

# PORTFOLIO

---

NLP Research Scientist

금빛나 (Bitna Keum)

# CONTENTS

- Profile
- Research Experiences
- Development Experiences

# Profile



금빛나 (Bitna Keum)

☎ +82-10-3446-3606

✉ [beausty23@naver.com](mailto:beausty23@naver.com)

🐙 [GitHub](#)

📖 [Tech Blog](#)

## Educations

2024 건국대학교 인공지능학과 석사 졸업  
(GPA: 4.37 / 4.5, Lab: 자연어처리연구실)

2022 광운대학교 컴퓨터정보공학부 학사 졸업  
(GPA: 4.16 / 4.5)

2017 청학고등학교 졸업

## Research Interests

- Chatbot
- Personalization
- Generative AI
- LLM

## Work Experiences

2021 TUNiB — Data & Software Engineer Internship

2019 GNC Solution — AI Developer Internship

## Skills

Programming Languages Python, C, Java, JavaScript, SQL

Deep Learning Frameworks PyTorch, Tensorflow

Cloud Computing Services AWS, GCP

DevOps Docker, Git

# Profile



금빛나 (Bitna Keum)

☎ +82-10-3446-3606

✉ [beausty23@naver.com](mailto:beausty23@naver.com)

🐙 [GitHub](#)

📖 [Tech Blog](#)

## Awards

2023 우수 논문상, 제35회 한글 및 한국어 정보처리 학술대회

2022 우수 논문상, 제34회 한글 및 한국어 정보처리 학술대회

2021 IITP 원장상, AI GRAND CHALLENGE

2020 총장상, 광운대학교 인공지능 활용 아이디어 공모전

## Scholarships

2023 우수연구장학금, 건국대학교

2023 수업조교장학금, 건국대학교

2023 RUS멘토장학금, 건국대학교

2022 우수연구장학금, 건국대학교

2021, 2020, 2018 성적우수장학금, 광운대학교

## Extracurricular activities

2023 컴퓨터공학부 기계학습 수업 TA, 건국대학교

2023 학부생 연구인턴 프로그램 멘토, 건국대학교

2023 데이터 크리에이터 캠프(DCC) 멘토, NIA

2022 Text-to-SQL 데이터셋 어노테이터, LG

2021, 2019 SW학교 튜터, 광운대학교

# Profile

## Papers

### International Journals

- [1] Bitna Keum, Juoh Sun, Woojin Lee, Seongheum Park, and Harksoo Kim, "Persona-Identified Chatbot through Small-Scale Modeling and Data Transformation", *Electronics (SI: Emerging Theory and Applications in NLP)*, Vol. 13, no. 8, 2024. (DOI: 10.3390/electronics13081409) (SCIE, IF: 2.9, Q2)

### International Conferences

- [1] Jeonghyun Kang, Hongjin Kim, Bitna Keum, Jaieun Kim, and Harksoo Kim, "Generation-Based and Emotion-Reflected Memory Update: Creating the KEEM Dataset for Long-Term Conversation", under review for the 62nd Annual Meeting of the Association for Computational Linguistics, 2024.

### Domestic Journals

- [1] 금빛나, 김홍진, 황금하, 권오욱, 김학수, "가이드 응답의 노이즈 제거 학습을 통한 개선된 오픈 도메인 대화 생성 모델", *정보과학회/논문지*, vol. 50, no. 10, pp. 851-858, 2023. (DOI: 10.5626/jok.2023.50.10.851)

