



구글 Gemini와 LLM의 세계



코딩목사 닥터윌(Dr Will)_양진욱

구글 Gemini 소개

1 혁신적 AI 모델

Gemini는 구글의 최신 AI 기술을 집약한 모델입니다.
다양한 분야에서 뛰어난 성능을 보입니다.

2 멀티모달 능력

텍스트뿐만 아니라 이미지, 오디오, 비디오 등 다양한 형태의 입력을
처리할 수 있습니다.

3 확장성

Ultra, Pro, Nano 버전으로 다양한 환경에 적용 가능합니다.
모바일 기기부터 대규모 서버까지 활용됩니다.



LLM의 기본 원리

확률 기반 모델

LLM은 텍스트의 확률을 계산하는 대규모 모델
 $P(\text{text})$ 를 정확하게 예측하는 것이 목표입니다
.

트랜스포머 아키텍처

대부분의 LLM은 트랜스포머 구조 기반
이는 병렬 처리와 장거리 의존성 포착에
효과적입니다.

대규모 학습

방대한 양의 텍스트 데이터로 학습합니다.
이를 통해 언어의 패턴과 구조를 이해합니다.





언어 모델링의 핵심

1

텍스트 확률 학습

언어 모델링은 텍스트 문치에 있는 모든 텍스트의 출현 확률을 학습합니다. 이는 자연스러운 언어 생성의 기초가 됩니다.

2

컨텍스트 이해

이전 텍스트를 기반으로 다음 토큰의 확률을 예측합니다. 이를 통해 문맥을 이해하고 연속성 있는 텍스트를 생성합니다.

3

표현 학습

언어 모델은 텍스트의 의미적 표현을 학습합니다. 이는 다양한 자연어 처리 작업에 활용됩니다.

언어 모델 학습 과정

1

데이터 수집

대규모 텍스트 코퍼스를 수집합니다.
웹 크롤링, 디지털 도서관 등 다양한 소스를 활용합니다.

2

전처리

텍스트를 정제하고 토큰화합니다.
특수문자 처리, 단어 분리 등의 작업을 수행합니다.

3

모델 훈련

준비된 데이터로 모델을 학습시킵니다.
관측된 텍스트의 확률은 높아지고, 그렇지 않은 텍스트의 확률은 낮아집니다.

4

평가 및 최적화

모델의 성능을 평가하고 하이퍼파라미터를 조정합니다.
필요시 추가 학습을 진행합니다.



언어 모델의 다양한 응용

자연어 생성

챗봇, 자동 작문, 시나리오 생성 등에 활용됩니다.
인간과 유사한 텍스트를 생성할 수 있습니다.



기계 번역

언어 간 번역 품질을 크게 향상시켰습니다.
문맥을 고려한 자연스러운 번역이 가능해졌습니다.



텍스트 요약

긴 문서를 간결하게 요약할 수 있습니다.
핵심 내용을 추출하고 재구성합니다.



감성 분석

텍스트의 감정과 의견을 분석합니다.
고객 리뷰 분석, 여론 조사 등에 활용됩니다.



Gemini 프로젝트를 위한 필수 도구



Google Colab

클라우드 기반 Python 환경입니다. GPU 지원으로 Gemini 모델 실험에 적합합니다.



Pandas

데이터 조작과 분석을 위한 라이브러리입니다. 대규모 데이터셋 처리에 효과적입니다.



