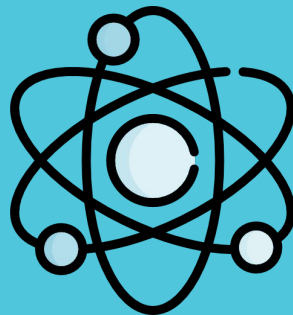




**Harvard** John A. Paulson  
**School of Engineering**  
and Applied Sciences

# ScienceTutor



October 2023

AC215 Project Midterm Presentation

Sijia Li, Ziqing Luo, Yuqing Pan, Jiashu Xu, Xiaohan Zhao

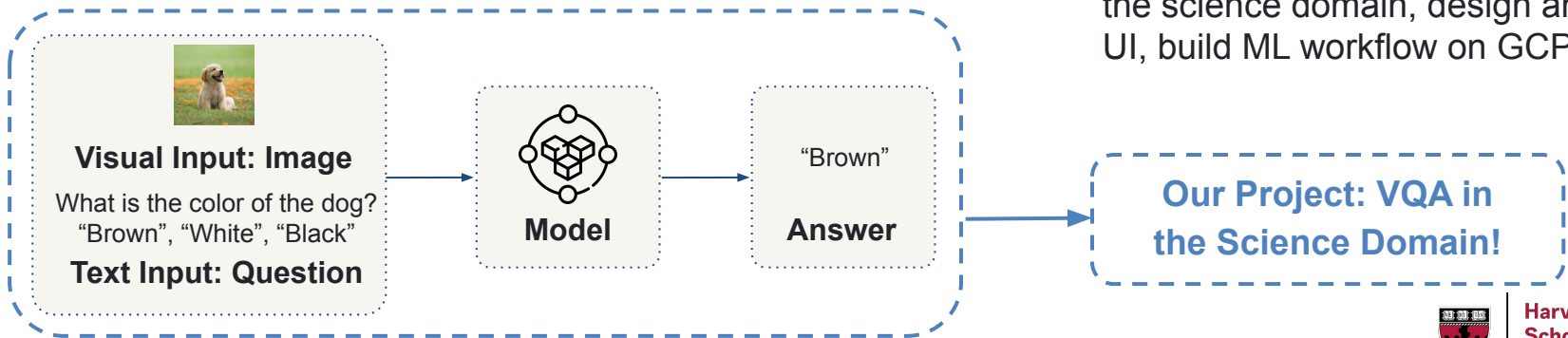
# Project Overview

## Background and Motivation:

- Growing interest in multi-modal models and their applications
- Visual Question Answering (VQA)
  - Provide answers to questions about input images in natural language





## Project Goal:

- Develop an educational application
  - Provides answers to science questions to children via a chatbot and potentially other features
- Technical objectives:
  - Collect VQA data, finetune LLaVA on the science domain, design and deploy UI, build ML workflow on GCP



# Data Source: ScienceQA

- ~21k multimodal multiple choice questions from elementary and high school science curricula
- **Relevant attributes:**
  - Image, question, choices, answer, subject, topic, category
- Highly relevant to our project
- High quality, mature VQA dataset

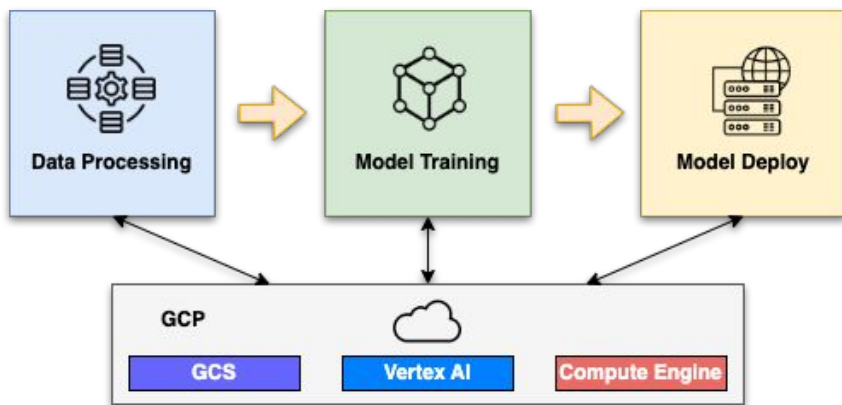
<b>Social Science</b>		<b>Geography</b> State capitals Geography Maps Oceania: geography Physical Geography The Americas: geography Oceans and continents Cities States	 <b>History</b> Colonial America English colonies in North America The American Revolution <b>World History</b> Greece Ancient Mesopotamia World religions American history Medieval Asia	 <b>Civics</b> Social skills Government The Constitution
		<b>Biology</b> Genes to traits Classification Adaptations Traits and heredity Ecosystems Classification Scientific names Heredity Ecological interactions Cells Plants Animals Plant reproduction	Velocity and forces Force and motion Particle motion and energy Heat and thermal energy States of matter Kinetic and potential energy Mixture	<b>Economics</b> Basic economic principles Supply and demand Banking and finance <b>Global Studies</b> Society and environment
<b>Language Science</b>		<b>Writing Strategies</b> Supporting arguments Sentences, fragments, and run-ons Word usage and nuance Creative techniques Audience, purpose, and tone Pronouns and antecedents Persuasive strategies Editing and revising Visual elements Opinion writing	<b>Earth Science</b> Weather and climate Rocks and minerals Astronomy Fossils Earth events Plate tectonics	<b>Chemistry</b> Solutions Physical and chemical change Atoms and molecules Chemical reactions
		Sentences and fragments Phrases and clauses <b>Figurative Language</b> Literary devices	 <b>Engineering</b> Designing experiments Engineering practices <b>Units and Measurement</b> Weather and climate	 <b>Physics</b> Rhyming <b>Reference</b> Research skills

