Capstone Design Project Report

Project Title : 이미지 생성형 AI 모델을 위한 인테리어 프롬프트 생성 LLM 모델 개발 및 레퍼런스 이미지 기반 프롬프트 추출을 위한 VLM 모델 서빙 웹 기반 플랫폼 구현

Lecture Class	Capstone Design 1-02 (2024 Fall)			
Term	2024.09.02. ~ 2025.06.14			
Supervisor	Class Advisor		김동우 교수님	
	Industry Mentor	Industry Mentor	㈜마이공사	
			윤홍선	
Team Member Students	Student ID No.		Name	E-mail
	2019112130		조양진	choyj1222@gmail.com
	2021111947		하유경	eeheueklf@dgu.ac.kr
	2018112525		오진호	rldmlgmfma@naver.com
	2021113359		김수빈	yeh4512@gmail.com

<Summary in English>

<table>
<tr><td>Project Goal and Objectives</td><td colspan="3">This research aims to develop an AI-based system for interior design, focusing on creating algorithms that process text and image data to generate optimized prompts. The system will provide users with prompts based on their input, offering efficient solutions during the initial stages of interior design planning. Specifically, the research will fine-tune LLM, R-CNN, and VLM models to generate the most suitable prompts for different data types. This system will help users visualize design ideas more easily and receive creative prompts in real time, enhancing efficiency in the interior design market and contributing to the related industry. The goal is to publish software, which will be released as open-source to promote innovation in the interior design field through technology sharing with the industry.</td></tr>
<tr><td>Project Contents</td><td colspan="3">This research consists of four sub tasks as belows.
【1st task】 The development of algorithms to generate Interior Prompts(LLM)
(1) To develop GPT-4 fine-tuning and optimization measures to generate interior design prompts based on user-populated text.
(2) To establish quality evaluation criteria for prompts generated from a variety of text data.
【2nd task】 The development of R-CNN-based Object Detection to analysis Interrior-Related Informations
(1) To develop an algorithm based on Mask R-CNN and Detectron2 models for detecting furniture objects in interior images, and analyzing the colors and textures of these objects.
(2) To construct and optimize a dataset for training the model, enabling precise object detection and text generation.
(3) To develop data preprocessing and model optimization strategies for accurately extracting style and color information from image analysis.
【3rd task】 The development of algorithm to generate an interior image as a prompt
(1) To develop a CLIP model-based algorithm for analyzing uploaded user images to extract interior styles and convert them into text prompts.
(2) To propose optimization methodologies for model training using a diverse image dataset to enable real-time prompt generation.
【3th task】 The development of a web-based platform
(1) To build a web-based platform that integrates and provides access to the algorithms developed in tasks 1 through 3.</td></tr>
<tr><td>Anticipated Results and Impacts on Industry</td><td colspan="3">This research aims to develop AI prompt generation technology optimized for interior design image creation. It will fine-tune LLM, R-CNN, and VLM models to enhance prompt generation and object detection based on user-provided text and images. The goal is to automate the interior design process and improve communication between small businesses and design companies. Additionally, the optimized models will be released as open-source, advancing AI in the interior design field.</td></tr>
<tr><td rowspan="3">Keywords</td><td>LLM(Large Language Model)</td><td>VLM(Vision-Language Model)</td><td>R-CNN(Region Based Convolutional Neural Networks)</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