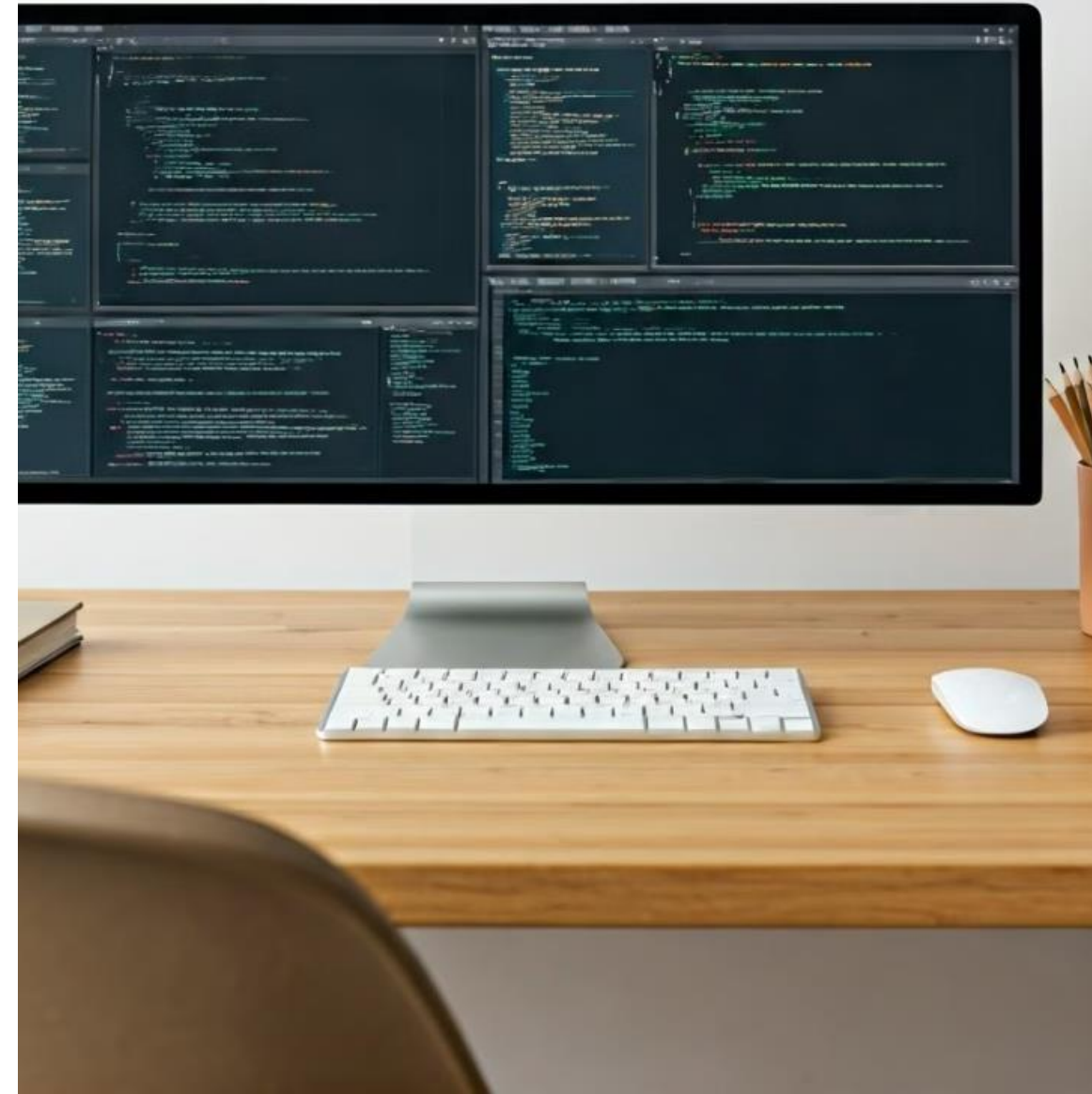
NumPy

수치 연산 라이브러리입니다. 행렬 연산과 수학적 함수를 제공합니다.



Matplotlib

데이터 시각화 라이브러리입니다. 결과를 그래프로 표현하는 데 사용됩니다.



Streamlit을 활용한 Gemini 애플리케이션 개발



특징	장점	활용 예시
간편한 UI 구축	빠른 프로토타이핑	Gemini 데모 페이지
데이터 시각화 지원	직관적인 결과 표현	LLM 성능 대시보드
실시간 업데이트	대화형 애플리케이션 구현	AI 챗봇 인터페이스



Gemini

Gemini를 위한
멀티모달 프롬프트 설계

멀티모달

프롬프트

이 사진을 바탕으로 짧고 흥미로운 블로그 게시물을 작성해 주세요. 사진에 식사에 대한 설명이 포함되어야 하고, 음식 준비에 관해 이야기해야 합니다.



응답

식사 준비는 시간과 비용을 절약할 수 있는 좋은 방법일 뿐만 아니라, 더 건강한 음식을 먹는 데도 도움이 될 수 있습니다. 이 식사는 미리 쉽게 준비할 수 있는 건강하고 맛있는 식사의 좋은 예입니다.

