

# VLM 등장 및 기술 동향 : LLM에서 산업 응용까지

□ 이건희, 정찬성, 윤동식, 이예진, 이준영, 이하은 / HDC 랩스

## 요약

본 연구는 Vision-Language Model(VLM)의 기술적 배경과 산업적 응용 동향을 종합적으로 고찰한다. 최근 대규모 언어모델(LLM, Large Language Model)의 자연어 처리 전반에서 새로운 패러다임을 제시하였으며, 이에 따라 시각·언어 정보를 통합적으로 처리하는 멀티모달(multimodal) 모델의 필요성이 부각되었다. VLM은 이미지와 텍스트를 공동 표현 공간에서 학습·추론함으로써, 단순한 인식 수준을 넘어 고차원적 인지 과업을 가능하게 한다. 본고에서는 LLM의 발전과 함께, VLM 연구의 핵심 기법 및 아키텍처의 진화를 살펴보고, 콘텐츠 제작, 제조 현장, 보안 감시, 방송·미디어 산업 등에서의 응용 사례를 분석하였다. 이를 통해 VLM이 학문적 의의뿐 아니라 경제·사회적으로도 큰 파급력을 지니고 있음을 확인하였다.

## I. 서론

최근 LLM의 발전은 자연어 처리(NLP) 분야 전반에서 획기적 성과를 도출하였다. 트랜스포머(Transformer)[1] 기반 아키텍처와 대규모 사전학습(pre-training) 기법의 결합은 기계 번역, 질의응답, 요약, 코드 생성 등 광범위한 과업에서 인간에 근접한 성능을 보이며 인공지능 연구의 새로운 패러다임을 정립하였다. 그러나 실제 세계의 정보는 텍스트에 국한되지 않고 시각·청각·센서 기반 신호 등의 멀티모달 형태로 존재한다. 이에 따라 언어 이외 모달리티와 언어 표현을 통합적으로 이해할 수 있는 지능의 필요성이 증대되었다.

VLM은 이러한 요구에 부응하여 시각 정보와 언어 정보를 공동 표현 공간(joint representation space)에서 학습·추론하도록 고안된 모델이다. VLM은 단순한 이미지 인식 수준을 넘어 복잡한 시각 장면에 대한 의미적 기술, 맥락 기반 질의응답, 텍스트-이미지 생성과 같은 고차원 인지 과업을 가능하게 한다. 이는 인간의 인지 메커니즘과 유사하게 다중 정보를 통합 처리한다는 점에서 인공지능 진화의 중요한 전환점으로 평가된다.

현재 VLM 연구는 대규모 이미지-텍스트 페어 구축, 대조(contrastive) 및 생성형(generative) 사전학습 전략, 인스트럭션 투닝(instruction tuning)과 강화학습을 활용한 정교한 모델 조정 기법 등으로 확장되고 있다. 동시에 의

료 영상 분석, 보안 감시, 전자상거래, 창의적 콘텐츠 생성 등 다양한 산업 영역에서 응용이 진전되고 있어 학문적 의의뿐 아니라 경제·사회적 파급력 또한 크다. 본고는 VLM 기술의 발전 배경과 주요 연구 동향을 개관하고, 산업 현장의 응용 사례를 중심으로 논의한다.

## II. VLM의 기술적 배경

### 1. LLM 기본 개념

대규모 언어모델, LLM은 방대한 텍스트 데이터로 학습된 딥러닝 아키텍처를 활용하여 인간과 유사한 수준으로 텍스트를 이해·생성하는 인공지능 시스템이다. 트랜스포머 구조가 언어모델의 사실상 표준으로 자리잡은 이후, 순환신경망(RNN, Recurrent Neural Network) 계열의 순차 연산 한계를 벗어나 대규모 병렬 처리를 통한 언어 학습이 가능해졌다. 트랜스포머는 셀프-어텐션(self-attention)만으로 문맥 의존성을 학습함으로써 데이터 및 파라미터(parameter)의 대규모 확장을 현실화하여 LLM 시대의 토대를 마련하였다.

