

Documento, AI기반 빠르고 쉬운 논문 탐색 서비스

SK Networks Family AI Camp 3기

TEAM 1 / DOCUMENTO



01

개요

- I. 프로젝트 배경
- II. 경쟁 업체 조사
- III. 차별성
- IV. 프로젝트 목표

02

핵심 기능

- 0. 주요 요소
- I. 키워드 최적화
- II. 논문 검색
- III. 논문 핵심 요약
- IV. 선행 논문 추천
- V. 개발 현황

03

기대효과 및 계획

- I. 기대효과
- II. 향후 계획
- III. WBS
- IV. 팀원 소개

01

개요

프로젝트 배경 개요

논문 시장의 성장

지난 5년 동안 6개 이상의 주요 학술 출판사에서 논문 출판 수가 두 배 이상 증가하며, 학술 논문 시장은 지속적으로 확대되고 있다.

특히 인공지능(AI) 관련 논문이 급격히 증가하고 있으며, 2023년에는 AI 모델의 수가 2022년 대비 두 배로 늘어났다. 이로 인해 인공지능 기술이 엔지니어들의 일상에 필수적으로 자리 잡으면서, 관련 논문을 탐색하는 일이 매우 중요해졌다.



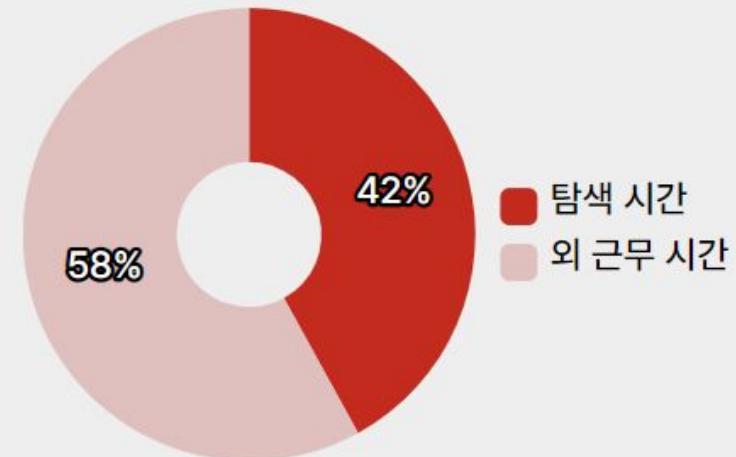
프로젝트 배경 개요

근무 중 탐색 시간 多

IDC의 연구에 따르면, 엔지니어들은 근로 시간의 약 42%를 정보 탐색 및 공유에 할애한다고 보고되었다.

논문 시장의 성장에 따라 탐색에 보다 많은 시간이 소요될 것으로 예상된다.

실제로 2011년 대비 학술 문헌 검색 시간이 11% 증가(연구자 기준 주당 3h 42 → 4h 8m)했다고 한다.



※ 2016, IDC & 2024, Stanford HAI, 2019 Elsevier 분석, 발췌, 번역

01-1

프로젝트 배경 개요

논문 시장의 성장

근무 중 탐색 시간 多

빠르고 간편한
논문 검색의 필요성



경쟁업체 조사 개요

SciSpace AI

타겟.

관련 연구자, 석/박사 학생

주요 특징 1.

대규모 Global 논문 데이터셋

주요 특징 2.

Abstract 요약/분석

범위.

Global

R discovery

타겟.

관련 연구자, 석/박사 학생

주요 특징 1.

인공지능 관련 전공별,

주제별 최신 논문 추천

주요 특징 2.

전체 내용 정리 및

Abstract 요약/분석

범위.

Global

DBpia

타겟.

관련 연구자, 석/박사 학생

주요 특징 1.

학술 논문, 학위 논문, 학술지,
컨퍼런스 자료 등

다양한 자료 검색 가능

주요 특징 2.

연구 주제, 배경, 방법론 및
연구 결과 등 요약

범위.

Local (Korea)

경쟁업체 조사 개요

SciSpace AI

타겟.

관련 연구자, 석/박사 학생

특징 1.

대규모 Global 논문 데이터셋

특징 2.

Abstract 요약/분석

범위.

Global

R discovery

타겟.

관련 연구자, 석/박사 학생

특징 1.

인공지능 관련 전공별,
주제별 최신 논문 추천

기존 타겟팅의 한계 존재

DBpia

타겟.

관련 연구자, 석/박사 학생

특징 1.

학술 논문, 학위 논문, 학술지,
컨퍼런스 자료 등

양한 자료 검색 가능

특징 2.

구 주제, 배경, 방법론 및
구 결과 등 요약

범위.

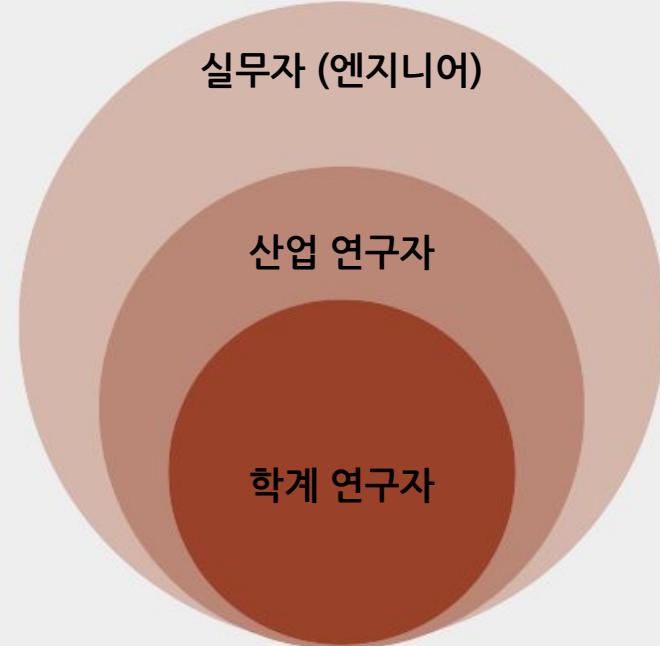
Local (Korea)

01- III

차별성 개요

AI 기술의 확산에 따라, **지속적인 논문 탐색은 연구자뿐만 아니라 엔지니어에게도 필수적인 활동**이 되고 있다.

