

## 팀 소개



팀명 : KBVQA

팀원 :

NO	이름	전공(학년)	주요역할
1	김민준	컴퓨터공학(4)	VQA 모델 개발
2	송승우	컴퓨터공학(4)	VQA 모델 개발
3	송지현	컴퓨터공학(4)	모델 배포 및 앱 개발

팀 소개(한줄) : 유아 교육을 위한 VQA 애플리케이션을 개발한 KBVQA 팀입니다.

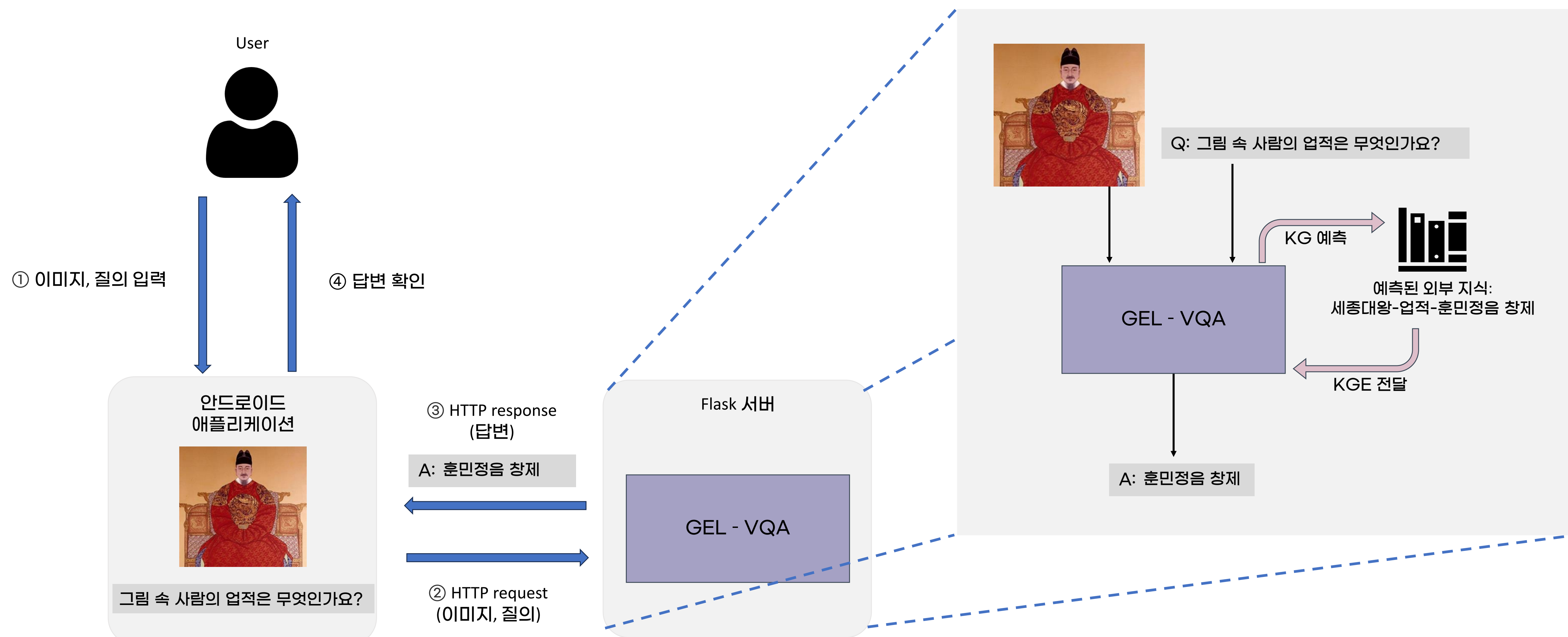
## 프로젝트 소개

프로젝트명 : 지식 베이스를 활용한 VQA 유아 교육 애플리케이션

주요기술 : 자연어처리, 컴퓨터비전, 멀티모달 학습, 지식 그래프 임베딩

작품소개(한줄) : VQA 모델에서 외부 지식 활용 방법을 고안하여 유아 교육에 활용

### 프로젝트 사진





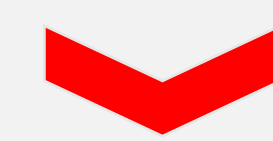
## 프로젝트 소개

### 프로젝트 개요 (개발배경, 목표 시스템 등)

VQA 시스템이란?	<ul style="list-style-type: none"> <li>주어지는 이미지와 질의에 대해 적절한 답변을 도출하는 시스템으로, <b>이미지와 질의에 대한 복합적 이해가 가능한 인공지능 시스템</b></li> <li>영유아 교육 도우미, 시각 장애인 보조 시스템 등으로 활용 가능</li> </ul>
VQA 시스템 작동 원리	<ul style="list-style-type: none"> <li>이미지 처리: CNN 기반의 이미지 특징 추출 모델 사용</li> <li>질의 처리: RNN 혹은 Transformer 기반 자연어 특징 추출 모델 사용</li> <li>답변에 이미지 특징과 자연어 특징을 모두 이해하도록 multi-modal fusion 방법을 적용하여 답변 분류 혹은 생성</li> </ul>
VQA 데이터 예시	 <div> <p>: 시계는 몇 시를 나타내나요? : 11시 55분</p> <p>Q : 이 운동의 이름은 무엇인가요? A : 테니스</p> <p>Q : 이 동물의 이름은 무엇인가요? A : 곰</p> </div>