[Link to Dataset](#)

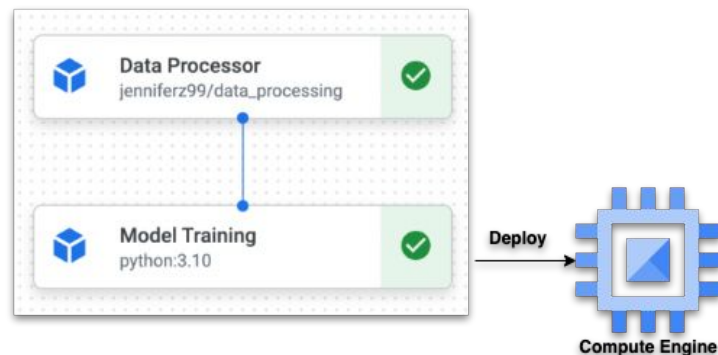


**Harvard** John A. Paulson  
**School of Engineering**  
 and Applied Sciences

# Current Progress



Implement containerized data management  
and model training pipeline



Build ML workflow with Vertex AI  
and Compute Engine



# Application Pipeline - Data Processing

- **Prepare Data in LLaVA Format**
  - Extract information like questions, context, and choices
  - Create conversations between human and AI
- **Upload to GCS for data version control**
- **Upload to Hugging Face**

ac215-sciencetutor

Location	Storage class	Public access	Protection
us-central1 (Iowa)	Standard	Not public	None

**OBJECTS** CONFIGURATION PERMISSIONS PRO

Buckets > ac215-sciencetutor

[UPLOAD FILES](#) [UPLOAD FOLDER](#) [CREATE FOLDER](#) [TR](#)

Filter by name prefix only ▾ **Filter** Filter objects and folders

<input type="checkbox"/>	Name	Size	Type
<input type="checkbox"/>	<a href="#">ScienceQA-LLAVA/</a>	—	Folder
<input type="checkbox"/>	<a href="#">dvc/</a>	—	Folder

Datasets: cnut1648, **ScienceQA-LLAVA** 0

**Dataset card** Files and versions Community 1

**Dataset Viewer** Auto-converted to Parquet [Go to dataset viewer](#)

Split: train (12.7k rows) ▾


Search this dataset

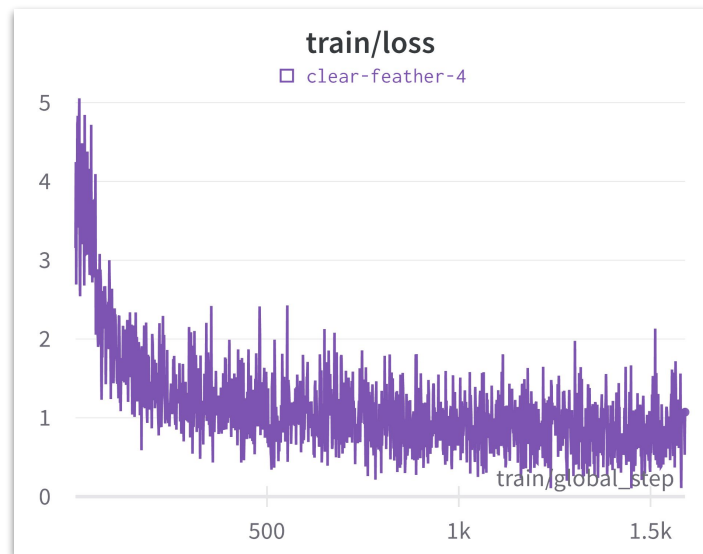
image	conversations	question	context
image	list	string · lengths	string · lengths
1		14	862
	[ { "from": "human", "value": "Context:..."	Which of these states is farthes...	N/A
	[ { "from": "human", "value": "Context: The..."	Identify the question that Tom...	The passage describes :