<국문 요약>

<table>
<tr><td>과제의
목적과
목표</td><td colspan="3">이 연구에서는 인테리어 디자인을 위한 AI 기반 시스템 개발을 목표로 하여, 텍스트 및 이미지 데이터를 처리하고 최적화된 프롬프트를 생성하는 알고리즘을 개발한다. 본 연구는 사용자가 입력한 텍스트나 이미지를 통해 프롬프트를 제공하여, 인테리어 디자인의 초기 구상 단계에서 효율적인 해결책을 제시하는 데 중점을 둔다. 특히, LLM, R-CNN, VLM을 파인튜닝하여 데이터 유형에 맞는 최적의 프롬프트를 생성하도록 한다. 이러한 시스템은 사용자가 보다 쉽게 디자인 아이디어를 시각화하고, 실시간으로 창의적인 프롬프트를 얻을 수 있게 하여, 인테리어 디자인 시장의 효율성을 높이고, 관련 산업에 실질적인 기여를 할 것으로 기대된다. 본 연구의 목표는 소프트웨어 1건을 발표하고, 이를 오픈소스로 공개하여 업계와의 기술 공유를 통해 인테리어 분야의 혁신을 촉진하는 것이다.</td></tr>
<tr><td>과제의
내용</td><td colspan="3">이 연구과제는 다음과 같이 네 개의 세부과제로 구성된다.
【1세부과제】 LLM 기반 텍스트 기반 인테리어 프롬프트 생성 알고리즘 개발
(1) 사용자가 입력한 텍스트를 기반으로 인테리어 디자인 프롬프트를 생성하기 위한 GPT-4 기반 LLM 모델의 파인튜닝 및 최적화 방안 개발
(2) 다양한 텍스트 데이터를 활용한 모델 학습 및 생성된 프롬프트의 품질 평가 기준 수립
【2세부과제】 R-CNN 기반 객체 탐지 및 인테리어 관련 정보 분석 알고리즘 개발
(1) Mask R-CNN 및 Detectron2 모델을 기반으로 인테리어 이미지에서 가구 객체를 탐지하고, 객체의 색감과 질감을 분석하는 알고리즘 개발
(2) 이미지 데이터를 기반으로 정확한 객체 탐지와 텍스트 생성을 위한 학습 데이터셋 구축 및 최적화
(3) 이미지 분석에서 발생하는 스타일 및 색상 정보를 정밀하게 추출하기 위한 데이터 전처리 및 모델 최적화 방안 개발
【3세부과제】 VLM 기반 이미지 기반 인테리어 프롬프트 생성 알고리즘 개발
(1) 사용자가 업로드한 이미지를 분석하여 인테리어 스타일을 추출하고, 텍스트 프롬프트로 변환하는 CLIP 모델 기반 알고리즘 개발
(2) 다양한 스타일의 이미지 데이터셋을 활용하여 모델 학습 및 실시간 프롬프트 생성을 위한 최적화 방법론 제시
【4세부과제】 웹 기반 플랫폼 개발
(1) 1~3 세부과제를 이용할 수 있는 웹 기반 플랫폼 구축</td></tr>
<tr><td>예상되는
결과와
파급효과</td><td colspan="3">이 연구에서는 인테리어 디자인 이미지 생성에 최적화된 AI 프롬프트 생성 기술을 개발하고, 프롬프트 생성에 최적화된 LLM, 이미지 분석에 R-CNN, VLM 모델을 파인튜닝하여 특화시키고, 이를 기반으로 사용자가 제공한 텍스트 및 이미지를 분석하여 인테리어 디자인 프롬프트 생성 및 가구 객체 탐지 기술을 개선할 것이다. 이를 통해 개인의 인테리어 디자인 프로세스를 자동화하고, 소상공인이 보다 효율적으로 인테리어 업체와 소통할 수 있도록 돕는 시스템을 구축할 예정이다. 또한, 프롬프트 생성에 최적화된 LLM 및 VLM 모델을 오픈소스로 공개하고, 이를 통해 인테리어 디자인 분야의 AI 기술 발전에 기여할 것이다. 연구 결과는 인테리어 산업뿐만 아니라, 관련 시장의 확장에도 긍정적인 영향을 미칠 것이다. 또한, 본 연구에서 개발한 기술은 관련 기업 및 개인들에게 기술 이전을 통해, 국내 시장을 중심으로 인테리어 디자인 분야에서 실질적인 변화를 가져올 것으로 기대된다.</td></tr>
<tr><td rowspan="3">중요단어</td><td>LLM(Large Language Model)</td><td>VLM(Vision-Language Model)</td><td>R-CNN(Region Based Convolutional Neural Networks)</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td></td><td></td><td></td></tr>
</table>

1. Project Goals and Objectives (과제의 목표와 목적)

The detailed objective of the project has been established under the theme of "Demonstrating Generative AI-Based Commercial Space Image Generation and Editing Functions for Improving the Business Environment of Small Enterprises."

1.1 Project Goal (과제의 목적)

One way to improve the business environment for small enterprises is by transforming the interior environment of their commercial spaces. Typically, this involves hiring an interior design expert to receive an estimate and design proposal, which is then reviewed before proceeding with construction.

However, if the design proposal does not meet expectations, additional revisions may be requested, often incurring extra costs and delays. This can be a significant burden for small business owners with limited time and financial resources.

To address this issue, providing a blueprint for the design proposal to the interior designer can be a practical solution. By basing the design proposal on this blueprint, both parties can save time and reduce unnecessary costs.

Therefore, the purpose of this project is to generate and deliver commercial space images using generative AI as a means to save time and expenses for small enterprises.

1.2 Project Objectives (과제의 목표)

한컷스튜디오(operated by My04) generates commercial images using diffusion-based generative AI. Users can upload space photos and input design prompts to create images reflecting their preferences, enabling small business owners to obtain design blueprints before consulting professionals.

However, diffusion models struggle with understanding long or detailed prompts, often leading to missing content. Additionally, users with limited design sense may find it difficult to articulate their ideas. To address this, a feature that extracts furniture and interior styles from reference images and converts them into prompts is proposed.

Our goal is to implement these improvements through two key strategies to enhance usability and accuracy.

1.2.1 Prompt Engineering through LLM

To optimize prompts for the industrial AI functionality, pre-processing will be conducted using an LLM that satisfies “semantic guidance” requirements. This involves refining user

prompts to highlight words closely related to interior design.

