대학원 입학설명회

# 진학과 입사 사이, 대학원에서의 선택

한양대학교 자연어처리 연구실 | 2025.09.23

컴퓨터·소프트웨어학과 박사과정 **윤예진**

# YEJIN YOON

NLP Researcher

*# Conversational AI*
*# Dialog System*

yejin-yoon- yjyoon-ai

HYU 한양대학교 HANYANG UNIVERSITY


NLP Researcher @ Hanyang University NLP Lab.
Seoul, Republic of Korea
@ stillwithyou@hanyang.ac.kr


CS PhD Candidate @ Hanyang University
HYU NLP Lab. Advisor : Prof. Taeuk Kim (2022.09 – , in progress)

NLP Researcher @ AI2XL Labs, KT (2023.07 – 2023.08)
NLP Researcher @ AI Labs, Saltlux (2020.03 – 2022.08)

## Publications

- **TACT:** Beyond Task-Oriented and Chitchat Dialogues: Proactive and Transition-Aware Conversational Agents (EMNLP2025, accepted)
- **BlendX:** Complex Multi-intent Detection with Blended Patterns (LREC-COLING2024)

## Patents

- Method And Apparatus for Generating Multi-Intent Utterance Datasets (2024, Patent Applied)
- Multi-Persona Classification Device Independent of Threshold Settings (2023, Patent Applied)

## Talks

- Talks titled “Practical Research on Overcoming the Limitations of LLM Dialogue Capacity” (SKT)
- Talks titled “Complex MID with Blended Patterns” at GITW2023 (Waseda University in Tokyo)

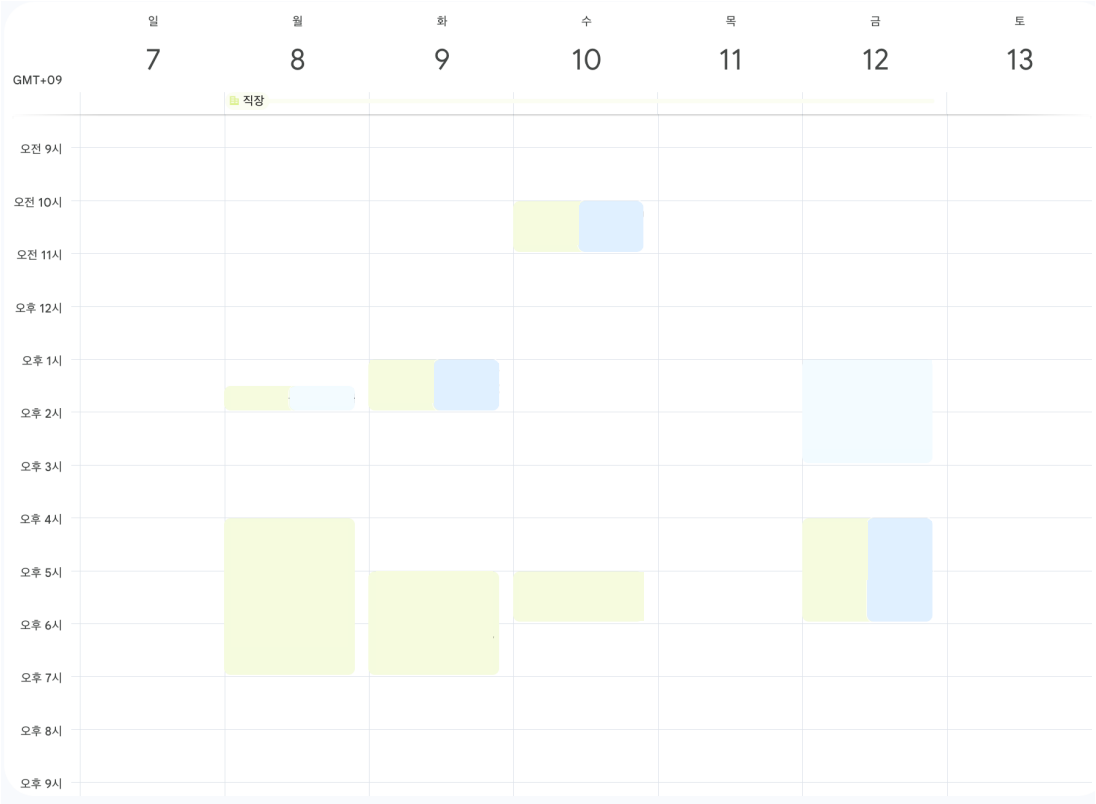
# 연구실에서 일상

- **연구실 정기 일정과 하루 일과**
  - 매일: 논문 읽기, 메일, 개인/프로젝트 연구(실험, 논문 쓰기, 발표 준비, …), course work

  - 주 1회 프로젝트 회의 (같은 연구 주제 팀 대상, 오프라인)
  - 주 1회 주간 연구 회의 (연구실 전체 대상, 오프라인/온라인)
  - 주 1회 연구 세미나 (연구실 전체 대상, 분기마다 1회 발표)

