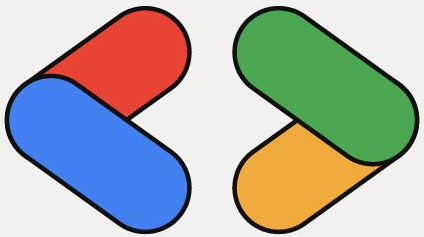
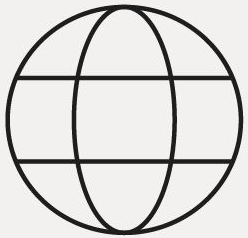
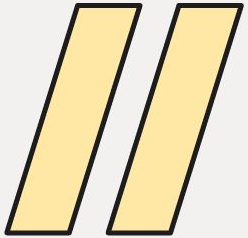
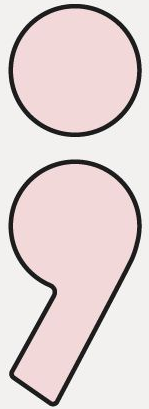
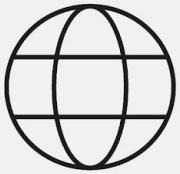
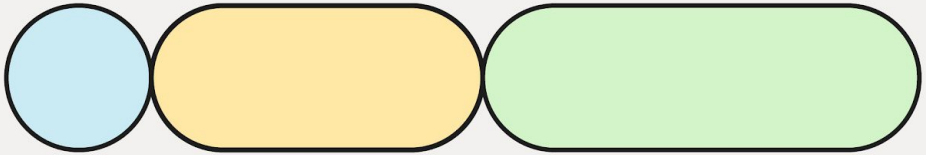





Beyond Text: Exploring Multimodality of Gemini



Google Developer Groups

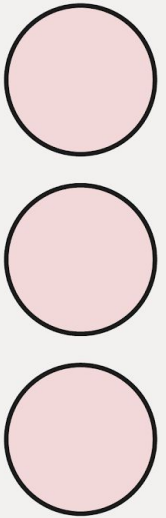




DevFest

Mauritius

Natively Multimodal한 Gemini의 기능을 데모를 통해서 살펴봅니다.



Google Developer Groups

Speaker Introduction

T Kim, Ph.D.

Machine Learning
Solution Lead

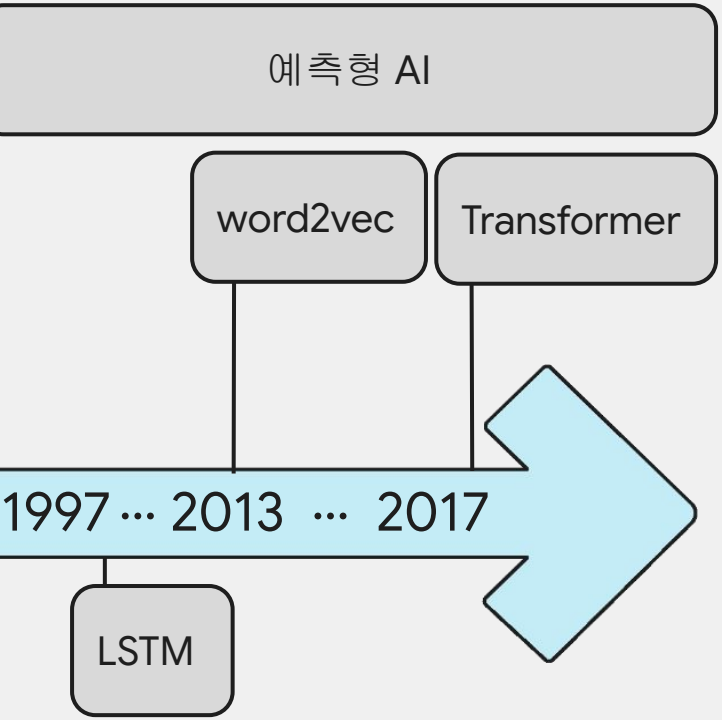
Google Cloud > APAC Customer Growth Engine

- Group of SMEs to meet the top customers in APAC
- AI Practice Team > AI Practice Specialist