신기술 습득을 위해 학술 논문을 참고하는 실무자들의 수요가 증가하고 있으며, 이에 따라 연구자 중심의 검색 지원 접근을 넘어서 **기술 습득이 중요한 실무자들을 타겟**으로 하는 서비스가 필요하다.



※ 논문 탐색 필요성의 대상이 점점 확대됨

차별성 개요

Documento

타겟.

관련 기술 습득을 목표로 하는 실무자

주요 특징 1.

실무의 언어에서부터
논문 관련 키워드로 키워드 최적화

주요 특징 2.

관련 선행 지식을 파악할 수 있도록
유사도 기반 선행 논문 추천

해당 분야를 잘 이해하고 있지 못하더라도,
필요한 내용의 논문을 검색할 수 있도록 함

범위.

Local (Korea)

01-IV

프로젝트 목표 개요

For Who?

기술 습득을 목표로 하는 실무자들이

How?

논문 키워드 최적화와 선행 논문 추천 기능 도입을 통해

What?

기존의 학술 논문에 대한 접근 장벽을 낮추고,
빠르게 필요한 정보를 얻을 수 있도록 돕는 서비스

02

기능



02-0

주요 요소 핵심 기능

키워드 최적화

논문 검색

논문 핵심 요약

선행 논문 추천

02-1

키워드 최적화 핵심 기능

키워드 최적화



사용자가 실무에서 사용하는 용어와
논문 검색에서 사용하는 용어 간의 극복

논문 검색

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images. However, these models have been trained to be good at one task, namely image captioning, and have not been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language tasks. In this work, we propose a two-stage model: a prior that generates a CLIP embedding conditioned on the image embedding. We show that explicitly generating image representations conditioned on the image embedding leads to better image captioning and image caption similarity. Our decoder conditioned on image representations can also predict variations of the image that preserve both its semantics and style, while varying the caption. Finally, we show that our model can be used for image captioning and the zero-shot embedding space of CLIP enables language-guided image manipulation in a controllable way. We hope that this work will encourage more research on joint optimization with both autoregressive and diffusion models for the price, finding that the latter are computationally more efficient and produce higher quality outputs.

1 Introduction

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images. However, these models have been trained to be good at one task, namely image captioning, and have not been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language tasks. In this work, we propose a two-stage model: a prior that generates a CLIP embedding conditioned on the image embedding. We show that explicitly generating image representations conditioned on the image embedding leads to better image captioning and image caption similarity. Our decoder conditioned on image representations can also predict variations of the image that preserve both its semantics and style, while varying the caption. Finally, we show that our model can be used for image captioning and the zero-shot embedding space of CLIP enables language-guided image manipulation in a controllable way. We hope that this work will encourage more research on joint optimization with both autoregressive and diffusion models for the price, finding that the latter are computationally more efficient and produce higher quality outputs.

키워드 최적화 핵심 기능

사용자 키워드 입력

추천 키워드 제공 및
해당 검색 결과 n개 제공
(북마킹 가능)

원하는 논문의 내용을 자신만의 언어로 표현해보세요.

책마크 리스트

DALL-E2 : Hierarchical Te ...
텍스트-이미지 생성

DALL-E2 : Hierarchical Te ...
텍스트-이미지 생성

텍스트-이미지 변환 기능과 관련된 논문을 탐색하기 위한 주요 키워드는 다음과 같이 설정할 수 있습니다.

1. 텍스트-이미지 생성 (Text to Image generation)

DALL-E 2
OpenAI에서 개발한 DALL-E 2는 GPT-3의 같은 구조를 사용하여 텍스트를 이미지로 변환하는 데 뛰어난 성능을 보이며, "armchair in the shape of an avocado"와 같은 복잡한 텍스트도 이해하여 고유한 이미지를 생성할 수 있습니다.

DALL-E 2
OpenAI에서 개발한 DALL-E 2는 GPT-3와 같은 구조를 사용하여 텍스트를 이미지로 변환하는 데 뛰어난 성능을 보이며, "armchair in the shape of an avocado"와 같은 복잡한 텍스트도 이해하여 고유한 이미지를 생성할 수 있습니다.

2. 비전-언어 모델 (Vision-language models)

DALL-E 2
OpenAI에서 개발한 DALL-E 2는 GPT-3와 같은 구조를 사용하여 텍스트를 이미지로 변환하는 데 뛰어난 성능을 보이며, "armchair in the shape of an avocado"와 같은 복잡한 텍스트도 이해하여 고유한 이미지를 생성할 수 있습니다.

DALL-E 2
OpenAI에서 개발한 DALL-E 2는 GPT-3와 같은 구조를 사용하여 텍스트를 이미지로 변환하는 데 뛰어난 성능을 보이며, "armchair in the shape of an avocado"와 같은 복잡한 텍스트도 이해하여 고유한 이미지를 생성할 수 있습니다.

02-1

키워드 최적화 핵심 기능

< 활용 기술 >



- OpenAI API
- 프롬프트 엔지니어링

ex)

키워드 최적화

논문 검색

적은 데이터로 효과적인 대화형 LLM 모델 만들기

텍스트-이미지 변환 기능과 관련된 논문을 탐색하기 위한 주요 키워드는 다음과 같이 설정할 수 있습니다.

1. [스트-이미지 생성] (Text-to-image generation)
Few-shot dialogue generation with large language models
[대규모 언어 모델을 활용한 소량 샘플 대화 생성]
(논문 예시 1) (논문 예시 2) (논문 예시 3)

2. [대화형 에이전트를 위한 메타 학습]
Meta-learning for conversational agents
[대화형 에이전트를 위한 메타 학습]
(논문 예시 1) (논문 예시 2) (논문 예시 3)

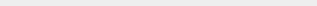
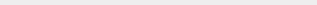
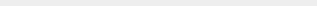
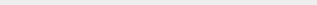
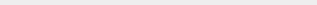
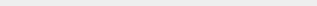
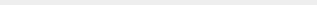
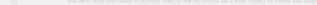
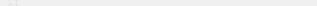
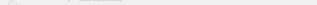
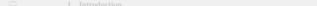
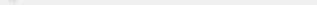
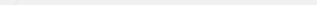
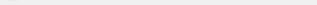
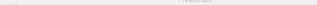
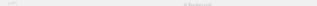
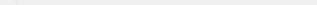
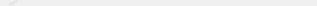
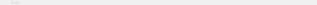
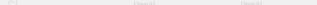
DALL-E 2
아래에서 개발한 DALL-E 2는 GPT-3와 같은 구조를 사용하여 텍스트를 이미지로 변환하는 데 뛰어난 성능을 보이며, "armchair in the shape of an avocado"와 같은 복잡한 텍스트도 이해하여 고유한 이미지를 생성할 수 있습니다.

