SKKU Lab Recommendation Service : FindMyLab

Team J (Last Lap) 김현진 송민석 장민석 조재희

2024-Fall Capstone Design Project

Contents

1. Motivation

“What are some ways to build
the competence of your field of interest?”

Team J (Last Lap) 김현진 송민석 장민석 조재희

1. Motivation

① Undergraduate Research Participation program (URP)

- System that provides undergraduate students with opportunities to participate in research
- System that recognizes credits by performing individual research projects

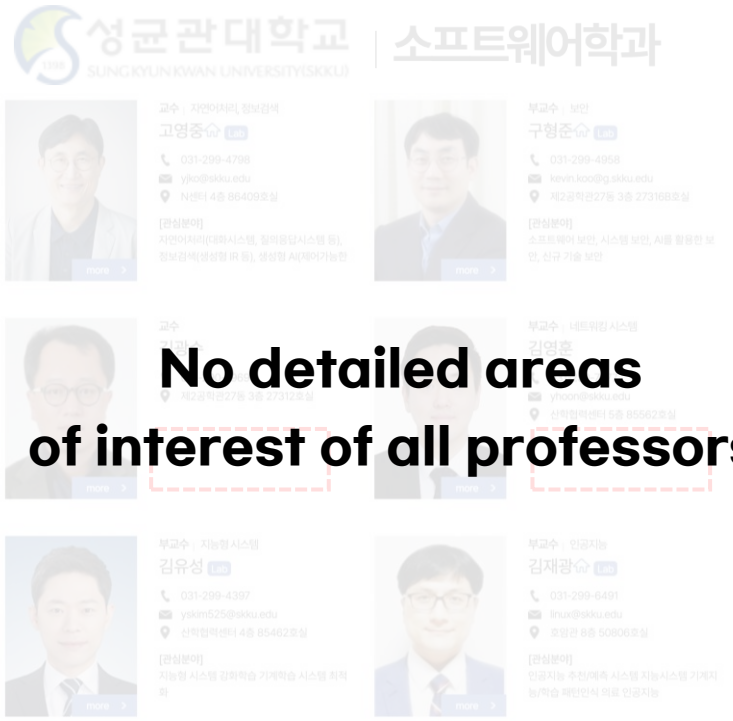
② In-school Research Practice(Co-op)

- Under the guidance of a professor, undergraduate students experience the research field of the supervisor at an understandable level in the short term.

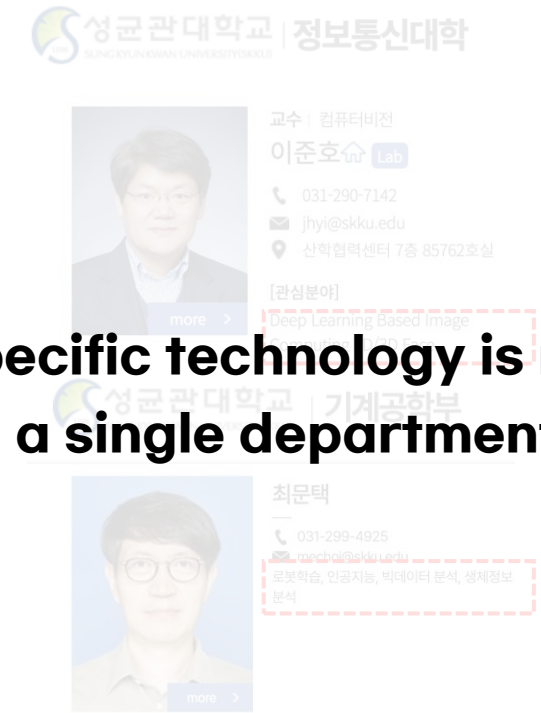
③ Undergraduate Research Student

- As an undergraduate student, he/she participates in the research of the professor in charge and works with graduate students in a laboratory to gain research experience

1. Motivation



No detailed areas of interest of all professors.



Specific technology is not only used in a single department laboratory.