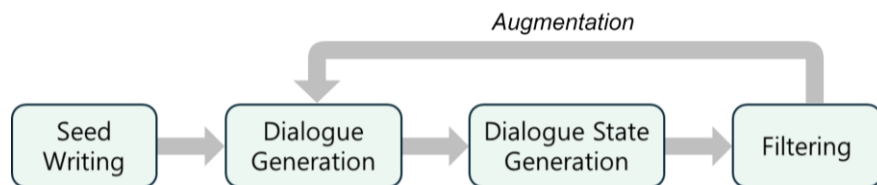
### Domestic Conferences

- [3] 금빛나, 김홍진, 황금하, 권오욱, 김학수, "페르소나 기반의 장기 대화를 위한 다각적 어텐션을 활용한 생성 모델", *제35회 한글 및 한국어 정보처리 학술대회*, pp. 281-286, 2023. (우수논문)
- [2] 박상민, 손유리, 금빛나, 김홍진, 김학수, 김재은, "멀티턴 대화를 활용한 레퍼런스 기반의 발화 생성 모델", *제34회 한글 및 한국어 정보처리 학술대회 논문집*, pp.88-91, 2022.
- [1] 금빛나, 김홍진, 박상민, 김재은, 황금하, 권오욱, 김학수, "오픈 도메인 대화를 위한 노이즈된 가이드 기반 생성 모델", *제34회 한글 및 한국어 정보처리 학술대회 논문집*, pp.82-87, 2022. (우수논문)

# Research Experiences

## 사용자 맞춤형 추천을 위한 LLM 기반 챗봇

- Dialogue state tracking을 활용해 사용자 맞춤형 추천을 하는 챗봇
  - 목표: LLM 기반 챗봇이 보편화된 추천을 하는 경향을 개선하기 위해, 사용자 중심의 정보를 바탕으로 맞춤형 추천을 제공
- Descriptions
  - 매 턴마다 dialogue state tracking을 통해 사용자로부터 얻은 정보를 축적하고 추가적으로 필요한 정보를 사용자에게 질의함으로써 맞춤형 추천 응답을 제공
  - Self-Instruct 방식을 차용해 instruction 데이터를 단계별 구축 및 증강



- 구축한 데이터로 Llama-2에 instruction-tuning 수행
- Unseen topics에서도 응답의 개선된 경향성을 확인

**[Original Chatbot]**

User: I'm going to have steaks with my family, Can you recommend a restaurant?

Chatbot: Certainly! Here are some highly-rated steakhouse restaurants that could be great options for your family dinner:  
1. Ruth's Chris Steak House: Known for its USDA Prime aged beef and sizzling plates, Ruth's Chris is a popular choice for steak lovers. Many locations offer an extensive wine list and upscale ambiance.  
2. The Capital Grille: With a focus on dry-aged steaks, fresh seafood and an impressive wine collection, The Capital Grille offers a classic and sophisticated dining experience.  
3. Morton's The Steakhouse: A longtime favorite for high-quality steaks, Morton's is known for its generous portions and extensive menu including appetizers, entrees and desserts.  
4. Fleming's Prime Steakhouse & Wine Bar: Fleming's offers a variety of prime aged USDA steaks, as well as seafood dishes and an extensive wine list. Their warm, inviting atmosphere makes for a great dining experience.  
5. Texas de Brazil: For a unique twist on the classic steakhouse experience, consider Texas de Brazil. They offer a variety of grilled meats (including steaks) as well as a salad bar and a dessert bar. The all-you-can-eat format is perfect for large groups and families!

These are just a few suggestions based on popular reviews and feedback from customers. Be sure to check if they have locations near you and make reservations in advance to ensure availability!

**[Instruction-tuned Chatbot]**

User: I'm going to have steaks with my family, Can you recommend a restaurant?