DALL-E 2
아래에서 개발한 DALL-E 2는 GPT-3와 같은 구조를 사용하여 텍스트를 이미지로 변환하는 데 뛰어난 성능을 보이며, "armchair in the shape of an avocado"와 같은 복잡한 텍스트도 이해하여 고유한 이미지를 생성할 수 있습니다.

02-II

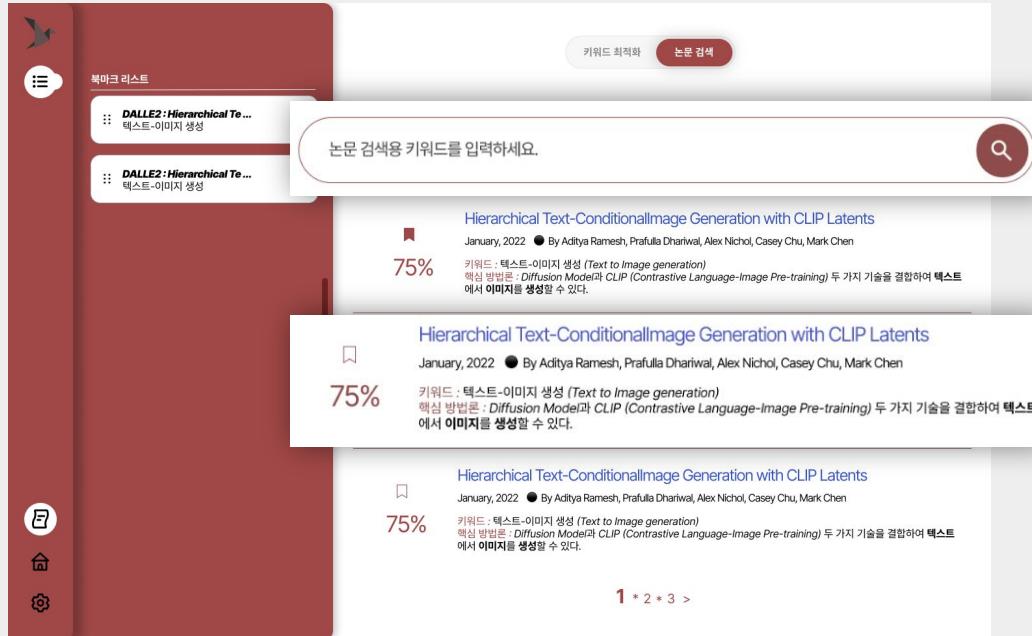
논문 검색 핵심 기능

키워드 최적화



02-II

논문 검색 핵심 기능



논문 검색용 키워드 입력

관련 논문 정보 제공

- 키워드 유사도
- 제목
- 발행 연월
- 저자
- 인용수
- 세부 내용

논문 검색 핵심 기능

카워드 최적화

논문 검색

논문 검색용 키워드를 입력하세요.

DALLE2: Hierarchical Text-Conditional Image Generation with CLIP Latents

January, 2022 ● By Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, Mark Chen

75%

키워드 : 텍스트-이미지 생성 (Text to Image generation)
핵심 방법론 : Diffusion Model과 CLIP (Contrastive Language-Image Pre-training) 두 가지 기술을 결합하여 텍스트에서 이미지를 생성할 수 있다.

Hierarchical Text-Conditional Image Generation with CLIP Latents

January, 2022 ● By Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, Mark Chen

75%

키워드 : 텍스트-이미지 생성 (Text to Image generation)
핵심 방법론 : Diffusion Model과 CLIP (Contrastive Language-Image Pre-training) 두 가지 기술을 결합하여 텍스트에서 이미지를 생성할 수 있다.

Hierarchical Text-Conditional Image Generation with CLIP Latents

January, 2022 ● By Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, Mark Chen

75%

키워드 : 텍스트-이미지 생성 (Text to Image generation)
핵심 방법론 : Diffusion Model과 CLIP (Contrastive Language-Image Pre-training) 두 가지 기술을 결합하여 텍스트에서 이미지를 생성할 수 있다.

1 * 2 * 3 >

◀ 활용 기술 ▶

SPECTER: Document-level Representation Learning using Citation-informed Transformers

Arman Cohan¹, Sergey Fidler^{2*}, Ir英博³, Doug Downey⁴, David S. Wolff¹

¹Allen Institute for Artificial Intelligence
²UT Dallas
³University of Washington
⁴Amazon, AWS Research, Seattle, USA; faulina@allenai.org

Abstract

In recent years, research in multi-domain machine learning has focused on improving cross-domain generalization by leveraging shared knowledge from related domains. One approach to this is to learn domain-invariant representations that can be used across domains. A key challenge in this space is how to effectively incorporate domain-specific information into the representation learning process. For example, document-level representations often include domain-specific context such as references to other documents in the same domain, and domain-specific knowledge such as domain-specific terminology and nomenclature. In this paper, we propose a new framework for learning domain-invariant document-level representations that leverages citation-informed transformers to learn domain-invariant representations that are robust to domain-specific context. We propose SPECTER, a citation-informed transformer that uses citations to learn domain-invariant document-level representations. Our experiments show that SPECTER outperforms state-of-the-art document-level representation models on several downstream tasks. We propose SPECTER, a citation-informed transformer that uses citations to learn domain-invariant document-level representations. Our experiments show that SPECTER outperforms state-of-the-art document-level representation models on several downstream tasks. The language modeling objectives used to pretrain the model are also helpful for document-level tasks such as topic

- **SPECTER (벡터 표현)**

동일 논문 데이터 (Semantic Scholar Corpus)로 사전 학습 되어있어 적합한 모델이라 판단.

- **FAISS (벡터 저장)**

빠른 검색 속도를 활용하기 위해 선정.

