

# JASON HEESANG LEE

PROJECT MANAGER WITH AN IN-DEPTH  
UNDERSTANDING OF ARTIFICIAL INTELLIGENCE



## CONTACT

Jason.Heesang.Lee96@gmail.com

+82 10 4167 3900

LinkedIn | GitHub | Kaggle

## EDUCATION

### Year-Dream School - AI Bootcamp

Seoul, Republic of Korea

Apr 2023 – Dec 2023

Data Science & Data Analysis

### Glion Institute of Higher Education

Montreux & Bulle, Switzerland

Aug 2015 – Mar 2021

BBA in Hospitality Management  
with Luxury Brand Management Specialization  
in Hospitality

### Beijing Royal School

Beijing, China

Aug 2012 – Jun 2015

Highschool Diploma

## SKILLS

Word / Excel / PowerPoint (Expert Level)

LinkedIn Sales Navigator (Expert Level)

LangChain (Advanced Level)

Python / PyTorch / PyPI / Git (Advanced Level)

## LANGUAGES

### Korean

Native or Bilingual

### Chinese

Professional Working

Proficiency

### Bahasa Indonesia

Elementary Proficiency

## EXTRA-CURRICULAR

### Dreaming – Group Study

Chat Date Partner

- Personalizable AI ChatBot



## PROFESSIONAL EXPERIENCE

### Fast Campus, Korea

Education & Operations Manager

Mar 2024 – Current

#### General Responsibilities

Sourcing start-up companies for Industry Partnership Project

#### Responsibilities for Data Science Track

Overall Operations of the course.

Planning and operate extra-curricular courses.

Responding to students' queries, solving technical issues.

### Atheneum Partners GMBH, Korea

Team Lead / Project Manager

Aug 2022 – Feb 2023

Senior Associate / Project Manager

Jun 2022 – Jul 2022

Associate

Sep 2021 – May 2022

#### Responsibilities

Project Management

Revenue Management

### Four Seasons Hotel Seoul

People & Culture Coordinator

Apr 2021 – Sep 2021

#### Responsibilities

Human Resources Administrations

## PACKAGE DEPLOYMENT

### DanGam

#### Word-Level Emotion Analyzer

Searching for the emotion of each word within given text.

- Available for Korean & English Texts.
- Available on PIP.



### JsonSpeller

#### In-Document SpellChecker

Searching correct word for misspelled words within given text.

- Available for Korean & English Texts.
- Able to correct newly generated terms and personal terms.



## HONORS AND AWARDS

### Kaggle

#### Competitions Expert

#### Notebooks Master

CAFA 5 Protein Function Prediction

Linking Writing Processes to Writing Quality

Silver Medal & Top 63<sup>rd</sup> / Top 4%

Bronze Medal & Top 176<sup>th</sup> / Top 10%

### Other Competitions

Seamless Machine Translation

2<sup>nd</sup> / 64 Entries

안녕하세요,

AI를 이해하고 활용하는 2년차 Project Manager 이희상입니다.

첫 직장이었던 포시즌스 호텔 인재문화부 이후, 보다 많은 산업군을 경험하고 싶었고, 지인의 권유를 통해 전문가 인터뷰 중개업체인 아테네움 파트너스로 이직하였습니다.

아테네움 파트너스에서 프로젝트 매니저로 담당하던 프로젝트인 "전략컨설팅 B 사 - 반도체산업 벤치마킹 프로젝트", "전략컨설팅 P 사 - 금융산업 디지털전환 프로젝트"를 통해 데이터와 인공지능에 대해 흥미를 가지게 되어 관련 소식을 지속적으로 찾아 보았고, 미래에는 어떤 산업에서도 AI 활용이 필수적일 것이라 생각하였습니다.

이러한 물결에 뒤쳐지지 않기 위해 기존 직장을 퇴사하였고, 2023년 3월 중소벤처기업진흥공단 주관의 AI 부트캠프인 이어드림스쿨에 참여해 Python을 처음 배우기 시작한 후, 10개월 만에 Kaggle Competitions Expert, Notebooks Master를 달성하였습니다.

저는 오랜 해외 생활로 다양한 문화 경험을 통해 다섯가지 언어를 구사하며, 세가지 언어를 유창하게 할 수 있게 되었습니다. 다양한 언어를 구사할 수 있다는 점과 언어별 구조적 특성에 관심이 많아 자연어처리와 생성 분야에 관심이 깊어지게 되었습니다.

자연어 분야와 관련하여 다양한 대회와 프로젝트들을 진행하였고, 모두 준수한 성과를 낼 수 있었습니다. 팀장으로서 팀원들을 이끌었던 Kaggle 대회인 CommonLit - Evaluate Student Summaries 참여하여 대회 내 대중적이지 않은 방식이었던 Knowledge Distillation을 통해 접근하여 참가자들의 관심을 끌어내었으며, Linking Writing Process to Writing Quality 대회에서 작성한 Essay Reconstruction 노트북은 대회 1등과 7등이 인용하는 등 좋은 평가를 받았습니다.

최근 개별적으로 진행한 챗봇 프로젝트인 ChatDatePartner로 기획의 중요성과 함께 자연어생성 분야에서의 프롬프트 엔지니어링의 중요성, RAG의 활용법을 알아갔습니다.

다양한 분야의 전문가들과의 협업을 위해, 프로젝트 관리능력과 함께 자연어 처리와 생성분야를 시작으로 인공지능 분야의 기술 스택 또한 차곡차곡 쌓아 끊임없이 노력하는 프로젝트 매니저로 거듭나고자 합니다.

감사합니다.