#### ✓ 기존 VQA 시스템의 한계점

- 기존 VQA 시스템은 이미지와 질의로부터 얻는 정보만 활용해서 답변하기 때문에 **외부 지식이 필요한 질문에 대해서는 답변이 어려움**
- 이러한 한계점은 다양한 질문에 대한 대응이 필요한 유아 교육용 애플리케이션에 적합하지 않음

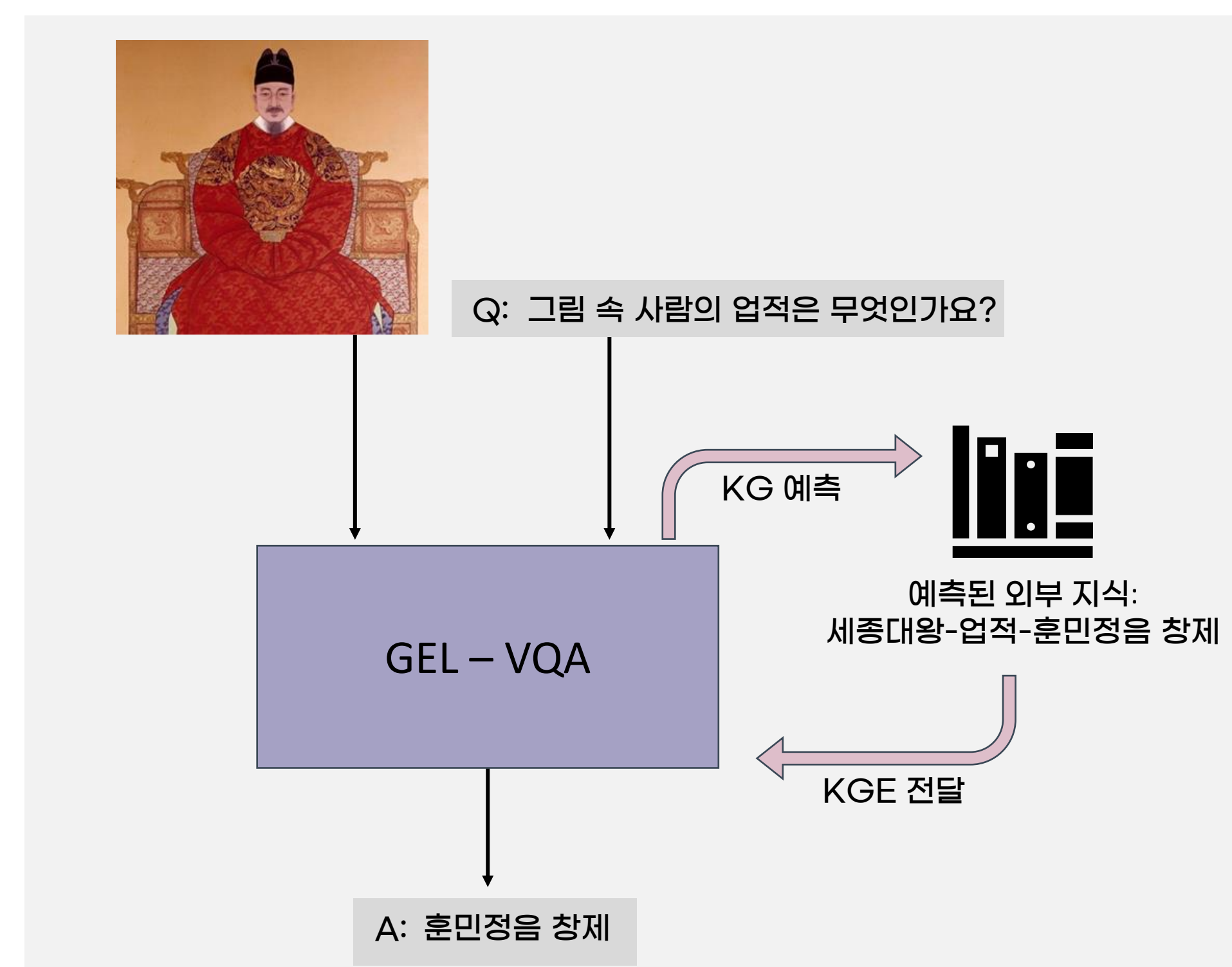


#### ✓ 지식 베이스 활용의 필요성

- 효과적인 유아 교육 솔루션을 제공하기 위해서는 이미지에 보이는 정보 뿐 아니라, 외부 지식을 활용해야 함
- 지식 베이스를 통해 유아들에게 보다 깊고, 넓은 배경 지식 전달 가능
- 유아 교육에서 중요한 것은 **정확한 정보의 전달** → 지식 베이스를 활용하면 정확한 과학적, 역사적 사실 제공 가능