2. Related Work

In korea

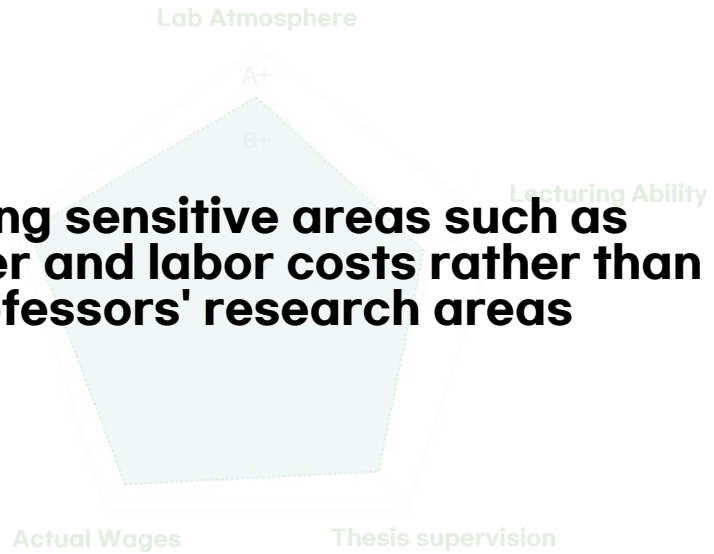
PhD.KIM NET

Abroad

RATE MY PROFESSORS

2. Related Work

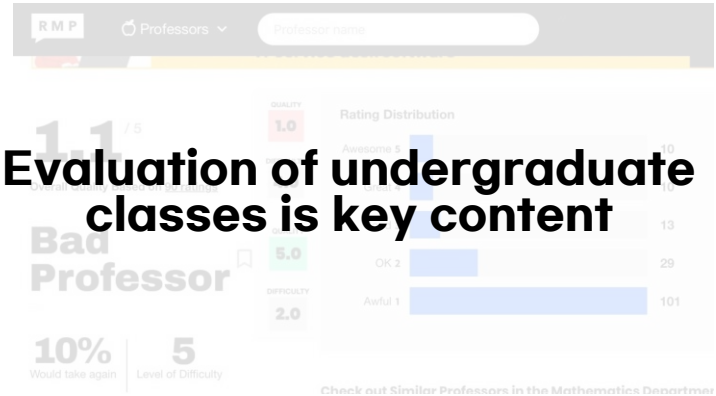
In korea



Sharing sensitive areas such as character and labor costs rather than professors' research areas

PhD.KIM NET

Abroad



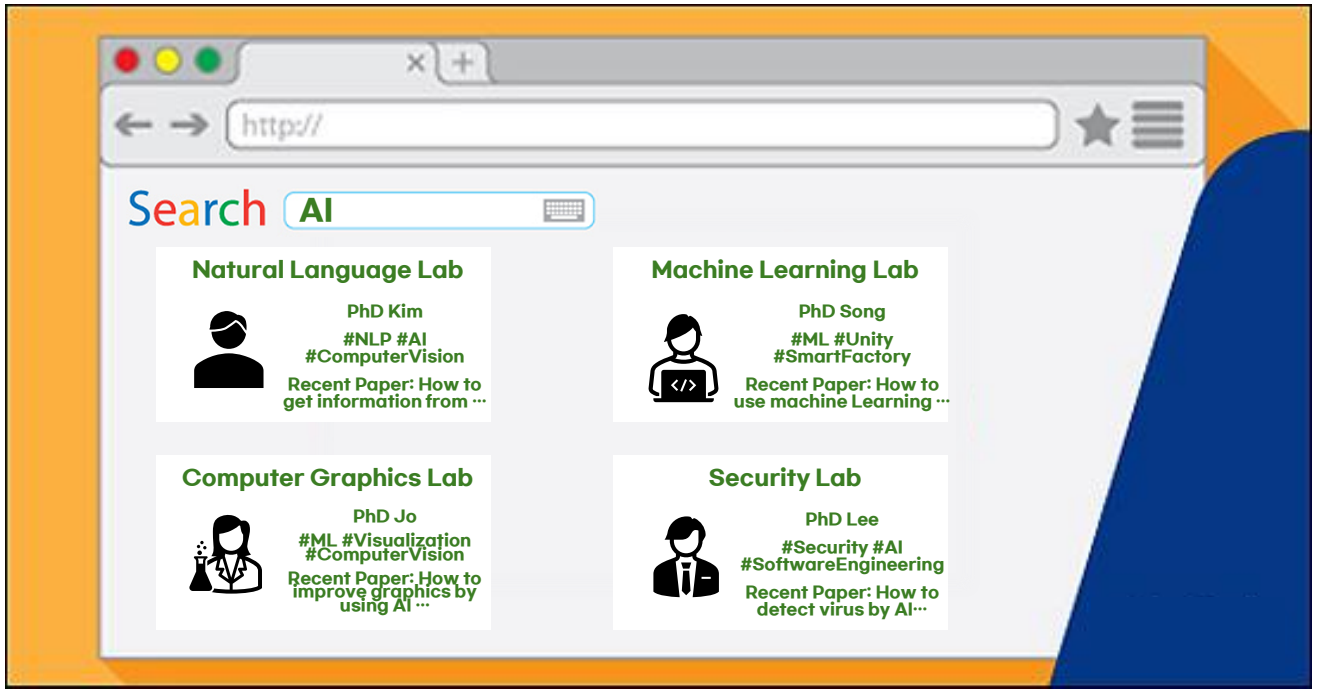
Evaluation of undergraduate classes is key content