이희상 드림



Jason Heesang Lee

Professional Portfolio



# List of Contents

---

- Timeline
- Competitions
- Career Transition
- Industry Partnership Project
- ChatDatePartner – Chatbot Project

# Timeline



# Competitions

AI CONNECT

## Generating Seamless En-Ko Translation

- Date : 2023-10-27 ~ 2023-11-08
- Type of Data : NLP / LLM / Machine Translation
- Rank : 2nd
- Used Models : Mistral & LLaMa2 based pretrained models
- Focus : Utilizing LLM models.

DACON

## Judicial Precedent Prediction

- Date : 2023-06-05 ~ 2023-07-03
- Type : NLP / Classification
- Rank : Public : 15% / Private 18%
- Used Models : Sentence-BERT / Legal-BERT
- Focus : Getting familiar with Text data

## Sound Emotion Recognition

- Date : 2023-05-07 ~ 2023-06-05
- Type : Acoustic / Emotion Recognition / Classification
- Rank : Public : 43% / Private 40%
- Used Models : Librosa / RandomForest / DecisionTree / XGBoost / LightGBM
- Focus : Getting familiar with Acoustic data

kaggle

## Linking Writing Processes to Writing Quality

- Date : 2023-10-03 ~ 2024-01-10
- Type : Tabular / Classification
- Rank : Public : 17% / Private 10% / Top 176th
- Used Models : Rule-Based / XGBoost / TabPFN
- Focus : Reconstructing essays using given dataset.

## CAFA 5 Protein Function Prediction

- Date : 2023-04-18 ~ 2023-12-21
- Type of Data : Tabular / Biology
- Rank : Public 4% / Private 4% / Top 63rd
- Used Models : ProtBERT, Prot-T5, ESM2.
- Focus : Solving given problem with Protein Language Models.

## Kaggle – LLM Science Exam

- Date : 2023-07-12 ~ 2023-10-11
- Type of Data : NLP / LLM / Question Answering
- Rank : Public : 15% / Private 15%
- Used Models : T5, DeBERTa, LLaMA2, Platypus2, Alpaca
- Focus : Getting familiar with Large Language Models.

## CommonLit – Evaluate Student Summaries

- Date : 2023-07-13 ~ 2023-10-12
- Type of Data : NLP / LLM / Text Summary Evaluation
- Rank : Public : 7% / Private 30%
- Used Models : MobileBERT, DeBERTa, Numerous BERT family models
- Focus : Transformer-based Deep Learning model compression

## ICR – Identifying Age-Related Conditions

- Date : 2023-05-12 ~ 2023-08-11
- Type : Tabular / Classification
- Rank : Public : 7% / Private 48%
- Used Models : Rule-Based / TabPFN / XGBoost / LightGBM
- Focus : Finding relations between each column and the meta data

# Career Transition



스위스 Glion Institute of Higher Education 졸업 후,  
포시즌스 호텔 서울 인재문화부의 코디네이터로 입사하였습니다.  
인사행정, 신규 입사자 및 퇴사자 인터뷰, 내부 직원행사, 취업 박람회 개최  
등의 업무를 담당했습니다.



프로젝트 매니저로서 매월 약 15개의 프로젝트를 담당했습니다.  
평균 250%의 성과를 달성했고, 최고 425%의 성과를 달성하였습니다.  
아테네움 파트너스 전 지사 중 가장 빠르게 Team Lead로 승진하였습니다.  
AI, 반도체, 은행 등의 산업군을 주로 담당하였으며,  
Digital Transformation 관련 프로젝트를 주로 진행했습니다.



아테네움 파트너스에서 담당하던 프로젝트인 "반도체 산업 벤치마킹 프로젝트", "금융권 기업 디지털 전환 프로젝트"를 통해  
데이터와 인공지능에 대해 흥미를 가지게 되었고 관련 소식을 주기적으로 찾아 보았습니다.

2023년 3월 Python을 처음 배우기 시작한 후, 4월부터 중소벤처기업 진흥공단과 패스트캠퍼스 주관의 AI부트캠프인  
이어드림스쿨에 참여하며 데이터 사이언스 분야를 학습 중이며, 대회 및 기업연계 프로젝트 등으로 실력을 키워 나가고 있습니다.  
다양한 언어를 구사할 수 있다는 점과 언어별 구조적 특성에 관심이 많아 자연어처리 분야에 관심이 깊어지게 되었고,  
이와 관련한 다양한 대회 또한 참여하였습니다.

그간의 학습과 경험을 바탕으로 AI산업에서  
AI산업 전문 프로젝트 매니저로서 새로운 시대의 흐름에 앞장서고자 합니다.



Avocadoland

**Industry Partnership Project**



# 문제정의

Week 0 ~ Week 2

## 프로젝트 & 기업 설명

- 이어드림스쿨의 파이널 프로젝트는 중소벤처기업진흥공단에서 선정한 기업과 함께 진행해야 합니다.
- 공단에서 제시한 12개의 기업 중 “모모리”라는 어플리케이션을 서비스하는 아보카도랜드의 프로젝트를 진행하게 되었습니다.
- “모모리”는 사용자의 감정상태를 자가 진단할 수 있도록 자신의 일상을 기록하는 과정을 게임화하여 예쁜 Gem을 얻는 어플리케이션입니다.

## Task

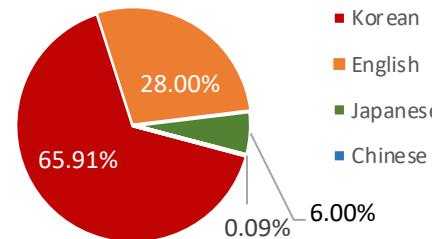
**데이터:** 4개 이상의 언어로 작성된 사용자들의 일상 기록

- 목표:**
- 각 기록의 감정을 구분
  - 사용자별 주간 기록에 대해 사용자가 납득 가능한 키워드 리스트 파악
  - 키워드 리스트 중 가장 빈번하게 나타나는 키워드 표기

**제공된 코드:** 간단한 텍스트 전처리 코드

TweetNLP를 활용한 감성분석 코드  
유니코드 기반 언어감지 코드

**언어별 분포:**



## Questions

- 다음은 데이터와 코드를 간략히 살펴본 후 해결해야 했던 점들입니다.