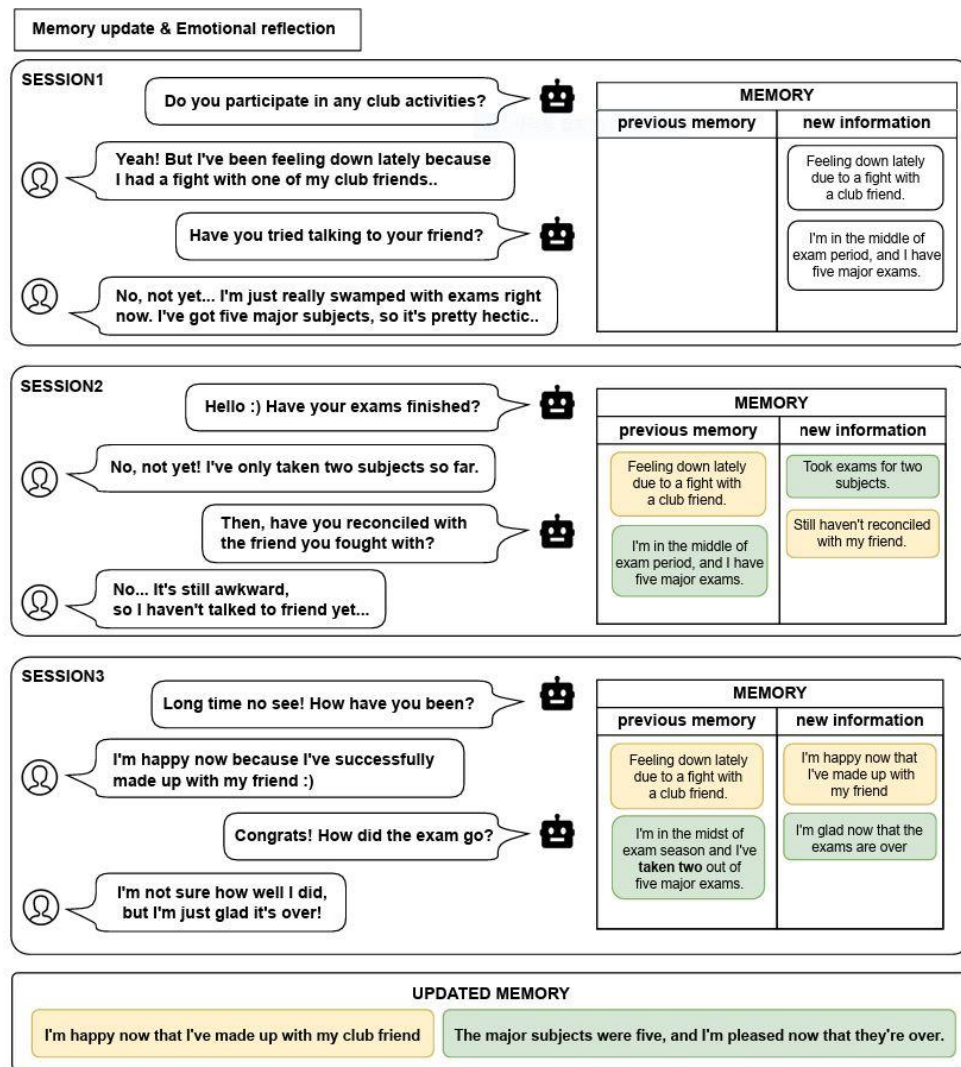
Chatbot: Are you looking for a specific type of steakhouse—perhaps something more upscale or a cozy family-friendly place?

Metadata:   
cuisine: Steakhouse  
group\_size: family  
location: None  
price\_range: None  
ambiance: None  
dining\_preference: None

# Research Experiences

Generation-based and Emotion-reflected Memory Update for Long-Term Conversation (*ACL 2024 under review*)