  - 월 1회 현대자동차 산학연구 정기보고
  - 월 1회 신진연구 연구노트 작성

  - 비정기 일정
    - Colloquium
    - 연구실 지원자 인터뷰
    - 연구 제안
    - 연구 성과 발표 (학회, 워크샵, …)



# 대학원 입학 동기

- **KoLegal-BERT: 법률 도메인 텍스트 마이닝을 위한 법률 언어 표상 모델**
  - Authors: 박상민*, **윤예진***, 이재윤, 김재은 (*: equal contribution)
  - 2022 한국컴퓨터종합학술대회 (KCC2022)논문 accepted

KoLegal-BERT: 법률 도메인 텍스트 마이닝을 위한 법률 언어 표상 모델*

박상민+, 윤예진+O, 이재윤, 김재은+
㈜솔트룩스, AI Labs
{smpark, yjyoon, jylee, jekim}@saltlux.com

KoLegal-BERT: Legal Language Representation Model for Legal Domain Text Mining

Sangmin Park+, Yejin Yoon+O, Jaeyun Lee, Jaieun Kim+
AI Labs, Saltlux Inc

**요 약**

BERT(Bidirectional Encoder Representations from Transformers) 모델은 양방향 트랜스포머 기반을 활용하여 대용량 말뭉치를 학습시켜 사전학습한 언어모델이며 다양한 자연어 처리 태스크에서 높은 성능 향상을 달성하였다. 하지만 사전학습을 위해 사용된 말뭉치는 위키피디아, 뉴스 등과 같은 말뭉치이기 때문에 전문적인 어휘나 문맥이 쓰이는 도메인에서는 향상된 성능을 기대하기 어렵다. 특히 최근 법률 서비스와 인공지능이 융합한 리걸테크(Legal Tech)의 중요성이 대두되고 있으나 언어적 한계점에 대한 문제가 지속적으로 제기되고 있다. 본 논문에서는 이와 같이 리걸테크에서 제기되는 법률 도메인의 언어적 한계점을 개선하고 법률 도메인을 잘 이해할 수 있는 법률 도메인에 특화된 KoLegal-BERT 모델을 제안한다. 또한 제안한 모델을 법률 문서 분류 평가를 통해 일반 BERT 모델과의 비교 평가를 수행하고 우수성을 입증한다.

1. 서 론

이며 두 번째 방안은 학습된 일반 BERT 모델에 법률 말뭉치를 추가로 사전학습 시키는 추가 사전학습(Adaptive

# 석사과정 진학 동기

- **퇴사, 한양대학교 AI 응용학과 (KT계약학과) 입학**

  - 승진, 급여인상, 안정적인 직장, …

  - 경력개발 과정에서 학위에 대한 니즈 확대
    - 연구직무에 종사한다면 더 좋은 일자리, 더 나은 대우
    - **전문가가 되고 싶은 욕심**

  - 안정적인 이직 고려
    - 석사과정 등록금 지원 + 장학금
    - 졸업 후 입사 확정형 전형

  - 석사학위 진학 목표
    - 연구 과정을 설계하고 경험하기

**KT-한양대 AI응용학과(채용연계형 계약학과)**

**2022년 하반기 석사과정 모집**

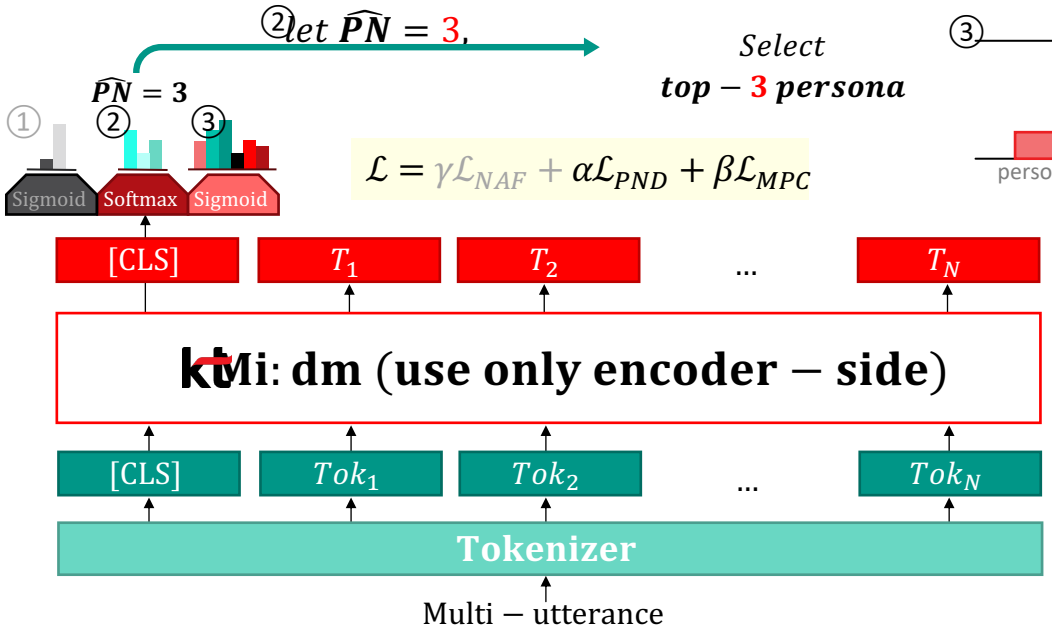
√ 입학금 · 등록금 전액 및 소정의 연구지원비 지급
√ 최신형 GPU 서버 이용권한 및 KT 현장실습/대규모 AI데이터 연구 기회 제공
√ 석사 졸업 후 KT 융합기술원에서 R&D 연구원으로 근무