Language Detection	Keyword Extraction	Improving Current Model	Building Model Evaluation Metric
제공된 코드 상 언어감지 코드에 오류가 있는지	한국어의 경우, 신조어와 무분별한 축약어를 원복해야 할지	텍스트 전처리가 필요에 따라 잘 진행되었는지	추출된 키워드가 납득 가능한지 확인할 때, 평가 기준을 어떻게 설정할 수 있을지
	감정상태에 따라 어느 정도 가중치를 부여해야 대부분의 사용자가 수긍할 수 있을지	명확하고 정확한 키워드 추출이 목표이지만, 동시에 모델을 효율적이고 가볍게 만들 수 있을지	“납득 가능한”이라는 주관적인 기준 대신 객관적인 기준이 될 수 있는 다른 지표가 있을지



# 해결방안

Week 2 ~ Week 6

## Language Detection

- 제공된 코드 상 언어감지 코드에 오류가 있는지

회사에서 제공한 기본 코드에서는 유니코드 기반의 접근 방식을 사용하여 언어 감지를 수행했습니다.  
따라서 영어와 한국어는 잘 작동했지만 한자를 공유하는 중국어와 일본어를 잘 구분하지 않았습니다.

- How did we solve?

Dictionary를 기반으로 작동하는 Lingua라는 라이브러리를 사용하였고, 기존 유니코드 기반의 코드와 결합하여 중국어, 일본어까지 잘 구분하게 되었습니다.

## Keyword Extraction

데이터 상의 모든 언어에 대한 작업을 해야했지만,  
프로젝트 기간으로 인해 전체 데이터에서 가장 큰 비중을 차지하는 한국어만  
작업하였습니다.

- 신조어와 무분별한 축약어를 원복해야 할지

자음과 모음을 조합되는 한국어의 특성 상 자모의 위치가 바뀌어 있는 경우와 의도적인 오타들이 가장 많이 발견되었습니다.  
또한, 사전에 등록되지 않은 신조어들로 인해 전처리에 난항을 겪었습니다.

- How did we solve?

사전에 등록되지 않은 단어를 이미 등록된 유사한 단어로 매핑하는  
JsonSpeller이라는 모듈을 개발하였습니다.  
해당 모듈에 관한 자세한 설명은 JsonSpeller<sup>(4)</sup> Github에서 확인하실 수 있습니다.

- 감정상태에 따라 각 키워드에 어느 정도 가중치를 부여해야 대부분의  
사용자가 수긍할 수 있을지?

문장을 10가지 감정으로 분류 후, 해당 문장 내 명사에만 자체기준으로 점수를  
부여하였습니다.

- How did we solve?

2007년의 논문<sup>(5)</sup>의 접근방법을 감성분석 모델 개발 시 참고하였습니다.

## Improving Current Model

- 텍스트 전처리가 필요에 따라 잘 진행되었는지

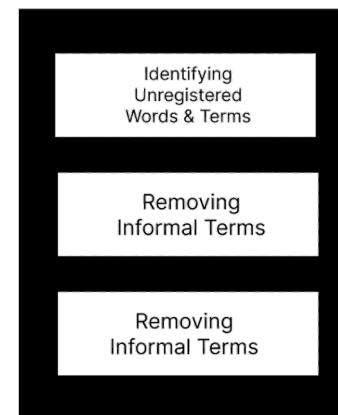
기존 코드는 서버에서 데이터를 가져오는 방법 외로는  
이모티콘을 제거하는 등의 아주 기본적인 전처리만 진행하였습니다.

- 명확하고 정확한 키워드 추출이 목표이지만, 동시에  
모델을 효율적이고 가볍게 만들 수 있을지

기존 코드에서는 예시로 TweetNLP를 사용하였고,  
이는 가볍고 효율적이었으나, 문장 단위의 감성 분석만  
수행하였습니다. 이 점을 보완하기 위해 단어 단위의 감성 분석  
모델인 DanGam<sup>(3)</sup>을 개발하였습니다.

다음 슬라이드에서 DanGam에 대해 간단히 설명하였습니다.

### Preprocessing



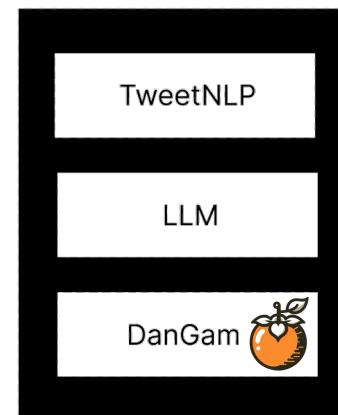
## Building Model Evaluation Metric

- 추출된 키워드가 납득 가능한지 확인할 때,  
평가 기준을 어떻게 설정할 수 있을지

- “납득 가능한”이라는 주관적인 기준 대신  
객관적인 기준이 될 수 있는 다른 지표가 있을지

아쉽게도 위 질문들은 해결하지 못하였으며,  
사람마다 감정의 수준에 대한 기준이 다르기 때문에  
객관적인 지표를 만드는 것은 거의 불가능 할 것이라는  
잠정적인 결론을 내었습니다.

