

Coding Task For Chat & Ad Gen

This task follows up from the previous high-level design task, where we'll focus on several aspects of ad generation - namely, the construction of the ad story and fitting of media elements (i.e. assets) into templates. Due to the time constraints imposed, it is very likely that you'll not be able to finish all sections. Kindly follow this order of priority in your answers, and skip sections of lower priorities - as in, emphasize depth over breadth in tackling the problems:

1. **A: Ad Storyboard Design Agent (Main Question)**
2. **C: Video Cutdown Generation (Main Question)**
3. *C: Video Cutdown Generation (Follow-up/Advanced/Further Discussion Questions)*
4. *A: Ad Storyboard Design Agent (Follow-up/Advanced/Further Discussion Questions)*
5. *B: Asset & Template Fitting Process*

A: Ad Storyboard Design Agent

You're tasked with creating a chatbot to gather basic information about an ad, and follow up with more specific questions as it guides the user through the entire ad generation process. The basic idea progresses through these steps:

1. Welcome message → "What would you like to do today?"
2. User replies "make me an ad" → triggers a routine (Agents or otherwise) to ask user for a series of basic information:
 - a. Ad Duration (in seconds, ranging from 0 (meaning static ad) to ~60s)
 - b. Ad Channel (Facebook, Instagram, TikTok etc.)
 - c. Ad Theme (a short ~50 word description on what the ad should be about)

User may provide all or partial information at this stage, and the chat agent should be able to parse and ask for missing inputs.

3. Once Step 2 is complete, trigger an LLM prompt completion (based on a well-crafted system prompt to describe the LLM's role as an ad writer, followed with user inputs) to generate a brief Ad Concept, which is a ~100 word elaboration of the Ad Theme provided by the user.
 - a. In this section, we have one more piece of information, which is the client's brand description (a paragraph of ~200 words or so describing the brand's value proposition and style etc.) that has been pre-saved and is a simple DB lookup
 - b. An example of the input and output of this step would be:

System Prompt:

You are an expert ad creative agent tasked with writing high-level ad story concepts...you are given information about the client's brand description and ad theme....

User Prompt:

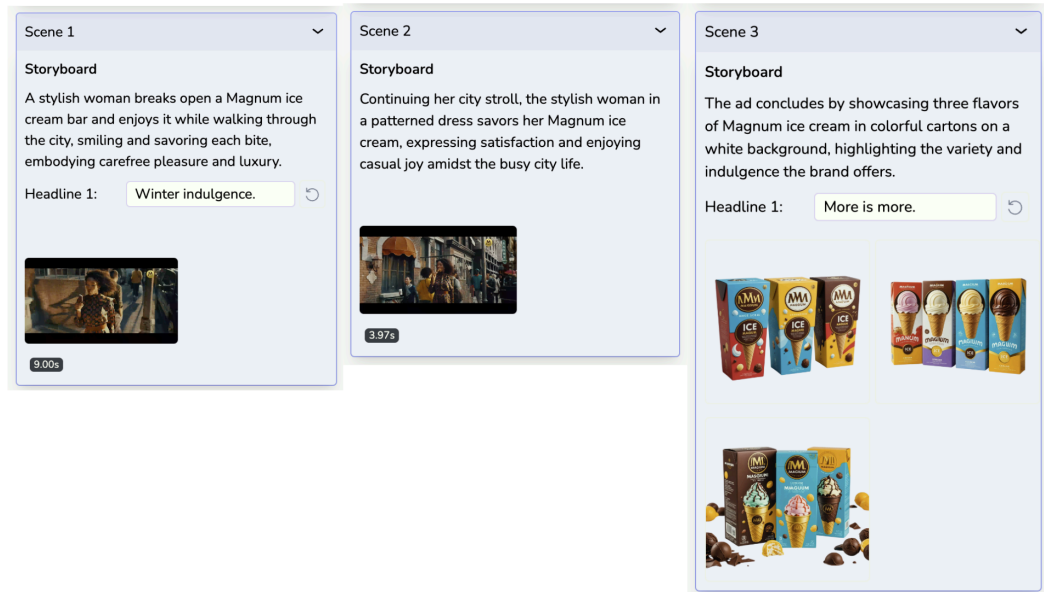
Ad Theme: <the ad theme provided in Step 2 e.g. "ice cream in winter">

Brand Description: <a DB lookup based on current user/client>

Output Ad Concept: <an output ad concept ~100 words, e.g. "Enjoy the delight of delicious ice cream in a winter wonderland...>

4. After Step 3, user will be presented with the generated Ad Concept. User should be prompted to validate it, or, if they dislike it, they may ask the chat agent to rewrite the story. User may or may not provide a reason for why they dislike the initial concept proposal, or direction they have in mind.
5. After confirmation, the process continues to scenewise storyboard generation. In this stage, we will use the input of Ad Duration to determine how many scenes to generate - for

simplicity, we can assume 0s = 1 scene (static ad), <10s → 3 scenes, <30s → 5 scenes, <60s → 7 scenes. This step will trigger another LLM call to generate scenewise storyboard. An example output would be something like in the screenshot below, where, for each Scene, we have a Storyboard:



Note that the story should flow from one scene to the next, and have a beginning, middle, and end.