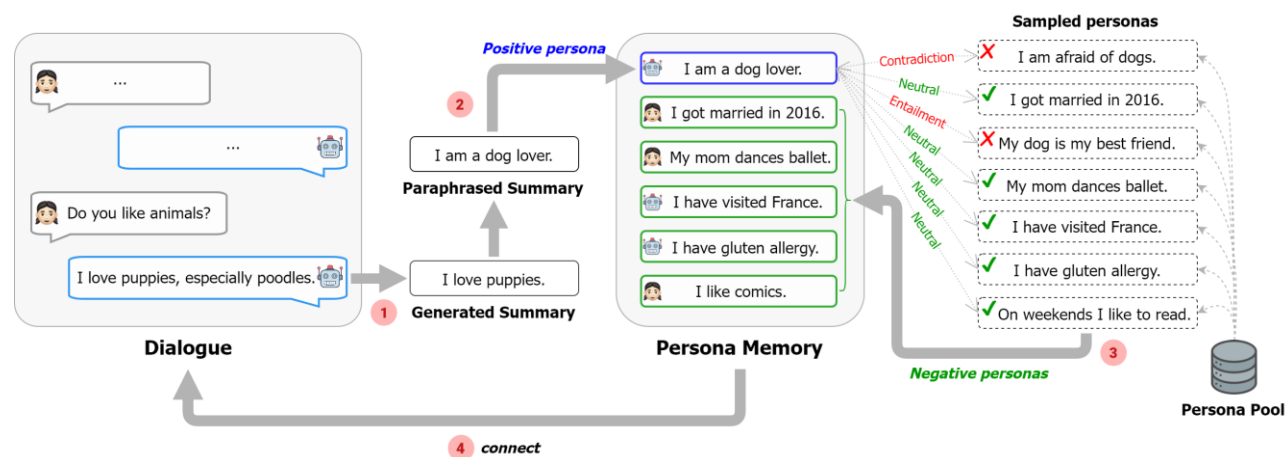
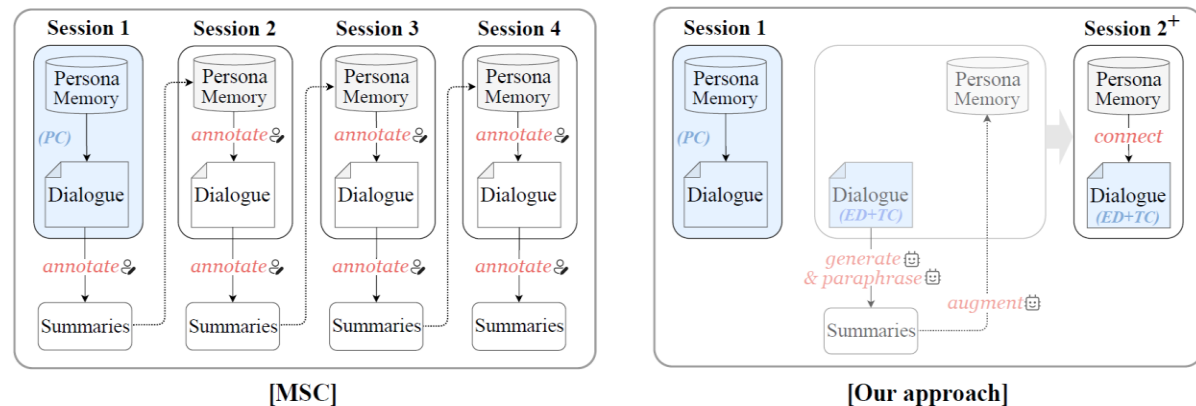
- 메모리에 저장된 정보에 감정과 원인이 반영되도록 수정하고, 메모리 업데이트를 위한 데이터셋 구축
  - 목표: 메모리에 저장되는 정보에 감정과 원인을 반영함으로써 기존 데이터셋의 한계점을 개선하며, 메모리 업데이트를 위한 데이터가 희소한 점에 기여
- Descriptions
  - 각 세션마다 대화 내용에 대한 요약이 메모리에 저장되는데, 기존 데이터셋은 주로 감정과 원인이 배제되어 있기 때문에 메모리의 품질이 낮은 경향이 있음
  - 감정과 원인을 반영한 대화 요약 seed 데이터를 작성하고 LLM에 few-shot으로 주어 기존 데이터셋의 요약을 수정
  - 메모리 업데이트 seed 데이터를 작성하고 LLM에 few-shot으로 주어 기존 데이터셋으로부터 메모리 업데이트 데이터를 만들어 냄
  - 정성평가 결과, 기존 데이터셋 대비 현저한 품질 향상을 이룸



# Research Experiences

## Cost-Efficient Data Synthesis Framework for Long-Term Personalized Conversations

- 저비용으로 장기 대화에 효과적인 데이터를 합성하는 framework
  - 목표: Multi-session으로 구성된 대화 데이터가 장기 대화에 효과적임이 알려져 있지만, 구축하는 데에 많은 비용이 드는 한계점을 해결
- Descriptions
  - 공개된 single-session 일상 대화를 대상으로, 다음의 프로세스를 거쳐 페르소나 기반 대화로 변형
    1. 대화로부터 personal traits에 관한 요약 generation
    2. 요약을 paraphrasing하여 메모리의 \*positive persona로 구성
    3. 샘플링한 페르소나들 중 positive persona와 Neutral 관계인 것들만 메모리의 \*negative persona로 증강
    4. 새로 구축된 페르소나 메모리를 기존 대화에 연결시킴
  - 평가 결과, 저비용으로 구축한 본 데이터가 고비용으로 구축된 MSC와 준하는 성능을 보임을 확인



