SES - Smart Evaluation Solution

Authors: Le Duc Minh, Le Nguyen Dang Khoa, Tran Ngoc Dai, and Pham Minh Man
Github: https://github.com/MinLee0210/Smart-Evaluation-Solution.git

Introduction

The project is part of the AngelHack competition. We took part in the track of digital empowerment, in which we played the role of a member of the Digital & Technology (D&T) team of a company. We used the suggested tools to solve several business problems. In our case, we serve for Heineken, and our task is to use AI to detect as many features as possible, so that we can use that information to evaluate the effectiveness of a marketing strategy as well as the branding awareness of the brand.

The structure of this document is as follows:

1. **About the project**: We have a brief summary of our projects.
2. **Technologies**: We present technologies that we have developed for our project as well as explain why we used them.
3. **Methodology**: We discuss how we develop our project in detail.
4. **Conclusion**: We summarize what we have achieved during the project, and suggest how it can be improved in the future.

About the project

Heineken Vietnam, the country's top beer maker, is trying to transform the brand experience for its customers. From local retail stores to beloved hangout areas with friends, the brand's promotional materials (banners, posters, LED signs, and so on) are everywhere, adding to engaging and memorable consumer experiences. However, manually assessing and evaluating these locations using photographs is becoming time-consuming and expensive for the organization. As a result, using AI becomes the best idea because it can handle a large amount of information and extract information as detailed as humans.

As a member of the D&T team, we want our solution to be:

- **F**unctionality: The project meets the requirements and specifications that it was designed for, and it behaves as expected when it is used in its intended environment.
- **U**sability: The project is easy to use and understand, and it provides a positive user experience.
- **C**ompetency: The project is always catching up the latest technologies, concepts, etc. so that it becomes mind-blowing.

Following the **FUC** guideline, we believe that our project will evolve into a prominent framework for analyzing a marketing campaign, especially for Heineken.

Heineken is tapping into the power of AI to transform its marketing and customer experiences. Personalization takes center stage. AI can analyze customer purchase history to recommend similar Heineken beers or suggest pairings with snacks, creating a more enjoyable drinking experience. Social media data can be mined to understand customer demographics and interests, allowing Heineken to deliver targeted advertising on relevant platforms. AI-powered chatbots can answer customer queries about Heineken products and even handle simple transactions, all contributing to a smooth and efficient brand interaction.

Beyond personalization, AI unlocks exciting new avenues for engagement. Imagine Heineken analyzing social media posts to see what people are drinking and then targeting them with relevant ads for Heineken beers. AR/VR experiences could let customers virtually "try" different Heineken brews or explore the brand's history in an interactive way. AI can even be used to create personalized content for each customer, optimize ad campaigns for maximum impact, and identify social media influencers who perfectly align with Heineken's target audience. The possibilities are vast, and for Heineken, AI represents a powerful tool to create a more engaging and effective marketing strategy.

Technologies

Our technique uses a multi-model architecture for business solutions, known as **MBS**. In particular, we favor compact models (often with less than 1.5 B parameters) since they have proven to be superior after fine-tuning for downstream

tasks. From the perspective of a businessman, we believe that our product should be affordable (supply and demand) and function properly for its intended purpose. Tiny models appear to be a "de-facto" choice since they require no light and perform exceptionally well after fine-tuning for downstream applications.
For models, we aim at cuting-ede AI-based tools that can optimize the performance of the image capture process. In detail, we use:

- **CLIP Model:** with the power of the constrastive learning, we can find the context of the image. For instance, normally, Vietnamese people drink with their friends at the weekends; the model is able to capture the context in which we set and predict the image's concept.
- **Object Detection Model (OD)**: this model is the core of the project, allowing for the detection of features (Heineken logo, person, promoter, etc.). We use DINO, which is an outstanding model for zero-shot learning object detection.
- **Optical Character Recognition (OCR)**: It is a technology that recognizes text within a digital image. It is commonly used to recognize text in scanned documents and images. We use it to help the OD model enrich the detection process. We use *DINO* families.
- **LLMs**: this AI builds upon the knowledge of previous models to provide informative text-based answers. This makes the system accessible to everyone - both non-technical users seeking natural interaction and tech professionals seeking to broaden their knowledge beyond their specialty. The model not only delivers insights but also explains them clearly, making them truly valuable.

To conclude, our system consists of two components: the *Imagen* and the *Intepretor*. The former is in charge of extracting desirable objects from an image(s), while the latter is in charge of interpreting information so that everyone can readily comprehend the insights it contains. Notably, it may assist us in evaluating the performance of a marketing plan as well as recommending ways to enhance our campaigns.

Methodology

The below graph illustrates the pipeline for extracting information that SES restricts.

