Seja bem vindo a avaliação técnica para o cargo de computer vision machine learning engineer. A seguir será apresentada uma simulação de um projeto que enfrentamos no dia a dia.

Proposta:

Um grande cliente do segmento fashion europeu precisa inovar a maneira como ele interage com seu público, após ler um post no Linkedin dizendo que machine learning é uma mina de ouro esse cliente entrou em contato com o google buscando apoio para fomentar essa tecnologia. Neste momento entra a Hvar Consulting para auxiliar nessa jornada.
Em um brainstorming de ideias uma chamou a atenção dos executivos, a proposta seria criar um sistema que fosse capaz de receber imagens / vídeos e através de machine learning identificar quais itens fashions existem naquela imagem e encontrar itens similares no catálogo da empresa.

A abordagem escolhida para essa PoC será VQA (Visual Question Answering) onde um algoritmo de machine learning deve ser capaz de responder através de prompts (inputs de texto como entrada) informações sobre o domínio fashion.

Exemplos de possíveis prompts:

- Quais itens estão contidos nesta imagem;

- Qual é a cor da calça;
- Descreva os detalhes na blusa;
- Qual é a cor da bolsa;

Objetivos:

- Realize uma avaliação dos modelos open source disponíveis que atendem a tarefa proposta e justifique sua decisão;
    - *Foram escolhidos para comparativo os modelos **blip-vqa-base, Salesforce** (https://huggingface.co/Salesforce/blip-vqa-base) e **matcha-chatqa, Google** (https://huggingface.co/google/matcha-chartqa). Ambos modelos VQA são bem avaliados pela comunidade e semelhantemente desenvolvidos em PyTorch.*
- Desenvolva uma metodologia para comparar a performance de diferentes modelos open source disponíveis;
    - *Procurou-se estabelecer a acurácia de cada modelo, tendo por base um comparativo semântico entre a resposta predita por um modelo em relação a resposta esperada para determinada pergunta.*
    - *Foi feito uso da **Distância de Levenshtein**. Em síntese, trata-se de uma métrica para estimar o número mínimo de operações necessárias para transformar uma string em outra, ou seja, a resposta predita para a resposta esperada.*
    - *Para o comparativo foi utilizado dataset **Locatized Narratives** (https://google.github.io/localized-narratives/, https://huggingface.co/datasets/vikhyatk/lnqa). Dentre vários atributos a base conta com imagens, perguntas e respostas esperadas para cada um dos registros. Para fins de objetividade foi utilizado um subdataset com cerca de 500 registros. A base totaliza 1,5 M de registros.*
    - *Foram obtidos os seguintes resultados:*

	blip-vqa-base	***matcha-chatqa***
mínimo	*13*	*12*
média	*50*	*52.7*
variância	*701.7*	*719.4*
desvio	*26.5*	*26.9*

    - *Com valores de média e ligeiramente menores, o modelo **blip-vqa-base** obteve uma melhor performance tendo em vista a metodologia adotada para comparativo.*

- Realize um finetune no modelo escolhido utilizando datasets open sources (DeepFashion,SaffalPoosh/deepFashion-with-masks, ldhnam/deepfashion_controlnet, glami-1m, etc).
    - *Como dito anteriormente, o fine-tune foi realizado no modelo **blip-vqa-base;***
    - *O modelo foi retreinado e adaptado para o dataset **deepFashion_with_masks** (https://huggingface.co/datasets/SaffalPoosh/deepFashion-with-masks). Por sua fácil aquisição e por atender os requisitos de atributos usados no comparativo de modelos.*
    - *Novamente para fins objetivos, o dataset usado foi reduzido para 1000 registros. Ao todo o dataset conta com 40 mil registros. E também seu treino foi aplicado em 10 épocas*
- Desenvolva uma interface gráfica usando gradio que seja capaz de receber inputs de imagens, prompts e output com a resposta do modelo.
    - Foi desenvolvida uma interface utilizando Streamlit, disponível em https://pocvapp-8jxedqwct7p9v3zphxhtu5.streamlit.app/ .
    - Entretanto por contratempos a integração da UI com o modelo não foi realizada.
- Documente o projeto utilizando o github;
- Para os experimentos utilize o huggingface hub para armazená-los
    - Modelo com fine-tune está disponível no Hungging Face: https://huggingface.co/davicn7/hvar_vaq_model .
- Bônus: Adapte a aplicação para receber entradas de vídeos

Eventuais dúvidas fiquem à vontade para solicitar via email ou agendar uma reunião rápida para tirar dúvidas, estarei disponível 😁. Gostaria só de reforçar que estamos interessados em avaliar sua capacidade de adaptação não necessariamente o resultado alcançado. Desejo que o projeto seja divertido e interessante.