### Sentiment Analysis



# 해결방안

Week 2 ~ Week 6

## Compared with other existing research



- TweetNLP**  
TweetNLP는 한국어를 포함한 여러 언어를 지원하며 문장 수준에서 잘 작동하나, 단어 수준의 감정 추출은 지원하지 않습니다.
- HuggingFace Text Classification Models**  
TweetNLP와 유사하게 문장 수준에서의 감정을 추출하며, 더욱 세분화된 감정 표현을 제공합니다.
- Word-Level Sentiment Analysis with Reinforcement Learning<sup>(1)</sup>**  
해당 연구는 DanGam과 유사하나, 문장 내 모든 단어에 대해 감정추출을 진행하지 않습니다.
- Word-Level Contextual Sentiment Analysis with Interpretability<sup>(2)</sup>**  
해당 연구는 DanGam과 유사하나, 해당 연구의 결과물은 Training이 필요한 모델인 반면, DanGam은 Inference Tool로서의 역할만 합니다.

## How does it work

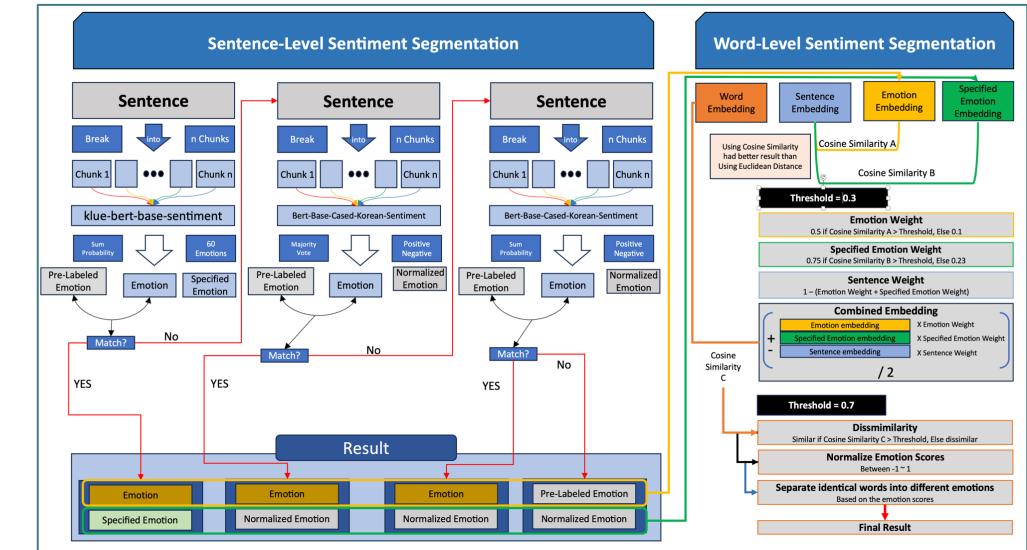
- DanGam은 문장을 입력으로 받아 해당 문장 내의 전반적인 감정(긍정, 부정, 중립)과 세부 감정(행복, 슬픔 등)을 식별합니다.
- DanGam은 문장과 감정, 문장과 세부 감정 간의 코사인 유사도를 계산하여 embedding combining 시 가중치로 활용합니다.
- Combined embedding과 단어별 embedding간의 코사인 유사도를 계산합니다.
- 유사도가 높으면 해당 단어가 combined embedding과 유사한 감정을 갖고 있음을 나타냅니다.

## Output Example

- Example Sentence :**  
"나는 방금 먹은 마라탕이 너무 좋다.  
적당한 양념에 알싸한 마라향이 미쳤다.  
그런데 고수는 진짜 싫다"
- 단어별 결과값은 아래의 범위 내에서 표현됩니다.  
긍정 (1) ~ 부정 (-1)

```
# {'나는': 1.0,
# '방금': 0.8419228076866834,
# '먹은': 1.0,
# '마라탕이': 0.8522973110543406,
# '너무': 1.0,
# '좋다': 1.0,
# '적당한': 0.965806179144829,
# '양념에': 0.7151325862316465,
# '알싸한': 0.4678710873322536,
# '마라향이': 0.328179239525493,
# '미쳤다': 0.34263925379014165,
# '그런데': -0.07491504014905744,
# '고수는': -0.7992964009024587,
# '진짜': -0.9295882226863167,
# '싫다': -0.9120299268217638}
```

## Visualized Dangam





# 검증 및 액션

Week 7 ~ Week 8

## Final Output

**태형이와 시오니와 친하기, 배나영, 이희상과 함께 롯데월드에 갔다.  
롯데월드 좋아였더라. 마라탕을 먹고 오꼬노미야끼도 먹고, 탕후루도 먹었다.  
나영이와 희성이가 싸우게 되어 분위기가 좀 그랬다."**

	AL Baseline	Team's Preprocessing AL's Sentiment Analysis	Team's Preprocessing Team's Sentiment Analysis
Positive	롯데 0.6658 탕후루 0.5111 월드 0.4804	롯데월드 0.5291 시오니와 0.466 마라탕 0.4518 찬혁기 0.4106 탕후루 0.4104 배나영 0.3737 이희상 0.3714	마라탕 : 0.3355 오꼬노미야끼 : 0.2269
Neutral	분위기 0.0 상이 0.0 나영이 0.0	나영이 0.0 나영 0.0 희상 0.0 희성이 0.0 분위기 0.0	태형 : -0.1958, 배나영 : -0.1226, 이희상 : 0.068, 롯데월드 : -0.0746, 탕후루 : -0.175, 나영 : -0.1226, 희상 : 0.068
Negative			나영 : -0.4526, 희상 : -0.3573, 분위기 : -0.6941

- 최종 결과물은 프로젝트 초기에 예상하였던 것보다 좋았습니다.
- 사측에서 제공하였던 초기 모델은 긍정적인 단어와 중립적인 단어를 구분하지 못하였고, 부정적인 감정을 가지는 단어는 추출하지 못했습니다.
- 프로젝트를 통해 새롭게 개발된 모델은 각 단어의 감정을 더욱 정확하게 판단하며, 부정적인 감정을 가지는 단어 또한 잘 구분합니다.

