



최규민(Gyumin Choi)



Contact

GitHub | <https://github.com/catuscio>
Phone | 82+ 10-2498-0398

Email | catuscio@hotmail.com

Greetings!



좋아서 개발하는 신입 ML 엔지니어, 최규민 입니다!

저는 ML엔지니어 및 관련 직무를 희망하고 있는 인공지능 관련 학과 3학년 재학생입니다.

자연어처리, 그 중에서도 LLM에 관심이 깊어, 자대 연구실에서 LLM을 활용한 Text-to-SQL 프로젝트를 진행하고 있습니다.

Technical Skills

ML / DL

- sLLM(gemma-2-2b-it / Llama3.1-8B-instruct)을 SQL 쿼리문 생성을 위해 QLoRA 파인튜닝 및 프롬프트 엔지니어링을 진행하였습니다.
- Huggingface사의 `transformer` 라이브러리의 `SFTTrainer` 를 이용해 LLM을 학습하고, `pipeline` 을 이용해 추론할 수 있습니다.
- `vLLM` 라이브러리를 사용하여 LLM 추론을 할 수 있습니다.
- 파인튜닝 및 프롬프트 엔지니어링을 진행한 모델을 TTQ 벤치마크인 Spider dataset에 따라 평가하였습니다.

Environment Settings

- 신생 연구실 인턴으로 활동하며 딥러닝 워크스테이션 견적서를 작성하고, SSH 포트 포워딩을 통해 원격 서버로 구축한 경험이 있습니다.

- `Conda` 를 이용한 가상 환경을 구축할 수 있고, 각 환경에 필요한 라이브러리를 버전에 맞추어 셋업할 수 있습니다. `CUDA` , `torch` , 기타 라이브러리 버전 충돌로 인한 의존성 문제를 해결한 경험이 있습니다.

Data Analysis

- TensorFlow에 익숙하고, Kaggle, DAICON과 같은 데이터 분석 플랫폼을 통한 데이터 분석 프로세스에 경험이 있습니다.
- `Numpy` , `Pandas` , `Matplotlib` , `Keras` , `Scikit-learn` 등을 이용해 EDA부터 모델 학습 및 시각화까지의 데이터 분석 워크플로우를 설계할 수 있습니다.
- LG에서 주치한 제품 이상 여부 판별 프로젝트에 데이터분석 및 모델링으로 참여한 경험이 있습니다.

Work Experience

세종대학교 AI로봇학과 IDEAL Lab ([link](#))

(Industrial Digitalization & Energy Applications with Learning system Laboratory)

Undergraduate Research Program (2023.12 -)

- `Speech-to-Text-to-Query` 프로젝트에 참여하여, 허깅페이스의 8B 이하 sLLM을 파인튜닝 및 프롬프트 튜닝을 통해 `Spider Data` 의 벤치마크를 향상하는 과제를 수행하였습니다.
- 4개월의 수습 기간 중 CNN, LSTM, VAE, Attention, Transformer등 다양한 논문을 리뷰하고 몇 건은 세미나에서 발표하였습니다.

학부 3학년 당시 신생랩에 스타팅 멤버로 합류하였습니다.

`Text-to-SQL` 과제를 수행하던 중, 가상환경에서 발생하는 의존성 문제들과 CUDA OOM 에러 등을 Github issue, huggingface forum와 같은 다양한 커뮤니티에서 인사이트를 얻어 해결하였고, 여러 시도를 거쳐 가용 가능한 메모리 안에서 최적의 파라미터를 찾아 학습을 진행하였습니다.

모델이 자연어 질의를 쿼리문으로 변환하는 과정을 추론할 수 있도록, 사용한 데이터인 `Spider Data` 의 Database schema를 함께 학습 데이터에 추가하려 하였으나, 이렇게 만들어진 데이터의 max token은 2400인 반면, 사용 가능한 `max_sequence_length` 는 200에 불과하였습니다.

추론에 필요한 정보를 상당 수 포기하더라도, `Spider Data` 에 레이블 된 정답 쿼리의 format이라도 학습하도록 하자는 생각으로, 학습 데이터 프롬프트에서 자연어 질의를 전진 배치하였습니다.

Projects

STOP-EAT!; ChatGPT API를 이용한 다이어트 식단 추천 서비스

- **Develop Period:** 2024.04 - 2024.06
- **Project Description**

전공 학과 학술제 참여를 위해, 키와 몸무게를 입력하면 BMI를 계산하고, 사용자가 원하는 다이어트 강도에 따라 식단을 추천하는 프로젝트였습니다.

- **Role**

- Project Manager

- PM으로서 프로젝트의 계획을 수립하고, 팀원들의 업무 관리와 프레젠테이션을 담당하였습니다.

