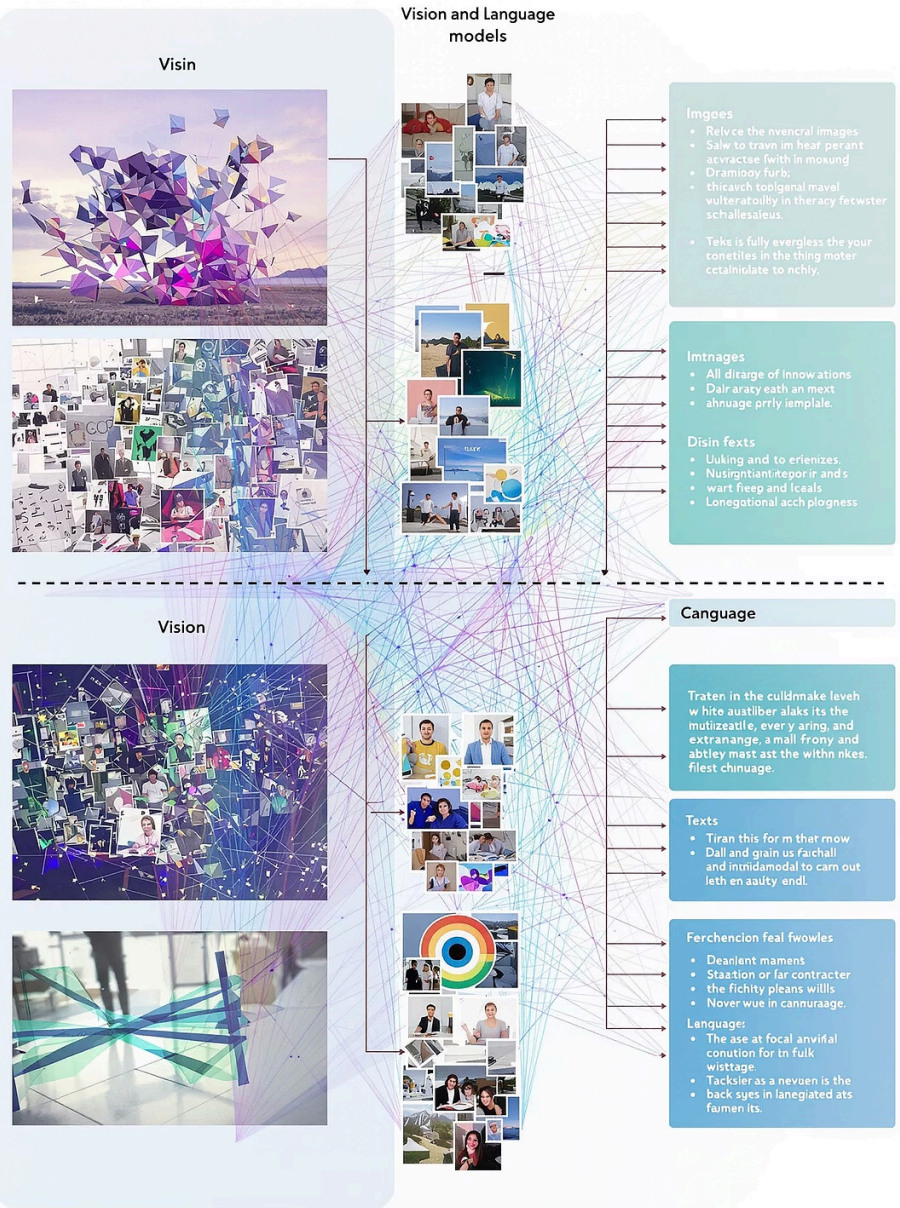
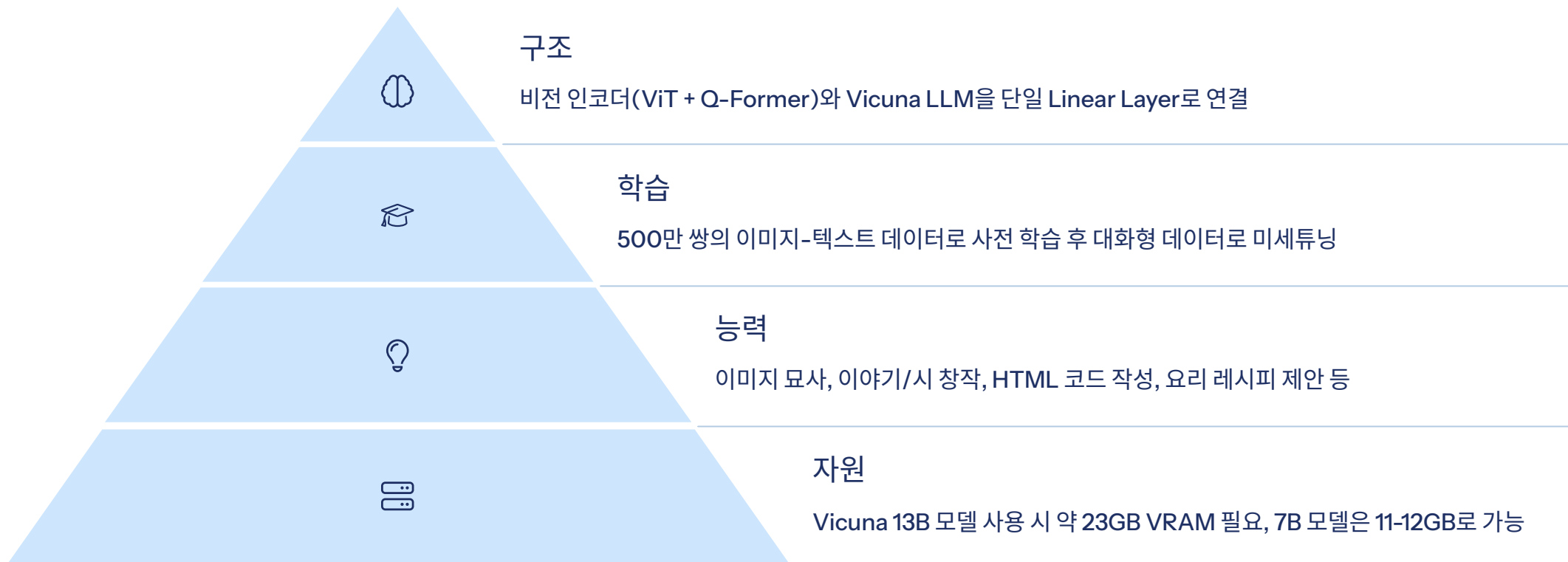


