Arcade

## Take-Home Challenge

**Improve Stable Diffusion for Modern, Prompt-Accurate Jewelry Images**

---

### 1. Scenario

Arcade's customers rely on generative imagery to prototype jewelry ideas. Today we see two pain-points:

| # | Pain-point | Symptoms in Production |
|---|---|---|
| 1 | **Prompt adherence** | Specific product language ("channel-set diamond", "threader earrings") is often ignored or only partially represented, especially in long prompts. |
| 2 | **Aesthetic drift** | Outputs skew toward fussy, vintage, or "cheap-catalog" styles instead of the clean, contemporary look current shoppers expect (e.g., Mejuri, Catbird, Vrai). |

Your mission is to **download an open-weights Stable Diffusion checkpoint and demonstrate concrete improvements on BOTH dimensions**. A submission need not fully solve both problems to shine — we care how you think, measure, and iterate.

---

### 2. Resources you may use

- Any public Stable Diffusion 1.x/2.x/XL checkpoint
- Any datasets, text embeddings, or LoRA libraries
- Python ≥3.10, diffusers or ComfyUI, and any open-source ML tooling

---

### 3. Required Deliverables

| File / Artifact | What we look for |
|---|---|
| `before_after/` **folder** | 12 paired images (see §4 Prompt Suite) — baseline SD output vs. your improved method, saved as PNG/JPEG & named `prompt01_baseline.png`, `prompt01_yours.png`, etc. |
| `report.md` **(max 800 words)** | • *Approach*: what you changed (fine-tuning, LoRA, embeddings, post-filtering, prompt engineering, reranking, etc.) • *Rationale*: why these choices attack pain-points 1 & 2 • *Evidence*: 2–3 quantitative indicators (e.g., CLIP similarity, aesthetic scoring)• *Next steps*: what you'd try with an extra week |
| `notebook_or_scripts/` | Code that reproduces at least one improved image on CPU in ≤10 min (we won't rerun full training, but would like to see graphs!). |
| `README.md` | Quickstart: environment setup, how to generate your sample set from scratch. |

Zip the project or share a private repo, and submit within **one week** of receiving the prompt.

---

**4. Prompt Suite (use verbatim)**

1. *"channel-set diamond eternity band, 2 mm width, hammered 18k yellow gold, product-only white background"*
2. *"14k rose-gold threader earrings, bezel-set round lab diamond ends, lifestyle macro shot, soft natural light"*
3. *"organic cluster ring with mixed-cut sapphires and diamonds, brushed platinum finish, modern aesthetic"*
4. *"A solid gold cuff bracelet with blue sapphire, with refined simplicity and intentionally crafted for everyday wear"*
5. *"modern signet ring, oval face, engraved gothic initial 'M', high-polish sterling silver, subtle reflection"*
6. *"delicate gold huggie hoops, contemporary styling, isolated on neutral background"*
7. *"stack of three slim rings: twisted gold, plain platinum, black rhodium pavé, editorial lighting"*
8. *"bypass ring with stones on it, with refined simplicity and intentionally crafted for everyday wear"*

(Feel free to fine-tune on broader datasets, but grade yourself on these prompts.) Good luck — we're excited to see how you push generative models toward on-trend, prompt-faithful jewelry imagery!

**Evaluation Rubric (internal — for Arcade reviewers)**

| Criterion | Weight | Signals |
|---|---|---|
| **Prompt adherence↑** | 35 % | Quantitative (text–image CLIP similarity, term-matching) + visual check for all key terms. |
| **Modern aesthetic↑** | 25 % | Use of current-market style references; lower ornamentation; referential mood-board alignment. |
| **Rigor & Creativity** | 20 % | Choice of technique (LoRA vs. textual inversion vs. reranker, etc.), justification, and awareness of trade-offs. |
| **Reproducibility** | 10 % | Clear instructions; deterministic seed; runnable code snippets. |
| **Communication** | 10 % | Report clarity, concision, insight into future work. |