fAIrConnect

A fair chance to get the opportunities you deserve

AlphaBetaGemma

jahnvikumari@   rajmaheshwari@   amishaaggarwal@
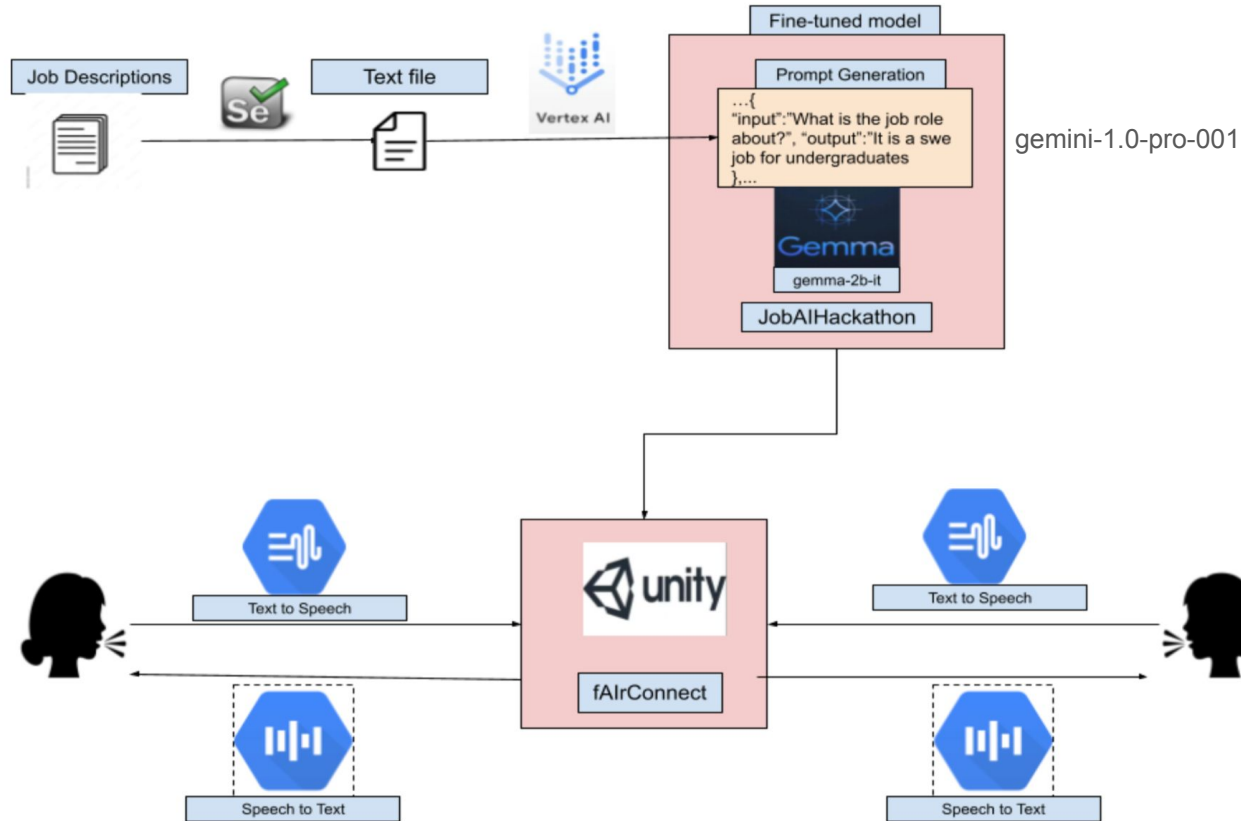
Problem statement

fAIrConnect is a virtual career fair with an AI-powered company spokesperson. In-person career fairs are often not accessible to people due to geographical, physical and financial constraints. fAIrConnect aims to bring together AI recruiters and candidates in a metaverse.

Advantages of using fAIrConnect

- Consistent and inclusive
- Available
- Accessible (Multilingual)
- Scalable

Solution Pipeline

We have used Google Cloud Technologies

Cloud Products We Used

Vertex AI for generating dataset to finetune the generative LLM.
(For training, we used the T4 GPU available for free in colab.)

