

## Technical Appendices

### A EXAMPLE OF THE ENHANCED ARTIFACT ATTRIBUTES

As a supplement to Section 3.1 and Table 1 in the main text, we provide a concrete example of the proposed expert attributes of the historical artifact “*yuhuchun vase in cobalt blue glaze*” in Table 5 of this appendix<sup>5</sup>. The last row is the arranged sequence of the artifact attributes separated by [SEP] (implemented as a Chinese comma), which forms the LLM-enhanced prompt input to our text-to-image models.

### B EXAMPLE OF THE PROMPT TEMPLATE FOR QUERYING GPT-3.5

As specified in Section 3.1 in the main text, we use *GPT-3.5-TURBO* as our knowledge-base LLM. The prompt template for querying GPT-3.5 is designed with a similar format following self-instruct [45]<sup>6</sup>. It consists of three parts: 1). A task statement that describes to GPT-3.5 the task to be done; 2). Two in-context examples sampled from our labeled pool of 54 artifacts written by archaeology experts; 3). The target artifacts whose “material”, “shape”, “pattern”, “type” and “type definition” are left blank and need to be answered by GPT-3.5. In between the parts and the different artifact samples, we use “###” as a separator. Figure 7 shows an example of our prompt for querying information about the artifact “*snuff bottle with intertwined floral decoration in fencai polychrome enamels on a yellow ground*”.

### C MORE ON IMPLEMENTATION DETAILS

Here we provide additional details with regard to the implementation of our method as a supplement to the model details specified in Section 4.1 of the main text: To get the canny edge map of images for the edge loss computation, we implement a canny filter [4] with a Gaussian kernel of size 3, Sobel filter kernel size of 3, a low threshold on pixel intensity of 0.15. We compute perceptual loss [15] using visual features extracted by the image encoder of *CLIP-ViT-L/14* [25]. The weights of additional losses used in our model training objective defined in (5) of the main text are  $\lambda_1 = 0.3$ ,  $\lambda_2 = 0.3$ , and  $\lambda_3 = 0.1$  (as can be seen in Table 2 in the main text). In addition, we employed the Min-SNR Weighting Strategy [11] to facilitate the training process, resulting in faster convergence using  $\gamma = 5.0$ . During training, we use an Adam [17] optimizer and set the batch size to 24 and learning rate to  $1.e^{-6}$  for all our experiments, which run on dual NVIDIA A100 GPUs.

### D MORE DETAILS ON EVALUATION METRICS

In this section, we offer a more extensive explanation of the automatic metrics (**CLIP Visual Similarity**, **Structural Similarity Index (SSIM)** [46] and **Learned Perceptual Image Patch Similarity (LPIPS)**) employed to quantitatively evaluate the performance of artifact image generation models. This serves to provide

additional technical details to the evaluation metrics stated in Section 4.1 in the main text.

**CLIP Visual Similarity.** Inspired by the CLIP Score [13], which is widely employed to assess the similarity between a text-image pair, we compute the visual similarity of two images (*i.e.*, the ground-truth image and the generated image) with the visual module of a pre-trained CLIP model (specifically, CLIP-ViT-L/14) [25]. We refer to this metric as Clip Visual Similarity, which is also utilized in [7, 19, 47]. Since the CLIP model has shown a strong ability in capturing the overall image contents and mapping those into a feature space, a higher similarity of the encoded visual features suggests a generally closer resemblance of the generated image to the ground truth.

**Structural Similarity Index (SSIM).** SSIM [46] is designed to primarily measure the similarity in terms of structural components between two images. It evaluates how well the synthesized output preserves the structural details present in the ground truth images. This includes important edges, boundaries, and overall structural coherence. By quantifying the degree of structural resemblance, SSIM provides valuable insights into the accuracy of artifact image synthesis in terms of preserving crucial details related to the formative appearance of artifacts, *e.g.*, their shape, and patterns.