현미, 구운 야채, 닭고기 데리야키가 포함된 식사입니다. 현미는 건강하고 복잡한 탄수화물로, 지속적인 에너지를 공급합니다. 구운 채소는 비타민과 미네랄을 매일 섭취하는 데 좋은 방법이며 치킨 데리야키는 맛있고 단백질이 풍부한 메뉴입니다.

이 식사는 준비도 매우 쉽습니다. 현미와 채소를 구우고 치킨 데리야키를 요리하기만 하면 됩니다. 모든 것이 조리되면 식사 준비 용기로 나누어 냉장고에 보관합니다. 그런 다음 컨테이너를 들고 바쁜 아침을 출근할 수 있습니다!

미리 쉽게 준비할 수 있는 건강하고 맛있는 식사를 찾고 있다면 이 식사가 좋은 선택입니다. 영양가가 가득하고 맛이 담겨 있으므로 포만감과 만족감을 느낄 수 있습니다.

이곳에서는 건강하고 맛있는 식사를 준비할 수 있습니다!

프롬프트 설계 기초

1. 구체적으로 안내: 오해의 여지가 최소화되도록 명확하고 간결한 안내를 작성합니다.
2. 프롬프트에 몇 가지 예시 추가: 현실적인 퓨샷 예시를 사용해 달성하려는 목표를 설명하세요.
3. 단계별 세분화: 복잡한 작업을 관리 가능한 하위 목표로 나누어 프로세스 전반에 걸쳐 모델을 안내합니다.
4. 출력 형식 지정: 프롬프트에서 마크다운, JSON, HTML 등 원하는 형식으로 출력하도록 요청합니다.
5. 단일 이미지 프롬프트를 위해 이미지를 먼저 배치:

>> Gemini는 이미지와 텍스트 입력을 원하는 순서로 처리할 수 있지만

단일 이미지가 포함된 프롬프트의 경우 해당 이미지를 텍스트 프롬프트 앞에 배치하면 성능이 더 좋을 수 있습니다.



구체적으로
요청하라!

프롬프트



이미지를 설명해 주세요.

모델 응답

공항 도착 및 출발 안내판을 보여주는 이미지입니다.

업데이트된 프롬프트



이 이미지에 표시된 공항 보드의 시간과 도시를 목록으로 파싱합니다.

개선된 대응

- 10:50 모스크바
- 11:05 에든버러
- 11:05 런던
- 11:10 부카레스트
- 11:30 키예프
- 11:35 더블린
- 11:45 이스트 미들랜즈
- 12:15 소피아
- 12:30 런던
- 12:30 뉴캐슬
- 12:40 상트페테르부르크
- 12:40 런던
- 12:45 맨체스터

Multi modal Fewshot Prompt

프롬프트



명소와 함께 도시를 확인합니다.

모델 응답

이 명소는 브라질 리우데자네이루에 있는 그리스도 구속자 상입니다.

명소와 함께 도시를 확인합니다.



도시: 로마, 명소: 콜로세움.





도시: 베이징, 명소: 자금성



도시: 리우데자네이루, 명소: 구세주 그리스도 동상

단계별 분석 요청

프롬프트	모델 응답
 <p>화장실 휴지는 언제 부족해?</p>	<p>곧 남은 롤 3개뿐입니다.</p>

업데이트된 프롬프트	개선된 대응
 <p>1. 먼저 이 사진에 화장지 롤이 몇 개인지 세세요. 2. 그런 다음 보통 사람이 하루에 사용하는 화장지 양을 결정합니다. 3. 이 화장지 롤의 지속 시간을 계산하세요.</p>	<p>1. 이 사진에 화장지 롤 3개가 있습니다. 2. 일반적으로 한 사람이 하루에 20장 정도 화장지를 사용합니다. 3. 각 주사위 굴리기의 시트가 200장인 경우 각 롤은 약 10일 동안 지속됩니다. 따라서 이 3개의 롤은 약 한 달 동안 지속됩니다.</p>

출력 형식 지정

프롬프트

Subject	Status	Time	Start date	Due on
Math	Done	1.5 hrs	11/29/2023	12/1/2023
Geography	Done	30 min	12/2/2023	12/4/2023
Physics	In progress	2 hrs	12/3/2023	12/5/2023
Biology	Not started	6 hrs	11/2/2016	12/6/2023

이 이미지의 테이블을 마크다운 형식으로 파싱합니다.