\*positive persona: 대화와 관련이 있는 페르소나

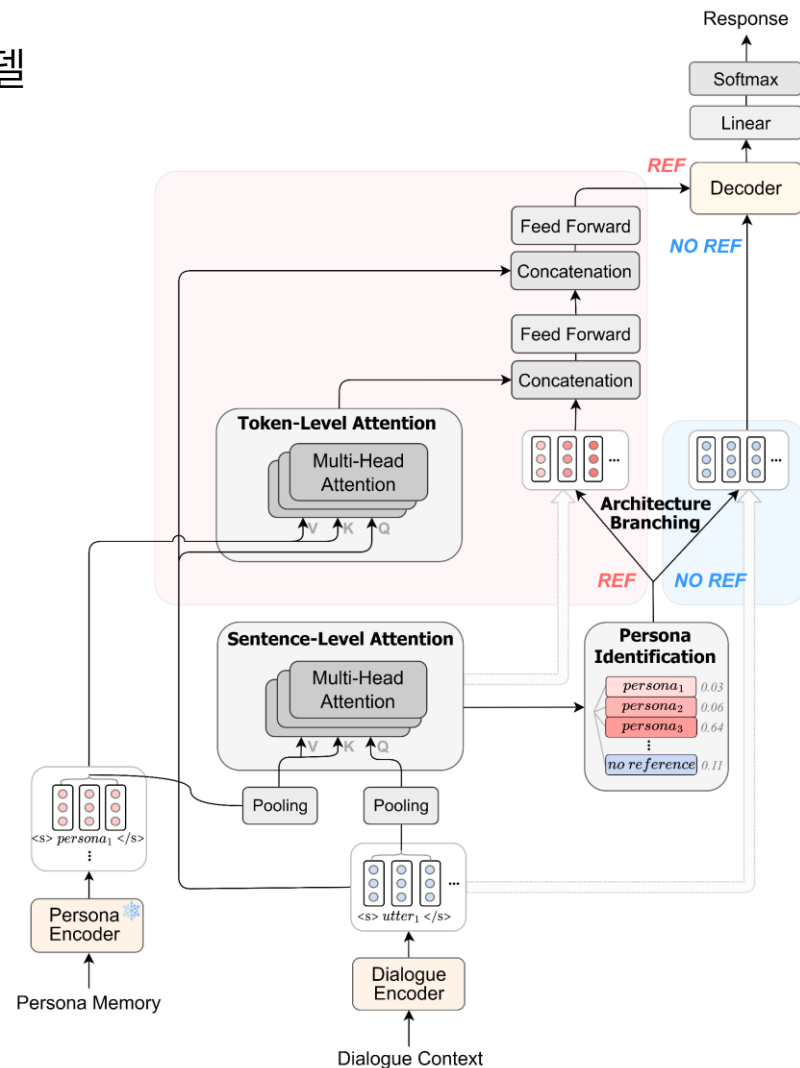
\*negative persona: 대화와 관련이 없는 페르소나



# Research Experiences

페르소나 기반의 장기 대화를 위한 다각적 어텐션을 활용한 생성 모델 (HCLT 2023)

- 대화와 관련 있는 페르소나 정보를 식별하고 이를 반영해 응답을 생성하는 모델
  - 목표: 검색 모델 없이 메모리로부터 관련된 페르소나를 찾을 수 있고, 메모리 참조의 필요성 여부를 판단 가능하게 함
  - HCLT 2023 우수논문상 수상
- Descriptions
  - 대화와 페르소나 간 심층적인 관계 파악을 위해 sentence-level 및 token-level attention 연산 활용
  - Attention weight를 기반으로, 메모리 참조 여부 및 참조할 정보에 대한 classification loss를 추가 학습
  - 검색 모델을 사용하지 않아 추론이 빠르고 효율적
  - 평가 결과, 장기적인 대화 환경에서도 잘 적응하는 것을 확인