사전학습(pre-training) 패러다임이 정착하며 BERT(Bidirectional Encoder Representations from Transformers)[2]는 양방향 인코딩(bidirectional Encoding)과 마스킹(masking) 기반 학습으로 범용 언어표현을 길러 소량의 파인튜닝(fine-tuning)만으로도 다양한 과제를 해결하는 방법론을 확립하였다. 반면 GPT(Generative Pre-trained Transformer)[3] 계열은 자가회귀(autoregressive) 방식의 다음 토큰 예측을 통해 자연스러운 생성 능력을 강화하여 생성형 언어모델의 계보를 열었다.

모델 규모를 175B로 확장한 OpenAI의 GPT-3는 별도 가중치 업데이트 없이 프롬프트(prompt)만으로 다양한 작업을 수행하는 제로샷(zero-shot)·퓨샷(few-shot)을 일 반화하여 대중적으로 각인시켰다. 이후 GPT-3.5는 인간

피드백 기반 강화학습(RLHF: Reinforcement Learning with Human Feedback)을 도입하여 응답의 유용성과 안전성을 향상시켰다. 나아가 GPT-4는 텍스트와 이미지를 함께 처리하는 멀티모달 입력을 지원함으로써 언어적 추론을 시각 이해로 확장하였다. 한편 메타(Meta)社의 LLaMA 시리즈, 구글(Google)社의 Gemini·Gemma, Anthropic의 Claude, Alibaba Cloud의 Qwen 등 주요 기업들이 LLM의 고도화 및 상용화를 둘러싸고 경쟁하고 있다[4].

사전학습된 LLM은 광범위한 언어 작업에서 뛰어난 성과를 보였으나, 텍스트만을 학습·처리하는 단일 모델 아키텍처는 서로 다른 양식 간 복잡한 관계를 이해해야 하는 현실 데이터에서 구조적 한계를 드러낸다. 이러한 한계는 시각적 데이터와 텍스트 데이터를 결합하여 공간 관계, 대상, 장면 및 추상 개념에 대한 포괄적 이해를 제공하는 VLM에 대한 요구로 이어졌다. VLM은 대량의 이미지-텍스트 데이터를 학습해 시각과 언어 정보를 동시에 처리할 수 있으며, 원시 시각 신호와 언어 표현을 공동 학습함으로써 이미지 캡셔닝(Captioning) 과정에서의 정보 손실을 보완하고 시각적 질의응답 및 자율주행 등 상황 인지적 판단을 가능하게 한다.

### 2. 멀티모달 학습의 기본 개념과 최신 VLM 연구

2020년대 초반, 원시 텍스트 기반의 대규모 사전학습은 NLP 분야의 혁신을 견인하였다. GPT, BERT, T5 등 의 등장은 방대한 웹 텍스트가 고품질 레이블 데이터셋을 대체하는 강력한 감독(supervision) 신호가 될 수 있음을 보여주었다. 반면 컴퓨터 비전 분야는 오랫동안 이미지넷(Imagenet) 등 제한적 수작업 레이블 데이터에 의존해 왔다. 이에 “웹에서 무한히 얻을 수 있는 텍스트를 비전 학습의 감독 신호로 활용할 수 있는가”라는 문제의식이 대두 되었다. 이미지는 시각적 풍부함을 제공하고 언어는 추상적 개념과 맥락을 표현하는데 이미지, 언어를 결합해 학습

하면 보다 일반화 가능하고 유연한 표현 학습이 가능하다.

Contrastive Language-Image Pre-training(CLIP)[5]은 이러한 배경에서 등장하였다. CLIP은 수억 개 이미지-텍스트 쌍을 웹에서 수집하여 대규모 contrastive 학습을 수행함으로써 지도학습 없이도 제로샷 전이가 가능한 새로운 패러다임을 제시하였다. 멀티모달 학습은 서로 다른 형태의 데이터를 동시에 이해·표현하기 위한 접근이며, 교차 모달 표현 학습(cross-modal representation learning) 기법은 두 모달리티를 동일 임베딩 공간에 정렬(alignment)하는 것을 목표로 한다. 이로써 “고양이”라는 단어와 고양이 이미지를 동일 개념으로 연계하거나, 이미지를 입력받아 관련 텍스트를 검색하는 것이 가능해진다.