RATE MY PROFESSORS

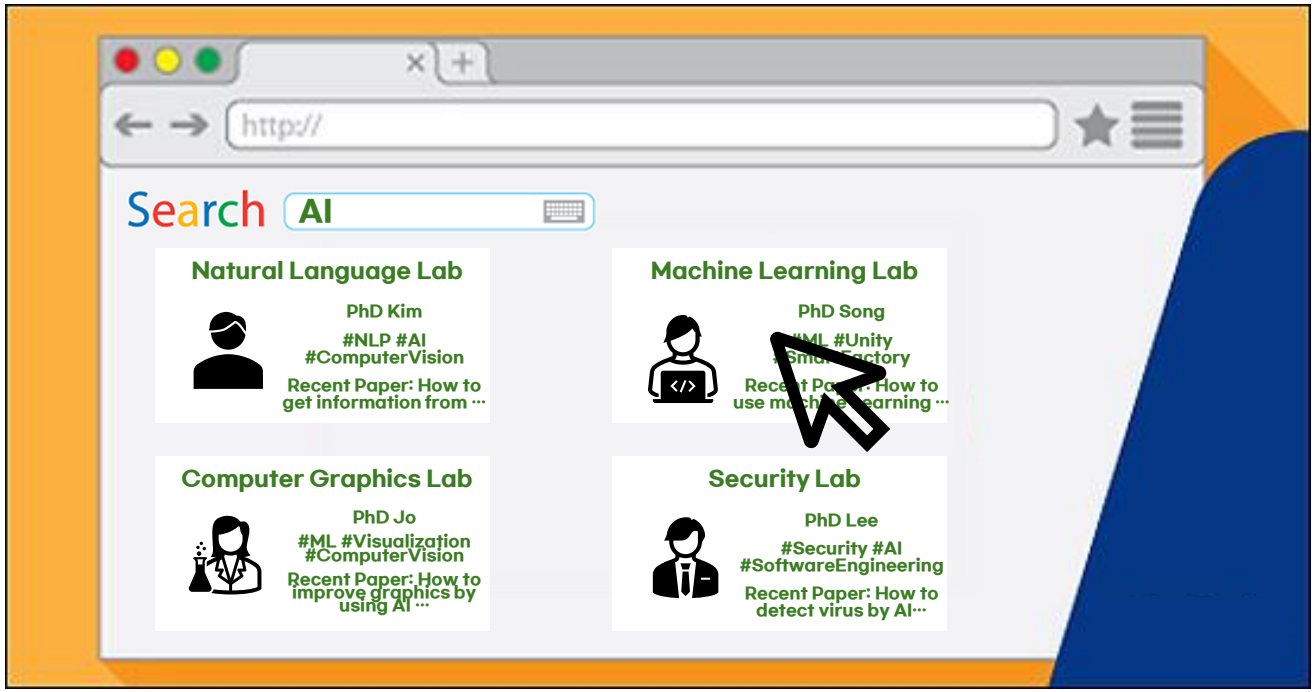
3. Problem Statement

Develop a program that analyzes the professor's published papers to effectively identify the research areas of professor!