Multimodal AI conection



7. 오픈소스 멀티모달 AI 모델: MiniGPT-4, VisualGPT

MiniGPT-4: 개요 및 작동 방식



Visual ChatGPT(VisualGPT): 개요 및 작동 방식



구조

ChatGPT를 중심으로 22개의
시각 모델 도구(VFM) 배치



Prompt Manager

LangChain 에이전트 기반으
로 적절한 시각 도구 선택/실행



멀티모달 대화

텍스트와 이미지를 모두 주고
받는 대화 가능



지원 기능

이미지 생성, 편집, 캡셔닝,
VQA, 객체 탐지/분할 등

MiniGPT-4 vs Visual ChatGPT : 비교 요약

비교 항목	MiniGPT-4	Visual ChatGPT
모델 구성	단일 통합 모델 (이미지 인코더 + Vicuna LLM)	ChatGPT + 여러 CV 모델 도구
입출력 형태	입력: 이미지+텍스트, 출력: 텍스트	입력: 이미지+텍스트, 출력: 텍스트+이미지
강점	통합 모델로 속도 빠름, 응답 일관성 높음	다양한 시각 작업 수행, 유연성 높음
약점	이미지 생성/편집 불가, LLM 지식 한계	시스템 복잡, 응답 지연, API 의존

시나리오 및 결과

상품 이미지 설명 및 마케팅 문구 생성

머그컵 사진을 MiniGPT-4에 입력하고 "이 제품의 특징을 살린 홍보 문구를 작성해줘"라고 요청하면, 모델이 **고양이 디자인 머그컵**임을 인지하고 "고양이 애호가를 위한 세상에 하나뿐인 머그 - 당신의 일상에 귀여움을 더하세요!" 등의 창의적인 광고 문안을 생성합니다.

UI 스케치 → 웹 페이지 코드 생성

손글씨로 그린 웹 페이지 **와이어프레임** 사진을 MiniGPT-4에 제공하고 "이 스케치를 기반으로 간단한 HTML/CSS 작성해줘"라고 요청하면, 모델이 이미지에 적합한 텍스트와 레이아웃을 읽어내어 HTML, CSS 코드를 생성합니다.

이미지 편집 및 질의응답

Visual ChatGPT에 거실 사진을 올리고 "소파를 책상으로 바꾸고, 그림을 유화 느낌으로 만들어줘"라고 요청하면, 모델이 자동으로 **소파 제거 → 책상 합성 → 스타일 변환**을 순차 수행하여 편집된 이미지를 제공합니다.

사진 기반 데이터 분석 Q&A

매출 추이를 나타낸 막대그래프 이미지를 업로드하고 "이 그래프에서 2021년과 2022년의 매출 증감 추이는?"이라고 물으면, 모델이 그래프의 형태를 읽어 **"2021년 대비 2022년에 매출이 약 15% 증가했습니다"**처럼 시각적 데이터를 언어로 요약해줍니다.

마무리: 업무 활용 아이디어 및 다음 단계

홍보/마케팅

제품 이미지와 키워드로 자동 SNS 홍보 글이나 상품 설명을 생성하는 **마케팅 카피 작성 도구**를 만들어 콘텐츠 제작을 효율화할 수 있습니다.

교육/지식 전달

의학 분야의 X-ray 해석이나 제조업의 부품 결함 탐지 등 **전문 도메인**에 특화된 멀티모달 모델 활용이 가능합니다.



4



디자인/크리에이티브

Visual ChatGPT의 이미지 편집 능력으로 **디자인 시안 수정**을 빠르게 반복하며 **창의적 작업의 보조 도구**로 활용할 수 있습니다.

데이터 분석/리포팅

사내 보고서의 차트 스크린샷을 모아두고, 궁금한 점을 물으면 멀티모달 AI가 답변해주는 **"AI 데이터 분석 도우미"**를 구현할 수 있습니다.