사전학습 기법은 크게 contrastive, generative 두 방식으로 나눌 수 있는데 첫 번째 contrastive 방식은 이미지와 텍스트의 올바른 쌍을 맞추는 데 초점을 둔다. 예로 CLIP과 ALIGN[6] 기법은 이미지와 설명 문장을 서로 끌어당기고 다른 쌍과는 멀어지도록 학습하여 강력한 제로샷 성능을 보였다. 둘째, generative 방식은 언어 생성 과정을 통해 시각 정보를 반영한다. Flamingo[7], BLIP-2[8] 기법 등은 비전 인코더를 통해 시각 정보를 추출한 뒤 언어모델과 결합하여 자연스러운 텍스트 생성을 구현한다.

아키텍처 관점에서도 두 흐름이 존재한다. 초기 모델은 이미지 인코더와 텍스트 인코더를 분리하는 dual-encoder 구조를 채택했으나, 최근에는 시각·언어 정보를 보다 긴밀히 융합하는 unified encoder-decoder 구조가 주목받고 있다. 이는 단순 매칭을 넘어 복잡한 문맥과 추론을 다루는 토대를 제공한다. 대화형 멀티모달 모델도 주목된다. BLIP-2는 비전 인코더와 대규모 언어모델 사이를 연결하기 위해 Q-Former를 도입하여 시각 정보를 효율적으로 언어 표현 공간으로 투영하였다. LLaVA[9]는 해당 방식을 계승하면서 인스트럭션 투닝 데이터를 활용해 시각 맥락을 반영한 대화 능력을 확보하였다. 이러한 접근은 “이미지를 설명”하는 수준을 넘어, 이미지를 기반으로 추론하며 자연스럽게 질의응답을 수행하는 인터페イス를 제공한다. 이 흐름의 정점으로 GPT-4V[10]와 같

은 범용 멀티모달 모델이 제시되며, 단일 이미지 해석을 넘어 복잡한 시각 정보와 언어적 맥락을 종합한 고차원 추론을 수행한다.

### III. 산업 응용 현황

#### 1. 콘텐츠 및 디자인 분야 응용

##### 1) 이미지 및 비디오 요약

VLM의 가장 직관적 응용 중 하나는 시각 콘텐츠에 대한 자연어 설명 생성이다. 단순 객체 인식을 넘어 이미지의 맥락, 감정, 스타일적 특징을 반영한 캡션 생성이 가능하며, 도표·그래프 해석과 이미지 내 텍스트 판독(OCR 유사 기능)도 지원한다. 특히 비디오 요약 분야에서 스트리밍 업계 전반의 활용도가 급증하고 있다. VLM 기반 시스템은 긴 영상에서 핵심 장면을 추출하고 시청자 맞춤형 하이라이트를 생성한다. 교육 콘텐츠·강의 영상의 경우, 중요한 개념이 등장하는 시점을 자동 태깅(tagging)하여 학습 효율을 높인다.

##### 2) 사용자 만족도 기반 평가 기법

창작 산업에서 AI 도구 활용이 증가함에 따라 기술적 품질과 실제 사용자 만족도 간 괴리가 문제로 대두되고 있다. 전통적 이미지 스타일 전송(style transfer)은 ‘원본 내용(인물·구도 등)을 보존하면서 별도 참조 스타일(색채 팔레트, 질감, 선형 과장 등)을 입히는 기법’을 의미한다. 최근 널리 공유된 ‘지브리풍(Ghibli-style)’ 변환 사례처럼, 사진의 내용은 유지하되 특정 스튜디오 미학을 덧입히는 방식이 전형적이다. 기존 평가는 콘텐츠 보존과 스타일 유사성을 별도 측정하였으나 사용자 만족도를 충분히 반영하지 못했다. 이를 보완하기 위해 VLM 기반으로 의미론적 콘텐츠와 예술적 스타일을 통합 고려하여 만족도를 예측하는 접근이 제안되고 있다.