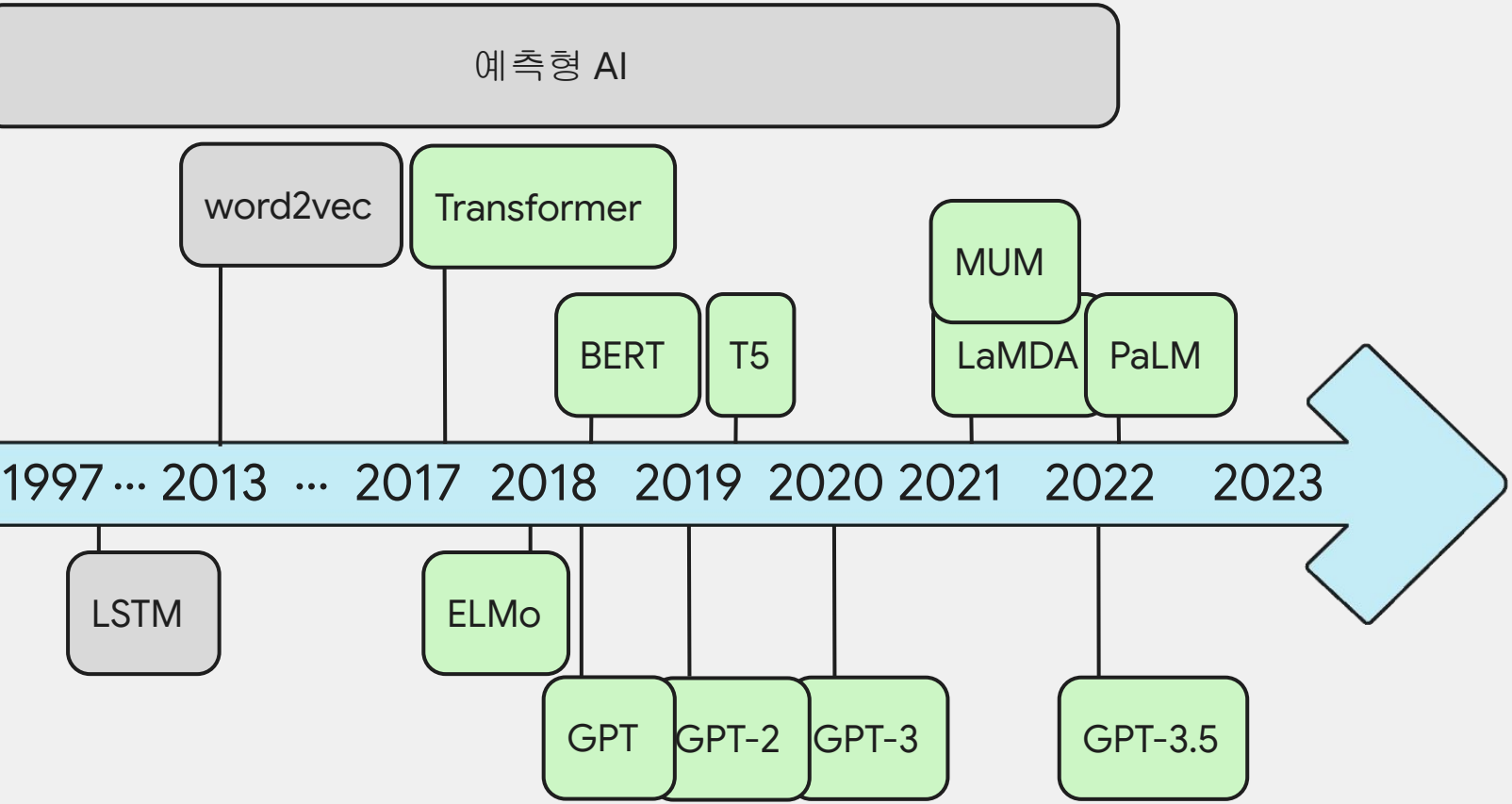
Agenda

1. Gemini 1.5 Pro: 멀티모달 생성형AI의 새로운 지평
2. 멀티모달 방식의 이해
   - Vertex AI Gemini 1.5 Pro 활용 가이드
3. 멀티모달 방식의 활용
   - Vertex AI 기반 멀티모달 Q&A 시스템 구축
4. 결론