- LLM Fine-Tuning : The prompts, which vary in format depending on the user, will be standardized. At Onecut Studio, prompts are expected to be translated into English and written as keyword-based inputs. Considering the interior design domain, the process will emphasize interior-relevant terms while masking or minimizing unrelated words to reduce noise. Additionally, a fine-tuned model will be prepared to pre-process prompts into a format that aligns with the semantic guidance requirements of the diffusion model (as referenced in related research).
- Calling the Fine-Tuned LLM Model : The fine-tuned LLM will be employed to pre-process user prompts, ensuring the generated input aligns with the diffusion model's capabilities for enhanced understanding and output accuracy.

1.2.2. Prompt Engineering Using VLM

Relevant terms related to interior design will be extracted from reference images using a Vision Language Model (VLM).

- Furniture Detection : Furniture in the reference images will be identified and extracted as tokens, which will then be incorporated into the prompt. This ensures that the identified furniture can be reflected in the space images uploaded to Onecut Studio.
- Style Extraction : By applying VLM, interior tokens embedded in the reference images will be extracted and provided. This enables even users with minimal knowledge of design to recognize key terms and include them in their prompts effectively.

2. Background of the Project (과제의 필요성)

2.1 Motivations (필요성)

- Technical Necessity

Generating interior design images with generative AI requires prompts, but individuals lacking design sensibility face challenges in expressing their preferences. A VLM (Vision Language Model) can extract design elements from reference images, simplifying prompt creation for beginners.

Diffusion models also struggle with context in longer prompts, often missing key content. Incorporating LLM fine-tuning with semantic guidance can improve text-to-image alignment by emphasizing interior-specific terms and ensuring coherent outputs.

- Economic Necessity

Small business owners often face conflicts and extra costs due to unsatisfactory design proposals caused by miscommunication. Using LLM and VLM-based prompt engineering, contractors can receive clear, AI-generated design blueprints, reducing misunderstandings, time, and costs.

For DIY enthusiasts, this service lowers barriers to creating design plans without professional help, potentially stimulating the interior market by encouraging purchases of furniture and decor.

- Social Necessity

Home decoration as a hobby offers comfort and accomplishment. This service simplifies design blueprint creation, promoting the trend of home decoration and enhancing societal well-being by providing a creative and accessible outlet.

2.2 Significances (중요성)

- Importance from a Market Perspective

To grow the interior design market, contractors must deliver proposals aligned with customer preferences. Generative AI enables customers to create design blueprints that clearly communicate their needs, reducing disputes and improving satisfaction. This creates additional demand and expands the market by simplifying the design process for all users.

- Importance from a Technical Perspective

The quality of AI-generated images heavily depends on prompt quality. By developing and sharing diffusion-based generative AI prompts optimized for interior design, non-major companies can enhance their service quality and competitiveness.

- Importance from a Social Perspective

Home decoration, often limited by design expertise, becomes more accessible with an interior prompt service. This technology allows anyone to create DIY blueprints, fostering a stronger connection to their living spaces and promoting emotional well-being.

3. Related Works (관련연구)

3.1 Related Techniques

3.1.1. LLM fine-tuning

- Advantages of Preceding Technology : Introduces semantic guidance, improving diffusion models' understanding of prompts by repeating key terms or masking irrelevant ones to enhance text-to-image alignment.

- Limitations of Preceding Technology : Requires joint training of the diffusion model and semantic guidance, and relies on natural language prompts, whereas 한컷스튜디오 prefers tokenized inputs.

- Functions of the Developed Technology : Standardizes prompts by splitting them into tokenized phrases, optimized for Onecut Studio's system. A fine-tuned LLM model emphasizes interior design-related tokens while excluding unrelated content, ensuring compatibility with diffusion models.

- Efficiency of the Developed Technology : Focuses solely on training the LLM for prompt optimization, making it cost-effective and practical for users without access to complex AI training.

- Convenience of the Developed Technology : Automatically applies semantic guidance to raw prompts, simplifying the process by emphasizing interior design-related elements, allowing even inexperienced users to create optimized inputs tailored specifically for the interior design domain effortlessly.