# Application Pipeline - Model Training

## Training Steps:

- **Fork** the **LLaVA** repository to  `cnut1648 / LLaVA`
- **Customize** the model to get compatible with preprocessed **LLaVA-ScienceQA** dataset
- **Train** 1 epoch and report to **wandb** by running `LLaVA/llava/train/train_mem.py`
- **Upload** model checkpoints to **GCS** / Hugging Face



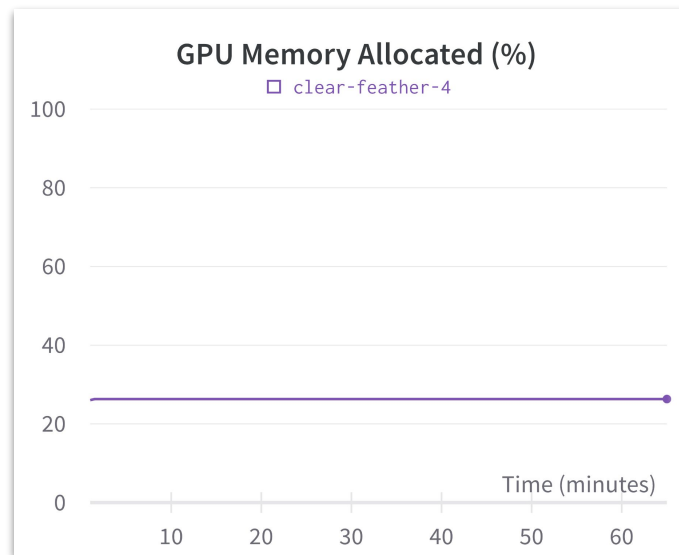
W&B Logging of Training Loss



# Application Pipeline - Model Training

## Optimization techniques to reduce memory usage:

- Choose smaller model (LLaVA 7B instead of 13B)
- Deepspeed ZERO-2 for multi-GPU
- LoRA (Low-Rank Adaptation)
- bf16 (bfloat16 floating-point format)
- Gradient checkpointing
- tf32 (TensorFloat-32 precision format)



W&B Logging of Memory Allocation



# Application Pipeline - Model Deployment & Inference

## Deployment:

- **Vertex AI is not suitable:** Vertex AI builds an API endpoint for a model but we have our **own service and API** (a web server)
- Customize **web\_server** Docker container and push image to **DockerHub**
- Deploy the container in an instance of **Google Compute Engine**