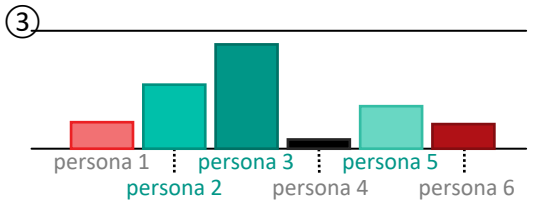
# 석사과정 진학

## • KT 인턴십

- 장기 기억을 위한 시니어 케어 대화의 다중 페르소나 분류
  Multi-Persona Classification of Senior Care Conversation for Long-term Memory

- Multi-label Classification w/ weighted loss
  - Multi-task learning: ① NA Filtering + ② Persona number detection + ③ Multi-persona classification
  - Backbone: KT Mi:dm (kt-ulm-base), using only encoder-side




① NA Filtering
- **Input**: multi-utterance
- **Output**: NA or Not

② **Persona number detection**
- **Input**: multi-utterance
- **Output**: 1 label (# of persona)

③ **Multi-persona classification**
- **Input**: multi-utterance
- **Output**: 1 or more label(s)

backbone PLM: KT Mi:dm
  * Mi:dm-ed : encoder-decoder
  * Mi:dm-e / Mi:dm-e-baseline : encoder-side only

max_sequence_length: 512
  * input: $[utterance_{t-2}] + [\,] + [utterance_{t-1}] + [\,] + [utterance_t]$
  * output: $[utterance_t]$ 에 대한 73-length binary tensor

epoch: 70

| Model \ Metric | Persona Number Detection | | Multi-Persona Classification | | |
|---|---|---|---|---|---|
| | **Accuracy** | **F1-Score** | **Accuracy** (instance) | **Accuracy** (samples) | **F1-Score** (samples) |
| **Mi:dm-ed** | - | - | | 85.19 | |
| **Mi:dm-e-baseline** | - | - | 0.61 | 98.57 | 0.55 |
| **Mi:dm-e** $\alpha=0.5\,\beta=1$ | 91.28 | 91.27 | **84.34** | **99.56** | **84.32** |
| **Mi:dm-e** $\alpha=0.7\,\beta=1$ | 90.62 | 90.61 | 83.84 | 99.54 | 83.82 |

# 석사과정 진학

- **현대자동차 산학연구: 자연어 기반 복합의도 알고리즘 연구**

**1. Accepted by LREC-COLING2024 main conference**

- *"BlendX: Complex Multi-intent Detection with Blended Patterns"*
- Publishing the Constructed Dataset, "BlendX" and "MixX"

**2. Patent Application** (Status: Pending)

- 단일 의도 발화를 다중 의도 발화로 확장하기 위한 규칙 활용 방식과 생성형 인공지능 모델 활용을 결합한 발화 병합기술

*An utterance merging method for extending single-intent utterances to multi-intent utterance*

**3. Attended GITW2023**

HYU 한양대학교 HANYANG UNIVERSITY HYUNDAI

LREC-COLING 2024

**BlendX: Complex Multi-Intent Detection with Blended Patterns**

**Anonymous submission**

**Abstract**

Task-Oriented Dialogue (TOD) systems typically suppose that a user utterance corresponds to a single intent. This assumption may be misaligned with real-world scenarios where users often express multiple intents simultaneously. While there is an emerging interest in Multi-Intent Detection (MID), existing in-domain datasets such as MixATIS and MixSNIPS have limitations in their formulation. To address these issues, we present BlendX, a suite of refined datasets featuring more diverse patterns than their predecessors, elevating both its complexity and difficulty. For dataset construction, we utilize both rule-based heuristics as well as a generative tool–OpenAI's ChatGPT—which is augmented with a similarity-driven strategy for utterance selection. To ensure the quality of the proposed datasets, we also introduce three novel metrics that assess statistical properties of an utterance related to word count, conjunction use, and pronoun usage. Extensive experiments on BlendX reveal that state-of-the-art MID models struggle with the challenges posed by the new datasets, highlighting the need to reexamine the current state of the MID field.