3.1.2 VLM

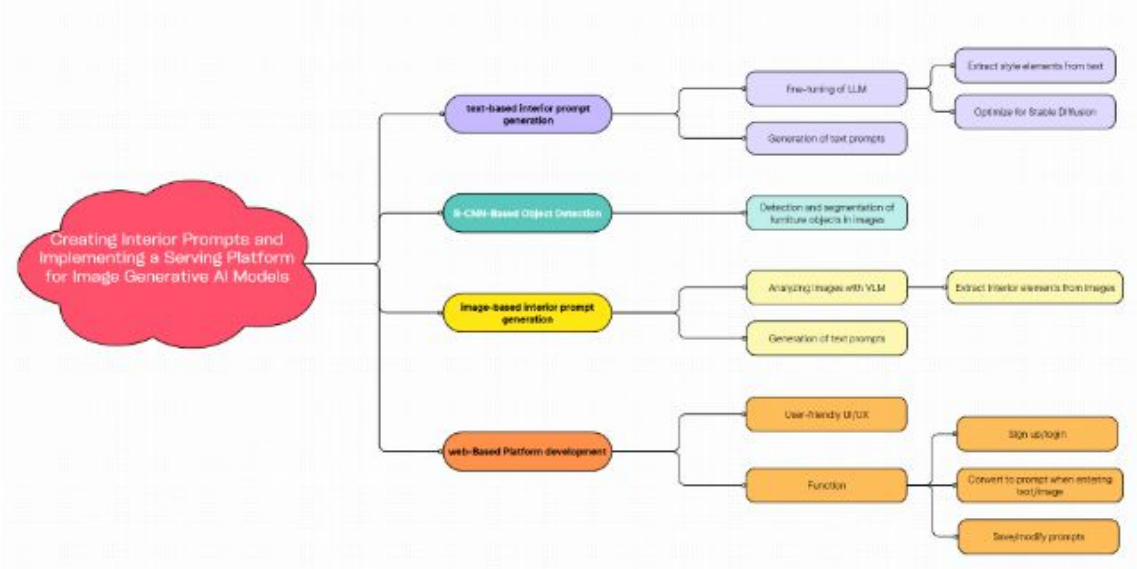
- A Visual Language Model (VLM) is an AI model capable of processing both visual information and textual data. It is widely used in tasks such as image caption generation, image-text retrieval, and multimodal learning. In this study, they utilize the CLIP (Contrastive Language-Image Pre-training) model, which excels at learning associations between images and text, to convert desired interior reference images into textual descriptions. To achieve this, the CLIP-based VLM is fine-tuned with a specialized dataset to enhance its ability to generate text specifically tailored for interior images.

3.1.3. Faster R-CNN

- Faster R-CNN is an effective model for detecting objects within images and classifying them into specific categories. However, it has limitations in accurately separating overlapping objects or objects within complex backgrounds, as it represents object boundaries only in rectangular form. In comparison, Mask R-CNN, an extension of Faster R-CNN, incorporates instance segmentation functionality to separate objects at the pixel level. This allows for precise recognition of furniture within room images, which can then be utilized for subsequent tasks such as texture analysis, color extraction, and size comparison.

4. Requirement Analysis (요구 분석)

4.1 Objective Tree (목적나무)



[Figure 1] Object Tree

In the figure, the project is categorized into four major functional features: text-based

interior prompt generation, R-CNN-based object detection, image-based interior prompt generation, and web-based platform development. Each function includes the necessary elements to implement and the goals to achieve.

4.2 Functional Requirement Analysis (요구 분석)

4.2.1 Function-Related Requirement (기능 관련 요구사항)

The text-based interior prompt generation feature should generate appropriate text prompts that include elements related to interior styles based on user input. To achieve this, the system will utilize an LLM (GPT-4) model for text processing, and the output prompt must be designed for use in an AI image generation model.

The R-CNN-based object detection feature will use Detectron2 and Mask R-CNN to detect objects in images and convert them into text. The attributes and textual data of the detected objects should be accessible to other system functions.

The image-based interior prompt generation feature should analyze user-uploaded images, extract features such as style, layout, and colors, and convert them into text prompts. In this process, a CLIP-based VLM model will analyze and textualize the images, and the generated prompts must be provided for AI models.

The web-based platform development will provide user accessibility to the above functions, enabling users to input text or upload images and view the prompt generation results through a web platform.

4.2.2 Data-Related Requirement (자료 관련 요구사항)

Input data must be provided in the form of text and images. The text input must consist of a string of up to 255 characters, while images must be in high-resolution JPEG or PNG format. Real-time processing must be performed based on this input data.

Output data will include generated text prompts and detected object information. The text prompts must be delivered as strings of up to 1000 characters, and detected object data must be in JSON format containing information such as object name, position, and size. Each user request must generate corresponding results.

Processed data, including generated prompts and analysis results, must be securely stored in a repository for reuse or verification purposes. Stored data should be retained for a specific period and securely managed.

4.2.3 Interface-Related Requirement (인터페이스 관련 요구사항)

The system must provide an intuitive web interface that allows users to input text or

upload images. Input data must be transmitted to the server through a RESTful API, where it will be processed in real-time.

The generated prompts and analysis results must be immediately visible in the web browser and also be available in JSON format. The input image data must be presented with a visual display of the results to enhance user experience. The UI/UX design must be simple and efficient to improve usability.

The system must seamlessly integrate with external AI systems such as Detectron2, CLIP, and LLM models. For data storage and management, cloud storage services like AWS S3 should be utilized.

4.2.4 User-Related Requirement (사용자 관련 요구사항)

Potential users of this technology include interior designers and non-experts seeking interior solutions. The system must cater to users with varying levels of technical expertise by providing an intuitive UI/UX. It should return accurate and fast results based on user input data (text or image).

