# Assignment

## Description

The overall objective of the assignment is to build a workflow involving multiple processes (or tasks) with multi-modal LLMs. The purpose of the assignment is to assess your problem-solving skills, i.e. how you would generally approach such a problem, and the current level of your technical competencies in Gen AI. The problem is framed within a neuromarketing context, where we want to obtain some insights pertaining to the most visually salient elements in marketing assets (images) and the potential perceptual (cognitive) load typically induced by the assets.

## Details of proposed workflow

**You will be provided with:**

1) Two images, and their corresponding attention heatmaps as predicted by our proprietary Neurons AI technology.
2) Four prompts to be used in the different processes listed below (see “Prompts” section at end of document).

The multiple processes are described below. Note that all the prompts already include the desired JSON format of the output. Your task is to set up the workflow as part of an API (e.g. FastAPI) using Docker.

**Overall flow:**

- In essence, there are three processes, namely A, B & C.
- Inputs to the API constitute an image (used in Processes A & B) and its corresponding heatmap (used in Process B).
- API returns the final JSON output from Process C (see “response-template” section in the prompt).

**Process A (multimodal) - Chain of prompts:**

- Prompt A1 - Provide image and ask model to describe key elements of advert (product, brand, CTA, etc.) and also identify purpose of advert (brand building or conversion).
- Prompt A2 - Provide corresponding attention heatmap and ask model to describe the most visually salient elements based on the heatmap.

**Process B (multimodal) - Single prompt:**

- Prompt B - Provide image and ask model to assess perceptual/cognitive load of the asset.

**Process C (text-based only) - Single prompt:**

- Prompt C - Feed output from Processes A & B and ask model to summarise the information and ensure that the final output has the desired JSON format.

## Further information

**Bonus points for the following:**

- Produce a demo using Gradio or Streamlit.
- Make use of a framework such as LangChain or your own preferred/custom framework for setting up the workflow.

**Final notes:**

- For the multimodal or text-based LLMs, you are encouraged to use any free-to-use models from HuggingFace or any APIs.
- To build the overall workflow, you are welcome to use any tools or framework that you deem suitable. We only have requirements for the inputs and final output of the whole workflow, such that you have the complete freedom to design the connecting blocks in your preferred way.
- Prior to the technical meeting, please send us the link to a repository (e.g. GitHub or GitLab).
- During the technical meeting, please walk us through the core components of the code and the results obtained with the two images provided. If you can demo the application using Gradio/Streamlit, that would be much appreciated.

Thank you very much in advance for investing your time and efforts in working through this assignment. Please schedule the technical meeting at your own convenience, and do not hesitate to let us know if you have any questions (d.ramanah@neuronsinc.com).

## Prompts

For the purpose of the assignment, we are providing rather basic prompts, but you are welcome to improve upon them if you wish.

**Prompt A1:**

```
<role>
You are a Senior Insights Manager with decades of experience, and a background
in marketing.
</role>

<input-overview>
You are provided with an image of a digital advertisement.
</input-overview>

<task>
You have two tasks:
```

1) Provide a detailed description of the advert. In other words, identify and describe the key elements such as the product being advertised, the brand name, and the call-to-action (CTA), where available.
2) Additionally, assess and determine the primary purpose of the advertisement, i.e. whether it is aimed at brand building or aimed at driving conversion.
</task>

<response-template>
Provide the output in the following JSON format
```
[
  {
    "ad_description":$description,
    "ad_purpose":$purpose
  }
]
```
In this format, $description is a placeholder for the description of the advert, $purpose can only be either "brand-building" or "conversion".
</response-template>.

### Prompt A2:

<input-overview>
You are now provided with the attention heatmap of the same image. The attention heatmap illustrates the distribution of attention as predicted by an AI model that was trained on eye-tracking data. Red colour indicates high attention, green implies moderate level and transparent colours mean low attention. Please do not confuse the heatmap colours, i.e. the red, yellow, green blobs etc. with the actual colours of the video frames.
</input-overview>

<task>
You have a single task:
Based on the provided heatmap, identify the most visually salient elements, i.e. the elements that catch the most attention. Please pay special attention to the product being advertised, the brand or logo, and call-to-action (CTA), where available.
</task>

<response-template>
Provide the output in the following JSON format
```
[
  {
    "saliency_description":$description
  }
]
```

In this format, $description is a placeholder for the description of the visually salient elements in the advertisement.

### Prompt B:

```
<role>
You are an expert in applied neuroscience and behavioural psychology.
</role>

<input-overview>
You are provided with an image of a digital advertisement.
</input-overview>

<task>
You have a single task:
Assess the perceptual or cognitive load of the image. This is a measure of the effort required for mental processing based on the visual complexity, such as diversity of colours, presence of patterns and the inclusion of text. In other words, assess how accessible the image will be to a viewer in terms of the brain processing capacity required to interpret and understand the advertisement.
</task>

<response-template>
Provide the output in the following JSON format
```
[
  {
    "cognitive_description":$description
  }
]
```
In this format, $description is a placeholder for assessment of the cognitive load induced by the advertisement in viewers.
```

### Prompt C:

```
<role>
You are a professional writer.
</role>

<input-overview>
You are provided with text descriptions that are outputs from two different multi-modal LLMs.
</input-overview>

<task>
You have a single task:
```

Summarise the text provided from the two different outputs and ensure that the JSON format of the final output matches the requirements provided below.
</task>

<response-template>
Provide the final output in the following JSON format
```
[
  {
    "ad_description":$ad_description,
    "ad_purpose":$ad_purpose,
    "saliency_description":$saliency_description,
    "cognitive_description":$cognitive_description
  }
]
```
In this format, $ad_description, $ad_purpose, $saliency_description and $cognitive_description are placeholders for the summarised versions of the text provided as input.