### 프로젝트 구성 및 상세 내용

학습 데이터 수집	<ul style="list-style-type: none"> <li>AI Hub의 “외부 지식 기반 멀티모달 질의응답 데이터”에서 17,836개의 <b>시각 질의응답 데이터</b> 수집</li> <li>DBpedia, ConceptNet에서 지식 그래프 임베딩 학습을 위한 282,533개의 <b>지식 그래프</b> 수집</li> </ul>
모델 학습	<ul style="list-style-type: none"> <li>지식 그래프 임베딩 : 수집한 282,533개의 지식 그래프를 <b>ConvKB 모델</b>을 사용하여 학습</li> <li>GEL-VQA : 초기 학습의 불안정성을 개선하기 위해 외부 지식 예측에 <b>Teacher Forcing</b> 방법 적용</li> <li>외부 지식 예측과 VQA 답변 두 작업을 동시에 학습하기 위해 GEL-VQA 모델 학습에 <b>multi-task loss</b> 사용</li> <li>예측한 지식 그래프에서, head-relation-tail의 상대적 중요도를 학습하기 위해 <b>self-attention</b> 적용</li> </ul>
안드로이드 애플리케이션 개발	<ul style="list-style-type: none"> <li>학습된 GEL-VQA 모델을 Flask 서버에 배포</li> <li>사용자가 이미지와 질문을 입력하면 클라이언트는 Flask 서버로 전송하고 GEL-VQA를 통해 답변 추론</li> <li>추론된 답변은 다시 클라이언트에게 전송되며 <b>사용자는 응답을 실시간으로 확인</b></li> </ul>



#### GEL-VQA 모델 답변 프로세스

- 이미지와 질의를 입력 받는다.
- GEL-VQA 내부에서 이미지와 질의와 관련된 지식 그래프 (세종대왕-업적-훈민정을 창제)를 **실시간으로 추론**한다.
- 예측된 지식 그래프에 지식 그래프 임베딩 방법을 적용하여 변환한 **지식 임베딩**을 **GEL-VQA 모델에 전달**한다.
- 모델에 전달된 지식 임베딩은 GEL-VQA 모델이 더욱 정확한 답변을 줄 수 있도록 보조한다.

### 개발 세부 내용

#### 지식 그래프 임베딩 (KGE)

- 지식 그래프 임베딩은 지식 그래프 내의 Entities(head, tail)와 Relations를 효율적으로 벡터 공간에 표현하는 방법
- 지식 그래프의 정보를 고밀도로 압축하고, **VQA 모델에 지식 정보를 주입**하는 것에 활용함 (그림1.)

#### GEL-VQA 구조: 외부지식 예측 + 답변 예측

- 이미지 특징 추출기는 ResNet50 모델을 사용하였고, 질의 특징 추출기로는 bilinguality를 위해 XLM-RoBERTa 사용
- 그림 2.에서 (a)는 VQA모듈, (b)는 지식 베이스(지식 그래프) 예측을 위한 KB-prediction 모듈, (c)는 사전 학습된 지식 그래프 임베딩 테이블을 나타냄. (b)에서 예측한 외부지식을 사용하여 (c)를 통해 임베딩 형태로 변환하고, (a)에서 보다 정확한 답변을 도출할 수 있도록 도움