# Research Experiences

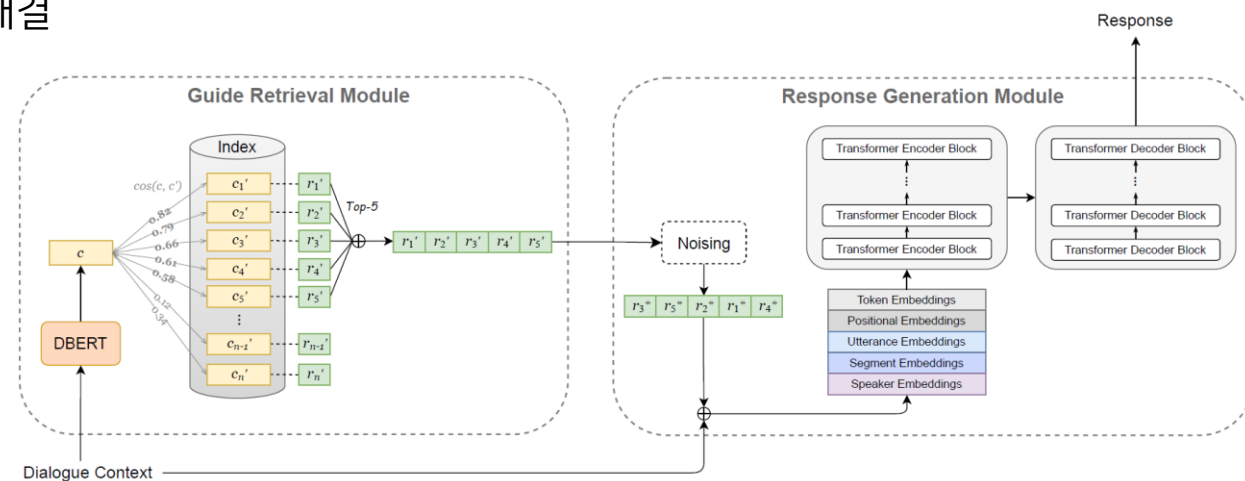
오픈 도메인 대화를 위한 노이즈된 가이드 기반 생성 모델 (HCLT 2022 & Journal of KIISE)

- 생성 모델 또는 검색 모델 한쪽에 의존하는 경향을 완화하고 조화롭게 활용하는 대화 모델

- 목표: 검색된 응답을 활용하지 않고 생성 모델에만 의존하거나, 검색된 응답을 그대로 복사하여 검색 모델에만 의존하는 문제를 해결
- HCLT 2022 우수논문상 수상
- 확장하여 정보과학회논문지에 출판

- Descriptions

- 구축해둔 FAISS Index로부터 입력 컨텍스트와 유사한 컨텍스트 및 응답을 k개 검색
- 검색된 응답들에 노이즈 기법 적용
  - 내용어의 15% masking
  - 언어 모델로 masking된 토큰을 예측해 filling
  - 응답 간 순서를 shuffle
- 생성 모델의 인코더 입력단에 Utterance / Segment / Speaker embeddings 추가



Input	<s>	<user>	좋아	<agent>	같이	가자	...	<guide>	몇	시	?	<guide>	그래	...	</s>
Token Embeddings	$E_{<s>}$	$E_{<user>}$	$E_{\text{좋아}}$	$E_{<agent>}$	$E_{\text{같이}}$	$E_{\text{가자}}$	...	$E_{<guide>}$	$E_{\text{몇}}$	$E_{\text{시}}$	$E_{?}$	$E_{<guide>}$	$E_{\text{그래}}$	...	$E_{</s>}$
Positional Embeddings	$E_0$	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	...	$E_m$	$E_{m+1}$	$E_{m+2}$	$E_{m+3}$	$E_{m+4}$	$E_{m+5}$	...	$E_n$
Utterance Embeddings	$E_{U_1}$	$E_{U_1}$	$E_{U_1}$	$E_{U_2}$	$E_{U_2}$	$E_{U_2}$	...	$E_{U_6}$	$E_{U_6}$	$E_{U_6}$	$E_{U_6}$	$E_{U_6}$	$E_{U_6}$	...	$E_{U_6}$
Segment Embeddings	$E_{CXT}$	$E_{CXT}$	$E_{CXT}$	$E_{CXT}$	$E_{CXT}$	$E_{CXT}$	...	$E_{GUD}$	$E_{GUD}$	$E_{GUD}$	$E_{GUD}$	$E_{GUD}$	$E_{GUD}$	...	$E_{GUD}$
Speaker Embeddings	$E_{USR}$	$E_{USR}$	$E_{USR}$	$E_{AGT}$	$E_{AGT}$	$E_{AGT}$	...	$E_{AGT}$	$E_{AGT}$	$E_{AGT}$	$E_{AGT}$	$E_{AGT}$	$E_{AGT}$	...	$E_{AGT}$

