GDGHack!

Build with AI Hackathon

May 2 - 5 · ICC Conference Hall

Study Guide

Partners

JAAI JUST ADD AI

lector.ai

Red Bull

What is a VLLM?

A vision-language model is a fusion of vision and natural language models. It ingests images and their respective textual descriptions as inputs and learns to associate the knowledge from the two modalities. The vision part of the model captures spatial features from the images, while the language model encodes information from the text.

The data from both modalities, including detected objects, the spatial layout of the image, and text embeddings, are mapped to each other. For example, if the image contains a bird, the model will learn to associate it with a similar keyword in the text descriptions.

How to make one?

As it may take a long time to implement a VLLM from scratch, our hackathon participants should use Google Technologies either for the Language Model, Vision Model or both. It is also possible to find already existing VLLMs and attempt to fine-tune them for a given task.

They can use other pre-trained Vision Models such as ResNet, ViT (Vision Transformers), or CLIP (Contrastive Language-Image Pretraining), or other pre-trained LLMs, such as GPT, BERT, as long as they use at least one Google Technology. The Vision Models are meant to convert raw images into vector representations (embeddings), which capture information about the image content. These embeddings can be fused with embeddings from the pre-trained Language Models, which transform text into input and then generate language-based outputs.

After processing and filtering the data, our participants should use it to train and test the LLM and the vision models. They then use the vision model as the image encoder and the language model as the text encoder. Finally, they have to use a fusion mechanism to create the VLLMs and fine-tune their model for the given tasks.

Possible Datasets

Our participants can use Kaggle (www.kaggle.com) as a resource for finding the datasets they need. They have to decide which datasets are useful for completing their task and training the VLLM. Other possible resources can be: LAION-5B, PMD, VQA, ImageNet, and OpenML. We, of course, encourage our participants to find their own additional sources and support them in doing so.

kaggle

Things to study beforehand!

This section will introduce you to four key Google AI technologies you can use for various AI applications. These technologies offer powerful tools for machine learning, AI model deployment, and generative AI capabilities.

The AI technologies covered in this guide are:

- Vertex AI
- Google Generative AI API
- Gemini & Gemma AI
- AI Studio

This guide provides a general overview of each technology and its potential applications, but you are not restricted to using only these!

Below are just a few examples of technologies you can use, but of course you are not limited to these! These are some of our recommendations to get you started.

- **Gen AI API**
- **Gemini**
- **Gemma AI**
- **AI Studio**
- **BERT**
- **Pytorch/TensorFlow**

Vertex AI – The AI Development Hub

Vertex AI is Google Cloud's unified AI platform that enables developers and data scientists to build, deploy, and manage machine learning models efficiently. It provides an end-to-end ML workflow, integrating data processing, model training, and deployment within a single environment.

Role in VLLM Development:
Vertex AI is Google Cloud's all-in-one platform for training and deploying machine learning models. It provides essential tools for managing the **end-to-end** lifecycle of a VLLM.

- **Data Preparation & Preprocessing:** Use Google Cloud Storage and BigQuery for dataset management.
- **Model Training:** Train custom VLLMs using **TPUs/GPUs** for high performance.
- **Fine-Tuning & Optimization:** Leverage **Vertex AI Training & AutoML** to refine models.
- **Deployment & Scalability:** Deploy VLLMs using **Vertex AI Endpoints** for real-world applications.

Why Vertex AI?
- Seamless integration with Google's compute power (TPUs & GPUs).
- MLOps tools for automation and scaling.
- Easy model monitoring and optimization.

Resources:
- Vertex AI Overview: https://cloud.google.com/vertex-ai
- Quickstart Guide: https://cloud.google.com/vertex-ai/docs/start
- Vertex AI Model Training: https://cloud.google.com/vertex-ai/docs/training
- Vertex AI Tutorials: https://cloud.google.com/vertex-ai/docs/tutorials

Google Generative AI API – Powering VLLM Applications

The **Google Generative AI API** provides access to state-of-the-art generative AI models, allowing developers to integrate advanced AI capabilities into their applications. It enables text generation, image synthesis, and other generative tasks.

Role in VLLM Development:
The **Google Generative AI API** provides direct access to Google's pre-trained generative models. Instead of training a VLLM from scratch, developers can **fine-tune and extend** existing models.

1. **Text Generation & Understanding:** Base VLLMs on powerful generative models.
2. **Custom Model Fine-Tuning:** Adjust AI responses with proprietary datasets.
3. **Multimodal Capabilities:** Enable VLLMs to work with **text, images, and code**.
4. **API-Based Integration:** Seamlessly integrate VLLMs into web and mobile applications.

Why Use the Generative AI API?

- Reduces computational costs of training from scratch.
- Provides instant access to state-of-the-art models.
- Can be extended for specific use cases (e.g., chatbots, summarization, creative writing).

Resources:

- Google Generative AI API Overview: https://ai.google.dev/
- Getting Started with Generative AI: https://ai.google.dev/tutorials
- API Documentation: https://ai.google.dev/docs

AI Studio – Experimenting & Prototyping

AI Studio is a user-friendly platform for experimenting with Google's AI models and developing AI applications without extensive coding experience. It provides an interactive environment for building and testing AI solutions.

Role in VLLM Development:
AI Studio provides a **no-code/low-code** environment for experimenting with AI models, making it useful for the **rapid prototyping** of VLLMs.

- **Testing AI Behaviors:** Validate model outputs before full-scale training.
- **Prompt Engineering:** Fine-tune text generation and improve responses.
- **API Exploration:** Experiment with Google AI's capabilities before building a full VLLM.
- **Deployment Previews:** Evaluate VLLMs in real-world applications.

Why Use AI Studio?
- Fast prototyping without extensive coding.
- Allows testing different AI models before committing to training.
- Simplifies the development process for beginners.

Resources:
- Google AI Studio: https://aistudio.google.com/
- Getting Started with AI Studio: https://ai.google.dev/tutorials/aistudio_quickstart
- Google AI Studio API: https://ai.google.dev/docs/aistudio

Gemini & Gemma AI

Gemini and Gemma AI are Google's advanced AI models designed for various applications, from generative AI to natural language processing and multimodal tasks.

Role in VLLM Development:

Gemini and **Gemma AI** are Google's next-gen AI models that can serve as a foundation for a VLLM.

Gemini AI:

- A powerful **multimodal** AI model (text, images, code, video).
- Ideal for training VLLMs that handle **complex reasoning tasks**.
- Can be fine-tuned for domain-specific applications.

Gemma AI:

- **Lightweight**, **open-weight models** optimized for efficiency.
- Suitable for **local VLLM development** with limited resources.
- Can be deployed on custom hardware for edge AI applications.

Why Use Gemini & Gemma?

- Gemini's multimodal capabilities make it an excellent base for VLLMs.
- Gemma's lightweight nature helps create **resource-efficient** LLMs.

Resources:

- Gemini AI Overview: https://deepmind.google/technologies/gemini/
- Google's Gemini AI Models: https://blog.google/technology/ai/google-gemini-ai/
- Gemma AI Overview: https://ai.google.dev/gemma
- Gemma Model Weights & Code: https://github.com/google/gemma