**Keywords:** Multi-Intent Detection, Task-Oriented Dialogue

**1. Introduction**

The successful implementation of Task-Oriented Dialogue (TOD) systems begins with the precise recognition of user intents. By accurately discerning the queries embedded in user inputs and routing them to the relevant components, the systems can adeptly respond, thereby effectively fulfilling user requests. In general, such systems are constructed on the assumption that each user utterance correlates exclusively with a single intent, which often diverges from practical scenarios.

Contrary to the conventional setting, the task of **Multi-Intent Detection (MID)** presents a more nuanced and comprehensive challenge for TOD systems, permitting users to express multiple intentions simultaneously. The problems posed by MID are not only more demanding but also more realistic—for reference, Gangadharaiah and Narayanaswamy (2019) reported that over half of the total instances (52%) from Amazon's in-house TOD dataset contain multiple intents, underscoring the practical significance of the task.

Despite sustained interest in MID, the availability of resources supporting the research direction is, to our surprise, notably limited. Most stud-

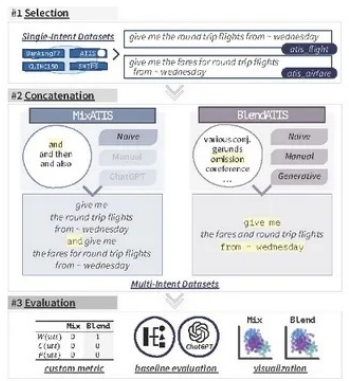
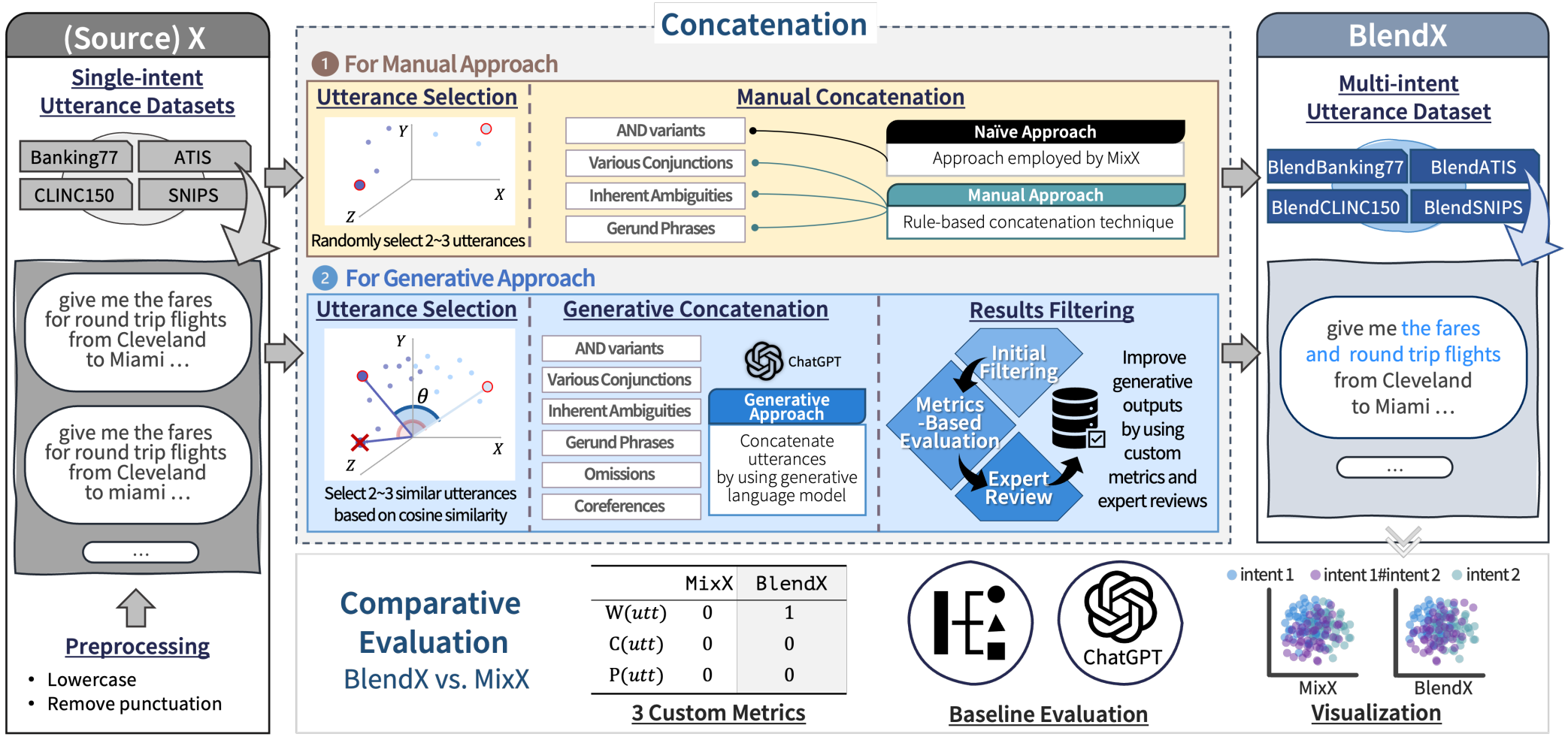