Encoder Input Representations

# Development Experiences

TUNiB — Data & Software Engineer Internship (2021)

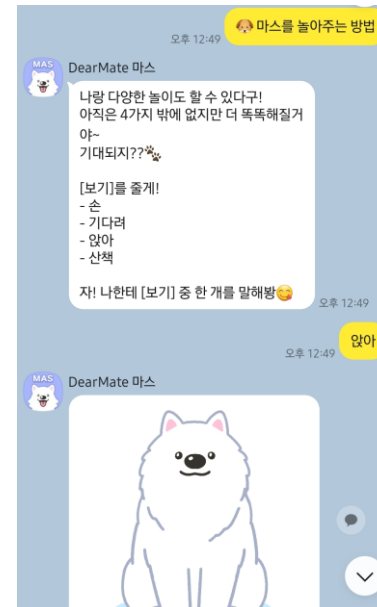
- 페르소나 기반 챗봇 서비스 - 마스, 코코
  - 사용자의 만족도를 향상시키기 위해 챗봇의 페르소나를 더 견고히 하는 역할 수행
- Descriptions
  - 카카오 i 오픈 빌더에서 스킵 서버를 연결해 API를 통한 챗봇 세부 기능 구현
    - 간식 주기, 놀아주기, 사용자 이름 설정, 대화 피드백, 말 가르치기, 출석체크 등
  - 기획팀과 긴밀히 협업하며 요청 받은 기능들을 구현
  - Jenkins를 통해 배포 환경을 자동화하여 베타 서비스를 진행



사용자 이름 설정



간식 주기



놀이 주기



대화 피드백

# Development Experiences

TUNiB — Data & Software Engineer Internship (2021)

---

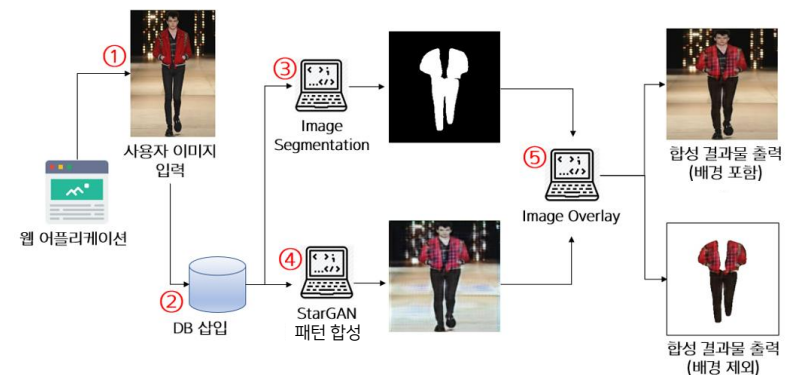
- Web Crawling을 통해 언어 모델 학습용 대규모 데이터셋 구축
- Descriptions
  - 블로그, 커뮤니티, GitHub 등 다양한 웹 사이트에 대한 crawler 제작
  - Multiprocessing을 통해 프로세스 효율화
  - 600GB 가량의 데이터를 수집 및 전처리
  - 수집한 데이터는 Polyglot-Ko, TUNiB-Electra 모델의 학습에 사용됨

# Development Experiences

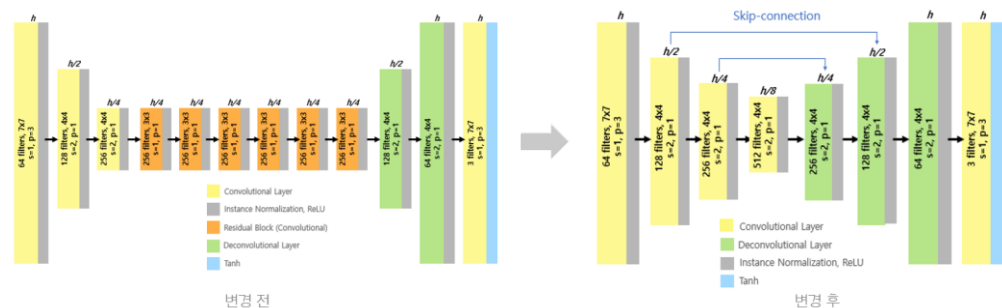
딥러닝 기반 의류 합성 프로그램 개발 (산학연계/SW프로젝트, 2020)