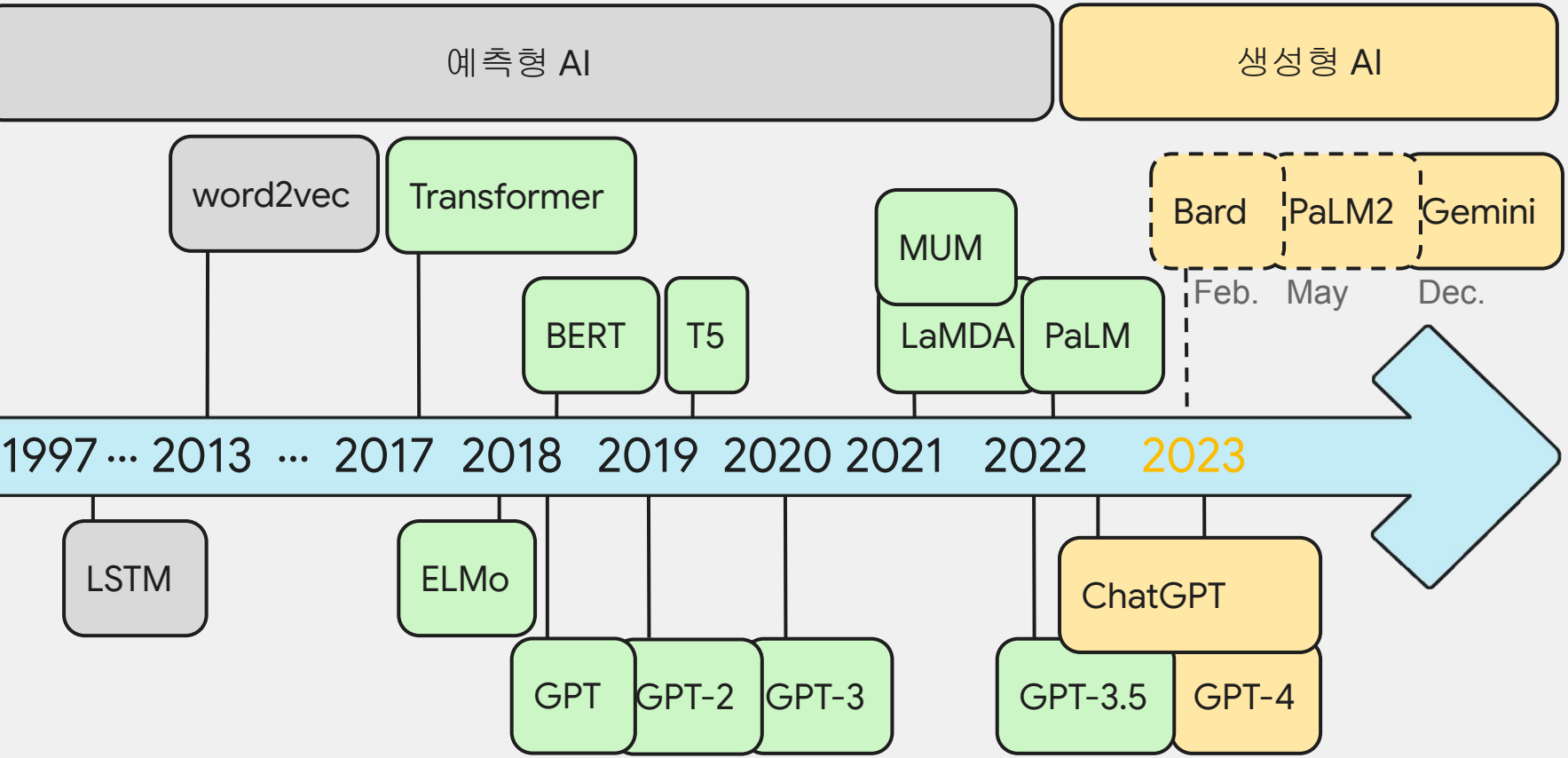
1. Gemini 1.5 Pro: 멀티모달 생성형AI의 새로운 지평



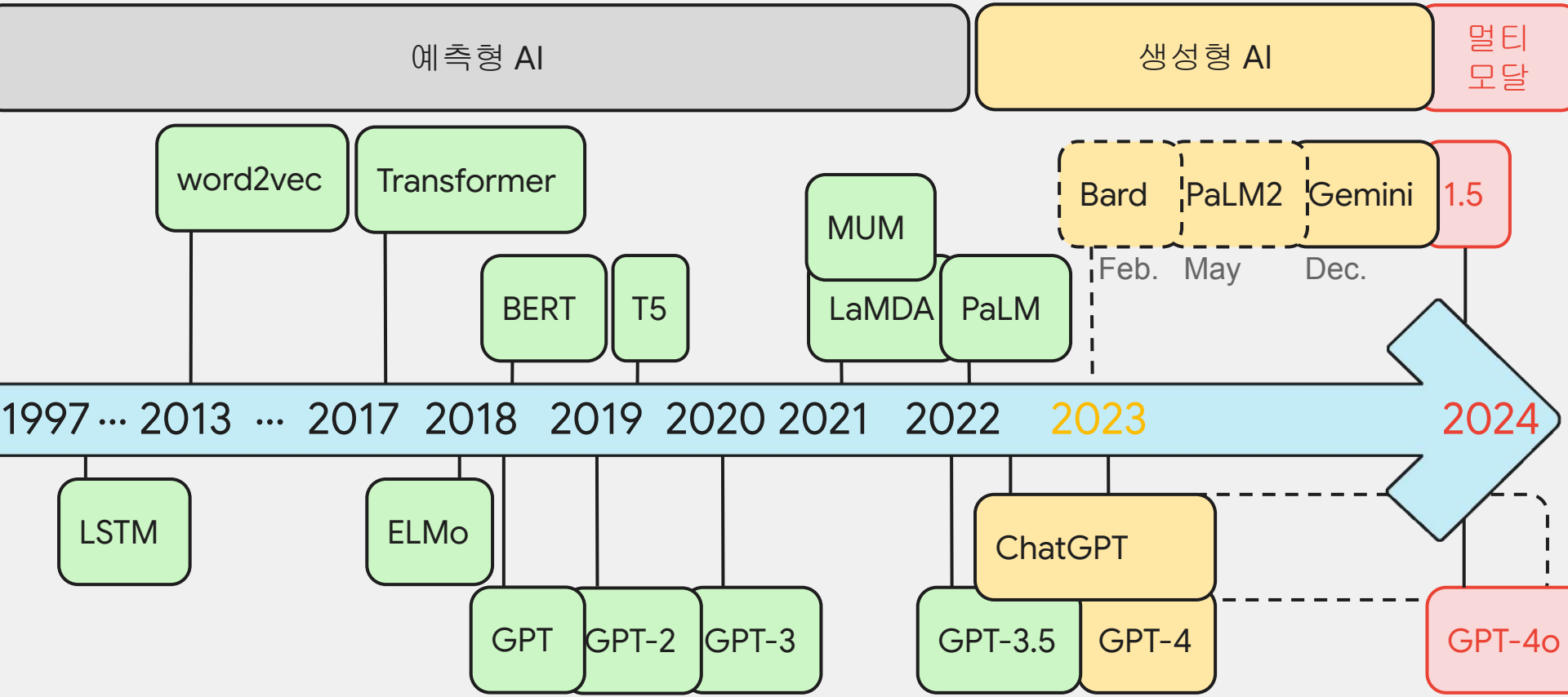
1. Gemini 1.5 Pro: 멀티모달 생성형AI의 새로운 지평



1. Gemini 1.5 Pro: 멀티모달 생성형AI의 새로운 지평



1. Gemini 1.5 Pro: 멀티모달 생성형AI의 새로운 지평



멀티모달 방식이 왜 필요할까요?

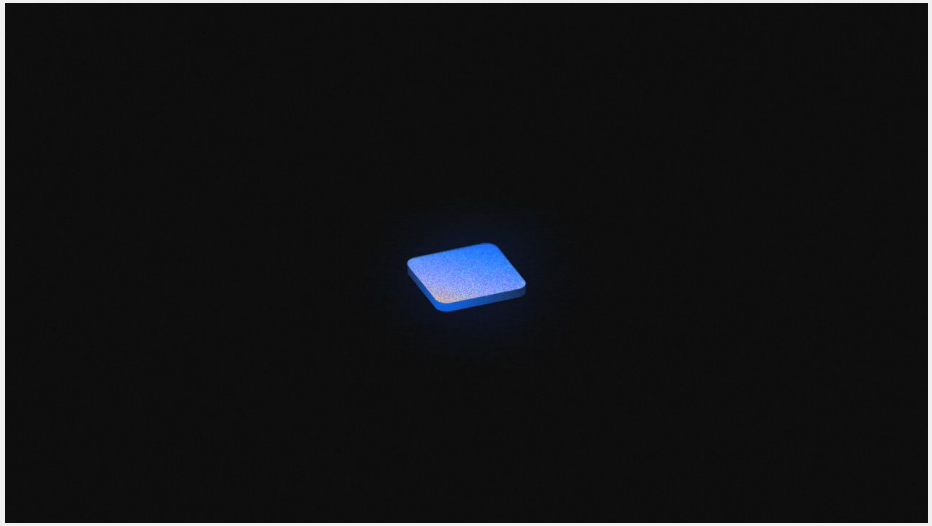
인간의 인지 능력과 유사함
AI 시스템이 인간의 사고방식과 유사하게 작동하려면, 멀티모달 방식으로 다양한 정보를 이해할 필요.

인간의 뇌는 문자, 시각 정보, 소리 등 다양한 정보를 종합해서 상황을 이해하고 의미를 파악.