Figure 1: The framework of **BlendX** (comparison with **MixX**). First, select single-intent utterances. Second, combine them by applying mixed methods. The right example in #2 shows a case of non-trivial merging in BlendX, leading to an **omission**. Finally,

27

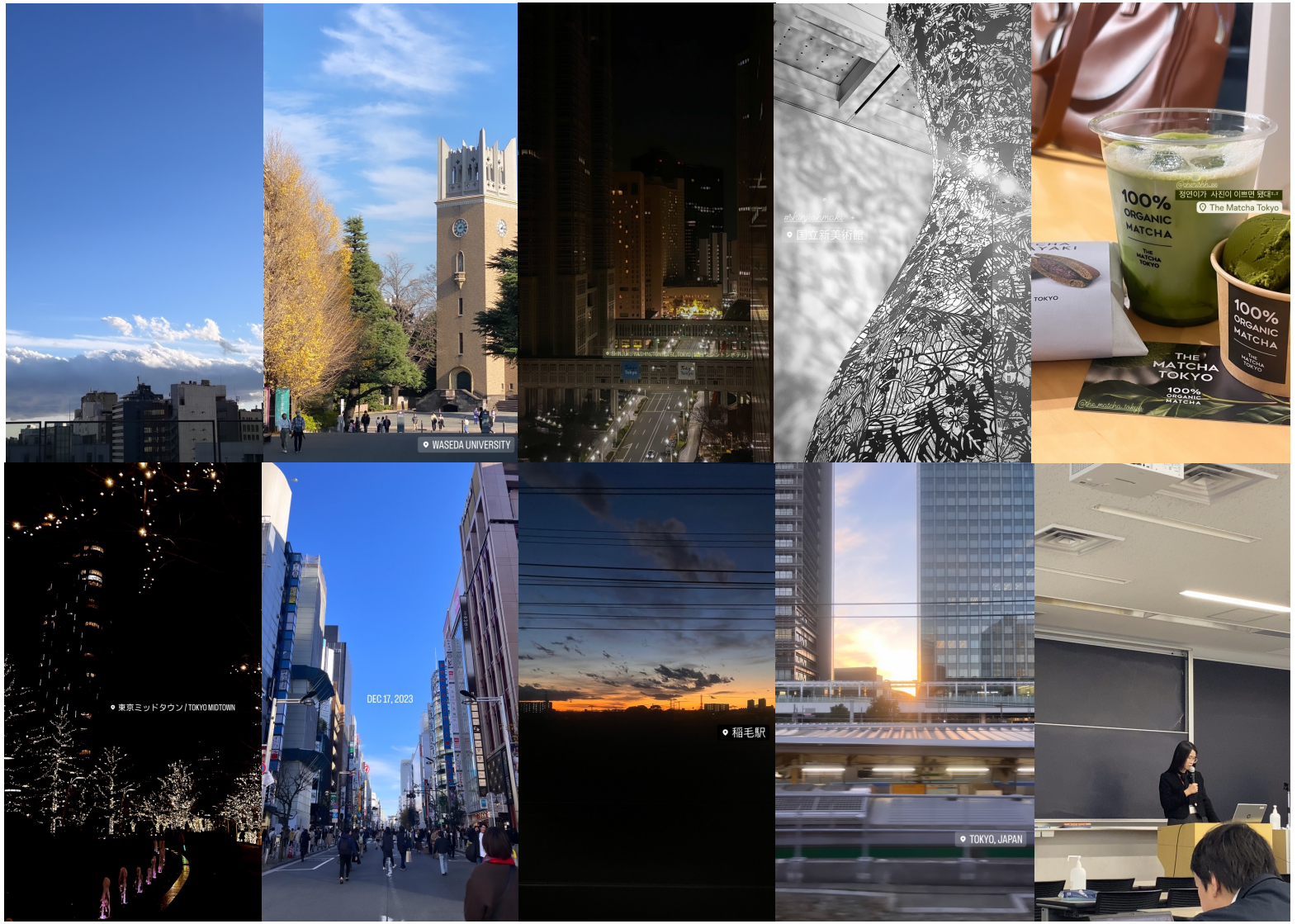
# 석사과정 진학

- **BlendX: Complex Multi-Intent Detection with Blended Patterns (COLING2024)**



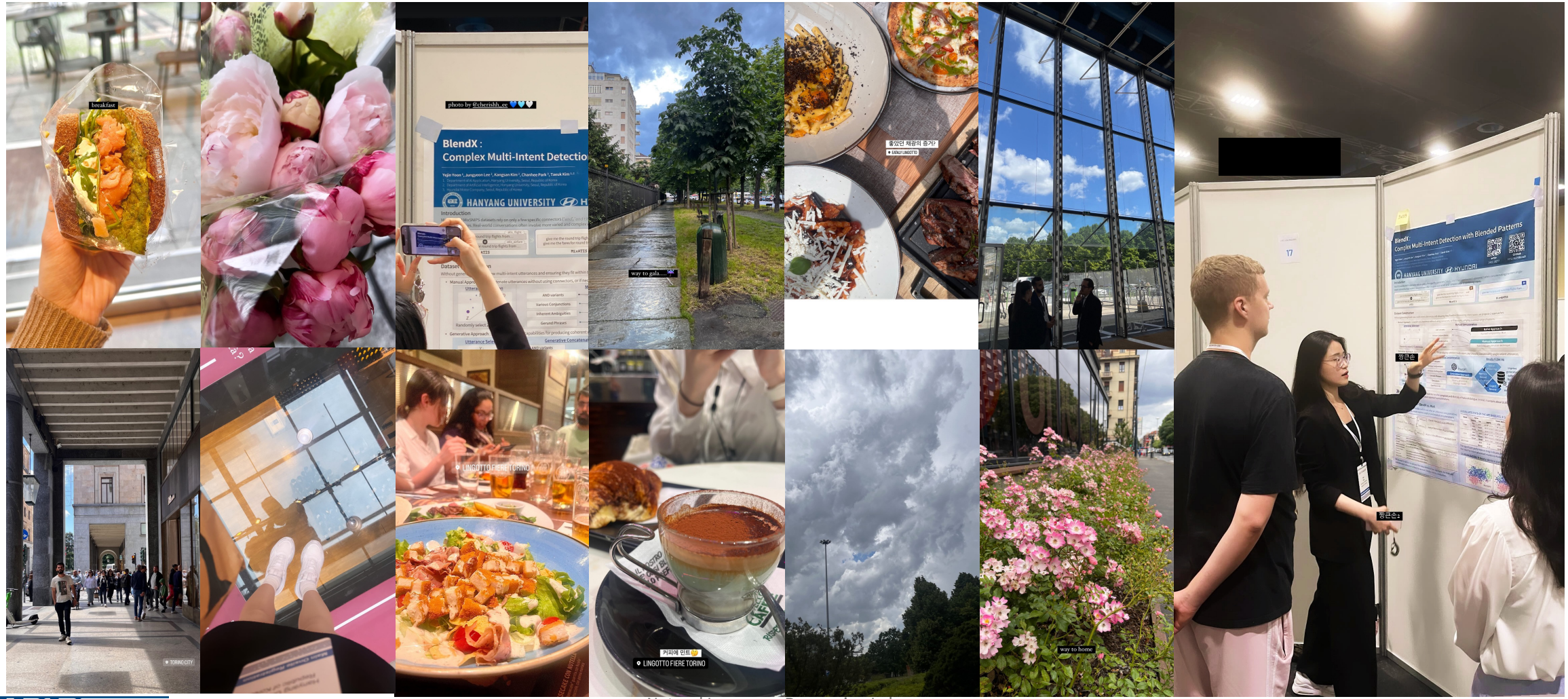
# 석사과정 진학

- **GITW2023 @ Tokyo, Japan**



# 석사과정 진학

- **LREC-COLING2025 @Torino, Italy**



# 석사과정 진학

- **KCC2024 @ Jeju**

한국어 발화의 다중 의도 감지 연구 🔗