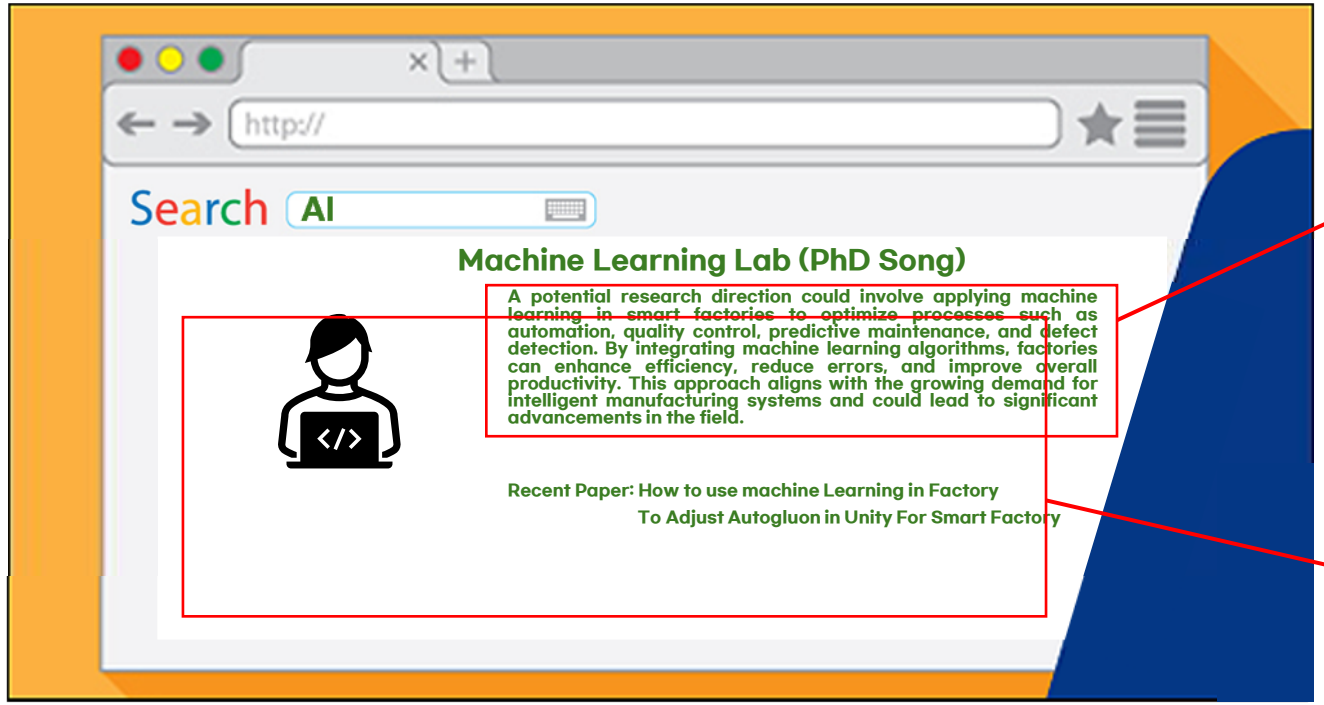
4. Proposed Solution



4. Proposed Solution



4. Proposed Solution



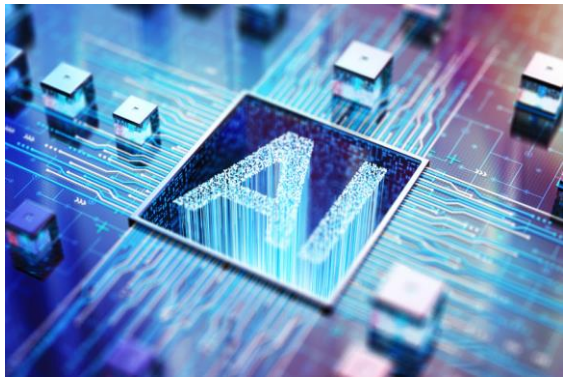
Detailed introduction and guidance on the field the professor is researching

Introduction to the laboratory conducting similar research with the professor

5. Technical Background

Prompt Engineering

Prompt engineering is the process of guiding generative AI solutions to produce the desired output. In prompt engineering, the most appropriate format, syntax, words, and symbols are selected to guide the AI to interact with users more meaningfully.



10.1109/TASLP.2023.3302238
Provide the keywords for the key technologies of the paper related to this DOI.