### 3) 자동화된 데이터 생성 활용

산업에서 고품질 데이터셋 확보는 시스템 성능과 직결되는 핵심 과제이다. 특수 도메인·희소 시점·특정 객체 카테고리의 데이터는 수집 비용이 높다. 이를 해결하기 위해 학습 모델 기반 도메인 특화 데이터 생성이 제안되었고, 합성 데이터의 품질 검증을 위해 VLM 기반 객체 검출·미학적 품질 평가·비전·언어 정렬을 통합한 다단계 검증 프로세스가 활용된다. 또한 사용자 선호도 분류기를 통해 주관적 선택 기준을 반영하여 최종 데이터셋을 구성할 수 있다. 이러한 기법은 실제 수집이 어려운 특수 환경 데이터 합성에 효과적임이 확인되고 있다.

## 2. 보안·안전 분야 응용

### 1) 제조 현장의 결함 검출

제조업에서 VLM은 품질 관리의 새로운 패러다임을 제시한다. 규칙 기반 검사 시스템과 달리 VLM은 다양한 각도와 조명 조건의 제품 이미지를 종합 분석하여 미세 결함 까지 감지할 수 있다. 반도체, 자동차, 전자부품 제조 현장에서는 VLM 기반 검사 시스템을 통해 검사 시간을 단축하면서도 검출 정확도를 향상시키는 연구가 활발하다. 복잡한 조립품이나 다양한 표면 처리를 요구하는 제품의 경우 기존 컴퓨터 비전 기술의 한계를 보완하는 대안이 된

다. 또한 텍스트 기반 품질 기준과 시각 정보를 결합함으로써 보다 정교한 품질 판정이 가능하다.

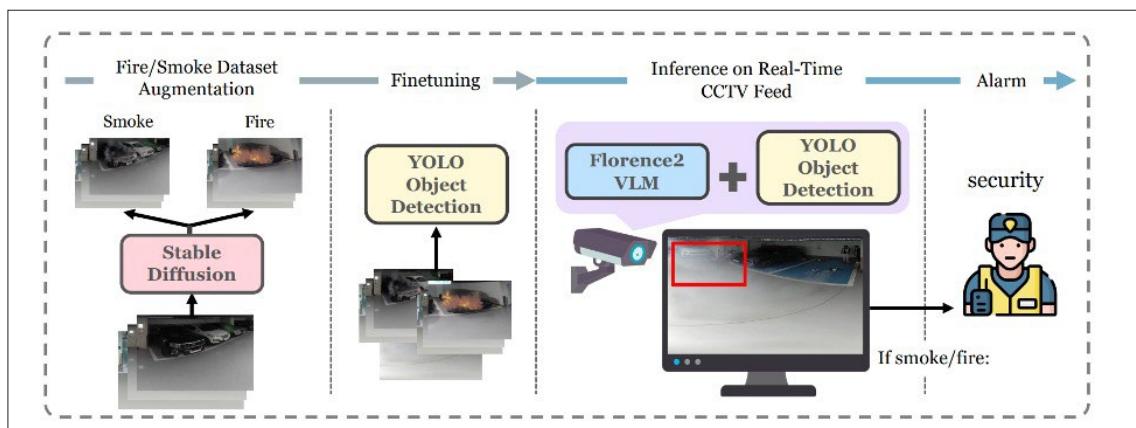
### 2) 지능형 감시 및 보안 시스템

보안 분야에서 VLM은 단순 움직임 감지를 넘어 상황 맥락을 이해하는 지능형 감시를 구현한다. AI CCTV 시스템은 이상 행동 패턴을 실시간 분석하여 절도, 폭력, 화재 등 위험 상황을 조기에 감지한다. 공항, 지하철역, 대형마트 등에서 시범 운영되는 VLM 기반 보안 시스템은 방대한 영상 데이터 중 실제 위험 상황만 선별하여 보안 요원에게 경보를 전송한다. 특히 화재 감지 분야에서, 기존 연기 감지기의 한계를 보완하는 비전 기반 화재 감지 시스템이 주목받고 있다. VLM은 연기의 색상, 확산 패턴, 주변 환경을 종합 분석하여 화재 초기 단계에서도 높은 정확도로 감지할 수 있다[11]. 야외나 대형 공간에서도 센서 방식 대비 빠르고 정확한 성능을 보이는 것으로 보고된다.