6. User is again asked for confirmation on whether they like the ad storyboard. At this stage, user may ask for specific alterations e.g. “change scene 1” or “add 1 more scene”. Once the user is satisfied with the final output, they would be prompted for final confirmation, and if they reply “Yes”, we’ll proceed with asset and template fitting.

Based on these steps, design a chat system that is able to handle steps 1 through 6, with this guiding principle of priorities to work out due to the time constraint imposed by this task:

1. Prioritize designing the “happy flow” where user will reply “OK” to validation steps #4 and #6.
2. Implement a form validation cycle in Step #2 if the user did not provide all information when replying.
3. If you have time, work out a rudimentary flow that will be able to regenerate the Ad Concept (with or without factoring in feedback) in step #4 if the user replies with a negative.
4. If you have even more time, work out a proposal on how you would handle the back-and-forth chat-based editing in Step #6 - e.g. user may ask many different things, and may even backtrack all the way to Step #1 (e.g. if the user says “I don’t like any of this, let’s change the Ad Theme”)

B: Asset & Template Fitting Process

This task follows from the output of Task A.

Assume that we have a set of templates. These templates contain multiple “scenes”, where each scene is a pre-designed visual containing placeholders that can be fitted with images, videos, and text assets. An ad is composed of a sequence of scene, where each scene transitions to the next

smoothly via pre-designed animations (these animations can also be applied to individual asset placeholders):

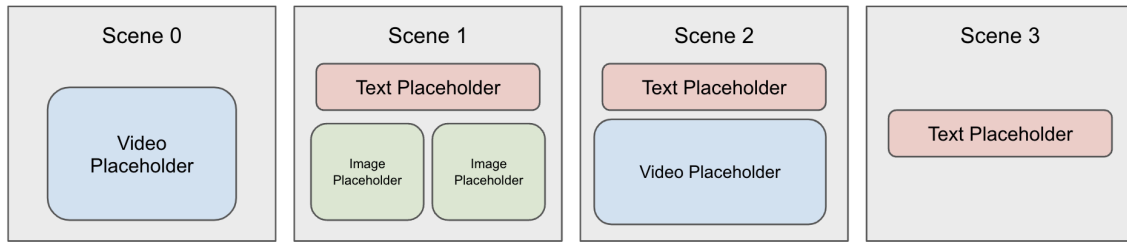


Figure 1: Example of a template with placeholders into which we can fit assets

Let's focus on a specific case where we have a template of N scenes containing a mixture of video and image placeholders. Our goal is to find video clips and images to fit into the placeholders in each scene of the template, together with text. We shall break the task down as follows:

1. Assuming the user has provided us with brand descriptions and theme/storyboard requirements, and a fixed template, use LLMs to generate advertisement "storyboard" i.e. a short paragraph describing the high-level theme of that scene. Each scene will have its own storyboard paragraph, and they should flow as a narrative. Example below:

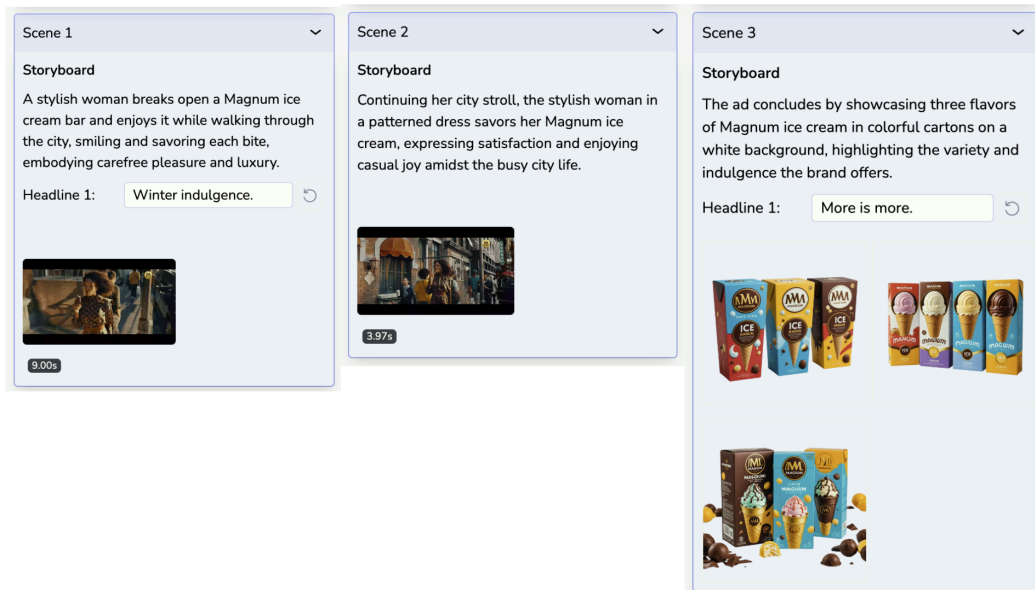


Figure 2: Example of a storyboard crafted given a template, and assets picked to **ideally** match the storyboard

2. Given the storyboards as guidance, we need to generate text slogans that are semantically relevant to the storyboard. In the example above, Scene #1 and Scene #3 have the text slogans "Winter Indulgence" and "More is More" (ideally, the storyboard and headlines should match - in the example above, they don't totally match yet)
3. We also need to pick (or generate) images and video clips that semantically match the storyboard of each scene, so that all elements within a scene appear congruent and consistent. Again referring to point #1, the images and videos there appears to match the storyboard presented.