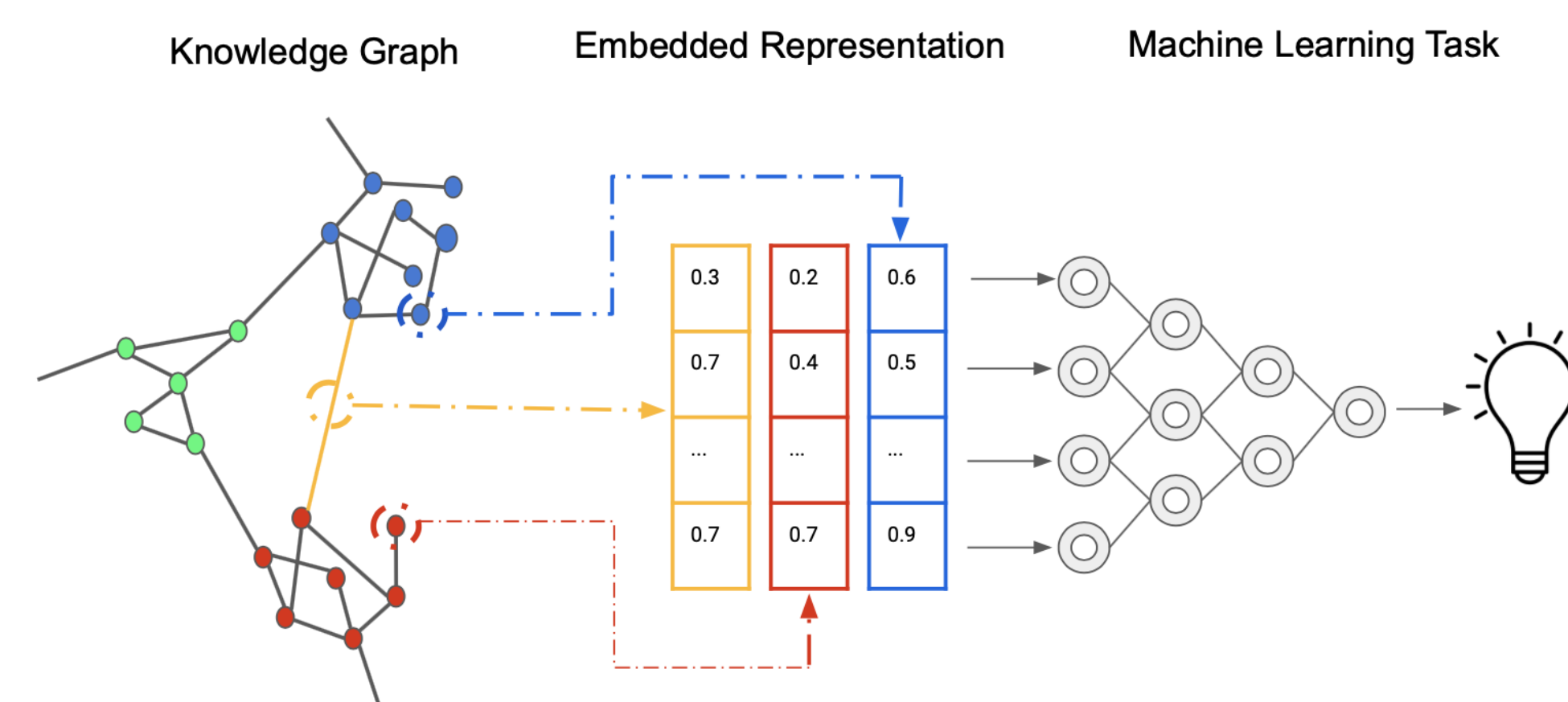


그림1. 지식 그래프 임베딩

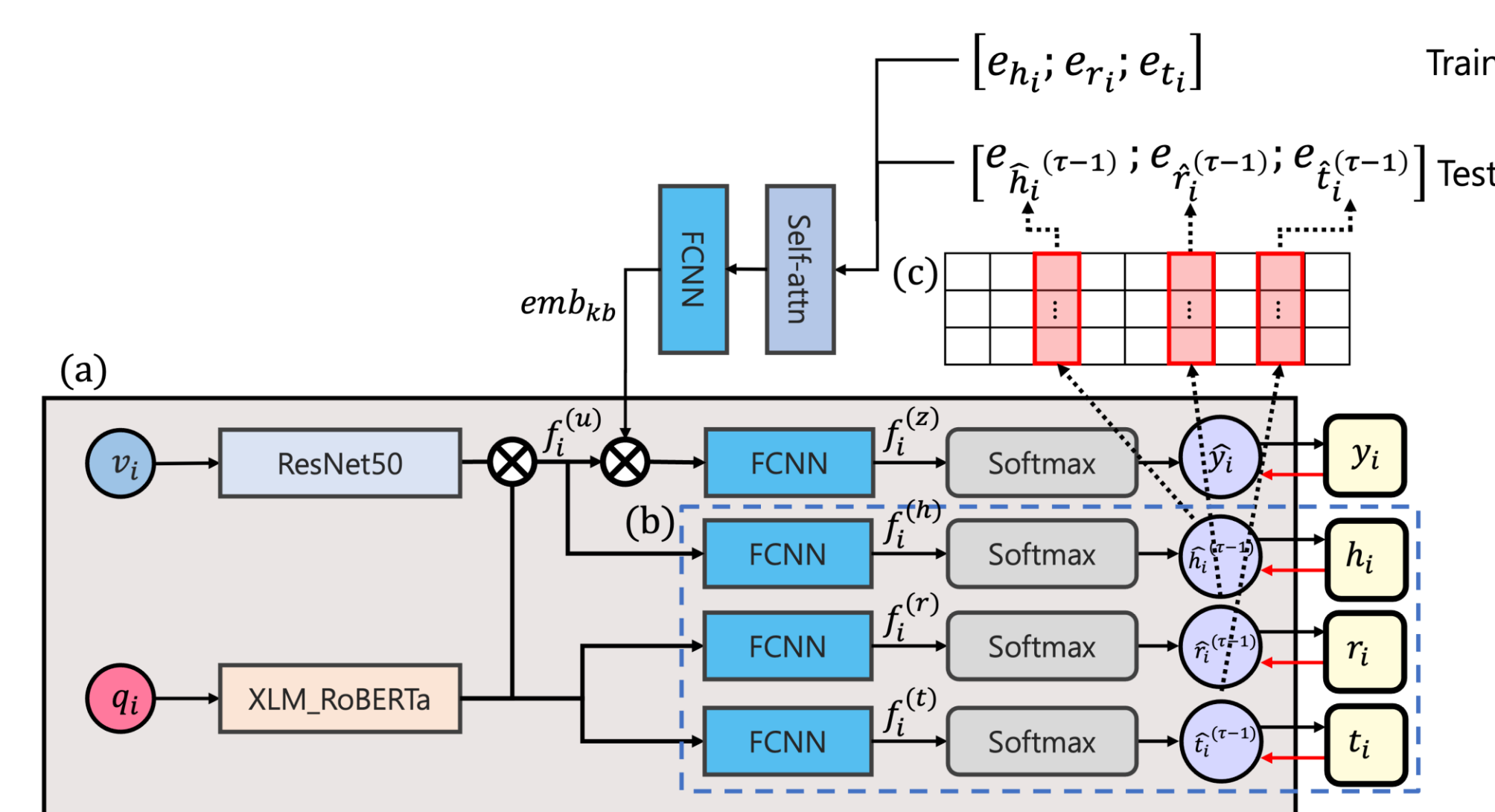


그림2. GEL-VQA 모델 구조



## 프로젝트 소개

### 구현 결과



Q : 이미지에 있는 나비는 어느 문에 속하는 동물인가요?

GEL - VQA

A : 절지동물 문  
Predict KB : 나비-문-절지동물



그림 3. Grad CAM과 attention visualization을 통한 결과 분석

#### GEL-VQA 모델 성능 시각화

- GEL-VQA 모델이 지식 그래프 “나비-문-절지동물”을 예측하였으며 그와 동시에 나비의 문에 해당하는 “절지동물 문” 답변 예측
- 그림 3.의 Grad CAM과 attention 시각화를 살펴보면, 이미지와 외부 지식 모두 적절한 정보를 중점적으로 활용하고 있는 것을 확인