[KCC2024] 한국어 발화의 다중 의도 감지 연구.pdf 1073.7KB

- **윤예진***, 김지수*, 임정민, 이정연, 김태욱
- KCC 2024 Accepted || 🏆 **Outstanding Paper Award**
- ▶ **citation**

다중 레이블 분류를 위한 프롬프팅 고도화 🔗

[KCC2024] 다중 레이블 분류를 위한 프롬프팅 고도화.pdf 949.3KB

- 이정연, 신영우, **윤예진**, 김태욱
- KCC 2024 Accepted
- ▶ **citation**

생략과 상호참조를 보강한 다중 의도 데이터셋 🔗

[KCC2024] 생략과 상호참조를 보강한 다중 의도 데이터셋.pdf 877.3KB

- 소성민*, 민지우*, **윤예진**, 이정연, 김태욱
- KCC 2024 Accepted
- ▶ **citation**

# 석사과정 진학

- **Invited Talk; 2024 SKT Market Top AI 과정 강연**

# 석사 졸업, 박사과정 진학 동기

- **KT 입사 포기, 박사과정 진학**

  - KT에 입사한 이후에도 (언젠가는) 박사학위 희망
    - 석사과정 진학 동기 일치: 전문가가 되는 한 가지 확실한 길
    - 교수님 권유 👀