모델 응답

| 제목 | 상태 | 시간 | 시작일 | 마감일 |

| ----- | ----- | --- | ----- | ----- |

| 수학 | 완료 | 1.5시간 | 2023년 11월 29일 | 2023년 12월 1일 |

| 지역 | 완료 | 30분 | 2023년 12월 2일 | 2023년 12월 4일 |

| 물리학 | 진행 중 | 2시간 | 2023년 12월 3일 | 2023년 12월 5일 |

| 생물학 | 시작하지 않음 | 6시간 | 2016년 11월 2일 | 2023년 12월 6일 |

Gemini 강점

사용 사례	설명
정보 탐색	세계 지식을 이미지 및 동영상에서 추출한 정보와 혼합합니다.
객체 인식	이미지 및 동영상에 포함된 객체에 대해 세밀하게 조정된 식별과 관련된 질문에 답변합니다.
디지털 콘텐츠 이해	질문에 답변하고 인포그래픽, 차트, 그림, 테이블, 웹페이지 등 다양한 콘텐츠로부터 정보를 추출합니다.
정형화된 콘텐츠 생성	제공된 프롬프트 지침에 따라 HTML 및 JSON과 같은 형식으로 응답을 생성합니다.
자막 생성/설명	다양한 수준의 세부정보를 사용해서 이미지 및 동영상에 대한 설명을 생성합니다. 이미지/동영상에 대해 다음 프롬프트에서 시작하고 반복 처리를 통해 보다 구체적인 설명을 얻는 것이 좋습니다. <ul style="list-style-type: none">• 이미지: "이 이미지에 대해 설명해줄 수 있나요?"• 동영상: "이 동영상에서 진행되고 있는 상황에 대해 설명해줄 수 있나요?"
외삽	위치, 이미지 또는 동영상에서 이전/다음/중간에 발생할 수 있는 상황에 따라 확인할 추가 항목을 제안하고 시각적 입력을 기반으로 이야기 쓰기와 같은 창의적인 사용을 지원합니다.



Gemini 한계

제한 사항	설명
공간 추론	이미지에서 정확한 객체/텍스트 지역화가 어려울 수 있습니다. 회전된 이미지를 이해하는 데에는 정확도가 저하될 수 있습니다.
개수 계산	특히 가려진 객체의 경우 대략적인 객체 수만 제공할 수 있습니다.
긴 동영상 이해	개별 형식(개별 이미지 처리와 다름)으로 동영상을 지원할 수 있습니다. 하지만 모델은 연속적인 동영상 자체(오디오 제외)가 아닌 비연속적인 이미지 프레임 집합으로부터 정보를 수신합니다. Gemini는 또한 동영상이 2분을 초과할 경우 어떠한 정보도 추출하지 않습니다. 조밀한 콘텐츠가 포함된 동영상의 성능을 향상시키려면 동영상을 줄여서 모델이 더 많은 동영상 콘텐츠 부분을 캡처하도록 합니다.
복잡한 지침 수행	여러 추론 단계가 필요한 태스크는 어려울 수 있습니다. 지침을 나누거나 효율적인 안내를 위해 퓨샷 예시를 제공하는 것이 좋습니다.
할루시네이션	경우에 따라 이미지/동영상에 실제로 포함된 내용을 벗어나거나 광범위한 텍스트 파싱으로 잘못된 콘텐츠를 생성할 수 있습니다. 강도를 줄이거나 더 짧은 설명을 요청하면 이 문제를 해결하는 데 도움이 될 수 있습니다.
의료 용도	의료용 영상(예: X선, CT 촬영)을 해석하거나 의료 조언을 제공하는 데 적합하지 않습니다.
멀티턴(멀티모달) 채팅	챗봇 기능 또는 채팅 톤으로 질문에 답변하도록 학습되지 않았으며 멀티턴 대화에서 효율이 저하될 수 있습니다.