멀티모달로 이해도와 정확도를 향상 가능

Gemini

는 Google이 개발한
모델 중 가장 강력한
범용 모델



is
natively
multimodal

처음부터 멀티모달로 설계되어
텍스트뿐만 아니라 이미지, 비디오,
오디오 등 다양한 형태의 정보를
동시에 이해하고 처리할 수 있습니다

Gemini 1.5 Pro

2M 컨텍스트 윈도우

다양한 작업에 사용

서비스 중인 베스트 모델

Gemini 1.5 Pro

2M 컨텍스트 윈도우
다양한 작업에 사용
서비스 중인 베스트 모델

Gemini 1.5 Flash

1M 컨텍스트 윈도우
대용량 작업을 위해 제작
더 빠른 속도+낮은 비용을 위한 모델

긴 컨텍스트 윈도우

→ 매우 긴 텍스트, 오디오 및 비디오
데이터를 기억하고 분석 가능

* 컨텍스트 캐싱 기능 제공

긴 컨텍스트 윈도우
→ 매우 긴 텍스트, 오디오 및 비디오
데이터를 기억하고 분석 가능

최대 동영상 길이
~50분 (오디오 포함)
60분 (오디오 미포함)
프롬프트당 최대 동영상 수: 10

Gen AI on Vertex AI > Doc. > Video understanding

- Samples and notebooks > Summarize a video file with audio with Gemini 1.5 Pro

2. 멀티모달 방식의 이해

여러 모드/형식의 데이터를 처리

- 텍스트, 코드
- 이미지
- PDF
- 오디오
- 비디오 (오디오 포함/미포함)

2. 멀티모달 방식의 이해

여러 모드/형식의 데이터를 처리

- 텍스트, 코드
- 이미지
- PDF
- 오디오
- 비디오 (오디오 포함/미포함)

다양한 형태의 정보를 동시에 이해하고 처리

- 비디오-오디오 인터리빙 기술
  비디오 프레임과 오디오 세그먼트를 번갈아서 입력
- 예: 영화의 한 장면을 보며 대사를 듣고, 인물의 감정,
  이야기의 흐름 등을 종합적으로 파악 가능

Vertex AI Gemini 1.5 Pro 활용 가이드

본 가이드에서는 Vertex AI API를 통해 Google의 멀티모달 LLM, Gemini 1.5 Pro를 활용하는 방법을 살펴봅니다. 텍스트 및 비디오 처리를 위한 실습 예제를 통해 Gemini 1.5 Pro의 강력한 기능을 경험하고, 멀티모달 AI 기술의 잠재력을 탐색합니다. 본 가이드를 통해 Gemini 1.5 Pro의 다양한 활용 가능성을 탐색하고, 멀티모달 AI기술을 이용한 혁신적인 애플리케이션 개발에 대한 영감을 얻으시기 바랍니다.

실습 내용

텍스트 처리에 특화된 LLM (translator_model)과 멀티모달 데이터 처리에 최적화된 LLM (model)을 구분하여 사용하며, 다음과 같은 내용을 다룹니다.

1. 설정
2. 모델 로딩: Gemini 1.5 Pro
3. 텍스트 이해: 영한 번역
4. 비디오 이해
5. 멀티모달 입력의 이해

설정

[1]:
```
!pip3 install --upgrade --quiet google-cloud-aiplatform
```

[2]:
```
# Restart runtime
import sys

if "google.colab" in sys.modules:
    import IPython

    app = IPython.Application.instance()
```

AI Solution

Contact Center AI | Risk AI | Healthcare Data Engine | Search for Retail, Media and Healthcare

Gemini for Google Cloud

Gemini for Google Workspace

Build your own generative AI-powered agent

Vertex AI Agent Builder

OOTB and custom Agents | Search
Orchestration | Extensions | Connectors | Document Processors | Retrieval engines | Rankers | Grounding

Vertex AI Model Builder

Prompt | Serve | Tune | Distill | Eval | Notebooks I Training | Feature Store | Pipelines | Monitoring