## 3. 검색 및 추천

### 1) 방송·미디어 산업에서의 RAG 필요성

RAG(Retrieval-Augmented Generation)는 대규모 모델의 답변 정확성과 해석 가능성을 높이기 위해 외부의 자료를 검색하여 답변에 활용하는 AI 기술이다. 외부 데이



<그림 1> 밀폐된 공간의 실시간 CCTV 영상에서 차량 화재/연기 감지를 위해 설계된 VLM 기반의 화재 감지 프레임워크 예시

데이터베이스에서 정보를 검색하여 답변을 생성하므로, 환각 현상을 줄이고 사용자의 질문에 정확하고 최신 정보까지 활용하여 유연하게 답변을 제공할 수 있다[12]. 매번 500 시간 이상의 동영상이 업로드 되는 방송 산업 환경에서, 이러한 방대한 데이터를 활용할 수 있는 RAG는 차별화된 서비스를 제공하기 적합한 기술이다.

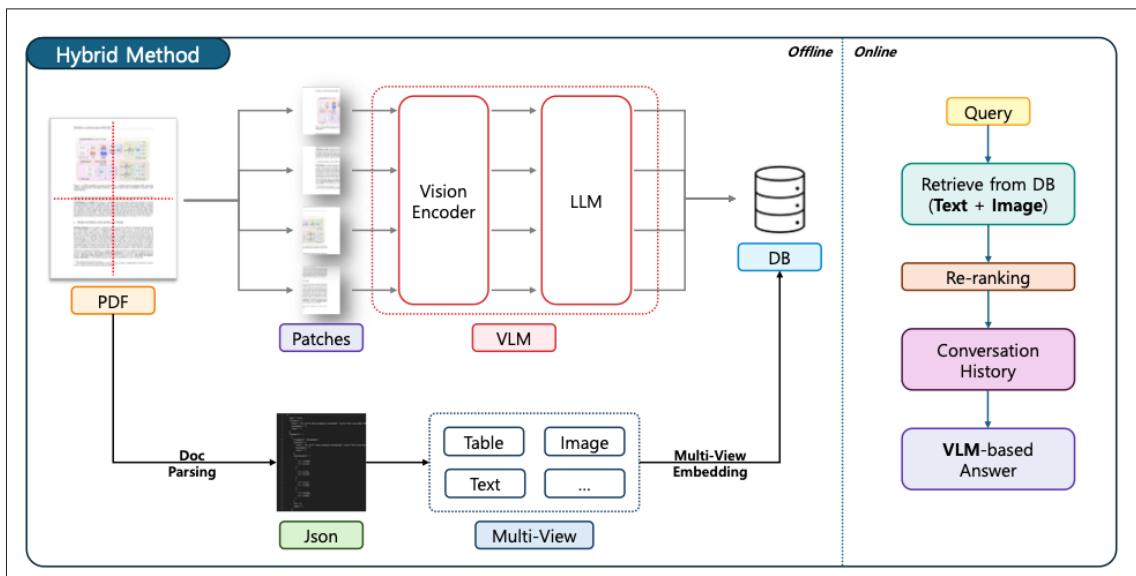
## 2) 기존 RAG 시스템의 한계와 새로운 접근법

기존의 RAG 파이프라인은 OCR, 레이아웃 감지, 청킹, 텍스트 임베딩 등 복잡한 단계를 거쳐 데이터를 데이터베이스에 저장하여 활용한다. 해당 방식은 텍스트 중심의 콘텐츠에는 유리하지만, 이미지나 영상 비중이 높을수록 정보 손실과 처리 복잡성이 증가한다. 이를 보완하기 위한 대안으로, VLM을 활용하여 이미지를 직접 임베딩함으로써 검색 과정을 단순화한 방식이 주목받고 있다[13]. 이는 문서 페이지를 이미지 페치 단위로 임베딩하고, 사용자의 질문은 토큰 단위로 임베딩한 뒤, 각 질문 토큰을 해당 페이지의 모든 페치 임베딩과 비교하여 가장 유사한 페치 하나만을 대표로 선택하고 토큰별 대표 유사도를 합산해 페이지 우선순위 점수를 매기는 late interaction 방

식을 취하고 있다[14]. 시각적 요소를 그대로 활용하기 때문에 인덱싱 구조가 단순해지고, 다양한 레이아웃이 혼재된 자료에서도 원활한 검색이 가능하다. 아울러 자막 등 텍스트 비중이 높은 영상 콘텐츠에서는 〈그림 2〉에 제시한 바와 같이 텍스트 임베딩과 이미지 임베딩을 상호보완적으로 결합한 하이브리드 구성이 최적의 검색 성능을 보일 수 있다.