**Learned Perceptual Image Patch Similarity (LPIPS).** LPIPS [53] judges the perceptual similarities between two images. In artifact image synthesis tasks, it is indispensable to assess not only the structural similarity but also the perceptual quality of the synthesized images. LPIPS is designed to align with human perception of image quality which also corresponds to the goal of artifact image synthesis tasks of generating historical objects that are visually convincing to human experts.

### E MORE EXAMPLES OF ARTIFACT IMAGES GENERATED BY OUR MODEL

Our model is capable of synthesizing images of a wide range of artifacts with high historical accuracy based on simple textual descriptions, as shown in Figure 6 in this appendix.

### F DIFFUSION MODELS AND STABLE DIFFUSION

In this appendix, we present a mathematical introduction to diffusion models and the Stable Diffusion architecture which serves as the backbone of our method (see Section 2 in the main text).

**Diffusion Models.** [14, 27, 35, 37, 38] are a family of probabilistic models which involve two processes: forward process and reverse process. Let  $p(x_0)$  be the image distribution and  $x_0 \sim p(x_0)$  be a real image in the distribution. The forward process  $q$  (also known as diffusion process) is a Markov chain of a length of a fixed timestep  $T$  that applies random noises from Gaussian distribution onto the previous state according to a variance schedule  $\beta_1, \beta_2, \dots, \beta_T$ , where

$$q(x_t|x_0) = \prod_{t=1}^T q(x_t|x_{t-1}) \quad (6)$$

<sup>5</sup>The numbering of tables, figures and equations in this Appendix follows that of the main text. So the number does not start fresh from 1 in this Appendix.

<sup>6</sup>The citations follow that of the main text. Please refer to the **References** in our main text to look up cited papers in this Appendix.