- GAN을 활용한 의류 합성 프로그램 개발
  - 목표: 의류 이미지에 대한 GAN 모델의 성능 저하 문제를 해결하고, 의류 디자인에 대한 접근성을 높임
  - 인공지능 활용 아이디어 공모전 총장상 수상
- Descriptions
  - Mask R-CNN 모델을 통해 의류 영역을 segmentation
  - 추출된 영역에 StarGAN 모델을 통해 의류 패턴이 합성된 이미지 생성
  - StarGAN 아키텍처를 개선하여 파라미터 수 35% 경량화
- 역할
  - 팀장으로서 프로젝트 총괄
  - StarGAN 아키텍처 개선 및 fine-tuning
  - 데이터 웹 크롤링 및 정제

## System Overview



## Improved Architecture



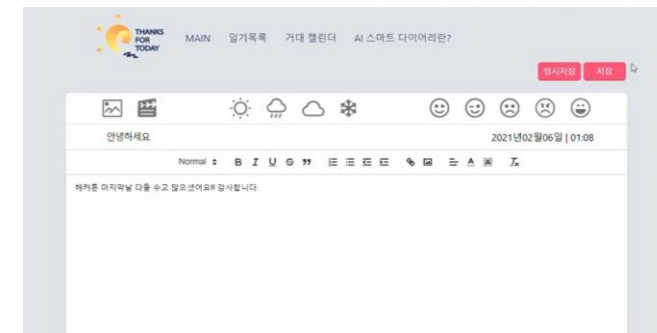
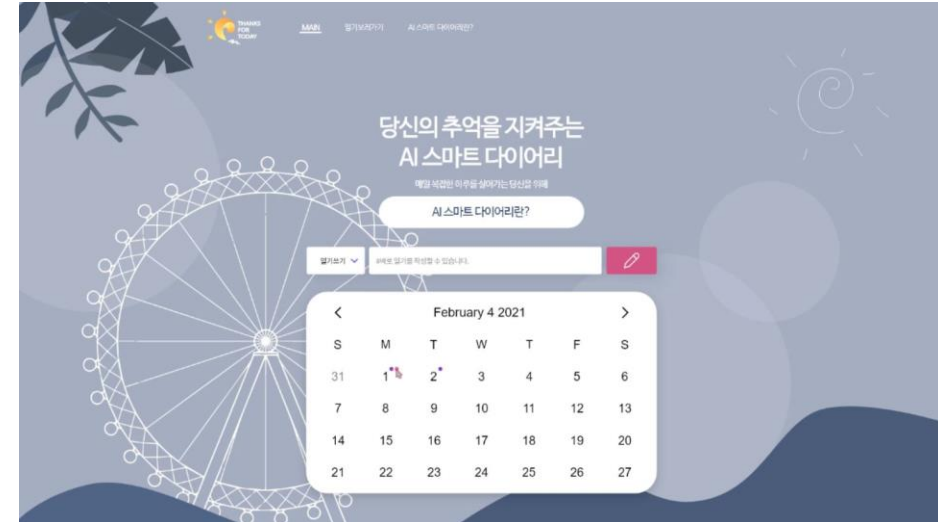
## Results



# Development Experiences

SW중심대학 공동 해커톤 (2021)

- Thanks For Today - 나를 알려주는 스마트한 다이어리
- Descriptions
  - 자신에 대해 더 알아갈 수 있는 all-in-one 다이어리 플랫폼
  - 작성한 일기로부터 중요한 문장과 키워드를 추출해 보여줌
  - 전체 일기로부터 자주 사용된 키워드를 wordcloud로 시각화 함으로써 개인의 관심사를 파악 가능
- 역할
  - NLP 기술 적용
  - Back-end 개발



# 감사합니다.

---

금빛나 (Bitna Keum)