

AI Image Engineer Take-Home Assessment

You will have 96 hours to complete this assessment.

The better your image pipeline and consistency results, the higher the salary and position offer. When you are finished, submit your assessment by sharing your GitHub repo with **0xtaozi** and **0xmihutao**.

Context

At Nectar AI, our core product revolves around immersion. A user shouldn't just talk to an AI; they should "see" them. The biggest challenge in AI companionship today is **Character Consistency**—ensuring that the character looks the same in every photo, regardless of the pose, lighting, or outfit. This assessment tests your ability to build a high-fidelity image generation pipeline that feels like a real "memory" of a person.

Your Task

Build a mini-application that allows a user to chat with an AI character who can send high-quality, consistent photos of themselves on demand.

1. The Multi-Modal Chat

Build a simple interface (NextJS, Streamlit, or even a robust CLI) where a user can converse with an LLM-powered character.

- **Intent Detection:** The character should "decide" when it's appropriate to send a photo based on the conversation context (e.g., if the user asks "What are you wearing?" or "Where are you right now?").

2. The Image Pipeline (The Core Task)

Create a workflow that generates images of the character with the following requirements:

- **Character Consistency:** Use an identity anchor (e.g., **InstantID**, **IP-Adapter-FaceID**, or a trained **LoRA**). The character must look like the same person across 5+ different generated scenes. Ensure the highest quality and include 2-3 NSFW examples in your generated scenes. You must have a mix of both male and female.
- **Context Awareness:** If the chat happens in a "coffee shop," the generated image must reflect that setting accurately.
- **Zero-Shot Flexibility:** Your pipeline must be able to handle an arbitrary reference photo of a character it has never seen before and maintain that identity.
- **NSFW Support:** The pipeline must be uncensored and capable of generating realistic NSFW content without "mushy" anatomy or artifacts.

3. Infrastructure

- **Hosted GPU:** You may host your workflows on a remote GPU (e.g., **RunPod**, **Lambda Labs**, or **Modal**).
- **Optimization:** We are looking for workflows that balance speed and quality. Mention if you are using **TensorRT**, **FlashAttention**, or specialized nodes (like **ComfyUI-Manager** custom nodes).

What We're Evaluating

Area	What we're looking for
Identity Persistence	Does the character's "visual DNA" (bone structure, eye color, signature features) stay identical across different prompts?
Pipeline Sophistication	Are you just hitting an API, or have you built a custom ComfyUI or Diffusers workflow? We value those who understand ControlNets and Inpainting.
Prompt Engineering	Your ability to translate LLM "intent" into a high-quality Diffusion prompt (Subject + Style + Composition + Lighting).
Anatomical Correctness	High-quality "NSFW" is hard. We're looking for realistic skin textures, correct hand/limb counts, and coherent backgrounds.
Technical Choices	Why did you choose certain base image models ? Why this specific LoRA? Your README should justify your "Stack."

Requirements

- **NextJS** (or preferred FE) + **Python** (for the AI logic).
- **Live Demo:** A video recording of the chat and image generation in action.
- **Workflow Export/Source Code:** If using ComfyUI, include the **.json** or **.png** API-format workflow in your repo.

- **README.md:** Explain your approach to "Character Consistency." What were the biggest hurdles, and how did you solve "identity drift"?

Notes:

1. You can create an account on Runpod and we will reimburse up to \$45 for this assignment. You should pick the gpu that you think is best for this kind of use case in production, remember to factor in cost and speed. You will be required to submit a receipt (and also validate that it was used for training and assessment purposes only). Don't just leave the machine on overnight for no reason.
2. This is mainly for you to come up with, but we strongly suggest carrying out a normal conversation or roleplay conversation to ensure the image responses feel natural.
3. No hard limitation, but the Lower latency the better, but quality is more important