1. An user inputs an image and sends to SES.
2. Extract the visual contexts (The *Imagen*):
   a. The OCR model is used to catch characters that relate to beer's brands, especially Heineken.
   b. The CLIP model is used to catch suroundings context.
   c. The image goes to the OD model, with the help of 2 previous models, it detects objects in the image.
3. The visual contexts associated with the role-centric actor (The *Intepretor*): generate a human-like answer that provides insights about the image.
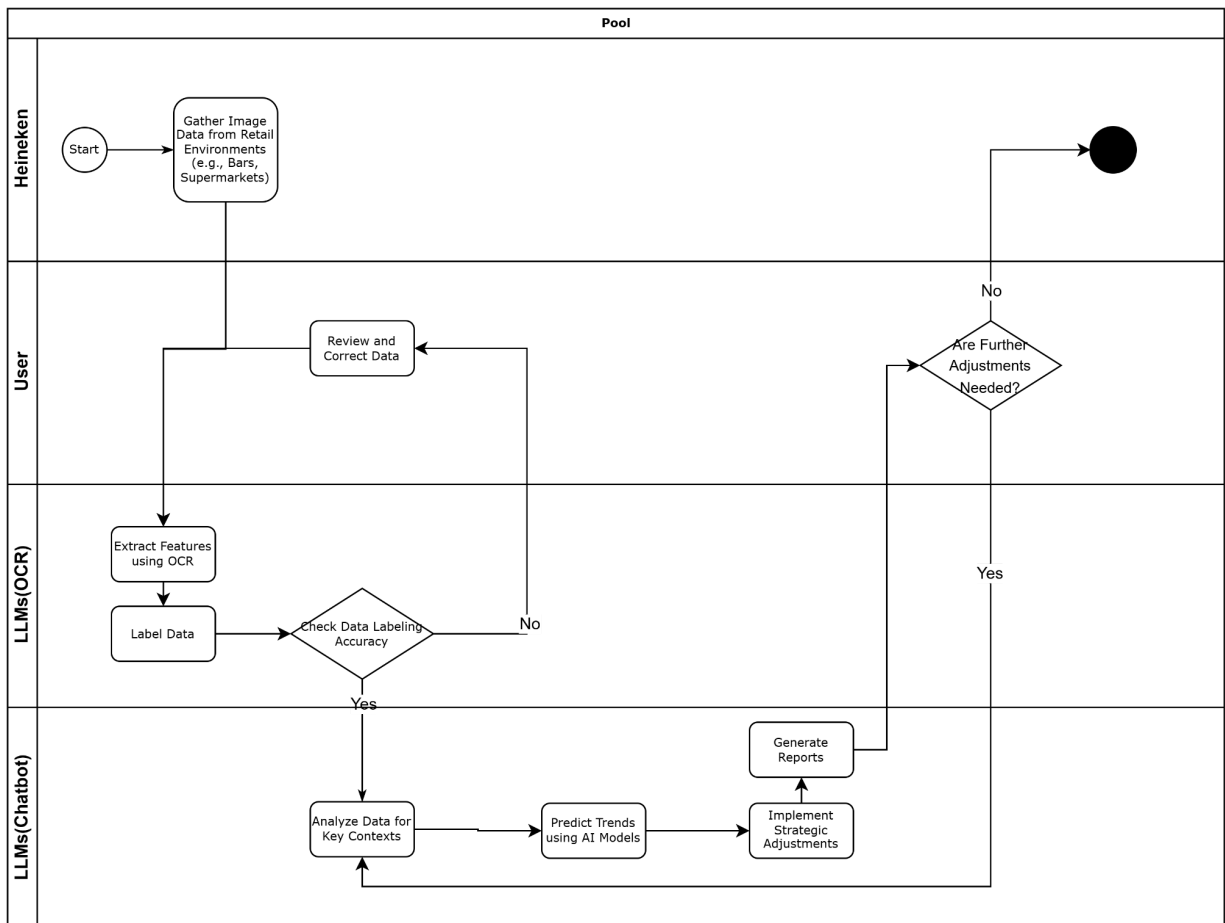
Note: Role-centric means that we give the interpreter a role. The role will be specialized based on the endpoint's characteristics. For example, if our endpoint is

a person from the marketing team, we build up a portfolio that showcases an ideal marketer, the *Intepretor* to learn the information needed to specialize a personality for that person. As a result, the generated answers would have more intimacy with that person but remain professional.

Conclusion

We set up guidelines for our project that help us move steadily toward our objectives.

In addition, we proposed the MBS architecture, which is specialized for business solutions, especially in evaluating branding awareness. The pipeline that makes the architecture happen is described in detail.

For future development, we aim at:

- Improving the *Imagen*: more high-quality labeled data will be fed into the mode, and it will have the capability to process audio formats. We will add a visual language model to illustrate the image with the best description (Gemini Flash, Llava, etc.).
- Improving the *Intepretor*: we apply RAG systems so that it can reach the latest knowledge, providing more useful knowledge for the endpoint user. More prompting techniques would be implemented, promising a clear and concise output.