- AI Engineer

- ChatGPT API를 불러와 사용자의 질문에 맞는 답변을 생성하는 역할을 담당하였습니다.

- **Learnig Point**

- 해당 과정에서는 팀원들 간의 기술적 이해도의 차이를 어떻게 극복할 수 있는지를 고민하였습니다. 깃허브, 웹개발 언어 등을 전혀 모르는 팀원들에게 개발 방법론을 제시하고, Notion과 Microsoft Teams를 활용한 계획 관리 및 주기적 미팅을 통해, 성공적이진 못하지만 팀원들 모두가 개발에 대해 조금이나마 이해를 가져갈 수 있도록 하였습니다.

True Love; ChatGPT API를 이용한 TTS 일상 대화 챗봇 서비스 ([GitHub](#))

- **Develop Period:** 2023.07 - 2023.09

- **Project Description**

코로나 이후 1인가구의 수가 급증하며, 자연스레 사회적 단절로 인한 우울증 등의 정신질환 발병 수가 증가하였습니다. 이에 따라, 타인과의 일상 대화를 나눌 기회가 적은 사용자들을 위해 STT 기술로 사용자의 음성을 입력 받은 후, OpenAI API를 이용해 텍스트 답변을 생성하여 다시 TTS로 해당 답변을 음성으로 출력하여 발화를 나눌 수 있는 챗봇을 개발하는 프로젝트였습니다. 동아리 활동의 일환으로 참여하였습니다.

- **Role**

- Project Manager

- PM으로서 프로젝트의 계획을 수립하고, 팀원들의 업무 관리와 프레젠테이션을 담당하였습니다.

- **Learnig Point**

- 해당 과정을 통해 Agile, Waterfall 등 다양한 소프트웨어 개발 방법론에 대해 학습하였고, 프로젝트 유형과 팀 유형에 따른 올바른 개발 방법론에 대해 배울 수 있었습니다. 또한 OpenAI API의 구동 방식과 Node.js에서의 개발 방식을 학습하였고, GitHub, Notion, Slack 등 소프트웨어 개발에 있어서 중요한 협업툴과 협업 방식을 경험하였습니다.

걸어서 맛집 속으로; 광진구 맛집 탐방 지도 서비스 ([GitHub](#))

- **Develop Period:** 2022.07 - 2022.09

- **Project Description**

광진구 거주 직장인/학생들을 대상으로, 인근 맛집의 위치와 요약 정보를 제공하는 동아리 활동의 일환으로 참여한 프로젝트였습니다.

- **Role**

- Frontend 개발 담당

- HTML, CSS를 이용하여 웹사이트의 Nav바 UI를 제작하였습니다.
 - 카카오맵 API를 이용하여, 웹사이트 상에서 지도를 표시하고, 지도 위에 마커 기능을 통해 음식점의 이미지 및 요약 정보를 표기하는 역할을 담당하였습니다.

- **Learnig Point**

- 해당 과정을 통해 깃허브를 이용한 협업 방식에 대해 알 수 있었습니다. 또한 HTML, CSS, JS의 문법과 다양한 함수의 사용을 학습하였고, 웹사이트 개발에서의 API의 기초적 사용 방법을 배웠습니다.

Awards

LG Aimers/Data Intelligence 5기 프로그램 교육과정 이수

- LG가 주최한 데이터분석전문가 양성을 위한 교육 프로그램을 수료하였습니다.
- 이후 해당 코스와 연계되는 **디스플레이 공정 과정에서의 제품 이상 판별 해커톤**에 참여하여 Baseline 이상의 점수를 획득 하였습니다.

세종대학교 AI로봇학과 창의SW기초설계 전체 3위

ESP32 WIFI통신과 카카오프API를 활용한 응급상황 감지/경보기 개발 ([GitHub](#))

- **Project Description**

- 1인가구 급증에 따라, 1인가구의 응급상황 발생 시 대처 방법에 대한 우려 또한 증가하였습니다. 가장 큰 문제로는 '응급상황 발생 시 발견자/신고자가 없어 사후 대응이 늦어진다'는 점입니다. 이에 따라, 혼자 사는 1인가구 거주민들도 응급상황 발생 시 제때에 대처가 이루어 질 수 있도록, 응급상황 감지/경보기를 제작하여, 응급상황임을 감지하고 이를 소방서 등에 알릴 수 있는 소형 가전을 제작하였습니다.