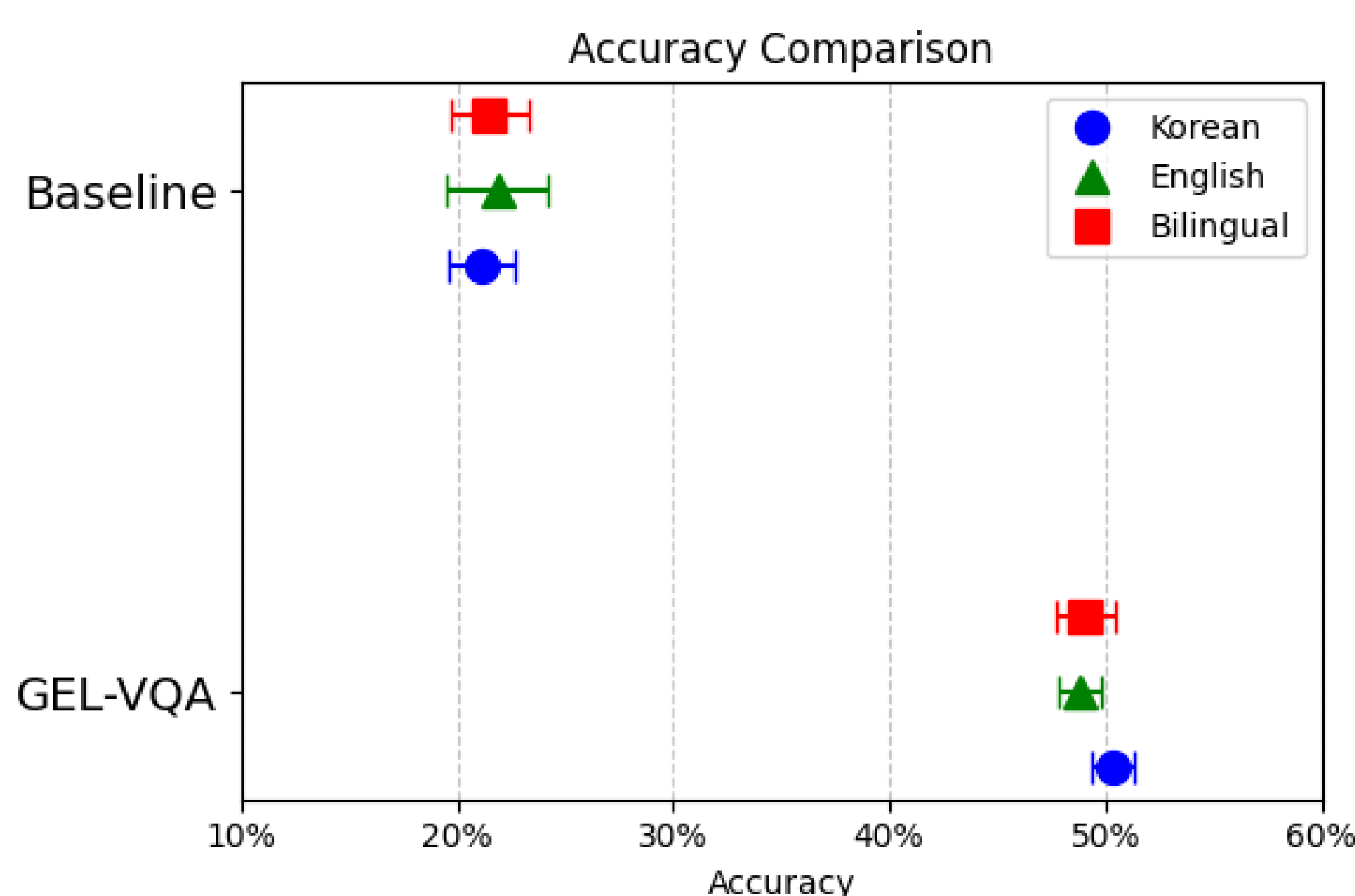