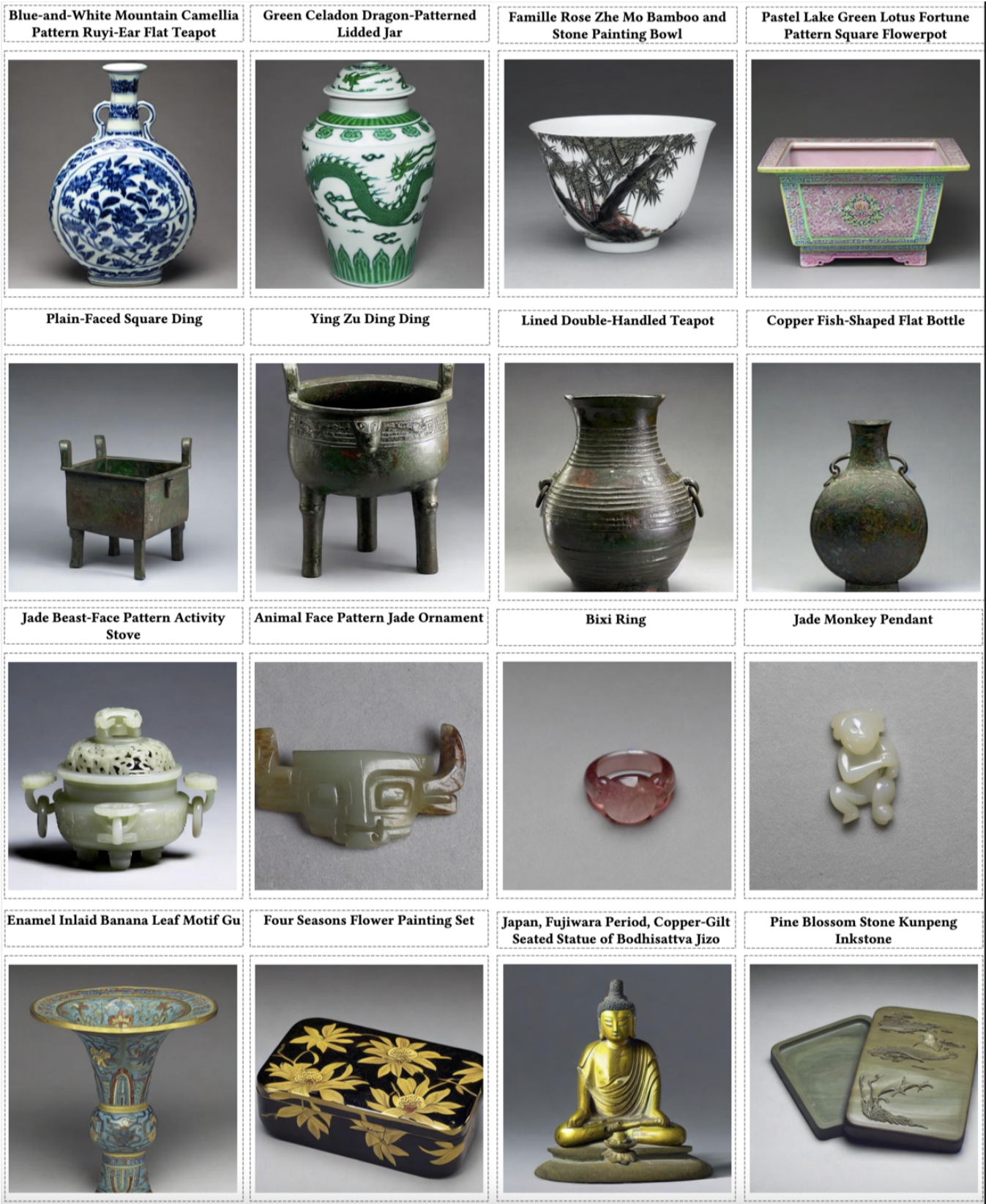


Figure 6: High-fidelity images of a wide range of artifacts generated by our model.

Expert Attribute	Example
<b>Name</b>	Yuhuchun vase in cobalt blue glaze
<b>Material</b>	Porcelain
<b>Time Period</b>	Qing Dynasty, Yongzheng reign, 1723-1735 AD
<b>Type</b>	Yuhuchun vase
<b>Type Definition</b>	Also known as "narrow-necked vase," yuhuchun vase is a practical commemorative ceramic widely popular in the northern regions. The vase consists of five parts: neck, shoulders, body, foot, and mouth. The neck is long and slender, the body is plump, and the foot can be a short circular footring or a horseshoe-shaped foot. Yuhuchun vases are created using various clay recipes and glaze techniques, resulting in distinct colors and surface effects for each piece
<b>Shape Pattern</b>	Flared mouth, slender neck, sloping shoulders, pear-shaped ample body, and a circular footring. The body of the vase is adorned with a cobalt blue glaze, which shines with a bright indigo color. The interior and the base of the vessel are covered in white glaze. The footring reveals the white body of the vase
<b>Size</b>	Height of 30.3 cm, mouth diameter of 8.5 cm, base diameter of 12.0 cm
<b>Enhanced Prompt Example</b>	Yuhuchun vase in cobalt blue glaze [SEP] Porcelain [SEP] Qing Dynasty, Yongzheng reign, 1723-1735 AD [SEP] Yuhuchun vase [SEP] Also known as "narrow-necked vase," yuhuchun vase is a practical commemorative ceramic widely popular in the northern regions. The vase consists of five parts: neck, shoulders, body, foot, and mouth. The neck is long and slender, the body is plump, and the foot can be a short circular footring or a horseshoe-shaped foot. Yuhuchun vases are created using various clay recipes and glaze techniques, resulting in distinct colors and surface effects for each piece [SEP] Flared mouth, slender neck, sloping shoulders, pear-shaped ample body, and a circular footring [SEP] The body of the vase is adorned with a cobalt blue glaze, which shines with a bright indigo color. The interior and the base of the vessel are covered in white glaze. The footring reveals the white body of the vase [SEP] Height of 30.3 cm, mouth diameter of 8.5 cm, base diameter of 12.0 cm

**Table 5: An example of the proposed expert attributes of artifacts. The last row forms the LLM-enhanced prompt input to our text-to-image model. The special delimiter [SEP] connecting attributes is implemented as a Chinese comma.**

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t|\sqrt{1-\beta_t}x_{t-1}; \beta_t \mathbf{I}) \quad (7)$$

