Controlled Image Generation for Reflecting Eating Habits in Virtual Avatars

Graduate

Aziz Hazeraj

Thashvar Uthayakumar

Problem: The Smart Eating Platform is a website that provides personalized meal plans tailored to user's dietary preferences and physical goals. The objective of this exploratory thesis is to visualize potential changes in a user's physical appearance resulting from adherence to such a plan. To generate these visual transformations, AI-based image generation and editing models should be used. A key component of this work is the evaluation and comparison of different models based on their ability to realistically simulate physical changes driven by dietary input. The resulting images could help motivate users to stay focused on their meal plan.

Approach: In a first step, manual testing of several image generation techniques was conducted to identify models with promising visual and semantic quality. Based on these tests, two open-source models were selected for deeper evaluation: ControlNet and Null-text Inversion. Both were systematically assessed by generating several hundred images using combinations of input images, predefined prompts, and parameter sets. This evaluation led to the identification of an optimal parameter set for each model.
To leverage its state-of-the-art image editing capabilities, we also integrated OpenAI's GPT-4o image generation model into the final system. All three models are supported, allowing users to select one based on preference.
To use these models in the Smart Eating Platform, we developed a FastAPI backend, which uses a Python pipeline. For testing and demonstration, a frontend was implemented using React.js, communicating with the backend via RESTful APIs. The frontend itself takes in a meal plan, the original image and selections of the large language model (LLM) and the image model. As for the LLM, there are two options: OpenAI GPT-4o and LLMHub Qwen3.
The backend is built with FastAPI in Python and acts as the bridge between the interface and the image generation engine. The core backend logic consists of three stages: parameter validation, prompt generation, and image generation. After validating user input, the system analyses the meal plan to generate a textual "reflection in appearance" description, which captures the potential change in appearance. This reflection is then used to construct prompts adapted to the syntax of the selected image model, which finally generates the resulting image (see Fig. 1).

Result: To evaluate the approach, a survey with 9 participants was conducted. Results showed that GPT-4o demonstrated the strongest performance both in rating level and consistency, making it the most reliable option from a user perception standpoint. ControlNet, although less consistent, showed a higher average score than Null-text Inversion, indicating greater practical effectiveness in this specific use case. These results showed that the visual changes are reflected in the images as expected (see Fig. 2 & 3). It also validates the decision to offer GPT-4o, while retaining the open-source pipelines for local or privacy-sensitive deployments. Overall, the system presents a valuable addition to the Smart Eating Platform by enhancing user engagement, providing intuitive visual feedback, and potentially increasing motivation to adhere to personalized dietary plans.

Figure 1: Activity diagram of one image generation request
Own presentment

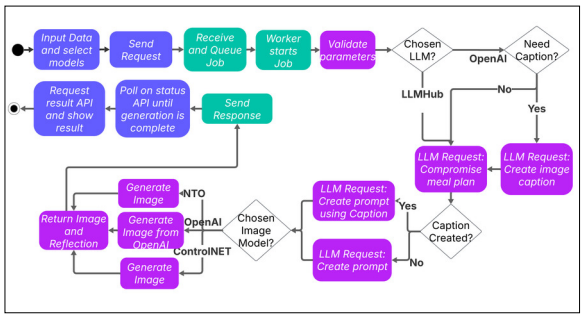

Figure 2: Transformation using ControlNET and an athletic and healthy meal plan. Left: Original, Right: Transformed
Photo by Clive Thibela on Unsplash

Figure 3: Transformation using Null-Text-Inversion and an unhealthy meal plan. Left: Original, Right: Transformed
Photo by Reza Biazar on Unsplash

Advisors
Prof. Dr. Mitra Purandare, Prof. Dr. Christoph Gebhardt

Co-Examiner
Dr. Sammy Christen, The Walt Disney Company (Schweiz) GmbH

Subject Area
Artificial Intelligence