The generated prompts must include options for users to review, edit, and save the results. Security features must be implemented to ensure user data protection.

4.3 Realistic Constrains (현실적 제한조건)

4.3.1 Constraints for Available Resources (가용 자원에 대한 제한조건)

The system development and execution environment must support GPU and cloud-based infrastructure capable of running models such as Mask R-CNN, Detectron2, CLIP, and LLM (GPT-4). The development and testing phases are planned for completion by the first half of 2025.

The current team consists of four members: two front-end developers and two back-end developers. However, most team members lack experience in AI-related technologies. Therefore, it will be essential to build understanding and hands-on experience with models like Mask R-CNN, Detectron2, and CLIP before the development begins in earnest.

4.3.2 Constraints for Minium Performance (최소 성능에 대한 제한조건)

The system must generate prompts for text input as quickly as possible and optimize them for Stable Diffusion to ensure greater accuracy than previous prompts. Image analysis must maintain a response time that does not compromise user experience.

4.3.3 Constraints for Security (보안에 대한 제한조건)

With the increasing importance of security issues, access to the system and its data must be strictly controlled. Both users and administrators must authenticate themselves to gain access, and access logs must be maintained to detect unauthorized activities.

All data stored in the system must be encrypted and securely protected during transmission to minimize the risk of data leakage or tampering. The system must ensure fault tolerance by implementing redundancy mechanisms to enable immediate recovery in case of failures. Periodic backups must be performed to prevent data loss, and the backups should be stored securely in a separate storage system. Regular data recovery simulations must be conducted to verify the stability of the recovery process.

Security policies must comply with current standards and regulations, including GDPR for data protection. Only the minimum necessary user data will be collected and managed strictly for project purposes.

4.3.4 Constraints for Compatibility (호환성에 대한 제한조건)

The system must be compatible with multiple operating systems, including Windows and macOS, and it must work stably on major browsers such as Chrome, Firefox, and Edge.

The system will adopt a modular architecture to facilitate future upgrades and maintenance. Standardized APIs will be used to ensure smooth interaction with external systems. The system's performance will be validated in various scenarios to identify and resolve potential compatibility issues in advance, ensuring long-term maintainability and operational efficiency.

4.3.5 Constraints for Ethics (윤리성에 대한 제한조건)

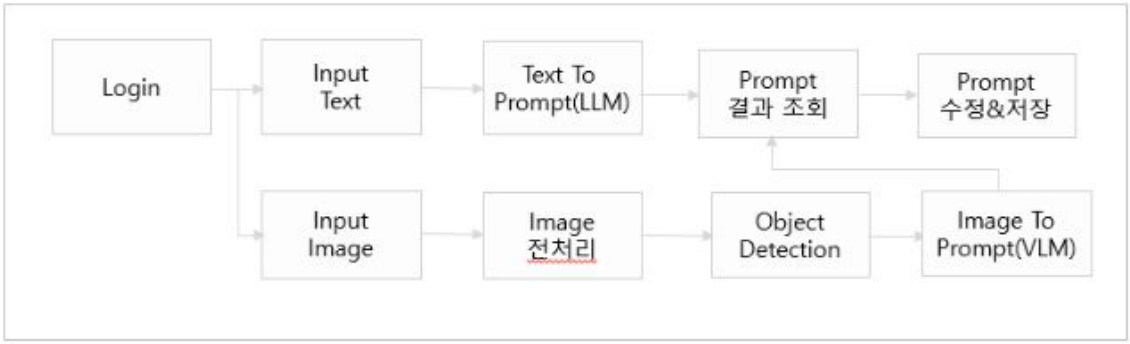
The project must strictly comply with intellectual property regulations. All AI models, datasets, and libraries used must be utilized under valid licenses, and the team must ensure adherence to these regulations throughout the project.

The team will collect only the minimum level of data required to achieve project goals. Continuous testing will be conducted to guarantee system stability and safety, preventing unexpected errors or design flaws from causing significant harm to users. Procedures for immediate corrections will be established, and any issues discovered will be promptly addressed.

5. Project Contents (과제 내용)

5.1 Design in Detail (상세 설계)

5.1.1 System Structure (시스템 구조)



[figure 2] 블록 다이어그램

5.1.2 Module Specification (모듈 설계)

1) Text-Based Interior Prompt Generation Module (텍스트 기반 인테리어 프롬프트 생성 모듈, LLM_001)

This module generates prompts for the interior design AI model based on user-input text. The input text provided by the user is delivered in JSON format, for example: {"user_input": "modern living room design"}. A fine-tuned GPT-4-mini model processes the input text, analyzes, and transforms it. The generated output is also returned in JSON format. An example of semantic guidance-applied output is as follows: {"output": "modern living room design, modern, living room, design"}.