The reverse process  $p_\theta$  is a Markov chain that aims to reverse the diffusion process by denoising the random noises at time step  $t$  where  $1 < t \leq T$  to eventually restore the real image  $x_0$ . The goal is to learn the parameter  $\theta$  for Gaussian transitions starting from  $p(x_T) = \mathcal{N}(x_T|0; \mathbf{I})$ , where

$$p_\theta(x_0) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1}|x_t) \quad (8)$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}|\mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (9)$$

The original diffusion model [14] set  $\Sigma_\theta(x_t, t) = \sigma_t^2 \mathbf{I}$  and  $\mu_\theta(x_t, t)$  to predict noise  $\epsilon_t$  at time step  $t$  with a noise predictor  $\epsilon_\theta$  parameterized by  $\theta$ , leading to the loss function

$$L(\theta) := \mathbb{E}_{t, x_0, \epsilon} \|\epsilon - \epsilon_\theta(\alpha_t x_0 + \sigma_t \epsilon, t)\|^2 \quad (10)$$

where  $\alpha_t = \sqrt{1 - \sigma_t^2}$  and  $\sigma_t^2 = \beta_t$  and  $\epsilon_\theta$  is a neural model using U-Net [29] as the backbone.

**Stable Diffusion (SD).** [27] introduces perceptual image compression which first maps the image to a latent space, then applies diffusion process and reverse process on the latent space rather than the pixel space which was used in earlier diffusion models. Overall, a Stable Diffusion model can be broken down into three parts: a

VAE [18], a text encoder and a U-Net [29]. VAE [18] contains two parts: an encoder  $\mathcal{E}$  and a decoder  $\mathcal{D}$ . In Stable Diffusion [27], the encoder part of VAE [18] is used to encode the image  $x$  into latent space  $\mathcal{Z}$  in the forward process during training. The decoder part of VAE [18] is used to decode the denoised latent representation into image at inference time. **Text Encoder**  $\mathcal{E}_{text}$  is responsible for mapping raw text into an embedding space that can be used to condition the backward denoising process. *i.e.*, For a raw text input  $S$ , the text encoder maps it to  $w$  such that  $w = \mathcal{E}_{text}(S)$ . Stable diffusion [27] uses a pre-trained CLIP [25] as the text encoder. **U-Net** [29] contains two parts: an encoder and a decoder, with ResNet [12] as the block structure. The encoder part projects the image into a low-resolution image presentation and the decoder part aims to restore the original image. Similar to the original diffusion model, in a Stable Diffusion model, the U-Net structure serves as  $\epsilon_\theta$  and aims to denoise the latent space at time step  $t$  where  $1 < t \leq T$  during the reverse process, conditioned on the text embedding using cross-attention [43]. The training objective for Stable Diffusion can be written as

$$L_{SD}(\theta) := \mathbb{E}_{t, \mathcal{E}(x_0), \epsilon} \|\epsilon - \epsilon_\theta(z_t, t, w)\|^2 \quad (11)$$

where  $z_t \in \mathcal{Z}$  is the representation of an image in the latent space at time step  $t$  and  $\mathcal{E}$  is the latent space encoder.  $w$  is the text representation encoded by  $\mathcal{E}_{text}$ .