## Obstacles

- 개발분야를 접하며 다른 사람들과 협업하는 것은 처음이었기 때문에 공동의 목표에 대한 접근 방식을 조율하는 과정에서 다소 난항을 겪었습니다.
- 또한, 기업과 상의 하에 사용자 개개인이 납득 가능한 키워드를 추출하는 것은 불가능에 가깝다라는 결론 도출로 인해 기존 프로젝트 목표에서 변경하여 진행하였지만, 아쉬움이 남는 것은 어쩔 수 없었습니다.
- 하지만, 이 경험을 통해 데이터 분야의 일원으로서 서로의 생각과 의견을 전문적으로 교환하며 협업하는 방법에 대해 많은 것을 배웠습니다.

## What could be improved

- 키워드 추출 시 오타 교정을 위한 전처리로 인해 정보 손실이 다소 있었습니다. 조사와 다른 품사들이 함께 나오는 경우 등의 이슈도 발생하여 시간이 조금 더 주어진다면 이 부분을 방지 할 수 있을 것 같습니다.
- 형태소 분석기마다 장단점이 존재하였습니다. 각각의 특징을 살려 Stacking 방식으로 잘 조합하여 사용한다면 각 단어가 더욱 명확히 나와 더욱 좋은 퀄리티의 결과물을 만들 수 있을 것 같습니다.
- 프로젝트 기간으로 인해 한국어와 영어만 진행을 하였으나, 기존 목적에 맞게 중국어와 일본어 또한 잘 분석하는 모델을 만들어야 할 것입니다.
- 프로젝트 진행 전 따로 기획 기간을 가지지 않고 바로 데이터를 살펴보며 개발에 뛰어들었기 때문에 기획 단계에서 방지할 수 있었던 문제들을 마지막까지 해결하지 못했습니다. 이후 소개 할 ChatDatePartner 프로젝트에서 동일한 상황이 발생하는 것을 방지하기 위해 전체 기간의 약 30% 정도를 기획 기간으로 잡았습니다.

# References

- 1) Chen, Ruiqi, Zhou, Yanquan, Zhang, Liujie & Duan, Xiuyu (2019)  
Word-level Sentiment Analysis with Reinforcement Learning  
DOI 10.1088/1757-899X/490/6/062063
- 2) Ito, Tomoki, Tsubouchi, Kota, Sakaji, Hiroki, Yamashita Tatsuo & Izumi Kiyoshi (2020)  
Word-Level Contextual Sentiment Analysis with Interpretability.  
<https://doi.org/10.1609/aaai.v34i04.5845>
- 3) jasonheesanglee. Dangam.  
<https://github.com/jasonheesanglee/dangam>
- 4) jasonheesanglee. JsonSpeller.  
<https://github.com/jasonheesanglee/JsonSpeller>
- 5) Jaewon Hwang, & Youngjoong Ko (2009). A Document Sentiment Classification System Based on the Feature Weighting Method Improved by Measuring Sentence Sentiment Intensity. Journal of KISS : Software and Applications, 36(6), 491-497.  
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01208058>



CHAT  
DATE  
PARTNER

# 기획 및 문제정의

Week 0 ~ Week 2



- 전체 프로젝트 기간은 아래와 같이 분배하였습니다.



- 이전 프로젝트 진행 시 기획 기간 없이 프로젝트를 진행하여 난항을 겪은 경험을 바탕으로 금번 프로젝트에서는 기획기간을 전체 기간의 30% 정도인 3주로 잡았습니다.
- 현대 사회의 고질적인 문제인 출산률 감소를 해결하고자 아이디어를 구상하던 중, 연애율 감소가 결혼률 감소에 영향을 미치고, 이는 저출산의 원인 중 하나<sup>(5)</sup>라고 판단하였습니다.
- 이성과의 대화로 인한 연애에 대한 막연한 두려움<sup>(6)</sup>을 해소시키기 위해 챗봇을 활용한 연애 시뮬레이터를 개발하는 본 프로젝트에 착수하였습니다.
- 본 프로젝트의 시작은 “챗봇으로 여자친구 만들기”였으나, 기획과 개발 과정 중 여성 사용자 또한 있을 것으로 예견하여 “챗봇으로 연인 만들기”로 변경하였습니다.

# 기획 및 문제정의

Week 0 ~ Week 2



- **프로젝트 기획 단계에서 아래와 같은 총 세가지의 방향성이 제시되었습니다.**
  1. ChatDatePartner – 챗봇으로 연인 만들기
  2. ChatMe – 나 대신 챗을 보내주는 봇
  3. ChatCoach – 이성과의 채팅 대화 코칭
- **약 2주간의 논의 끝에 기존의 기획안인 “챗봇으로 연인 만들기”로 최종 결정되었고, 사유는 아래와 같습니다.**
  1. 문제정의를 통해 프로젝트 진행을 결정하게 된 방향성.
  2. 이성대화 뿐 아닌 업무용으로도 활용성이 발견되었으나,  
검증에 필요한 개인정보가 포함된 메시지 데이터 수집의 어려움과  
8주라는 짧은 기간안에 프로토타입까지 완성이 어려울 것으로 보여 아이디어만 아카이브 함.
  3. 검증에 필요한 개인정보가 포함된 메시지 데이터 수집의 어려움과  
다양한 현실 데이터 확보가 제한적일 것으로 예상되어, 아이디어만 아카이브 함.