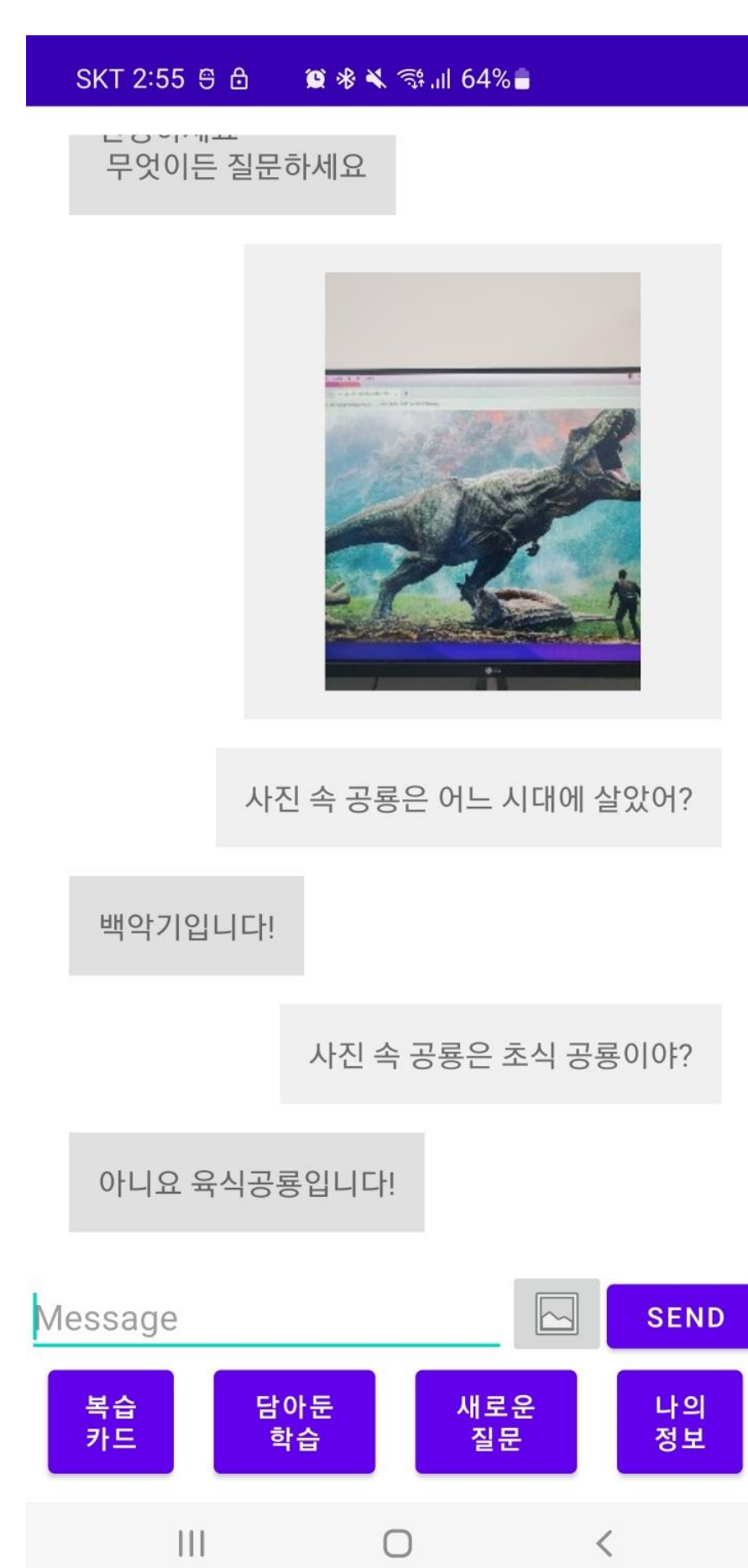
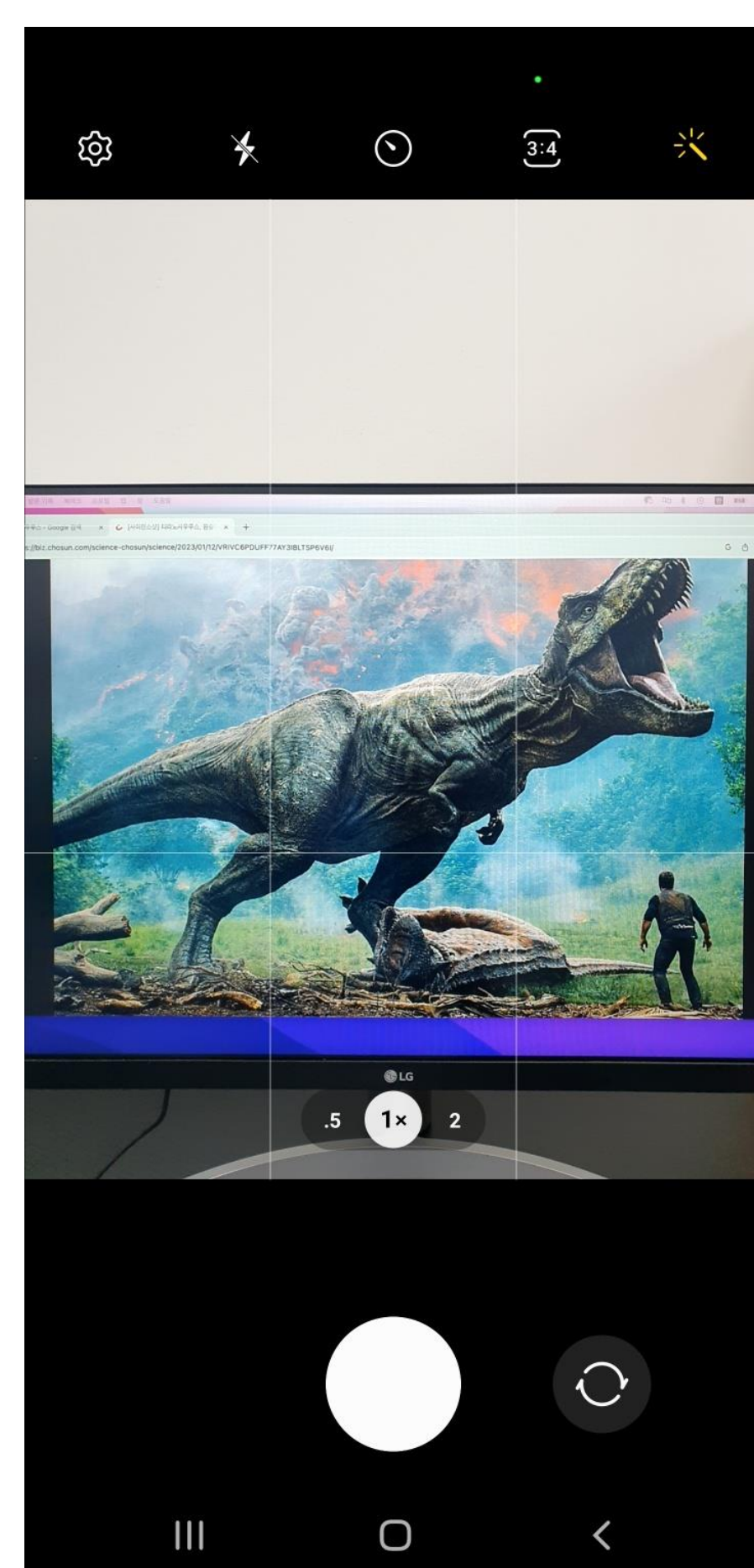