2) Image Preprocessing Module (이미지 전처리 모듈)

The image preprocessing module processes input images to optimize the performance of the AI models. Uploaded images undergo resizing and pixel value normalization to match the required input format of the model. Additionally, color normalization is applied to meet the input requirements of the AI model. The primary algorithms used include OpenCV and PyTorch. The input data is an image file path, and the output is a preprocessed image used by the R-CNN-based object detection module.

3) R-CNN-Based Object Detection Module (R-CNN 기반 오브젝트 디텍션 모듈, DET_001)

This module analyzes uploaded images to detect furniture and decorative objects and converts them into textual data. Using Mask R-CNN, powered by Detectron2, the module performs instance segmentation of furniture within the image. Additionally, the detection

information is used to analyze the color and texture of the furniture. The detection results are provided in JSON format, for example: {"objects": [{"name": "sofa", "color": "gray", "material": "fabric"}]}. The detected data is then delivered to the image-based interior prompt generation module.

4) Image-Based Interior Prompt Generation Module (이미지 기반 인테리어 프롬프트 생성 모듈, VLM_001)

The image-based interior prompt generation module analyzes uploaded images to generate relevant interior styles and prompts. Using a VLM-based CLIP model, the content of the image is converted into text. The results are provided in JSON format, for example: {"prompt": "a Scandinavian living room with natural wood furniture and indoor plants"}.

5) Data Storage and Management Module (데이터 저장 및 관리 모듈)

The data storage and management module efficiently stores and manages user input data, AI-generated prompts, and detection results.

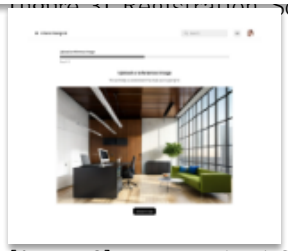
5.2 Implementation (구현)

5.2.1 Implementation Result (구현 결과)

[figure 3] Registration Screen

[figure 4] Login Screen

[figure 5] Main Screen



[figure 6] Image Upload Screen

[figure 7] Prompt Result Screen

[figure 8] Prompt Edit and Save Screen

[figure 3] The registration screen includes a feature for account creation. [figure 4] The login screen provides the login functionality. The initial screen upon accessing the service is the [figure 5] Main Screen, where users can select desired menus via the top navigation bar. Two key features are provided: Text-To-Prompt and Image-To-Prompt. The Image-To-Prompt function allows users to upload an image on [figure 6] Image Upload Screen. The uploaded image is processed into a prompt, which is displayed on [figure 7] Prompt Result Screen. The generated prompt can be edited and saved on [figure 8] Prompt

Edit and Save Screen.

5.2.2 Implementation Tools (구현 도구)

The project utilizes React.js, Express, MySQL, and AI models (ChatGPT-4-mini, Detectron2, CLIP) as software tools.

React.js handles the implementation of the user interface (UI), offering a fast and flexible library for frontend features such as text input, image upload, and viewing/editing prompt results. The backend server is designed using Express, ensuring stable performance for API development and data processing, while efficiently managing database interactions. MySQL is employed as the relational database management system to systematically store and manage text, image data, and user input/output results.

The AI models utilized include:

- LLM (Large Language Model): GPT-4-mini, which processes user-input text into prompts optimized for generating interior images. With its excellent natural language processing capabilities, it accurately understands user intent and delivers optimized results.

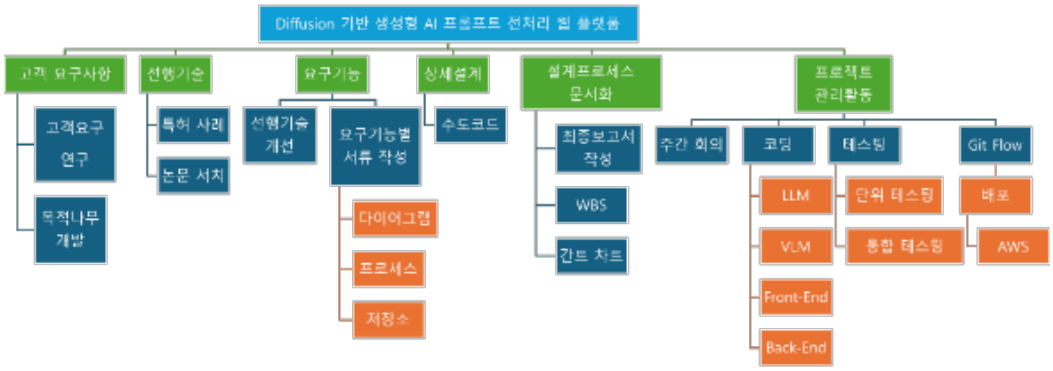
- Mask R-CNN-Based Model: Detectron2, a state-of-the-art object detection model leveraging deep learning frameworks to accurately and reliably detect key objects, such as furniture and decorations, in uploaded images.