  - **독립적인 연구자로 성장**
    - 연구 == 문제해결

  - 자신감
    - KT 말고 다른 기업 이직에도, …

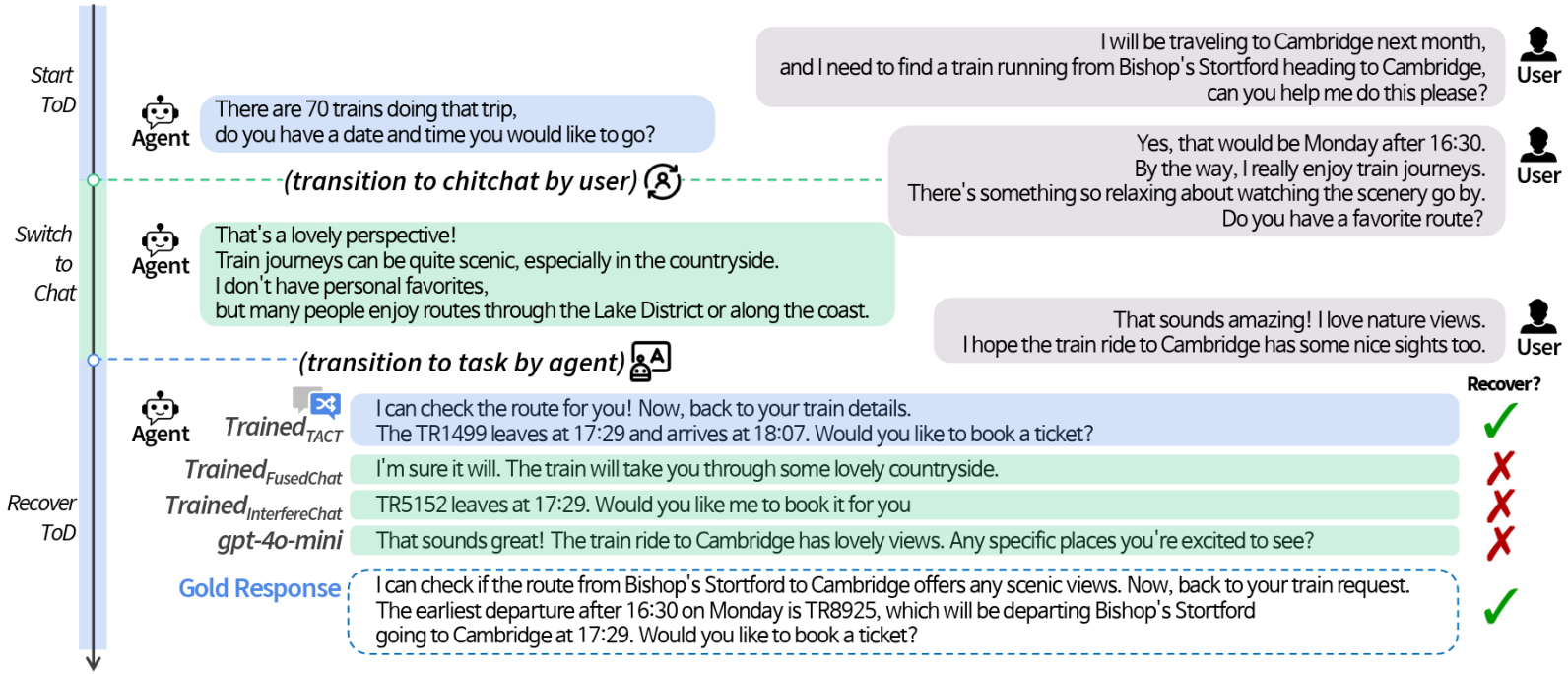
  - ~~연봉협상 포기~~
  - ~~장학금 반환~~
  - ~~이탈리아에서 박사 면접,,~~ 😇

# 박사과정 진학

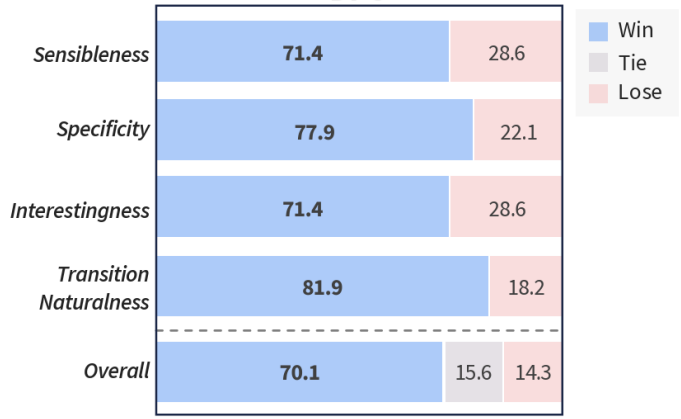
## • 현대자동차 산학연구: 음성인식 고도화를 위한 복합의도 모델 연구