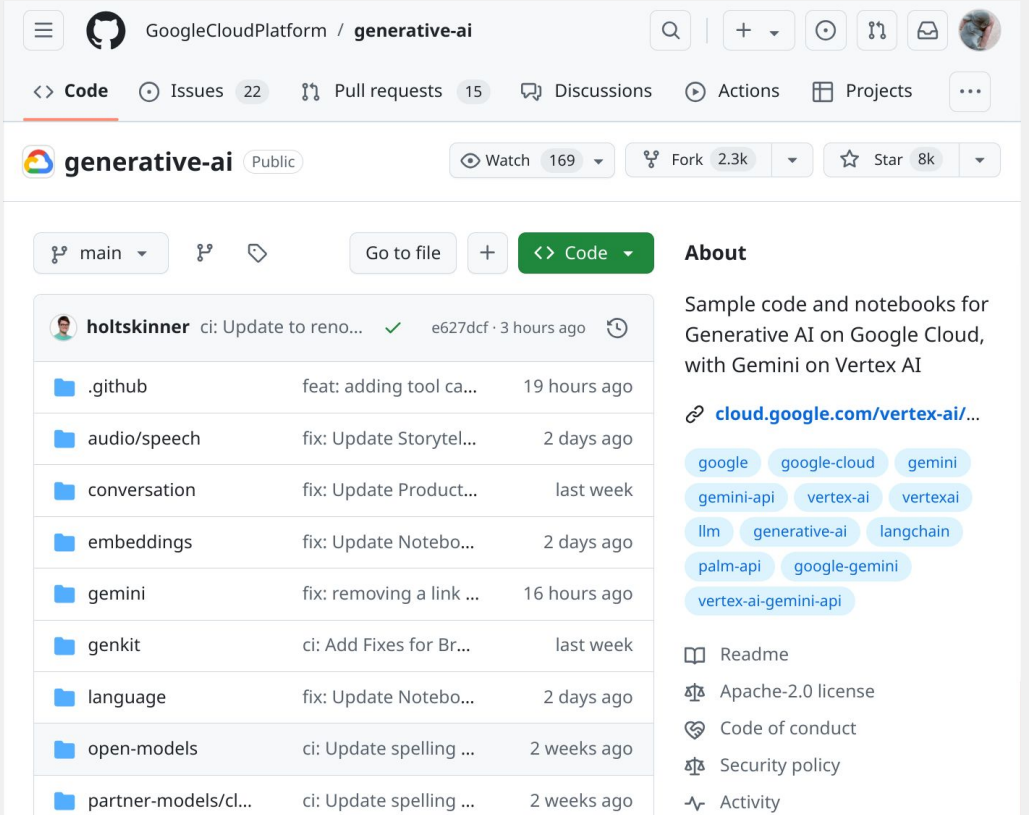
Vertex AI Model Garden

Google | Open | Partner

Google Cloud Infrastructure (GPU/TPU) | Google Data Cloud

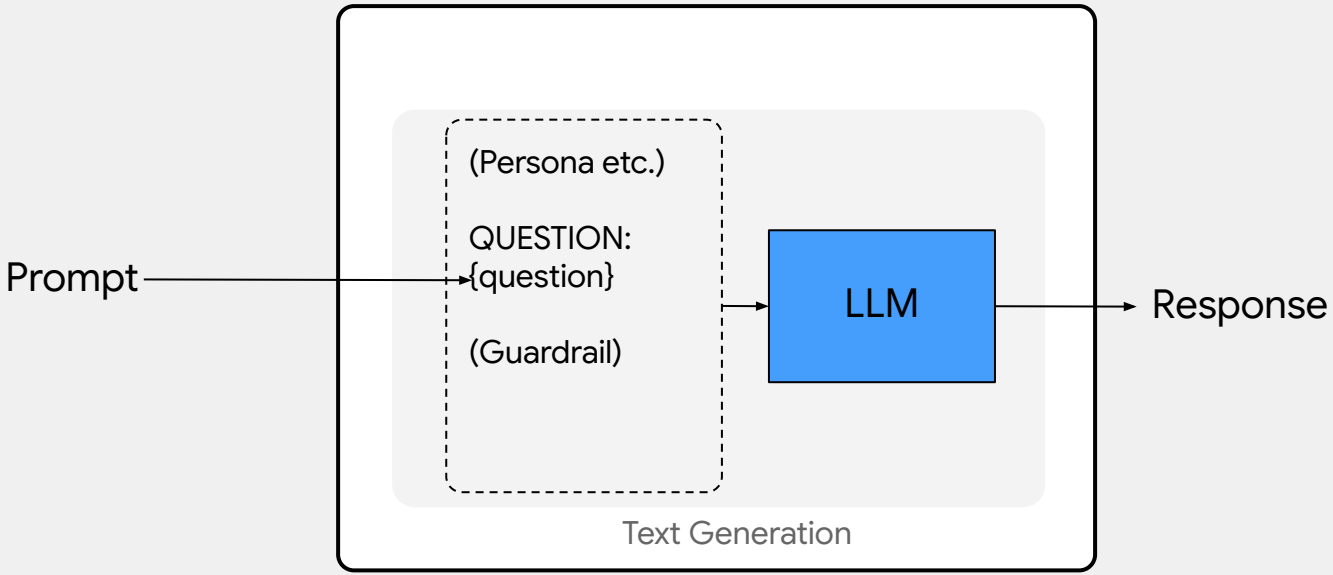
github.com/GoogleCloudPlatform/generative-ai