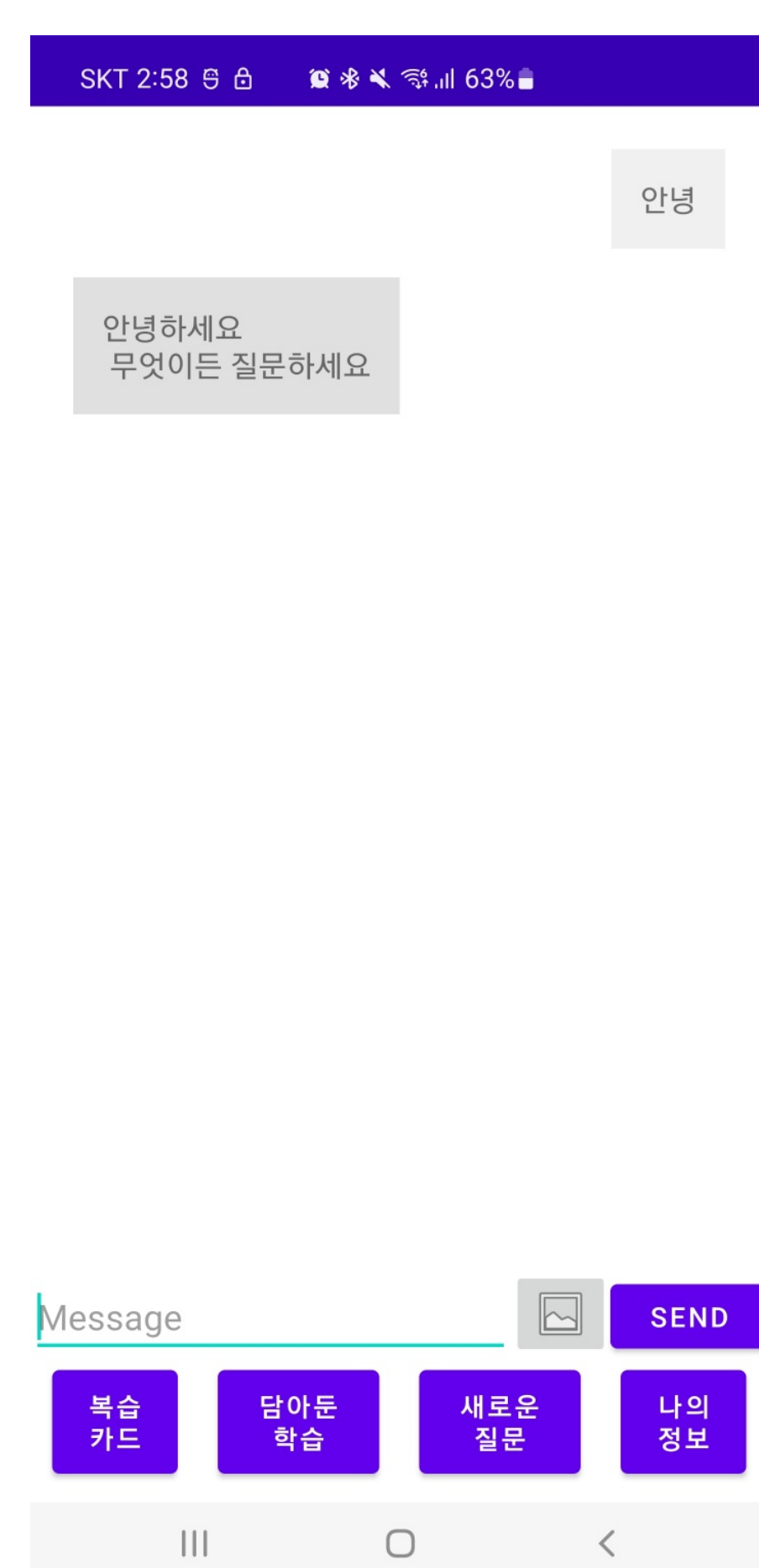