## 3) 실제 산업 현황 및 응용 시나리오

방송·미디어 현장은 지능형 광고, 생성형 크리에이티브, 접근성 서비스에서 이미 컴퓨터비전·그래픽스·생성형 AI가 활발히 쓰이고 있으며, 여기에 RAG를 결합하면 품질과 준법, 운영 효율을 동시에 끌어올릴 수 있다. 예를 들어 장면 삽입형 가상 광고(VPP)는 본편 맥락을 보존한 채 노출을 최적화하지만, RAG가 브랜드 가이드·법규·금지어·과거 집행 사례를 검색해 근거 기반의 삽입 전후 검증(brand safety)과 유사 장면 검색을 지원하면 운영 신뢰도가 높아진다. 생성형 크리에이티브의 경우, 이미지·영상 생성 자체는 생성형 AI가 담당하되 RAG가 최신 캐릭터로그, 상품 스펙, 스타일 가이드, 법적 고지 사항을 주입해



〈그림 2〉 이미지-텍스트 하이브리드 임베딩 기반 검색 구조 예시

프롬프트와 결과물을 정합·준법·일관성 측면에서 자동 점검·보강할 수 있다. 접근성 측면에서도 실시간 수어 아바타에 RAG를 결합하면 지명·전문용어·절차 매뉴얼과 과거 방송 코퍼스를 조회해 표준 표현을 제시함으로써 용어 정확도와 문맥 일치를 개선할 수 있다. 특히 방송 데이터는 자막·대본 같은 텍스트와 그래픽·표·장면 전환 같은 시각 신호가 공존하므로, 텍스트 기반 점수와 이미지 패치 기반 점수를 결합하는 하이브리드 RAG를 채택하면 콘텐츠 유형과 맥락에 따라 두 신호가 상호보완적으로 작동해 전반적 성능을 안정적으로 끌어올릴 수 있다.

## IV. 결론 및 미래 전망

본 연구에서는 VLM 기술의 발전 배경과 최신 연구 동향을 정리하고, 다양한 산업 응용 사례를 분석하였다. VLM

은 언어와 시각 정보를 동시에 처리할 수 있는 능력을 바탕으로, 콘텐츠 요약·생성, 사용자 만족도 예측, 제조업 결함 검출, 지능형 감시, 방송 미디어 검색 등 다각적 활용 가능성을 보여주고 있다. 이는 기존 단일 모달 기반 AI의 한계를 극복하고, 인간과 유사한 인지적 추론을 수행하는 지능형 시스템으로 발전하는 중요한 이정표라 할 수 있다.

향후 연구는 대규모 멀티모달 데이터셋의 확보와 품질 관리, 모델 경량화와 실시간 처리 가능성, 프라이버시·윤리 문제에 대한 대응을 중심으로 전개될 것이다. 또한 VLM을 기반으로 한 가상광고 및 미디어 제작 분야에서는 실시간 상호작용과 사용자 맞춤형 서비스가 더욱 정교화될 것으로 예상된다. 나아가 하드웨어 및 알고리즘적 최적화의 진전과 더불어, VLM의 산업적 적용 범위는 의료, 교육, 국방, 문화예술 등 사회 전반으로 확산될 것으로 전망된다.