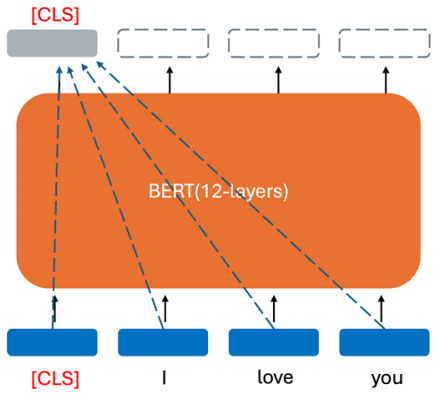
The paper titled "Two-Step Masked Language Model for Domain-Adapting Multi-Modal Task-Oriented Dialogue Systems" focuses on several key technologies, including:

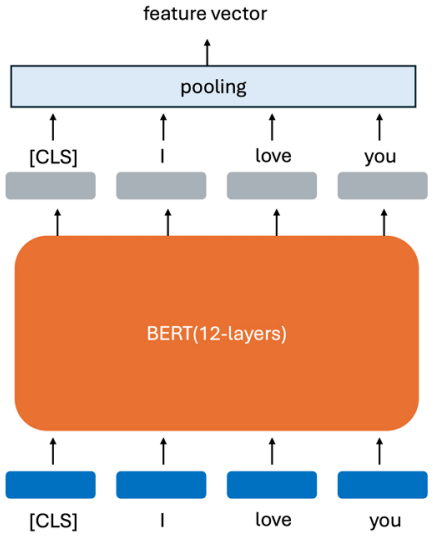
1. **Masked Language Models (MLM)** - a technique used to train models by predicting missing parts of input data.
2. **Domain Adaptation** - methods to tailor models for specific domains, enhancing their performance in targeted applications.
3. **Multi-Modal Systems** - systems that process and integrate different types of data, such as text and images.
4. **Task-Oriented Dialogue Systems** - dialogue systems specifically designed to accomplish user-defined tasks.

5. Technical Background

BERT

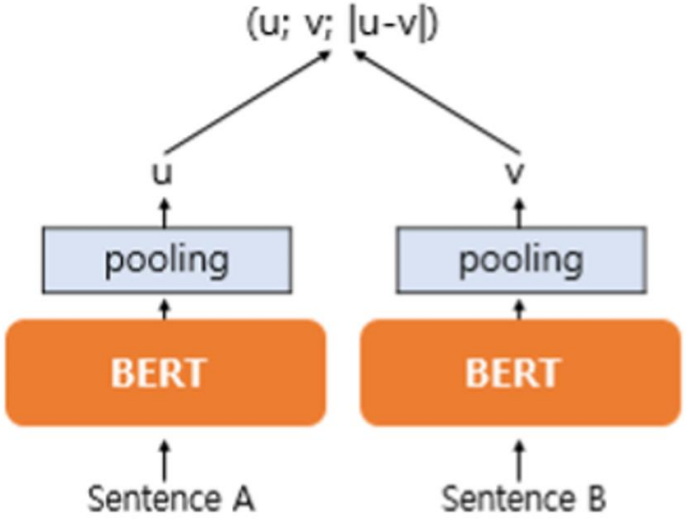
BERT is a bidirectional model that uses multiple layers of the transformer encoder, allowing attention to be applied to both past and future tokens.