"*Beyond Task-Oriented and Chitchat Dialogues: Proactive and Transition-Aware Conversational Agents*" (EMNLP2025)

- 선호 기반 전환 인지 대화 모델링 시스템 (출원 진행중)
- 전환 인지형 대화 데이터셋 생성 및 AI-인간 융합 검증 시스템 (출원 진행중)




| Dataset | TACT | |
|---|---|---|
| Seed | MultiWOZ2.2 | SLURP |
| # Intents | 11 | 50* |
| # Dialog | 7,199 | 9,936 |
| # Avg. Turn | 15.04 | 16.42 |
| # Avg. Switch | 1.93 | 2.06 |
| # Avg. Recov. | 0.93 | 1.07 |
| # Uniq. Flow | 11 | 12 |
| Flow Types | TCT, CTC, TCTCT, etc. | |



# 연구실에서 일상

## • 대학원 생활의 현실(?)

- 독창성에 대한 환상 깨기
  - 내 아이디어가 세상에 처음 있는 생각일 리가 없다…!
  - 수많은 선행연구 사이에서 작은 틈새 관찰, 그 안에서 의미 있는 질문을 뽑아내는 일 → 관찰과 정리

- 연구는 종합예술
  - 코드 짜기, 수식 세우기는 기본 바탕
  - 실험 설계, 데이터 처리, 모델링, 결과해석, 논문 쓰기, 성과 발표, 모든 게 맞물려 한 편의 연구
  - 비연구 업무, 행정 업무도 다수

- 연구와 응용의 허들
  - (인공지능) 연구와 응용 사이의 간극이 매우 좁다…

- 일정 관리: 연구만 하는 것은 아니었다..!
  - 논문 마감일 12시간 후 산학연구 중간보고
  - 세미나 발표날 외부 발표 일정
  - 논문 마감일에 제안서 회의

- 소통과 네트워킹
  - 논문도 언어, 논문을 읽을 또 다른 연구자, 리뷰어와 소통의 영역
  - ~~업계가 정말 좁다…~~

- 일과 삶의 균형이나 연구와 삶의 균형이나…
  - 출퇴근 개념 모호
  - 장거리 통근..

- 정체성 고민
  - 학생 신분을 이어간다는 것은 또래 친구들과는 전혀 다른 길을 걷는다는 이야기

# 대학원 진학 고민 : 추천할 만한 사람

- **내가 주도하는 깊은 몰입을 경험하고 싶다** ⭐⭐⭐⭐⭐
  - 연구는 내가 문제를 설계하고 질문을 만들고 해결해내는 과정 == 문제해결력
  - 단순히 “기술을 쓰는 것”이 아니라 “왜 이렇게 동작하는지, 더 잘하려면 어떻게 바꿀 수 있을지”를 궁금한 사람!
- **몸과 마음이 건강하고 정신적 여유가 있다** ⭐⭐⭐⭐⭐
  - 빠른 결과보다 같은 문제를 ~~오랫동안~~(?) 붙잡고 실험하고 개선할 수 있는 사람
- **자율성과 불확실성을 견딜 수 있다** ⭐⭐⭐⭐⭐
  - 매뉴얼 없이 스스로 주제를 찾고 길을 만들어갈 수 있는 사람
  - 장기적 목표와 현재를 연결할 수 있는 사람
  - (인공지능)

# 대학원 진학 고민 : 추천하기 어려운 사람

- **취직에서의 스펙이 필요한 사람** ❌❌❌❌❌
  - 단기간에 안정적인 커리어와 금전적 보상은 바로 오지 않는 경우도 잦기 때문에…
- **정해진 과제를 받으면서 일하는 게 편한 사람** 🤔🤔🤔🤔🤔
  - 연구는 무에서 유를 창조하는 일
  - 주어진 틀 안에서의 성취감이 중요한 사람에게는 힘들 수 있다…
- **실패와 불확실성에 영향을 많이 받는 사람** 🥺
  - 처음 세운 연구 가설이 성공하는 경우는 거의 없다…

# 대학원 진학 고민

- **스스로에 대해 많이 고민해보기**
- **연구실에서 하는 모든 일은 자기 성장이랑 연결된다**
  - 내 성장은 내가 시간 쓴 만큼, 신경 쓴 만큼에 (항상은 아니지만) 비례하는 경향이 있다
  - 많이 헤맨 만큼 내 땅
- **자기 성장이 곧 자신감**
  - 취업에 불안하지도 않고 조급하지도 않고
  - 취업 말고 사실 무엇이든!
- **연구실은 작은 사회, 스타트업?**
  - 같이 연구하고 싶은 사람이 (아마도) 같이 일하고 싶은 사람

# Thank You

**Yejin Yoon**

HYU NLP Lab.
Hanyang University, South Korea

stillwithyou@hanyang.ac.kr