goo.gle/gen-ai-github

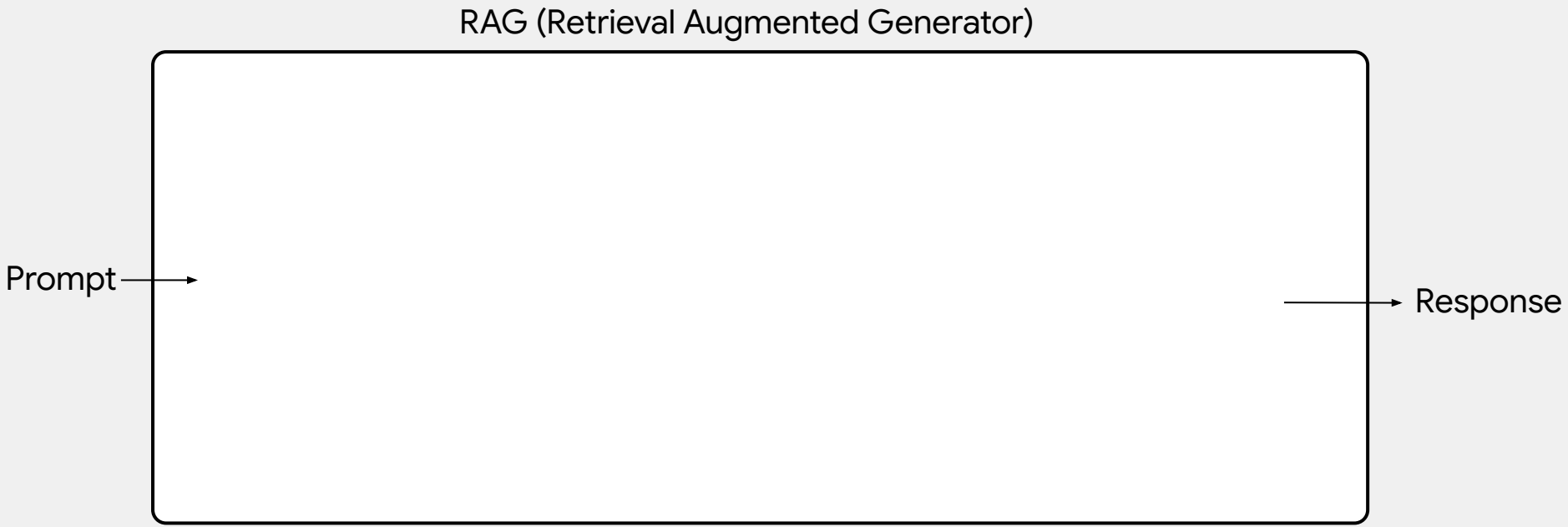



순환
내리실 문은 왼쪽입니다
종합운동장
삼성
선릉
역삼

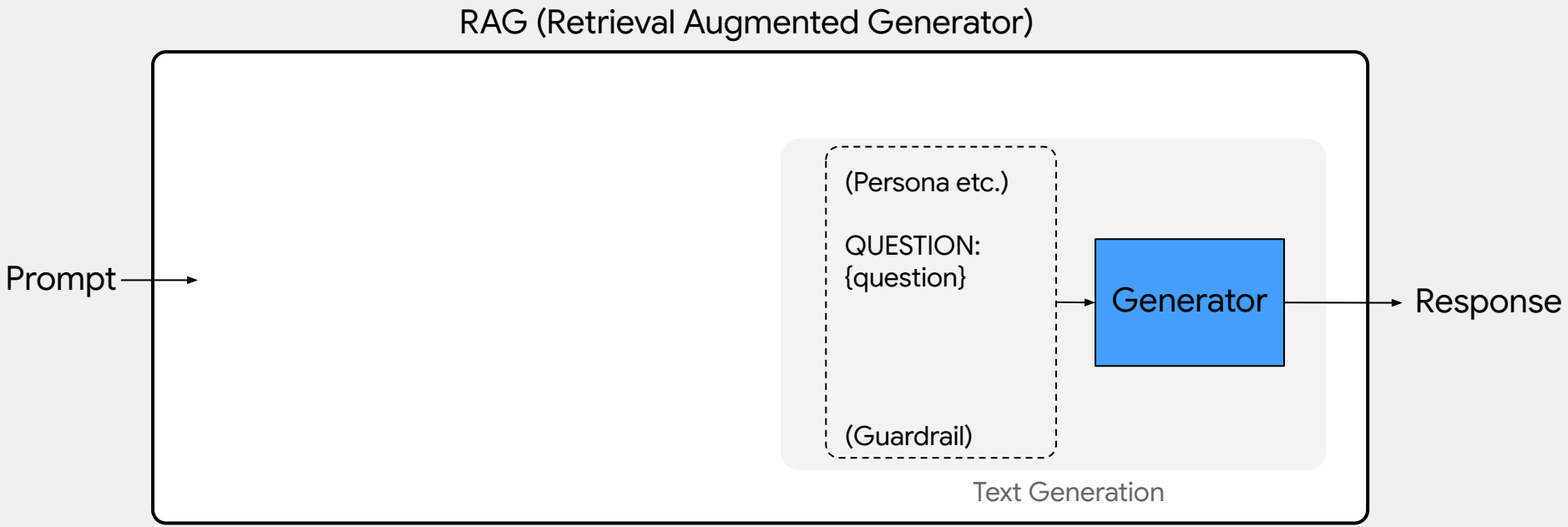
3. 멀티모달 방식의 활용 > 멀티모달 RAG



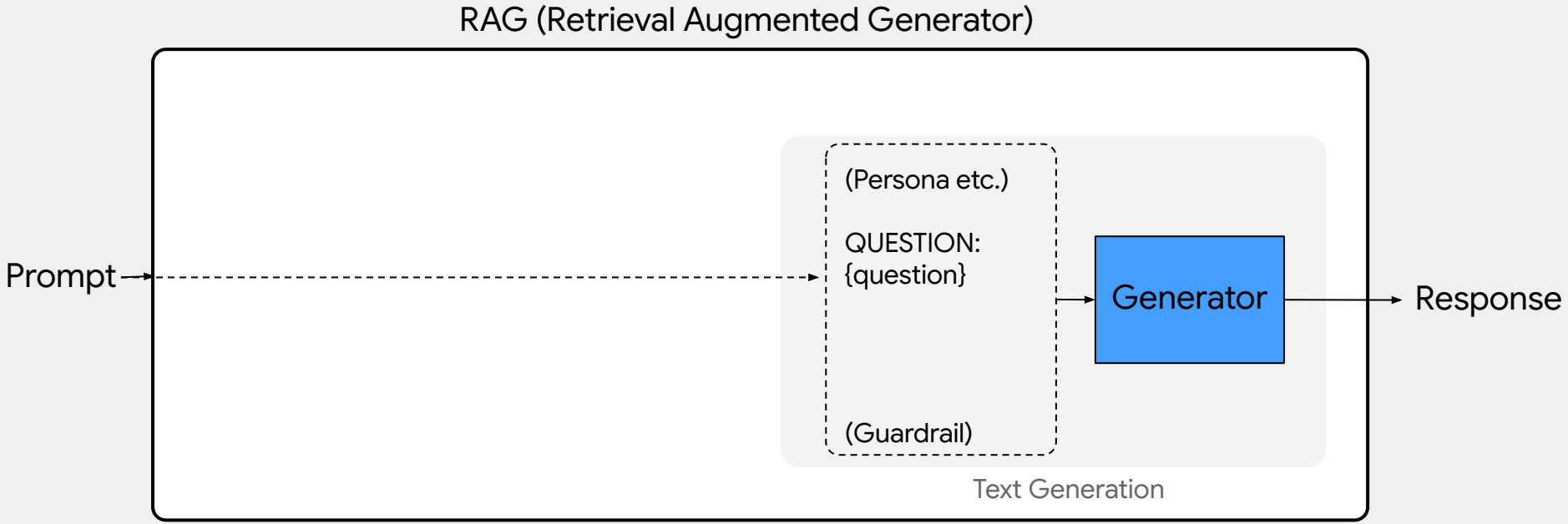
3. 멀티모달 방식의 활용 > 멀티모달 RAG



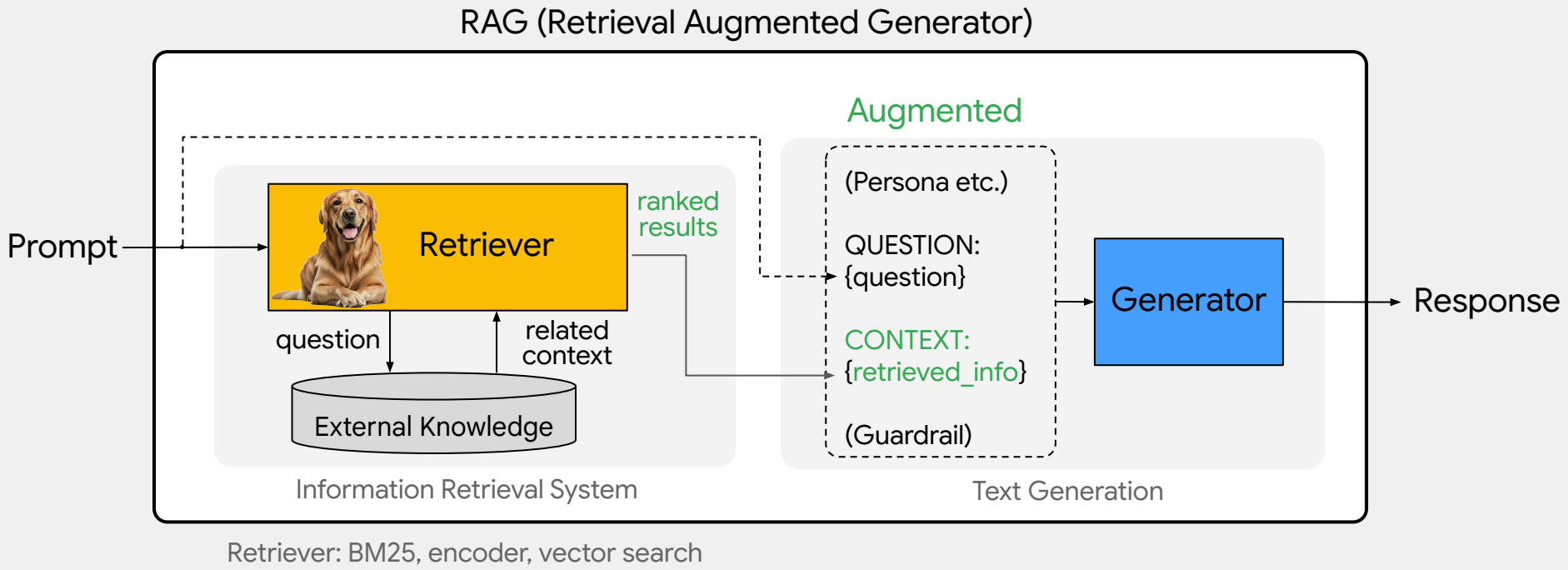
3. 멀티모달 방식의 활용 > 멀티모달 RAG



3. 멀티모달 방식의 활용 > 멀티모달 RAG



3. 멀티모달 방식의 활용 > 멀티모달 RAG



Retriever: BM25, encoder, vector search

Image prompt, square, no style: Image of a retriever. Use a white background

3. 멀티모달 방식의 활용 > 멀티모달 RAG

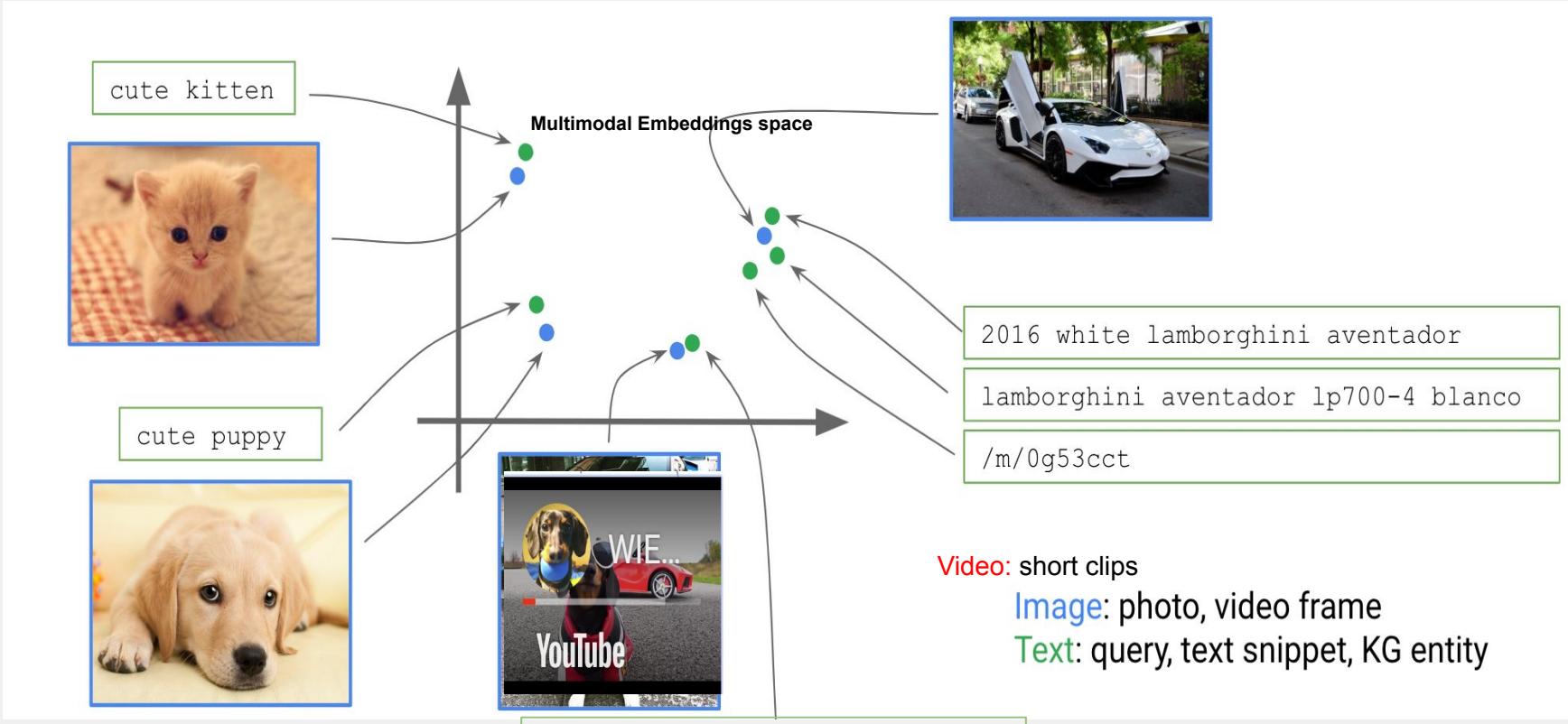
Task: Answer the following questions in detail, providing clear reasoning and evidence from the images and text in bullet points.
Instructions:

1. **Analyze:** Carefully examine the provided images and text context.
2. **Synthesize:** Integrate information from both the visual and textual elements.
3. **Reason:** Deduce logical connections and inferences to address the question.
4. **Respond:** Provide a concise, accurate answer in the following format:

   * **Question:** [Question]
   * **Answer:** [Direct response to the question]
   * **Explanation:** [Bullet-point reasoning steps if applicable]
   * **Source** [name of the file, page, image from where the information is cited]

5. **Ambiguity:** If the context is insufficient to answer, respond "Not enough context to answer."

3. 멀티모달 방식의 활용 > 멀티모달 임베딩



Vertex AI 기반 멀티모달 Q&A 시스템 구축하기

본 가이드에서는 Vertex AI API를 통해 Google의 멀티모달 LLM, Gemini를 이용해서 텍스트와 이미지를 모두 이해하는 멀티모달 Q&A 시스템 구축 방법을 살펴봅니다.