For steps #1 and #2, propose a more detailed design solution (compared to our previous task) that can achieve this. Minimally, pseudocode or flow diagrams with discussions will do. Prompt examples or working demo would be great, though we're aware not all LLMs are practically accessible for this task. You may craft your own assumption on how user-supplied assets should be processed, if needed, to extract semantic knowledge (e.g. embeddings or otherwise) that is useful for your solution

C: Video Cutdown Generation

Given an input video, the traditional way of cutting it up into smaller pieces involves frame HSV changes or similar - this is fine for separating our disparate sets of scenes. However, in ads, we need to piece together these broken up pieces of video segments in order to fit them into a video ad of a fixed duration e.g. 20s:

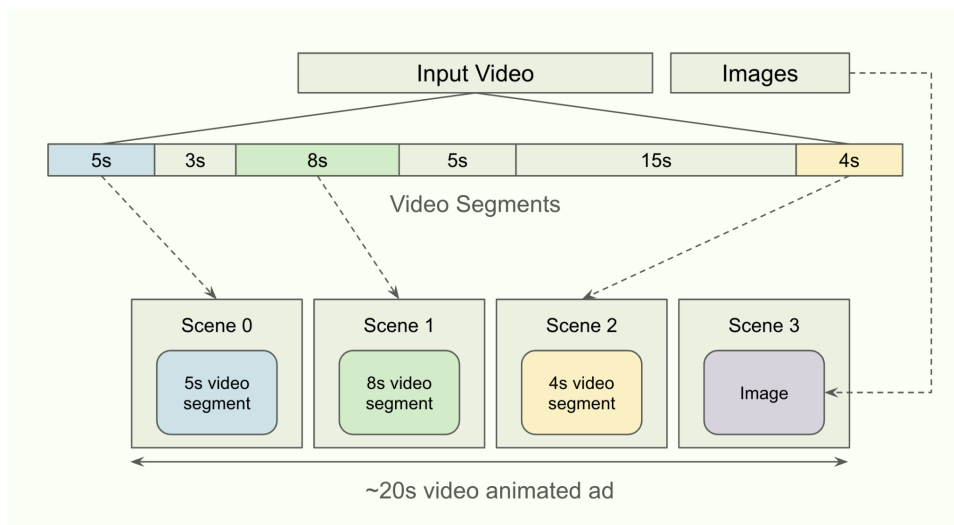


Figure 3: Video cutdown and fitting of segments into template placeholders

Propose and code up a simple solution that will allow you to find a video cutdown from a longer piece of video that will fit into an ad of a requested duration. E.g. given a 60s video, find segments of that video that will fit into 8-10s of ad. Break this task up into two steps:

1. In the first step, we need a method to split up a long video into shorter clip segments. You do not need to code this up. However, do write a simple set of classes to represent a video and its constituent clips:
 - a. Every video clip should have an original Video ID
 - b. Every video clip should have a start and end timestamp
 - c. Optionally, video clips may have a "scene change threshold" which describes how much the scene changes visually from the end of the current scene to the start of the next scene. You may choose to use or ignore such hypothetical information.
 - d. Optionally, video clips can have text descriptions, and/or contextual multimodal embeddings, seeing that these are easily obtainable.
2. In the second step, given these video segments, we need to pick and choose those that can sum up to a specific duration as required. Write a runnable algorithm, assuming you have inputs from Step 1, that can pick a subset of video clips that sums up to a required ad duration. E.g. if a user asks for a 15s cutdown from a 1-minute video clip, we may choose to pick 3 video segments of {3s, 4s, 6s} that sums up to ~13s. Note that we may not always be able to find cutdowns with precise total durations.

Are you able to improve on your solution based on one or more of the considerations below?


1. If we want user to be able to ask for cutdowns obeying semantics e.g. “give me a 10s cutdown of people walking in the city”
2. If we want to perform an $N \times M$ matching e.g. given $N=3$ scenes as below, each with a storyboard, pick the best 3 video clips from the set of M clips ($>N$) that can match the best 3 storyboards optimally - there can be one or more source videos::

Scene 1

Storyboard

A stylish woman breaks open a Magnum ice cream bar and enjoys it while walking through the city, smiling and savoring each bite, embodying carefree pleasure and luxury.

Headline 1: Winter indulgence.




9.00s

Scene 2

Storyboard

Continuing her city stroll, the stylish woman in a patterned dress savors her Magnum ice cream, expressing satisfaction and enjoying casual joy amidst the busy city life.



3.97s

Scene 3

Storyboard

The ad concludes by showcasing three flavors of Magnum ice cream in colorful cartons on a white background, highlighting the variety and indulgence the brand offers.

Headline 1: More is more.