现有3件文物，并有名称和描述对各件文物做出介绍。请根据文物的名称和描述，回答各件文物的材质、形态、纹饰、器型，并给出该种器型的器型基本外形定义。前2件是例子，第3件请按同样的格式输出。	There are three cultural relics, each with a name and a description. Please answer the material, shape, pattern, and type of each relic based on its name and description, and give the basic external shape definition of the type of vessel. The first two are examples, and the third one should be in the same format.
###	###
名称：青花云龙纹双耳扁壶	Name: Flask with handles and decoration of cloud and dragon in underglaze blue
描述：扁壶，小唇口，直颈，器腹呈扁圆形，浅圈足呈椭圆形，颈肩处饰以螭形双耳。口沿及圈足壁饰以青地白纹的海波纹样；颈饰变形蕉叶纹，肩部饰如意云头纹一周；瓶身腹部，正背两面皆饰圆形团龙云纹，外绕一圈留白。龙龙纹作正面龙形，四之四展，五爪特张；龙身鳞片勾勒细腻，云纹穿绕隙地。满布缠枝花卉纹，近足处为一圈变形莲瓣纹。底有「大清乾隆年制」六字款。	Description: Flat pot, small lip mouth, straight neck, flat oval belly, shallow round foot in an elliptical shape, and adorned with dragon-shaped double ears at the neck and shoulder. The mouth and foot are decorated with white wave patterns on a blue background; the neck is adorned with deformed banana leaf patterns, and the shoulder is adorned with ruyi cloud patterns all around; both the front and back of the belly are adorned with round dragon cloud patterns, surrounded by a circle of white space. The dragon pattern is in the form of a front-facing dragon, with four legs spread out and five claws open; the dragon's scales are delicately outlined, and the cloud pattern weaves through the gaps. The entire body is covered with entwined floral patterns, and near the foot is a circle of deformed lotus petal patterns. The bottom has a six-character inscription "Da Qing Qianlong Year Made".
材质：青花瓷	Material: Blue and white porcelain
形态：扁壶，小唇口，直颈，器腹呈扁圆形，浅圈足呈椭圆形，颈肩处饰以螭形双耳。	Shape: Flat pot, small lip mouth, straight neck, flat oval belly, shallow round foot in an elliptical shape, and adorned with dragon-shaped double ears at the neck and shoulder.
纹饰：口沿及圈足壁饰以青地白纹的海波纹样；颈饰变形蕉叶纹，肩部饰如意云头纹一周；瓶身腹部，正背两面皆饰圆形团龙云纹，外绕一圈留白。龙龙纹作正面龙形，四之四展，五爪特张；龙身鳞片勾勒细腻，云纹穿绕隙地。满布缠枝花卉纹，近足处为一圈变形莲瓣纹。	Pattern: Mouth and foot are decorated with white wave patterns on a blue background; the neck is adorned with deformed banana leaf patterns, and the shoulder is adorned with ruyi cloud patterns all around; both the front and back of the belly are adorned with round dragon cloud patterns, surrounded by a circle of white space.
器型：双耳扁壶	Type: Double-eared flat pot
器型基本外形定义：以扁壶而有双耳得名。小口，细颈，扁圆腹，颈肩处置双耳，椭圆形圈足或长方倭角圈足。	Basic external shape definition: Named for its flat pot with double ears. Small mouth, thin neck, flat round belly, double ears at the neck and shoulder, and elliptical or rectangular foot.
###	###
名称：茶叶末釉螭耳花浇	Name: Pitcher with a chi-dragon handles in tea-dust glaze
描述：花浇圆口，周缘切平，一侧向外突出形成短尖流。短颈，圆硕腹，口腹接一道边棱，侧置一螭形把，口下划一道，定出螭首衔贴位置。底内挖成卧足，底心浅印「雍正年制」两行四字篆款。全器罩施茶叶末釉，口部釉层下流，形成褐色边。釉面清楚可见橘皮棕眼。此器祖型为西亚玉器或金属器，十五世纪明朝永乐、宣德官窑临仿之烧制成青花瓷器。由于永乐作品较常出现双首螭纹柄的造型，故以为此品当是雍正仿永乐之作。回溯明朝，因有与西亚交流的背景，因此瓷花浇的烧制，可视为是反映史实的具体例证。相对于此，雍干两朝档案倒是出现有永宣花浇的纪录，遂让人从中推想今日将此类器皿称作花浇，当是沿袭清宫旧称。据此也能明白花浇加设出水流口的设计，或是为了进一步落实使用功能，而进行的改良。那么始见于康熙朝的茶叶末釉加施于外，也让此品兼具仿古器形和新釉彩的特色。	