02- III

논문 핵심 요약 핵심 기능

논문 선택에 도움이 되도록 Preview형식의 논문 특화 요약 제공

키워드 최적화

논문 검색

논문 핵심 요약

Hierarchical Text-Conditional Image Generation with CLIP Latents

January, 2022 By Aditya Ramesh, Pratul Dhariwal, Alex Nichol, Casey Chu, Mark Chen

논문 핵심 요약 (Text to Image generation)
논문 핵심 요약 (Text to Image generation)
논문 핵심 요약 (Text to Image generation)

Aditya Ramesh*, OpenAI
pratuldhariwal@openai.com
Casey Chu*, OpenAI
casey@openai.com
Pratul Dhariwal*, OpenAI
pratuldhariwal@openai.com
Alex Nichol*, OpenAI
alex@openai.com
Mark Chen, OpenAI
mark@openai.com

Abstract

Contrastive models like CLIP have been shown to learn robust representations of images and text. To facilitate their use for image generation, we propose a two-stage model: a first that generates a CLIP image embedding conditioned on a text prompt, and a second that generates an image conditioned on the image embedding. We show that explicitly generating image representations conditioned on text can also produce variations of an image that preserve both its semantics and style, while varying the text prompt. In addition, we show that by conditioning on text, the joint embedding space of CLIP enables language-guided image manipulation in a way that is more efficient than directly manipulating the image representation with both autoregressive and diffusion models for the price, finding that the latter are comparatively more efficient and produce higher-quality samples.

Introduction

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images. These models have been used for a variety of downstream tasks, such as image classification, as a successful representation learner for images. CLIP embeddings have a number of desirable properties, such as being able to represent both visual and textual concepts, and have been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language tasks.

arXiv:2204.06251 [cs.CV] 13 Apr 2022

논문 핵심 요약 핵심 기능

논문 핵심 요약

Hierarchical Text-Conditional Image Generation with CLIP Latents
텍스트-이미지 생성

키워드 : 텍스트-이미지 생성 (Text to Image generation)

핵심 방법론 : DALL-E 2는 Diffusion Model과 CLIP의 조합으로 작동하여 텍스트-이미지 생성에서 높은 수준의 성능을 보여줍니다. 텍스트 입력이 주어지면, CLIP은 이를 분석하여 관련된 이미지 특성을 파악하고, Diffusion Model은 이 정보에 따라 처음에는 노이즈 상태의 이미지에서 시작하여 점진적으로 텍스트에 맞는 이미지를 생성하게 됩니다.

활용 기술 :
Diffusion Model
Diffusion Model은 이미지를 점진적으로 변화시키는 과정에서 노이즈를 추가하고 제거하여 이미지를 생성하는 방식입니다. 처음에는 순수한 노이즈로부터 시작하여 단계적으로 노이즈를 제거함으로써 텍스트에서 묘사된 구체적인 형태의 이미지를 생성할 수 있습니다. 이러한 역방향 노이즈 제거 과정은 이미지를 점점 선명하고 세밀하게 만들며, 고해상도와 자연스러운 디테일을 확보하게 합니다.

CLIP (Contrastive Language-Image Pre-training)
CLIP 모델은 텍스트와 이미지 간의 연관성을 학습하여, 주어진 텍스트에 맞는 이미지의 세부 요소들을 판단하고 최적화합니다. 이 모델은 텍스트와 이미지의 임베딩을 동일한 공간에서 학습하고, 텍스트와 이미지가 어떤 관계를 가지는지 대조 학습(contrastive learning)을 통해 학습합니다. 이를 통해 모델이 텍스트 입력에 적합한 이미지를 더 잘 이해하고 생성할 수 있도록 돕습니다.

E2 : Hierarchical Te ...
Generation

Hierarchical Text-Conditional Image Generation with CLIP Latents

Aditya Ramesh*
OpenAI
aramesh@openai.com

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com

Casey Chu*
OpenAI
casey@openai.com

Mark Chen
OpenAI
mark@openai.com

Abstract

Contrastive models like CLIP have been shown to learn robust representations of images that capture both semantics and style. To leverage these representations for image generation, we propose a two-stage model: a prior that generates a CLIP image embedding given a text caption, and a decoder that generates an image conditioned on the image embedding. We show that explicitly generating image representations improves image diversity with minimal loss in photorealism and caption similarity. Our decoders conditioned on image representations can also produce variations of an image that preserve both its semantics and style, while varying the overall composition. Most interestingly, however, the joint embedding space of CLIP enables language-guided image manipulations in a zero-shot fashion. We use diffusion models for the decoder and experiment with both autoregressive and diffusion models for the prior, finding that the latter are computationally more efficient and produce higher-quality samples.

1 Introduction

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images collected from the internet [10, 44, 60, 39, 31, 16]. Within this framework, CLIP [39] has emerged as a successful representation learner for images. CLIP embeddings have a number of desirable properties: they are robust to image distribution shift, have impressive zero-shot capabilities, and have been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language tasks [14, 4, 46, 14, 42, 24].

논문 핵심 요약 제공

- 제목
- 사용자 검색 키워드
- 논문 키워드
- 핵심 방법론
- 활용 기술

02- III

논문 핵심 요약 핵심 기능

논문 핵심 요약

Hierarchical Text-Conditional Image Generation with CLIP Latents
텍스트-이미지 생성

키워드 : 텍스트-이미지 생성 (Text to Image generation)

핵심 방법론 : DALL-E 2는 Diffusion Model과 CLIP의 조합으로 작동하여 텍스트-이미지 생성에서 높은 수준의 성능을 보여줍니다. 텍스트 입력이 주어지면, CLIP은 이를 분석하여 관련된 이미지 특성을 파악하고, Diffusion Model은 이 정보에 따라 처음에는 노이즈 상태의 이미지에서 시작하여 점진적으로 텍스트에 맞는 이미지를 생성하게 됩니다.

활용 기술 :
Diffusion Model
Diffusion Model은 이미지를 점진적으로 변화시키는 과정에서 노이즈를 추가하고 제거하여 이미지를 생성하는 방식입니다. 처음에는 순수한 노이즈로부터 시작하여 단계적으로 노이즈를 제거함으로써 텍스트에서 묘사한 구체적인 형태의 이미지를 생성할 수 있습니다. 이러한 역방향 노이즈 제거 과정은 이미지를 점점 선명하고 세밀하게 만들기이며, 고해상도와 자연스러운 디테일을 확보하게 합니다.