- **Role**

- Project Manager
 - PM으로서 프로젝트의 계획을 수립하고, 팀원들의 업무 관리와 프레젠테이션을 담당하였습니다.
- CAD
 - Fusion360을 이용한 제품의 3D모델링을 담당하였습니다.
- SW Develop
 - PIR센서, 가스감지센서, LED, 부저 등의 센서가 사용자가 응급상황에 처했음을 감지하고, 감지된 경우 카카오톡API를 사용하여 메시지를 전송할 수 있는 기능을 구현하였습니다.
- Circuit Develop
 - ESP32보드를 사용하여, 아두이노 개발 환경에서 제품이 잘 작동할 수 있도록 전반적인 회로를 설계하였습니다.

- **Learnig Point**

- 해당 과정을 통해 임베디드에 관한 지식을 학습하였으며, IoT 통신 기술의 기초적인 지식을 배울 수 있었습니다.

세종대학교 소프트웨어융합대학 SW멘토링 대상 (1위)

<YOLO를 활용한 교통환경에서의 객체 검출 기술 동향 및 전망> 소논문 작성

- **Project Description**

- 교내 공통필수 과목에서 요구하는 팀프로젝트의 일환으로, 전공 방향성에 맞는 소논문을 작성하는 프로젝트였습니다. 논문 주제로 YOLO를 활용한 객체 검출을 선정하였고, 7개의 국내 저널과 4개의 국제

저널을 참고하여 소논문을 작성하였습니다.

- **Role**

- Project Manager

- PM으로서 프로젝트의 계획을 수립하고, 팀원들의 업무 관리와 프레젠테이션을 담당하였습니다.

- Paper Writing

- CNN, RNN에 관한 부분을 작성하였습니다.
 - Abstract, Conclusion을 작성하였습니다.
 - YOLOv3를 ROS환경에서 구동하여, Checkbox로 객체를 인식하는 실습을 진행하였습니다.

- **Learnig Point**

- 해당 과정을 통해 논문의 작성 방식 및 문법을 학습하였고, Object Detection 등 Computer Vision에 관한 지식과 CNN, RNN 등의 인공지능망에 관한 지식을 쌓을 수 있었습니다.

세종대학교 창의나눔SW튜터링 우수상(3위)

- C언어를 사용하는 전공 필수 과목 <고급C프로그래밍>을 위한 스터디에 멘토로 참여하였습니다.

Teaching Experiment & Extracurricular Activities

세종대학교 AI로봇학과 인공지능 융합 학술동아리 SMARCLE 집행부원 (2023) ([link](#))

- 교내 학술동아리 집행부원으로 활동하며, 동아리의 연간 스터디 커리큘럼을 계획하고 운영하는 한편, 동아리 내 해커톤 및 공모전 등의 대회를 기획하였습니다. 동아리 소식지 <월간 스마클>을 작성 및 편집하여 배포하였습니다.
- 2년 간 인공지능 학술동아리 부원으로 활동하며, Arduino 스터디, 데이터 분석 스터디, 딥러닝 스터디, Kaggle 스터디에 참여하였습니다. 이를 통해 인공지능에 관한 지식을 학습할 수 있었으며, 동아리 내부 경진대회인 Makers' Day에 2회 참여하며 프로젝트 경험을 쌓을 수 있었습니다.

코드클럽 SW교육봉사단 대학생 멘토 (2023)

- 라즈베리파이 재단 소속의 교육 단체 코드클럽에 대학생 교육 봉사자로 가입하여, 초등학교 4학년 학생들을 대상으로 스크래치를 이용한 블록코딩을 주제로 16회의 수업을 진행하였습니다.

세종대학교 AI로봇학과 제8대 MATE 학생회 학술부장 (2024)

- 과 학생회의 학술부장으로 활동하며, 학과 학생들을 대상으로 한 스터디, 세미나 등을 기획하고, 해커톤 등의 대회를 계획하고 있습니다.

Lectures

기계학습 Machine Learning (전공필수)

인공지능 Artificial Intelligence (전공필수)

자연어처리 Natural Language Processing (전공선택)

Research Interest

- LLM Fine-Tuning for Specific downstream tasks
- RAG, CoT, Prompt Engineering

Technical Stacks

Language

Python, C, HTML, CSS, JavaScript

Libraries & Frameworks

TensorFlow, Pytorch, Transformers, Numpy, Pandas, DataFrame, Scikitlearn, Keras, etc.

Etc.

LaTeX, Markdown

Collaboration & Tools

- Familiar with Jupyter notebook, Colab, Kaggle, DAICON
- Familiar with Visual Studio, VS code, Eclipse
- Familiar with Slack, Notion, Microsoft Teams, Git

Education

Major

세종대학교 AI로봇학과 (2022.03 - 재학 중(6학기))

Sejong University Artificial Intelligence & Robotics

GPA 3.15/4.5

Language

TOEIC 865 (RC380 LC 485)