### 참 고 문 헌

- [1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” NIPS’17: Proceedings of the 31st International Conference on Neural Information Processing Systems (2017), pp. 6000 - 6010. doi: <https://doi.org/10.48550/arXiv.1706.03762> (accessed Dec, 27, 2024)
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies 1 (2019), pp. 4171-4186. doi: <https://doi.org/10.18653/v1/N19-1423> (accessed Dec, 27, 2024)
- [3] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I., Improving Language Understanding by Generative Pre-Training, OpenAI Technical Report, 2018.
- [4] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., ... & Lample, G. (2023). Llama: Open and efficient foundation language models, arXiv preprint arXiv:2302.13971.
- [5] Radford, Alec, et al. “Learning transferable visual models from natural language supervision.” International conference on machine learning. PMLR, 2021.
- [6] Jia, Chao, et al. “Scaling up visual and vision-language representation learning with noisy text supervision.” International conference on machine learning. PMLR, 2021.
- [7] Alayrac, Jean-Baptiste, et al. “Flamingo: a visual language model for few-shot learning.” Advances in neural information processing systems 35 (2022): 23716-23736.
- [8] Li, Junnan, et al. “Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models.” International

## 참 고 문 헌

conference on machine learning, PMLR, 2023.

[9] Liu, Haotian, et al. “Visual instruction tuning.” Advances in neural information processing systems 36 (2023): 34892-34916.

[10] OpenAI, “GPT-4V(json) system card,” OpenAI, Sep. 2023. [Online]. Available: <https://openai.com/research/gpt-4v-system-card>

[11] Kim, Joanne, et al. “An Integrated YOLO and VLM System for Fire Detection in Enclosed Environments.” I Can’t Believe It’s Not Better: Challenges in Applied Deep Learning.

[12] Lewis, Patrick, et al. “Retrieval-augmented generation for knowledge-intensive nlp tasks.” Advances in neural information processing systems 33 (2020): 9459-9474.

[13] Faysse, Manuel, et al. “Colpal: Efficient document retrieval with vision language models.” arXiv preprint arXiv:2407.01449 (2024).

[14] Khattab, Omar, and Matei Zaharia. “Colbert: Efficient and effective passage search via contextualized late interaction over bert.” Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, 2020.

## 저 자 소 개



### 이 건 희

- 2010년 : 경희대학교 전자공학 석사
- 2010년 ~ 2019년 : 삼성전자, SR(Samsung Reaserch) AI 센터
- 2019년 ~ 2024년 : 삼성전자, MX사업부 Immersive S/W개발그룹
- 2024년 ~ 현재 : HDC 랩스 기술혁신연구소 AI LAB 장
- 주관심분야 : 컴퓨터 비전, 자연어처리, 데이터사이언스



### 정 찬 성

- 2016년 : 한양대학교 지능형로봇학과 석사
- 2017년 ~ 2020년 : 한국전자기술연구원, 지능로보틱스연구센터
- 2020년 ~ 2024년 : 두산 디지털 이노베이션, AI전략팀
- 2024년 ~ 현재 : HDC 랩스 기술혁신연구소 AI LAB
- 주관심분야 : 컴퓨터 비전, 이미지 처리, VLM



### 윤 동 식

- 2021년 : 백석대학교 ICT학부 멀티미디어학과 학사
- 2023년 : 고려대학교 전기전자공학과 석사
- 2023년 : 메타버스엔터테인먼트 TD실 TD팀 AI파트
- 2024년 ~ 현재 : HDC 랩스 기술연구소 AI LAB
- 주관심분야 : 컴퓨터 비전, 이미지 처리

## 저자 소개

### 이예진



- 2022년 : 아주대학교 전자공학과 학사
- 2024년 : 한국과학기술원 전기및전자공학부 석사
- 2024년 ~ 현재 : HDC 랩스 기술연구소 AI LAB
- 주관심분야 : 영상처리

### 이준영



- 2021년 : 고려대학교 바이오의공학부 학사
- 2024년 : 고려대학교 인공지능학과 석사
- 2025년 ~ 현재 : HDC 랩스 기술연구소 AI LAB
- 주관심분야 : 멀티모달, RAG, VLM

### 이하은



- 2023년 : 성균관대학교 소프트웨어학과 학사
- 2025년 : 성균관대학교 소프트웨어학과 석사
- 2025년 ~ 현재 : HDC 랩스 기술연구소 AI LAB
- 주관심분야 : 자연어처리, 정보 검색, 대화 시스템