CLIP (Contrastive Language-Image Pre-training)
CLIP 모델은 텍스트와 이미지 간의 연관성을 학습하여, 주어진 텍스트에 맞는 이미지의 세부 요소들을 판단하고 최적화합니다. 이 모델은 텍스트와 이미지의 임베딩을 동일한 공간에서 학습하고, 텍스트와 이미지가 어떤 관계를 가지는지 대조 학습(contrastive learning)을 통해 학습합니다. 이를 통해 모델이 텍스트 입력에 적합한 이미지를 더 잘 이해하고 생성할 수 있도록 돕습니다.

E2 : Hierarchical Te...
Generation

Hierarchical Text-Conditional Image Generation with CLIP Latents

Aditya Ramesh*
OpenAI
aramesh@openai.com

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com

Casey Chu*
OpenAI
casey@openai.com

Mark Chen
OpenAI
mark@openai.com

Abstract

Contrastive models like CLIP have been shown to learn robust representations of images that capture both semantics and style. To leverage these representations for image generation, we propose a two-stage model: a prior that generates a CLIP image embedding given a text caption, and a decoder that generates an image conditioned on the image embedding. We show that explicitly generating image representations improves image diversity with minimal loss in photorealism and caption similarity. Our decoders conditioned on image representations can also produce variations of an image that preserve both its semantics and style, while varying its overall composition. Finally, we show that the joint embedding space of CLIP enables language-guided image manipulations in a zero-shot fashion. We use diffusion models for the decoder and experiment with both autoregressive and diffusion models for the prior, finding that the latter are computationally more efficient and produce higher-quality samples.

1 Introduction

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images collected from the internet [10, 44, 60, 39, 31, 16]. Within this framework, CLIP [39] has emerged as a successful representation learner for images. CLIP embeddings have a number of desirable properties: they are robust to image distribution shift, have impressive zero-shot capabilities, and have been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language



< 활용 기술 >



Llama 3.1 8B

Meta

- Llama 3.1 8B

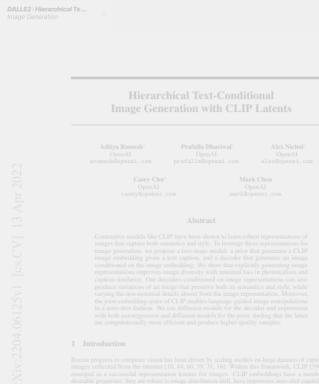
무료 모델 대비 높은 자유도로 프롬프트를 유연하게 수정 가능

- 프롬프트 엔지니어링

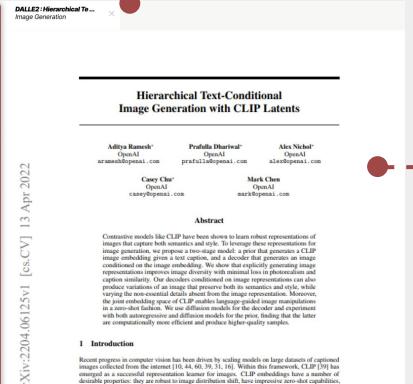
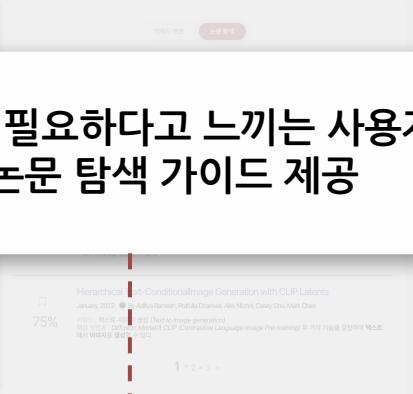
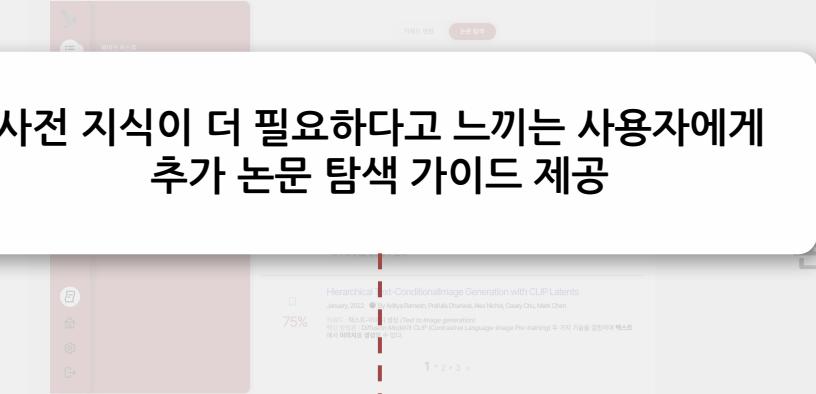
02-IV

선행 논문 추천 핵심 기능

키워드 최적화



사전 지식이 더 필요하다고 느끼는 사용자에게
추가 논문 탐색 가이드 제공



선행 논문 추천

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images. However, the resulting models have become increasingly complex and difficult to interpret. This has emerged as a successful representation learner for images. CLIP embeddings have a number of notable properties, such as being language-guided, being able to represent both visual and semantic modalities, and have been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language tasks.

1 Introduction

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images. However, the resulting models have become increasingly complex and difficult to interpret. This has emerged as a successful representation learner for images. CLIP embeddings have a number of notable properties, such as being language-guided, being able to represent both visual and semantic modalities, and have been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language tasks.

02-IV

선행 논문 추천 핵심 기능

선행 논문 추천

DALLE2 : Hierarchical Te ... 텍스트-이미지 생성

75% DALLE2 : Hierarch ... 텍스트-이미지 생성

키워드 : 텍스트-이미지 생성 (Text to Image generation)

핵심 방법론 : Diffusion Model과 CLIP (Contrastive Language-Image Pre-training) 두 가지 기술을 결합하여 텍스트에서 이미지를 생성할 수 있다.

유사 부분 요약 : CLIP Embedding을 활용하여 어찌구 저찌구 이렇게 하는 것이 일일하게 적용되며 어찌구 저찌구를 활용하여 이렇게 저렇게 하는 부분이 유사하다.