- VLM (Vision-Language Model): CLIP, which links the detected image objects from Detectron2 with text to extract keywords related to the image's style. This enables users to receive furniture and decor recommendations based on information extracted from reference images.

These AI models run in a GPU environment, ensuring performance optimization and efficient execution.

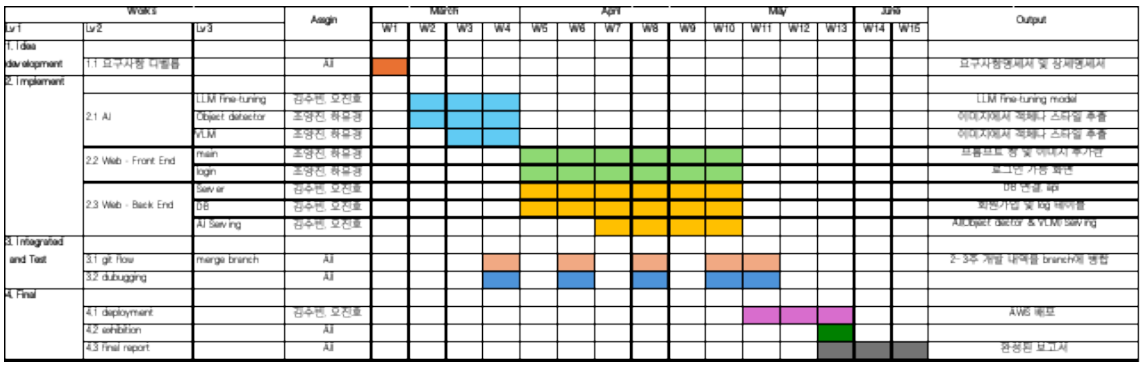
5.3 Tasks and Task Assignments (과업 및 과업 분담)

5.3.1 Work Breakdown Structure (과업 분해)



[figure 9] Work Breakdown Structure

5.3.2 Task Assignment and Project Timeline (과업 분담 및 과제 진행)



[figure 10] TimeTable

Tasks	Detail Description	Weight (1-10)	조양진 (%)	하유경 (%)	오진호 (%)	김수빈 (%)	Total (%)
Project Lead	Make meetings	2	40	20	20	20	100
Meeting	Meeting Facilitation	3	30	25	30	15	100
Minutes	Meeting Minutes Preparation	3	10	10	70	10	100
Presentation	Final Presentation	4	70	10	10	10	100
Report	Document Preparation	7	25	25	25	25	100
Individual Sum			**655**	**360**	**555**	**330**	

[figure 11] Work Division Contribution

6. Anticipated Results and Contributions (예상 결과 및 기대 효과)

6.1 Anticipated Results

Upon project completion, a web platform optimized for diffusion-based generative AI and interior design-specific prompt generation will be available. This platform will integrate both LLM and VLM functionalities. The software will be released as open-source, along with the dataset used for LLM fine-tuning, providing access for broader community use and further development.

6.2 Impact

- Impact on the Market

Our project offers a service for creating design blueprints that can be shared with interior contractors before initiating construction. This helps small business owners prevent disputes caused by unsatisfactory design proposals, allowing them to focus more on business operations despite their limited time and budget.

Additionally, this service can be applied not only to commercial spaces but also to residential interior design. With the interior market growing from KRW 7 trillion in 2008 to KRW 18 trillion in 2023, the trend of "home decorating" has become increasingly popular. By lowering the entry barrier for creating design blueprints, the service aligns with current trends and encourages consumers to purchase interior items, thereby contributing to the expansion of the related market.

- Impact on Technology

By open-sourcing the technology for generating interior-specific prompts, this project provides a global reference point, benefiting not only Korea but also other countries seeking to advance their generative AI applications in interior design.

- Impact on Human Life

Amid the current trend of increased interest in home decoration following the COVID-19 pandemic, this service allows individuals, regardless of their aesthetic sensibilities, to create design blueprints. This facilitates personalized home decoration and offers an opportunity to enjoy a more relaxed and fulfilling lifestyle.

7. References (참고 서적)

Ankur Agrawal, Vinod Koppurapu, Shirish Arashanapalli. “Systems and Methods for

Generating Interior Designs Using Artificial Intelligence", 2024

Chang Yu, "Seek for Incantations: Towards Accurate Text-to-Image Diffusion Synthesis through Prompt Engineering", 2024

박희수, "A i 기반 인테리어 공간설계를 위한 데이터 가공 서비스 제공 시스템", 2024

김진성, "딥러닝 기반 이미지 자동인식 기술을 활용한 사무집기 자동인식과 정보관리 시스템과의 연동방안", 한양대학교, 2017

8. Budget (예산)

The estimated cost of resources and literature acquisition required for the project is presented as follows. Resources include expenses for AWS deployment, LLM fine-tuning, and API call usage.