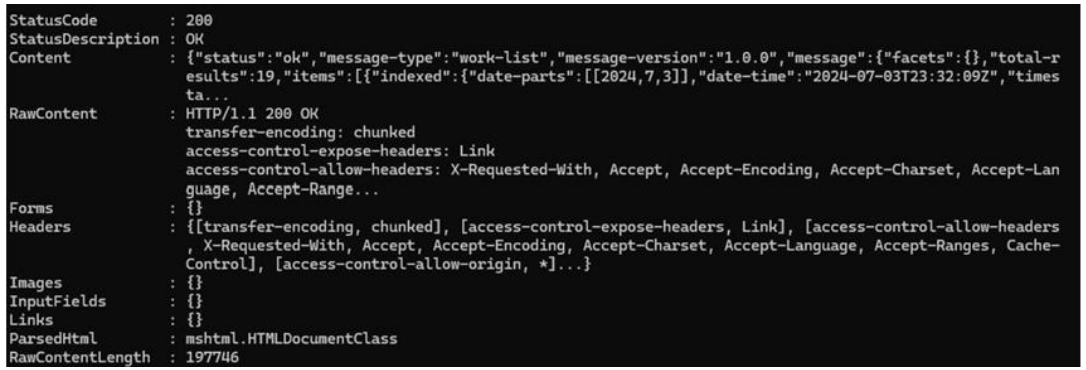


SBERT

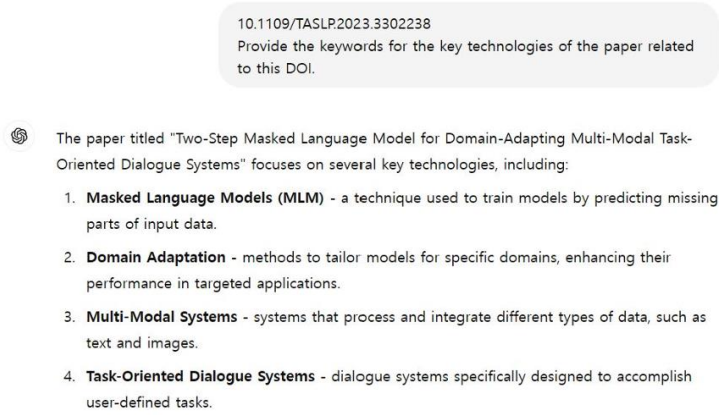
SBERT enhances BERT by adding a pooling layer to use embeddings and processes each sentence separately, unlike BERT, which uses a single Transformer for both.



5. Technical Background



```
StatusCode        : 200
StatusDescription : OK
Content           : {"status":"ok","message-type":"work-list","message-version":"1.0.0","message":{"facets":{},"total-r
                    esults":19,"items":[{"indexed":{"date-parts":[[2024,7,3]],"date-time":"2024-07-03T23:32:09Z","times
                    ta...
RawContent        : HTTP/1.1 200 OK
                    transfer-encoding: chunked
                    access-control-expose-headers: Link
                    access-control-allow-headers: X-Requested-With, Accept, Accept-Encoding, Accept-Charset, Accept-Lan
                    guage, Accept-Range...
Forms             : {}
Headers           : {[transfer-encoding, chunked], [access-control-expose-headers, Link], [access-control-allow-headers
                    , X-Requested-With, Accept, Accept-Encoding, Accept-Charset, Accept-Language, Accept-Ranges, Cache-
                    Control], [access-control-allow-origin, *]...}
Images            : {}
InputFields       : {}
Links             : {}
ParsedHtml        : mshtml.HTMLDocumentClass
RawContentLength  : 197746
```



10.1109/TASLP.2023.3302238
Provide the keywords for the key technologies of the paper related to this DOI.

The paper titled "Two-Step Masked Language Model for Domain-Adapting Multi-Modal Task-Oriented Dialogue Systems" focuses on several key technologies, including:

1. **Masked Language Models (MLM)** - a technique used to train models by predicting missing parts of input data.
2. **Domain Adaptation** - methods to tailor models for specific domains, enhancing their performance in targeted applications.
3. **Multi-Modal Systems** - systems that process and integrate different types of data, such as text and images.
4. **Task-Oriented Dialogue Systems** - dialogue systems specifically designed to accomplish user-defined tasks.

The Message returned by Crossref allows us to obtain the DOI through the work-list in the content section

Core keywords of technologies are extracted through DOI using GPT prompting

5. Technical Background

Column Name	Data Type	Description
professor_id	INT	prefessor ID
name	TEXT	prefessor Name
department	TEXT	department
lab_name	TEXT	labs
research_area	TEXT[]	interst of research (key workds extracted from paper abstract)
research_embedding	VECTOR	comprehensive embedding vector of the professor's research

Column Name	Data Type	Description
paper_id	INT	paper ID
professor_id	INT	prefessor ID (Foreign Key)
abstract	TEXT	paper abstract
title	TEXT	paper title
keywords	TEXT[]	key word
embedding	VECTOR	embedding vectors from paper abstract

Shape of the dataset after preprocessing

$$AP@K = \frac{1}{m}\sum_{i=1}^{K} Precision@i \cdot rel(i)$$

$$MAP@K = \frac{1}{|U|}\sum_{u=1}^{|U|}(AP@K)_u$$

Mean Average Precision @ K (mAP) as the evaluation metric, which takes the order of recommendations into account.

6. Role & Project Plan

Frontend
Data Collection

김현진

장민석

Backend,
Data Preprocessing

송민석

AI Model Design, Data Preprocessing

조재희

6. Role & Project Plan

	weekly plan					
Week	2 3 4	5	6 7 8	9	10 11 12 13	14 15
Overall	Project Proposal	Proposal LaTeX		Midterm Presentation		Final Presentation
AI			data collection data preprocessing		Model implementation	Beta Test
Web/App			UI/UX design Database setting		frontend and backend work	Debugging

Q & A