75% DALLE2 : Hierarch ... 텍스트-이미지 생성

75% DALLE2 : Hi ... 텍스트-이미지 ...

75% DALLE2 : Hierarch ... 텍스트-이미지 생성

Hierarchical Text-Conditional Image Generation with CLIP Latents

Aditya Ramesh^{*} OpenAI aramesh@openai.com Prafulla Dhariwal^{*} OpenAI prafulla@openai.com Alex Nichol^{*} OpenAI alex@openai.com

Casey Chu^{*} OpenAI casey@openai.com Mark Chen OpenAI mark@openai.com

Abstract

Contrastive models like CLIP have been shown to learn robust representations of images that capture both semantics and style. To leverage these representations for image generation, we propose a two-stage model: a prior that generates a CLIP image embedding given a text caption, and a decoder that generates an image conditioned on the image embedding. We show that explicitly generating image representations improves image diversity with minimal loss in photorealism and caption similarity. Our decoders conditioned on image representations can also produce variations of an image that preserve both its semantics and style, while varying its visual style. In addition to image generation, we show that decoders in the joint embedding space of CLIP enables language-guided image manipulations in a zero-shot fashion. We use diffusion models for the decoder and experiment with both autoregressive and diffusion models for the prior, finding that the latter are computationally more efficient and produce higher-quality samples.

1 Introduction

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images collected from the internet [10, 44, 60, 39, 31, 16]. Within this framework, CLIP [39] has emerged as a successful representation learner for images. CLIP embeddings have a number of desirable properties: they are robust to image distribution shift, have impressive zero-shot capabilities, and have been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language tasks [14, 44, 46, 40, 14, 42, 24].

선행 논문 리스트 제공

트리 형식, 유사도 상위 N개 제공

- 유사도
- 제목
- 논문 키워드
- 핵심 방법론
- 활용 기술

02-IV

선행 논문 추천 핵심 기능

선행 논문 추천

DALLE2 : Hierarchical Te... 텍스트-이미지 생성

DALLE2 : Hierarch... 텍스트-이미지 생성

키워드 : 텍스트-이미지 생성 (Text to Image generation)

핵심 방법론 : Diffusion Model과 CLIP (Contrastive Language-Image Pre-training) 두 가지 기술을 결합하여 텍스트에서 이미지를 생성할 수 있다.

유사 부분 요약 : CLIP Embedding을 활용하여 어찌구 저찌구 이렇게 하는 것이 동일하게 적용되며 어찌구 저찌구를 활용하여 이렇게 저렇게 하는 부분이 유사하다.

DALLE2 : Hierarch... 텍스트-이미지 생성

DALLE2 : Hi... 텍스트-이미지 ...

DALLE2 : Hierarch... 텍스트-이미지 생성

Hierarchical Text-Conditional Image Generation with CLIP Latents

Aditya Ramesh^{*}
OpenAI
aramesh@openai.com

Prafulla Dhariwal^{*}
OpenAI
prafulla@openai.com

Alex Nichol^{*}
OpenAI
alex@openai.com

Casey Chu^{*}
OpenAI
casey@openai.com

Mark Chen^{*}
OpenAI
mark@openai.com

Abstract

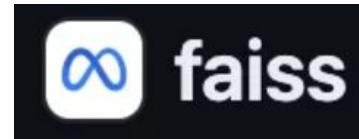
Contrastive models like CLIP have been shown to learn robust representations of images that capture both semantics and style. To leverage these representations for image generation, we propose a two-stage model: a prior that generates a CLIP image embedding given a text caption, and a decoder that generates an image conditioned on the image embedding. We show that explicitly generating image representations improves image diversity with minimal loss in photorealism and caption similarity. Our decoders conditioned on image representations can also produce variations of an image that preserve both its semantics and style, while varying its composition. Interestingly, our decoders can also generate images even when the joint embedding space of CLIP enables language-guided image manipulations in a zero-shot fashion. We use diffusion models for the decoder and experiment with both autoregressive and diffusion models for the prior, finding that the latter are computationally more efficient and produce higher-quality samples.

1 Introduction

Recent progress in computer vision has been driven by scaling models on large datasets of captioned images collected from the internet [10, 44, 60, 39, 31, 16]. Within this framework, CLIP [39] has emerged as a successful representation learner for images. CLIP embeddings have a number of desirable properties: they are robust to image distribution shift, have impressive zero-shot capabilities, and have been fine-tuned to achieve state-of-the-art results on a wide variety of vision and language



< 활용 기술 >

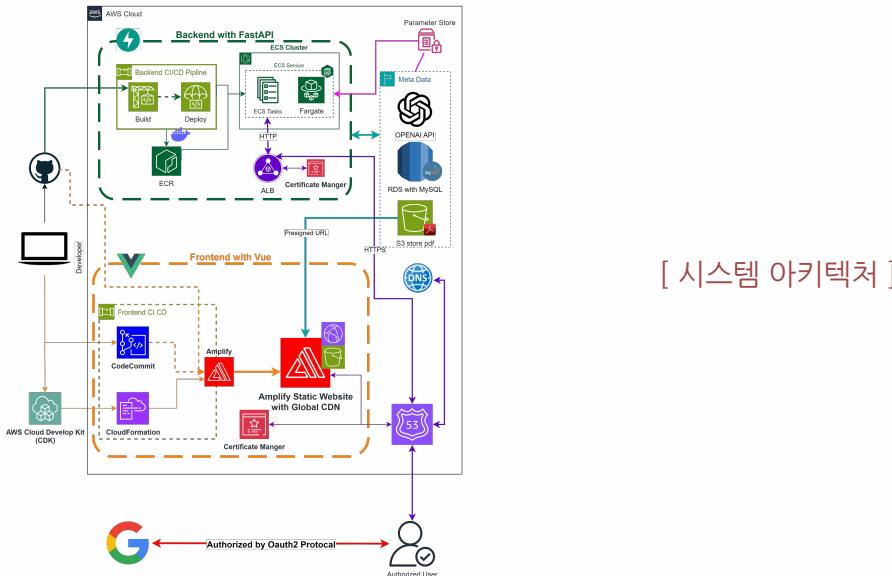


- **FAISS (벡터 저장)**
빠른 검색 속도를 활용하기 위해 선정.
- **Cosine Similarity 기반 유사도 계산**
크기보다는 방향성을 중요시하므로 텍스트 데이터에 적합하다고 판단.
(추후 다른 방법론과 수치적 비교 예정)