Description: The flower sprinkler has a round mouth, flat edge, and one side protrudes outward to form a short, pointed spout. Short neck, round and full belly, a rib connects the mouth and belly, and a dragon-shaped handle is placed on the side. The bottom is carved into a lying foot, and the shallow center is marked with "Yongzheng Year Made" in two rows of four-character seal script. The entire piece is covered with a tea-dust glaze, and the glaze layer flows down from the mouth, forming a brown edge. The glaze surface clearly shows orange-peel brown eyes. The prototype of this piece is a Western Asian jade or metal piece, and the Ming Dynasty official kiln in the 15th century made it into blue and white porcelain. As Yongle's works often feature double dragon pattern handles, it is believed that this piece is an imitation of Yongle during the Yongzheng period. Tracing back to the Ming Dynasty, due to the background of exchanges with Western Asia, the firing of porcelain flower sprinklers can be seen as a concrete example reflecting historical facts. In contrast, the Yong and Xuan archives have records of Yongxuan flower sprinklers, which led people to speculate that the name "flower sprinkler" today is inherited from the old name in the Qing Palace. It can also be understood that the design of adding a water outlet to the flower sprinkler is an improvement for further implementing the use function. Then, the tea-dust glaze applied from the Kangxi period onward gives this piece a combination of antique shape and new glaze color features.
材质：茶叶末釉瓷	Material: Tea-dust glazed porcelain
形态：花浇圆口，周缘切平，一侧向外突出形成短尖流。短颈，圆硕腹，口腹接一道边棱，侧置一螭形把，口下划一道，定出螭首衔贴位置。底内挖成卧足。	Shape: Round mouth flower sprinkler, flat edge, one side protrudes outward to form a short, pointed spout. Short neck, round and full belly, a rib connects the mouth and belly, and a dragon-shaped handle is placed on the side. The bottom is carved into a lying foot.
纹饰：全器罩施茶叶末釉，口部釉层下流，形成褐色边。釉面清楚可见橘皮棕眼。	Pattern: The entire piece is covered with a tea-dust glaze, and the glaze layer flows down from the mouth, forming a brown edge. The glaze surface clearly shows orange-peel brown eyes.
器型：花浇	Type: Flower sprinkler
器型基本外形定义：室内浇花用的瓷壶、瓷杯。罐形腹，口沿处有流，相对处有把柄。	Basic external shape definition: A porcelain pot or cup used for indoor flower watering. Jar-shaped belly, with a spout at the mouth and a handle on the opposite side.
###	###
名称：粉彩黄地缠枝花卉纹鼻烟壶	Name: Snuff bottle with intertwined floral decoration in fencai polychrome enamels on a yellow ground
描述：鼻烟壶，口外撇，短颈，扁圆腹下敛，上厚下薄，前后微鼓，平底，附红色盖，盖上突起小钮，下接木塞、牙匙，用来舀壶里的鼻烟，或盛在小碟上，或直接放在拇指背上，入鼻嗅用。全器满布黄地彩绘卷枝花纹，花心伸出桃实；盖沿、口沿及器缘均描金边。底外有一行四字篆款「嘉庆年制」。乾隆、嘉庆年间，常成套烧制釉上彩鼻烟壶，此器为一套十件的其中一件。	Description: Snuff bottle, outward-flaring mouth, short neck, flat round belly that tapers down, thick at the top and thin at the bottom, slightly bulging front and back, flat bottom, with a red lid, the lid has a small raised knob, connected to a wooden plug and ivory scoop, used to scoop snuff from the bottle, or put it on a small plate, or directly on the back of the thumb for sniffing. The entire piece is covered with yellow ground famille rose entwined floral patterns, with peach fruit extending from the flower center; the lid edge, mouth edge, and vessel edge are all outlined in gold. The bottom has a row of four-character seal script "Jiaqing Year Made". During the Qianlong and Jiaqing periods, sets of overglaze painted snuff bottles were often made, and this piece is one of a set of ten.
材质：	Material:
形态：	Shape:
纹饰：	Pattern:
器型：	Type:
器型基本外形定义：	Basic external shape definition:

Figure 7: An example of our GPT-3.5 querying prompt. We use Chinese by default because of the origin language of our data. The right-hand side is an English translation also done by the same *GPT-3.5-TURBO* engine.