표 1. 기본 VQA 모델과 GEL-VQA 모델의 성능 비교

#### 일반 VQA 모델 vs GEL-VQA 모델 성능 비교 및 분석

- 일반적인 VQA 모델인 Baseline과 외부 지식을 활용하는 GEL-VQA 모델의 입력 언어별 테스트 성능을 비교한 표
- 모든 언어(한국어, 영어, 이중언어)에서 GEL-VQA의 성능이 BASELINE보다 약 24~29% 높은 것을 확인. 이러한 실험 결과는 GEL-VQA 모델이 효과적으로 외부 지식을 활용한다는 것을 의미함



#### 지식 베이스를 활용한 VQA 유아 교육 애플리케이션

- GEL-VQA 모델을 활용하여 만든 유아 교육 애플리케이션으로 ‘티라노사우르스’에 대해 질문을 했을 때 외부 지식에 대한 질문에 대해서도 적절하게 답변함

### 기대 효과

- GEL-VQA 모델을 활용한 안드로이드 애플리케이션은 유아의 시각적 미디어에 대한 **궁금증을 실시간으로 해결**함으로써 호기심을 유발하고 탐구 심리 향상
- 교실에서 뿐만 아니라 다양한 환경까지 **유아의 학습 기회 확장** 가능. 가정, 야외 활동, 여행 중 등 다양한 활동에서 궁금증을 해소하며 학습을 촉진
- 외부 지식을 활용하기 때문에 **단순한 질의 응답 이상의 깊이 있는 정보**를 얻을 수 있음