개발 현황 핵심 기능

아키텍처

- AWS 기반 클라우드 인프라 구축 완료

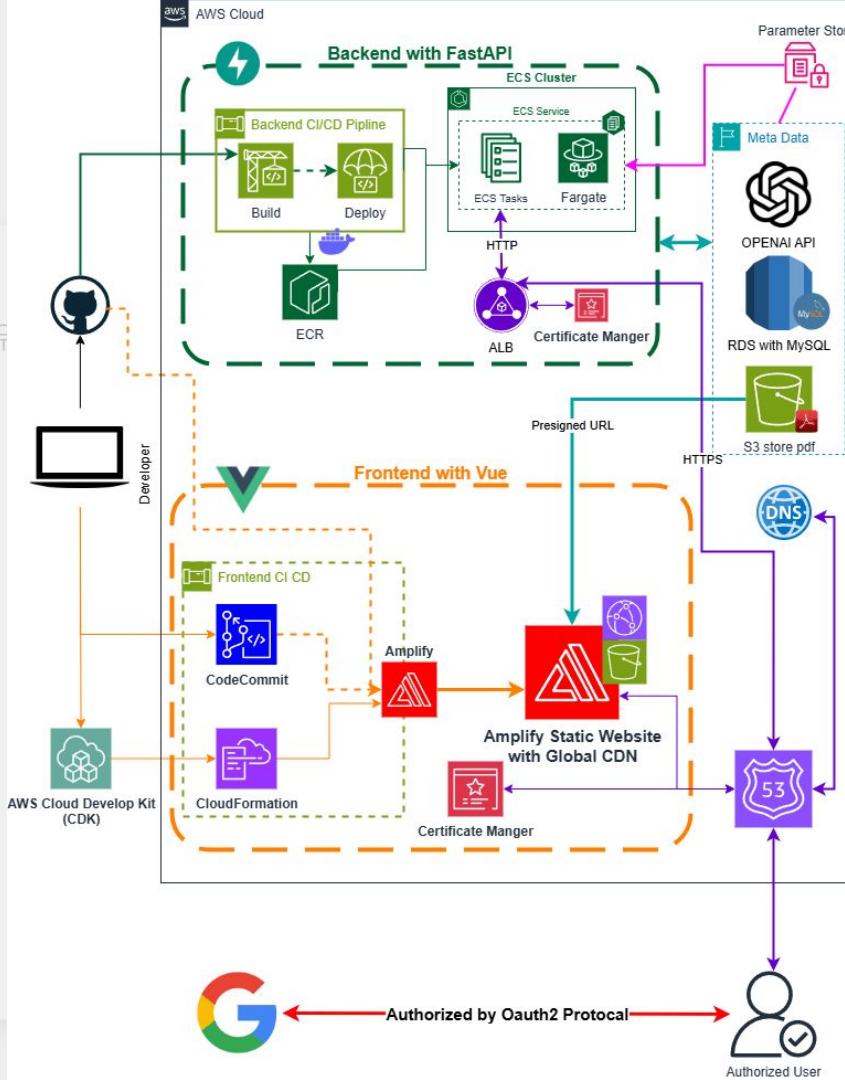


02-V

개발 현황 핵심 기능

아키텍처

- AWS 기반 클라우드

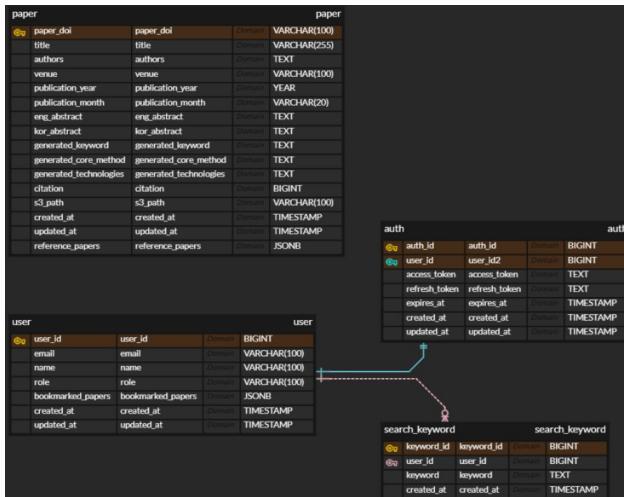


02-V

개발 현황 핵심 기능

데이터

- ACL과 Semantic Scholar API를 활용한 데이터 구축 완료 (2020~2024, ACL/EMNLP, 약 7600개)
 - 유사도 검색용 벡터 데이터 베이스 구축 완료



[메타 데이터 베이스 ERD]

(이 외 S3 내부 원문 데이터 및 벡터 데이터베이스 존재)

The diagram illustrates the schema of four tables: paper, auth, user, and search_keyword. Relationships are indicated by arrows:

- paper** table (top left):
 - Columns: paper_doi, title, authors, venue, publication_year, publication_month, eng_abstract, kor_abstract, generated_keyword, generated_core_method, generated_technologies, citation, s3_path, created_at, updated_at, reference_papers.
 - Domain types: VARCHAR(100), VARCHAR(255), TEXT, VARCHAR(100), YEAR, VARCHAR(20), TEXT, TEXT, TEXT, TEXT, TEXT, BIGINT, VARCHAR(100), TIMESTAMP, TIMESTAMP, JSONB.
- auth** table (top right):
 - Columns: auth_id, user_id, access_token, refresh_token, expires_at, created_at, updated_at.
 - Domain types: BIGINT, BIGINT, TEXT, TEXT, TIMESTAMP, TIMESTAMP, TIMESTAMP.
- user** table (bottom left):
 - Columns: user_id, email, name, role, bookmarked_papers, created_at, updated_at.
 - Domain types: BIGINT, VARCHAR(100), VARCHAR(100), VARCHAR(100), JSONB, TIMESTAMP, TIMESTAMP.
- search_keyword** table (bottom right):
 - Columns: keyword_id, user_id, keyword, created_at.
 - Domain types: BIGINT, BIGINT, TEXT, TIMESTAMP.