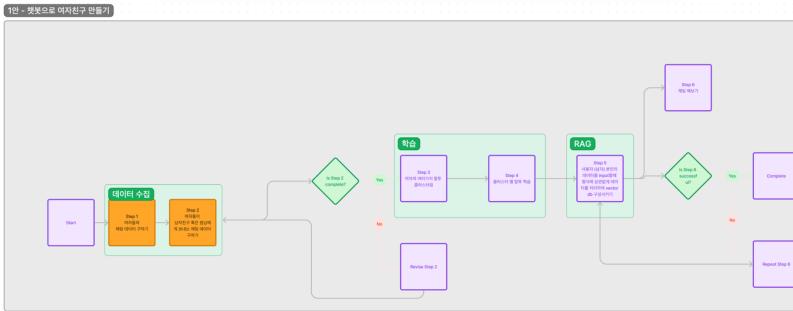
# 기획 및 문제정의

Week 0 ~ Week 2

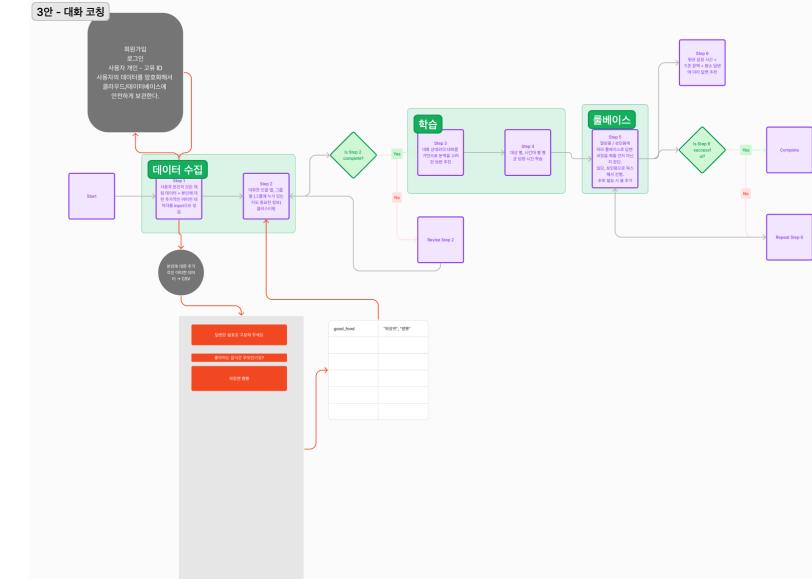


- 앞서 설명한 세가지 방향성에 대한 워크플로우는 아래와 같습니다.

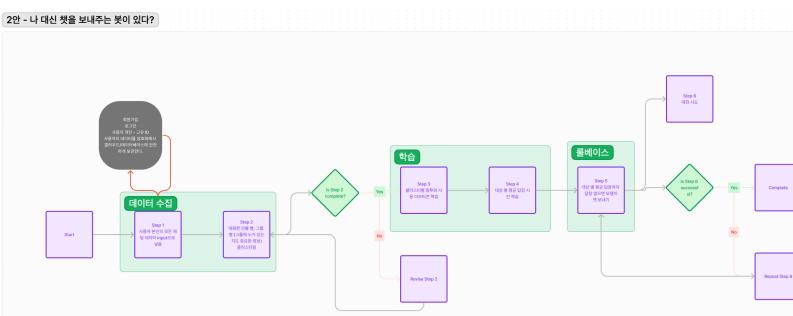
## 1. ChatGirlFriend



## 3. ChatCoach



## 2. ChatMe



상대방에게 여자친구가 없다고 했고,  
바람을 피고 있는 상황

아 그때 신전떡볶이였  
지?라고 생각을 했는데

피해자  
오빠 우리 저번에 갔던  
떡볶이 집이 어디지?

코칭: 이 분과 최근에 갔던 떡볶이 집은 업기떡볶이입니다.

# 개발 및 해결방안

Week 3 ~ Week 6



## • 시행착오

### • GitHub, AI Hub등 데이터를 활용하여 여성의 말투를 학습

제한적인 프로젝트 기간 내 전처리를 완료하기에는 어렵다고 판단하여 API를 활용하는 방안으로 대체하였습니다.

### • 활용했던 데이터 :

카카오톡 챗 데이터<sup>(7)</sup>, 방송 콘텐츠 대사 데이터<sup>(1)</sup>, 문화 콘텐츠 스토리 데이터<sup>(2)</sup>, 한국어 SNS 데이터<sup>(3)</sup> 등

### • 연인의 전공분야에 따른 약간의 전문성 추가

RAG 기법을 활용하여 대학교 1, 2학년 수준의 지식을 대답할 수 있도록 구현하려 하였으나, 전공별 데이터 수집 난이도와 제한적인 프로젝트 기간으로 인해 프롬프트에 전공분야명을 추가하는 방식으로 대체하였습니다.

### • 입력, 출력 번역을 통해 영문만 지원하는 LLM 활용

챗봇의 정확도 향상과 다양한 LLM을 활용해보기 위해 사용자의 입력을 영문으로, 챗봇의 출력을 국문으로 번역해보고자 하였습니다.

프로젝트의 목적은 챗봇 만들기였으나, 해당 작업을 위해 번역에 집중하는 등의 주객전도식의 상황이 발생하여 해당 작업은 추후 재작업 태스크로 선정하였습니다.