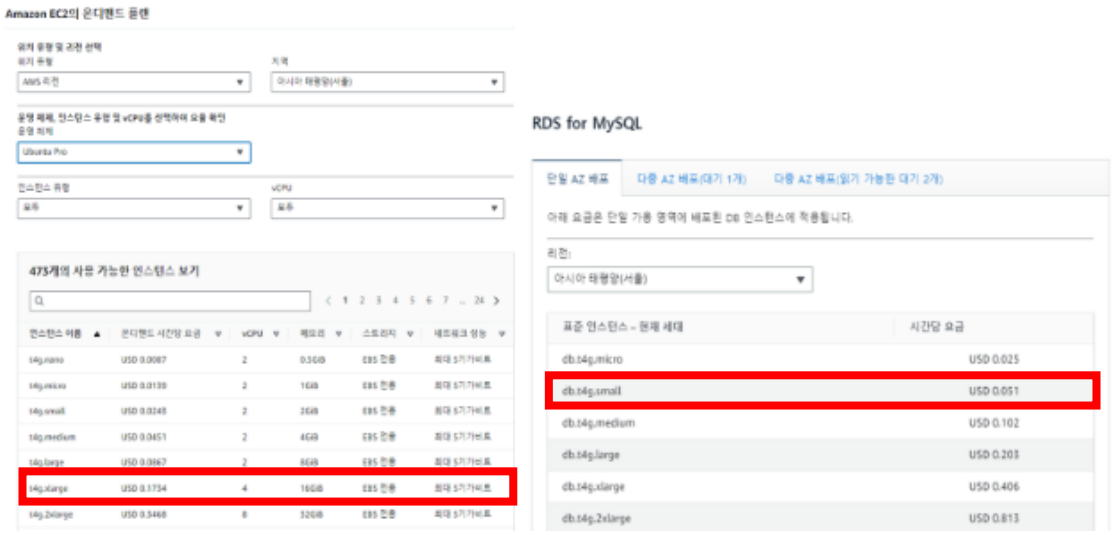
8.1. Resource(AWS Deployment)

8.1.1. AWS EC2 Instance

To run the project in a cloud environment, an EC2 instance is required. Based on the Gantt chart, deployment tasks are planned for 3 weeks, with usage estimated at 6 hours per day for 21 days. The expected cost is $21.8484 (0.1734 * 6 * 21).

8.1.2. AWS RDS Instance

A database is necessary to store prompts and other data during project execution. Based on the Gantt chart, deployment tasks will take place over 3 weeks, with usage estimated at 6 hours per day for 21 days. The expected cost is $6.426 (0.051 * 6 * 21).



[figure 12] AWS EC2 and RDS On-Demand Hourly Rates

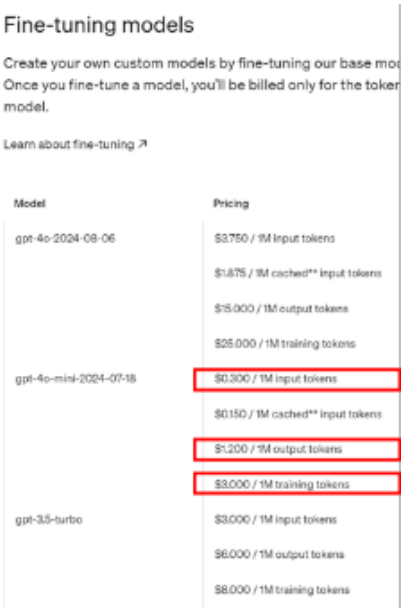
8.2. Resource(LLM fine-tuning)

8.2.1. GPT-4o-mini fine-tuning

For fine-tuning, 5M tokens are expected to be used, with an estimated cost of $15 (5 * 3).

8.2.2. GPT-4o-mini API Usage

Using the fine-tuned LLM for input and output requires $0.3 per 1M input tokens and $1.2 per 1M output tokens. For both training analysis and practical use, 3M tokens are needed for input and output respectively. The estimated cost is $4.50 (0.3 * 3 + 1.2 * 3).



[figure 13] GPT-4o-mini Training and API Input/Output Token Costs

8.3. Literature Acquisition Costs

8.3.1. 러닝 깃허브 액션

This book is essential for acquiring pre-requisite knowledge to manage the project using GitHub Actions. The price is KRW 27,000.

[figure 14] 러닝 깃허브 액션 (Yes24 price)

8.3.2 OpenAI API와 파이썬으로 나만의 챗GPT 만들기

Since fine-tuning for the LLM will be performed using the OpenAI API, this book is necessary. The price is KRW 21,600.

[figure 15] OpenAI API와 파이썬으로 나만의 챗GPT 만들기 (kyobobook price)

8.3.3 한 권으로 끝내는 실전 LLM 파인튜닝

For the VLM, direct fine-tuning is required, making this book essential for focused study. The price is KRW 25,200.

[figure 16] 한 권으로 끝내는 실전 LLM 파인튜닝 (kyobobook price)