GCP's Speech to text API

GCP's Text to Speech API

Pipeline Overview

Backend (ML model)

- Preparing a custom-dataset for a given company from their website.https://huggingface.co/jahnvik13/JobAIForHackathon
- Fine-tuning gemma-2B-it model with the data to have an AI spokesperson for the company.

Frontend (3D Unity metaverse frontend):

- User walks up to the AI spokesperson and asks a question, the voice data is converted to text.
- The text is fed into the AI model to generate an answer.
- The answer is spoken by the AI model and visible in text.

```
GemmaForCausalLM(
  (model): GemmaModel(
    (embed_tokens): Embedding(256000, 3072, padding_idx=0)
    (layers): ModuleList(
      (0-27): 28 x GemmaDecoderLayer(
        (self_attn): GemmaSdpaAttention(
          (q_proj): Linear4bit(in_features=3072, out_features=4096, bias=False)
          (k_proj): Linear4bit(in_features=3072, out_features=4096, bias=False)
          (v_proj): Linear4bit(in_features=3072, out_features=4096, bias=False)
          (o_proj): Linear4bit(in_features=4096, out_features=3072, bias=False)
          (rotary_emb): GemmaRotaryEmbedding()
        )
        (mlp): GemmaMLP(
          (gate_proj): Linear4bit(in_features=3072, out_features=24576, bias=False)
          (up_proj): Linear4bit(in_features=3072, out_features=24576, bias=False)
          (down_proj): Linear4bit(in_features=24576, out_features=3072, bias=False)
          (act_fn): GELUActivation()
        )
        (input_layernorm): GemmaRMSNorm()
        (post_attention_layernorm): GemmaRMSNorm()
      )
    )
    (norm): GemmaRMSNorm()
  )
  (lm_head): Linear(in_features=3072, out_features=256000, bias=False)
)
```

Impact

- **Accessibility**: Expanded access to career opportunities regardless of geography, financial resources for travel, or physical limitations.
- **Low costs**: Organization gets triple benefits of increased reach, reduced costs and attracting diverse talent than in-person fairs.
- **Scalability and Concurrency**: Streamlined operations with no wait time. The design itself is concurrent. This increases the scalability.
- **Availability**: The spokesperson won't go on lunch breaks :)
- **Consistency**: No human bias is introduced including tiredness and repetition.

Innovation and Novelty

- **Novelty**: While virtual career fairs exist, a fully AI-driven spokesperson does not.

- **Market differentiation:** Stands out in a crowded field by addressing scalability AND accessibility, with the multilingual support being a major plus.

- **Technical Challenge:** Creating an AI capable of conversational interactions across formats (speech, text, potentially sign language) is a complex and innovative pursuit.

- Existing AI models include Olivia by Paradox and Mya, both of which are used to screen candidates and schedule interviews. None of the existing models is focussed on career fairs and accessibility

Project Demo

- https://drive.google.com/file/d/1zFLsbfgs_RRYQc65B6GZrx7zu2YK2DSa/view?pli=1
- https://drive.google.com/file/d/1Xk7CeF3OcTQDWmbz8xrntiVcM9dBd_5K/view?usp=sharing

Challenges we faced

1. Metaverse development in Unity and C#

2. quotas absent and exceeded for free tier limit of GCP. For example custom model serving requires Nvidia L4 GPUs

3. Integrating unity with AI models and cloud APIs

☐	Custom model serving Nvidia A100 80GB GPUs per region	Quota	region : us-central1
☐	Custom model serving Nvidia A100 GPUs per region	Quota	region : us-central1
☐	Custom model serving Nvidia H100 80GB GPUs per region	Quota	region : us-central1
☐	Custom model serving Nvidia K80 GPUs per region	Quota	region : us-central1
☐	Custom model serving Nvidia L4 GPUs per	Quota	region : us-central1

Future Scope

- Multiplayer will make career fair more immersive.
- Add capability for sign language
- Fine tune the dataset on a larger data for more companies
- AI based screening of candidates based on their experience and skills