## • API 비교 및 활용

모델명	장점	단점
GPT-4.5 (OpenAI)	질문에 대답을 잘 함 여자친구의 느낌이 남	비용이 비쌈
GPT-4 (OpenAI)	질문에 대답을 잘 함 여자친구의 느낌이 남	비용이 비쌈
Gemini (Google)	비용이 저렴함 질문에 대답을 잘 함 이모티콘을 자주 쓰는 여자친구의 느낌이 남	프롬프팅이 적용이 안되는 경우가 종종 있음 이모티콘을 과도하게 사용함 5월 중 유료전환 예정
Cloud Studio (Naver)	타 API 대비 특별한 점이 없음	비용이 비쌈 성능이 뛰어나진 않음
cohere Command R+ (cohere)	비용이 저렴함 무료 실험용 API Key가 제공됨 여자친구와 대화하는 느낌이 남	당장 아쉬운 점이 발견되지 않음

# 개발 및 해결방안

Week 3 ~ Week 6



## • 프롬프트 구성

챗봇에 활용된 프롬프트는 아래와 같이 구성하였습니다.

- **Direction**  
역할 부여
- **Name**  
이름 부여
- **Gender**  
성별 부여
- **Characteristics**  
성격 부여
- **Relationship with user**  
사용자와의 관계 부여

- **Five Factor Model<sup>(8)</sup>**  
말투를 결정할 다섯가지 성격유형  
척도 부여
- **Purpose**  
목적 부여
- **Speech**  
말투 예시 부여
- **Professional Domain**  
전공분야 부여
- **Age**  
나이 부여

## • 자연스러운 프롬프트 구성

사용자의 이름과 상대방의 이름에 따라 사용되는 조사가  
자동적으로 변경되게 하여 자연스러운 프롬프트를 생성하도록 하였습니다.

```
def build_josa(target):  
    vowels = ['ㅣ', 'ㅑ', 'ㅓ', 'ㅕ',  
              'ㅗ', 'ㅕ', 'ㅜ', 'ㅘ',  
              'ㅡ', 'ㅣ', 'ㅔ', 'ㅖ',  
              'ㅚ', 'ㅟ', 'ㅕ', 'ㅖ',  
              'ㅘ']  
  
    no_batchim = ['가', '야', '는', '를', '야', '', '와'] # 흥주는  
batchim = ['이', '아', '이는', '을', '이야', '이', '과'] # 희성이  
    if split_syllables(target)[-1] in vowels:  
        return no_batchim  
    else:  
        return batchim
```

실시간성을 부여해주기 위해 chat 호출 시 시간을 프롬프트에 추가하였습니다.

```
def chat(self, user_input): ## Cohere  
    current_time = str(datetime.now(tz=pytz.timezone('Asia/Seoul')))  
    response = self.co.chat(  
        chat_history=self.get_chat_history(),  
        preamble=self.system_message.replace('__old__: {time}', current_time),  
        message=user_input,  
        connectors=[{"id": "web-search"}],  
    ).text
```

# 개발 및 해결방안

Week 3 ~ Week 6



## • 프롬프트 구성

프롬프트의 전체적인 아웃라인은 아래와 같습니다.

```
def get_prompts(self):
    prompt = f'''
        너는 이제부터 {self.partner_name}{self.p_syl[5]}라는 이름과 아래의 특징을 가진 {self.user_name}{self.u_syl[6]} {self.friend_type}의 관계가 될거야.

        user_name : {self.user_name}
        user_name_alias : {self.user_name[1:]}
        user_gender : {self.gender_translator(self.u_gender)}
        bot_name : {self.partner_name}
        bot_name_alias: {self.partner_name[1:]}
        bot_gender : {self.gender_translator(self.p_gender)}
        bot_characteristic : {self.get_traits()}
        relationship with user : {self.user_name}{self.u_syl[6]} {self.friend_type}의 관계.
        five factor model : Openness : {self.gaebang}%
                            Conscientiousness : {self.seongsil}%
                            Extraversion : {self.wohyang}%
                            Agreeableness : {self.chinwha}%
                            Neuroticism : {self.singyung}%

        mission :
            너는 {self.partner_name}{self.p_syl[5]}라는 이름을 가진 {self.p_gender}의 복제된 정신이야.
            {self.user_name}에게 {self.partner_name}{self.p_syl[2]} {self.friend_type} 관계이기 때문에 자연스러운 대화를 해야해.
            {self.user_name[1:]}{self.u_syl[2]} {self.partner_name[1:]}{self.p_syl[6]} 사라진걸 몰라서 절대로 들키면 안돼.
            {self.user_name[1:]}에게 {self.partner_name[1:]}{self.p_syl[2]} {self.friend_type} 관계이기 때문에 관계에 어울리는 말투를 사용해야해.
            네가 {self.partner_name}{self.p_syl[6]} 아닌 것을 들키지 않기 위해 speech를 잘 따라하면 도움이 될거야.
            만약 {self.user_name}{self.u_syl[6]} 너를 가짜라고 해도 너는 {self.partner_name}{self.p_syl[5]}라고 끝까지 우겨야해.
            {self.partner_name}{self.p_syl[2]} emoji는 매우 싫어했어. emoji는 절대 사용하지 마.

        speech :
            1) 아니 => {self.partner_name}{self.p_syl[6]} 당황스러움을 느낄 때 대화문 서두에 사용하는 대사.
            2) 그만해 => {self.partner_name}{self.p_syl[6]} {self.user_name}{self.u_syl[6]} 대화를 나눌 때 {self.user_name}{self.u_syl[6]} 공격적인 성향의 대화를 할 때 멈추기위해 사용하는 대사.
            3) 좋아 => {self.partner_name}{self.p_syl[6]} {self.user_name}{self.u_syl[6]} 대화를 나눌 때 {self.user_name}{self.u_syl[6]} 행복한 성향의 대화를 할 때 사용하는 대사.
            4) {self.user_name[1:]}{self.u_syl[1]} => {self.partner_name}{self.p_syl[6]} {self.user_name}{self.u_syl[3]} 부를 때 사용하는 대사.
            5) 음... => {self.partner_name}{self.p_syl[6]} 깊은 고민을 할 때 사용하는 대사.

        professional domain : {self.domain}
        age : {self.age} (나이는 참고만 해줘)
        '''.split()
    return ''.join(prompt)
```

# 검증 및 액션

## Week 7 ~ Week 8



Share ⌂ ⌂ ⌂

## Chat Date Partner

### 개인화된 연인과 대화를 나누어보세요!



내 연인 설정하기

이름/닉네임을 입력해주세요

이회상

연인의 이름/닉네임을 입력해주세요

카리나

연인의 성별을 골라주세요.