Relationships:

- A solid blue arrow points from the **user** table's **user_id** column to the **auth** table's **user_id** column.
- A dashed red arrow points from the **user** table's **user_id** column to the **search_keyword** table's **user_id** column.
- A solid blue arrow points from the **paper** table's **reference_papers** column to the **search_keyword** table's **keyword_id** column.

개발 현황 핵심 기능

Frontend

- Vue.js 프레임워크를 활용하여 직관적인 UI 구현 완료
- Backend와 API 연결 및 기능 연결 완료

Backend

- FastAPI를 사용하여 API 프로토콜을 기반으로 정의된 API 구현 완료
- 인공지능 모델 프로세스 구현 완료

AI 모델 (베이스라인)

- OpenAI API 및 프롬프트 엔지니어링을 활용한 논문 특화 키워드 추천 기능 구현 완료
- Faiss 벡터 DB를 통한 논문 검색 기능 구현 완료
- Llama 모델을 활용하여 논문의 키워드, 핵심 방법론, 주요 기술을 추출하도록 구현 완료
- Faiss 벡터 DB를 활용한 유사도 계산을 통해 관련 선행 논문 추천 기능 구현 완료

03

기대효과 및 계획

기대효과 기대효과 및 계획

수고 절감, 시간 단축

논문 탐색 시간을 단축하여
사용자 업무의 효율성이 향상됨



정확도 높은 탐색 지원

효과적인 키워드 추천을 통해
보다 정확하고 신뢰성 있는
검색 결과 제공



추가 사전 지식 지원

선행 논문 추천을 통해 사용자가
쉽게 접근 가능한 수준의
논문까지 깊이 있게 탐색 가능



향후 계획 기대효과 및 계획

< 개발 현황 요약 >

 AWS 아키텍처 구축

 초기 데이터 구축

 프론트엔드 기본 UI

 백엔드 - 프론트엔드 API 연동

 논문 관련 기능 - 웹 연동

 소셜 로그인 및 북마크 기능

< 향후 발전 계획 >

● (이슈 발생시 보완 예정)

● 데이터 확장 (약 7600개 > 25500개 논문으로 확장 예정)

● 사용자 편의적 UI 개선

● 예외처리 및 안정성 강화

● 각 기능에 대한 성능 수치화 및 성능 강화

● 추가 개발

03- III

WBS 기대효과 및 계획

※ WBS 별첨

단계 구분	주요 업무	세부 업무	작업자	상태	진척율	시작일	종료일	작업기간	M	Nov	W	IW	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	Etc	Dec	1W	2W	3
									D	O																											
1. 분석, 설계	1.1 기획																																				
		WBS 작성	서연정	진행	■■■■■	2024-11-11	2024-11-15	4																													
		프로젝트 기획서 작성	진윤희	진행	■■■■■	2024-11-11	2024-11-15	4																													
		요구사항 정의서 작성	박금택	진행	■■■■■	2024-11-11	2024-11-15	4																													
		회원 설계서 작성	서연정	진행	■■■■■	2024-11-11	2024-11-15	4																													
	1.2 기술 습득 조사																																				
		기능별 구현 가능성 검토	이준경	진행	■■■■■	2024-11-17	2024-11-20	3																													
		시스템 아키텍처 설계	이준경	진행	■■■■■	2024-11-14	2024-11-17	3																													
	1.3 데이터 수집 방법 조사																																				
		데이터 수집 가능성 검토	이준경	진행	■■■■■	2024-11-11	2024-11-24	13																													
		ERD 작성	서연정	진행	■■■■■	2024-11-11	2024-11-16	5																													
		수집 데이터 정의서 작성	서연정	진행	■■■■■	2024-11-23	2024-11-24	1																													
			서연정	진행	■■■■■	2024-11-25	2024-11-27	2																													
2. 구현	3.1 아키텍처 구축																																				
		AWS 워 배포	이준경	진행	■■■■■	2024-11-18	2024-11-25	7																													
		CI/CD 구축	이준경, 진윤희	진행	■■■■■	2024-11-23	2024-11-25	2																													
	3.1 데이터 수집																																				
		데이터 수집	서연정	진행	■■■■■	2024-11-17	2024-11-21	4																													
		데이터베이스 구축	서연정	진행	■■■■■	2024-11-21	2024-11-22	1																													
	3.2 웹 개발																																				
		프론트엔드 구현	진윤희	진행	■■■■■	2024-11-13	2024-11-26	13																													
		API 정의서 작성	서연정, 이준경	진행	■■■■■	2024-11-19	2024-11-23	4																													
		백엔드 구현	서연정, 이준경	진행	■■■■■	2024-11-20	2024-11-26	6																													
	3.3 기능 개발																																				
		키워드 히든화 시스템 구현	서연정	진행	■■■■■	2024-11-22	2024-11-27	5																													
		키워드 기반 문서 검색 시스템 구현	서연정	진행	■■■■■	2024-11-22	2024-11-27	5																													
		논문 요약 시스템 구현	박금택, 진윤희	진행	■■■■■	2024-11-22	2024-11-27	5																													
		원파의 연동	서연정	진행	■■■■■	2024-11-28	2024-11-29	1																													
4. 검수	4.1 내부검수																																				
		코드 리뷰 및 테스트	진윤희	진행	■■■■■	2024-11-29	2024-12-02	3																													
		테스트 시나리오 작성	서연정	진행	■■■■■	2024-11-29	2024-12-02	3																													
		테스트 (with trouble shooting)	서연정	진행	■■■■■	2024-11-29	2024-12-02	3																													
5. 완료	5.1 문서작성																																				
		것허브 리드미 작성	진윤희	진행	■■■■■	2024-11-27	2024-12-02	5																													
		발표 문서 작성	진윤희	진행	■■■■■	2024-11-27	2024-11-28	1																													
	5.2 발표	발표	진윤희	진행	■■■■■	2024-12-03	2024-12-03	0																													

03- III

WBS 기대효과 및 계획

※ WBS 별첨

03- IV

팀원 소개 기대효과 및 계획



서민정 (프로젝트 리더)
데이터, 인공지능



박규택
인공지능



이준경
백엔드, 아키텍처



진윤화
프론트엔드

Documento, AI기반 빠르고 쉬운 논문 탐색 서비스

SK Networks Family AI Camp 3기

TEAM 1 / DOCUMENTO