목적

- Google Cloud의 Vertex AI 플랫폼을 기반으로 시스템을 단계별로 구현하여, 외부 도구에 의존하지 않고 핵심 원리를 명확하게 이해할 수 있도록 합니다.
- 강력한 성능을 자랑하는 대규모 언어 모델 (LLM)은 종종 내부 작동 방식이 불투명한 "블랙박스"처럼 느껴질 수 있습니다.
- 멀티모달 LLM 기반 Q&A 시스템을 직접 구축하는 과정에서 핵심 개념의 심층적 이해를 도모합니다.

주의사항: 본 가이드는 멀티모달 RAG (Retrieval Augmented Generation)를 이용한 멀티모달 Q&A 시스템 구축에 대한 기본적인 이해를 돕기 위한 예시입니다.

- 실제 시스템 구축 시에는 다양한 접근 방식을 고려해야 하며, 필요에 따라 구성 요소를 직접 코딩하거나 외부 라이브러리를 활용할 수 있습니다.
- 본 가이드의 내용을 참고하여 시스템의 목적과 요구사항에 맞는 최적의 구현 방안을 모색하시기 바랍니다.

실습 내용

(텍스트와 이미지를 포함한) PDF 문서의 테스트, 이미지, 표 등으로 **메타데이터 저장소**를 구성한 다음, 다양한 방식으로 **검색, 비교, 추론**을 합니다.

1. 노트북 환경 설정

Vertex AI API만을 사용해서 빌딩 블록을 완벽하게 제어하고 이해할 수 있습니다.

- Vertex AI Embeddings API
  - 텍스트 임베딩

Google Cloud Summit Seoul ’24 >
Invitation only |
Special Hands on
(총 2시간 30분 분량)

Google Developer Groups

Google Cloud Summit Seoul '24 > Workshops > GenAI Live and Labs

• **Question:** Gemini 1.5 Pro의 전문가 혼합(MoE) 아키텍처는 핵심 기능의 성능을 유지하면서 긴 컨텍스트를 처리하는 능력에 어떻게 기여합니까? 관련된 잠재적인 trade-off에 대해 알려주세요.
• **Answer:** Gemini 1.5 Pro의 전문가 혼합(MoE) 아키텍처는 **모델 내에서 다양한 작업 및 데이터 유형에 대한 특정 전문가 모듈을 활성화**하여 핵심 성능을 저하시키지 않고도 우수한 장기 컨텍스트 처리를 가능하게 합니다.
• **Explanation:**
    • **Specialized Expertise:** MoE를 통해 Gemini 1.5 Pro는 모델 내의 전문화된 "전문가"를 활용할 수 있습니다. 각 전문가는 특정 도메인이나 작업에 집중하여 다양한 데이터와 복잡한 쿼리를 보다 효율적으로 처리할 수 있습니다.
    • **Dynamic Routing:** 모든 입력에 대해 모델의 모든 부분을 활성화하는 대신 MoE는 입력 데이터를 가장 관련성 있는 전문가에게 동적으로 라우팅합니다. 이 선택적 활성화는 컴퓨팅 리소스를 보존하고 더 긴 컨텍스트 처리를 허용합니다.
    • **Maintaining Core Capabilities (핵심 역량 유지):** 핵심 작업에 집중하는 전문가를 유지하고 긴 컨텍스트 처리를 위해 새로운 전문가를 통합함으로써 Gemini 1.5 Pro는 기존 영역에서 성능을 희생하지 않고도 역량을 확장할 수 있습니다.
• **Source:** gemini_v1_5_report_technical.pdf (Pages 1, 3)
• **Trade-offs:**
    • **복잡성 :** MoE 아키텍처는 모델의 복잡성을 증가시켜 잠재적으로 훈련 및 미세 조정을 더욱 어렵게 만듭니다.
    • **메모리 증가:** 여러 전문가 모듈을 저장하려면 기존의 모놀리식 모델에 비해 더 많은 메모리가 필요합니다.

4. 결론

Gemini 1.5의 natively multimodal한 기능과 특성을 살펴봄

- 노트북 다운로드: https://github.com/aimldl/genai

4. 결론

Gemini 1.5의 natively multimodal한 기능과 특성을 살펴봄

- 노트북 다운로드: https://github.com/aimldl/genai

텍스트 vs. 멀티모달 RAG

- 인간의 인지 과정은 문자, 시각 정보, 소리 등 다양한 양식의 정보를 통합적으로 처리하여 상황을 이해하고 의미를 도출하는 복잡한 메커니즘을 기반으로 합니다.

- 이 인간의 사고 방식을 모방하여 AI 시스템의 성능을 향상시키기 위해서는 멀티모달 접근 방식이 필수적입니다.

- 텍스트 기반 RAG가 텍스트 데이터에 국한되는 반면, 멀티모달 RAG는 다양한 양식의 정보를 활용하여 AI 시스템의 이해도와 정확도를 향상시킬 수 있습니다.

4. 결론

멀티모달 방식의 주요 이점

- 향상된 지식 접근성
  멀티모달 RAG는 텍스트, 이미지, 오디오, 비디오 등 다양한 형태의 데이터를 포괄적으로 검색하고 활용할 수 있는 기반 기술을 제공합니다. 이는 AI 시스템이 더욱 광범위하고 심층적인 지식 베이스에 접근하여 보다 정확하고 포괄적인 답변을 생성할 수 있도록 지원합니다.

- 추론 능력 향상
  멀티모달 RAG는 다양한 유형의 데이터 간의 상관관계를 분석하고, 이를 기반으로 복잡한 추론을 수행할 수 있습니다. 예를 들어, 이미지와 텍스트 정보를 결합하여 특정 상황에 대한 맥락을 더욱 정확하게 파악하여, 향상된 추론 능력을 보여줄 수 있습니다.

4. 결론

결론적으로...

- 멀티모달 RAG는 AI 시스템의 인지 능력을 향상시키는 핵심 기술로, 인간과 유사한 방식으로 정보를 처리하고 이해할 수 있도록 지원합니다.
- 이는 AI 시스템이 더욱 복잡하고 다양한 작업을 수행하고, 인간과 자연스러운 상호 작용을 가능하게 하는 미래 GenAI 기술의 초석이 될 것입니다.

Gemini

감사합니다

Image prompt, square, no style: A structure made of letters "Natively Multimodal". Text, images, video, audio and code icons are placed around the structure. Use a white background