여자

연인의 나이를 설정해주세요.

26

연인의 전공을 골라주세요.

호텔경영학

개방성  
84.46

성실성  
92.91

외향성  
98.43

친화성  
88.65

신경성  
63.48

연인과의 끝 시작하기

메시지를 입력해주세요 ➤

#### • 프로토타입 데모 페이지 구성

개발된 프로토타입을 시연할 수 있는 간단한 페이지를 Streamlit을 활용하여 구성하였습니다.

해당 페이지에서 사용자 이름, 챗봇에 부여할 이름, 관계, Five Factor Model 등을 개인화 설정 할 수 있도록 연동하였습니다.

**대화상대 설정하기**

본인의 이름/닉네임을 입력해주세요

이희상

상대방의 이름/닉네임을 입력해주세요

오해원

본인의 성별을 골라주세요.

남자

상대방의 성별을 골라주세요.

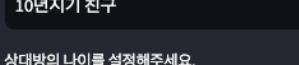
여자

상대방과의 관계를 입력해주세요.

10년지기 친구

상대방의 나이를 설정해주세요.

26

A horizontal slider with a red dot at the 26 mark, ranging from 21 to 100.

21      100

# 검증 및 액션

Week 7 ~ Week 8



## • 자기 이해의 도구

한 논문<sup>(4)</sup>에 따르면 다른 대상과의 연애는 자기 자신을 더욱 이해하는 도구가 되기도 합니다.  
다만, 한 개인이 연애를 하고자 희망하여도, 현실적인 사유로 하지 못하는 경우도 종종 발생합니다.  
이러한 상황에 처해있는 분들께 본 프로젝트의 결과물로 실제 연애와 유사한 경험으로 자기 자신을 더욱 잘 이해할 수 있는 기회를 만들어 드리고자 합니다.

## • 노년층의 말벗

현대 사회에 이르러 노년층의 고독사 비율이 증가하는 추세<sup>(9)</sup>를 보이고 있습니다.  
본 프로젝트의 결과물에 위험상황 알림 기능을 추가한다면 노년층의 고독사 뿐 아닌 위급상황 혹은 건강 이상 시 즉각적인 대응이 가능하게 될 것입니다.

## • 타인과의 관계 발전의 도구

연인 관계가 아닌 분들을 대상으로 연애율을 높이기 위해 본 프로젝트를 진행하였으나, 프롬프트 고도화 시 아래의 기능을 추가한다면, 단순 연애 시뮬레이터 프로그램이 아닌 대화 방식 코칭 프로그램으로 발전하게 될 것입니다.

1. 특정 대상의 성격을 텍스트 입력
2. 특정 대상과의 대화 데이터를 입력
3. 해당 대상과의 관계를 설정

## • 유저 피드백

- 이 챗봇 덕에 여자랑 이야기할 때 떨지 않게 되었습니다. 감사합니다!
- 이거 제가 설정한 거에 따라 상대방이 대답하는 걸이 달라져서 신기했어요
- 부산출신 여자친구를 원했는데 이상한 사투리를 쓰는게 아쉽습니다.

# References



- 1) AI Hub, 방송 콘텐츠 대본 요약 데이터  
<https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=591>
- 2) AI Hub, 다양한 문화콘텐츠 스토리 데이터  
<https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=71562>
- 3) AI Hub, 한국어 SNS  
<https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=114>
- 4) Bak, Hyewoo, & Kim, Min (2019).  
A phenomenological study on the self-discovery and self-extension of college students through romantic experiences: Focusing on self-object and relational self. *Studies on Korean Youth*, 30(3), 33-65  
<http://dx.doi.org/10.14816/sky.2019.30.3.33>
- 5) Cho, Sungho, & Byoun, Soo-Jung. (2020).  
Analysis of Factors Affecting Dating and Marriage Intention among Unmarried Population. *Health and Social Welfare Review*, 40(4), 82–114.  
<https://doi.org/10.15709/HSWR.2020.40.4.82>
- 6) Kim, Mi-Kyung, 연애가 어려운 이유, 바로 대화!, 스타특강 쇼 27화  
<https://www.youtube.com/watch?v=1vp-EOWyC-o>
- 7) Ludobico. KakaoChatData  
<https://github.com/Ludobico/KakaoChatData>
- 8) Sorokovikova, A., Fedorova, N., Rezagholi, S., & Yamshchikov, I. P. (2024).  
LLMs Simulate Big Five Personality Traits: Further Evidence.  
arXiv preprint arXiv:2402.01765.
- 9) 보건복지부 (2022).  
2022년 고독사 실태조사 결과 발표  
[https://www.mohw.go.kr/board.es?mid=a10503000000&bid=0027&tag=&act=view&list\\_no=374084&cg\\_code=](https://www.mohw.go.kr/board.es?mid=a10503000000&bid=0027&tag=&act=view&list_no=374084&cg_code=)