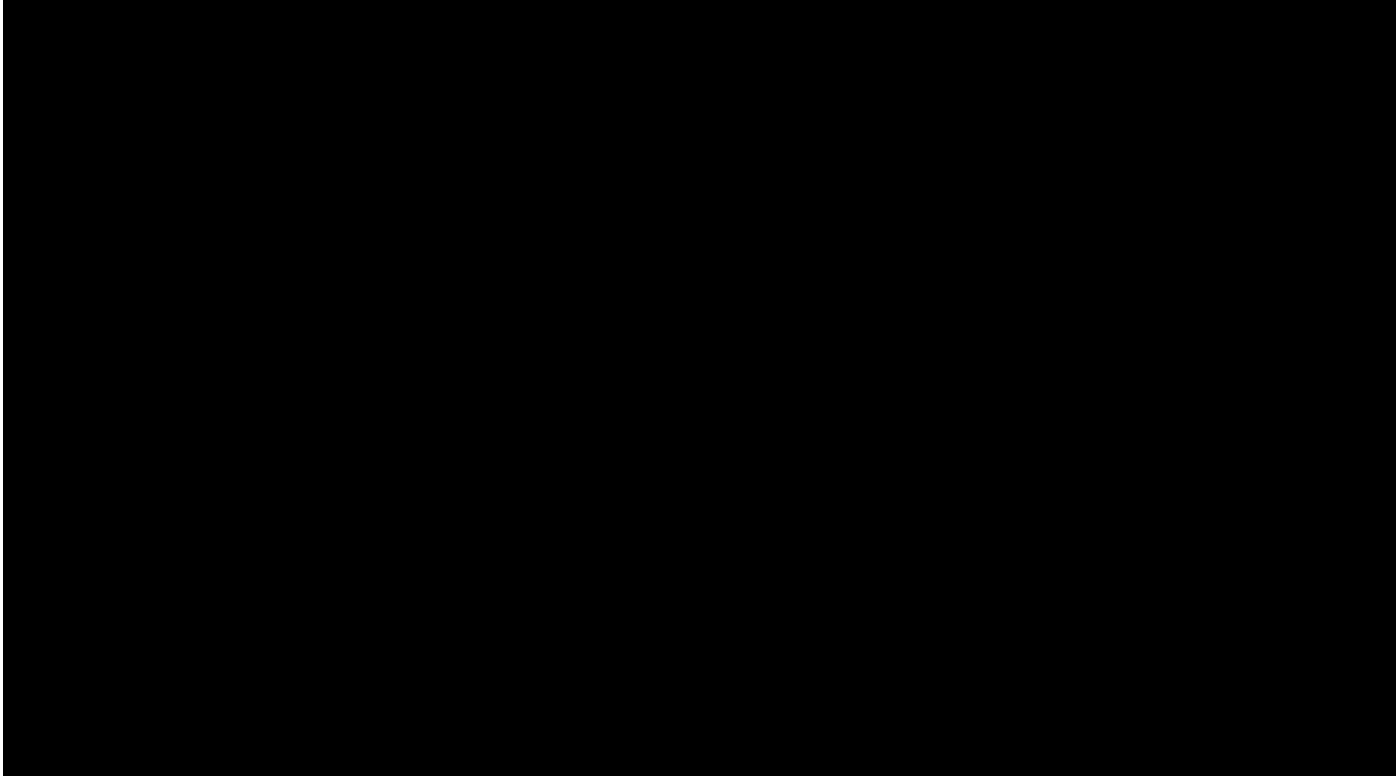
## Inference:

- Utilize **multi-GPU** inference and load the **4-bit quantized** model
- Achieve **accuracy 65.53%** on the testset of **LLaVA-ScienceQA**







# Live Demo

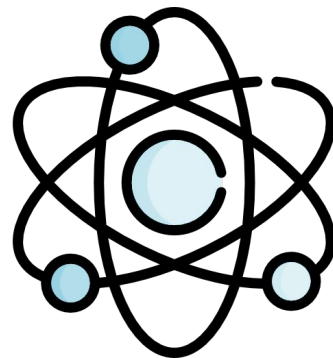


**Harvard** John A. Paulson  
**School of Engineering**  
and Applied Sciences

# Next Steps

- Customize UI with `gradio.app`
- Integrate potential **extensions**:
  - Quiz me 
  - Recommend Resources 
- Set up **CI/CD pipelines** to facilitate users' access to the app
- Scale with Kubernetes and deploy with Ansible

**Science  
Tutor**





**Harvard** John A. Paulson  
**School of Engineering**  
and Applied Sciences

# Thank You!

# Agenda

1. **Project Background and Overview (50s)**
2. **Walk-Through of Current Progress**
  - **Overview: Application Pipeline Flow; Vertex AI Pipeline Flow (25s)**
    - Data Processing (25s)
    - Model Training (30s)
    - Model Deployment & Inference (30s)
  - **Live Demo (40s)**
3. **Next Steps (30s)**
  - **Chatbot Web App**
  - **Potential Extensions**



# Project

## Background and

- Growing in models and
- Visual Que

- Provide input



## Visual Input: Image

What is the color of the dog's fur?  
"Brown", "White", "Black"

## Text Input: Question

+ New Chat

Does this passage...

What is the capital of...

Is the following trait...

What is the probability...

Select the organism in...

In this food chain, the...

Compare the average...

Which organ controls...

?

Quiz Me!

👍

Recommend Resources!

Science Tutor

Sample A

Mass of each particle: 100g  
Average particle speed: 1,500 m/s

Sample B

Mass of each particle: 400g  
Average particle speed: 1,500 m/s

Compare the average kinetic energies of the particles in each sample. Which sample has the higher temperature?

"neither; the samples have the same temperature", "sample A", "sample B"

"sample B"

Can you provide some explanation?

The particles in both samples have the same average speed, but each particle in sample B has more mass than each particle in sample A. So, the particles in sample B have a higher average kinetic energy than the particles in sample A.

📷

Add Image

Type your question...

➤

# Application Mockup Design

(Subject to Change)

## Final application

s to science  
dren via a chatbot  
enhance  
gs

VQA in  
omain!

Harvard John A. Paulson  
School of Engineering  
and Applied Sciences