이터를 넣어준다.

# 04

## 한계점 및 추후방안

---



# 한계점

## 1. 데이터의 한계

좀 더 다양하고 방대한 데이터가  
있으면 좋겠지만 채팅기록은  
정보보호상 쉽게 구할수 있는  
데이터가 아니기 때문에 아쉬울수  
밖에 없다.

## 2. 모델 이해의 어려움

복잡한 모델이기 때문에  
한정적인 시간안에 모델을  
깊이있게 이해하기  
어려웠다.

# 추후 방안

## 1. 지속적인 채팅기록 찾기

일상 대화기록이 아니더라도  
특정 도메인에서 공개하는  
데이터라도 찾아보고 그걸  
이용해 학습을 시켜보자.

## 2. 카톡 찜 수작업

대화 기록을 구할 수 없다면  
다양한 대화 예시를 보면서  
직접 수작업 하는 방법을  
사용해 보는 것도  
도움이 될 거 같다.

**THANKS**

